

---

# **MMPose**

**发布 0.29.0**

**MMPose Authors**

**2022 年 10 月 24 日**



---

## Contents

---

<b>1</b>	<b>依赖环境</b>	<b>3</b>
<b>2</b>	<b>安装</b>	<b>5</b>
2.1	最佳实践 . . . . .	5
2.2	验证安装 . . . . .	6
2.3	自定义安装 . . . . .	7
2.4	故障解决 . . . . .	9
<b>3</b>	<b>基础教程</b>	<b>11</b>
3.1	准备数据集 . . . . .	12
3.2	使用预训练模型进行推理 . . . . .	12
3.3	如何训练模型 . . . . .	15
3.4	基准测试 . . . . .	18
3.5	进阶教程 . . . . .	18
<b>4</b>	<b>示例</b>	<b>19</b>
4.1	2D Animal Pose Demo . . . . .	19
4.2	2D Face Keypoint Demo . . . . .	23
4.3	2D Hand Keypoint Demo . . . . .	25
4.4	2D Human Pose Demo . . . . .	28
4.5	2D Pose Tracking Demo . . . . .	33
4.6	2D Human Whole-Body Pose Demo . . . . .	36
4.7	3D Mesh Demo . . . . .	39
4.8	3D Hand Demo . . . . .	40
4.9	3D Human Pose Demo . . . . .	41
4.10	3D Multiview Human Pose Demo . . . . .	44
4.11	Hand Gesture Recognition Demo . . . . .	45
4.12	Webcam Demo . . . . .	45

<b>5</b>	<b>基准测试</b>	<b>49</b>
<b>6</b>	<b>推理速度总结</b>	<b>51</b>
6.1	比较规则	51
6.2	MMPose 中主要模型的复杂度信息和推理速度总结	52
<b>7</b>	<b>概览</b>	<b>53</b>
7.1	2D 动物关键点数据集	53
7.2	2D 人体关键点数据集	53
7.3	2D 人脸关键点数据集	54
7.4	2D 服装关键点数据集	54
7.5	2D 手部关键点数据集	54
7.6	2D 全身人体关键点数据集	54
7.7	3D 人体关键点数据集	54
7.8	3D 人体网格模型数据集	54
7.9	3D 手部关键点数据集	55
<b>8</b>	<b>2D 人体关键点数据集</b>	<b>57</b>
8.1	COCO	58
8.2	MPII	59
8.3	MPII-TRB	60
8.4	AIC	60
8.5	CrowdPose	61
8.6	OCHuman	62
8.7	MHP	63
8.8	PoseTrack18	64
8.9	sub-JHMDB dataset	66
<b>9</b>	<b>2D 全身人体关键点数据集</b>	<b>67</b>
<b>10</b>	<b>2D 人脸关键点数据集</b>	<b>69</b>
<b>11</b>	<b>2D 手部关键点数据集</b>	<b>71</b>
<b>12</b>	<b>2D 服装关键点数据集</b>	<b>73</b>
<b>13</b>	<b>2D 动物关键点数据集</b>	<b>75</b>
<b>14</b>	<b>3D 人体关键点数据集</b>	<b>77</b>
<b>15</b>	<b>3D 人体网格模型数据集</b>	<b>79</b>
<b>16</b>	<b>3D 手部关键点数据集</b>	<b>81</b>
<b>17</b>	<b>概览</b>	<b>83</b>
17.1	Animal	83

17.2	Body(2D,Kpt,Sview,Img)	84
17.3	Body(2D,Kpt,Sview,Vid)	88
17.4	Body(3D,Kpt,Mview,Img)	89
17.5	Body(3D,Kpt,Sview,Img)	89
17.6	Body(3D,Kpt,Sview,Vid)	90
17.7	Body(3D,Mesh,Sview,Img)	90
17.8	Face	90
17.9	Fashion	92
17.10	Hand(2D,Kpt,Rgb,Img)	92
17.11	Hand(3D,Kpt,Rgb,Img)	94
17.12	Hand(Gesture,Rgb,Vid)	94
17.13	Wholebody	95
<b>18</b>	<b>Animal</b>	<b>97</b>
18.1	Animalpose Dataset	97
18.2	Ap10k Dataset	98
18.3	Atrw Dataset	99
18.4	Fly Dataset	100
18.5	Horse10 Dataset	101
18.6	Locust Dataset	102
18.7	Macaque Dataset	103
18.8	Zebra Dataset	104
<b>19</b>	<b>Body(2D,Kpt,Sview,Img)</b>	<b>105</b>
19.1	Aic Dataset	105
19.2	Coco Dataset	108
19.3	Crowdpose Dataset	135
19.4	H36m Dataset	138
19.5	JHMDB Dataset	139
19.6	MHP Dataset	140
19.7	Mpii Dataset	142
19.8	Mpii_trb Dataset	151
19.9	Ochuman Dataset	152
19.10	Posetrack18 Dataset	153
<b>20</b>	<b>Body(2D,Kpt,Sview,Vid)</b>	<b>157</b>
20.1	Posetrack18 Dataset	157
<b>21</b>	<b>Body(3D,Kpt,Sview,Img)</b>	<b>159</b>
21.1	H36m Dataset	159
21.2	Mpi_inf_3dhp Dataset	160
<b>22</b>	<b>Body(3D,Kpt,Sview,Vid)</b>	<b>161</b>

22.1	H36m Dataset . . . . .	161
22.2	Mpi_inf_3dhp Dataset . . . . .	162
<b>23</b>	<b>Body(3D,Kpt,Mview,Img)</b>	<b>165</b>
23.1	Campus Dataset . . . . .	165
23.2	Panoptic Dataset . . . . .	166
23.3	Shelf Dataset . . . . .	166
<b>24</b>	<b>Body(3D,Mesh,Sview,Img)</b>	<b>169</b>
24.1	Mixed Dataset . . . . .	169
<b>25</b>	<b>Face</b>	<b>171</b>
25.1	300w Dataset . . . . .	171
25.2	Aflw Dataset . . . . .	172
25.3	Coco_wholebody_face Dataset . . . . .	173
25.4	Cofw Dataset . . . . .	177
25.5	WFLW Dataset . . . . .	178
<b>26</b>	<b>Fashion</b>	<b>185</b>
26.1	Deepfashion Dataset . . . . .	185
<b>27</b>	<b>Hand(2D,Kpt,Rgb,Img)</b>	<b>189</b>
27.1	Coco_wholebody_hand Dataset . . . . .	189
27.2	Freihand2d Dataset . . . . .	193
27.3	Interhand2d Dataset . . . . .	194
27.4	Onehand10k Dataset . . . . .	195
27.5	Panoptic2d Dataset . . . . .	199
27.6	Rhd2d Dataset . . . . .	204
<b>28</b>	<b>Hand(3D,Kpt,Rgb,Img)</b>	<b>209</b>
28.1	Interhand3d Dataset . . . . .	209
<b>29</b>	<b>Hand(Gesture,Rgbd,Vid)</b>	<b>211</b>
29.1	Nvgesture Dataset . . . . .	211
<b>30</b>	<b>Wholebody</b>	<b>213</b>
30.1	Coco-Wholebody Dataset . . . . .	213
30.2	Halpe Dataset . . . . .	218
<b>31</b>	<b>Algorithms</b>	<b>221</b>
31.1	MTUT (CVPR' 2019) . . . . .	221
31.2	MSPN (ArXiv' 2019) . . . . .	222
31.3	InterNet (ECCV' 2020) . . . . .	223
31.4	DEKR (CVPR' 2021) . . . . .	223
31.5	HigherHRNet (CVPR' 2020) . . . . .	225

31.6	DeepPose (CVPR' 2014)	230
31.7	RLE (ICCV' 2021)	239
31.8	SoftWingloss (TIP' 2021)	241
31.9	VideoPose3D (CVPR' 2019)	243
31.10	Hourglass (ECCV' 2016)	244
31.11	LiteHRNet (CVPR' 2021)	247
31.12	AdaptiveWingloss (ICCV' 2019)	248
31.13	SimpleBaseline2D (ECCV' 2018)	249
31.14	PoseWarper (NeurIPS' 2019)	270
31.15	SimpleBaseline3D (ICCV' 2017)	271
31.16	HMR (CVPR' 2018)	273
31.17	UDP (CVPR' 2020)	274
31.18	ViPNAS (CVPR' 2021)	279
31.19	Wingloss (CVPR' 2018)	281
31.20	DarkPose (CVPR' 2020)	282
31.21	Associative Embedding (NIPS' 2017)	292
31.22	VoxelPose (ECCV' 2020)	303
31.23	RSN (ECCV' 2020)	304
31.24	CID (CVPR' 2022)	305
31.25	CPM (CVPR' 2016)	306
31.26	HRNet (CVPR' 2019)	308
31.27	HRNetv2 (TPAMI' 2019)	327
31.28	SCNet (CVPR' 2020)	341
<b>32</b>	<b>Backbones</b>	<b>345</b>
32.1	MSPN (ArXiv' 2019)	345
32.2	MobilenetV2 (CVPR' 2018)	346
32.3	HigherHRNet (CVPR' 2020)	351
32.4	ResNeSt (ArXiv' 2020)	355
32.5	ResNext (CVPR' 2017)	356
32.6	ResNet (CVPR' 2016)	357
32.7	Hourglass (ECCV' 2016)	384
32.8	ShufflenetV1 (CVPR' 2018)	386
32.9	Swin (ICCV' 2021)	388
32.10	LiteHRNet (CVPR' 2021)	389
32.11	I3D (CVPR' 2017)	390
32.12	ShufflenetV2 (ECCV' 2018)	391
32.13	TCFormer (CVPR' 2022)	392
32.14	ResNetV1D (CVPR' 2019)	393
32.15	SEResNet (CVPR' 2018)	394
32.16	ViPNAS (CVPR' 2021)	395
32.17	PVTv2 (CVMIJ' 2022)	397

32.18 PVT (ICCV' 2021) . . . . .	398
32.19 RSN (ECCV' 2020) . . . . .	399
32.20 CPM (CVPR' 2016) . . . . .	399
32.21 HRNet (CVPR' 2019) . . . . .	401
32.22 HRNetv2 (TPAMI' 2019) . . . . .	421
32.23 HRFormer (NIPS' 2021) . . . . .	435
32.24 AlexNet (NeurIPS' 2012) . . . . .	436
32.25 VGG (ICLR' 2015) . . . . .	436
32.26 SCNet (CVPR' 2020) . . . . .	437
<b>33 Datasets</b>	<b>441</b>
33.1 InterHand2.6M (ECCV' 2020) . . . . .	441
33.2 MPII (CVPR' 2014) . . . . .	443
33.3 NVGesture (CVPR' 2016) . . . . .	452
33.4 MacaquePose (bioRxiv' 2020) . . . . .	453
33.5 ATRW (ACM MM' 2020) . . . . .	454
33.6 Animal-Pose (ICCV' 2019) . . . . .	455
33.7 Human3.6M (TPAMI' 2014) . . . . .	456
33.8 OneHand10K (TCSVT' 2019) . . . . .	460
33.9 JHMDB (ICCV' 2013) . . . . .	464
33.10 CMU Panoptic (ICCV' 2015) . . . . .	466
33.11 CMU Panoptic HandDB (CVPR' 2017) . . . . .	466
33.12 AFLW (ICCVW' 2011) . . . . .	470
33.13 300W (IMAVIS' 2016) . . . . .	472
33.14 FreiHand (ICCV' 2019) . . . . .	473
33.15 MHP (ACM MM' 2018) . . . . .	473
33.16 Vinegar Fly (Nature Methods' 2019) . . . . .	475
33.17 CrowdPose (CVPR' 2019) . . . . .	476
33.18 Halpe (CVPR' 2020) . . . . .	479
33.19 OCHuman (CVPR' 2019) . . . . .	479
33.20 AP-10K (NeurIPS' 2021) . . . . .	481
33.21 Desert Locust (Elife' 2019) . . . . .	482
33.22 COCO-WholeBody-Face (ECCV' 2020) . . . . .	483
33.23 MPI-INF-3DHP (3DV' 2017) . . . . .	487
33.24 AI Challenger (ArXiv' 2017) . . . . .	488
33.25 MPII-TRB (ICCV' 2019) . . . . .	491
33.26 WFLW (CVPR' 2018) . . . . .	492
33.27 PoseTrack18 (CVPR' 2018) . . . . .	497
33.28 COCO-WholeBody-Hand (ECCV' 2020) . . . . .	500
33.29 COFW (ICCV' 2013) . . . . .	504
33.30 DeepFashion (CVPR' 2016) . . . . .	505
33.31 COCO-WholeBody (ECCV' 2020) . . . . .	507



33.32 Horse-10 (WACV' 2021) . . . . .	512
33.33 Grévy' s Zebra (Elife' 2019) . . . . .	513
33.34 COCO (ECCV' 2014) . . . . .	514
33.35 RHD (ICCV' 2017) . . . . .	542
<b>34 Techniques</b>	<b>547</b>
34.1 SoftWingloss (TIP' 2021) . . . . .	547
34.2 FP16 (ArXiv' 2017) . . . . .	548
34.3 Albumentations (Information' 2020) . . . . .	550
34.4 AdaptiveWingloss (ICCV' 2019) . . . . .	551
34.5 FPN (CVPR' 2017) . . . . .	552
34.6 UDP (CVPR' 2020) . . . . .	553
34.7 Wingloss (CVPR' 2018) . . . . .	558
34.8 DarkPose (CVPR' 2020) . . . . .	559
<b>35 教程 0: 模型配置文件</b>	<b>571</b>
35.1 通过脚本参数修改配置 . . . . .	571
35.2 配置文件命名约定 . . . . .	572
35.3 常见问题 . . . . .	577
<b>36 教程 1: 如何微调模型</b>	<b>579</b>
36.1 概要 . . . . .	579
36.2 修改网络头 . . . . .	580
36.3 支持自己的数据集 . . . . .	581
36.4 修改训练策略 . . . . .	582
36.5 使用预训练模型 . . . . .	583
<b>37 教程 2: 增加新的数据集</b>	<b>585</b>
37.1 将数据集转化为 COCO 格式 . . . . .	585
37.2 为自定义数据集创建 dataset_info 数据集配置文件 . . . . .	586
37.3 创建自定义数据集类 . . . . .	591
37.4 创建和修改训练配置文件 . . . . .	591
<b>38 教程 3: 自定义数据前处理流水线</b>	<b>593</b>
38.1 设计数据前处理流水线 . . . . .	593
38.2 扩展和使用自定义流水线 . . . . .	595
<b>39 教程 4: 增加新的模块</b>	<b>597</b>
39.1 自定义优化器 . . . . .	597
39.2 自定义优化器构造器 . . . . .	598
39.3 开发新组件 . . . . .	599
<b>40 教程 5: 如何导出模型为 onnx 格式</b>	<b>603</b>
40.1 支持的模型 . . . . .	603

40.2	如何使用	604
<b>41</b>	<b>教程 6: 自定义运行配置</b>	<b>605</b>
41.1	自定义优化方法	606
41.2	自定义训练策略	609
41.3	自定义工作流	609
41.4	自定义钩子	610
<b>42</b>	<b>教程 7: 摄像头应用接口</b>	<b>615</b>
42.1	摄像头应用接口概览	616
42.2	摄像头应用示例	616
42.3	通过定义新节点扩展功能	618
<b>43</b>	<b>常用工具</b>	<b>623</b>
<b>44</b>	<b>常见问题</b>	<b>625</b>
44.1	安装	625
44.2	开发	626
44.3	数据	627
44.4	训练	628
44.5	评测	628
44.6	推理	628
44.7	部署	629
<b>45</b>	<b>mmpose.apis</b>	<b>631</b>
<b>46</b>	<b>mmpose.apis.webcam</b>	<b>641</b>
46.1	Executor	642
46.2	Nodes	643
46.3	Utils	659
<b>47</b>	<b>mmpose.core</b>	<b>669</b>
47.1	evaluation	669
47.2	fp16	680
47.3	utils	683
47.4	post_processing	684
<b>48</b>	<b>mmpose.models</b>	<b>691</b>
48.1	backbones	691
48.2	necks	728
48.3	detectors	732
48.4	heads	767
48.5	losses	791
48.6	misc	805

<b>49 mmpose.datasets</b>	<b>807</b>
49.1 datasets . . . . .	869
49.2 pipelines . . . . .	891
49.3 samplers . . . . .	907
<b>50 mmpose.utils</b>	<b>909</b>
<b>51 Indices and tables</b>	<b>911</b>
<b>Python 模块索引</b>	<b>913</b>
<b>索引</b>	<b>915</b>



您可以在页面左下角切换文档语言。

You can change the documentation language at the lower-left corner of the page.

- 依赖环境
- 安装
  - 最佳实践
    - \* 从源码安装
    - \* 作为 *Python* 包安装
  - 验证安装
  - 自定义安装
    - \* *CUDA* 版本
    - \* 不使用 *MIM* 安装 *MMCV*
    - \* 在 *CPU* 环境中安装
    - \* 在 *Google Colab* 中安装
    - \* 通过 *Docker* 使用 *MMPose*
  - 故障解决



# CHAPTER 1

## 依赖环境

在本节中，我们将演示如何准备 PyTorch 相关的依赖环境。

MMPose 适用于 Linux、Windows 和 macOS。它需要 Python 3.6+、CUDA 9.2+ 和 PyTorch 1.5+。

**注解：**如果您对配置 PyTorch 环境已经很熟悉，并且已经完成了配置，可以直接进入[下一节](#)。否则，请依照以下步骤完成配置。

**第 1 步**从[官网](#) 下载并安装 Miniconda。

**第 2 步**创建一个 conda 虚拟环境并激活它。

```
conda create --name openmmlab python=3.8 -y
conda activate openmmlab
```

**第 3 步**按照[官方指南](#) 安装 PyTorch。例如：

在 GPU 平台：

```
conda install pytorch torchvision -c pytorch
```

**警告：** 以上命令会自动安装最新版的 PyTorch 与对应的 cudatoolkit，请检查它们是否与您的环境匹配。

在 CPU 平台：

```
conda install pytorch torchvision cpuonly -c pytorch
```



我们推荐用户按照我们的最佳实践来安装 MMPose。但除此之外，如果您想根据您的习惯完成安装流程，也可以参见[自定义安装](#)一节来获取更多信息。

## 2.1 最佳实践

### 第 1 步使用 MIM 安装 MMCV

```
pip install -U openmim
mim install mmcv-full
```

### 第 2 步安装 MMPose

根据具体需求，我们支持两种安装模式：

- [从源码安装（推荐）](#)：如果基于 MMPose 框架开发自己的任务，需要添加新的功能，比如新的模型或是数据集，或者使用我们提供的各种工具。
- [作为 Python 包安装](#)：只是希望调用 MMPose 的接口，或者在自己的项目中导入 MMPose 中的模块。

## 2.1.1 从源码安装

这种情况下，从源码按如下方式安装 mmpose：

```
git clone https://github.com/open-mmlab/mmpose.git
cd mmpose
pip install -r requirements.txt
pip install -v -e .
# "-v" 表示输出更多安装相关的信息
# "-e" 表示以可编辑形式安装，这样可以在不重新安装的情况下，让本地修改直接生效
```

## 2.1.2 作为 Python 包安装

直接使用 pip 安装即可。

```
pip install mmpose
```

## 2.2 验证安装

为了验证 MMPose 的安装是否正确，我们提供了一些示例代码来执行模型推理。

**第 1 步**我们需要下载配置文件和模型权重文件

```
mim download mmpose --config associative_embedding_hrnet_w32_coco_512x512 --dest .
```

下载过程往往需要几秒或更多的时间，这取决于您的网络环境。完成之后，您会在当前目录下找到这两个文件：associative\_embedding\_hrnet\_w32\_coco\_512x512.py, hrnet\_w32\_coco\_512x512-bcb8c247\_20200816.pth, 分别是配置文件和对应的模型权重文件。

**第 2 步**验证推理示例

如果您是从源码安装的 mmpose，那么直接运行以下命令进行验证：

```
python demo/bottom_up_img_demo.py associative_embedding_hrnet_w32_coco_512x512.py
↪hrnet_w32_coco_512x512-bcb8c247_20200816.pth --img-path tests/data/coco/ --out-img-
↪root vis_results
```

您可以在 vis\_results 这个目录下看到输出的图片，这些图片展示了人体姿态估计的结果。

如果您是作为 Python 包安装，那么可以打开您的 Python 解释器，复制并粘贴如下代码：

```
from mmpose.apis import (init_pose_model, inference_bottom_up_pose_model, vis_pose_
↪result)
```

(下页继续)

(续上页)

```

config_file = 'associative_embedding_hrnet_w32_coco_512x512.py'
checkpoint_file = 'hrnet_w32_coco_512x512-bcb8c247_20200816.pth'
pose_model = init_pose_model(config_file, checkpoint_file, device='cpu') # or device=
→ 'cuda:0'

image_name = 'demo/persons.jpg'
# test a single image
pose_results, _ = inference_bottom_up_pose_model(pose_model, image_name)

# show the results
vis_pose_result(pose_model, image_name, pose_results, out_file='demo/vis_persons.jpg')

```

准备好一张带有人的图片，并放置在合适的位置，然后运行以上代码，您将会在输出的图片上看到检测到的人体姿态结果。

## 2.3 自定义安装

### 2.3.1 CUDA 版本

安装 PyTorch 时，需要指定 CUDA 版本。如果您不清楚选择哪个，请遵循我们的建议：

- 对于 Ampere 架构的 NVIDIA GPU，例如 GeForce 30 系列以及 NVIDIA A100，CUDA 11 是必需的。
- 对于更早的 NVIDIA GPU，CUDA 11 是向后兼容 (backward compatible) 的，但 CUDA 10.2 能够提供更好的兼容性，也更加轻量。

请确保您的 GPU 驱动版本满足最低的版本需求，参阅[这张表](#)。

---

**注解：**如果按照我们的最佳实践进行安装，CUDA 运行时库就足够了，因为我们提供相关 CUDA 代码的预编译，您不需要进行本地编译。但如果您希望从源码进行 MMCV 的编译，或是进行其他 CUDA 算子的开发，那么就必須安装完整的 CUDA 工具链，参见 [NVIDIA 官网](#)，另外还需要确保该 CUDA 工具链的版本与 PyTorch 安装时的配置相匹配（如用 `conda install` 安装 PyTorch 时指定的 `cuda-toolkit` 版本）。

---

### 2.3.2 不使用 MIM 安装 MMCV

MMCV 包含 C++ 和 CUDA 扩展，因此其对 PyTorch 的依赖比较复杂。MIM 会自动解析这些依赖，选择合适的 MMCV 预编译包，使安装更简单，但它并不是必需的。

要使用 `pip` 而不是 MIM 来安装 MMCV，请遵照 [MMCV 安装指南](#)。它需要您用指定 `url` 的形式手动指定对应的 PyTorch 和 CUDA 版本。

举个例子，如下命令将会安装基于 PyTorch 1.10.x 和 CUDA 11.3 编译的 `mmcv-full`。

```
pip install mmdcv-full -f https://download.openmmlab.com/mmdcv/dist/cu113/torch1.10/
↪ index.html
```

### 2.3.3 在 CPU 环境中安装

MMPose 可以仅在 CPU 环境中安装，在 CPU 模式下，您可以完成训练（需要 MMCV 版本  $\geq 1.4.4$ ）、测试和模型推理等所有操作。

在 CPU 模式下，MMCV 的部分功能将不可用，通常是一些 GPU 编译的算子，如 Deformable Convolution。MMPose 中大部分的模型都不会依赖这些算子，但是如果您尝试使用包含这些算子的模型来运行训练、测试或推理，将会报错。

### 2.3.4 在 Google Colab 中安装

Google Colab 通常已经包含了 PyTorch 环境，因此我们只需要安装 MMCV 和 MMPose 即可，命令如下：

#### 第 1 步使用 MIM 安装 MMCV

```
!pip3 install openmim
!mim install mmdcv-full
```

#### 第 2 步从源码安装 mmpose

```
!git clone https://github.com/open-mmlab/mmpose.git
%cd mmpose
!pip install -e .
```

#### 第 3 步验证

```
import mmpose
print(mmpose.__version__)
# 预期输出: 0.26.0 或其他版本号
```

---

**注解：**在 Jupyter 中，感叹号 `!` 用于执行外部命令，而 `%cd` 是一个魔术命令，用于切换 Python 的工作路径。

---

### 2.3.5 通过 Docker 使用 MMPose

MMPose 提供 Dockerfile 用于构建镜像。请确保您的 Docker 版本  $\geq 19.03$ 。

```
# 构建默认的 PyTorch 1.6.0, CUDA 10.1 版本镜像
# 如果您希望使用其他版本, 请修改 Dockerfile
docker build -t mmpose docker/
```

**注意：**请确保您已经安装了 `nvidia-container-toolkit`。

用以下命令运行 Docker 镜像：

```
docker run --gpus all --shm-size=8g -it -v {DATA_DIR}:/mmpose/data mmpose
```

{DATA\_DIR} 是您本地存放用于 MMPose 训练、测试、推理等流程的数据目录。

## 2.4 故障解决

如果您在安装过程中遇到了什么问题，请先查阅[常见问题](#)。如果没有找到解决方法，可以在 [GitHub](#) 上提出 [issue](#)。



本文档提供 MMPose 的基础使用教程。请先参阅[安装指南](#)，进行 MMPose 的安装。

- 准备数据集
- 使用预训练模型进行推理
  - 测试某个数据集
  - 运行演示
- 如何训练模型
  - 训练配置
  - 使用单个 GPU 训练
  - 使用 CPU 训练
  - 使用多个 GPU 训练
  - 使用多台机器训练
  - 使用单台机器启动多个任务
- 基准测试
- 进阶教程

## 3.1 准备数据集

MMPose 支持各种不同的任务。请根据需要，查阅对应的数据集准备教程。

- 2D 人体关键点检测
- 3D 人体关键点检测
- 3D 人体形状恢复
- 2D 人手关键点检测
- 3D 人手关键点检测
- 2D 人脸关键点检测
- 2D 全身人体关键点检测
- 2D 服饰关键点检测
- 2D 动物关键点检测

## 3.2 使用预训练模型进行推理

MMPose 提供了一些测试脚本用于测试数据集上的指标（如 COCO, MAPII 等），并提供了一些高级 API，使您可以轻松使用 MMPose。

### 3.2.1 测试某个数据集

- [x] 单 GPU 测试
- [x] CPU 测试
- [x] 单节点多 GPU 测试
- [x] 多节点测试

用户可使用以下命令测试数据集

```
# 单 GPU 测试
python tools/test.py ${CONFIG_FILE} ${CHECKPOINT_FILE} [--out ${RESULT_FILE}] [--fuse-
↪conv-bn] \
    [--eval ${EVAL_METRICS}] [--gpu_collect] [--tmpdir ${TMPDIR}] [--cfg-options $
↪{CFG_OPTIONS}] \
    [--launcher ${JOB_LAUNCHER}] [--local_rank ${LOCAL_RANK}]

# CPU 测试：禁用 GPU 并运行测试脚本
export CUDA_VISIBLE_DEVICES=-1
```

(下页继续)



(续上页)

```
python tools/test.py ${CONFIG_FILE} ${CHECKPOINT_FILE} [--out ${RESULT_FILE}] \
    [--eval ${EVAL_METRICS}]

# 多 GPU 测试
./tools/dist_test.sh ${CONFIG_FILE} ${CHECKPOINT_FILE} ${GPU_NUM} [--out ${RESULT_
→FILE}] [--eval ${EVAL_METRICS}] \
    [--gpu-collect] [--tmpdir ${TMPDIR}] [--options ${OPTIONS}] [--average-clips $
→{AVG_TYPE}] \
    [--launcher ${JOB_LAUNCHER}] [--local_rank ${LOCAL_RANK}]
```

此处的 CHECKPOINT\_FILE 可以是本地的模型权重文件的路径，也可以是模型的下载链接。

可选参数:

- RESULT\_FILE: 输出结果文件名。如果没有被指定，则不会保存测试结果。
- --fuse-conv-bn: 是否融合 BN 和 Conv 层。该操作会略微提升模型推理速度。
- EVAL\_METRICS: 测试指标。其可选值与对应数据集相关，如 mAP，适用于 COCO 等数据集，PCK AUC EPE 适用于 OneHand10K 等数据集等。
- --gpu-collect: 如果被指定，姿态估计结果将会通过 GPU 通信进行收集。否则，它将被存储到不同 GPU 上的 TMPDIR 文件夹中，并在 rank 0 的进程中被收集。
- TMPDIR: 用于存储不同进程收集的结果文件的临时文件夹。该变量仅当 --gpu-collect 没有被指定时有效。
- CFG\_OPTIONS: 覆盖配置文件中的一些实验设置。比如，可以设置 'cfg-options model.backbone.depth=18 model.backbone.with\_cp=True'，在线修改配置文件内容。
- JOB\_LAUNCHER: 分布式任务初始化启动器选项。可选值有 none, pytorch, slurm, mpi。特别地，如果被设置为 none，则会以非分布式模式进行测试。
- LOCAL\_RANK: 本地 rank 的 ID。如果没有被指定，则会被设置为 0。

例子:

假定用户将下载的模型权重文件放置在 checkpoints/ 目录下。

1. 在 COCO 数据集下测试 ResNet50（不存储测试结果为文件），并验证 mAP 指标

```
./tools/dist_test.sh configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/res50_
→coco_256x192.py \
    checkpoints/SOME_CHECKPOINT.pth 1 \
    --eval mAP
```

2. 使用 8 块 GPU 在 COCO 数据集下测试 ResNet。在线下载模型权重，并验证 mAP 指标。

```
./tools/dist_test.sh configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/res50_
↪coco_256x192.py \
    https://download.openmmlab.com/mmpose/top_down/resnet/res50_coco_256x192-
↪ec54d7f3_20200709.pth 8 \
    --eval mAP
```

### 3. 在 slurm 分布式环境中测试 ResNet50 在 COCO 数据集下的 mAP 指标

```
./tools/slurm_test.sh slurm_partition test_job \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/res50_coco_256x192.py \
    checkpoints/SOME_CHECKPOINT.pth \
    --eval mAP
```

## 3.2.2 运行演示

我们提供了丰富的脚本，方便大家快速运行演示。下面是多人人体姿态估计的演示示例，此处我们使用了人工标注的人体框作为输入。

```
python demo/top_down_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --json-file ${JSON_FILE} \
    --out-img-root ${OUTPUT_DIR} \
    [--show --device ${GPU_ID}] \
    [--kpt-thr ${KPT_SCORE_THR}]
```

例子：

```
python demo/top_down_img_demo.py \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/hrnet_w48_coco_256x192.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_256x192-
↪b9e0b3ab_20200708.pth \
    --img-root tests/data/coco/ --json-file tests/data/coco/test_coco.json \
    --out-img-root vis_results
```

更多实例和细节可以查看 [demo 文件夹](#) 和 [demo 文档](#)。

## 3.3 如何训练模型

MMPose 使用 `MMDistributedDataParallel` 进行分布式训练, 使用 `MMDDataParallel` 进行非分布式训练。

对于单机多卡与多台机器的情况, MMPose 使用分布式训练。假设服务器有 8 块 GPU, 则会启动 8 个进程, 并且每台 GPU 对应一个进程。

每个进程拥有一个独立的模型, 以及对应的数据加载器和优化器。模型参数同步只发生于最开始。之后, 每经过一次前向与后向计算, 所有 GPU 中梯度都执行一次 `allreduce` 操作, 而后优化器将更新模型参数。由于梯度执行了 `allreduce` 操作, 因此不同 GPU 中模型参数将保持一致。

### 3.3.1 训练配置

所有的输出 (日志文件和模型权重文件) 会被将保存到工作目录下。工作目录通过配置文件中的参数 `work_dir` 指定。

默认情况下, MMPose 在每轮训练轮后会在验证集上评估模型, 可以通过在训练配置中修改 `interval` 参数来更改评估间隔

```
evaluation = dict(interval=5) # 每 5 轮训练进行一次模型评估
```

根据 [Linear Scaling Rule](#), 当 GPU 数量或每个 GPU 上的视频批大小改变时, 用户可根据批大小按比例地调整学习率, 如, 当 4 GPUs x 2 video/gpu 时, `lr=0.01`; 当 16 GPUs x 4 video/gpu 时, `lr=0.08`。

### 3.3.2 使用单个 GPU 训练

```
python tools/train.py ${CONFIG_FILE} [optional arguments]
```

如果用户想在命令中指定工作目录, 则需要增加参数 `--work-dir ${YOUR_WORK_DIR}`

### 3.3.3 使用 CPU 训练

使用 CPU 训练的流程和使用单 GPU 训练的流程一致, 我们仅需要在训练流程开始前禁用 GPU。

```
export CUDA_VISIBLE_DEVICES=-1
```

之后运行单 GPU 训练脚本即可。

注意:

我们不推荐用户使用 CPU 进行训练, 这太过缓慢。我们支持这个功能是为了方便用户在没有 GPU 的机器上进行调试。

### 3.3.4 使用多个 GPU 训练

```
./tools/dist_train.sh ${CONFIG_FILE} ${GPU_NUM} [optional arguments]
```

可选参数为：

- `--work-dir ${WORK_DIR}`: 覆盖配置文件中指定的工作目录。
- `--resume-from ${CHECKPOINT_FILE}`: 从以前的模型权重文件恢复训练。
- `--no-validate`: 在训练过程中, 不进行验证。
- `--gpus ${GPU_NUM}`: 使用的 GPU 数量, 仅适用于非分布式训练。
- `--gpu-ids ${GPU_IDS}`: 使用的 GPU ID, 仅适用于非分布式训练。
- `--seed ${SEED}`: 设置 python, numpy 和 pytorch 里的种子 ID, 已用于生成随机数。
- `--deterministic`: 如果被指定, 程序将设置 CUDNN 后端的确定化选项。
- `--cfg-options CFG_OPTIONS`: 覆盖配置文件中的某些实验设置。比如, 可以设置 `'--cfg-options model.backbone.depth=18 model.backbone.with_cp=True'`, 在线修改配置文件内容。
- `--launcher ${JOB_LAUNCHER}`: 分布式任务初始化启动器选项。可选值有 `none`, `pytorch`, `slurm`, `mpi`。特别地, 如果被设置为 `none`, 则会以非分布式模式进行测试。
- `--autoscale-lr`: 根据 [Linear Scaling Rule](#), 当 GPU 数量或每个 GPU 上的视频批大小改变时, 用户可根据批大小按比例地调整学习率。
- `LOCAL_RANK`: 本地 rank 的 ID。如果没有被指定, 则会被设置为 0。

`resume-from` 和 `load-from` 的区别: `resume-from` 加载模型参数和优化器状态, 并且保留检查点所在的训练轮数, 常被用于恢复意外被中断的训练。`load-from` 只加载模型参数, 但训练轮数从 0 开始计数, 常被用于微调模型。

这里提供一个使用 8 块 GPU 加载 ResNet50 模型权重文件的例子。

```
./tools/dist_train.sh configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/res50_
→coco_256x192.py 8 --resume_from work_dirs/res50_coco_256x192/latest.pth
```

### 3.3.5 使用多台机器训练

如果用户在 `slurm` 集群上运行 MMPose, 可使用 `slurm_train.sh` 脚本。(该脚本也支持单台机器上训练)

```
[GPUS=${GPUS}] ./tools/slurm_train.sh ${PARTITION} ${JOB_NAME} ${CONFIG_FILE} [--work-
→dir ${WORK_DIR}]
```

这里给出一个在 `slurm` 集群上的 `dev` 分区使用 16 块 GPU 训练 ResNet50 的例子。使用 `GPUS_PER_NODE=8` 参数来指定一个有 8 块 GPUS 的 `slurm` 集群节点, 使用 `CPUS_PER_TASK=2` 来指定每个任务拥有 2 块 cpu。

```
GPUS=16 GPUS_PER_NODE=8 CPUS_PER_TASK=2 ./tools/slurm_train.sh Test res50 configs/
↳ body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/res50_coco_256x192.py work_dirs/
↳ res50_coco_256x192
```

用户可以查看 `slurm_train.sh` 文件来检查完整的参数和环境变量。

如果您想使用由 ethernet 连接起来的多台机器，您可以使用以下命令：

```
# 在第一台机器上：
NNODES=2 NODE_RANK=0 PORT=$MASTER_PORT MASTER_ADDR=$MASTER_ADDR sh tools/dist_train.
↳ sh $CONFIG $GPUS

# 在第二台机器上：
NNODES=2 NODE_RANK=1 PORT=$MASTER_PORT MASTER_ADDR=$MASTER_ADDR sh tools/dist_train.
↳ sh $CONFIG $GPUS
```

但是，如果您不使用高速网路连接这几台机器的话，训练将会非常慢。

### 3.3.6 使用单台机器启动多个任务

如果用使用单台机器启动多个任务，如有 8 块 GPU 的单台机器上启动 2 个需要 4 块 GPU 的训练任务，则需要为每个任务指定不同端口，以避免通信冲突。

如果用户使用 `dist_train.sh` 脚本启动训练任务，则可以通过以下命令指定端口

```
CUDA_VISIBLE_DEVICES=0,1,2,3 PORT=29500 ./tools/dist_train.sh ${CONFIG_FILE} 4
CUDA_VISIBLE_DEVICES=4,5,6,7 PORT=29501 ./tools/dist_train.sh ${CONFIG_FILE} 4
```

如果用户在 `slurm` 集群下启动多个训练任务，则需要修改配置文件（通常是配置文件的第 4 行）中的 `dist_params` 变量，以设置不同的通信端口。

在 `config1.py` 中，

```
dist_params = dict(backend='nccl', port=29500)
```

在 `config2.py` 中，

```
dist_params = dict(backend='nccl', port=29501)
```

之后便可启动两个任务，分别对应 `config1.py` 和 `config2.py`。

```
CUDA_VISIBLE_DEVICES=0,1,2,3 GPUS=4 ./tools/slurm_train.sh ${PARTITION} ${JOB_NAME} _
↳ config1.py [--work-dir ${WORK_DIR}]
CUDA_VISIBLE_DEVICES=4,5,6,7 GPUS=4 ./tools/slurm_train.sh ${PARTITION} ${JOB_NAME} _
↳ config2.py [--work-dir ${WORK_DIR}]
```

## 3.4 基准测试

您可以使用以下脚本获得平均推理速度。请注意，它不包括 IO 时间和预处理时间。

```
python tools/analysis/benchmark_inference.py ${MMPOSE_CONFIG_FILE}
```

## 3.5 进阶教程

目前, MMPose 提供了以下更详细的教程:

- 如何编写配置文件
- 如何微调模型
- 如何增加新数据集
- 如何设计数据处理流程
- 如何增加新模块
- 如何导出模型为 *onnx* 格式
- 如何自定义运行配置
- 如何使用摄像头应用接口 (*Webcam API*)

## 4.1 2D Animal Pose Demo

### 4.1.1 2D Animal Pose Image Demo

#### Using gt bounding boxes as input

We provide a demo script to test a single image, given gt json file.

*Pose Model Preparation:* The pre-trained pose estimation model can be downloaded from [model zoo](#). Take [macaque model](#) as an example:

```
python demo/top_down_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --json-file ${JSON_FILE} \
    --out-img-root ${OUTPUT_DIR} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR}]
```

Examples:

```
python demo/top_down_img_demo.py \
    configs/animal/2d_kpt_sview_rgb_img/topdown_heatmap/macaque/res50_macaque_256x192.
↪py \
    https://download.openmmlab.com/mmpose/animal/resnet/res50_macaque_256x192-
↪98f1dd3a_20210407.pth \
```

(下页继续)

(续上页)

```
--img-root tests/data/monkey/ --json-file tests/data/monkey/test_monkey.json \
--out-img-root vis_results
```

To run demos on CPU:

```
python demo/top_down_img_demo.py \
    configs/animal/2d_kpt_sview_rgb_img/topdown_heatmap/monkey/res50_monkey_256x192-
↪py \
    https://download.openmmlab.com/mmpose/animal/resnet/res50_monkey_256x192-
↪98f1dd3a_20210407.pth \
    --img-root tests/data/monkey/ --json-file tests/data/monkey/test_monkey.json \
    --out-img-root vis_results \
    --device=cpu
```

## 4.1.2 2D Animal Pose Video Demo

We also provide video demos to illustrate the results.

### Using the full image as input

If the video is cropped with the object centered in the screen, we can simply use the full image as the model input (without object detection).

```
python demo/top_down_video_demo_full_frame_without_det.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR}]
```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```
python demo/top_down_video_demo_full_frame_without_det.py \
    configs/animal/2d_kpt_sview_rgb_img/topdown_heatmap/fly/res152_fly_192x192.py \
    https://download.openmmlab.com/mmpose/animal/resnet/res152_fly_192x192-fcafb5a-
↪20210407.pth \
    --video-path https://user-images.githubusercontent.com/87690686/165095600-
↪f68e0d42-830d-4c22-8940-c90c9f3bb817.mp4 \
    --out-video-root vis_results
```



## Using MMDetection to detect animals

Assume that you have already installed `mmdet`.

### COCO-animals

In COCO dataset, there are 80 object categories, including 10 common animal categories (15: 'bird', 16: 'cat', 17: 'dog', 18: 'horse', 19: 'sheep', 20: 'cow', 21: 'elephant', 22: 'bear', 23: 'zebra', 24: 'giraffe') For these COCO-animals, please download the COCO pre-trained detection model from [MMDetection Model Zoo](#).

```
python demo/top_down_video_demo_with_mmdet.py \
    ${MMDET_CONFIG_FILE} ${MMDET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    --det-cat-id ${CATEGORY_ID}
[--show --device ${GPU_ID or CPU}] \
[--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}]
```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```
python demo/top_down_video_demo_with_mmdet.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪2x_coco/faster_rcnn_r50_fpn_2x_coco_bbox_mAP-0.384_20200504_210434-a5d8aa15.pth \
    configs/animal/2d_kpt_sview_rgb_img/topdown_heatmap/horse10/res50_horse10_256x256-
↪split1.py \
    https://download.openmmlab.com/mmpose/animal/resnet/res50_horse10_256x256_split1-
↪3a3dc37e_20210405.pth \
    --video-path https://user-images.githubusercontent.com/15977946/173124855-
↪c626835e-1863-4003-8184-315bc0b7b561.mp4 \
    --out-video-root vis_results \
    --bbox-thr 0.1 \
    --kpt-thr 0.4 \
    --det-cat-id 18
```

### Other Animals

For other animals, we have also provided some pre-trained animal detection models (1-class models). Supported models can be found in [det model zoo](#). The pre-trained animal pose estimation model can be found in [pose model zoo](#).

```
python demo/top_down_video_demo_with_mmdet.py \
    ${MMDET_CONFIG_FILE} ${MMDET_CHECKPOINT_FILE} \
```

(下页继续)

(续上页)

```

    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--det-cat-id ${CATEGORY_ID}]
    [--show --device ${GPU_ID or CPU}] \
    [--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}]

```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```

python demo/top_down_video_demo_with_mmdet.py \
    demo/mmdetection_cfg/cascade_rcnn_x101_64x4d_fpn_1class.py \
    https://download.openmmlab.com/mmpose/mmdet_pretrained/cascade_rcnn_x101_64x4d_
↪fpn_20e_macaque-e45e36f5_20210409.pth \
    configs/animal/2d_kpt_sview_rgb_img/topdown_heatmap/macaque/hrnet_w32_macaque_
↪256x192.py \
    https://download.openmmlab.com/mmpose/animal/hrnet/hrnet_w32_macaque_256x192-
↪f7e9e04f_20210407.pth \
    --video-path https://user-images.githubusercontent.com/15977946/173135633-
↪1c54a944-4f01-4747-8c2e-55b8c83be533.mp4 \
    --out-video-root vis_results \
    --bbox-thr 0.5 \
    --kpt-thr 0.3 \
    --radius 9 \
    --thickness 3

```

### 4.1.3 Speed Up Inference

Some tips to speed up MMPose inference:

For 2D animal pose estimation models, try to edit the config file. For example,

1. set `flip_test=False` in `macaque-res50`.
2. set `post_process='default'` in `macaque-res50`.

## 4.2 2D Face Keypoint Demo

### 4.2.1 2D Face Image Demo

#### Using gt face bounding boxes as input

We provide a demo script to test a single image, given gt json file.

*Face Keypoint Model Preparation:* The pre-trained face keypoint estimation model can be found from [model zoo](#). Take [aflw model](#) as an example:

```
python demo/top_down_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --json-file ${JSON_FILE} \
    --out-img-root ${OUTPUT_DIR} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR}]
```

Examples:

```
python demo/top_down_img_demo.py \
    configs/face/2d_kpt_sview_rgb_img/topdown_heatmap/aflw/hrnetv2_w18_aflw_256x256.
↪py \
    https://download.openmmlab.com/mmpose/face/hrnetv2/hrnetv2_w18_aflw_256x256-
↪f2bbc62b_20210125.pth \
    --img-root tests/data/aflw/ --json-file tests/data/aflw/test_aflw.json \
    --out-img-root vis_results
```

To run demos on CPU:

```
python demo/top_down_img_demo.py \
    configs/face/2d_kpt_sview_rgb_img/topdown_heatmap/aflw/hrnetv2_w18_aflw_256x256.
↪py \
    https://download.openmmlab.com/mmpose/face/hrnetv2/hrnetv2_w18_aflw_256x256-
↪f2bbc62b_20210125.pth \
    --img-root tests/data/aflw/ --json-file tests/data/aflw/test_aflw.json \
    --out-img-root vis_results \
    --device=cpu
```

## Using face bounding box detectors

We provide a demo script to run face detection and face keypoint estimation.

Please install `face_recognition` before running the demo, by `pip install face_recognition`. For more details, please refer to [https://github.com/ageitgey/face\\_recognition](https://github.com/ageitgey/face_recognition).

```
python demo/face_img_demo.py \  
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \  
    --img-root ${IMG_ROOT} --img ${IMG_FILE} \  
    --out-img-root ${OUTPUT_DIR} \  
    [--show --device ${GPU_ID or CPU}] \  
    [--kpt-thr ${KPT_SCORE_THR}]
```

```
python demo/face_img_demo.py \  
    configs/face/2d_kpt_sview_rgb_img/topdown_heatmap/aflw/hrnetv2_w18_aflw_256x256.  
→py \  
    https://download.openmmlab.com/mmpose/face/hrnetv2/hrnetv2_w18_aflw_256x256-  
→f2bbc62b_20210125.pth \  
    --img-root tests/data/aflw/ \  
    --img image04476.jpg \  
    --out-img-root vis_results
```

## 4.2.2 2D Face Video Demo

We also provide a video demo to illustrate the results.

Please install `face_recognition` before running the demo, by `pip install face_recognition`. For more details, please refer to [https://github.com/ageitgey/face\\_recognition](https://github.com/ageitgey/face_recognition).

```
python demo/face_video_demo.py \  
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \  
    --video-path ${VIDEO_PATH} \  
    --out-video-root ${OUTPUT_VIDEO_ROOT} \  
    [--show --device ${GPU_ID or CPU}] \  
    [--kpt-thr ${KPT_SCORE_THR}]
```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```
python demo/face_video_demo.py \  
    configs/face/2d_kpt_sview_rgb_img/topdown_heatmap/aflw/hrnetv2_w18_aflw_256x256.  
→py \  
    https://download.openmmlab.com/mmpose/face/hrnetv2/hrnetv2_w18_aflw_256x256-  
→f2bbc62b_20210125.pth
```

(下页继续)

(续上页)

```

--video-path https://user-images.githubusercontent.com/87690686/137441355-
↪ec4da09c-3a8f-421b-bee9-b8b26f8c2dd0.mp4 \
--out-video-root vis_results

```

## 4.2.3 Speed Up Inference

Some tips to speed up MMPose inference:

For 2D face keypoint estimation models, try to edit the config file. For example,

1. set `flip_test=False` in `face-hrnetv2_w18`.
2. set `post_process='default'` in `face-hrnetv2_w18`.

## 4.3 2D Hand Keypoint Demo

### 4.3.1 2D Hand Image Demo

#### Using gt hand bounding boxes as input

We provide a demo script to test a single image, given gt json file.

*Hand Pose Model Preparation:* The pre-trained hand pose estimation model can be downloaded from [model zoo](#). Take [onehand10k model](#) as an example:

```

python demo/top_down_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --json-file ${JSON_FILE} \
    --out-img-root ${OUTPUT_DIR} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR}]

```

Examples:

```

python demo/top_down_img_demo.py \
    configs/hand/2d_kpt_sview_rgb_img/topdown_heatmap/onehand10k/res50_onehand10k_
↪256x256.py \
    https://download.openmmlab.com/mmpose/top_down/resnet/res50_onehand10k_256x256-
↪e67998f6_20200813.pth \

```

(下页继续)

(续上页)

```
--img-root tests/data/onehand10k/ --json-file tests/data/onehand10k/test_
↪onehand10k.json \
--out-img-root vis_results
```

To run demos on CPU:

```
python demo/top_down_img_demo.py \
    configs/hand/2d_kpt_sview_rgb_img/topdown_heatmap/onehand10k/res50_onehand10k_
↪256x256.py \
    https://download.openmmlab.com/mmpose/top_down/resnet/res50_onehand10k_256x256-
↪e67998f6_20200813.pth \
    --img-root tests/data/onehand10k/ --json-file tests/data/onehand10k/test_
↪onehand10k.json \
    --out-img-root vis_results \
    --device=cpu
```

## Using mmdet for hand bounding box detection

We provide a demo script to run mmdet for hand detection, and mmpose for hand pose estimation.

Assume that you have already installed `mmdet`.

*Hand Box Model Preparation:* The pre-trained hand box estimation model can be found in [det model zoo](#).

*Hand Pose Model Preparation:* The pre-trained hand pose estimation model can be downloaded from [pose model zoo](#).

```
python demo/top_down_img_demo_with_mmdet.py \
    ${MMDET_CONFIG_FILE} ${MMDET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --img ${IMG_FILE} \
    --out-img-root ${OUTPUT_DIR} \
    [--show --device ${GPU_ID or CPU}] \
    [--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}]
```

```
python demo/top_down_img_demo_with_mmdet.py demo/mmdetection_cfg/cascade_rcnn_x101_
↪64x4d_fpn_1class.py \
    https://download.openmmlab.com/mmpose/mmdet_pretrained/cascade_rcnn_x101_64x4d_
↪fpn_20e_onehand10k-dac19597_20201030.pth \
    configs/hand/2d_kpt_sview_rgb_img/topdown_heatmap/onehand10k/res50_onehand10k_
↪256x256.py \
    https://download.openmmlab.com/mmpose/top_down/resnet/res50_onehand10k_256x256-
↪e67998f6_20200813.pth \
    --img-root tests/data/onehand10k/ \
```

(下页继续)

(续上页)

```
--img 9.jpg \
--out-img-root vis_results
```

### 4.3.2 2D Hand Video Demo

We also provide a video demo to illustrate the results.

Assume that you have already installed `mmdet`.

*Hand Box Model Preparation:* The pre-trained hand box estimation model can be found in [det model zoo](#).

*Hand Pose Model Preparation:* The pre-trained hand pose estimation model can be found in [pose model zoo](#).

```
python demo/top_down_video_demo_with_mmdet.py \
    ${MMDET_CONFIG_FILE} ${MMDET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
    [--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}]
```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```
python demo/top_down_video_demo_with_mmdet.py demo/mmdetection_cfg/cascade_rcnn_x101_
↪64x4d_fpn_1class.py \
    https://download.openmmlab.com/mmpose/mmdet_pretrained/cascade_rcnn_x101_64x4d_
↪fpn_20e_onehand10k-dac19597_20201030.pth \
    configs/hand/2d_kpt_sview_rgb_img/topdown_heatmap/onehand10k/res50_onehand10k_
↪256x256.py \
    https://download.openmmlab.com/mmpose/top_down/resnet/res50_onehand10k_256x256-
↪e67998f6_20200813.pth \
    --video-path https://user-images.githubusercontent.com/87690686/137441388-
↪3ea93d26-5445-4184-829e-bf7011def9e4.mp4 \
    --out-video-root vis_results
```

### 4.3.3 Speed Up Inference

Some tips to speed up MMPose inference:

For 2D hand pose estimation models, try to edit the config file. For example,

1. set `flip_test=False` in `hand-res50`.
2. set `post_process='default'` in `hand-res50`.

## 4.4 2D Human Pose Demo

### 4.4.1 2D Human Pose Top-Down Image Demo

#### Using gt human bounding boxes as input

We provide a demo script to test a single image, given gt json file.

```
python demo/top_down_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --json-file ${JSON_FILE} \
    --out-img-root ${OUTPUT_DIR} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR}]
```

Examples:

```
python demo/top_down_img_demo.py \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/hrnet_w48_coco_256x192.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_256x192-
    ↪b9e0b3ab_20200708.pth \
    --img-root tests/data/coco/ --json-file tests/data/coco/test_coco.json \
    --out-img-root vis_results
```

To run demos on CPU:

```
python demo/top_down_img_demo.py \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/hrnet_w48_coco_256x192.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_256x192-
    ↪b9e0b3ab_20200708.pth \
    --img-root tests/data/coco/ --json-file tests/data/coco/test_coco.json \
    --out-img-root vis_results \
    --device=cpu
```



## Using mmdet for human bounding box detection

We provide a demo script to run mmdet for human detection, and mmpose for pose estimation.

Assume that you have already installed mmdet.

```
python demo/top_down_img_demo_with_mmdet.py \
    ${MMDNET_CONFIG_FILE} ${MMDNET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --img ${IMG_FILE} \
    --out-img-root ${OUTPUT_DIR} \
    [--show --device ${GPU_ID or CPU}] \
    [--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}]
```

Examples:

```
python demo/top_down_img_demo_with_mmdet.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
    ↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/hrnet_w48_coco_256x192.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_256x192-
    ↪b9e0b3ab_20200708.pth \
    --img-root tests/data/coco/ \
    --img 000000196141.jpg \
    --out-img-root vis_results
```

## 4.4.2 2D Human Pose Top-Down Video Demo

We also provide a video demo to illustrate the results.

Assume that you have already installed mmdet.

```
python demo/top_down_video_demo_with_mmdet.py \
    ${MMDNET_CONFIG_FILE} ${MMDNET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
    [--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}] \
    [--use-multi-frames] [--online]
```

Note that

1. \${VIDEO\_PATH} can be the local path or **URL** link to video file.

2. You can turn on the `[--use-multi-frames]` option to use multi frames for inference in the pose estimation stage.
3. If the `[--online]` option is set to **True**, future frame information can **not** be used when using multi frames for inference in the pose estimation stage.

Examples:

For single-frame inference that do not rely on extra frames to get the final results of the current frame, try this:

```
python demo/top_down_video_demo_with_mmdet.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/hrnet_w48_coco_256x192.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_256x192-
↪b9e0b3ab_20200708.pth \
    --video-path demo/resources/demo.mp4 \
    --out-video-root vis_results
```

For multi-frame inference that rely on extra frames to get the final results of the current frame, try this:

```
python demo/top_down_video_demo_with_mmdet.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/body/2d_kpt_sview_rgb_vid/posewarper/posetrack18/hrnet_w48_posetrack18_
↪384x288_posewarper_stage2.py \
    https://download.openmmlab.com/mmpose/top_down/posewarper/hrnet_w48_posetrack18_
↪384x288_posewarper_stage2-4abf88db_20211130.pth \
    --video-path https://user-images.githubusercontent.com/87690686/137440639-
↪fb08603d-9a35-474e-b65f-46b5c06b68d6.mp4 \
    --out-video-root vis_results \
    --use-multi-frames --online
```

## Using the full image as input

We also provide a video demo which does not require human bounding box detection. If the video is cropped with the human centered in the screen, we can simply use the full image as the model input.

```
python demo/top_down_video_demo_full_frame_without_det.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR}]
```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```
python demo/top_down_video_demo_full_frame_without_det.py \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/vipnas_res50_coco_256x192.
↪py \
    https://download.openmmlab.com/mmpose/top_down/vipnas/vipnas_res50_coco_256x192-
↪cc43b466_20210624.pth \
    --video-path https://user-images.githubusercontent.com/87690686/169808764-
↪29e5678c-6762-4f43-8666-c3e60f94338f.mp4 \
    --show
```

We also provide a GPU version which can accelerate inference and save CPU workload. Assume that you have already installed `ffmpegcv`. If the `--nvdecode` option is turned on, the video reader can support NVIDIA-VIDEO-DECODING for some qualified Nvidia GPUs, which can further accelerate the inference.

```
python demo/top_down_video_demo_full_frame_without_det_gpuaccel.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR}] \
    [--nvdecode]
```

Examples:

```
python demo/top_down_video_demo_full_frame_without_det_gpuaccel.py \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/vipnas_res50_coco_256x192.
↪py \
    https://download.openmmlab.com/mmpose/top_down/vipnas/vipnas_res50_coco_256x192-
↪cc43b466_20210624.pth \
    --video-path https://user-images.githubusercontent.com/87690686/169808764-
↪29e5678c-6762-4f43-8666-c3e60f94338f.mp4 \
    --out-video-root vis_results
```

### 4.4.3 2D Human Pose Bottom-Up Image Demo

We provide a demo script to test a single image.

```
python demo/bottom_up_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-path ${IMG_PATH} \
    --out-img-root ${OUTPUT_DIR} \
```

(下页继续)

(续上页)

```
[--show --device ${GPU_ID or CPU}] \
[--kpt-thr ${KPT_SCORE_THR} --pose-nms-thr ${POSE_NMS_THR}]
```

Examples:

```
python demo/bottom_up_img_demo.py \
    configs/body/2d_kpt_sview_rgb_img/associative_embedding/coco/hrnet_w32_coco_
↪512x512.py \
    https://download.openmmlab.com/mmpose/bottom_up/hrnet_w32_coco_512x512-bcb8c247_
↪20200816.pth \
    --img-path tests/data/coco/ \
    --out-img-root vis_results
```

#### 4.4.4 2D Human Pose Bottom-Up Video Demo

We also provide a video demo to illustrate the results.

```
python demo/bottom_up_video_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR} --pose-nms-thr ${POSE_NMS_THR}]
```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```
python demo/bottom_up_video_demo.py \
    configs/body/2d_kpt_sview_rgb_img/associative_embedding/coco/hrnet_w32_coco_
↪512x512.py \
    https://download.openmmlab.com/mmpose/bottom_up/hrnet_w32_coco_512x512-bcb8c247_
↪20200816.pth \
    --video-path demo/resources/demo.mp4 \
    --out-video-root vis_results
```

## 4.4.5 Speed Up Inference

Some tips to speed up MMPose inference:

For top-down models, try to edit the config file. For example,

1. set `flip_test=False` in `topdown-res50`.
2. set `post_process='default'` in `topdown-res50`.
3. use faster human bounding box detector, see `MMDetection`.

For bottom-up models, try to edit the config file. For example,

1. set `flip_test=False` in `AE-res50`.
2. set `adjust=False` in `AE-res50`.
3. set `refine=False` in `AE-res50`.
4. use smaller input image size in `AE-res50`.

## 4.5 2D Pose Tracking Demo

### 4.5.1 2D Top-Down Video Human Pose Tracking Demo

We provide a video demo to illustrate the pose tracking results.

Assume that you have already installed `mmdet`.

```
python demo/top_down_pose_tracking_demo_with_mmdet.py \
    ${MMDNET_CONFIG_FILE} ${MMDNET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
    [--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}] \
    [--use-oks-tracking --tracking-thr ${TRACKING_THR} --euro] \
    [--use-multi-frames] [--online]
```

Note that

1. `${VIDEO_PATH}` can be the local path or **URL** link to video file.
2. You can turn on the `[--use-multi-frames]` option to use multi frames for inference in the pose estimation stage.

3. If the `--online` option is set to **True**, future frame information can **not** be used when using multi frames for inference in the pose estimation stage.

Examples:

For single-frame inference that do not rely on extra frames to get the final results of the current frame, try this:

```
python demo/top_down_pose_tracking_demo_with_mmdet.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/res50_coco_256x192.py \
    https://download.openmmlab.com/mmpose/top_down/resnet/res50_coco_256x192-ec54d7f3_
↪20200709.pth \
    --video-path demo/resources/demo.mp4 \
    --out-video-root vis_results
```

For multi-frame inference that rely on extra frames to get the final results of the current frame, try this:

```
python demo/top_down_pose_tracking_demo_with_mmdet.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/body/2d_kpt_sview_rgb_vid/posewarper/posetrack18/hrnet_w48_posetrack18_
↪384x288_posewarper_stage2.py \
    https://download.openmmlab.com/mmpose/top_down/posewarper/hrnet_w48_posetrack18_
↪384x288_posewarper_stage2-4abf88db_20211130.pth \
    --video-path https://user-images.githubusercontent.com/87690686/137440639-
↪fb08603d-9a35-474e-b65f-46b5c06b68d6.mp4 \
    --out-video-root vis_results \
    --use-multi-frames --online
```

## 4.5.2 2D Top-Down Video Human Pose Tracking Demo with MMTracking

MMTracking is an open source video perception toolbox based on PyTorch for tracking related tasks. Here we show how to utilize MMTracking and MMPose to achieve human pose tracking.

Assume that you have already installed `mmtracking`.

```
python demo/top_down_video_demo_with_mmtracking.py \
    ${MMTRACKING_CONFIG_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
```

(下页继续)

(续上页)

```
[--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}] \
[--use-multi-frames] [--online]
```

Note that

1. `${VIDEO_PATH}` can be the local path or **URL** link to video file.
2. You can turn on the `--use-multi-frames` option to use multi frames for inference in the pose estimation stage.
3. If the `--online` option is set to **True**, future frame information can **not** be used when using multi frames for inference in the pose estimation stage.

Examples:

For single-frame inference that do not rely on extra frames to get the final results of the current frame, try this:

```
python demo/top_down_pose_tracking_demo_with_mmtracking.py \
demo/mmtracking_cfg/tracktor_faster-rcnn_r50_fpn_4e_mot17-private.py \
configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/res50_coco_256x192.py \
https://download.openmmlab.com/mmpose/top_down/resnet/res50_coco_256x192-ec54d7f3_
↪20200709.pth \
--video-path demo/resources/demo.mp4 \
--out-video-root vis_results
```

For multi-frame inference that rely on extra frames to get the final results of the current frame, try this:

```
python demo/top_down_pose_tracking_demo_with_mmtracking.py \
demo/mmtracking_cfg/tracktor_faster-rcnn_r50_fpn_4e_mot17-private.py \
configs/body/2d_kpt_sview_rgb_vid/posewarper/posetrack18/hrnet_w48_posetrack18_
↪384x288_posewarper_stage2.py \
https://download.openmmlab.com/mmpose/top_down/posewarper/hrnet_w48_posetrack18_
↪384x288_posewarper_stage2-4abf88db_20211130.pth \
--video-path demo/resources/demo.mp4 \
--out-video-root vis_results \
--use-multi-frames --online
```

### 4.5.3 2D Bottom-Up Video Human Pose Tracking Demo

We also provide a pose tracking demo with bottom-up pose estimation methods.

```
python demo/bottom_up_pose_tracking_demo.py \
${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
--video-path ${VIDEO_PATH} \
--out-video-root ${OUTPUT_VIDEO_ROOT} \
```

(下页继续)

(续上页)

```
[--show --device ${GPU_ID or CPU}] \  
[--kpt-thr ${KPT_SCORE_THR} --pose-nms-thr ${POSE_NMS_THR}]  
[--use-oks-tracking --tracking-thr ${TRACKING_THR} --euro]
```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```
python demo/bottom_up_pose_tracking_demo.py \  
    configs/body/2d_kpt_sview_rgb_img/associative_embedding/coco/hrnet_w32_coco_  
↪512x512.py \  
    https://download.openmmlab.com/mmpose/bottom_up/hrnet_w32_coco_512x512-bcb8c247_  
↪20200816.pth \  
    --video-path demo/resources/demo.mp4 \  
    --out-video-root vis_results
```

## 4.5.4 Speed Up Inference

Some tips to speed up MMPose inference:

For top-down models, try to edit the config file. For example,

1. set `flip_test=False` in `topdown-res50`.
2. set `post_process='default'` in `topdown-res50`.
3. use faster human detector or human tracker, see `MMDetection` or `MMTracking`.

For bottom-up models, try to edit the config file. For example,

1. set `flip_test=False` in `AE-res50`.
2. set `adjust=False` in `AE-res50`.
3. set `refine=False` in `AE-res50`.
4. use smaller input image size in `AE-res50`.

## 4.6 2D Human Whole-Body Pose Demo



## 4.6.1 2D Human Whole-Body Pose Top-Down Image Demo

### Using gt human bounding boxes as input

We provide a demo script to test a single image, given gt json file.

```
python demo/top_down_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --json-file ${JSON_FILE} \
    --out-img-root ${OUTPUT_DIR} \
    [--show --device ${GPU_ID or CPU}] \
    [--kpt-thr ${KPT_SCORE_THR}]
```

Examples:

```
python demo/top_down_img_demo.py \
    configs/wholebody/2d_kpt_sview_rgb_img/topdown_heatmap/coco-wholebody/hrnet_w48_
    →coco_wholebody_384x288_dark_plus.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_wholebody_
    →384x288_dark-f5726563_20200918.pth \
    --img-root tests/data/coco/ --json-file tests/data/coco/test_coco.json \
    --out-img-root vis_results
```

To run demos on CPU:

```
python demo/top_down_img_demo.py \
    configs/wholebody/2d_kpt_sview_rgb_img/topdown_heatmap/coco-wholebody/hrnet_w48_
    →coco_wholebody_384x288_dark_plus.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_wholebody_
    →384x288_dark-f5726563_20200918.pth \
    --img-root tests/data/coco/ --json-file tests/data/coco/test_coco.json \
    --out-img-root vis_results \
    --device=cpu
```

### Using mmdet for human bounding box detection

We provide a demo script to run mmdet for human detection, and mmpose for pose estimation.

Assume that you have already installed mmdet.

```
python demo/top_down_img_demo_with_mmdet.py \
    ${MMDET_CONFIG_FILE} ${MMDET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --img-root ${IMG_ROOT} --img ${IMG_FILE} \
```

(下页继续)

(续上页)

```
--out-img-root ${OUTPUT_DIR} \
[--show --device ${GPU_ID or CPU}] \
[--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}]
```

Examples:

```
python demo/top_down_img_demo_with_mmdet.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/wholebody/2d_kpt_sview_rgb_img/topdown_heatmap/coco-wholebody/hrnet_w48_
↪coco_wholebody_384x288_dark_plus.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_wholebody_
↪384x288_dark-f5726563_20200918.pth \
    --img-root tests/data/coco/ \
    --img 000000196141.jpg \
    --out-img-root vis_results
```

## 4.6.2 2D Human Whole-Body Pose Top-Down Video Demo

We also provide a video demo to illustrate the results.

Assume that you have already installed `mmdet`.

```
python demo/top_down_video_demo_with_mmdet.py \
    ${MMDET_CONFIG_FILE} ${MMDET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --video-path ${VIDEO_PATH} \
    --out-video-root ${OUTPUT_VIDEO_ROOT} \
    [--show --device ${GPU_ID or CPU}] \
    [--bbox-thr ${BBOX_SCORE_THR} --kpt-thr ${KPT_SCORE_THR}]
```

Note that `${VIDEO_PATH}` can be the local path or **URL** link to video file.

Examples:

```
python demo/top_down_video_demo_with_mmdet.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/wholebody/2d_kpt_sview_rgb_img/topdown_heatmap/coco-wholebody/hrnet_w48_
↪coco_wholebody_384x288_dark_plus.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_wholebody_
↪384x288_dark-f5726563_20200918.pth \
```

(下页继续)

(续上页)

```
--video-path https://user-images.githubusercontent.com/87690686/137440639-
→fb08603d-9a35-474e-b65f-46b5c06b68d6.mp4 \
--out-video-root vis_results
```

### 4.6.3 Speed Up Inference

Some tips to speed up MMPose inference:

For top-down models, try to edit the config file. For example,

1. set `flip_test=False` in `pose_hrnet_w48_dark+`.
2. set `post_process='default'` in `pose_hrnet_w48_dark+`.
3. use faster human bounding box detector, see [MMDetection](#).

## 4.7 3D Mesh Demo

### 4.7.1 3D Mesh Recovery Demo

We provide a demo script to recover human 3D mesh from a single image.

```
python demo/mesh_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --json-file ${JSON_FILE} \
    --img-root ${IMG_ROOT} \
    [--show] \
    [--device ${GPU_ID or CPU}] \
    [--out-img-root ${OUTPUT_DIR}]
```

Example:

```
python demo/mesh_img_demo.py \
    configs/body/3d_mesh_sview_rgb_img/hmr/mixed/res50_mixed_224x224.py \
    https://download.openmmlab.com/mmpose/mesh/hmr/hmr_mesh_224x224-c21e8229_20201015.
→pth \
    --json-file tests/data/h36m/h36m_coco.json \
    --img-root tests/data/h36m \
    --out-img-root vis_results
```

## 4.8 3D Hand Demo

### 4.8.1 3D Hand Estimation Image Demo

#### Using gt hand bounding boxes as input

We provide a demo script to test a single image, given gt json file.

```
python demo/interhand3d_img_demo.py \
    ${MMPOSE_CONFIG_FILE} ${MMPOSE_CHECKPOINT_FILE} \
    --json-file ${JSON_FILE} \
    --img-root ${IMG_ROOT} \
    [--camera-param-file ${CAMERA_PARAM_FILE}] \
    [--gt-joints-file ${GT_JOINTS_FILE}] \
    [--show] \
    [--device ${GPU_ID} or CPU] \
    [--out-img-root ${OUTPUT_DIR}] \
    [--rebase-keypoint-height] \
    [--show-ground-truth]
```

Example with gt keypoints and camera parameters:

```
python demo/interhand3d_img_demo.py \
    configs/hand/3d_kpt_sview_rgb_img/internet/interhand3d/res50_interhand3d_all_
↪256x256.py \
    https://download.openmmlab.com/mmpose/hand3d/internet/res50_intehand3d_all_
↪256x256-b9c1cf4c_20210506.pth \
    --json-file tests/data/interhand2.6m/test_interhand2.6m_data.json \
    --img-root tests/data/interhand2.6m \
    --camera-param-file tests/data/interhand2.6m/test_interhand2.6m_camera.json \
    --gt-joints-file tests/data/interhand2.6m/test_interhand2.6m_joint_3d.json \
    --out-img-root vis_results \
    --rebase-keypoint-height \
    --show-ground-truth
```

Example without gt keypoints and camera parameters:

```
python demo/interhand3d_img_demo.py \
    configs/hand/3d_kpt_sview_rgb_img/internet/interhand3d/res50_interhand3d_all_
↪256x256.py \
    https://download.openmmlab.com/mmpose/hand3d/internet/res50_intehand3d_all_
↪256x256-b9c1cf4c_20210506.pth \
```

(下页继续)

(续上页)

```
--json-file tests/data/interhand2.6m/test_interhand2.6m_data.json \
--img-root tests/data/interhand2.6m \
--out-img-root vis_results \
--rebase-keypoint-height
```

## 4.9 3D Human Pose Demo

### 4.9.1 3D Human Pose Two-stage Estimation Image Demo

Using ground truth 2D poses as the 1st stage (pose detection) result, and inference the 2nd stage (2D-to-3D lifting)

We provide a demo script to test on single images with a given ground-truth Json file.

```
python demo/body3d_two_stage_img_demo.py \
    ${MMPOSE_CONFIG_FILE_3D} \
    ${MMPOSE_CHECKPOINT_FILE_3D} \
    --json-file ${JSON_FILE} \
    --img-root ${IMG_ROOT} \
    --only-second-stage \
    [--show] \
    [--device ${GPU_ID} or CPU] \
    [--out-img-root ${OUTPUT_DIR}] \
    [--rebase-keypoint-height] \
    [--show-ground-truth]
```

Example:

```
python demo/body3d_two_stage_img_demo.py \
    configs/body/3d_kpt_sview_rgb_img/pose_lift/h36m/simplebaseline3d_h36m.py \
    https://download.openmmlab.com/mmpose/body3d/simple_baseline/simple3Dbaseline_
↪h36m-f0ad73a4_20210419.pth \
    --json-file tests/data/h36m/h36m_coco.json \
    --img-root tests/data/h36m \
    --camera-param-file tests/data/h36m/cameras.pkl \
    --only-second-stage \
    --out-img-root vis_results \
    --rebase-keypoint-height \
    --show-ground-truth
```

## 4.9.2 3D Human Pose Two-stage Estimation Video Demo

Using mmdet for human bounding box detection and top-down model for the 1st stage (2D pose detection), and inference the 2nd stage (2D-to-3D lifting)

Assume that you have already installed `mmdet`.

```
python demo/body3d_two_stage_video_demo.py \
    ${MMDET_CONFIG_FILE} \
    ${MMDET_CHECKPOINT_FILE} \
    ${MMPOSE_CONFIG_FILE_2D} \
    ${MMPOSE_CHECKPOINT_FILE_2D} \
    ${MMPOSE_CONFIG_FILE_3D} \
    ${MMPOSE_CHECKPOINT_FILE_3D} \
    --video-path ${VIDEO_PATH} \
    [--rebase-keypoint-height] \
    [--norm-pose-2d] \
    [--num-poses-vis NUM_POSES_VIS] \
    [--show] \
    [--out-video-root ${OUT_VIDEO_ROOT}] \
    [--device ${GPU_ID or CPU}] \
    [--det-cat-id DET_CAT_ID] \
    [--bbox-thr BBOX_THR] \
    [--kpt-thr KPT_THR] \
    [--use-oks-tracking] \
    [--tracking-thr TRACKING_THR] \
    [--euro] \
    [--radius RADIUS] \
    [--thickness THICKNESS] \
    [--use-multi-frames] [--online]
```

Note that

1. `${VIDEO_PATH}` can be the local path or **URL** link to video file.
2. You can turn on the `[--use-multi-frames]` option to use multi frames for inference in the 2D pose detection stage.
3. If the `[--online]` option is set to **True**, future frame information can **not** be used when using multi frames for inference in the 2D pose detection stage.

Examples:

During 2D pose detection, for single-frame inference that do not rely on extra frames to get the final results of the current frame, try this:

```
python demo/body3d_two_stage_video_demo.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/body/2d_kpt_sview_rgb_img/topdown_heatmap/coco/hrnet_w48_coco_256x192.py \
    https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_256x192-
↪b9e0b3ab_20200708.pth \
    configs/body/3d_kpt_sview_rgb_vid/video_pose_lift/h36m/videopose3d_h36m_243frames_
↪fullconv_supervised_cpn_ft.py \
    https://download.openmmlab.com/mmpose/body3d/videopose/videopose_h36m_243frames_
↪fullconv_supervised_cpn_ft-88f5abbb_20210527.pth \
    --video-path https://user-images.githubusercontent.com/87690686/164970135-
↪b14e424c-765a-4180-9bc8-fa8d6abc5510.mp4 \
    --out-video-root vis_results \
    --rebase-keypoint-height
```

During 2D pose detection, for multi-frame inference that rely on extra frames to get the final results of the current frame, try this:

```
python demo/body3d_two_stage_video_demo.py \
    demo/mmdetection_cfg/faster_rcnn_r50_fpn_coco.py \
    https://download.openmmlab.com/mmdetection/v2.0/faster_rcnn/faster_rcnn_r50_fpn_
↪1x_coco/faster_rcnn_r50_fpn_1x_coco_20200130-047c8118.pth \
    configs/body/2d_kpt_sview_rgb_vid/posewarper/posetrack18/hrnet_w48_posetrack18_
↪384x288_posewarper_stage2.py \
    https://download.openmmlab.com/mmpose/top_down/posewarper/hrnet_w48_posetrack18_
↪384x288_posewarper_stage2-4abf88db_20211130.pth \
    configs/body/3d_kpt_sview_rgb_vid/video_pose_lift/h36m/videopose3d_h36m_243frames_
↪fullconv_supervised_cpn_ft.py \
    https://download.openmmlab.com/mmpose/body3d/videopose/videopose_h36m_243frames_
↪fullconv_supervised_cpn_ft-88f5abbb_20210527.pth \
    --video-path https://user-images.githubusercontent.com/87690686/164970135-
↪b14e424c-765a-4180-9bc8-fa8d6abc5510.mp4 \
    --out-video-root vis_results \
    --rebase-keypoint-height \
    --use-multi-frames --online
```

## 4.10 3D Multiview Human Pose Demo

### 4.10.1 3D Multiview Human Pose Estimation Image Demo

#### VoxelPose

We provide a demo script to test on multiview images with given camera parameters. To run the demo:

```
python demo/body3d_multiview_detect_and_regress_img_demo.py \
    ${MMPOSE_CONFIG_FILE} \
    ${MMPOSE_CHECKPOINT_FILE} \
    --out-img-root ${OUT_IMG_ROOT} \
    --camera-param-file ${CAMERA_FILE} \
    [--img-root ${IMG_ROOT}] \
    [--visualize-single-view ${VIS_SINGLE_IMG}] \
    [--device ${GPU_ID or CPU}] \
    [--out-img-root ${OUTPUT_DIR}]
```

Example:

```
python demo/body3d_multiview_detect_and_regress_img_demo.py \
    configs/body/3d_kpt_mview_rgb_img/voxelpose/panoptic/voxelpose_prn64x64x64_
    ↪cpn80x80x20_panoptic_cam5.py \
    https://download.openmmlab.com/mmpose/body3d/voxelpose/voxelpose_prn64x64x64_
    ↪cpn80x80x20_panoptic_cam5-545c150e_20211103.pth \
    --out-img-root vis_results \
    --camera-param-file tests/data/panoptic_body3d/demo/camera_parameters.json \
    --visualize-single-view
```

#### Data Preparation

Currently, we only support CMU Panoptic data format. Users can leave the argument `--img-root` unset to automatically download our [default demo data](#) (~6M). Users can also use custom data, which should be organized as follow:

```
|— ${IMG_ROOT}
|   |— camera_parameters.json
|   |— camera0
|       |-- 0.jpg
|       |-- ...
|   |— camera1
|   |— ...
```

The camera parameters should be a dictionary that include a key “cameras”. Under the key “cameras” should be a



list of dictionaries containing the camera parameters. Each dictionary under the list should include a key “name” , the value of which is the directory name of images of a certain camera view.

```
{
  "cameras": [
    {"name": "camera0", ...},
    {"name": "camera1", ...},
    ...
  ]
}
```

## 4.11 Hand Gesture Recognition Demo

We provide a demo for gesture recognition with MMPose. This demo is built upon [MMPose Webcam API](#).

### 4.11.1 Get started

Launch the demo from the mmpose root directory:

```
python demo/webcam_demo.py --config demo/webcam_cfg/gesture_recognition.py
```

### 4.11.2 Hotkeys

Note that the demo will automatically save the output video into a file `gesture.mp4`.

### 4.11.3 Configurations

Detailed configurations can be found in the [config file](#). And more information about the gesture recognition model used in the demo can be found at the [model page](#).

## 4.12 Webcam Demo

We provide a webcam demo tool which integrates detection and 2D pose estimation for humans and animals. It can also apply fun effects like putting on sunglasses or enlarging the eyes, based on the pose estimation results.

### 4.12.1 Get started

Launch the demo from the mmpose root directory:

```
## Run webcam demo with GPU
python demo/webcam_demo.py

## Run webcam demo with CPU
python demo/webcam_demo.py --cpu
```

The command above will use the default config file `demo/webcam_cfg/pose_estimation.py`. You can also specify the config file in the command:

```
## Use the config "pose_tracking.py" for higher inference speed
python demo/webcam_demo.py --config demo/webcam_cfg/pose_tracking.py
```

### 4.12.2 Hotkeys

Note that the demo will automatically save the output video into a file `webcam_demo.mp4`.

### 4.12.3 Usage and configurations

Detailed configurations can be found in the config file.

- **Configure detection models** Users can choose detection models from the [MMDetection Model Zoo](#). Just set the `model_config` and `model_checkpoint` in the detector node accordingly, and the model will be automatically downloaded and loaded.

```
## 'DetectorNode':
## This node performs object detection from the frame image using an
## MMDetection model.
dict(
    type='DetectorNode',
    name='detector',
    model_config='demo/mmdetection_cfg/'
    'ssdlite_mobilenetv2_scratch_600e_coco.py',
    model_checkpoint='https://download.openmmlab.com'
    '/mmdetection/v2.0/ssd/'
    'ssdlite_mobilenetv2_scratch_600e_coco/ssdlite_mobilenetv2_'
    'scratch_600e_coco_20210629_110627-974d9307.pth',
    input_buffer='_input_',
    output_buffer='det_result')
```

- **Configure pose estimation models** In this demo we use two **top-down** pose estimation models for humans and animals respectively. Users can choose models from the **MMPose Model Zoo**. To apply different pose models on different instance types, you can add multiple pose estimator nodes with `cls_names` set accordingly.

```
## 'TopDownPoseEstimatorNode':
## This node performs keypoint detection from the frame image using an
## MMPose top-down model. Detection results is needed.
dict(
    type='TopDownPoseEstimatorNode',
    name='human pose estimator',
    model_config='configs/wholebody/2d_kpt_sview_rgb_img/'
    'topdown_heatmap/coco-wholebody/'
    'vipnas_mbv3_coco_wholebody_256x192_dark.py',
    model_checkpoint='https://openmmlab-share.oss-cn-hangz'
    'hou.aliyuncs.com/mmpose/top_down/vipnas/vipnas_mbv3_co'
    'co_wholebody_256x192_dark-e2158108_20211205.pth',
    labels=['person'],
    input_buffer='det_result',
    output_buffer='human_pose'),
dict(
    type='TopDownPoseEstimatorNode',
    name='animal pose estimator',
    model_config='configs/animal/2d_kpt_sview_rgb_img/topdown_heatmap'
    '/animalpose/hrnet_w32_animalpose_256x256.py',
    model_checkpoint='https://download.openmmlab.com/mmpose/animal/'
    'hrnet/hrnet_w32_animalpose_256x256-1aa7f075_20210426.pth',
    labels=['cat', 'dog', 'horse', 'sheep', 'cow'],
    input_buffer='human_pose',
    output_buffer='animal_pose')
```

- **Run the demo on a local video file** You can use local video files as the demo input by set `camera_id` to the file path.
- **The computer doesn't have a camera?** A smart phone can serve as a webcam via apps like **Camo** or **DroidCam**.
- **Test the camera and display** Run follow command for a quick test of video capturing and displaying.

```
python demo/webcam_demo.py --config demo/webcam_cfg/test_camera.py
```



# CHAPTER 5

---

## 基准测试

---

内容建设中……



这里总结了 MMPose 中主要模型的复杂度信息和推理速度，包括模型的计算复杂度、参数数量，以及以不同的批处理大小在 CPU 和 GPU 上的推理速度。还比较了不同模型在 COCO 人体关键点数据集上的全类别平均正确率，展示了模型性能和模型复杂度之间的折中。

### 6.1 比较规则

为了保证比较的公平性，在相同的硬件和软件环境下使用相同的数据集进行了比较实验。还列出了模型在 COCO 人体关键点数据集上的全类别平均正确率以及相应的配置文件。

对于模型复杂度信息，计算具有相应输入形状的模型的浮点数运算次数和参数数量。请注意，当前某些网络层或算子还未支持，如 DeformConv2d，因此您可能需要检查是否所有操作都已支持，并验证浮点数运算次数和参数数量的计算是否正确。

对于推理速度，忽略了数据预处理的时间，只测量模型前向计算和数据后处理的时间。对于每个模型设置，保持相同的数据预处理方法，以确保相同的特征输入。分别测量了在 CPU 和 GPU 设备上的推理速度。对于自上而下的热图模型，我们还测试了批处理量较大（例如，10）情况，以测试拥挤场景下的模型性能。

推断速度是用每秒处理的帧数 (FPS) 来衡量的，即每秒模型的平均迭代次数，它可以显示模型处理输入的速度。这个数值越高，表示推理速度越快，模型性能越好。

### 6.1.1 硬件

- GPU: GeForce GTX 1660 SUPER
- CPU: Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz

### 6.1.2 软件环境

- Ubuntu 16.04
- Python 3.8
- PyTorch 1.10
- CUDA 10.2
- mmcv-full 1.3.17
- mmpose 0.20.0

## 6.2 MMPose 中主要模型的复杂度信息和推理速度总结

1 注意，这里运行迭代多次，并记录每次迭代的时间，同时展示了 FPS 数值的平均值和标准差。

2 FPS 定义为每秒的平均迭代次数，与此迭代中的批处理大小无关。



- 论文数量: 9
  - DATASET: 9

已支持的算法详细信息请见 [模型池](#).

## 7.1 2D 动物关键点数据集

- 论文数量: 0

## 7.2 2D 人体关键点数据集

- 论文数量: 9
  - [DATASET] 2d Human Pose Estimation: New Benchmark and State of the Art Analysis (MPII  $\Rightarrow$ )
  - [DATASET] Ai Challenger: A Large-Scale Dataset for Going Deeper in Image Understanding (AIC  $\Rightarrow$ )
  - [DATASET] Crowdpose: Efficient Crowded Scenes Pose Estimation and a New Benchmark (CrowdPose  $\Rightarrow$ )
  - [DATASET] Learning Delicate Local Representations for Multi-Person Pose Estimation (sub-JHMDB dataset  $\Rightarrow$ )
  - [DATASET] Microsoft Coco: Common Objects in Context (COCO  $\Rightarrow$ )
  - [DATASET] Pose2seg: Detection Free Human Instance Segmentation (OCHuman  $\Rightarrow$ )

- [DATASET] Posetrack: A Benchmark for Human Pose Estimation and Tracking (PoseTrack18  $\Rightarrow$ )
- [DATASET] Trb: A Novel Triplet Representation for Understanding 2d Human Body (MPII-TRB  $\Rightarrow$ )
- [DATASET] Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and a New Benchmark for Multi-Human Parsing (MHP  $\Rightarrow$ )

## 7.3 2D 人脸关键点数据集

- 论文数量: 0

## 7.4 2D 服装关键点数据集

- 论文数量: 0

## 7.5 2D 手部关键点数据集

- 论文数量: 0

## 7.6 2D 全身人体关键点数据集

- 论文数量: 0

## 7.7 3D 人体关键点数据集

- 论文数量: 0

## 7.8 3D 人体网格模型数据集

- 论文数量: 0

## 7.9 3D 手部关键点数据集

- 论文数量: 0



---

### 2D 人体关键点数据集

---

我们建议您将数据集的根目录放置在 `$MMPOSE/data` 下。如果您的文件结构比较特别，您需要在配置文件中修改相应的路径。

MMPose 支持的数据集如下所示：

- 图像
  - *COCO* [ 主页 ]
  - *MPII* [ 主页 ]
  - *MPII-TRB* [ 主页 ]
  - *AI Challenger* [ 主页 ]
  - *CrowdPose* [ 主页 ]
  - *OCHuman* [ 主页 ]
  - *MHP* [ 主页 ]
- 视频
  - *PoseTrack18* [ 主页 ]
  - *sub-JHMDB* [ 主页 ]

## 8.1 COCO

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

请从此链接 [COCO download](#) 下载数据集。请注意，2017 Train/Val 对于 COCO 关键点的训练和评估是非常必要的。[HRNet-Human-Pose-Estimation](#) 提供了 COCO val2017 的检测结果，可用于复现我们的多人姿态估计的结果。请从 [OneDrive](#) 或 [GoogleDrive](#) 下载。可选地，为了在 COCO' 2017 test-dev 上评估，请下载 [image-info](#)。请将数据置于 \$MMPose/data 目录下，并整理成如下的格式：

```
mmpose
├─ mmpose
├─ docs
├─ tests
├─ tools
├─ configs
└─ data
    └─ coco
        ├── annotations
        │   ├── person_keypoints_train2017.json
        │   ├── person_keypoints_val2017.json
        │   └── person_keypoints_test-dev-2017.json
        ├── person_detection_results
        │   ├── COCO_val2017_detections_AP_H_56_person.json
        │   └── COCO_test-dev2017_detections_AP_H_609_person.json
        ├── train2017
        │   ├── 000000000009.jpg
        │   ├── 000000000025.jpg
        │   ├── 000000000030.jpg
        │   └── ...
        └── val2017
            ├── 000000000139.jpg
            ├── 000000000285.jpg
            ├── 000000000632.jpg
            └── ...
```

## 8.2 MPII

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

请从此链接 [MPII Human Pose Dataset](#) 下载数据集。我们已经将原来的标注文件转成了 json 格式，请从此链接 [mpii\\_annotations](#) 下载。请将数据置于 \$MMPOSE/data 目录下，并整理成如下的格式：

```
mmpose
├─ mmpose
├─ docs
├─ tests
├─ tools
├─ configs
└─ data
    └─ mpii
        ├── annotations
        │   ├── mpii_gt_val.mat
        │   ├── mpii_test.json
        │   ├── mpii_train.json
        │   ├── mpii_trainval.json
        │   └─ mpii_val.json
        └─ images
            ├── 000001163.jpg
            └─ 000003072.jpg
```

在训练和推理过程中，预测结果将会被默认保存为 ‘.mat’ 的格式。我们提供了一个工具将这种 ‘.mat’ 的格式转换成更加易读的 ‘.json’ 格式。

```
python tools/dataset/mat2json ${PRED_MAT_FILE} ${GT_JSON_FILE} ${OUTPUT_PRED_JSON_
↪FILE}
```

比如，

```
python tools/dataset/mat2json work_dirs/res50_mpii_256x256/pred.mat data/mpii/
↪annotations/mpii_val.json pred.json
```

## 8.3 MPII-TRB

```
@inproceedings{duan2019trb,
  title={TRB: A Novel Triplet Representation for Understanding 2D Human Body},
  author={Duan, Haodong and Lin, Kwan-Yee and Jin, Sheng and Liu, Wentao and Qian,
↵Chen and Ouyang, Wanli},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={9479--9488},
  year={2019}
}
```

请从此链接[MPII Human Pose Dataset](#)下载数据集，并从此链接 [mpii\\_trb\\_annotations](#) 下载标注文件。请将数据置于 \$MMPPOSE/data 目录下，并整理成如下的格式：

```
mmpose
├── mmpose
├── docs
├── tests
├── tools
├── configs
└── data
    ├── mpii
    │   ├── annotations
    │   │   ├── mpii_trb_train.json
    │   │   └── mpii_trb_val.json
    │   └── images
    │       ├── 000001163.jpg
    │       └── 000003072.jpg
```

## 8.4 AIC

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↵,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↵Liang, Rui and Wang, Wenjia and Zhou, Shiwei and Lin, Guosen and Fu, Yanwei and
↵others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

请从此链接 [AI Challenger 2017](#) 下载 AIC 数据集。请注意，2017 Train/Val 对于关键点的训练和评估是必要的。



请从此链接 [aic\\_annotations](#) 下载标注文件。请将数据置于 \$MMPOSE/data 目录下，并整理成如下的格式：

```

mmpose
├── mmpose
├── docs
├── tests
├── tools
├── configs
└── data
    ├── aic
    │   ├── annotations
    │   │   ├── aic_train.json
    │   │   ├── aic_val.json
    │   ├── ai_challenger_keypoint_train_20170902
    │   │   ├── keypoint_train_images_20170902
    │   │   │   ├── 0000252aea98840a550dac9a78c476ecb9f47ffa.jpg
    │   │   │   ├── 000050f770985ac9653198495ef9b5c82435d49c.jpg
    │   │   │   └── ...
    │   ├── ai_challenger_keypoint_validation_20170911
    │   │   ├── keypoint_validation_images_20170911
    │   │   │   ├── 0002605c53fb92109a3f2de4fc3ce06425c3b61f.jpg
    │   │   │   ├── 0003b55a2c991223e6d8b4b820045bd49507bf6d.jpg
    │   │   │   └── ...

```

## 8.5 CrowdPose

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

请从此链接 [CrowdPose](#) 下载数据集，并从此链接 [crowdpose\\_annotations](#) 下载标注文件和人体检测结果。对于 top-down 方法，我们仿照 [CrowdPose](#)，使用 YOLOv3 的预训练权重 来产生人体的检测框。对于模型训练，我们仿照 [HigherHRNet](#)，在 CrowdPose 训练/验证数据集上训练模型，并在 CrowdPose 测试集上评估模型。请将数据置于 \$MMPOSE/data 目录下，并整理成如下的格式：

```

mmpose
├── mmpose
├── docs
├── tests

```

(下页继续)

(续上页)

```

├─ tools
├─ configs
└─ data
    └─ crowdpose
        ├── annotations
        │   ├── mmpose_crowdpose_train.json
        │   ├── mmpose_crowdpose_val.json
        │   ├── mmpose_crowdpose_trainval.json
        │   ├── mmpose_crowdpose_test.json
        │   └── det_for_crowd_test_0.1_0.5.json
        ├── images
        │   ├── 100000.jpg
        │   ├── 100001.jpg
        │   ├── 100002.jpg
        │   └── ...

```

## 8.6 OCHuman

```

@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},
  author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↪and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={889--898},
  year={2019}
}

```

请从此链接 [OCHuman](#) 下载数据集的图像和标注文件。请将数据置于 \$MMPose/data 目录下，并整理成如下格式：

```

mmpose
├─ mmpose
├─ docs
├─ tests
├─ tools
├─ configs
└─ data
    └─ ochuman
        ├── annotations
        │   └── ochuman_coco_format_val_range_0.00_1.00.json

```

(下页继续)

(续上页)

```

| |-- ochuman_coco_format_test_range_0.00_1.00.json
|-- images
|   |-- 000001.jpg
|   |-- 000002.jpg
|   |-- 000003.jpg
|   |-- ...

```

## 8.7 MHP

```

@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan, ↪
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}

```

请从此链接 [MHP](#) 下载数据文件，并从此链接 [mhp\\_annotations](#) 下载标注文件。请将数据置于 \$MMPOSE/data 目录下，并整理成如下的格式：

```

mmpose
├── mmpose
├── docs
├── tests
├── tools
├── configs
└── data
    ├── mhp
    │   ├── annotations
    │   │   ├── mhp_train.json
    │   │   ├── mhp_val.json
    │   │   └──
    │   ├── train
    │   │   ├── images
    │   │   │   ├── 1004.jpg
    │   │   │   ├── 10050.jpg
    │   │   │   └── ...
    │   │   └──
    │   └── val

```

(下页继续)

(续上页)

```

|   |-- images
|   |   |-- 10059.jpg
|   |   |-- 10068.jpg
|   |   |-- ...
|
|-- test
|   |-- images
|   |   |-- 1005.jpg
|   |   |-- 10052.jpg
|   |   |-- ...~~~~~

```

## 8.8 PoseTrack18

```

@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin, Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={5167--5176},
  year={2018}
}

```

请从此链接 [PoseTrack18](#) 下载数据文件，并从此链接 [posetrack18\\_annotations](#) 下载标注文件。我们已将官方提供的所有单视频标注文件合并为两个 json 文件 (posetrack18\_train & posetrack18\_val.json)，并生成了 [mask files](#) 来加速训练。对于 top-down 的方法，我们使用 [MMDetection](#) 的预训练 [Cascade R-CNN \(X-101-64x4d-FPN\)](#) 来生成人体的检测框。请将数据置于 \$MMPOSE/data 目录下，并整理成如下的格式：

```

mmpose
├── mmpose
├── docs
├── tests
├── tools
├── configs
└── data
    ├── posetrack18
    │   ├── annotations
    │   │   ├── posetrack18_train.json
    │   │   ├── posetrack18_val.json
    │   │   ├── posetrack18_val_human_detections.json
    │   └── train

```

(下页继续)

(续上页)

```

| | | -- 000001_bonn_train.json
| | | -- 000002_bonn_train.json
| | | -- ...
| | -- val
| | | -- 000342_mpii_test.json
| | | -- 000522_mpii_test.json
| | | -- ...
| | `-- test
| | | -- 000001_mpiinew_test.json
| | | -- 000002_mpiinew_test.json
| | | -- ...
| |
| | `-- images
| | | -- train
| | | | -- 000001_bonn_train
| | | | | -- 000000.jpg
| | | | | -- 000001.jpg
| | | | | -- ...
| | | | -- ...
| | | -- val
| | | | -- 000342_mpii_test
| | | | | -- 000000.jpg
| | | | | -- 000001.jpg
| | | | | -- ...
| | | | -- ...
| | | `-- test
| | | | -- 000001_mpiinew_test
| | | | | -- 000000.jpg
| | | | | -- 000001.jpg
| | | | | -- ...
| | | | -- ...
| |
| | `-- mask
| | | -- train
| | | | -- 000002_bonn_train
| | | | | -- 000000.jpg
| | | | | -- 000001.jpg
| | | | | -- ...
| | | | -- ...
| | | `-- val
| | | | -- 000522_mpii_test
| | | | | -- 000000.jpg
| | | | | -- 000001.jpg
| | | | | -- ...
| | | | -- ...

```

请从 Github 上安装 PoseTrack 官方评估工具：

```
pip install git+https://github.com/svenkreiss/poseval.git
```

## 8.9 sub-JHMDB dataset

```
@misc{cai2020learning,
  title={Learning Delicate Local Representations for Multi-Person Pose Estimation},
  author={Yuanhao Cai and Zhicheng Wang and Zhengxiong Luo and Binyi Yin and Angang
↪Du and Haoqian Wang and Xinyu Zhou and Erjin Zhou and Xiangyu Zhang and Jian Sun},
  year={2020},
  eprint={2003.04030},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}
```

对于 sub-JHMDB 数据集，请从此链接 [JHMDB](#) 下载 images，请从此链接 [jhmdb\\_annotations](#) 下载标注文件。将它们移至 \$MMPPOSE/data 目录下，使得文件呈如下的格式：

```
mmpose
├── mmpose
├── docs
├── tests
├── tools
├── configs
└── data
    ├── jhmdb
    │   ├── annotations
    │   │   ├── Sub1_train.json
    │   │   ├── Sub1_test.json
    │   │   ├── Sub2_train.json
    │   │   ├── Sub2_test.json
    │   │   ├── Sub3_train.json
    │   │   └── Sub3_test.json
    │   ├── Rename_Images
    │   │   ├── brush_hair
    │   │   │   ├── April_09_brush_hair_u_nm_np1_ba_goo_0
    │   │   │   ├── 00001.png
    │   │   │   └── 00002.png
    │   │   ├── catch
    │   │   └── ...
```

## CHAPTER 9

---

### 2D 全身人体关键点数据集

---

内容建设中……





## CHAPTER 10

---

### 2D 人脸关键点数据集

---

内容建设中……



## CHAPTER 11

---

### 2D 手部关键点数据集

---

内容建设中……



## CHAPTER 12

---

### 2D 服装关键点数据集

---

内容建设中……



## CHAPTER 13

---

### 2D 动物关键点数据集

---

内容建设中……





## CHAPTER 14

---

### 3D 人体关键点数据集

---

内容建设中……



## CHAPTER 15

---

### 3D 人体网格模型数据集

---

内容建设中……



## CHAPTER 16

---

### 3D 手部关键点数据集

---

内容建设中……



- 模型权重文件数量: 322
- 配置文件数量: 342
- 论文数量: 82
  - ALGORITHM: 30
  - BACKBONE: 15
  - DATASET: 34
  - OTHERS: 3

已支持的数据集详细信息请见[数据集](#).

## 17.1 Animal

- 模型权重文件数量: 43
- 配置文件数量: 43
- 论文数量: 9
  - [ALGORITHM] Deep High-Resolution Representation Learning for Human Pose Estimation (Topdown Heatmap + Hrnet on Macaque  $\Rightarrow$ , Topdown Heatmap + Hrnet on Horse10  $\Rightarrow$ , Topdown Heatmap + Hrnet on Atrw  $\Rightarrow$ , Topdown Heatmap + Hrnet on Ap10k  $\Rightarrow$ , Topdown Heatmap + Hrnet on Animalpose  $\Rightarrow$ )

- [ALGORITHM] Simple Baselines for Human Pose Estimation and Tracking (Topdown Heatmap + Resnet on Zebra  $\Rightarrow$ , Topdown Heatmap + Resnet on Macaque  $\Rightarrow$ , Topdown Heatmap + Resnet on Locust  $\Rightarrow$ , Topdown Heatmap + Resnet on Horse10  $\Rightarrow$ , Topdown Heatmap + Resnet on Fly  $\Rightarrow$ , Topdown Heatmap + Resnet on Atrw  $\Rightarrow$ , Topdown Heatmap + Resnet on Ap10k  $\Rightarrow$ , Topdown Heatmap + Resnet on Animalpose  $\Rightarrow$ )
- [DATASET] Ap-10k: A Benchmark for Animal Pose Estimation in the Wild (Topdown Heatmap + Resnet on Ap10k  $\Rightarrow$ , Topdown Heatmap + Hrnet on Ap10k  $\Rightarrow$ )
- [DATASET] Atrw: A Benchmark for Amur Tiger Re-Identification in the Wild (Topdown Heatmap + Resnet on Atrw  $\Rightarrow$ , Topdown Heatmap + Hrnet on Atrw  $\Rightarrow$ )
- [DATASET] Cross-Domain Adaptation for Animal Pose Estimation (Topdown Heatmap + Resnet on Animalpose  $\Rightarrow$ , Topdown Heatmap + Hrnet on Animalpose  $\Rightarrow$ )
- [DATASET] Deepposekit, a Software Toolkit for Fast and Robust Animal Pose Estimation Using Deep Learning (Topdown Heatmap + Resnet on Zebra  $\Rightarrow$ , Topdown Heatmap + Resnet on Locust  $\Rightarrow$ )
- [DATASET] Fast Animal Pose Estimation Using Deep Neural Networks (Topdown Heatmap + Resnet on Fly  $\Rightarrow$ )
- [DATASET] Macaquepose: A Novel ‘In the Wild’ macaque Monkey Pose Dataset for Markerless Motion Capture (Topdown Heatmap + Hrnet on Macaque  $\Rightarrow$ , Topdown Heatmap + Resnet on Macaque  $\Rightarrow$ )
- [DATASET] Pretraining Boosts Out-of-Domain Robustness for Pose Estimation (Topdown Heatmap + Hrnet on Horse10  $\Rightarrow$ , Topdown Heatmap + Resnet on Horse10  $\Rightarrow$ )

## 17.2 Body(2D,Kpt,Sview,Img)

- 模型权重文件数量: 184
- 配置文件数量: 192
- 论文数量: 45
  - [ALGORITHM] Associative Embedding: End-to-End Learning for Joint Detection and Grouping (Associative Embedding + Hrnet on MHP  $\Rightarrow$ , Associative Embedding + Higherhrnet on Crowdpose  $\Rightarrow$ , Associative Embedding + Hourglass + Ae on Coco  $\Rightarrow$ , Associative Embedding + Hrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Resnet on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet on Coco  $\Rightarrow$ , Associative Embedding + Hrnet on Coco  $\Rightarrow$ , Associative Embedding + Mobilenetv2 on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet on Aic  $\Rightarrow$ , Associative Embedding + Hrnet on Aic  $\Rightarrow$ )
  - [ALGORITHM] Bottom-Up Human Pose Estimation via Disentangled Keypoint Regression (Dekr + Hrnet on Crowdpose  $\Rightarrow$ , Dekr + Hrnet on Coco  $\Rightarrow$ )
  - [ALGORITHM] Contextual Instance Decoupling for Robust Multi-Person Pose Estimation (Cid + Hrnet on Coco  $\Rightarrow$ )



- [ALGORITHM] Convolutional Pose Machines (Topdown Heatmap + CPM on Mpii  $\Rightarrow$ , Topdown Heatmap + CPM on JHMDB  $\Rightarrow$ , Topdown Heatmap + CPM on Coco  $\Rightarrow$ )
- [ALGORITHM] Deep High-Resolution Representation Learning for Human Pose Estimation (Topdown Heatmap + Hrnet on Posetrack18  $\Rightarrow$ , Topdown Heatmap + Hrnet on Ochuman  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Mpii  $\Rightarrow$ , Topdown Heatmap + Hrnet on Mpii  $\Rightarrow$ , Associative Embedding + Hrnet on MHP  $\Rightarrow$ , Topdown Heatmap + Hrnet on H36m  $\Rightarrow$ , Topdown Heatmap + Hrnet on Crowdpose  $\Rightarrow$ , Dekr + Hrnet on Crowdpose  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet + Udp on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet + Augmentation on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet + Fp16 on Coco  $\Rightarrow$ , Dekr + Hrnet on Coco  $\Rightarrow$ , Associative Embedding + Hrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Hrnet on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet on Aic  $\Rightarrow$ , Associative Embedding + Hrnet on Aic  $\Rightarrow$ )
- [ALGORITHM] Deeppose: Human Pose Estimation via Deep Neural Networks (Deeppose + Resnet + Rle on Mpii  $\Rightarrow$ , Deeppose + Resnet on Mpii  $\Rightarrow$ , Deeppose + Resnet + Rle on Coco  $\Rightarrow$ , Deeppose + Resnet on Coco  $\Rightarrow$ )
- [ALGORITHM] Distribution-Aware Coordinate Representation for Human Pose Estimation (Topdown Heatmap + Hrnet + Dark on Mpii  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Coco  $\Rightarrow$ , Topdown Heatmap + Resnet + Dark on Coco  $\Rightarrow$ )
- [ALGORITHM] Higherhrnet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation (Associative Embedding + Higherhrnet on Crowdpose  $\Rightarrow$ , Associative Embedding + Higherhrnet on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet on Aic  $\Rightarrow$ )
- [ALGORITHM] Hrformer: High-Resolution Vision Transformer for Dense Predict (Topdown Heatmap + Hrformer on Coco  $\Rightarrow$ )
- [ALGORITHM] Human Pose Regression With Residual Log-Likelihood Estimation (Deeppose + Resnet + Rle on Mpii  $\Rightarrow$ , Deeppose + Resnet + Rle on Coco  $\Rightarrow$ )
- [ALGORITHM] Improving Convolutional Networks With Self-Calibrated Convolutions (Topdown Heatmap + Scnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Scnet on Coco  $\Rightarrow$ )
- [ALGORITHM] Learning Delicate Local Representations for Multi-Person Pose Estimation (Topdown Heatmap + RSN on Coco  $\Rightarrow$ )
- [ALGORITHM] Lite-Hrnet: A Lightweight High-Resolution Network (Topdown Heatmap + Litehrnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Litehrnet on Coco  $\Rightarrow$ )
- [ALGORITHM] Rethinking on Multi-Stage Networks for Human Pose Estimation (Topdown Heatmap + MSPN on Coco  $\Rightarrow$ )
- [ALGORITHM] Simple Baselines for Human Pose Estimation and Tracking (Topdown Heatmap + Resnet on Posetrack18  $\Rightarrow$ , Topdown Heatmap + Resnet on Ochuman  $\Rightarrow$ , Topdown Heatmap + Resnet + Mpii on Mpii\_trb  $\Rightarrow$ , Topdown Heatmap + Resnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Resnet on MHP  $\Rightarrow$ , Topdown Heatmap + Resnet on JHMDB  $\Rightarrow$ , Topdown Heatmap + Resnet on Crowdpose  $\Rightarrow$ , Topdown Heatmap +

- Resnet on Coco  $\Rightarrow$ , Topdown Heatmap + Swin on Coco  $\Rightarrow$ , Topdown Heatmap + Resnet + Dark on Coco  $\Rightarrow$ , Topdown Heatmap + Resnet + Fp16 on Coco  $\Rightarrow$ , Topdown Heatmap + Resnet on Aic  $\Rightarrow$ )
- [ALGORITHM] Stacked Hourglass Networks for Human Pose Estimation (Topdown Heatmap + Hourglass on Mpii  $\Rightarrow$ , Topdown Heatmap + Hourglass on Coco  $\Rightarrow$ )
  - [ALGORITHM] The Devil Is in the Details: Delving Into Unbiased Data Processing for Human Pose Estimation (Topdown Heatmap + Hrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Hrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet + Udp on Coco  $\Rightarrow$ )
  - [ALGORITHM] Vipnas: Efficient Video Pose Estimation via Neural Architecture Search (Topdown Heatmap + Vipnas on Coco  $\Rightarrow$ )
  - [BACKBONE] Aggregated Residual Transformations for Deep Neural Networks (Topdown Heatmap + Resnext on Mpii  $\Rightarrow$ , Topdown Heatmap + Resnext on Coco  $\Rightarrow$ )
  - [BACKBONE] Associative Embedding: End-to-End Learning for Joint Detection and Grouping (Associative Embedding + Hrnet on MHP  $\Rightarrow$ , Associative Embedding + Higherhrnet on Crowdpose  $\Rightarrow$ , Associative Embedding + Hourglass + Ae on Coco  $\Rightarrow$ , Associative Embedding + Hrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Resnet on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet on Coco  $\Rightarrow$ , Associative Embedding + Hrnet on Coco  $\Rightarrow$ , Associative Embedding + Mobilenetv2 on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Higherhrnet on Aic  $\Rightarrow$ , Associative Embedding + Hrnet on Aic  $\Rightarrow$ )
  - [BACKBONE] Bag of Tricks for Image Classification With Convolutional Neural Networks (Topdown Heatmap + Resnetv1d on Mpii  $\Rightarrow$ , Topdown Heatmap + Resnetv1d on Coco  $\Rightarrow$ )
  - [BACKBONE] Deep High-Resolution Representation Learning for Human Pose Estimation (Topdown Heatmap + Hrnet on Posetrack18  $\Rightarrow$ , Topdown Heatmap + Hrnet on Ochuman  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Mpii  $\Rightarrow$ , Topdown Heatmap + Hrnet on Mpii  $\Rightarrow$ , Associative Embedding + Hrnet on MHP  $\Rightarrow$ , Topdown Heatmap + Hrnet on H36m  $\Rightarrow$ , Topdown Heatmap + Hrnet on Crowdpose  $\Rightarrow$ , Dekr + Hrnet on Crowdpose  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet + Udp on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet + Augmentation on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet + Fp16 on Coco  $\Rightarrow$ , Dekr + Hrnet on Coco  $\Rightarrow$ , Associative Embedding + Hrnet + Udp on Coco  $\Rightarrow$ , Associative Embedding + Hrnet on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet on Aic  $\Rightarrow$ , Associative Embedding + Hrnet on Aic  $\Rightarrow$ )
  - [BACKBONE] Deep Residual Learning for Image Recognition (Topdown Heatmap + Resnet on Posetrack18  $\Rightarrow$ , Topdown Heatmap + Resnet on Ochuman  $\Rightarrow$ , Topdown Heatmap + Resnet + Mpii on Mpii\_trb  $\Rightarrow$ , Topdown Heatmap + Resnet on Mpii  $\Rightarrow$ , Deeppose + Resnet + Rle on Mpii  $\Rightarrow$ , Deeppose + Resnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Resnet on MHP  $\Rightarrow$ , Topdown Heatmap + Resnet on JHMDB  $\Rightarrow$ , Topdown Heatmap + Resnet on Crowdpose  $\Rightarrow$ , Topdown Heatmap + Resnet on Coco  $\Rightarrow$ , Topdown Heatmap + Resnet + Dark on Coco  $\Rightarrow$ , Topdown Heatmap + Resnet + Fp16 on Coco  $\Rightarrow$ , Deeppose + Resnet + Rle on Coco  $\Rightarrow$ , Deeppose + Resnet on Coco  $\Rightarrow$ , Associative Embedding + Resnet on Coco  $\Rightarrow$ , Topdown Heatmap + Resnet on Aic  $\Rightarrow$ )
  - [BACKBONE] Imagenet Classification With Deep Convolutional Neural Networks (Topdown Heatmap + Alexnet on Coco  $\Rightarrow$ )

- [BACKBONE] Mobilenetv2: Inverted Residuals and Linear Bottlenecks (Topdown Heatmap + Mobilenetv2 on Mpii  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 on Coco  $\Rightarrow$ , Associative Embedding + Mobilenetv2 on Coco  $\Rightarrow$ )
- [BACKBONE] Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction Without Convolutions (Topdown Heatmap + PVT on Coco  $\Rightarrow$ )
- [BACKBONE] Resnest: Split-Attention Networks (Topdown Heatmap + Resnest on Coco  $\Rightarrow$ )
- [BACKBONE] Shufflenet V2: Practical Guidelines for Efficient CNN Architecture Design (Topdown Heatmap + Shufflenetv2 on Mpii  $\Rightarrow$ , Topdown Heatmap + Shufflenetv2 on Coco  $\Rightarrow$ )
- [BACKBONE] Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices (Topdown Heatmap + Shufflenetv1 on Mpii  $\Rightarrow$ , Topdown Heatmap + Shufflenetv1 on Coco  $\Rightarrow$ )
- [BACKBONE] Squeeze-and-Excitation Networks (Topdown Heatmap + Seresnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Seresnet on Coco  $\Rightarrow$ )
- [BACKBONE] Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows (Topdown Heatmap + Swin on Coco  $\Rightarrow$ )
- [BACKBONE] Very Deep Convolutional Networks for Large-Scale Image Recognition (Topdown Heatmap + VGG on Coco  $\Rightarrow$ )
- [DATASET] 2d Human Pose Estimation: New Benchmark and State of the Art Analysis (Topdown Heatmap + Hourglass on Mpii  $\Rightarrow$ , Topdown Heatmap + Resnet1d on Mpii  $\Rightarrow$ , Topdown Heatmap + Seresnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Scnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Mpii  $\Rightarrow$ , Topdown Heatmap + Shufflenetv1 on Mpii  $\Rightarrow$ , Topdown Heatmap + Resnext on Mpii  $\Rightarrow$ , Topdown Heatmap + Litehrnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Shufflenetv2 on Mpii  $\Rightarrow$ , Topdown Heatmap + CPM on Mpii  $\Rightarrow$ , Topdown Heatmap + Resnet on Mpii  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 on Mpii  $\Rightarrow$ , Topdown Heatmap + Hrnet on Mpii  $\Rightarrow$ , Deeppose + Resnet + Rle on Mpii  $\Rightarrow$ , Deeppose + Resnet on Mpii  $\Rightarrow$ )
- [DATASET] Ai Challenger: A Large-Scale Dataset for Going Deeper in Image Understanding (Topdown Heatmap + Resnet on Aic  $\Rightarrow$ , Topdown Heatmap + Hrnet on Aic  $\Rightarrow$ , Associative Embedding + Higherhrnet on Aic  $\Rightarrow$ , Associative Embedding + Hrnet on Aic  $\Rightarrow$ )
- [DATASET] Crowdpose: Efficient Crowded Scenes Pose Estimation and a New Benchmark (Topdown Heatmap + Resnet on Crowdpose  $\Rightarrow$ , Topdown Heatmap + Hrnet on Crowdpose  $\Rightarrow$ , Dekr + Hrnet on Crowdpose  $\Rightarrow$ , Associative Embedding + Higherhrnet on Crowdpose  $\Rightarrow$ )
- [DATASET] Human3.6m: Large Scale Datasets and Predictive Methods for 3d Human Sensing in Natural Environments (Topdown Heatmap + Hrnet on H36m  $\Rightarrow$ )
- [DATASET] Microsoft Coco: Common Objects in Context (Cid + Hrnet on Coco  $\Rightarrow$ , Topdown Heatmap + CPM on Coco  $\Rightarrow$ , Topdown Heatmap + Vipnas on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet + Udp on Coco  $\Rightarrow$ , Topdown Heatmap + Alexnet on Coco  $\Rightarrow$ , Topdown Heatmap + Resnet on Coco  $\Rightarrow$ , Topdown Heatmap + PVT on Coco  $\Rightarrow$ , Topdown Heatmap + Resnest on Coco  $\Rightarrow$ , Topdown Heatmap + Hrformer on Coco  $\Rightarrow$ , Topdown Heatmap + Scnet on Coco  $\Rightarrow$ , Topdown Heatmap + Litehrnet on Coco  $\Rightarrow$ , Topdown Heatmap + Swin on Coco  $\Rightarrow$ , Topdown Heatmap + Hrnet +

- Augmentation on Coco ⇒, Topdown Heatmap + Shufflenetv2 on Coco ⇒, Topdown Heatmap + VGG on Coco ⇒, Topdown Heatmap + Resnet + Dark on Coco ⇒, Topdown Heatmap + Resnext on Coco ⇒, Topdown Heatmap + Hrnet on Coco ⇒, Topdown Heatmap + Hrnet + Fp16 on Coco ⇒, Topdown Heatmap + MSPN on Coco ⇒, Topdown Heatmap + Shufflenetv1 on Coco ⇒, Topdown Heatmap + Mobilenetv2 on Coco ⇒, Topdown Heatmap + Resnet + Fp16 on Coco ⇒, Topdown Heatmap + RSN on Coco ⇒, Topdown Heatmap + Hourglass on Coco ⇒, Topdown Heatmap + Resnetv1d on Coco ⇒, Topdown Heatmap + Seresnet on Coco ⇒, Dekr + Hrnet on Coco ⇒, Deeppose + Resnet + Rle on Coco ⇒, Deeppose + Resnet on Coco ⇒, Associative Embedding + Hourglass + Ae on Coco ⇒, Associative Embedding + Hrnet + Udp on Coco ⇒, Associative Embedding + Resnet on Coco ⇒, Associative Embedding + Higherhrnet on Coco ⇒, Associative Embedding + Hrnet on Coco ⇒, Associative Embedding + Mobilenetv2 on Coco ⇒, Associative Embedding + Higherhrnet + Udp on Coco ⇒)
- [DATASET] Pose2seg: Detection Free Human Instance Segmentation (Topdown Heatmap + Hrnet on Ochuman ⇒, Topdown Heatmap + Resnet on Ochuman ⇒)
  - [DATASET] Posetrack: A Benchmark for Human Pose Estimation and Tracking (Topdown Heatmap + Resnet on Posetrack18 ⇒, Topdown Heatmap + Hrnet on Posetrack18 ⇒)
  - [DATASET] Towards Understanding Action Recognition (Topdown Heatmap + Resnet on JHMDB ⇒, Topdown Heatmap + CPM on JHMDB ⇒)
  - [DATASET] Trb: A Novel Triplet Representation for Understanding 2d Human Body (Topdown Heatmap + Resnet + Mpii on Mpii\_trb ⇒)
  - [DATASET] Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and a New Benchmark for Multi-Human Parsing (Topdown Heatmap + Resnet on MHP ⇒, Associative Embedding + Hrnet on MHP ⇒)
  - [OTHERS] Albumentations: Fast and Flexible Image Augmentations (Topdown Heatmap + Hrnet + Augmentation on Coco ⇒)
  - [OTHERS] Feature Pyramid Networks for Object Detection (Topdown Heatmap + Swin on Coco ⇒)
  - [OTHERS] Mixed Precision Training (Topdown Heatmap + Hrnet + Fp16 on Coco ⇒, Topdown Heatmap + Resnet + Fp16 on Coco ⇒)

## 17.3 Body(2D,Kpt,Sview,Vid)

- 模型权重文件数量: 3
- 配置文件数量: 2
- 论文数量: 4
  - [ALGORITHM] Deep High-Resolution Representation Learning for Human Pose Estimation (Posewarper + Hrnet + Posetrack18 on Posetrack18 ⇒)

- [ALGORITHM] Learning Temporal Pose Estimation From Sparsely Labeled Videos (Posewarper + Hrnet + Posetrack18 on Posetrack18  $\Rightarrow$ )
- [DATASET] Microsoft Coco: Common Objects in Context (Posewarper + Hrnet + Posetrack18 on Posetrack18  $\Rightarrow$ )
- [DATASET] Posetrack: A Benchmark for Human Pose Estimation and Tracking (Posewarper + Hrnet + Posetrack18 on Posetrack18  $\Rightarrow$ )

## 17.4 Body(3D,Kpt,Mview,Img)

- 模型权重文件数量: 5
- 配置文件数量: 5
- 论文数量: 3
  - [ALGORITHM] Voxelpose: Towards Multi-Camera 3d Human Pose Estimation in Wild Environment (Voxelpose + Voxelpose on Shelf  $\Rightarrow$ , Voxelpose + Voxelpose + Prn64x64x64 + Cpn80x80x20 + Panoptic on Panoptic  $\Rightarrow$ , Voxelpose + Voxelpose on Campus  $\Rightarrow$ )
  - [DATASET] Panoptic Studio: A Massively Multiview System for Social Motion Capture (Voxelpose + Voxelpose + Prn64x64x64 + Cpn80x80x20 + Panoptic on Panoptic  $\Rightarrow$ )
  - [DATASET] {3d (Voxelpose + Voxelpose on Shelf  $\Rightarrow$ , Voxelpose + Voxelpose on Campus  $\Rightarrow$ )

## 17.5 Body(3D,Kpt,Sview,Img)

- 模型权重文件数量: 2
- 配置文件数量: 2
- 论文数量: 3
  - [ALGORITHM] A Simple Yet Effective Baseline for 3d Human Pose Estimation (Pose Lift + Simplebaseline3d on Mpi\_inf\_3dhp  $\Rightarrow$ , Pose Lift + Simplebaseline3d on H36m  $\Rightarrow$ )
  - [DATASET] Human3.6m: Large Scale Datasets and Predictive Methods for 3d Human Sensing in Natural Environments (Pose Lift + Simplebaseline3d on H36m  $\Rightarrow$ )
  - [DATASET] Monocular 3d Human Pose Estimation in the Wild Using Improved CNN Supervision (Pose Lift + Simplebaseline3d on Mpi\_inf\_3dhp  $\Rightarrow$ )

## 17.6 Body(3D,Kpt,Sview,Vid)

- 模型权重文件数量: 8
- 配置文件数量: 8
- 论文数量: 3
  - [ALGORITHM] 3d Human Pose Estimation in Video With Temporal Convolutions and Semi-Supervised Training (Video Pose Lift + Videopose3d on Mpi\_inf\_3dhp  $\Rightarrow$ , Video Pose Lift + Videopose3d on H36m  $\Rightarrow$ )
  - [DATASET] Human3.6m: Large Scale Datasets and Predictive Methods for 3d Human Sensing in Natural Environments (Video Pose Lift + Videopose3d on H36m  $\Rightarrow$ )
  - [DATASET] Monocular 3d Human Pose Estimation in the Wild Using Improved CNN Supervision (Video Pose Lift + Videopose3d on Mpi\_inf\_3dhp  $\Rightarrow$ )

## 17.7 Body(3D,Mesh,Sview,Img)

- 模型权重文件数量: 1
- 配置文件数量: 1
- 论文数量: 3
  - [ALGORITHM] End-to-End Recovery of Human Shape and Pose (HMR + Resnet on Mixed  $\Rightarrow$ )
  - [BACKBONE] Deep Residual Learning for Image Recognition (HMR + Resnet on Mixed  $\Rightarrow$ )
  - [DATASET] Human3.6m: Large Scale Datasets and Predictive Methods for 3d Human Sensing in Natural Environments (HMR + Resnet on Mixed  $\Rightarrow$ )

## 17.8 Face

- 模型权重文件数量: 16
- 配置文件数量: 16
- 论文数量: 16
  - [ALGORITHM] Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression (Topdown Heatmap + Hrnetv2 + Awing on WFLW  $\Rightarrow$ )
  - [ALGORITHM] Deep High-Resolution Representation Learning for Visual Recognition (Topdown Heatmap + Hrnetv2 on WFLW  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on WFLW  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Awing on WFLW  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 on Cofw  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Coco + Wholebody on

- Coco\_wholebody\_face  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 on Aflw  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Aflw  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 on 300w  $\Rightarrow$ )
- [ALGORITHM] Deeppose: Human Pose Estimation via Deep Neural Networks (Deeppose + Resnet + Wingloss on WFLW  $\Rightarrow$ , Deeppose + Resnet + Softwingloss on WFLW  $\Rightarrow$ , Deeppose + Resnet on WFLW  $\Rightarrow$ )
  - [ALGORITHM] Distribution-Aware Coordinate Representation for Human Pose Estimation (Topdown Heatmap + Hrnetv2 + Dark on WFLW  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Aflw  $\Rightarrow$ )
  - [ALGORITHM] Improving Convolutional Networks With Self-Calibrated Convolutions (Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ )
  - [ALGORITHM] Simple Baselines for Human Pose Estimation and Tracking (Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ )
  - [ALGORITHM] Stacked Hourglass Networks for Human Pose Estimation (Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ )
  - [ALGORITHM] Structure-Coherent Deep Feature Learning for Robust Face Alignment (Deeppose + Resnet + Softwingloss on WFLW  $\Rightarrow$ )
  - [ALGORITHM] Wing Loss for Robust Facial Landmark Localisation With Convolutional Neural Networks (Deeppose + Resnet + Wingloss on WFLW  $\Rightarrow$ )
  - [BACKBONE] Deep Residual Learning for Image Recognition (Deeppose + Resnet + Wingloss on WFLW  $\Rightarrow$ , Deeppose + Resnet + Softwingloss on WFLW  $\Rightarrow$ , Deeppose + Resnet on WFLW  $\Rightarrow$ , Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ )
  - [BACKBONE] Mobilenetv2: Inverted Residuals and Linear Bottlenecks (Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ )
  - [DATASET] 300 Faces in-the-Wild Challenge: Database and Results (Topdown Heatmap + Hrnetv2 on 300w  $\Rightarrow$ )
  - [DATASET] Annotated Facial Landmarks in the Wild: A Large-Scale, Real-World Database for Facial Landmark Localization (Topdown Heatmap + Hrnetv2 on Aflw  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Aflw  $\Rightarrow$ )
  - [DATASET] Look at Boundary: A Boundary-Aware Face Alignment Algorithm (Topdown Heatmap + Hrnetv2 on WFLW  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on WFLW  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Awing on WFLW  $\Rightarrow$ , Deeppose + Resnet + Wingloss on WFLW  $\Rightarrow$ , Deeppose + Resnet + Softwingloss on WFLW  $\Rightarrow$ , Deeppose + Resnet on WFLW  $\Rightarrow$ )
  - [DATASET] Robust Face Landmark Estimation Under Occlusion (Topdown Heatmap + Hrnetv2 on Cofw  $\Rightarrow$ )
  - [DATASET] Whole-Body Human Pose Estimation in the Wild (Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ , Topdown Heatmap + Hourglass + Coco + Wholebody

on Coco\_wholebody\_face  $\Rightarrow$ , Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ , Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_face  $\Rightarrow$ )

## 17.9 Fashion

- 模型权重文件数量: 6
- 配置文件数量: 6
- 论文数量: 5
  - [ALGORITHM] Deeppose: Human Pose Estimation via Deep Neural Networks (Deeppose + Resnet on Deepfashion  $\Rightarrow$ )
  - [ALGORITHM] Simple Baselines for Human Pose Estimation and Tracking (Topdown Heatmap + Resnet on Deepfashion  $\Rightarrow$ )
  - [BACKBONE] Deep Residual Learning for Image Recognition (Topdown Heatmap + Resnet on Deepfashion  $\Rightarrow$ , Deeppose + Resnet on Deepfashion  $\Rightarrow$ )
  - [DATASET] Deepfashion: Powering Robust Clothes Recognition and Retrieval With Rich Annotations (Topdown Heatmap + Resnet on Deepfashion  $\Rightarrow$ , Deeppose + Resnet on Deepfashion  $\Rightarrow$ )
  - [DATASET] Fashion Landmark Detection in the Wild (Topdown Heatmap + Resnet on Deepfashion  $\Rightarrow$ , Deeppose + Resnet on Deepfashion  $\Rightarrow$ )

## 17.10 Hand(2D,Kpt,Rgb,Img)

- 模型权重文件数量: 29
- 配置文件数量: 39
- 论文数量: 16
  - [ALGORITHM] Deep High-Resolution Representation Learning for Visual Recognition (Topdown Heatmap + Hrnetv2 on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Udp on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Udp on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )
  - [ALGORITHM] Deeppose: Human Pose Estimation via Deep Neural Networks (Deeppose + Resnet on Rhd2d  $\Rightarrow$ , Deeppose + Resnet on Panoptic2d  $\Rightarrow$ , Deeppose + Resnet on Onehand10k  $\Rightarrow$ )



- [ALGORITHM] Distribution-Aware Coordinate Representation for Human Pose Estimation (Topdown Heatmap + Hrnetv2 + Dark on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )
- [ALGORITHM] Improving Convolutional Networks With Self-Calibrated Convolutions (Topdown Heatmap + Snet + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )
- [ALGORITHM] Lite-Hrnet: A Lightweight High-Resolution Network (Topdown Heatmap + Litehrnet + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )
- [ALGORITHM] Simple Baselines for Human Pose Estimation and Tracking (Topdown Heatmap + Resnet on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Resnet on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Resnet on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Resnet on Interhand2d  $\Rightarrow$ , Topdown Heatmap + Resnet on Freihand2d  $\Rightarrow$ , Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )
- [ALGORITHM] Stacked Hourglass Networks for Human Pose Estimation (Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )
- [ALGORITHM] The Devil Is in the Details: Delving Into Unbiased Data Processing for Human Pose Estimation (Topdown Heatmap + Hrnetv2 + Udp on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Udp on Onehand10k  $\Rightarrow$ )
- [BACKBONE] Deep Residual Learning for Image Recognition (Topdown Heatmap + Resnet on Rhd2d  $\Rightarrow$ , Deeppose + Resnet on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Resnet on Panoptic2d  $\Rightarrow$ , Deeppose + Resnet on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Resnet on Onehand10k  $\Rightarrow$ , Deeppose + Resnet on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Resnet on Interhand2d  $\Rightarrow$ , Topdown Heatmap + Resnet on Freihand2d  $\Rightarrow$ , Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )
- [BACKBONE] Mobilenetv2: Inverted Residuals and Linear Bottlenecks (Topdown Heatmap + Mobilenetv2 on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )
- [DATASET] Freihand: A Dataset for Markerless Capture of Hand Pose and Shape From Single RGB Images (Topdown Heatmap + Resnet on Freihand2d  $\Rightarrow$ )
- [DATASET] Hand Keypoint Detection in Single Images Using Multiview Bootstrapping (Topdown Heatmap + Hrnetv2 on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Resnet on Panoptic2d  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 on Panoptic2d  $\Rightarrow$ , Deeppose + Resnet on Panoptic2d  $\Rightarrow$ )
- [DATASET] Interhand2.6m: A Dataset and Baseline for 3d Interacting Hand Pose Estimation From a Single RGB Image (Topdown Heatmap + Resnet on Interhand2d  $\Rightarrow$ )
- [DATASET] Learning to Estimate 3d Hand Pose From Single RGB Images (Topdown Heatmap + Hrnetv2 on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Udp on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Resnet on Rhd2d  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 on Rhd2d  $\Rightarrow$ , Deeppose + Resnet on Rhd2d  $\Rightarrow$ )

- [DATASET] Mask-Pose Cascaded CNN for 2d Hand Pose Estimation From Single Color Image (Topdown Heatmap + Hrnetv2 + Udp on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Resnet on Onehand10k  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark on Onehand10k  $\Rightarrow$ , Deeppose + Resnet on Onehand10k  $\Rightarrow$ )
- [DATASET] Whole-Body Human Pose Estimation in the Wild (Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ , Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ , Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ , Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ , Topdown Heatmap + Litehrnet + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ , Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ , Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand  $\Rightarrow$ )

## 17.11 Hand(3D,Kpt,Rgb,Img)

- 模型权重文件数量: 1
- 配置文件数量: 2
- 论文数量: 3
  - [ALGORITHM] Interhand2.6m: A Dataset and Baseline for 3d Interacting Hand Pose Estimation From a Single RGB Image (Internet + Internet on Interhand3d  $\Rightarrow$ )
  - [BACKBONE] Deep Residual Learning for Image Recognition (Internet + Internet on Interhand3d  $\Rightarrow$ )
  - [DATASET] Interhand2.6m: A Dataset and Baseline for 3d Interacting Hand Pose Estimation From a Single RGB Image (Internet + Internet on Interhand3d  $\Rightarrow$ )

## 17.12 Hand(Gesture,Rgbd,Vid)

- 模型权重文件数量: 3
- 配置文件数量: 4
- 论文数量: 3
  - [ALGORITHM] Improving the Performance of Unimodal Dynamic Hand-Gesture Recognition With Multimodal Training (Mtut + I3d on Nvgesture  $\Rightarrow$ )
  - [BACKBONE] Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset ()
  - [DATASET] Online Detection and Classification of Dynamic Hand Gestures With Recurrent 3d Convolutional Neural Network (Mtut + I3d on Nvgesture  $\Rightarrow$ )

## 17.13 Wholebody

- 模型权重文件数量: 22
- 配置文件数量: 22
- 论文数量: 9
  - [ALGORITHM] Associative Embedding: End-to-End Learning for Joint Detection and Grouping (Associative Embedding + Higherhrnet on Coco-Wholebody  $\Rightarrow$ , Associative Embedding + Hrnet on Coco-Wholebody  $\Rightarrow$ )
  - [ALGORITHM] Deep High-Resolution Representation Learning for Human Pose Estimation (Topdown Heatmap + Hrnet + Dark on Halpe  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Coco-Wholebody  $\Rightarrow$ , Topdown Heatmap + Hrnet on Coco-Wholebody  $\Rightarrow$ , Associative Embedding + Hrnet on Coco-Wholebody  $\Rightarrow$ )
  - [ALGORITHM] Distribution-Aware Coordinate Representation for Human Pose Estimation (Topdown Heatmap + Hrnet + Dark on Halpe  $\Rightarrow$ , Topdown Heatmap + Hrnet + Dark on Coco-Wholebody  $\Rightarrow$ , Topdown Heatmap + Vipnas + Dark on Coco-Wholebody  $\Rightarrow$ )
  - [ALGORITHM] Higherhrnet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation (Associative Embedding + Higherhrnet on Coco-Wholebody  $\Rightarrow$ )
  - [ALGORITHM] Not All Tokens Are Equal: Human-Centric Visual Analysis via Token Clustering Transformer (Topdown Heatmap + Tcformer on Coco-Wholebody  $\Rightarrow$ )
  - [ALGORITHM] Simple Baselines for Human Pose Estimation and Tracking (Topdown Heatmap + Resnet on Coco-Wholebody  $\Rightarrow$ )
  - [ALGORITHM] Vipnas: Efficient Video Pose Estimation via Neural Architecture Search (Topdown Heatmap + Vipnas + Dark on Coco-Wholebody  $\Rightarrow$ , Topdown Heatmap + Vipnas on Coco-Wholebody  $\Rightarrow$ )
  - [DATASET] Pastanet: Toward Human Activity Knowledge Engine (Topdown Heatmap + Hrnet + Dark on Halpe  $\Rightarrow$ )
  - [DATASET] Whole-Body Human Pose Estimation in the Wild (Topdown Heatmap + Hrnet + Dark on Coco-Wholebody  $\Rightarrow$ , Topdown Heatmap + Vipnas + Dark on Coco-Wholebody  $\Rightarrow$ , Topdown Heatmap + Vipnas on Coco-Wholebody  $\Rightarrow$ , Topdown Heatmap + Tcformer on Coco-Wholebody  $\Rightarrow$ , Topdown Heatmap + Hrnet on Coco-Wholebody  $\Rightarrow$ , Topdown Heatmap + Resnet on Coco-Wholebody  $\Rightarrow$ , Associative Embedding + Higherhrnet on Coco-Wholebody  $\Rightarrow$ , Associative Embedding + Hrnet on Coco-Wholebody  $\Rightarrow$ )



## 18.1 Animalpose Dataset

### 18.1.1 Topdown Heatmap + Hrnet on Animalpose

```
@inproceedings{sun2019deep,  
  title={Deep high-resolution representation learning for human pose estimation},  
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={5693--5703},  
  year={2019}  
}
```

```
@InProceedings{Cao_2019_ICCV,  
  author = {Cao, Jinkun and Tang, Hongyang and Fang, Hao-Shu and Shen, Xiaoyong and  
↪Lu, Cewu and Tai, Yu-Wing},  
  title = {Cross-Domain Adaptation for Animal Pose Estimation},  
  booktitle = {The IEEE International Conference on Computer Vision (ICCV)},  
  month = {October},  
  year = {2019}  
}
```

Results on AnimalPose validation set (1117 instances)

## 18.1.2 Topdown Heatmap + Resnet on Animalpose

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@InProceedings{Cao_2019_ICCV,
  author = {Cao, Jinkun and Tang, Hongyang and Fang, Hao-Shu and Shen, Xiaoyong and
↪Lu, Cewu and Tai, Yu-Wing},
  title = {Cross-Domain Adaptation for Animal Pose Estimation},
  booktitle = {The IEEE International Conference on Computer Vision (ICCV)},
  month = {October},
  year = {2019}
}
```

Results on AnimalPose validation set (1117 instances)

## 18.2 Ap10k Dataset

### 18.2.1 Topdown Heatmap + Hrnet on Ap10k

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@misc{yu2021ap10k,
  title={AP-10K: A Benchmark for Animal Pose Estimation in the Wild},
  author={Hang Yu and Yufei Xu and Jing Zhang and Wei Zhao and Ziyu Guan and
↪Dacheng Tao},
  year={2021},
  eprint={2108.12617},
  archivePrefix={arXiv},
}
```

(下页继续)

(续上页)

```

    primaryClass={cs.CV}
}

```

Results on AP-10K validation set

## 18.2.2 Topdown Heatmap + Resnet on Ap10k

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@misc{yu2021ap10k,
  title={AP-10K: A Benchmark for Animal Pose Estimation in the Wild},
  author={Hang Yu and Yufei Xu and Jing Zhang and Wei Zhao and Ziyu Guan and
↪Dacheng Tao},
  year={2021},
  eprint={2108.12617},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}

```

Results on AP-10K validation set

## 18.3 Atrw Dataset

### 18.3.1 Topdown Heatmap + Hrnet on Atrw

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```
@inproceedings{li2020atrw,
  title={ATRW: A Benchmark for Amur Tiger Re-identification in the Wild},
  author={Li, Shuyuan and Li, Jianguo and Tang, Hanlin and Qian, Rui and Lin, Weiyao},
  booktitle={Proceedings of the 28th ACM International Conference on Multimedia},
  pages={2590--2598},
  year={2020}
}
```

Results on ATRW validation set

## 18.3.2 Topdown Heatmap + Resnet on Atrw

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{li2020atrw,
  title={ATRW: A Benchmark for Amur Tiger Re-identification in the Wild},
  author={Li, Shuyuan and Li, Jianguo and Tang, Hanlin and Qian, Rui and Lin, Weiyao},
  booktitle={Proceedings of the 28th ACM International Conference on Multimedia},
  pages={2590--2598},
  year={2020}
}
```

Results on ATRW validation set

## 18.4 Fly Dataset

### 18.4.1 Topdown Heatmap + Resnet on Fly

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```



```
@article{pereira2019fast,
  title={Fast animal pose estimation using deep neural networks},
  author={Pereira, Talmo D and Aldarondo, Diego E and Willmore, Lindsay and Kislin,
↪Mikhail and Wang, Samuel S-H and Murthy, Mala and Shaevitz, Joshua W},
  journal={Nature methods},
  volume={16},
  number={1},
  pages={117--125},
  year={2019},
  publisher={Nature Publishing Group}
}
```

Results on Vinegar Fly test set

## 18.5 Horse10 Dataset

### 18.5.1 Topdown Heatmap + Resnet on Horse10

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{mathis2021pretraining,
  title={Pretraining boosts out-of-domain robustness for pose estimation},
  author={Mathis, Alexander and Biasi, Thomas and Schneider, Steffen and Yuksekgonul,
↪Mert and Rogers, Byron and Bethge, Matthias and Mathis, Mackenzie W},
  booktitle={Proceedings of the IEEE/CVF Winter Conference on Applications of
↪Computer Vision},
  pages={1859--1868},
  year={2021}
}
```

Results on Horse-10 test set

## 18.5.2 Topdown Heatmap + Hrnet on Horse10

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{mathis2021pretraining,
  title={Pretraining boosts out-of-domain robustness for pose estimation},
  author={Mathis, Alexander and Biasi, Thomas and Schneider, Steffen and Yuksekgonul,
↪Mert and Rogers, Byron and Bethge, Matthias and Mathis, Mackenzie W},
  booktitle={Proceedings of the IEEE/CVF Winter Conference on Applications of
↪Computer Vision},
  pages={1859--1868},
  year={2021}
}
```

Results on Horse-10 test set

## 18.6 Locust Dataset

### 18.6.1 Topdown Heatmap + Resnet on Locust

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@article{graving2019deepposekit,
  title={DeepPoseKit, a software toolkit for fast and robust animal pose estimation
↪using deep learning},
  author={Graving, Jacob M and Chae, Daniel and Naik, Hemal and Li, Liang and Koger,
↪Benjamin and Costelloe, Blair R and Couzin, Iain D},
  journal={Elife},

```

(下页继续)

(续上页)

```

    volume={8},
    pages={e47994},
    year={2019},
    publisher={eLife Sciences Publications Limited}
}

```

Results on Desert Locust test set

## 18.7 Macaque Dataset

### 18.7.1 Topdown Heatmap + Resnet on Macaque

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@article{labuguen2020macaquepose,
  title={MacaquePose: A novel ‘in the wild’ macaque monkey pose dataset for↵
↵markerless motion capture},
  author={Labuguen, Rollyn and Matsumoto, Jumpei and Negrete, Salvador and Nishimaru,↵
↵Hiroshi and Nishijo, Hisao and Takada, Masahiko and Go, Yasuhiro and Inoue, Ken-↵
↵ichi and Shibata, Tomohiro},
  journal={bioRxiv},
  year={2020},
  publisher={Cold Spring Harbor Laboratory}
}

```

Results on MacaquePose with ground-truth detection bounding boxes

### 18.7.2 Topdown Heatmap + Hrnet on Macaque

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern↵
↵recognition},

```

(下页继续)

(续上页)

```

    pages={5693--5703},
    year={2019}
  }

```

```

@article{labuguen2020macaquepose,
  title={MacaquePose: A novel 'in the wild' macaque monkey pose dataset for↵
↵markerless motion capture},
  author={Labuguen, Rollyn and Matsumoto, Jumpei and Negrete, Salvador and Nishimaru,↵
↵Hiroshi and Nishijo, Hisao and Takada, Masahiko and Go, Yasuhiro and Inoue, Ken-↵
↵ichi and Shibata, Tomohiro},
  journal={bioRxiv},
  year={2020},
  publisher={Cold Spring Harbor Laboratory}
}

```

Results on MacaquePose with ground-truth detection bounding boxes

## 18.8 Zebra Dataset

### 18.8.1 Topdown Heatmap + Resnet on Zebra

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@article{graving2019deepposekit,
  title={DeepPoseKit, a software toolkit for fast and robust animal pose estimation↵
↵using deep learning},
  author={Graving, Jacob M and Chae, Daniel and Naik, Hemal and Li, Liang and Koger,↵
↵Benjamin and Costelloe, Blair R and Couzin, Iain D},
  journal={Elife},
  volume={8},
  pages={e47994},
  year={2019},
  publisher={eLife Sciences Publications Limited}
}

```

Results on Grévy' s Zebra test set

---

Body(2D,Kpt,Sview,Img)

---

## 19.1 Aic Dataset

### 19.1.1 Associative Embedding + Hrnet on Aic

```
@inproceedings{newell2017associative,  
  title={Associative embedding: End-to-end learning for joint detection and grouping},  
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},  
  booktitle={Advances in neural information processing systems},  
  pages={2277--2287},  
  year={2017}  
}
```

```
@inproceedings{sun2019deep,  
  title={Deep high-resolution representation learning for human pose estimation},  
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={5693--5703},  
  year={2019}  
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding},
  ↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and ↪
  ↪Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and ↪
  ↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 19.1.2 Associative Embedding + Higherhrnet on Aic

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose ↪
  ↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang, ↪
  ↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↪
  ↪Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
  ↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and ↪
  ↪Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and ↪
  ↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

(下页继续)

(续上页)

}

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 19.1.3 Topdown Heatmap + Hrnet on Aic

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shipai and Lin, Guosen and Fu, Yanwei and
↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC val set with ground-truth bounding boxes

### 19.1.4 Topdown Heatmap + Resnet on Aic

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
```

(下页继续)

(续上页)

```

    author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@article{wu2017ai,
    title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
    author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and
↪others},
    journal={arXiv preprint arXiv:1711.06475},
    year={2017}
}

```

Results on AIC val set with ground-truth bounding boxes

## 19.2 Coco Dataset

### 19.2.1 Associative Embedding + Higherhrnet + Udp on Coco

```

@inproceedings{newell2017associative,
    title={Associative embedding: End-to-end learning for joint detection and grouping},
    author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
    booktitle={Advances in neural information processing systems},
    pages={2277--2287},
    year={2017}
}

```

```

@inproceedings{cheng2020higherhrnet,
    title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose
↪Estimation},
    author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,
↪Thomas S and Zhang, Lei},
    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
    pages={5386--5395},
    year={2020}
}

```



```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

## 19.2.2 Associative Embedding + Mobilenetv2 on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 19.2.3 Associative Embedding + Hrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

## 19.2.4 Associative Embedding + Higherhrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪ Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪ Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪ Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪ Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

## 19.2.5 Associative Embedding + Resnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

## 19.2.6 Associative Embedding + Hrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

## 19.2.7 Associative Embedding + Hourglass + Ae on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

## 19.2.8 Deeppose + Resnet on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.9 Deeppose + Resnet + Rle on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.10 Dekr + Hrnet on Coco

```
@inproceedings{geng2021bottom,
  title={Bottom-up human pose estimation via disentangled keypoint regression},
  author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,
↪Jingdong},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={14676--14686},
  year={2021}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```



Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

The results of models provided by the authors on COCO val2017 using the same evaluation protocol

The discrepancy between these results and that shown in paper is attributed to the differences in implementation details in evaluation process.

### 19.2.11 Topdown Heatmap + Seresnet on Coco

```
@inproceedings{hu2018squeeze,
  title={Squeeze-and-excitation networks},
  author={Hu, Jie and Shen, Li and Sun, Gang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={7132--7141},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that \* means without imagenet pre-training.

### 19.2.12 Topdown Heatmap + Resnetv1d on Coco

```
@inproceedings{he2019bag,
  title={Bag of tricks for image classification with convolutional neural networks},
  author={He, Tong and Zhang, Zhi and Zhang, Hang and Zhang, Zhongyue and Xie,
↪Junyuan and Li, Mu},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition},
  pages={558--567},
```

(下页继续)

(续上页)

```

    year={2019}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.13 Topdown Heatmap + Hourglass on Coco

```

@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.14 Topdown Heatmap + RSN on Coco

```
@misc{cai2020learning,
  title={Learning Delicate Local Representations for Multi-Person Pose Estimation},
  author={Yuanhao Cai and Zhicheng Wang and Zhengxiong Luo and Binyi Yin and Angang
↪Du and Haoqian Wang and Xinyu Zhou and Erjin Zhou and Xiangyu Zhang and Jian Sun},
  year={2020},
  eprint={2003.04030},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.15 Topdown Heatmap + Resnet + Fp16 on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{micikevicius2017mixed,
  title={Mixed precision training},
  author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
  journal={arXiv preprint arXiv:1710.03740},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.16 Topdown Heatmap + Mobilenetv2 on Coco

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.17 Topdown Heatmap + Shufflenetv1 on Coco

```
@inproceedings{zhang2018shufflenet,
  title={Shufflenet: An extremely efficient convolutional neural network for mobile↵
↵devices},
  author={Zhang, Xiangyu and Zhou, Xinyu and Lin, Mengxiao and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern↵
↵recognition},
  pages={6848--6856},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and↵
↵Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.18 Topdown Heatmap + MSPN on Coco

```
@article{li2019rethinking,
  title={Rethinking on Multi-Stage Networks for Human Pose Estimation},
  author={Li, Wenbo and Wang, Zhicheng and Yin, Binyi and Peng, Qixiang and Du,↵
↵Yuming and Xiao, Tianzi and Yu, Gang and Lu, Hongtao and Wei, Yichen and Sun, Jian},
  journal={arXiv preprint arXiv:1901.00148},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and↵
↵Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
}
```

(下页继续)

(续上页)

```

year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 19.2.19 Topdown Heatmap + Hrnet + Fp16 on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{micikevicius2017mixed,
  title={Mixed precision training},
  author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
  journal={arXiv preprint arXiv:1710.03740},
  year={2017}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 19.2.20 Topdown Heatmap + Hrnet on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 19.2.21 Topdown Heatmap + Resnext on Coco

```
@inproceedings{xie2017aggregated,
  title={Aggregated residual transformations for deep neural networks},
  author={Xie, Saining and Girshick, Ross and Doll{'a}r, Piotr and Tu, Zhuowen and
↪He, Kaiming},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1492--1500},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
}
```

(下页继续)

(续上页)

```
organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 19.2.22 Topdown Heatmap + Resnet + Dark on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```



Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.23 Topdown Heatmap + VGG on Coco

```
@article{simonyan2014very,
  title={Very deep convolutional networks for large-scale image recognition},
  author={Simonyan, Karen and Zisserman, Andrew},
  journal={arXiv preprint arXiv:1409.1556},
  year={2014}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.24 Topdown Heatmap + Shufflenetv2 on Coco

```
@inproceedings{ma2018shufflenet,
  title={Shufflenet v2: Practical guidelines for efficient cnn architecture design},
  author={Ma, Ningning and Zhang, Xiangyu and Zheng, Hai-Tao and Sun, Jian},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={116--131},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.25 Topdown Heatmap + Hrnet + Augmentation on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{buslaev2020albumentations,
  title={Albumentations: fast and flexible image augmentations},
  author={Buslaev, Alexander and Iglovikov, Vladimir I and Khvedchenya, Eugene and
↪Parinov, Alex and Druzhinin, Mikhail and Kalinin, Alexandr A},
  journal={Information},
  volume={11},
  number={2},
  pages={125},
  year={2020},
  publisher={Multidisciplinary Digital Publishing Institute}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 19.2.26 Topdown Heatmap + Swin on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{liu2021swin,
  title={Swin transformer: Hierarchical vision transformer using shifted windows},
  author={Liu, Ze and Lin, Yutong and Cao, Yue and Hu, Han and Wei, Yixuan and Zhang,
↪Zheng and Lin, Stephen and Guo, Baining},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={10012--10022},
  year={2021}
}
```

```
@inproceedings{lin2017feature,
  title={Feature pyramid networks for object detection},
  author={Lin, Tsung-Yi and Doll{\`a}r, Piotr and Girshick, Ross and He, Kaiming and
↪Hariharan, Bharath and Belongie, Serge},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2117--2125},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 19.2.27 Topdown Heatmap + Litehrnet on Coco

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 19.2.28 Topdown Heatmap + Snet on Coco

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.29 Topdown Heatmap + Hrformer on Coco

```
@article{yuan2021hrformer,
  title={HRFormer: High-Resolution Vision Transformer for Dense Predict},
  author={Yuan, Yuhui and Fu, Rao and Huang, Lang and Lin, Weihong and Zhang, Chao_
↪and Chen, Xilin and Wang, Jingdong},
  journal={Advances in Neural Information Processing Systems},
  volume={34},
  year={2021}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.30 Topdown Heatmap + Resnest on Coco

```
@article{zhang2020resnest,
  title={ResNeSt: Split-Attention Networks},
  author={Zhang, Hang and Wu, Chongruo and Zhang, Zhongyue and Zhu, Yi and Zhang, Zhi_
↪and Lin, Haibin and Sun, Yue and He, Tong and Muller, Jonas and Manmatha, R. and Li,
↪Mu and Smola, Alexander},
  journal={arXiv preprint arXiv:2004.08955},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
}
```

(下页继续)

(续上页)

```

year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.31 Topdown Heatmap + PVT on Coco

```

@inproceedings{wang2021pyramid,
  title={Pyramid vision transformer: A versatile backbone for dense prediction↵
↵without convolutions},
  author={Wang, Wenhai and Xie, Enze and Li, Xiang and Fan, Deng-Ping and Song,↵
↵Kaitao and Liang, Ding and Lu, Tong and Luo, Ping and Shao, Ling},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={568--578},
  year={2021}
}

```

```

@article{wang2022pvt,
  title={PVT v2: Improved baselines with Pyramid Vision Transformer},
  author={Wang, Wenhai and Xie, Enze and Li, Xiang and Fan, Deng-Ping and Song,↵
↵Kaitao and Liang, Ding and Lu, Tong and Luo, Ping and Shao, Ling},
  journal={Computational Visual Media},
  pages={1--10},
  year={2022},
  publisher={Springer}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and↵
↵Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.32 Topdown Heatmap + Resnet on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.33 Topdown Heatmap + Alexnet on Coco

```
@inproceedings{krizhevsky2012imagenet,
  title={Imagenet classification with deep convolutional neural networks},
  author={Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E},
  booktitle={Advances in neural information processing systems},
  pages={1097--1105},
  year={2012}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.34 Topdown Heatmap + Hrnet + Udp on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```



Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that, UDP also adopts the unbiased encoding/decoding algorithm of [DARK](#).

### 19.2.35 Topdown Heatmap + Hrnet + Dark on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.36 Topdown Heatmap + Vipnas on Coco

```
@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen_
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  year={2021}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.37 Topdown Heatmap + CPM on Coco

```
@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern_
↪Recognition},
  pages={4724--4732},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 19.2.38 Cid + Hrnet on Coco

```
@InProceedings{Wang_2022_CVPR,
  author    = {Wang, Dongkai and Zhang, Shiliang},
  title     = {Contextual Instance Decoupling for Robust Multi-Person Pose_
↵Estimation},
  booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and_
↵Pattern Recognition (CVPR)},
  month     = {June},
  year      = {2022},
  pages     = {11060-11068}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↵Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

## 19.3 Crowdpose Dataset

### 19.3.1 Associative Embedding + Higherhrnet on Crowdpose

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↵Estimation},
```

(下页继续)

(续上页)

```

    author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,
↪Thomas S and Zhang, Lei},
    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
    pages={5386--5395},
    year={2020}
}

```

```

@article{li2018crowdpose,
    title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
    author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
    journal={arXiv preprint arXiv:1812.00324},
    year={2018}
}

```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 2 scales ([2, 1]) are used

### 19.3.2 Dekr + Hrnet on Crowdpose

```

@inproceedings{geng2021bottom,
    title={Bottom-up human pose estimation via disentangled keypoint regression},
    author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,
↪Jingdong},
    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
    pages={14676--14686},
    year={2021}
}

```

```

@inproceedings{sun2019deep,
    title={Deep high-resolution representation learning for human pose estimation},
    author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={5693--5703},
    year={2019}
}

```

```
@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}
```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

### 19.3.3 Topdown Heatmap + Hrnet on Crowdpose

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}
```

Results on CrowdPose test with YOLOv3 human detector

### 19.3.4 Topdown Heatmap + Resnet on Crowdpose

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
```

(下页继续)

(续上页)

```

    year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

Results on CrowdPose test with YOLOv3 human detector

## 19.4 H36m Dataset

### 19.4.1 Topdown Heatmap + Hrnet on H36m

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu,
↪Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human
↪Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},

```

(下页继续)

(续上页)

```

publisher = {IEEE Computer Society},
volume = {36},
number = {7},
pages = {1325-1339},
month = {jul},
year = {2014}
}

```

Results on Human3.6M test set with ground truth 2D detections

## 19.5 JHMDB Dataset

### 19.5.1 Topdown Heatmap + CPM on JHMDB

```

@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={4724--4732},
  year={2016}
}

```

```

@inproceedings{Jhuang:ICCV:2013,
  title = {Towards understanding action recognition},
  author = {H. Jhuang and J. Gall and S. Zuffi and C. Schmid and M. J. Black},
  booktitle = {International Conf. on Computer Vision (ICCV)},
  month = Dec,
  pages = {3192-3199},
  year = {2013}
}

```

Results on Sub-JHMDB dataset

The models are pre-trained on MPII dataset only. NO test-time augmentation (multi-scale /rotation testing) is used.

- Normalized by Person Size
- Normalized by Torso Size

## 19.5.2 Topdown Heatmap + Resnet on JHMDB

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{Jhuang:ICCV:2013,
  title = {Towards understanding action recognition},
  author = {H. Jhuang and J. Gall and S. Zuffi and C. Schmid and M. J. Black},
  booktitle = {International Conf. on Computer Vision (ICCV)},
  month = Dec,
  pages = {3192-3199},
  year = {2013}
}
```

Results on Sub-JHMDB dataset

The models are pre-trained on MPII dataset only. *NO* test-time augmentation (multi-scale /rotation testing) is used.

- Normalized by Person Size
- Normalized by Torso Size

## 19.6 MHP Dataset

### 19.6.1 Associative Embedding + Hrnet on MHP

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
```

(下页继续)



(续上页)

```

booktitle={Advances in neural information processing systems},
pages={2277--2287},
year={2017}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}

```

Results on MHP v2.0 validation set without multi-scale test

Results on MHP v2.0 validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

## 19.6.2 Topdown Heatmap + Resnet on MHP

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},

```

(下页继续)

(续上页)

```

    pages={770--778},
    year={2016}
}

```

```

@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}

```

Results on MHP v2.0 val set

Note that, the evaluation metric used here is mAP (adapted from COCO), which may be different from the official evaluation codes. Please be cautious if you use the results in papers.

## 19.7 Mpii Dataset

### 19.7.1 Deeppose + Resnet on Mpii

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 19.7.2 Deeppose + Resnet + Rle on Mpii

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo ↵
↵and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
```

(下页继续)

(续上页)

```

    title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
    booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
    year = {2014},
    month = {June}
}

```

Results on MPII val set

### 19.7.3 Topdown Heatmap + Hrnet on Mpii

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 19.7.4 Topdown Heatmap + Mobilenetv2 on Mpii

```

@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}

```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 19.7.5 Topdown Heatmap + Resnet on Mpii

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 19.7.6 Topdown Heatmap + CPM on Mpii

```
@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={4724--4732},
  year={2016}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↪
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 19.7.7 Topdown Heatmap + Shufflenetv2 on Mpii

```
@inproceedings{ma2018shufflenet,
  title={Shufflenet v2: Practical guidelines for efficient cnn architecture design},
  author={Ma, Ningning and Zhang, Xiangyu and Zheng, Hai-Tao and Sun, Jian},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={116--131},
  year={2018}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↪
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 19.7.8 Topdown Heatmap + Litehrnet on Mpii

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 19.7.9 Topdown Heatmap + Resnext on Mpii

```
@inproceedings{xie2017aggregated,
  title={Aggregated residual transformations for deep neural networks},
  author={Xie, Saining and Girshick, Ross and Doll{'a}r, Piotr and Tu, Zhuowen and
↪He, Kaiming},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1492--1500},
  year={2017}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 19.7.10 Topdown Heatmap + Shufflenetv1 on Mpii

```
@inproceedings{zhang2018shufflenet,
  title={Shufflenet: An extremely efficient convolutional neural network for mobile↵
↵devices},
  author={Zhang, Xiangyu and Zhou, Xinyu and Lin, Mengxiao and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern↵
↵recognition},
  pages={6848--6856},
  year={2018}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 19.7.11 Topdown Heatmap + Hrnet + Dark on Mpii

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern↵
↵recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern↵
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```



```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 19.7.12 Topdown Heatmap + Scnet on Mpii

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and ↵
↵Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↵
↵Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 19.7.13 Topdown Heatmap + Seresnet on Mpii

```
@inproceedings{hu2018squeeze,
  title={Squeeze-and-excitation networks},
  author={Hu, Jie and Shen, Li and Sun, Gang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={7132--7141},
```

(下页继续)

(续上页)

```

    year={2018}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

Note that \* means without imagenet pre-training.

### 19.7.14 Topdown Heatmap + Resnetv1d on Mpii

```

@inproceedings{he2019bag,
  title={Bag of tricks for image classification with convolutional neural networks},
  author={He, Tong and Zhang, Zhi and Zhang, Hang and Zhang, Zhongyue and Xie, ↵
↵Junyuan and Li, Mu},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern ↵
↵Recognition},
  pages={558--567},
  year={2019}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

## 19.7.15 Topdown Heatmap + Hourglass on Mpii

```
@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 19.8 Mpii\_trb Dataset

### 19.8.1 Topdown Heatmap + Resnet + Mpii on Mpii\_trb

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{duan2019trb,
  title={TRB: A Novel Triplet Representation for Understanding 2D Human Body},
  author={Duan, Haodong and Lin, Kwan-Yee and Jin, Sheng and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={9479--9488},
  year={2019}
}
```

Results on MPII-TRB val set

## 19.9 Ochuman Dataset

### 19.9.1 Topdown Heatmap + Resnet on Ochuman

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},
  author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↪and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={889--898},
  year={2019}
}
```

Results on OCHuman test dataset with ground-truth bounding boxes

Following the common setting, the models are trained on COCO train dataset, and evaluate on OCHuman dataset.

## 19.9.2 Topdown Heatmap + Hrnet on Ochuman

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},
  author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↪and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={889--898},
  year={2019}
}
```

Results on OCHuman test dataset with ground-truth bounding boxes

Following the common setting, the models are trained on COCO train dataset, and evaluate on OCHuman dataset.

## 19.10 Posetrack18 Dataset

### 19.10.1 Topdown Heatmap + Hrnet on Posetrack18

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andrilluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin, Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={5167--5176},
  year={2018}
}
```

Results on PoseTrack2018 val with ground-truth bounding boxes

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

Results on PoseTrack2018 val with MMDetection pre-trained Cascade R-CNN (X-101-64x4d-FPN) human detector

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

## 19.10.2 Topdown Heatmap + Resnet on Posetrack18

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andrilluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin, Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={5167--5176},
}
```

(下页继续)

(续上页)

```
year={2018}  
}
```

Results on PoseTrack2018 val with ground-truth bounding boxes

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

Results on PoseTrack2018 val with [MMDetection](#) pre-trained [Cascade R-CNN](#) (X-101-64x4d-FPN) human detector

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.





Body(2D,Kpt,Sview,Vid)

## 20.1 Posetrack18 Dataset

### 20.1.1 Posewarper + Hrnet + Posetrack18 on Posetrack18

```
@inproceedings{NIPS2019_gberta,  
  title = {Learning Temporal Pose Estimation from Sparsely Labeled Videos},  
  author = {Bertasius, Gedas and Feichtenhofer, Christoph, and Tran, Du and Shi, Jianbo,  
    ↪ and Torresani, Lorenzo},  
  booktitle = {Advances in Neural Information Processing Systems 33},  
  year = {2019},  
}
```

```
@inproceedings{sun2019deep,  
  title={Deep high-resolution representation learning for human pose estimation},  
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
    ↪ recognition},  
  pages={5693--5703},  
  year={2019}  
}
```

```
@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin, Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={5167--5176},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Note that the training of PoseWarper can be split into two stages.

The first-stage is trained with the pre-trained [checkpoint](#) on COCO dataset, and the main backbone is fine-tuned on PoseTrack18 in a single-frame setting.

The second-stage is trained with the last [checkpoint](#) from the first stage, and the warping offsets are learned in a multi-frame setting while the backbone is frozen.

Results on PoseTrack2018 val with ground-truth bounding boxes

Results on PoseTrack2018 val with precomputed human bounding boxes from PoseWarper supplementary data files from [this link](#)<sup>1</sup>.

<sup>1</sup> Please download the precomputed human bounding boxes on PoseTrack2018 val from `$PoseWarper_supp_files/posetrack18_precomputed_boxes/val_boxes.json` and place it here: `$mmpose/data/posetrack18/posetrack18_precomputed_boxes/val_boxes.json` to be consistent with the [config](#). Please refer to [DATA Preparation](#) for more detail about data preparation.

Body(3D,Kpt,Sview,Img)

## 21.1 H36m Dataset

### 21.1.1 Pose Lift + Simplebaseline3d on H36m

```
@inproceedings{martinez_2017_3dbaseline,  
  title={A simple yet effective baseline for 3d human pose estimation},  
  author={Martinez, Julieta and Hossain, Rayat and Romero, Javier and Little, James J.  
→},  
  booktitle={ICCV},  
  year={2017}  
}
```

```
@article{h36m_pami,  
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, C  
→Cristian},  
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human  
→Sensing in Natural Environments},  
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},  
  publisher = {IEEE Computer Society},  
  volume = {36},  
  number = {7},  
  pages = {1325-1339},  
  month = {jul},
```

(下页继续)

(续上页)

```

year = {2014}
}

```

Results on Human3.6M dataset with ground truth 2D detections

1 Differing from the original paper, we didn't apply the max-norm constraint because we found this led to a better convergence and performance.

## 21.2 Mpi\_inf\_3dhp Dataset

### 21.2.1 Pose Lift + Simplebaseline3d on Mpi\_inf\_3dhp

```

@inproceedings{martinez_2017_3dbaseline,
  title={A simple yet effective baseline for 3d human pose estimation},
  author={Martinez, Julieta and Hossain, Rayat and Romero, Javier and Little, James J.
↪},
  booktitle={ICCV},
  year={2017}
}

```

```

@inproceedings{mono-3dhp2017,
  author = {Mehta, Dushyant and Rhodin, Helge and Casas, Dan and Fua, Pascal and
↪Sotnychenko, Oleksandr and Xu, Weipeng and Theobalt, Christian},
  title = {Monocular 3D Human Pose Estimation In The Wild Using Improved CNN
↪Supervision},
  booktitle = {3D Vision (3DV), 2017 Fifth International Conference on},
  url = {http://gvv.mpi-inf.mpg.de/3dhp_dataset},
  year = {2017},
  organization={IEEE},
  doi={10.1109/3dv.2017.00064},
}

```

Results on MPI-INF-3DHP dataset with ground truth 2D detections

1 Differing from the original paper, we didn't apply the max-norm constraint because we found this led to a better convergence and performance.

Body(3D,Kpt,Sview,Vid)

## 22.1 H36m Dataset

### 22.1.1 Video Pose Lift + Videopose3d on H36m

```
@inproceedings{pav1lo20193d,  
  title={3d human pose estimation in video with temporal convolutions and semi-  
↪supervised training},  
  author={Pav1lo, Dario and Feichtenhofer, Christoph and Grangier, David and Auli, ↪  
↪Michael},  
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↪  
↪Recognition},  
  pages={7753--7762},  
  year={2019}  
}
```

```
@article{h36m_pami,  
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, ↪  
↪Cristian},  
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human ↪  
↪Sensing in Natural Environments},  
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},  
  publisher = {IEEE Computer Society},  
  volume = {36},
```

(下页继续)

(续上页)

```

    number = {7},
    pages = {1325-1339},
    month = {jul},
    year = {2014}
}

```

Results on Human3.6M dataset with ground truth 2D detections, supervised training

Results on Human3.6M dataset with CPN 2D detections<sup>1</sup>, supervised training

Results on Human3.6M dataset with ground truth 2D detections, semi-supervised training

Results on Human3.6M dataset with CPN 2D detections<sup>1</sup>, semi-supervised training

<sup>1</sup> CPN 2D detections are provided by [official repo](#). The reformatted version used in this repository can be downloaded from [train\\_detection](#) and [test\\_detection](#).

## 22.2 Mpi\_inf\_3dhp Dataset

### 22.2.1 Video Pose Lift + Videopose3d on Mpi\_inf\_3dhp

```

@inproceedings{pav11o20193d,
  title={3d human pose estimation in video with temporal convolutions and semi-
↪supervised training},
  author={Pav11o, Dario and Feichtenhofer, Christoph and Grangier, David and Auli, ↪
↪Michael},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↪
↪Recognition},
  pages={7753--7762},
  year={2019}
}

```

```

@inproceedings{mono-3dhp2017,
  author = {Mehta, Dushyant and Rhodin, Helge and Casas, Dan and Fua, Pascal and ↪
↪Sotnychenko, Oleksandr and Xu, Weipeng and Theobalt, Christian},
  title = {Monocular 3D Human Pose Estimation In The Wild Using Improved CNN ↪
↪Supervision},
  booktitle = {3D Vision (3DV), 2017 Fifth International Conference on},
  url = {http://gvv.mpi-inf.mpg.de/3dhp_dataset},
  year = {2017},
  organization={IEEE},
  doi={10.1109/3dv.2017.00064},
}

```

Results on MPI-INF-3DHP dataset with ground truth 2D detections, supervised training





---

Body(3D,Kpt,Mview,Img)

---

## 23.1 Campus Dataset

### 23.1.1 Voxelpose + Voxelpose on Campus

```
@inproceedings{tumultipose,
  title={VoxelPose: Towards Multi-Camera 3D Human Pose Estimation in Wild Environment},
  ↪,
  author={Tu, Hanyue and Wang, Chunyu and Zeng, Wenjun},
  booktitle={ECCV},
  year={2020}
}
```

```
@inproceedings {belagian14multi,
  title = {{3D} Pictorial Structures for Multiple Human Pose Estimation},
  author = {Belagiannis, Vasileios and Amin, Sikandar and Andriluka, Mykhaylo and ↪
  ↪Schiele, Bernt and Navab
  Nassir and Ilic, Slobodan},
  booktitle = {IEEE Computer Society Conference on Computer Vision and Pattern ↪
  ↪Recognition (CVPR)},
  year = {2014},
  month = {June},
  organization={IEEE}
}
```

Results on Campus dataset.

## 23.2 Panoptic Dataset

### 23.2.1 Voxelpose + Voxelpose + Prn64x64x64 + Cpn80x80x20 + Panoptic on Panoptic

```
@inproceedings{tumultipose,
  title={VoxelPose: Towards Multi-Camera 3D Human Pose Estimation in Wild Environment},
  ↪,
  author={Tu, Hanyue and Wang, Chunyu and Zeng, Wenjun},
  booktitle={ECCV},
  year={2020}
}
```

```
@Article = {joo_iccv_2015,
author = {Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo_
↪Kanade, Shohei Nobuhara, and Yaser Sheikh},
title = {Panoptic Studio: A Massively Multiview System for Social Motion Capture},
booktitle = {ICCV},
year = {2015}
}
```

Results on CMU Panoptic dataset.

## 23.3 Shelf Dataset

### 23.3.1 Voxelpose + Voxelpose on Shelf

```
@inproceedings{tumultipose,
  title={VoxelPose: Towards Multi-Camera 3D Human Pose Estimation in Wild Environment},
  ↪,
  author={Tu, Hanyue and Wang, Chunyu and Zeng, Wenjun},
  booktitle={ECCV},
  year={2020}
}
```

```
@inproceedings {belagian14multi,
  title = {{3D} Pictorial Structures for Multiple Human Pose Estimation},
  author = {Belagiannis, Vasileios and Amin, Sikandar and Andriluka, Mykhaylo and_
↪Schiele, Bernt and Navab
(下页继续)
```

(续上页)

```
Nassir and Ilic, Sloba
booktitle = {IEEE Computer Society Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
year = {2014},
month = {June},
organization={IEEE}
}
```

Results on Shelf dataset.



Body(3D,Mesh,Sview,Img)

## 24.1 Mixed Dataset

### 24.1.1 HMR + Resnet on Mixed

```
@inProceedings{kanazawaHMR18,  
  title={End-to-end Recovery of Human Shape and Pose},  
  author = {Angjoo Kanazawa  
    and Michael J. Black  
    and David W. Jacobs  
    and Jitendra Malik},  
  booktitle={Computer Vision and Pattern Recognition (CVPR)},  
  year={2018}  
}
```

```
@inproceedings{he2016deep,  
  title={Deep residual learning for image recognition},  
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
→recognition},  
  pages={770--778},  
  year={2016}  
}
```

```
@article{h36m_pami,  
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, ↵  
↵Cristian},  
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human↵  
↵Sensing in Natural Environments},  
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},  
  publisher = {IEEE Computer Society},  
  volume = {36},  
  number = {7},  
  pages = {1325-1339},  
  month = {jul},  
  year = {2014}  
}
```

Results on Human3.6M with ground-truth bounding box having MPJPE-PA of 52.60 mm on Protocol2

## 25.1 300w Dataset

### 25.1.1 Topdown Heatmap + Hrnetv2 on 300w

```
@article{WangSCJDZLMTWLX19,  
  title={Deep High-Resolution Representation Learning for Visual Recognition},  
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and  
          Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and  
          Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},  
  journal={TPAMI},  
  year={2019}  
}
```

```
@article{sagonas2016300,  
  title={300 faces in-the-wild challenge: Database and results},  
  author={Sagonas, Christos and Antonakos, Epameinondas and Tzimiropoulos, Georgios  
↪and Zafeiriou, Stefanos and Pantic, Maja},  
  journal={Image and vision computing},  
  volume={47},  
  pages={3--18},  
  year={2016},  
  publisher={Elsevier}  
}
```

Results on 300W dataset

The model is trained on 300W train.

## 25.2 Aflw Dataset

### 25.2.1 Topdown Heatmap + Hrnetv2 + Dark on Aflw

```
@article{WangSCJDZLMTWLX19,  
  title={Deep High-Resolution Representation Learning for Visual Recognition},  
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and  
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and  
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},  
  journal={TPAMI},  
  year={2019}  
}
```

```
@inproceedings{zhang2020distribution,  
  title={Distribution-aware coordinate representation for human pose estimation},  
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},  
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern  
↪Recognition},  
  pages={7093--7102},  
  year={2020}  
}
```

```
@inproceedings{koestinger2011annotated,  
  title={Annotated facial landmarks in the wild: A large-scale, real-world database  
↪for facial landmark localization},  
  author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},  
  booktitle={2011 IEEE international conference on computer vision workshops (ICCV  
↪workshops)},  
  pages={2144--2151},  
  year={2011},  
  organization={IEEE}  
}
```

Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.



## 25.2.2 Topdown Heatmap + Hrnetv2 on Aflw

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database_
↵for facial landmark localization},
  author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
  booktitle={2011 IEEE international conference on computer vision workshops (ICCV_
↵workshops)},
  pages={2144--2151},
  year={2011},
  organization={IEEE}
}
```

Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

## 25.3 Coco\_wholebody\_face Dataset

### 25.3.1 Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_face

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↵and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↵recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 25.3.2 Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_face

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 25.3.3 Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_face

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
    ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 25.3.4 Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_face

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
    ↪ Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
    ↪ Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
    ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 25.3.5 Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_face

```
@inproceedings{newell12016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 25.3.6 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

## 25.4 Cofw Dataset

### 25.4.1 Topdown Heatmap + Hrnetv2 on Cofw

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{burgos2013robust,
  title={Robust face landmark estimation under occlusion},
  author={Burgos-Artizzu, Xavier P and Perona, Pietro and Doll{'a}r, Piotr},
  booktitle={Proceedings of the IEEE international conference on computer vision},
  pages={1513--1520},
  year={2013}
}
```

Results on COFW dataset

The model is trained on COFW train.

## 25.5 WFLW Dataset

### 25.5.1 Deeppose + Resnet on WFLW

```
@inproceedings{toshev2014deeppose,  
  title={Deeppose: Human pose estimation via deep neural networks},  
  author={Toshev, Alexander and Szegedy, Christian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_  
↪recognition},  
  pages={1653--1660},  
  year={2014}  
}
```

```
@inproceedings{he2016deep,  
  title={Deep residual learning for image recognition},  
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_  
↪recognition},  
  pages={770--778},  
  year={2016}  
}
```

```
@inproceedings{wu2018look,  
  title={Look at boundary: A boundary-aware face alignment algorithm},  
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and_  
↪Zhou, Qiang},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_  
↪recognition},  
  pages={2129--2138},  
  year={2018}  
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 25.5.2 Deeppose + Resnet + Softwingloss on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{lin2021structure,
  title={Structure-Coherent Deep Feature Learning for Robust Face Alignment},
  author={Lin, Chunze and Zhu, Beier and Wang, Quan and Liao, Renjie and Qian, Chen
↪and Lu, Jiwen and Zhou, Jie},
  journal={IEEE Transactions on Image Processing},
  year={2021},
  publisher={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 25.5.3 Deeppose + Resnet + Wingloss on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{feng2018wing,
  title={Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural
↪Networks},
  author={Feng, Zhen-Hua and Kittler, Josef and Awais, Muhammad and Huber, Patrik and
↪Wu, Xiao-Jun},
  booktitle={Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on},
  year={2018},
  pages = {2235-2245},
  organization={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.



## 25.5.4 Topdown Heatmap + Hrnetv2 + Awing on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wang2019adaptive,
  title={Adaptive wing loss for robust face alignment via heatmap regression},
  author={Wang, Xinyao and Bo, Liefeng and Fuxin, Li},
  booktitle={Proceedings of the IEEE/CVF international conference on computer vision},
  pages={6971--6981},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
    Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
    recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 25.5.5 Topdown Heatmap + Hrnetv2 + Dark on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 25.5.6 Topdown Heatmap + Hrnetv2 on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.



## 26.1 Deepfashion Dataset

### 26.1.1 Deeppose + Resnet on Deepfashion

```
@inproceedings{toshev2014deeppose,  
  title={Deeppose: Human pose estimation via deep neural networks},  
  author={Toshev, Alexander and Szegedy, Christian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={1653--1660},  
  year={2014}  
}
```

```
@inproceedings{he2016deep,  
  title={Deep residual learning for image recognition},  
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={770--778},  
  year={2016}  
}
```

```
@inproceedings{liuLQWtcvpr16DeepFashion,
  author = {Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou},
  title = {DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich
↪Annotations},
  booktitle = {Proceedings of IEEE Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
  month = {June},
  year = {2016}
}
```

```
@inproceedings{liuYLWTeccv16FashionLandmark,
  author = {Liu, Ziwei and Yan, Sijie and Luo, Ping and Wang, Xiaogang and Tang,
↪Xiaoou},
  title = {Fashion Landmark Detection in the Wild},
  booktitle = {European Conference on Computer Vision (ECCV)},
  month = {October},
  year = {2016}
}
```

Results on DeepFashion val set

## 26.1.2 Topdown Heatmap + Resnet on Deepfashion

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{liuLQWtcvpr16DeepFashion,
  author = {Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou},
```

(下页继续)

(续上页)

```
title = {DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich↵
↵Annotations},
booktitle = {Proceedings of IEEE Conference on Computer Vision and Pattern↵
↵Recognition (CVPR)},
month = {June},
year = {2016}
}
```

```
@inproceedings{liuYLWTeccv16FashionLandmark,
author = {Liu, Ziwei and Yan, Sijie and Luo, Ping and Wang, Xiaogang and Tang,↵
↵Xiaoou},
title = {Fashion Landmark Detection in the Wild},
booktitle = {European Conference on Computer Vision (ECCV)},
month = {October},
year = {2016}
}
```

Results on DeepFashion val set





---

Hand(2D,Kpt,Rgb,Img)

---

## 27.1 Coco\_wholebody\_hand Dataset

### 27.1.1 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,  
  title={Deep High-Resolution Representation Learning for Visual Recognition},  
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and  
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and  
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},  
  journal={TPAMI},  
  year={2019}  
}
```

```
@inproceedings{zhang2020distribution,  
  title={Distribution-aware coordinate representation for human pose estimation},  
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},  
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern  
↪Recognition},  
  pages={7093--7102},  
  year={2020}  
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 27.1.2 Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey↪
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern↪
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 27.1.3 Topdown Heatmap + Litehrnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei↪
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
```

(下页继续)

(续上页)

```

    year={2021}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Hand val set

#### 27.1.4 Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_hand

```

@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Hand val set

### 27.1.5 Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 27.1.6 Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
  ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

## 27.1.7 Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
  Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
  Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
  ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

## 27.2 Freihand2d Dataset

### 27.2.1 Topdown Heatmap + Resnet on Freihand2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
}
```

(下页继续)

(续上页)

```

pages={466--481},
year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{zimmermann2019freihand,
  title={Freihand: A dataset for markerless capture of hand pose and shape from
↪single rgb images},
  author={Zimmermann, Christian and Ceylan, Duygu and Yang, Jimei and Russell, Bryan
↪and Argus, Max and Brox, Thomas},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={813--822},
  year={2019}
}

```

Results on FreiHand val &amp; test set

## 27.3 Interhand2d Dataset

### 27.3.1 Topdown Heatmap + Resnet on Interhand2d

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},

```

(下页继续)

(续上页)

```

    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}

```

Results on InterHand2.6M val & test set

## 27.4 Onehand10k Dataset

### 27.4.1 Deeppose + Resnet on Onehand10k

```

@inproceedings{toshev2014deeppose,
    title={Deeppose: Human pose estimation via deep neural networks},
    author={Toshev, Alexander and Szegedy, Christian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={1653--1660},
    year={2014}
}

```

```

@inproceedings{he2016deep,
    title={Deep residual learning for image recognition},
    author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@article{wang2018mask,
    title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},

```

(下页继续)

(续上页)

```

author={Wang, Yangang and Peng, Cong and Liu, Yebin},
journal={IEEE Transactions on Circuits and Systems for Video Technology},
volume={29},
number={11},
pages={3258--3268},
year={2018},
publisher={IEEE}
}

```

Results on OneHand10K val set

## 27.4.2 Topdown Heatmap + Hrnetv2 + Dark on Onehand10k

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}

```

Results on OneHand10K val set



### 27.4.3 Topdown Heatmap + Resnet on Onehand10k

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 27.4.4 Topdown Heatmap + Hrnetv2 on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 27.4.5 Topdown Heatmap + Mobilenetv2 on Onehand10k

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

## 27.4.6 Topdown Heatmap + Hrnetv2 + Udp on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

## 27.5 Panoptic2d Dataset

### 27.5.1 Deeppose + Resnet on Panoptic2d

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
```

(下页继续)

(续上页)

```

    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={1653--1660},
    year={2014}
}

```

```

@inproceedings{he2016deep,
    title={Deep residual learning for image recognition},
    author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@inproceedings{simon2017hand,
    title={Hand keypoint detection in single images using multiview bootstrapping},
    author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
    booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
    pages={1145--1153},
    year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

## 27.5.2 Topdown Heatmap + Mobilenetv2 on Panoptic2d

```

@inproceedings{sandler2018mobilenetv2,
    title={Mobilenetv2: Inverted residuals and linear bottlenecks},
    author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={4510--4520},
    year={2018}
}

```

```

@inproceedings{simon2017hand,
    title={Hand keypoint detection in single images using multiview bootstrapping},
    author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},

```

(下页继续)

(续上页)

```

    booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
    pages={1145--1153},
    year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 27.5.3 Topdown Heatmap + Resnet on Panoptic2d

```

@inproceedings{xiao2018simple,
    title={Simple baselines for human pose estimation and tracking},
    author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
    booktitle={Proceedings of the European conference on computer vision (ECCV)},
    pages={466--481},
    year={2018}
}

```

```

@inproceedings{he2016deep,
    title={Deep residual learning for image recognition},
    author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@inproceedings{simon2017hand,
    title={Hand keypoint detection in single images using multiview bootstrapping},
    author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
    booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
    pages={1145--1153},
    year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

## 27.5.4 Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

## 27.5.5 Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern_
↪Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

## 27.5.6 Topdown Heatmap + Hrnetv2 on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern_
↪Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

## 27.6 Rhd2d Dataset

### 27.6.1 Deeppose + Resnet on Rhd2d

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 27.6.2 Topdown Heatmap + Mobilenetv2 on Rhd2d

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
}
```

(下页继续)



(续上页)

```

    year={2018}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}

```

Results on RHD test set

### 27.6.3 Topdown Heatmap + Resnet on Rhd2d

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}

```

Results on RHD test set

## 27.6.4 Topdown Heatmap + Hrnetv2 + Udp on Rhd2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

## 27.6.5 Topdown Heatmap + Hrnetv2 + Dark on Rhd2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
```

(下页继续)

(续上页)

```

    year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}

```

Results on RHD test set

## 27.6.6 Topdown Heatmap + Hrnetv2 on Rhd2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}

```

Results on RHD test set

Hand(3D,Kpt,Rgb,Img)

## 28.1 Interhand3d Dataset

### 28.1.1 Internet + Internet on Interhand3d

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,  
  author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee, ↵  
    ↪Kyoung Mu},  
  title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose ↵  
    ↪Estimation from a Single RGB Image},  
  booktitle = {European Conference on Computer Vision (ECCV)},  
  year = {2020}  
}
```

```
@inproceedings{he2016deep,  
  title={Deep residual learning for image recognition},  
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵  
    ↪recognition},  
  pages={770--778},  
  year={2016}  
}
```

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,  
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee, ↵  
↵Kyoung Mu},  
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose ↵  
↵Estimation from a Single RGB Image},  
booktitle = {European Conference on Computer Vision (ECCV)},  
year = {2020}  
}
```

Results on InterHand2.6M val & test set

### 29.1 Nvgesture Dataset

#### 29.1.1 Mtut + l3d on Nvgesture

```
@InProceedings{Abavisani_2019_CVPR,  
  author = {Abavisani, Mahdi and Joze, Hamid Reza Vaezi and Patel, Vishal M.},  
  title = {Improving the Performance of Unimodal Dynamic Hand-Gesture Recognition  
↪With Multimodal Training},  
  booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern  
↪Recognition (CVPR)},  
  month = {June},  
  year = {2019}  
}
```

```
@InProceedings{Carreira_2017_CVPR,  
  author = {Carreira, Joao and Zisserman, Andrew},  
  title = {Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset},  
  booktitle = {Proceedings of the IEEE Conference on Computer Vision and Pattern  
↪Recognition (CVPR)},  
  month = {July},  
  year = {2017}  
}
```

```
@InProceedings{Molchanov_2016_CVPR,
  author = {Molchanov, Pavlo and Yang, Xiaodong and Gupta, Shalini and Kim, Kihwan_
↪and Tyree, Stephen and Kautz, Jan},
  title = {Online Detection and Classification of Dynamic Hand Gestures With_
↪Recurrent 3D Convolutional Neural Network},
  booktitle = {Proceedings of the IEEE Conference on Computer Vision and Pattern_
↪Recognition (CVPR)},
  month = {June},
  year = {2016}
}
```

Results on NVGesture test set

\*: MTUT supports multi-modal training and uni-modal testing. Model trained with this config can be used to recognize gestures in rgb videos with [inference config](#).



### 30.1 Coco-Wholebody Dataset

#### 30.1.1 Associative Embedding + Hrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,  
  title={Associative embedding: End-to-end learning for joint detection and grouping},  
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},  
  booktitle={Advances in neural information processing systems},  
  pages={2277--2287},  
  year={2017}  
}
```

```
@inproceedings{sun2019deep,  
  title={Deep high-resolution representation learning for human pose estimation},  
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={5693--5703},  
  year={2019}  
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 30.1.2 Associative Embedding + Higherhrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪ Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang, _
↪ Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪ Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 30.1.3 Topdown Heatmap + Resnet on Coco-Wholebody

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
  ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 30.1.4 Topdown Heatmap + Hrnet on Coco-Wholebody

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
  ↪ recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
  ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 30.1.5 Topdown Heatmap + Tcformer on Coco-Wholebody

```
@inproceedings{zeng2022not,  
  title={Not All Tokens Are Equal: Human-centric Visual Analysis via Token Clustering_  
↪Transformer},  
  author={Zeng, Wang and Jin, Sheng and Liu, Wentao and Qian, Chen and Luo, Ping and_  
↪Ouyang, Wanli and Wang, Xiaogang},  
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_  
↪Recognition},  
  pages={11101--11111},  
  year={2022}  
}
```

```
@inproceedings{jin2020whole,  
  title={Whole-Body Human Pose Estimation in the Wild},  
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,  
↪Chen and Ouyang, Wanli and Luo, Ping},  
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},  
  year={2020}  
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 30.1.6 Topdown Heatmap + Vipnas on Coco-Wholebody

```
@article{xu2021vipnas,  
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},  
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen_  
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_  
↪recognition},  
  year={2021}  
}
```

```
@inproceedings{jin2020whole,  
  title={Whole-Body Human Pose Estimation in the Wild},  
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,  
↪Chen and Ouyang, Wanli and Luo, Ping},  
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},  
  year={2020}  
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 30.1.7 Topdown Heatmap + Vipnas + Dark on Coco-Wholebody

```
@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen_
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  year={2021}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 30.1.8 Topdown Heatmap + Hrnet + Dark on Coco-Wholebody

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

## 30.2 Halpe Dataset

### 30.2.1 Topdown Heatmap + Hrnet + Dark on Halpe

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
```

(下页继续)

(续上页)

```
    year={2020}  
  }
```

```
@inproceedings{li2020pastanet,  
  title={PaStaNet: Toward Human Activity Knowledge Engine},  
  author={Li, Yong-Lu and Xu, Liang and Liu, Xinpeng and Huang, Xijie and Xu, Yue and  
↪Wang, Shiyi and Fang, Hao-Shu and Ma, Ze and Chen, Mingyang and Lu, Cewu},  
  booktitle={CVPR},  
  year={2020}  
}
```

Results on Halpe v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on Halpe dataset. We find this will lead to better performance.





### 31.1 MTUT (CVPR' 2019)

#### 31.1.1 Mtut + l3d on Nvgesture

```
@InProceedings{Abavisani_2019_CVPR,  
  author = {Abavisani, Mahdi and Joze, Hamid Reza Vaezi and Patel, Vishal M.},  
  title = {Improving the Performance of Unimodal Dynamic Hand-Gesture Recognition  
↪With Multimodal Training},  
  booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern  
↪Recognition (CVPR)},  
  month = {June},  
  year = {2019}  
}
```

```
@InProceedings{Carreira_2017_CVPR,  
  author = {Carreira, Joao and Zisserman, Andrew},  
  title = {Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset},  
  booktitle = {Proceedings of the IEEE Conference on Computer Vision and Pattern  
↪Recognition (CVPR)},  
  month = {July},  
  year = {2017}  
}
```

```
@InProceedings{Molchanov_2016_CVPR,
  author = {Molchanov, Pavlo and Yang, Xiaodong and Gupta, Shalini and Kim, Kihwan_
↪and Tyree, Stephen and Kautz, Jan},
  title = {Online Detection and Classification of Dynamic Hand Gestures With_
↪Recurrent 3D Convolutional Neural Network},
  booktitle = {Proceedings of the IEEE Conference on Computer Vision and Pattern_
↪Recognition (CVPR)},
  month = {June},
  year = {2016}
}
```

Results on NVGesture test set

\*: MTUT supports multi-modal training and uni-modal testing. Model trained with this config can be used to recognize gestures in rgb videos with [inference config](#).

## 31.2 MSPN (ArXiv' 2019)

### 31.2.1 Topdown Heatmap + MSPN on Coco

```
@article{li2019rethinking,
  title={Rethinking on Multi-Stage Networks for Human Pose Estimation},
  author={Li, Wenbo and Wang, Zhicheng and Yin, Binyi and Peng, Qixiang and Du,_
↪Yuming and Xiao, Tianzi and Yu, Gang and Lu, Hongtao and Wei, Yichen and Sun, Jian},
  journal={arXiv preprint arXiv:1901.00148},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 31.3 InterNet (ECCV' 2020)

### 31.3.1 Internet + Internet on Interhand3d

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}
```

```
@inproceedings{he2016deep,
title={Deep residual learning for image recognition},
author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
pages={770--778},
year={2016}
}
```

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}
```

Results on InterHand2.6M val & test set

## 31.4 DEKR (CVPR' 2021)

### 31.4.1 Dekr + Hrnet on Coco

```
@inproceedings{geng2021bottom,
title={Bottom-up human pose estimation via disentangled keypoint regression},
author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,
↪Jingdong},
```

(下页继续)

(续上页)

```

    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
    pages={14676--14686},
    year={2021}
}

```

```

@inproceedings{sun2019deep,
    title={Deep high-resolution representation learning for human pose estimation},
    author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={5693--5703},
    year={2019}
}

```

```

@inproceedings{lin2014microsoft,
    title={Microsoft coco: Common objects in context},
    author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
    booktitle={European conference on computer vision},
    pages={740--755},
    year={2014},
    organization={Springer}
}

```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

The results of models provided by the authors on COCO val2017 using the same evaluation protocol

The discrepancy between these results and that shown in paper is attributed to the differences in implementation details in evaluation process.

### 31.4.2 Dekr + Hrnet on Crowdpose

```

@inproceedings{geng2021bottom,
    title={Bottom-up human pose estimation via disentangled keypoint regression},
    author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,
↪Jingdong},
    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},

```

(下页继续)

(续上页)

```

pages={14676--14686},
year={2021}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

## 31.5 HigherHRNet (CVPR' 2020)

### 31.5.1 Associative Embedding + Higherhrnet on Aic

```

@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}

```

```

@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose
↪Estimation},

```

(下页继续)

(续上页)

```

    author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,
    ↪Thomas S and Zhang, Lei},
    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
    ↪Recognition},
    pages={5386--5395},
    year={2020}
}

```

```

@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
  ↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
  ↪Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and
  ↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}

```

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.5.2 Associative Embedding + Higherhrnet + Udp on Coco

```

@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}

```

```

@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose
  ↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,
  ↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
  ↪Recognition},
  pages={5386--5395},
  year={2020}
}

```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

### 31.5.3 Associative Embedding + Higherhrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.5.4 Associative Embedding + Higherhrnet on Crowdpose

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose
↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,
↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}
```

Results on CrowdPose test without multi-scale test



Results on CrowdPose test with multi-scale test. 2 scales ([2, 1]) are used

### 31.5.5 Associative Embedding + Higherhrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪ Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪ Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪ Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

## 31.6 DeepPose (CVPR' 2014)

### 31.6.1 Deeppose + Resnet on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.6.2 Deeppose + Resnet + Rle on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
```

(下页继续)

(续上页)

```

    year={2014}
}

```

```

@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.6.3 Deeppose + Resnet on Mpii

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},

```

(下页继续)

(续上页)

```

    year={2014}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 31.6.4 Deeppose + Resnet + Rle on Mpii

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}

```

```

@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}

```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 31.6.5 Deeppose + Resnet on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
```

(下页继续)

(续上页)

```

    author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={2129--2138},
    year={2018}
}

```

## Results on WFLW dataset

The model is trained on WFLW train.

### 31.6.6 Deeppose + Resnet + Softwingloss on WFLW

```

@inproceedings{toshev2014deeppose,
    title={Deeppose: Human pose estimation via deep neural networks},
    author={Toshev, Alexander and Szegedy, Christian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={1653--1660},
    year={2014}
}

```

```

@inproceedings{he2016deep,
    title={Deep residual learning for image recognition},
    author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@article{lin2021structure,
    title={Structure-Coherent Deep Feature Learning for Robust Face Alignment},
    author={Lin, Chunze and Zhu, Beier and Wang, Quan and Liao, Renjie and Qian, Chen
↪and Lu, Jiwen and Zhou, Jie},
    journal={IEEE Transactions on Image Processing},
    year={2021},
    publisher={IEEE}
}

```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 31.6.7 Deeppose + Resnet + Wingloss on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{feng2018wing,
  title={Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural
↪Networks},
  author={Feng, Zhen-Hua and Kittler, Josef and Awais, Muhammad and Huber, Patrik and
↪Wu, Xiao-Jun},
  booktitle={Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on},
  year={2018},
  pages = {2235-2245},
  organization={IEEE}
```

(下页继续)

(续上页)

}

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 31.6.8 Deeppose + Resnet on Deepfashion

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{liuLQWtcvpr16DeepFashion,
  author = {Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou},
  title = {DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich
↪Annotations},
  booktitle = {Proceedings of IEEE Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
```

(下页继续)



(续上页)

```

month = {June},
year = {2016}
}

```

```

@inproceedings{liuYLWTeccv16FashionLandmark,
  author = {Liu, Ziwei and Yan, Sijie and Luo, Ping and Wang, Xiaogang and Tang,
↪Xiaoou},
  title = {Fashion Landmark Detection in the Wild},
  booktitle = {European Conference on Computer Vision (ECCV)},
  month = {October},
  year = {2016}
}

```

Results on DeepFashion val set

### 31.6.9 Deeppose + Resnet on Onehand10k

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},

```

(下页继续)

(续上页)

```

    pages={3258--3268},
    year={2018},
    publisher={IEEE}
}

```

Results on OneHand10K val set

### 31.6.10 Deeppose + Resnet on Panoptic2d

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={1145--1153},
  year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 31.6.11 Deeppose + Resnet on Rhd2d

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

## 31.7 RLE (ICCV' 2021)

### 31.7.1 Deeppose + Resnet + Rle on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
```

(下页继续)

(续上页)

```

    year={2014}
}

```

```

@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.7.2 Deeppose + Resnet + Rle on Mpii

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},

```

(下页继续)

(续上页)

```

    year={2014}
}

```

```

@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

## 31.8 SoftWingloss (TIP' 2021)

### 31.8.1 Deeppose + Resnet + Softwingloss on WFLW

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},

```

(下页继续)

(续上页)

```
pages={1653--1660},
year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{lin2021structure,
  title={Structure-Coherent Deep Feature Learning for Robust Face Alignment},
  author={Lin, Chunze and Zhu, Beier and Wang, Quan and Liao, Renjie and Chen
↪and Lu, Jiwen and Zhou, Jie},
  journal={IEEE Transactions on Image Processing},
  year={2021},
  publisher={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 31.9 VideoPose3D (CVPR' 2019)

### 31.9.1 Video Pose Lift + Videopose3d on H36m

```
@inproceedings{pav1lo20193d,
  title={3d human pose estimation in video with temporal convolutions and semi-
↪supervised training},
  author={Pav1lo, Dario and Feichtenhofer, Christoph and Grangier, David and Auli, ↪
↪Michael},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↪
↪Recognition},
  pages={7753--7762},
  year={2019}
}
```

```
@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, ↪
↪Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human ↪
↪Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
  publisher = {IEEE Computer Society},
  volume = {36},
  number = {7},
  pages = {1325-1339},
  month = {jul},
  year = {2014}
}
```

Results on Human3.6M dataset with ground truth 2D detections, supervised training

Results on Human3.6M dataset with CPN 2D detections<sup>1</sup>, supervised training

Results on Human3.6M dataset with ground truth 2D detections, semi-supervised training

Results on Human3.6M dataset with CPN 2D detections<sup>1</sup>, semi-supervised training

<sup>1</sup> CPN 2D detections are provided by [official repo](#). The reformatted version used in this repository can be downloaded from [train\\_detection](#) and [test\\_detection](#).

## 31.9.2 Video Pose Lift + Videopose3d on Mpi\_inf\_3dhp

```
@inproceedings{pav1lo20193d,
  title={3d human pose estimation in video with temporal convolutions and semi-
↪supervised training},
  author={Pav1lo, Dario and Feichtenhofer, Christoph and Grangier, David and Auli, ↪
↪Michael},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↪
↪Recognition},
  pages={7753--7762},
  year={2019}
}
```

```
@inproceedings{mono-3dhp2017,
  author = {Mehta, Dushyant and Rhodin, Helge and Casas, Dan and Fua, Pascal and ↪
↪Sotnychenko, Oleksandr and Xu, Weipeng and Theobalt, Christian},
  title = {Monocular 3D Human Pose Estimation In The Wild Using Improved CNN ↪
↪Supervision},
  booktitle = {3D Vision (3DV), 2017 Fifth International Conference on},
  url = {http://gvv.mpi-inf.mpg.de/3dhp_dataset},
  year = {2017},
  organization={IEEE},
  doi={10.1109/3dv.2017.00064},
}
```

Results on MPI-INF-3DHP dataset with ground truth 2D detections, supervised training

## 31.10 Hourglass (ECCV' 2016)

### 31.10.1 Topdown Heatmap + Hourglass on Coco

```
@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
```

(下页继续)



(续上页)

```

author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
booktitle={European conference on computer vision},
pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.10.2 Topdown Heatmap + Hourglass on Mpii

```

@inproceedings{newell12016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 31.10.3 Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_face

```

@inproceedings{newell12016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},

```

(下页继续)

(续上页)

```
organization={Springer}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 31.10.4 Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

## 31.11 LiteHRNet (CVPR' 2021)

### 31.11.1 Topdown Heatmap + Litehrnet on Coco

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.11.2 Topdown Heatmap + Litehrnet on Mpii

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 31.11.3 Topdown Heatmap + Litehrnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{Yulitehrnet21,  
  title={Lite-HRNet: A Lightweight High-Resolution Network},  
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei  
→ and Sang, Nong and Wang, Jingdong},  
  booktitle={CVPR},  
  year={2021}  
}
```

```
@inproceedings{jin2020whole,  
  title={Whole-Body Human Pose Estimation in the Wild},  
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,  
→ Chen and Ouyang, Wanli and Luo, Ping},  
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},  
  year={2020}  
}
```

Results on COCO-WholeBody-Hand val set

## 31.12 AdaptiveWingloss (ICCV' 2019)

### 31.12.1 Topdown Heatmap + Hrnetv2 + Awing on WFLW

```
@article{WangSCJDZLMTWLX19,  
  title={Deep High-Resolution Representation Learning for Visual Recognition},  
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and  
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and  
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},  
  journal={TPAMI},  
  year={2019}  
}
```

```
@inproceedings{wang2019adaptive,  
  title={Adaptive wing loss for robust face alignment via heatmap regression},  
  author={Wang, Xinyao and Bo, Liefeng and Fuxin, Li},  
  booktitle={Proceedings of the IEEE/CVF international conference on computer vision},  
  pages={6971--6981},  
}
```

(下页继续)

(续上页)

```

    year={2019}
}

```

```

@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}

```

Results on WFLW dataset

The model is trained on WFLW train.

## 31.13 SimpleBaseline2D (ECCV' 2018)

### 31.13.1 Topdown Heatmap + Resnet on Animalpose

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@InProceedings{Cao_2019_ICCV,
  author = {Cao, Jinkun and Tang, Hongyang and Fang, Hao-Shu and Shen, Xiaoyong and
↪Lu, Cewu and Tai, Yu-Wing},
  title = {Cross-Domain Adaptation for Animal Pose Estimation},
  booktitle = {The IEEE International Conference on Computer Vision (ICCV)},
  month = {October},
  year = {2019}
}

```

Results on AnimalPose validation set (1117 instances)

### 31.13.2 Topdown Heatmap + Resnet on Ap10k

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@misc{yu2021ap10k,
  title={AP-10K: A Benchmark for Animal Pose Estimation in the Wild},
  author={Hang Yu and Yufei Xu and Jing Zhang and Wei Zhao and Ziyu Guan and
↪ Dacheng Tao},
  year={2021},
  eprint={2108.12617},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}
```

Results on AP-10K validation set

### 31.13.3 Topdown Heatmap + Resnet on Atrw

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{li2020atrw,
  title={ATRW: A Benchmark for Amur Tiger Re-identification in the Wild},
  author={Li, Shuyuan and Li, Jianguo and Tang, Hanlin and Qian, Rui and Lin, Weiyao},
  booktitle={Proceedings of the 28th ACM International Conference on Multimedia},
  pages={2590--2598},
  year={2020}
}
```

Results on ATRW validation set

### 31.13.4 Topdown Heatmap + Resnet on Fly

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@article{pereira2019fast,
  title={Fast animal pose estimation using deep neural networks},
  author={Pereira, Talmo D and Aldarondo, Diego E and Willmore, Lindsay and Kislin,
↪Mikhail and Wang, Samuel S-H and Murthy, Mala and Shaevitz, Joshua W},
  journal={Nature methods},
  volume={16},
  number={1},
  pages={117--125},
  year={2019},
  publisher={Nature Publishing Group}
}
```

Results on Vinegar Fly test set

### 31.13.5 Topdown Heatmap + Resnet on Horse10

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{mathis2021pretraining,
  title={Pretraining boosts out-of-domain robustness for pose estimation},
  author={Mathis, Alexander and Biasi, Thomas and Schneider, Steffen and Yuksekgonul,
↪Mert and Rogers, Byron and Bethge, Matthias and Mathis, Mackenzie W},
  booktitle={Proceedings of the IEEE/CVF Winter Conference on Applications of
↪Computer Vision},
  pages={1859--1868},
  year={2021}
}
```

Results on Horse-10 test set

### 31.13.6 Topdown Heatmap + Resnet on Locust

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@article{graving2019deepposekit,
  title={DeepPoseKit, a software toolkit for fast and robust animal pose estimation↵
↵using deep learning},
  author={Graving, Jacob M and Chae, Daniel and Naik, Hemal and Li, Liang and Koger,↵
↵Benjamin and Costelloe, Blair R and Couzin, Iain D},
  journal={Elife},
  volume={8},
  pages={e47994},
  year={2019},
  publisher={eLife Sciences Publications Limited}
}
```

Results on Desert Locust test set

### 31.13.7 Topdown Heatmap + Resnet on Macaque

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@article{labuguen2020macaquepose,
  title={MacaquePose: A novel 'in the wild' macaque monkey pose dataset for↵
↵markerless motion capture},
  author={Labuguen, Rollyn and Matsumoto, Jumpei and Negrete, Salvador and Nishimaru,↵
↵Hiroshi and Nishijo, Hisao and Takada, Masahiko and Go, Yasuhiro and Inoue, Ken-
↵ichi and Shibata, Tomohiro},

```

(下页继续)



(续上页)

```

journal={bioRxiv},
year={2020},
publisher={Cold Spring Harbor Laboratory}
}

```

Results on MacaquePose with ground-truth detection bounding boxes

### 31.13.8 Topdown Heatmap + Resnet on Zebra

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@article{graving2019deepposekit,
  title={DeepPoseKit, a software toolkit for fast and robust animal pose estimation.↵
↵using deep learning},
  author={Graving, Jacob M and Chae, Daniel and Naik, Hemal and Li, Liang and Koger,↵
↵Benjamin and Costelloe, Blair R and Couzin, Iain D},
  journal={Elife},
  volume={8},
  pages={e47994},
  year={2019},
  publisher={eLife Sciences Publications Limited}
}

```

Results on Grévy' s Zebra test set

### 31.13.9 Topdown Heatmap + Resnet on Aic

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and
↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC val set with ground-truth bounding boxes

### 31.13.10 Topdown Heatmap + Resnet + Fp16 on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{micikevicius2017mixed,
  title={Mixed precision training},
```

(下页继续)

(续上页)

```

    author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
    ↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
    ↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
    journal={arXiv preprint arXiv:1710.03740},
    year={2017}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
    ↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.13.11 Topdown Heatmap + Resnet + Dark on Coco

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
    ↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},

```

(下页继续)

(续上页)

```

    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
    pages={7093--7102},
    year={2020}
}

```

```

@inproceedings{lin2014microsoft,
    title={Microsoft coco: Common objects in context},
    author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
    booktitle={European conference on computer vision},
    pages={740--755},
    year={2014},
    organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.13.12 Topdown Heatmap + Swin on Coco

```

@inproceedings{xiao2018simple,
    title={Simple baselines for human pose estimation and tracking},
    author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
    booktitle={Proceedings of the European conference on computer vision (ECCV)},
    pages={466--481},
    year={2018}
}

```

```

@inproceedings{liu2021swin,
    title={Swin transformer: Hierarchical vision transformer using shifted windows},
    author={Liu, Ze and Lin, Yutong and Cao, Yue and Hu, Han and Wei, Yixuan and Zhang,
↪Zheng and Lin, Stephen and Guo, Baining},
    booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
    pages={10012--10022},
    year={2021}
}

```

```

@inproceedings{lin2017feature,
    title={Feature pyramid networks for object detection},
    author={Lin, Tsung-Yi and Doll{'a}r, Piotr and Girshick, Ross and He, Kaiming and
↪Hariharan, Bharath and Belongie, Serge},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},

```

(下页继续)

(续上页)

```

    pages={2117--2125},
    year={2017}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.13.13 Topdown Heatmap + Resnet on Coco

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},

```

(下页继续)

(续上页)

```
year={2014},
organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.13.14 Topdown Heatmap + Resnet on Crowdpose

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}
```

Results on CrowdPose test with YOLOv3 human detector

### 31.13.15 Topdown Heatmap + Resnet on JHMDB

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{Jhuang:ICCV:2013,
  title = {Towards understanding action recognition},
  author = {H. Jhuang and J. Gall and S. Zuffi and C. Schmid and M. J. Black},
  booktitle = {International Conf. on Computer Vision (ICCV)},
  month = Dec,
  pages = {3192-3199},
  year = {2013}
}
```

Results on Sub-JHMDB dataset

The models are pre-trained on MPII dataset only. *NO* test-time augmentation (multi-scale /rotation testing) is used.

- Normalized by Person Size
- Normalized by Torso Size

### 31.13.16 Topdown Heatmap + Resnet on MHP

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
```

(下页继续)

(续上页)

```

    year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}

```

Results on MHP v2.0 val set

Note that, the evaluation metric used here is mAP (adapted from COCO), which may be different from the official evaluation codes. Please be cautious if you use the results in papers.

### 31.13.17 Topdown Heatmap + Resnet on Mpii

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},

```

(下页继续)



(续上页)

```

pages={770--778},
year={2016}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 31.13.18 Topdown Heatmap + Resnet + Mpii on Mpii\_trb

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{duan2019trb,
  title={TRB: A Novel Triplet Representation for Understanding 2D Human Body},
  author={Duan, Haodong and Lin, Kwan-Yee and Jin, Sheng and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={9479--9488},
  year={2019}
}

```

Results on MPII-TRB val set

### 31.13.19 Topdown Heatmap + Resnet on Ochuman

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},
  author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↪and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={889--898},
  year={2019}
}
```

Results on OCHuman test dataset with ground-truth bounding boxes

Following the common setting, the models are trained on COCO train dataset, and evaluate on OCHuman dataset.

### 31.13.20 Topdown Heatmap + Resnet on Posetrack18

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
```

(下页继续)

(续上页)

```

    year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{andriluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin,
↪Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition},
  pages={5167--5176},
  year={2018}
}

```

Results on PoseTrack2018 val with ground-truth bounding boxes

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

Results on PoseTrack2018 val with [MMDetection](#) pre-trained [Cascade R-CNN](#) (X-101-64x4d-FPN) human detector

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

### 31.13.21 Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_face

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},

```

(下页继续)

(续上页)

```

    author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@inproceedings{jin2020whole,
    title={Whole-Body Human Pose Estimation in the Wild},
    author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
    booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
    year={2020}
}

```

Results on COCO-WholeBody-Face val set

### 31.13.22 Topdown Heatmap + Resnet on Deepfashion

```

@inproceedings{xiao2018simple,
    title={Simple baselines for human pose estimation and tracking},
    author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
    booktitle={Proceedings of the European conference on computer vision (ECCV)},
    pages={466--481},
    year={2018}
}

```

```

@inproceedings{he2016deep,
    title={Deep residual learning for image recognition},
    author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@inproceedings{liuLQWtcvpr16DeepFashion,
    author = {Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou},
    title = {DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich
↪Annotations},
    booktitle = {Proceedings of IEEE Conference on Computer Vision and Pattern
↪Recognition (CVPR)},

```

(下页继续)

(续上页)

```
month = {June},
year = {2016}
}
```

```
@inproceedings{liuYLWTeccv16FashionLandmark,
  author = {Liu, Ziwei and Yan, Sijie and Luo, Ping and Wang, Xiaogang and Tang,
↪Xiaoou},
  title = {Fashion Landmark Detection in the Wild},
  booktitle = {European Conference on Computer Vision (ECCV)},
  month = {October},
  year = {2016}
}
```

Results on DeepFashion val set

### 31.13.23 Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 31.13.24 Topdown Heatmap + Resnet on Freihand2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{zimmermann2019freihand,
  title={Freihand: A dataset for markerless capture of hand pose and shape from
↪single rgb images},
  author={Zimmermann, Christian and Ceylan, Duygu and Yang, Jimei and Russell, Bryan
↪and Argus, Max and Brox, Thomas},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={813--822},
  year={2019}
}
```

Results on FreiHand val & test set

### 31.13.25 Topdown Heatmap + Resnet on Interhand2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,↪
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose↪
↪Estimation from a Single RGB Image},
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}
```

Results on InterHand2.6M val & test set

### 31.13.26 Topdown Heatmap + Resnet on Onehand10k

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
}
```

(下页继续)

(续上页)

```

journal={IEEE Transactions on Circuits and Systems for Video Technology},
volume={29},
number={11},
pages={3258--3268},
year={2018},
publisher={IEEE}
}

```

Results on OneHand10K val set

### 31.13.27 Topdown Heatmap + Resnet on Panoptic2d

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={1145--1153},
  year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)



### 31.13.28 Topdown Heatmap + Resnet on Rhd2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 31.13.29 Topdown Heatmap + Resnet on Coco-Wholebody

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
```

(下页继续)

(续上页)

```

author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

## 31.14 PoseWarper (NeurIPS' 2019)

### 31.14.1 Posewarper + Hrnet + Posetrack18 on Posetrack18

```

@inproceedings{NIPS2019_gberta,
title = {Learning Temporal Pose Estimation from Sparsely Labeled Videos},
author = {Bertasius, Gedas and Feichtenhofer, Christoph, and Tran, Du and Shi, Jianbo,
↪ and Torresani, Lorenzo},
booktitle = {Advances in Neural Information Processing Systems 33},
year = {2019},
}

```

```

@inproceedings{sun2019deep,
title={Deep high-resolution representation learning for human pose estimation},
author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
pages={5693--5703},
year={2019}
}

```

```

@inproceedings{andrilluka2018posetrack,
title={Posetrack: A benchmark for human pose estimation and tracking},
author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin,
↪ Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪ Recognition},
pages={5167--5176},
year={2018}
}

```

```

@inproceedings{lin2014microsoft,
title={Microsoft coco: Common objects in context},

```

(下页继续)

(续上页)

```

    author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
    booktitle={European conference on computer vision},
    pages={740--755},
    year={2014},
    organization={Springer}
}

```

Note that the training of PoseWarper can be split into two stages.

The first-stage is trained with the pre-trained [checkpoint](#) on COCO dataset, and the main backbone is fine-tuned on PoseTrack18 in a single-frame setting.

The second-stage is trained with the last [checkpoint](#) from the first stage, and the warping offsets are learned in a multi-frame setting while the backbone is frozen.

Results on PoseTrack2018 val with ground-truth bounding boxes

Results on PoseTrack2018 val with precomputed human bounding boxes from PoseWarper supplementary data files from [this link](#)<sup>1</sup>.

<sup>1</sup> Please download the precomputed human bounding boxes on PoseTrack2018 val from `$PoseWarper_supp_files/posetrack18_precomputed_boxes/val_boxes.json` and place it here: `$mmpose/data/posetrack18/posetrack18_precomputed_boxes/val_boxes.json` to be consistent with the [config](#). Please refer to [DATA Preparation](#) for more detail about data preparation.

## 31.15 SimpleBaseline3D (ICCV' 2017)

### 31.15.1 Pose Lift + Simplebaseline3d on H36m

```

@inproceedings{martinez_2017_3dbaseline,
    title={A simple yet effective baseline for 3d human pose estimation},
    author={Martinez, Julieta and Hossain, Rayat and Romero, Javier and Little, James J.
↪},
    booktitle={ICCV},
    year={2017}
}

```

```

@article{h36m_pami,
    author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, 
↪Cristian},
    title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human
↪Sensing in Natural Environments},

```

(下页继续)

(续上页)

```

journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
publisher = {IEEE Computer Society},
volume = {36},
number = {7},
pages = {1325-1339},
month = {jul},
year = {2014}
}

```

Results on Human3.6M dataset with ground truth 2D detections

1 Differing from the original paper, we didn't apply the max-norm constraint because we found this led to a better convergence and performance.

### 31.15.2 Pose Lift + Simplebaseline3d on Mpi\_inf\_3dhp

```

@inproceedings{martinez_2017_3dbaseline,
  title={A simple yet effective baseline for 3d human pose estimation},
  author={Martinez, Julieta and Hossain, Rayat and Romero, Javier and Little, James J.},
  ↪},
  booktitle={ICCV},
  year={2017}
}

```

```

@inproceedings{mono-3dhp2017,
  author = {Mehta, Dushyant and Rhodin, Helge and Casas, Dan and Fua, Pascal and ↪
  ↪Sotnychenko, Oleksandr and Xu, Weipeng and Theobalt, Christian},
  title = {Monocular 3D Human Pose Estimation In The Wild Using Improved CNN ↪
  ↪Supervision},
  booktitle = {3D Vision (3DV), 2017 Fifth International Conference on},
  url = {http://gvv.mpi-inf.mpg.de/3dhp_dataset},
  year = {2017},
  organization={IEEE},
  doi={10.1109/3dv.2017.00064},
}

```

Results on MPI-INF-3DHP dataset with ground truth 2D detections

1 Differing from the original paper, we didn't apply the max-norm constraint because we found this led to a better convergence and performance.

## 31.16 HMR (CVPR' 2018)

### 31.16.1 HMR + Resnet on Mixed

```
@inProceedings{kanazawaHMR18,
  title={End-to-end Recovery of Human Shape and Pose},
  author = {Angjoo Kanazawa
    and Michael J. Black
    and David W. Jacobs
    and Jitendra Malik},
  booktitle={Computer Vision and Pattern Recognition (CVPR)},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, ↪
↪Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human
↪Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
  publisher = {IEEE Computer Society},
  volume = {36},
  number = {7},
  pages = {1325-1339},
  month = {jul},
  year = {2014}
}
```

Results on Human3.6M with ground-truth bounding box having MPJPE-PA of 52.60 mm on Protocol2

## 31.17 UDP (CVPR' 2020)

### 31.17.1 Associative Embedding + Higherhrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↵Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↵Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↵(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↵Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

### 31.17.2 Associative Embedding + Hrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

### 31.17.3 Topdown Heatmap + Hrnet + Udp on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that, UDP also adopts the unbiased encoding/decoding algorithm of [DARK](#).

### 31.17.4 Topdown Heatmap + Hrnetv2 + Udp on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
```

(下页继续)



(续上页)

```

    year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}

```

```

@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}

```

Results on OneHand10K val set

### 31.17.5 Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},

```

(下页继续)

(续上页)

```

month = {June},
year = {2020}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={1145--1153},
  year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 31.17.6 Topdown Heatmap + Hrnetv2 + Udp on Rhd2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
  year = {2020}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
}

```

(下页继续)

(续上页)

```

note="https://arxiv.org/abs/1705.01389",
url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}

```

Results on RHD test set

## 31.18 ViPNAS (CVPR' 2021)

### 31.18.1 Topdown Heatmap + Vipnas on Coco

```

@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen_
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  year={2021}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.18.2 Topdown Heatmap + Vipnas on Coco-Wholebody

```

@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen_
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  year={2021}
}

```

(下页继续)

(续上页)

}

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 31.18.3 Topdown Heatmap + Vipnas + Dark on Coco-Wholebody

```
@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen
↪ and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  year={2021}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

## 31.19 Wingloss (CVPR' 2018)

### 31.19.1 Deeppose + Resnet + Wingloss on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{feng2018wing,
  title={Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural
↪Networks},
  author={Feng, Zhen-Hua and Kittler, Josef and Awais, Muhammad and Huber, Patrik and
↪Wu, Xiao-Jun},
  booktitle={Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on},
  year={2018},
  pages = {2235-2245},
  organization={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 31.20 DarkPose (CVPR' 2020)

### 31.20.1 Topdown Heatmap + Resnet + Dark on Coco

```
@inproceedings{xiao2018simple,  
  title={Simple baselines for human pose estimation and tracking},  
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},  
  booktitle={Proceedings of the European conference on computer vision (ECCV)},  
  pages={466--481},  
  year={2018}  
}
```

```
@inproceedings{he2016deep,  
  title={Deep residual learning for image recognition},  
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
→recognition},  
  pages={770--778},  
  year={2016}  
}
```

```
@inproceedings{zhang2020distribution,  
  title={Distribution-aware coordinate representation for human pose estimation},  
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},  
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern  
→Recognition},  
  pages={7093--7102},  
  year={2020}  
}
```

```
@inproceedings{lin2014microsoft,  
  title={Microsoft coco: Common objects in context},  
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and  
→Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},  
  booktitle={European conference on computer vision},  
  pages={740--755},  
  year={2014},  
  organization={Springer}
```

(下页继续)

(续上页)

}

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.20.2 Topdown Heatmap + Hrnet + Dark on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.20.3 Topdown Heatmap + Hrnet + Dark on Mpii

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 31.20.4 Topdown Heatmap + Hrnetv2 + Dark on Aflw

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```



```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database
↵for facial landmark localization},
  author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
  booktitle={2011 IEEE international conference on computer vision workshops (ICCV
↵workshops)},
  pages={2144--2151},
  year={2011},
  organization={IEEE}
}
```

Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

### 31.20.5 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
```

(下页继续)

(续上页)

```

    year={2020}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Face val set

### 31.20.6 Topdown Heatmap + Hrnetv2 + Dark on WFLW

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪ Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={2129--2138},
  year={2018}
}

```

Results on WFLW dataset

The model is trained on WFLW train.

### 31.20.7 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 31.20.8 Topdown Heatmap + Hrnetv2 + Dark on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  year={2019}
}
```

(下页继续)

(续上页)

```

journal={TPAMI},
year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}

```

Results on OneHand10K val set

### 31.20.9 Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},

```

(下页继续)

(续上页)

```

pages={7093--7102},
year={2020}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},
  year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 31.20.10 Topdown Heatmap + Hrnetv2 + Dark on Rhd2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
}

```

(下页继续)

(续上页)

```
url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 31.20.11 Topdown Heatmap + Vipnas + Dark on Coco-Wholebody

```
@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  year={2021}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 31.20.12 Topdown Heatmap + Hrnet + Dark on Coco-Wholebody

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 31.20.13 Topdown Heatmap + Hrnet + Dark on Halpe

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{li2020pastanet,
  title={PaStaNet: Toward Human Activity Knowledge Engine},
  author={Li, Yong-Lu and Xu, Liang and Liu, Xinpeng and Huang, Xijie and Xu, Yue and
↪Wang, Shiyi and Fang, Hao-Shu and Ma, Ze and Chen, Mingyang and Lu, Cewu},
  booktitle={CVPR},
  year={2020}
}
```

Results on Halpe v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on Halpe dataset. We find this will lead to better performance.

## 31.21 Associative Embedding (NIPS' 2017)

### 31.21.1 Associative Embedding + Hrnet on Aic

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```



```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding},
  ↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and ↪
  ↪Liang, Rui and Wang, Wenjia and Zhou, Shipai and Lin, Guosen and Fu, Yanwei and ↪
  ↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.21.2 Associative Embedding + Higherhrnet on Aic

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose ↪
  ↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang, ↪
  ↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↪
  ↪Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
  ↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and ↪
  ↪Liang, Rui and Wang, Wenjia and Zhou, Shipai and Lin, Guosen and Fu, Yanwei and ↪
  ↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

(下页继续)

(续上页)

}

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.21.3 Associative Embedding + Higherhrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose↵
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,↵
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern↵
↵Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for↵
↵Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition↵
↵(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and↵
↵Perona, Pietro and Ramanan, Deva and Doll{'a'}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},

```

(下页继续)

(续上页)

```

pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 without multi-scale test

### 31.21.4 Associative Embedding + Mobilenetv2 on Coco

```

@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}

```

```

@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.21.5 Associative Embedding + Hrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.21.6 Associative Embedding + Higherhrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.21.7 Associative Embedding + Resnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.21.8 Associative Embedding + Hrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

### 31.21.9 Associative Embedding + Hourglass + Ae on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.21.10 Associative Embedding + Higherhrnet on Crowdpose

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu_
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}
```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 2 scales ([2, 1]) are used

### 31.21.11 Associative Embedding + Hrnet on MHP

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```



```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}
```

Results on MHP v2.0 validation set without multi-scale test

Results on MHP v2.0 validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.21.12 Associative Embedding + Hrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 31.21.13 Associative Embedding + Higherhrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪ Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang, _
↪ Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪ Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

## 31.22 VoxelPose (ECCV' 2020)

### 31.22.1 Voxelpose + Voxelpose on Campus

```
@inproceedings{tumultipose,
  title={VoxelPose: Towards Multi-Camera 3D Human Pose Estimation in Wild Environment},
  ↪,
  author={Tu, Hanyue and Wang, Chunyu and Zeng, Wenjun},
  booktitle={ECCV},
  year={2020}
}
```

```
@inproceedings {belagian14multi,
  title = {{3D} Pictorial Structures for Multiple Human Pose Estimation},
  author = {Belagiannis, Vasileios and Amin, Sikandar and Andriluka, Mykhaylo and ↪
  ↪Schiele, Bernt and Navab
  ↪Nassir and Ilic, Slobodan},
  booktitle = {IEEE Computer Society Conference on Computer Vision and Pattern ↪
  ↪Recognition (CVPR)},
  year = {2014},
  month = {June},
  organization={IEEE}
}
```

Results on Campus dataset.

### 31.22.2 Voxelpose + Voxelpose + Prn64x64x64 + Cpn80x80x20 + Panoptic on Panoptic

```
@inproceedings{tumultipose,
  title={VoxelPose: Towards Multi-Camera 3D Human Pose Estimation in Wild Environment}
  ↪,
  author={Tu, Hanyue and Wang, Chunyu and Zeng, Wenjun},
  booktitle={ECCV},
  year={2020}
}
```

```
@Article = {joo_iccv_2015,
author = {Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo_
↪Kanade, Shohei Nobuhara, and Yaser Sheikh},
title = {Panoptic Studio: A Massively Multiview System for Social Motion Capture},
booktitle = {ICCV},
year = {2015}
}
```

Results on CMU Panoptic dataset.

### 31.22.3 Voxelpose + Voxelpose on Shelf

```
@inproceedings{tumultipose,
title={VoxelPose: Towards Multi-Camera 3D Human Pose Estimation in Wild Environment}
↪,
author={Tu, Hanyue and Wang, Chunyu and Zeng, Wenjun},
booktitle={ECCV},
year={2020}
}
```

```
@inproceedings {belagian14multi,
title = {{3D} Pictorial Structures for Multiple Human Pose Estimation},
author = {Belagiannis, Vasileios and Amin, Sikandar and Andriluka, Mykhaylo and_
↪Schiele, Bernt and Navab
Nassir and Ilic, Slobodan},
booktitle = {IEEE Computer Society Conference on Computer Vision and Pattern_
↪Recognition (CVPR)},
year = {2014},
month = {June},
organization={IEEE}
}
```

Results on Shelf dataset.

## 31.23 RSN (ECCV' 2020)

### 31.23.1 Topdown Heatmap + RSN on Coco

```
@misc{cai2020learning,
title={Learning Delicate Local Representations for Multi-Person Pose Estimation},
author={Yuanhao Cai and Zhicheng Wang and Zhengxiong Luo and Binyi Yin and Angang_
↪Du and Haoqian Wang and Xinyu Zhou and Erjin Zhou and Xiangyu Zhang and Jian S(下页继续)
```

(续上页)

```

    year={2020},
    eprint={2003.04030},
    archivePrefix={arXiv},
    primaryClass={cs.CV}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 31.24 CID (CVPR' 2022)

### 31.24.1 Cid + Hrnet on Coco

```

@InProceedings{Wang_2022_CVPR,
  author    = {Wang, Dongkai and Zhang, Shiliang},
  title     = {Contextual Instance Decoupling for Robust Multi-Person Pose
↪Estimation},
  booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and
↪Pattern Recognition (CVPR)},
  month     = {June},
  year      = {2022},
  pages     = {11060-11068}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},

```

(下页继续)

(续上页)

```
organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

## 31.25 CPM (CVPR' 2016)

### 31.25.1 Topdown Heatmap + CPM on Coco

```
@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={4724--4732},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.25.2 Topdown Heatmap + CPM on JHMDB

```
@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={4724--4732},
  year={2016}
}
```

```
@inproceedings{Jhuang:ICCV:2013,
  title = {Towards understanding action recognition},
  author = {H. Jhuang and J. Gall and S. Zuffi and C. Schmid and M. J. Black},
  booktitle = {International Conf. on Computer Vision (ICCV)},
  month = Dec,
  pages = {3192-3199},
  year = {2013}
}
```

Results on Sub-JHMDB dataset

The models are pre-trained on MPII dataset only. NO test-time augmentation (multi-scale /rotation testing) is used.

- Normalized by Person Size
- Normalized by Torso Size

### 31.25.3 Topdown Heatmap + CPM on Mpii

```
@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={4724--4732},
  year={2016}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 31.26 HRNet (CVPR' 2019)

### 31.26.1 Topdown Heatmap + Hrnet on Animalpose

```
@inproceedings{sun2019deep,  
  title={Deep high-resolution representation learning for human pose estimation},  
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={5693--5703},  
  year={2019}  
}
```

```
@InProceedings{Cao_2019_ICCV,  
  author = {Cao, Jinkun and Tang, Hongyang and Fang, Hao-Shu and Shen, Xiaoyong and  
↪Lu, Cewu and Tai, Yu-Wing},  
  title = {Cross-Domain Adaptation for Animal Pose Estimation},  
  booktitle = {The IEEE International Conference on Computer Vision (ICCV)},  
  month = {October},  
  year = {2019}  
}
```

Results on AnimalPose validation set (1117 instances)

### 31.26.2 Topdown Heatmap + Hrnet on Ap10k

```
@inproceedings{sun2019deep,  
  title={Deep high-resolution representation learning for human pose estimation},  
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={5693--5703},  
  year={2019}  
}
```

```
@misc{yu2021ap10k,  
  title={AP-10K: A Benchmark for Animal Pose Estimation in the Wild},  
  author={Hang Yu and Yufei Xu and Jing Zhang and Wei Zhao and Ziyu Guan and  
↪Dacheng Tao},  
  year={2021},  
  eprint={2108.12617},  
  archivePrefix={arXiv},
```

(下页继续)



(续上页)

```
primaryClass={cs.CV}
}
```

Results on AP-10K validation set

### 31.26.3 Topdown Heatmap + Hrnet on Atrw

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{li2020atrw,
  title={ATRW: A Benchmark for Amur Tiger Re-identification in the Wild},
  author={Li, Shuyuan and Li, Jianguo and Tang, Hanlin and Qian, Rui and Lin, Weiyao},
  booktitle={Proceedings of the 28th ACM International Conference on Multimedia},
  pages={2590--2598},
  year={2020}
}
```

Results on ATRW validation set

### 31.26.4 Topdown Heatmap + Hrnet on Horse10

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{mathis2021pretraining,
  title={Pretraining boosts out-of-domain robustness for pose estimation},
  author={Mathis, Alexander and Biasi, Thomas and Schneider, Steffen and Yuksekgonul,
↪Mert and Rogers, Byron and Bethge, Matthias and Mathis, Mackenzie W},
```

(下页继续)

(续上页)

```

    booktitle={Proceedings of the IEEE/CVF Winter Conference on Applications of
↪Computer Vision},
    pages={1859--1868},
    year={2021}
}

```

Results on Horse-10 test set

### 31.26.5 Topdown Heatmap + Hrnet on Macaque

```

@inproceedings{sun2019deep,
    title={Deep high-resolution representation learning for human pose estimation},
    author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={5693--5703},
    year={2019}
}

```

```

@article{labuguen2020macaquepose,
    title={MacaquePose: A novel ‘in the wild’ macaque monkey pose dataset for
↪markerless motion capture},
    author={Labuguen, Rollyn and Matsumoto, Jumpei and Negrete, Salvador and Nishimaru,
↪Hiroshi and Nishijo, Hisao and Takada, Masahiko and Go, Yasuhiro and Inoue, Ken-
↪ichi and Shibata, Tomohiro},
    journal={bioRxiv},
    year={2020},
    publisher={Cold Spring Harbor Laboratory}
}

```

Results on MacaquePose with ground-truth detection bounding boxes

### 31.26.6 Associative Embedding + Hrnet on Aic

```

@inproceedings{newell2017associative,
    title={Associative embedding: End-to-end learning for joint detection and grouping},
    author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
    booktitle={Advances in neural information processing systems},
    pages={2277--2287},
    year={2017}
}

```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and
↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.26.7 Topdown Heatmap + Hrnet on Aic

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and
↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC val set with ground-truth bounding boxes

### 31.26.8 Associative Embedding + Hrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.26.9 Associative Embedding + Hrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
```

(下页继续)

(续上页)

```

    year={2017}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 without multi-scale test

### 31.26.10 Dekr + Hrnet on Coco

```

@inproceedings{geng2021bottom,
  title={Bottom-up human pose estimation via disentangled keypoint regression},
  author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,_
↪Jingdong},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},

```

(下页继续)

(续上页)

```

pages={14676--14686},
year={2021}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

The results of models provided by the authors on COCO val2017 using the same evaluation protocol

The discrepancy between these results and that shown in paper is attributed to the differences in implementation details in evaluation process.

### 31.26.11 Topdown Heatmap + Hrnet + Fp16 on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```
@article{micikevicius2017mixed,
  title={Mixed precision training},
  author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
  journal={arXiv preprint arXiv:1710.03740},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.26.12 Topdown Heatmap + Hrnet on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.26.13 Topdown Heatmap + Hrnet + Augmentation on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{buslaev2020albumentations,
  title={Albumentations: fast and flexible image augmentations},
  author={Buslaev, Alexander and Iglovikov, Vladimir I and Khvedchenya, Eugene and
↪Parinov, Alex and Druzhinin, Mikhail and Kalinin, Alexandr A},
  journal={Information},
  volume={11},
  number={2},
  pages={125},
  year={2020},
  publisher={Multidisciplinary Digital Publishing Institute}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.26.14 Topdown Heatmap + Hrnet + Udp on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
```

(下页继续)



(续上页)

```

    year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that, UDP also adopts the unbiased encoding/decoding algorithm of [DARK](#).

### 31.26.15 Topdown Heatmap + Hrnet + Dark on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},

```

(下页继续)

(续上页)

```

    pages={7093--7102},
    year={2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.26.16 Dekr + Hrnet on Crowdpose

```

@inproceedings{geng2021bottom,
  title={Bottom-up human pose estimation via disentangled keypoint regression},
  author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,
↪Jingdong},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={14676--14686},
  year={2021}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},

```

(下页继续)

(续上页)

```

    year={2018}
}

```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

### 31.26.17 Topdown Heatmap + Hrnet on Crowdpose

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

Results on CrowdPose test with YOLOv3 human detector

### 31.26.18 Topdown Heatmap + Hrnet on H36m

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```
@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, ↵
↵Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human↵
↵Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
  publisher = {IEEE Computer Society},
  volume = {36},
  number = {7},
  pages = {1325-1339},
  month = {jul},
  year = {2014}
}
```

Results on Human3.6M test set with ground truth 2D detections

### 31.26.19 Associative Embedding + Hrnet on MHP

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern↵
↵recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and↵
↵a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,↵
↵Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
}
```

(下页继续)

(续上页)

```

    year={2018}
}

```

Results on MHP v2.0 validation set without multi-scale test

Results on MHP v2.0 validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 31.26.20 Topdown Heatmap + Hrnet on Mpii

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 31.26.21 Topdown Heatmap + Hrnet + Dark on Mpii

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 31.26.22 Topdown Heatmap + Hrnet on Ochuman

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↵recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},
  author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↵and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↵recognition},
  pages={889--898},
  year={2019}
}
```

Results on OCHuman test dataset with ground-truth bounding boxes

Following the common setting, the models are trained on COCO train dataset, and evaluate on OCHuman dataset.

### 31.26.23 Topdown Heatmap + Hrnet on Posetrack18

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin,
↪Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition},
  pages={5167--5176},
  year={2018}
}
```

Results on PoseTrack2018 val with ground-truth bounding boxes

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

Results on PoseTrack2018 val with [MMDetection](#) pre-trained [Cascade R-CNN](#) (X-101-64x4d-FPN) human detector

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

### 31.26.24 Posewarper + Hrnet + Posetrack18 on Posetrack18

```
@inproceedings{NIPS2019_gberta,
  title = {Learning Temporal Pose Estimation from Sparsely Labeled Videos},
  author = {Bertasius, Gedas and Feichtenhofer, Christoph, and Tran, Du and Shi, Jianbo,
↪ and Torresani, Lorenzo},
  booktitle = {Advances in Neural Information Processing Systems 33},
  year = {2019},
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
```

(下页继续)

(续上页)

```

    pages={5693--5703},
    year={2019}
}

```

```

@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin, Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={5167--5176},
  year={2018}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Note that the training of PoseWarper can be split into two stages.

The first-stage is trained with the pre-trained [checkpoint](#) on COCO dataset, and the main backbone is fine-tuned on PoseTrack18 in a single-frame setting.

The second-stage is trained with the last [checkpoint](#) from the first stage, and the warping offsets are learned in a multi-frame setting while the backbone is frozen.

Results on PoseTrack2018 val with ground-truth bounding boxes

Results on PoseTrack2018 val with precomputed human bounding boxes from PoseWarper supplementary data files from [this link](#)<sup>1</sup>.

<sup>1</sup> Please download the precomputed human bounding boxes on PoseTrack2018 val from `$PoseWarper_supp_files/posetrack18_precomputed_boxes/val_boxes.json` and place it here: `$mmpose/data/posetrack18/posetrack18_precomputed_boxes/val_boxes.json` to be consistent with the `config`. Please refer to [DATA Preparation](#) for more detail about data preparation.



### 31.26.25 Associative Embedding + Hrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 31.26.26 Topdown Heatmap + Hrnet on Coco-Wholebody

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 31.26.27 Topdown Heatmap + Hrnet + Dark on Coco-Wholebody

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

## 31.26.28 Topdown Heatmap + Hrnet + Dark on Halpe

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{li2020pastanet,
  title={PaStaNet: Toward Human Activity Knowledge Engine},
  author={Li, Yong-Lu and Xu, Liang and Liu, Xinpeng and Huang, Xijie and Xu, Yue and
↪Wang, Shiyi and Fang, Hao-Shu and Ma, Ze and Chen, Mingyang and Lu, Cewu},
  booktitle={CVPR},
  year={2020}
}
```

Results on Halpe v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on Halpe dataset. We find this will lead to better performance.

## 31.27 HRNetv2 (TPAMI' 2019)

### 31.27.1 Topdown Heatmap + Hrnetv2 on 300w

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  year={2019}
```

(下页继续)

(续上页)

```

journal={TPAMI},
year={2019}
}

```

```

@article{sagonas2016300,
  title={300 faces in-the-wild challenge: Database and results},
  author={Sagonas, Christos and Antonakos, Epameinondas and Tzimiropoulos, Georgios_
↪and Zafeiriou, Stefanos and Pantic, Maja},
  journal={Image and vision computing},
  volume={47},
  pages={3--18},
  year={2016},
  publisher={Elsevier}
}

```

Results on 300W dataset

The model is trained on 300W train.

### 31.27.2 Topdown Heatmap + Hrnetv2 + Dark on Aflw

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database_
↪for facial landmark localization},

```

(下页继续)

(续上页)

```

    author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
    booktitle={2011 IEEE international conference on computer vision workshops (ICCV_
↪workshops)},
    pages={2144--2151},
    year={2011},
    organization={IEEE}
}

```

Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

### 31.27.3 Topdown Heatmap + Hrnetv2 on Aflw

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database_
↪for facial landmark localization},
  author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
  booktitle={2011 IEEE international conference on computer vision workshops (ICCV_
↪workshops)},
  pages={2144--2151},
  year={2011},
  organization={IEEE}
}

```

Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

### 31.27.4 Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_face

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
    ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 31.27.5 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
    ↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 31.27.6 Topdown Heatmap + Hrnetv2 on Cofw

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{burgos2013robust,
  title={Robust face landmark estimation under occlusion},
  author={Burgos-Artizzu, Xavier P and Perona, Pietro and Doll{'a}r, Piotr},
  booktitle={Proceedings of the IEEE international conference on computer vision},
  pages={1513--1520},
  year={2013}
}
```

Results on COFW dataset

The model is trained on COFW train.

### 31.27.7 Topdown Heatmap + Hrnetv2 + Awing on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wang2019adaptive,
  title={Adaptive wing loss for robust face alignment via heatmap regression},
  author={Wang, Xinyao and Bo, Liefeng and Fuxin, Li},
  booktitle={Proceedings of the IEEE/CVF international conference on computer vision},
  pages={6971--6981},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 31.27.8 Topdown Heatmap + Hrnetv2 + Dark on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```



```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 31.27.9 Topdown Heatmap + Hrnetv2 on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 31.27.10 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↵ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 31.27.11 Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 31.27.12 Topdown Heatmap + Hrnetv2 + Dark on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 31.27.13 Topdown Heatmap + Hrnetv2 on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 31.27.14 Topdown Heatmap + Hrnetv2 + Udp on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪ Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪ (CVPR)},
  month = {June},
```

(下页继续)

(续上页)

```

    year = {2020}
}

```

```

@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}

```

Results on OneHand10K val set

### 31.27.15 Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},

```

(下页继续)

(续上页)

```

    year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 31.27.16 Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern_
↪Recognition},
  pages={1145--1153},
  year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 31.27.17 Topdown Heatmap + Hrnetv2 on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 31.27.18 Topdown Heatmap + Hrnetv2 + Udp on Rh2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↵Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↵(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 31.27.19 Topdown Heatmap + Hrnetv2 + Dark on Rhd2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set



## 31.27.20 Topdown Heatmap + Hrnetv2 on Rhd2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

## 31.28 SCNet (CVPR' 2020)

### 31.28.1 Topdown Heatmap + Snet on Coco

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
    ↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
    ↪Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
    ↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},

```

(下页继续)

(续上页)

```

pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 31.28.2 Topdown Heatmap + Snet on Mpii

```

@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 31.28.3 Topdown Heatmap + Snet + Coco + Wholebody on Coco\_wholebody\_face

```

@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}

```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 31.28.4 Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪ Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set



### 32.1 MSPN (ArXiv' 2019)

#### 32.1.1 Topdown Heatmap + MSPN on Coco

```
@article{li2019rethinking,
  title={Rethinking on Multi-Stage Networks for Human Pose Estimation},
  author={Li, Wenbo and Wang, Zhicheng and Yin, Binyi and Peng, Qixiang and Du,
↪Yuming and Xiao, Tianzi and Yu, Gang and Lu, Hongtao and Wei, Yichen and Sun, Jian},
  journal={arXiv preprint arXiv:1901.00148},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.2 MobilenetV2 (CVPR' 2018)

### 32.2.1 Associative Embedding + Mobilenetv2 on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 32.2.2 Topdown Heatmap + Mobilenetv2 on Coco

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.2.3 Topdown Heatmap + Mobilenetv2 on Mpii

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
}
```

(下页继续)

(续上页)

```

    month = {June}
}

```

Results on MPII val set

### 32.2.4 Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_face

```

@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Face val set

### 32.2.5 Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_hand

```

@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}

```



```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 32.2.6 Topdown Heatmap + Mobilenetv2 on Onehand10k

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey↪
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern↪
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 32.2.7 Topdown Heatmap + Mobilenetv2 on Panoptic2d

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern_
↪Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 32.2.8 Topdown Heatmap + Mobilenetv2 on Rhd2d

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

## 32.3 HigherHRNet (CVPR' 2020)

### 32.3.1 Associative Embedding + Higherhrnet on Aic

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose↵
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,↵
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern↵
↵Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}↵
↵,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and↵
↵Liang, Rui and Wang, Wenjia and Zhou, Shiwei and Lin, Guosen and Fu, Yanwei and↵
↵others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 32.3.2 Associative Embedding + Higherhrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

### 32.3.3 Associative Embedding + Higherhrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 32.3.4 Associative Embedding + Higherhrnet on Crowdpote

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
```

(下页继续)

(续上页)

```

    year={2017}
}

```

```

@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose↵
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,↵
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern↵
↵Recognition},
  pages={5386--5395},
  year={2020}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu↵
↵and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 2 scales ([2, 1]) are used

### 32.3.5 Associative Embedding + Higherhrnet on Coco-Wholebody

```

@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}

```

```

@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose↵
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,↵
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern↵
↵Recognition},

```

(下页继续)

(续上页)

```

pages={5386--5395},
year={2020}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

## 32.4 ResNeSt (ArXiv' 2020)

### 32.4.1 Topdown Heatmap + Resnest on Coco

```

@article{zhang2020resnest,
  title={ResNeSt: Split-Attention Networks},
  author={Zhang, Hang and Wu, Chongruo and Zhang, Zhongyue and Zhu, Yi and Zhang, Zhi-
↪ and Lin, Haibin and Sun, Yue and He, Tong and Muller, Jonas and Manmatha, R. and Li,
↪ Mu and Smola, Alexander},
  journal={arXiv preprint arXiv:2004.08955},
  year={2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪ Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.5 ResNext (CVPR' 2017)

### 32.5.1 Topdown Heatmap + Resnext on Coco

```
@inproceedings{xie2017aggregated,
  title={Aggregated residual transformations for deep neural networks},
  author={Xie, Saining and Girshick, Ross and Doll{\`a}r, Piotr and Tu, Zhuowen and
↪He, Kaiming},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1492--1500},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.5.2 Topdown Heatmap + Resnext on Mpii

```
@inproceedings{xie2017aggregated,
  title={Aggregated residual transformations for deep neural networks},
  author={Xie, Saining and Girshick, Ross and Doll{\`a}r, Piotr and Tu, Zhuowen and
↪He, Kaiming},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1492--1500},
  year={2017}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
```

(下页继续)



(续上页)

```

booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
year = {2014},
month = {June}
}

```

Results on MPII val set

## 32.6 ResNet (CVPR' 2016)

### 32.6.1 Topdown Heatmap + Resnet on Aic

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and
↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}

```

Results on AIC val set with ground-truth bounding boxes

### 32.6.2 Associative Embedding + Resnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 32.6.3 Deeppose + Resnet on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.6.4 Deeppose + Resnet + Rle on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.6.5 Topdown Heatmap + Resnet + Fp16 on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{micikevicius2017mixed,
  title={Mixed precision training},
```

(下页继续)

(续上页)

```

    author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
    ↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
    ↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
    journal={arXiv preprint arXiv:1710.03740},
    year={2017}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
    ↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.6.6 Topdown Heatmap + Resnet + Dark on Coco

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
    ↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},

```

(下页继续)

(续上页)

```

    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
    pages={7093--7102},
    year={2020}
}

```

```

@inproceedings{lin2014microsoft,
    title={Microsoft coco: Common objects in context},
    author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
    booktitle={European conference on computer vision},
    pages={740--755},
    year={2014},
    organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.6.7 Topdown Heatmap + Resnet on Coco

```

@inproceedings{xiao2018simple,
    title={Simple baselines for human pose estimation and tracking},
    author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
    booktitle={Proceedings of the European conference on computer vision (ECCV)},
    pages={466--481},
    year={2018}
}

```

```

@inproceedings{he2016deep,
    title={Deep residual learning for image recognition},
    author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={770--778},
    year={2016}
}

```

```

@inproceedings{lin2014microsoft,
    title={Microsoft coco: Common objects in context},
    author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
    booktitle={European conference on computer vision},

```

(下页继续)

(续上页)

```

pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.6.8 Topdown Heatmap + Resnet on Crowdpose

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

Results on CrowdPose test with YOLOv3 human detector

### 32.6.9 Topdown Heatmap + Resnet on JHMDB

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{Jhuang:ICCV:2013,
  title = {Towards understanding action recognition},
  author = {H. Jhuang and J. Gall and S. Zuffi and C. Schmid and M. J. Black},
  booktitle = {International Conf. on Computer Vision (ICCV)},
  month = Dec,
  pages = {3192-3199},
  year = {2013}
}
```

Results on Sub-JHMDB dataset

The models are pre-trained on MPII dataset only. *NO* test-time augmentation (multi-scale /rotation testing) is used.

- Normalized by Person Size
- Normalized by Torso Size

### 32.6.10 Topdown Heatmap + Resnet on MHP

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
```

(下页继续)



(续上页)

```

    year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}

```

Results on MHP v2.0 val set

Note that, the evaluation metric used here is mAP (adapted from COCO), which may be different from the official evaluation codes. Please be cautious if you use the results in papers.

### 32.6.11 Deeppose + Resnet on Mpii

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},

```

(下页继续)

(续上页)

```

    pages={770--778},
    year={2016}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 32.6.12 Deeppose + Resnet + Rle on Mpii

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}

```

```

@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},

```

(下页继续)

(续上页)

```

    year={2016}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 32.6.13 Topdown Heatmap + Resnet on Mpii

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 32.6.14 Topdown Heatmap + Resnet + Mpii on Mpii\_trb

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{duan2019trb,
  title={TRB: A Novel Triplet Representation for Understanding 2D Human Body},
  author={Duan, Haodong and Lin, Kwan-Yee and Jin, Sheng and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={9479--9488},
  year={2019}
}
```

Results on MPII-TRB val set

### 32.6.15 Topdown Heatmap + Resnet on Ochuman

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},
  author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↪and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={889--898},
  year={2019}
}
```

Results on OCHuman test dataset with ground-truth bounding boxes

Following the common setting, the models are trained on COCO train dataset, and evaluate on OCHuman dataset.

### 32.6.16 Topdown Heatmap + Resnet on Posetrack18

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andrilluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin, Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={5167--5176},
  year={2018}
}
```

Results on PoseTrack2018 val with ground-truth bounding boxes

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

Results on PoseTrack2018 val with MMDetection pre-trained Cascade R-CNN (X-101-64x4d-FPN) human detector

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

### 32.6.17 HMR + Resnet on Mixed

```
@inProceedings{kanazawaHMR18,
  title={End-to-end Recovery of Human Shape and Pose},
  author = {Angjoo Kanazawa
    and Michael J. Black
    and David W. Jacobs
    and Jitendra Malik},
  booktitle={Computer Vision and Pattern Recognition (CVPR)},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments},
}
```

(下页继续)

(续上页)

```

journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
publisher = {IEEE Computer Society},
volume = {36},
number = {7},
pages = {1325-1339},
month = {jul},
year = {2014}
}

```

Results on Human3.6M with ground-truth bounding box having MPJPE-PA of 52.60 mm on Protocol2

### 32.6.18 Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_face

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Face val set

### 32.6.19 Deeppose + Resnet on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 32.6.20 Deeppose + Resnet + Softwingloss on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
```

(下页继续)



(续上页)

```

    year={2014}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@article{lin2021structure,
  title={Structure-Coherent Deep Feature Learning for Robust Face Alignment},
  author={Lin, Chunze and Zhu, Beier and Wang, Quan and Liao, Renjie and Qian, Chen
↪and Lu, Jiwen and Zhou, Jie},
  journal={IEEE Transactions on Image Processing},
  year={2021},
  publisher={IEEE}
}

```

```

@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}

```

Results on WFLW dataset

The model is trained on WFLW train.

### 32.6.21 Deeppose + Resnet + Wingloss on WFLW

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},

```

(下页继续)

(续上页)

```
pages={1653--1660},
year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↵recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{feng2018wing,
  title={Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural_
↵Networks},
  author={Feng, Zhen-Hua and Kittler, Josef and Awais, Muhammad and Huber, Patrik and_
↵Wu, Xiao-Jun},
  booktitle={Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on},
  year={2018},
  pages = {2235-2245},
  organization={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and_
↵Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↵recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 32.6.22 Deeppose + Resnet on Deepfashion

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{liuLQWtcvpr16DeepFashion,
  author = {Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou},
  title = {DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich
↪Annotations},
  booktitle = {Proceedings of IEEE Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
  month = {June},
  year = {2016}
}
```

```
@inproceedings{liuYLWTeccv16FashionLandmark,
  author = {Liu, Ziwei and Yan, Sijie and Luo, Ping and Wang, Xiaogang and Tang,
↪Xiaoou},
  title = {Fashion Landmark Detection in the Wild},
  booktitle = {European Conference on Computer Vision (ECCV)},
  month = {October},
  year = {2016}
}
```

Results on DeepFashion val set

### 32.6.23 Topdown Heatmap + Resnet on Deepfashion

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{liuLQWtcvpr16DeepFashion,
  author = {Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou},
  title = {DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich
↪Annotations},
  booktitle = {Proceedings of IEEE Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
  month = {June},
  year = {2016}
}
```

```
@inproceedings{liuYLWTeccv16FashionLandmark,
  author = {Liu, Ziwei and Yan, Sijie and Luo, Ping and Wang, Xiaogang and Tang,
↪Xiaoou},
  title = {Fashion Landmark Detection in the Wild},
  booktitle = {European Conference on Computer Vision (ECCV)},
  month = {October},
  year = {2016}
}
```

Results on DeepFashion val set

### 32.6.24 Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 32.6.25 Topdown Heatmap + Resnet on Freihand2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
```

(下页继续)

(续上页)

```

author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
pages={770--778},
year={2016}
}

```

```

@inproceedings{zimmermann2019freihand,
  title={Freihand: A dataset for markerless capture of hand pose and shape from
↪single rgb images},
  author={Zimmermann, Christian and Ceylan, Duygu and Yang, Jimei and Russell, Bryan
↪and Argus, Max and Brox, Thomas},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={813--822},
  year={2019}
}

```

Results on FreiHand val &amp; test set

### 32.6.26 Topdown Heatmap + Resnet on Interhand2d

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},

```

(下页继续)

(续上页)

```
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}
```

Results on InterHand2.6M val & test set

### 32.6.27 Deeppose + Resnet on Onehand10k

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 32.6.28 Topdown Heatmap + Resnet on Onehand10k

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 32.6.29 Deeppose + Resnet on Panoptic2d

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```



```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 32.6.30 Topdown Heatmap + Resnet on Panoptic2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
```

(下页继续)

(续上页)

```

    booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
    pages={1145--1153},
    year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 32.6.31 Deeppose + Resnet on Rhd2d

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}

```

Results on RHD test set

### 32.6.32 Topdown Heatmap + Resnet on Rhd2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 32.6.33 Internet + Internet on Interhand3d

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,
  author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
  title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},
  booktitle = {European Conference on Computer Vision (ECCV)},
  year = {2020}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}
```

Results on InterHand2.6M val & test set

## 32.7 Hourglass (ECCV' 2016)

### 32.7.1 Topdown Heatmap + Hourglass on Coco

```
@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a'}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.7.2 Topdown Heatmap + Hourglass on Mpii

```
@inproceedings{newell12016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 32.7.3 Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_face

```
@inproceedings{newell12016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
```

(下页继续)

(续上页)

```
year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 32.7.4 Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
  ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

## 32.8 ShufflenetV1 (CVPR' 2018)

### 32.8.1 Topdown Heatmap + Shufflenetv1 on Coco

```
@inproceedings{zhang2018shufflenet,
  title={Shufflenet: An extremely efficient convolutional neural network for mobile_
  ↪ devices},
  author={Zhang, Xiangyu and Zhou, Xinyu and Lin, Mengxiao and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
  ↪ recognition},
  pages={6848--6856},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.8.2 Topdown Heatmap + Shufflenetv1 on Mpii

```
@inproceedings{zhang2018shufflenet,
  title={Shufflenet: An extremely efficient convolutional neural network for mobile
↪devices},
  author={Zhang, Xiangyu and Zhou, Xinyu and Lin, Mengxiao and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={6848--6856},
  year={2018}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 32.9 Swin (ICCV' 2021)

### 32.9.1 Topdown Heatmap + Swin on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{liu2021swin,
  title={Swin transformer: Hierarchical vision transformer using shifted windows},
  author={Liu, Ze and Lin, Yutong and Cao, Yue and Hu, Han and Wei, Yixuan and Zhang, ↵
↵Zheng and Lin, Stephen and Guo, Baining},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={10012--10022},
  year={2021}
}
```

```
@inproceedings{lin2017feature,
  title={Feature pyramid networks for object detection},
  author={Lin, Tsung-Yi and Doll{'a}r, Piotr and Girshick, Ross and He, Kaiming and ↵
↵Hariharan, Bharath and Belongie, Serge},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={2117--2125},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and ↵
↵Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset



## 32.10 LiteHRNet (CVPR' 2021)

### 32.10.1 Topdown Heatmap + Litehrnet on Coco

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.10.2 Topdown Heatmap + Litehrnet on Mpii

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 32.10.3 Topdown Heatmap + Litehrnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

## 32.11 I3D (CVPR' 2017)

### 32.11.1 Mtut + I3d on Nvgesture

```
@InProceedings{Abavisani_2019_CVPR,
  author = {Abavisani, Mahdi and Joze, Hamid Reza Vaezi and Patel, Vishal M.},
  title = {Improving the Performance of Unimodal Dynamic Hand-Gesture Recognition
↪With Multimodal Training},
  booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
  month = {June},
  year = {2019}
}
```

```
@InProceedings{Carreira_2017_CVPR,
  author = {Carreira, Joao and Zisserman, Andrew},
  title = {Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset},
  booktitle = {Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
  (下页继续)
```

(续上页)

```

month = {July},
year = {2017}
}

```

```

@InProceedings{Molchanov_2016_CVPR,
  author = {Molchanov, Pavlo and Yang, Xiaodong and Gupta, Shalini and Kim, Kihwan
↪and Tyree, Stephen and Kautz, Jan},
  title = {Online Detection and Classification of Dynamic Hand Gestures With
↪Recurrent 3D Convolutional Neural Network},
  booktitle = {Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
  month = {June},
  year = {2016}
}

```

Results on NVGesture test set

\*: MTUT supports multi-modal training and uni-modal testing. Model trained with this config can be used to recognize gestures in rgb videos with [inference config](#).

## 32.12 ShufflenetV2 (ECCV' 2018)

### 32.12.1 Topdown Heatmap + Shufflenetv2 on Coco

```

@inproceedings{ma2018shufflenet,
  title={Shufflenet v2: Practical guidelines for efficient cnn architecture design},
  author={Ma, Ningning and Zhang, Xiangyu and Zheng, Hai-Tao and Sun, Jian},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={116--131},
  year={2018}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.12.2 Topdown Heatmap + Shufflenetv2 on Mpii

```
@inproceedings{ma2018shufflenet,
  title={Shufflenet v2: Practical guidelines for efficient cnn architecture design},
  author={Ma, Ningning and Zhang, Xiangyu and Zheng, Hai-Tao and Sun, Jian},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={116--131},
  year={2018}
}
```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 32.13 TCFormer (CVPR' 2022)

### 32.13.1 Topdown Heatmap + Tcformer on Coco-Wholebody

```
@inproceedings{zeng2022not,
  title={Not All Tokens Are Equal: Human-centric Visual Analysis via Token Clustering
↪Transformer},
  author={Zeng, Wang and Jin, Sheng and Liu, Wentao and Qian, Chen and Luo, Ping and
↪Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={11101--11111},
  year={2022}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  (下页继续)
```

(续上页)

```
booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

## 32.14 ResNetV1D (CVPR' 2019)

### 32.14.1 Topdown Heatmap + Resnetv1d on Coco

```
@inproceedings{he2019bag,
  title={Bag of tricks for image classification with convolutional neural networks},
  author={He, Tong and Zhang, Zhi and Zhang, Hang and Zhang, Zhongyue and Xie, ↵
↵Junyuan and Li, Mu},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern ↵
↵Recognition},
  pages={558--567},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and ↵
↵Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.14.2 Topdown Heatmap + Resnetv1d on Mpii

```
@inproceedings{he2019bag,
  title={Bag of tricks for image classification with convolutional neural networks},
  author={He, Tong and Zhang, Zhi and Zhang, Hang and Zhang, Zhongyue and Xie, ↵
↵Junyuan and Li, Mu},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern ↵
↵Recognition},
```

(下页继续)

(续上页)

```

pages={558--567},
year={2019}
}

```

```

@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

## 32.15 SEResNet (CVPR' 2018)

### 32.15.1 Topdown Heatmap + Seresnet on Coco

```

@inproceedings{hu2018squeeze,
  title={Squeeze-and-excitation networks},
  author={Hu, Jie and Shen, Li and Sun, Gang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={7132--7141},
  year={2018}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that \* means without imagenet pre-training.

## 32.15.2 Topdown Heatmap + Seresnet on Mpii

```
@inproceedings{hu2018squeeze,
  title={Squeeze-and-excitation networks},
  author={Hu, Jie and Shen, Li and Sun, Gang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={7132--7141},
  year={2018}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

Note that \* means without imagenet pre-training.

## 32.16 ViPNAS (CVPR' 2021)

### 32.16.1 Topdown Heatmap + Vipnas on Coco

```
@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  year={2021}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},

```

(下页继续)

(续上页)

```

pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.16.2 Topdown Heatmap + Vipnas on Coco-Wholebody

```

@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  year={2021}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 32.16.3 Topdown Heatmap + Vipnas + Dark on Coco-Wholebody

```

@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  year={2021}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},

```

(下页继续)



(续上页)

```

    author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
    booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
    pages={7093--7102},
    year={2020}
}

```

```

@inproceedings{jin2020whole,
    title={Whole-Body Human Pose Estimation in the Wild},
    author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
    booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
    year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

## 32.17 PVTv2 (CVMJ' 2022)

### 32.17.1 Topdown Heatmap + PVT on Coco

```

@inproceedings{wang2021pyramid,
    title={Pyramid vision transformer: A versatile backbone for dense prediction
↪without convolutions},
    author={Wang, Wenhai and Xie, Enze and Li, Xiang and Fan, Deng-Ping and Song,
↪Kaitao and Liang, Ding and Lu, Tong and Luo, Ping and Shao, Ling},
    booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
    pages={568--578},
    year={2021}
}

```

```

@article{wang2022pvt,
    title={PVT v2: Improved baselines with Pyramid Vision Transformer},
    author={Wang, Wenhai and Xie, Enze and Li, Xiang and Fan, Deng-Ping and Song,
↪Kaitao and Liang, Ding and Lu, Tong and Luo, Ping and Shao, Ling},
    journal={Computational Visual Media},
    pages={1--10},
    year={2022},
    publisher={Springer}
}

```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.18 PVT (ICCV' 2021)

### 32.18.1 Topdown Heatmap + PVT on Coco

```
@inproceedings{wang2021pyramid,
  title={Pyramid vision transformer: A versatile backbone for dense prediction
↪without convolutions},
  author={Wang, Wenhai and Xie, Enze and Li, Xiang and Fan, Deng-Ping and Song,
↪Kaitao and Liang, Ding and Lu, Tong and Luo, Ping and Shao, Ling},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={568--578},
  year={2021}
}
```

```
@article{wang2022pvt,
  title={PVT v2: Improved baselines with Pyramid Vision Transformer},
  author={Wang, Wenhai and Xie, Enze and Li, Xiang and Fan, Deng-Ping and Song,
↪Kaitao and Liang, Ding and Lu, Tong and Luo, Ping and Shao, Ling},
  journal={Computational Visual Media},
  pages={1--10},
  year={2022},
  publisher={Springer}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
```

(下页继续)

(续上页)

```

pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.19 RSN (ECCV' 2020)

### 32.19.1 Topdown Heatmap + RSN on Coco

```

@misc{cai2020learning,
  title={Learning Delicate Local Representations for Multi-Person Pose Estimation},
  author={Yuanhao Cai and Zhicheng Wang and Zhengxiong Luo and Binyi Yin and Angang_
↪Du and Haoqian Wang and Xinyu Zhou and Erjin Zhou and Xiangyu Zhang and Jian Sun},
  year={2020},
  eprint={2003.04030},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.20 CPM (CVPR' 2016)

### 32.20.1 Topdown Heatmap + CPM on Coco

```

@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},

```

(下页继续)

(续上页)

```

    author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
    booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
    pages={4724--4732},
    year={2016}
}

```

```

@inproceedings{lin2014microsoft,
    title={Microsoft coco: Common objects in context},
    author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
    booktitle={European conference on computer vision},
    pages={740--755},
    year={2014},
    organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.20.2 Topdown Heatmap + CPM on JHMDB

```

@inproceedings{wei2016convolutional,
    title={Convolutional pose machines},
    author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
    booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
    pages={4724--4732},
    year={2016}
}

```

```

@inproceedings{Jhuang:ICCV:2013,
    title = {Towards understanding action recognition},
    author = {H. Jhuang and J. Gall and S. Zuffi and C. Schmid and M. J. Black},
    booktitle = {International Conf. on Computer Vision (ICCV)},
    month = Dec,
    pages = {3192-3199},
    year = {2013}
}

```

Results on Sub-JHMDB dataset

The models are pre-trained on MPII dataset only. NO test-time augmentation (multi-scale /rotation testing) is used.

- Normalized by Person Size

- Normalized by Torso Size

### 32.20.3 Topdown Heatmap + CPM on Mpii

```
@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={4724--4732},
  year={2016}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 32.21 HRNet (CVPR' 2019)

### 32.21.1 Topdown Heatmap + Hrnet on Animalpose

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Cao_2019_ICCV,
  author = {Cao, Jinkun and Tang, Hongyang and Fang, Hao-Shu and Shen, Xiaoyong and
↪Lu, Cewu and Tai, Yu-Wing},
  title = {Cross-Domain Adaptation for Animal Pose Estimation},
```

(下页继续)

(续上页)

```

booktitle = {The IEEE International Conference on Computer Vision (ICCV)},
month = {October},
year = {2019}
}

```

Results on AnimalPose validation set (1117 instances)

### 32.21.2 Topdown Heatmap + Hrnet on Ap10k

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@misc{yu2021ap10k,
  title={AP-10K: A Benchmark for Animal Pose Estimation in the Wild},
  author={Hang Yu and Yufei Xu and Jing Zhang and Wei Zhao and Ziyu Guan and_
↪Dacheng Tao},
  year={2021},
  eprint={2108.12617},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}

```

Results on AP-10K validation set

### 32.21.3 Topdown Heatmap + Hrnet on Atrw

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```
@inproceedings{li2020atrw,
  title={ATRW: A Benchmark for Amur Tiger Re-identification in the Wild},
  author={Li, Shuyuan and Li, Jianguo and Tang, Hanlin and Qian, Rui and Lin, Weiyao},
  booktitle={Proceedings of the 28th ACM International Conference on Multimedia},
  pages={2590--2598},
  year={2020}
}
```

Results on ATRW validation set

### 32.21.4 Topdown Heatmap + Hrnet on Horse10

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{mathis2021pretraining,
  title={Pretraining boosts out-of-domain robustness for pose estimation},
  author={Mathis, Alexander and Biasi, Thomas and Schneider, Steffen and Yuksekgonul,
↪Mert and Rogers, Byron and Bethge, Matthias and Mathis, Mackenzie W},
  booktitle={Proceedings of the IEEE/CVF Winter Conference on Applications of
↪Computer Vision},
  pages={1859--1868},
  year={2021}
}
```

Results on Horse-10 test set

### 32.21.5 Topdown Heatmap + Hrnet on Macaque

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
```

(下页继续)

(续上页)

```

    year={2019}
}

```

```

@article{labuguen2020macaquepose,
  title={MacaquePose: A novel 'in the wild' macaque monkey pose dataset for↵
↵markerless motion capture},
  author={Labuguen, Rollyn and Matsumoto, Jumpei and Negrete, Salvador and Nishimaru,↵
↵Hiroshi and Nishijo, Hisao and Takada, Masahiko and Go, Yasuhiro and Inoue, Ken-
↵ichi and Shibata, Tomohiro},
  journal={bioRxiv},
  year={2020},
  publisher={Cold Spring Harbor Laboratory}
}

```

Results on MacaquePose with ground-truth detection bounding boxes

### 32.21.6 Associative Embedding + Hrnet on Aic

```

@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern↵
↵recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↵,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and↵
↵Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and↵
↵others},
  journal={arXiv preprint arXiv:1711.06475},

```

(下页继续)



(续上页)

```

    year={2017}
}

```

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 32.21.7 Topdown Heatmap + Hrnet on Aic

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shiwei and Lin, Guosen and Fu, Yanwei and
↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}

```

Results on AIC val set with ground-truth bounding boxes

### 32.21.8 Associative Embedding + Hrnet on Coco

```

@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}

```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 32.21.9 Associative Embedding + Hrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

### 32.21.10 Dekr + Hrnet on Coco

```
@inproceedings{geng2021bottom,
  title={Bottom-up human pose estimation via disentangled keypoint regression},
  author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,_
↪Jingdong},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={14676--14686},
  year={2021}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

The results of models provided by the authors on COCO val2017 using the same evaluation protocol

The discrepancy between these results and that shown in paper is attributed to the differences in implementation details in evaluation process.

### 32.21.11 Topdown Heatmap + Hrnet + Fp16 on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{micikevicius2017mixed,
  title={Mixed precision training},
  author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
  journal={arXiv preprint arXiv:1710.03740},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
}
```

(下页继续)

(续上页)

```

booktitle={European conference on computer vision},
pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.21.12 Topdown Heatmap + Hrnet on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.21.13 Topdown Heatmap + Hrnet + Augmentation on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```
@article{buslaev2020albumentations,
  title={Albumentations: fast and flexible image augmentations},
  author={Buslaev, Alexander and Iglovikov, Vladimir I and Khvedchenya, Eugene and
↪Parinov, Alex and Druzhinin, Mikhail and Kalinin, Alexandr A},
  journal={Information},
  volume={11},
  number={2},
  pages={125},
  year={2020},
  publisher={Multidisciplinary Digital Publishing Institute}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.21.14 Topdown Heatmap + Hrnet + Udp on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
```

(下页继续)

(续上页)

```

    year = {2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that, UDP also adopts the unbiased encoding/decoding algorithm of [DARK](#).

### 32.21.15 Topdown Heatmap + Hrnet + Dark on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},

```

(下页继续)

(续上页)

```

pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.21.16 Dekr + Hrnet on Crowdpose

```

@inproceedings{geng2021bottom,
  title={Bottom-up human pose estimation via disentangled keypoint regression},
  author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang, Jingdong},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition},
  pages={14676--14686},
  year={2021}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.



### 32.21.17 Topdown Heatmap + Hrnet on Crowdpose

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}
```

Results on CrowdPose test with YOLOv3 human detector

### 32.21.18 Topdown Heatmap + Hrnet on H36m

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu,
↪Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human
↪Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
  publisher = {IEEE Computer Society},
  volume = {36},
  number = {7},
  pages = {1325-1339},
}
```

(下页继续)

(续上页)

```

month = {jul},
year = {2014}
}

```

Results on Human3.6M test set with ground truth 2D detections

### 32.21.19 Associative Embedding + Hrnet on MHP

```

@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}

```

Results on MHP v2.0 validation set without multi-scale test

Results on MHP v2.0 validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 32.21.20 Topdown Heatmap + Hrnet on Mpii

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 32.21.21 Topdown Heatmap + Hrnet + Dark on Mpii

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 32.21.22 Topdown Heatmap + Hrnet on Ochuman

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},
  author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↪and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={889--898},
  year={2019}
}
```

Results on OCHuman test dataset with ground-truth bounding boxes

Following the common setting, the models are trained on COCO train dataset, and evaluate on OCHuman dataset.

### 32.21.23 Topdown Heatmap + Hrnet on Posetrack18

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin,
↪Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition},
  pages={5167--5176},
  year={2018}
}
```

Results on PoseTrack2018 val with ground-truth bounding boxes

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

Results on PoseTrack2018 val with [MMDetection](#) pre-trained [Cascade R-CNN](#) (X-101-64x4d-FPN) human detector

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

### 32.21.24 Posewarper + Hrnet + Posetrack18 on Posetrack18

```
@inproceedings{NIPS2019_gberta,
  title = {Learning Temporal Pose Estimation from Sparsely Labeled Videos},
  author = {Bertasius, Gedas and Feichtenhofer, Christoph, and Tran, Du and Shi, Jianbo,
↪ and Torresani, Lorenzo},
  booktitle = {Advances in Neural Information Processing Systems 33},
  year = {2019},
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
```

(下页继续)

(续上页)

```

    pages={5693--5703},
    year={2019}
}

```

```

@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin, Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={5167--5176},
  year={2018}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Note that the training of PoseWarper can be split into two stages.

The first-stage is trained with the pre-trained [checkpoint](#) on COCO dataset, and the main backbone is fine-tuned on PoseTrack18 in a single-frame setting.

The second-stage is trained with the last [checkpoint](#) from the first stage, and the warping offsets are learned in a multi-frame setting while the backbone is frozen.

Results on PoseTrack2018 val with ground-truth bounding boxes

Results on PoseTrack2018 val with precomputed human bounding boxes from PoseWarper supplementary data files from [this link](#)<sup>1</sup>.

<sup>1</sup> Please download the precomputed human bounding boxes on PoseTrack2018 val from `$PoseWarper_supp_files/posetrack18_precomputed_boxes/val_boxes.json` and place it here: `$mmpose/data/posetrack18/posetrack18_precomputed_boxes/val_boxes.json` to be consistent with the `config`. Please refer to [DATA Preparation](#) for more detail about data preparation.

### 32.21.25 Associative Embedding + Hrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 32.21.26 Topdown Heatmap + Hrnet on Coco-Wholebody

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 32.21.27 Topdown Heatmap + Hrnet + Dark on Coco-Wholebody

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.



## 32.21.28 Topdown Heatmap + Hrnet + Dark on Halpe

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{li2020pastanet,
  title={PaStaNet: Toward Human Activity Knowledge Engine},
  author={Li, Yong-Lu and Xu, Liang and Liu, Xinpeng and Huang, Xijie and Xu, Yue and
↪Wang, Shiyi and Fang, Hao-Shu and Ma, Ze and Chen, Mingyang and Lu, Cewu},
  booktitle={CVPR},
  year={2020}
}
```

Results on Halpe v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on Halpe dataset. We find this will lead to better performance.

## 32.22 HRNetv2 (TPAMI' 2019)

### 32.22.1 Topdown Heatmap + Hrnetv2 on 300w

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  year={2019}
```

(下页继续)

(续上页)

```

journal={TPAMI},
year={2019}
}

```

```

@article{sagonas2016300,
  title={300 faces in-the-wild challenge: Database and results},
  author={Sagonas, Christos and Antonakos, Epameinondas and Tzimiropoulos, Georgios_
↪and Zafeiriou, Stefanos and Pantic, Maja},
  journal={Image and vision computing},
  volume={47},
  pages={3--18},
  year={2016},
  publisher={Elsevier}
}

```

Results on 300W dataset

The model is trained on 300W train.

### 32.22.2 Topdown Heatmap + Hrnetv2 + Dark on Aflw

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database_
↪for facial landmark localization},

```

(下页继续)

(续上页)

```

    author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
    booktitle={2011 IEEE international conference on computer vision workshops (ICCV_
↪workshops)},
    pages={2144--2151},
    year={2011},
    organization={IEEE}
}

```

#### Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

### 32.22.3 Topdown Heatmap + Hrnetv2 on Aflw

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database_
↪for facial landmark localization},
  author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
  booktitle={2011 IEEE international conference on computer vision workshops (ICCV_
↪workshops)},
  pages={2144--2151},
  year={2011},
  organization={IEEE}
}

```

#### Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

### 32.22.4 Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_face

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
    ↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 32.22.5 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
    ↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 32.22.6 Topdown Heatmap + Hrnetv2 on Cofw

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{burgos2013robust,
  title={Robust face landmark estimation under occlusion},
  author={Burgos-Artizzu, Xavier P and Perona, Pietro and Doll{'a}r, Piotr},
  booktitle={Proceedings of the IEEE international conference on computer vision},
  pages={1513--1520},
  year={2013}
}
```

Results on COFW dataset

The model is trained on COFW train.

### 32.22.7 Topdown Heatmap + Hrnetv2 + Awing on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wang2019adaptive,
  title={Adaptive wing loss for robust face alignment via heatmap regression},
  author={Wang, Xinyao and Bo, Liefeng and Fuxin, Li},
  booktitle={Proceedings of the IEEE/CVF international conference on computer vision},
  pages={6971--6981},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 32.22.8 Topdown Heatmap + Hrnetv2 + Dark on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 32.22.9 Topdown Heatmap + Hrnetv2 on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 32.22.10 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↵Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 32.22.11 Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```



```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 32.22.12 Topdown Heatmap + Hrnetv2 + Dark on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 32.22.13 Topdown Heatmap + Hrnetv2 on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 32.22.14 Topdown Heatmap + Hrnetv2 + Udp on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
```

(下页继续)

(续上页)

```

    year = {2020}
}

```

```

@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}

```

Results on OneHand10K val set

### 32.22.15 Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
          Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
          Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},

```

(下页继续)

(续上页)

```

    year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 32.22.16 Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern_
↪Recognition},
  pages={1145--1153},
  year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 32.22.17 Topdown Heatmap + Hrnetv2 on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 32.22.18 Topdown Heatmap + Hrnetv2 + Udp on Rh2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↵Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↵(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 32.22.19 Topdown Heatmap + Hrnetv2 + Dark on Rhd2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

## 32.22.20 Topdown Heatmap + Hrnetv2 on Rhd2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

## 32.23 HRFormer (NIPS' 2021)

### 32.23.1 Topdown Heatmap + Hrformer on Coco

```
@article{yuan2021hrformer,
  title={HRFormer: High-Resolution Vision Transformer for Dense Predict},
  author={Yuan, Yuhui and Fu, Rao and Huang, Lang and Lin, Weihong and Zhang, Chao_
↪and Chen, Xilin and Wang, Jingdong},
  journal={Advances in Neural Information Processing Systems},
  volume={34},
  year={2021}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
}
```

(下页继续)

(续上页)

```

year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.24 AlexNet (NeurIPS' 2012)

### 32.24.1 Topdown Heatmap + Alexnet on Coco

```

@inproceedings{krizhevsky2012imagenet,
  title={Imagenet classification with deep convolutional neural networks},
  author={Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E},
  booktitle={Advances in neural information processing systems},
  pages={1097--1105},
  year={2012}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.25 VGG (ICLR' 2015)

### 32.25.1 Topdown Heatmap + VGG on Coco

```

@article{simonyan2014very,
  title={Very deep convolutional networks for large-scale image recognition},
  author={Simonyan, Karen and Zisserman, Andrew},
  journal={arXiv preprint arXiv:1409.1556},

```

(下页继续)



(续上页)

```

    year={2014}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 32.26 SCNet (CVPR' 2020)

### 32.26.1 Topdown Heatmap + Snet on Coco

```

@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 32.26.2 Topdown Heatmap + Scnet on Mpii

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 32.26.3 Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_face

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 32.26.4 Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set



### 33.1 InterHand2.6M (ECCV' 2020)

#### 33.1.1 Topdown Heatmap + Resnet on Interhand2d

```
@inproceedings{xiao2018simple,  
  title={Simple baselines for human pose estimation and tracking},  
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},  
  booktitle={Proceedings of the European conference on computer vision (ECCV)},  
  pages={466--481},  
  year={2018}  
}
```

```
@inproceedings{he2016deep,  
  title={Deep residual learning for image recognition},  
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
→recognition},  
  pages={770--778},  
  year={2016}  
}
```

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}
```

Results on InterHand2.6M val & test set

### 33.1.2 Internet + Internet on Interhand3d

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}
```

```
@inproceedings{he2016deep,
title={Deep residual learning for image recognition},
author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
pages={770--778},
year={2016}
}
```

```
@InProceedings{Moon_2020_ECCV_InterHand2.6M,
author = {Moon, Gyeongsik and Yu, Shoou-I and Wen, He and Shiratori, Takaaki and Lee,
↪Kyoung Mu},
title = {InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose
↪Estimation from a Single RGB Image},
booktitle = {European Conference on Computer Vision (ECCV)},
year = {2020}
}
```

Results on InterHand2.6M val & test set

## 33.2 Mpii (CVPR' 2014)

### 33.2.1 Deeppose + Resnet on Mpii

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on Mpii val set

### 33.2.2 Deeppose + Resnet + Rle on Mpii

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
```

(下页继续)

(续上页)

```

    year={2014}
}

```

```

@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 33.2.3 Topdown Heatmap + Hrnet on Mpii

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```



```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 33.2.4 Topdown Heatmap + Mobilenetv2 on Mpii

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey ↵
↵and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{andrilluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 33.2.5 Topdown Heatmap + Resnet on Mpii

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
```

(下页继续)

(续上页)

```

    year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 33.2.6 Topdown Heatmap + CPM on Mpii

```

@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={4724--4732},
  year={2016}
}

```

```

@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 33.2.7 Topdown Heatmap + Shufflenetv2 on Mpii

```
@inproceedings{ma2018shufflenet,
  title={Shufflenet v2: Practical guidelines for efficient cnn architecture design},
  author={Ma, Ningning and Zhang, Xiangyu and Zheng, Hai-Tao and Sun, Jian},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={116--131},
  year={2018}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 33.2.8 Topdown Heatmap + Litehrnet on Mpii

```
@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 33.2.9 Topdown Heatmap + Resnext on Mpii

```
@inproceedings{xie2017aggregated,
  title={Aggregated residual transformations for deep neural networks},
  author={Xie, Saining and Girshick, Ross and Doll{\`a}r, Piotr and Tu, Zhuowen and
↪He, Kaiming},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1492--1500},
  year={2017}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 33.2.10 Topdown Heatmap + Shufflenetv1 on Mpii

```
@inproceedings{zhang2018shufflenet,
  title={Shufflenet: An extremely efficient convolutional neural network for mobile
↪devices},
  author={Zhang, Xiangyu and Zhou, Xinyu and Lin, Mengxiao and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={6848--6856},
  year={2018}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 33.2.11 Topdown Heatmap + Hrnet + Dark on Mpii

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 33.2.12 Topdown Heatmap + Scnet on Mpii

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
```

(下页继续)

(续上页)

```

    year={2020}
}

```

```

@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

### 33.2.13 Topdown Heatmap + Seresnet on Mpii

```

@inproceedings{hu2018squeeze,
  title={Squeeze-and-excitation networks},
  author={Hu, Jie and Shen, Li and Sun, Gang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={7132--7141},
  year={2018}
}

```

```

@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele, ↵
↵Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}

```

Results on MPII val set

Note that \* means without imagenet pre-training.

### 33.2.14 Topdown Heatmap + Resnetv1d on Mpii

```
@inproceedings{he2019bag,
  title={Bag of tricks for image classification with convolutional neural networks},
  author={He, Tong and Zhang, Zhi and Zhang, Hang and Zhang, Zhongyue and Xie,
↪Junyuan and Li, Mu},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition},
  pages={558--567},
  year={2019}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

### 33.2.15 Topdown Heatmap + Hourglass on Mpii

```
@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
  month = {June}
}
```

Results on MPII val set

## 33.3 NVGesture (CVPR' 2016)

### 33.3.1 Mtut + l3d on Nvgesture

```
@InProceedings{Abavisani_2019_CVPR,
  author = {Abavisani, Mahdi and Joze, Hamid Reza Vaezi and Patel, Vishal M.},
  title = {Improving the Performance of Unimodal Dynamic Hand-Gesture Recognition_
↵With Multimodal Training},
  booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↵Recognition (CVPR)},
  month = {June},
  year = {2019}
}
```

```
@InProceedings{Carreira_2017_CVPR,
  author = {Carreira, Joao and Zisserman, Andrew},
  title = {Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset},
  booktitle = {Proceedings of the IEEE Conference on Computer Vision and Pattern_
↵Recognition (CVPR)},
  month = {July},
  year = {2017}
}
```

```
@InProceedings{Molchanov_2016_CVPR,
  author = {Molchanov, Pavlo and Yang, Xiaodong and Gupta, Shalini and Kim, Kihwan_
↵and Tyree, Stephen and Kautz, Jan},
  title = {Online Detection and Classification of Dynamic Hand Gestures With_
↵Recurrent 3D Convolutional Neural Network},
  booktitle = {Proceedings of the IEEE Conference on Computer Vision and Pattern_
↵Recognition (CVPR)},
  month = {June},
  year = {2016}
}
```

Results on NVGesture test set

\*: MTUT supports multi-modal training and uni-modal testing. Model trained with this config can be used to recognize gestures in rgb videos with [inference config](#).



## 33.4 MacaquePose (bioRxiv' 2020)

### 33.4.1 Topdown Heatmap + Resnet on Macaque

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@article{labuguen2020macaquepose,
  title={MacaquePose: A novel 'in the wild' macaque monkey pose dataset for_
↪markerless motion capture},
  author={Labuguen, Rollyn and Matsumoto, Jumpei and Negrete, Salvador and Nishimaru,_
↪Hiroshi and Nishijo, Hisao and Takada, Masahiko and Go, Yasuhiro and Inoue, Ken-
↪ichi and Shibata, Tomohiro},
  journal={bioRxiv},
  year={2020},
  publisher={Cold Spring Harbor Laboratory}
}
```

Results on MacaquePose with ground-truth detection bounding boxes

### 33.4.2 Topdown Heatmap + Hrnet on Macaque

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{labuguen2020macaquepose,
  title={MacaquePose: A novel 'in the wild' macaque monkey pose dataset for_
↪markerless motion capture},
  author={Labuguen, Rollyn and Matsumoto, Jumpei and Negrete, Salvador and Nishimaru,_
↪Hiroshi and Nishijo, Hisao and Takada, Masahiko and Go, Yasuhiro and Inoue, Ken-
↪ichi and Shibata, Tomohiro},

```

(下页继续)

(续上页)

```
journal={bioRxiv},
year={2020},
publisher={Cold Spring Harbor Laboratory}
}
```

Results on MacaquePose with ground-truth detection bounding boxes

## 33.5 ATRW (ACM MM' 2020)

### 33.5.1 Topdown Heatmap + Hrnet on Atrw

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
  ↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{li2020atrw,
  title={ATRW: A Benchmark for Amur Tiger Re-identification in the Wild},
  author={Li, Shuyuan and Li, Jianguo and Tang, Hanlin and Qian, Rui and Lin, Weiyao},
  booktitle={Proceedings of the 28th ACM International Conference on Multimedia},
  pages={2590--2598},
  year={2020}
}
```

Results on ATRW validation set

### 33.5.2 Topdown Heatmap + Resnet on Atrw

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{li2020atrw,
  title={ATRW: A Benchmark for Amur Tiger Re-identification in the Wild},
  author={Li, Shuyuan and Li, Jianguo and Tang, Hanlin and Qian, Rui and Lin, Weiyao},
  booktitle={Proceedings of the 28th ACM International Conference on Multimedia},
  pages={2590--2598},
  year={2020}
}
```

Results on ATRW validation set

## 33.6 Animal-Pose (ICCV' 2019)

### 33.6.1 Topdown Heatmap + Hrnet on Animalpose

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Cao_2019_ICCV,
  author = {Cao, Jinkun and Tang, Hongyang and Fang, Hao-Shu and Shen, Xiaoyong and
↪Lu, Cewu and Tai, Yu-Wing},
  title = {Cross-Domain Adaptation for Animal Pose Estimation},
  booktitle = {The IEEE International Conference on Computer Vision (ICCV)},
  month = {October},
  year = {2019}
}
```

Results on AnimalPose validation set (1117 instances)

### 33.6.2 Topdown Heatmap + Resnet on Animalpose

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@InProceedings{Cao_2019_ICCV,
  author = {Cao, Jinkun and Tang, Hongyang and Fang, Hao-Shu and Shen, Xiaoyong and
↪Lu, Cewu and Tai, Yu-Wing},
  title = {Cross-Domain Adaptation for Animal Pose Estimation},
  booktitle = {The IEEE International Conference on Computer Vision (ICCV)},
  month = {October},
  year = {2019}
}
```

Results on AnimalPose validation set (1117 instances)

## 33.7 Human3.6M (TPAMI' 2014)

### 33.7.1 Topdown Heatmap + Hrnet on H36m

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu,
↪Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human
↪Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
  publisher = {IEEE Computer Society},
}
```

(下页继续)

(续上页)

```

    volume = {36},
    number = {7},
    pages = {1325-1339},
    month = {jul},
    year = {2014}
}

```

Results on Human3.6M test set with ground truth 2D detections

### 33.7.2 Pose Lift + Simplebaseline3d on H36m

```

@inproceedings{martinez_2017_3dbaseline,
  title={A simple yet effective baseline for 3d human pose estimation},
  author={Martinez, Julieta and Hossain, Rayat and Romero, Javier and Little, James J.
↵},
  booktitle={ICCV},
  year={2017}
}

```

```

@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, ↵
↵Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human↵
↵Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
  publisher = {IEEE Computer Society},
  volume = {36},
  number = {7},
  pages = {1325-1339},
  month = {jul},
  year = {2014}
}

```

Results on Human3.6M dataset with ground truth 2D detections

1 Differing from the original paper, we didn't apply the max-norm constraint because we found this led to a better convergence and performance.

### 33.7.3 Video Pose Lift + Videopose3d on H36m

```
@inproceedings{pav11o20193d,
  title={3d human pose estimation in video with temporal convolutions and semi-
↪supervised training},
  author={Pav11o, Dario and Feichtenhofer, Christoph and Grangier, David and Auli, ↪
↪Michael},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↪
↪Recognition},
  pages={7753--7762},
  year={2019}
}
```

```
@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu, ↪
↪Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human ↪
↪Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
  publisher = {IEEE Computer Society},
  volume = {36},
  number = {7},
  pages = {1325-1339},
  month = {jul},
  year = {2014}
}
```

Results on Human3.6M dataset with ground truth 2D detections, supervised training

Results on Human3.6M dataset with CPN 2D detections<sup>1</sup>, supervised training

Results on Human3.6M dataset with ground truth 2D detections, semi-supervised training

Results on Human3.6M dataset with CPN 2D detections<sup>1</sup>, semi-supervised training

<sup>1</sup> CPN 2D detections are provided by [official repo](#). The reformatted version used in this repository can be downloaded from [train\\_detection](#) and [test\\_detection](#).

### 33.7.4 HMR + Resnet on Mixed

```
@inProceedings{kanazawaHMR18,
  title={End-to-end Recovery of Human Shape and Pose},
  author = {Angjoo Kanazawa
    and Michael J. Black
    and David W. Jacobs
    and Jitendra Malik},
  booktitle={Computer Vision and Pattern Recognition (CVPR)},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↵recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{h36m_pami,
  author = {Ionescu, Catalin and Papava, Dragos and Olaru, Vlad and Sminchisescu,
↵Cristian},
  title = {Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human
↵Sensing in Natural Environments},
  journal = {IEEE Transactions on Pattern Analysis and Machine Intelligence},
  publisher = {IEEE Computer Society},
  volume = {36},
  number = {7},
  pages = {1325-1339},
  month = {jul},
  year = {2014}
}
```

Results on Human3.6M with ground-truth bounding box having MPJPE-PA of 52.60 mm on Protocol2

## 33.8 OneHand10K (TCSVT' 2019)

### 33.8.1 Deeppose + Resnet on Onehand10k

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set



### 33.8.2 Topdown Heatmap + Hrnetv2 + Dark on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 33.8.3 Topdown Heatmap + Resnet on Onehand10k

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 33.8.4 Topdown Heatmap + Hrnetv2 on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 33.8.5 Topdown Heatmap + Mobilenetv2 on Onehand10k

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 33.8.6 Topdown Heatmap + Hrnetv2 + Udp on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
```

(下页继续)

(续上页)

```

booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition,
↪(CVPR)},
month = {June},
year = {2020}
}

```

```

@article{wang2018mask,
title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
author={Wang, Yangang and Peng, Cong and Liu, Yebin},
journal={IEEE Transactions on Circuits and Systems for Video Technology},
volume={29},
number={11},
pages={3258--3268},
year={2018},
publisher={IEEE}
}

```

Results on OneHand10K val set

## 33.9 JHMDB (ICCV' 2013)

### 33.9.1 Topdown Heatmap + CPM on JHMDB

```

@inproceedings{wei2016convolutional,
title={Convolutional pose machines},
author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
pages={4724--4732},
year={2016}
}

```

```

@inproceedings{Jhuang:ICCV:2013,
title = {Towards understanding action recognition},
author = {H. Jhuang and J. Gall and S. Zuffi and C. Schmid and M. J. Black},
booktitle = {International Conf. on Computer Vision (ICCV)},
month = Dec,
pages = {3192-3199},
year = {2013}
}

```

Results on Sub-JHMDB dataset

The models are pre-trained on MPII dataset only. NO test-time augmentation (multi-scale /rotation testing) is used.

- Normalized by Person Size
- Normalized by Torso Size

### 33.9.2 Topdown Heatmap + Resnet on JHMDB

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{Jhuang:ICCV:2013,
  title = {Towards understanding action recognition},
  author = {H. Jhuang and J. Gall and S. Zuffi and C. Schmid and M. J. Black},
  booktitle = {International Conf. on Computer Vision (ICCV)},
  month = Dec,
  pages = {3192-3199},
  year = {2013}
}
```

Results on Sub-JHMDB dataset

The models are pre-trained on MPII dataset only. *NO* test-time augmentation (multi-scale /rotation testing) is used.

- Normalized by Person Size
- Normalized by Torso Size

## 33.10 CMU Panoptic (ICCV' 2015)

### 33.10.1 Voxelpose + Voxelpose + Prn64x64x64 + Cpn80x80x20 + Panoptic on Panoptic

```
@inproceedings{tumultipose,
  title={VoxelPose: Towards Multi-Camera 3D Human Pose Estimation in Wild Environment},
  ↪,
  author={Tu, Hanyue and Wang, Chunyu and Zeng, Wenjun},
  booktitle={ECCV},
  year={2020}
}
```

```
@Article = {joo_iccv_2015,
author = {Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo_
↪Kanade, Shohei Nobuhara, and Yaser Sheikh},
title = {Panoptic Studio: A Massively Multiview System for Social Motion Capture},
booktitle = {ICCV},
year = {2015}
}
```

Results on CMU Panoptic dataset.

## 33.11 CMU Panoptic HandDB (CVPR' 2017)

### 33.11.1 Deeppose + Resnet on Panoptic2d

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
```

(下页继续)

(续上页)

```

pages={770--778},
year={2016}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={1145--1153},
  year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 33.11.2 Topdown Heatmap + Mobilenetv2 on Panoptic2d

```

@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}

```

```

@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={1145--1153},
  year={2017}
}

```

Results on CMU Panoptic (MPII+NZSL val set)

### 33.11.3 Topdown Heatmap + Resnet on Panoptic2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 33.11.4 Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```



```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 33.11.5 Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 33.11.6 Topdown Heatmap + Hrnetv2 on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

## 33.12 AFLW (ICCVW' 2011)

### 33.12.1 Topdown Heatmap + Hrnetv2 + Dark on Aflw

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
```

(下页继续)

(续上页)

```

    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
    journal={TPAMI},
    year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database
↪for facial landmark localization},
  author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
  booktitle={2011 IEEE international conference on computer vision workshops (ICCV
↪workshops)},
  pages={2144--2151},
  year={2011},
  organization={IEEE}
}

```

### Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

## 33.12.2 Topdown Heatmap + Hrnetv2 on Aflw

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database
↪for facial landmark localization},

```

(下页继续)

(续上页)

```

author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
booktitle={2011 IEEE international conference on computer vision workshops (ICCV_
↪workshops)},
pages={2144--2151},
year={2011},
organization={IEEE}
}

```

Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

## 33.13 300W (IMAVIS' 2016)

### 33.13.1 Topdown Heatmap + Hrnetv2 on 300w

```

@article{WangSCJDZLMTWLX19,
title={Deep High-Resolution Representation Learning for Visual Recognition},
author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
journal={TPAMI},
year={2019}
}

```

```

@article{sagonas2016300,
title={300 faces in-the-wild challenge: Database and results},
author={Sagonas, Christos and Antonakos, Epameinondas and Tzimiropoulos, Georgios_
↪and Zafeiriou, Stefanos and Pantic, Maja},
journal={Image and vision computing},
volume={47},
pages={3--18},
year={2016},
publisher={Elsevier}
}

```

Results on 300W dataset

The model is trained on 300W train.

## 33.14 FreiHand (ICCV' 2019)

### 33.14.1 Topdown Heatmap + Resnet on FreiHand2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{zimmermann2019freihand,
  title={FreiHand: A dataset for markerless capture of hand pose and shape from_
↪single rgb images},
  author={Zimmermann, Christian and Ceylan, Duygu and Yang, Jimei and Russell, Bryan_
↪and Argus, Max and Brox, Thomas},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={813--822},
  year={2019}
}
```

Results on FreiHand val & test set

## 33.15 MHP (ACM MM' 2018)

### 33.15.1 Associative Embedding + Hrnet on MHP

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
```

(下页继续)

(续上页)

```

pages={2277--2287},
year={2017}
}

```

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}

```

Results on MHP v2.0 validation set without multi-scale test

Results on MHP v2.0 validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 33.15.2 Topdown Heatmap + Resnet on MHP

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},

```

(下页继续)

(续上页)

```

pages={770--778},
year={2016}
}

```

```

@inproceedings{zhao2018understanding,
  title={Understanding humans in crowded scenes: Deep nested adversarial learning and
↪a new benchmark for multi-human parsing},
  author={Zhao, Jian and Li, Jianshu and Cheng, Yu and Sim, Terence and Yan,
↪Shuicheng and Feng, Jiashi},
  booktitle={Proceedings of the 26th ACM international conference on Multimedia},
  pages={792--800},
  year={2018}
}

```

Results on MHP v2.0 val set

Note that, the evaluation metric used here is mAP (adapted from COCO), which may be different from the official evaluation codes. Please be cautious if you use the results in papers.

## 33.16 Vinegar Fly (Nature Methods' 2019)

### 33.16.1 Topdown Heatmap + Resnet on Fly

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@article{pereira2019fast,
  title={Fast animal pose estimation using deep neural networks},
  author={Pereira, Talmo D and Aldarondo, Diego E and Willmore, Lindsay and Kislin,
↪Mikhail and Wang, Samuel S-H and Murthy, Mala and Shaevitz, Joshua W},
  journal={Nature methods},
  volume={16},
  number={1},
  pages={117--125},
  year={2019},
  publisher={Nature Publishing Group}
}

```

(下页继续)

(续上页)

}

Results on Vinegar Fly test set

## 33.17 CrowdPose (CVPR' 2019)

### 33.17.1 Associative Embedding + Higherhrnet on Crowdpose

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose↵
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,↵
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern↵
↵Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu↵
↵and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}
```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 2 scales ([2, 1]) are used



### 33.17.2 Dekr + Hrnet on Crowdpose

```
@inproceedings{geng2021bottom,
  title={Bottom-up human pose estimation via disentangled keypoint regression},
  author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,
↪Jingdong},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={14676--14686},
  year={2021}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}
```

Results on CrowdPose test without multi-scale test

Results on CrowdPose test with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

### 33.17.3 Topdown Heatmap + Hrnet on Crowdpose

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
```

(下页继续)

(续上页)

```

year={2019}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

Results on CrowdPose test with YOLOv3 human detector

### 33.17.4 Topdown Heatmap + Resnet on Crowdpose

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@article{li2018crowdpose,
  title={CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark},
  author={Li, Jiefeng and Wang, Can and Zhu, Hao and Mao, Yihuan and Fang, Hao-Shu
↪and Lu, Cewu},
  journal={arXiv preprint arXiv:1812.00324},
  year={2018}
}

```

Results on CrowdPose test with YOLOv3 human detector

## 33.18 Halpe (CVPR' 2020)

### 33.18.1 Topdown Heatmap + Hrnet + Dark on Halpe

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{li2020pastanet,
  title={PaStaNet: Toward Human Activity Knowledge Engine},
  author={Li, Yong-Lu and Xu, Liang and Liu, Xinpeng and Huang, Xijie and Xu, Yue and
↪Wang, Shiyi and Fang, Hao-Shu and Ma, Ze and Chen, Mingyang and Lu, Cewu},
  booktitle={CVPR},
  year={2020}
}
```

Results on Halpe v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on Halpe dataset. We find this will lead to better performance.

## 33.19 OCHuman (CVPR' 2019)

### 33.19.1 Topdown Heatmap + Resnet on Ochuman

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
```

(下页继续)

(续上页)

```

author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
booktitle={Proceedings of the European conference on computer vision (ECCV)},
pages={466--481},
year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},
  author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↪and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={889--898},
  year={2019}
}

```

Results on OCHuman test dataset with ground-truth bounding boxes

Following the common setting, the models are trained on COCO train dataset, and evaluate on OCHuman dataset.

### 33.19.2 Topdown Heatmap + Hrnet on Ochuman

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhang2019pose2seg,
  title={Pose2seg: Detection free human instance segmentation},

```

(下页继续)

(续上页)

```

    author={Zhang, Song-Hai and Li, Ruilong and Dong, Xin and Rosin, Paul and Cai, Zixi
↪and Han, Xi and Yang, Dingcheng and Huang, Haozhi and Hu, Shi-Min},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={889--898},
    year={2019}
}

```

Results on OCHuman test dataset with ground-truth bounding boxes

Following the common setting, the models are trained on COCO train dataset, and evaluate on OCHuman dataset.

## 33.20 AP-10K (NeurIPS' 2021)

### 33.20.1 Topdown Heatmap + Hrnet on Ap10k

```

@inproceedings{sun2019deep,
    title={Deep high-resolution representation learning for human pose estimation},
    author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={5693--5703},
    year={2019}
}

```

```

@misc{yu2021ap10k,
    title={AP-10K: A Benchmark for Animal Pose Estimation in the Wild},
    author={Hang Yu and Yufei Xu and Jing Zhang and Wei Zhao and Ziyu Guan and
↪Dacheng Tao},
    year={2021},
    eprint={2108.12617},
    archivePrefix={arXiv},
    primaryClass={cs.CV}
}

```

Results on AP-10K validation set

### 33.20.2 Topdown Heatmap + Resnet on Ap10k

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@misc{yu2021ap10k,
  title={AP-10K: A Benchmark for Animal Pose Estimation in the Wild},
  author={Hang Yu and Yufei Xu and Jing Zhang and Wei Zhao and Ziyu Guan and
↪Dacheng Tao},
  year={2021},
  eprint={2108.12617},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}
```

Results on AP-10K validation set

## 33.21 Desert Locust (Elife' 2019)

### 33.21.1 Topdown Heatmap + Resnet on Locust

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@article{graving2019deepposekit,
  title={DeepPoseKit, a software toolkit for fast and robust animal pose estimation
↪using deep learning},
  author={Graving, Jacob M and Chae, Daniel and Naik, Hemal and Li, Liang and Koger,
↪Benjamin and Costelloe, Blair R and Couzin, Iain D},
  journal={Elife},
  volume={8},
}
```

(下页继续)

(续上页)

```

pages={e47994},
year={2019},
publisher={eLife Sciences Publications Limited}
}

```

Results on Desert Locust test set

## 33.22 COCO-WholeBody-Face (ECCV' 2020)

### 33.22.1 Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_face

```

@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Face val set

### 33.22.2 Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_face

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},

```

(下页继续)

(续上页)

```

pages={466--481},
year={2018}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={770--778},
  year={2016}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Face val set

### 33.22.3 Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_face

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```



Results on COCO-WholeBody-Face val set

### 33.22.4 Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_face

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 33.22.5 Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_face

```
@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

(下页继续)

(续上页)

```

    year={2020}
}

```

Results on COCO-WholeBody-Face val set

### 33.22.6 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Face val set

## 33.23 MPI-INF-3DHP (3DV' 2017)

### 33.23.1 Pose Lift + Simplebaseline3d on Mpi\_inf\_3dhp

```
@inproceedings{martinez_2017_3dbaseline,
  title={A simple yet effective baseline for 3d human pose estimation},
  author={Martinez, Julieta and Hossain, Rayat and Romero, Javier and Little, James J.},
  ↪},
  booktitle={ICCV},
  year={2017}
}
```

```
@inproceedings{mono-3dhp2017,
  author = {Mehta, Dushyant and Rhodin, Helge and Casas, Dan and Fua, Pascal and ↪
  ↪Sotnychenko, Oleksandr and Xu, Weipeng and Theobalt, Christian},
  title = {Monocular 3D Human Pose Estimation In The Wild Using Improved CNN ↪
  ↪Supervision},
  booktitle = {3D Vision (3DV), 2017 Fifth International Conference on},
  url = {http://gvv.mpi-inf.mpg.de/3dhp_dataset},
  year = {2017},
  organization={IEEE},
  doi={10.1109/3dv.2017.00064},
}
```

Results on MPI-INF-3DHP dataset with ground truth 2D detections

1 Differing from the original paper, we didn't apply the max-norm constraint because we found this led to a better convergence and performance.

### 33.23.2 Video Pose Lift + Videopose3d on Mpi\_inf\_3dhp

```
@inproceedings{pavullo20193d,
  title={3d human pose estimation in video with temporal convolutions and semi- ↪
  ↪supervised training},
  author={Pavullo, Dario and Feichtenhofer, Christoph and Grangier, David and Auli, ↪
  ↪Michael},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern ↪
  ↪Recognition},
  pages={7753--7762},
  year={2019}
}
```

```
@inproceedings{mono-3dhp2017,
  author = {Mehta, Dushyant and Rhodin, Helge and Casas, Dan and Fua, Pascal and
↪Sotnychenko, Oleksandr and Xu, Weipeng and Theobalt, Christian},
  title = {Monocular 3D Human Pose Estimation In The Wild Using Improved CNN
↪Supervision},
  booktitle = {3D Vision (3DV), 2017 Fifth International Conference on},
  url = {http://gvv.mpi-inf.mpg.de/3dhp_dataset},
  year = {2017},
  organization={IEEE},
  doi={10.1109/3dv.2017.00064},
}
```

Results on MPI-INF-3DHP dataset with ground truth 2D detections, supervised training

## 33.24 AI Challenger (ArXiv' 2017)

### 33.24.1 Associative Embedding + Hrnet on Aic

```
@inproceedings{newell12017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shiwei and Lin, Guosen and Fu, Yanwei and
↪others},
}
```

(下页继续)

(续上页)

```

journal={arXiv preprint arXiv:1711.06475},
year={2017}
}

```

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 33.24.2 Associative Embedding + Higherhrnet on Aic

```

@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}

```

```

@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪ Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪ Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪ Recognition},
  pages={5386--5395},
  year={2020}
}

```

```

@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪ ,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and_
↪ Liang, Rui and Wang, Wenjia and Zhou, Shipei and Lin, Guosen and Fu, Yanwei and_
↪ others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}

```

Results on AIC validation set without multi-scale test

Results on AIC validation set with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 33.24.3 Topdown Heatmap + Hrnet on Aic

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding}
↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and
↪Liang, Rui and Wang, Wenjia and Zhou, Shipai and Lin, Guosen and Fu, Yanwei and
↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC val set with ground-truth bounding boxes

### 33.24.4 Topdown Heatmap + Resnet on Aic

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{wu2017ai,
  title={Ai challenger: A large-scale dataset for going deeper in image understanding},
  ↪,
  author={Wu, Jiahong and Zheng, He and Zhao, Bo and Li, Yixin and Yan, Baoming and ↪
  ↪Liang, Rui and Wang, Wenjia and Zhou, Shipai and Lin, Guosen and Fu, Yanwei and ↪
  ↪others},
  journal={arXiv preprint arXiv:1711.06475},
  year={2017}
}
```

Results on AIC val set with ground-truth bounding boxes

## 33.25 Mpii-TRB (ICCV' 2019)

### 33.25.1 Topdown Heatmap + Resnet + Mpii on Mpii\_trb

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↪
  ↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{duan2019trb,
  title={TRB: A Novel Triplet Representation for Understanding 2D Human Body},
  author={Duan, Haodong and Lin, Kwan-Yee and Jin, Sheng and Liu, Wentao and Qian, ↪
  ↪Chen and Ouyang, Wanli},
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},
  pages={9479--9488},
  year={2019}
}
```

Results on MPII-TRB val set

## 33.26 WFLW (CVPR' 2018)

### 33.26.1 Deeppose + Resnet on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.



### 33.26.2 Deeppose + Resnet + Softwingloss on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{lin2021structure,
  title={Structure-Coherent Deep Feature Learning for Robust Face Alignment},
  author={Lin, Chunze and Zhu, Beier and Wang, Quan and Liao, Renjie and Qian, Chen
↪and Lu, Jiwen and Zhou, Jie},
  journal={IEEE Transactions on Image Processing},
  year={2021},
  publisher={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 33.26.3 Deeppose + Resnet + Wingloss on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{feng2018wing,
  title={Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural
↪Networks},
  author={Feng, Zhen-Hua and Kittler, Josef and Awais, Muhammad and Huber, Patrik and
↪Wu, Xiao-Jun},
  booktitle={Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on},
  year={2018},
  pages = {2235-2245},
  organization={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 33.26.4 Topdown Heatmap + Hrnetv2 + Awing on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wang2019adaptive,
  title={Adaptive wing loss for robust face alignment via heatmap regression},
  author={Wang, Xinyao and Bo, Liefeng and Fuxin, Li},
  booktitle={Proceedings of the IEEE/CVF international conference on computer vision},
  pages={6971--6981},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
    Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
    recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 33.26.5 Topdown Heatmap + Hrnetv2 + Dark on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 33.26.6 Topdown Heatmap + Hrnetv2 on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 33.27 PoseTrack18 (CVPR' 2018)

### 33.27.1 Topdown Heatmap + Hrnet on Posetrack18

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin,
↪Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition},
  pages={5167--5176},
  year={2018}
}
```

Results on PoseTrack2018 val with ground-truth bounding boxes

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

Results on PoseTrack2018 val with [MMDetection](#) pre-trained [Cascade R-CNN](#) (X-101-64x4d-FPN) human detector

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

### 33.27.2 Topdown Heatmap + Resnet on Posetrack18

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{andriluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin,
↪Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition},
  pages={5167--5176},
  year={2018}
}
```

Results on PoseTrack2018 val with ground-truth bounding boxes

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

Results on PoseTrack2018 val with [MMDetection](#) pre-trained [Cascade R-CNN](#) (X-101-64x4d-FPN) human detector

The models are first pre-trained on COCO dataset, and then fine-tuned on PoseTrack18.

### 33.27.3 Posewarper + Hrnet + Posetrack18 on Posetrack18

```
@inproceedings{NIPS2019_gberta,
  title = {Learning Temporal Pose Estimation from Sparsely Labeled Videos},
  author = {Bertasius, Gedas and Feichtenhofer, Christoph, and Tran, Du and Shi, Jianbo,
↪ and Torresani, Lorenzo},
  booktitle = {Advances in Neural Information Processing Systems 33},
  year = {2019},
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
}
```

(下页继续)

(续上页)

```
year={2019}
}
```

```
@inproceedings{andrilluka2018posetrack,
  title={Posetrack: A benchmark for human pose estimation and tracking},
  author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin, Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={5167--5176},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Note that the training of PoseWarper can be split into two stages.

The first-stage is trained with the pre-trained [checkpoint](#) on COCO dataset, and the main backbone is fine-tuned on PoseTrack18 in a single-frame setting.

The second-stage is trained with the last [checkpoint](#) from the first stage, and the warping offsets are learned in a multi-frame setting while the backbone is frozen.

Results on PoseTrack2018 val with ground-truth bounding boxes

Results on PoseTrack2018 val with precomputed human bounding boxes from PoseWarper supplementary data files from [this link](#)<sup>1</sup>.

<sup>1</sup> Please download the precomputed human bounding boxes on PoseTrack2018 val from `$PoseWarper_supp_files/posetrack18_precomputed_boxes/val_boxes.json` and place it here: `$mmpose/data/posetrack18/posetrack18_precomputed_boxes/val_boxes.json` to be consistent with the [config](#). Please refer to [DATA Preparation](#) for more detail about data preparation.

## 33.28 COCO-WholeBody-Hand (ECCV' 2020)

### 33.28.1 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 33.28.2 Topdown Heatmap + Mobilenetv2 + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪ and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
```

(下页继续)



(续上页)

```

pages={4510--4520},
year={2018}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Hand val set

### 33.28.3 Topdown Heatmap + Litehrnet + Coco + Wholebody on Coco\_wholebody\_hand

```

@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪ and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody-Hand val set

### 33.28.4 Topdown Heatmap + Hourglass + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{newell12016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
  organization={Springer}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 33.28.5 Topdown Heatmap + Scnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪ Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 33.28.6 Topdown Heatmap + Resnet + Coco + Wholebody on Coco\_wholebody\_hand

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 33.28.7 Topdown Heatmap + Hrnetv2 + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
```

(下页继续)

(续上页)

```
year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

## 33.29 COFW (ICCV' 2013)

### 33.29.1 Topdown Heatmap + Hrnetv2 on Cofw

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{burgos2013robust,
  title={Robust face landmark estimation under occlusion},
  author={Burgos-Artizzu, Xavier P and Perona, Pietro and Doll{'a}r, Piotr},
  booktitle={Proceedings of the IEEE international conference on computer vision},
  pages={1513--1520},
  year={2013}
}
```

Results on COFW dataset

The model is trained on COFW train.

## 33.30 DeepFashion (CVPR' 2016)

### 33.30.1 Deeppose + Resnet on Deepfashion

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{liuLQWtcvpr16DeepFashion,
  author = {Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou},
  title = {DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich_
↪Annotations},
  booktitle = {Proceedings of IEEE Conference on Computer Vision and Pattern_
↪Recognition (CVPR)},
  month = {June},
  year = {2016}
}
```

```
@inproceedings{liuYLWTeccv16FashionLandmark,
  author = {Liu, Ziwei and Yan, Sijie and Luo, Ping and Wang, Xiaogang and Tang, _
↪Xiaoou},
  title = {Fashion Landmark Detection in the Wild},
  booktitle = {European Conference on Computer Vision (ECCV)},
  month = {October},
  year = {2016}
}
```

Results on DeepFashion val set

### 33.30.2 Topdown Heatmap + Resnet on Deepfashion

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{liuLQWtcvpr16DeepFashion,
  author = {Liu, Ziwei and Luo, Ping and Qiu, Shi and Wang, Xiaogang and Tang, Xiaoou},
  title = {DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich
↪Annotations},
  booktitle = {Proceedings of IEEE Conference on Computer Vision and Pattern
↪Recognition (CVPR)},
  month = {June},
  year = {2016}
}
```

```
@inproceedings{liuYLWTeccv16FashionLandmark,
  author = {Liu, Ziwei and Yan, Sijie and Luo, Ping and Wang, Xiaogang and Tang,
↪Xiaoou},
  title = {Fashion Landmark Detection in the Wild},
  booktitle = {European Conference on Computer Vision (ECCV)},
  month = {October},
  year = {2016}
}
```

Results on DeepFashion val set

## 33.31 COCO-WholeBody (ECCV' 2020)

### 33.31.1 Associative Embedding + Hrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 33.31.2 Associative Embedding + Higherhrnet on Coco-Wholebody

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
```

(下页继续)

(续上页)

```

    year={2017}
}

```

```

@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↪ Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↪ Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪ Recognition},
  pages={5386--5395},
  year={2020}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody v1.0 val without multi-scale test

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 33.31.3 Topdown Heatmap + Resnet on Coco-Wholebody

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},

```

(下页继续)



(续上页)

```

year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 33.31.4 Topdown Heatmap + Hrnet on Coco-Wholebody

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 33.31.5 Topdown Heatmap + Tcformer on Coco-Wholebody

```

@inproceedings{zeng2022not,
  title={Not All Tokens Are Equal: Human-centric Visual Analysis via Token Clustering
↪Transformer},
  author={Zeng, Wang and Jin, Sheng and Liu, Wentao and Qian, Chen and Luo, Ping and
↪Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={11101--11111},
  year={2022}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},

```

(下页继续)

(续上页)

```

author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 33.31.6 Topdown Heatmap + Vipnas on Coco-Wholebody

```

@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen.
↪ and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern.
↪ recognition},
  year={2021}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 33.31.7 Topdown Heatmap + Vipnas + Dark on Coco-Wholebody

```

@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen.
↪ and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern.
↪ recognition},
  year={2021}
}

```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 33.31.8 Topdown Heatmap + Hrnet + Dark on Coco-Wholebody

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  year={2020}
}
```

(下页继续)

(续上页)

```
booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
year={2020}
}
```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

## 33.32 Horse-10 (WACV' 2021)

### 33.32.1 Topdown Heatmap + Resnet on Horse10

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{mathis2021pretraining,
  title={Pretraining boosts out-of-domain robustness for pose estimation},
  author={Mathis, Alexander and Biasi, Thomas and Schneider, Steffen and Yuksekgonul, Mert and Rogers, Byron and Bethge, Matthias and Mathis, Mackenzie W},
  booktitle={Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision},
  pages={1859--1868},
  year={2021}
}
```

Results on Horse-10 test set

### 33.32.2 Topdown Heatmap + Hrnet on Horse10

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern recognition},
  pages={5693--5703},
}
```

(下页继续)

(续上页)

```

    year={2019}
}

```

```

@inproceedings{mathis2021pretraining,
  title={Pretraining boosts out-of-domain robustness for pose estimation},
  author={Mathis, Alexander and Biasi, Thomas and Schneider, Steffen and Yuksekgonul, Mert and Rogers, Byron and Bethge, Matthias and Mathis, Mackenzie W},
  booktitle={Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision},
  pages={1859--1868},
  year={2021}
}

```

Results on Horse-10 test set

## 33.33 Grévy' s Zebra (Elife' 2019)

### 33.33.1 Topdown Heatmap + Resnet on Zebra

```

@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}

```

```

@article{graving2019deepposekit,
  title={DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning},
  author={Graving, Jacob M and Chae, Daniel and Naik, Hemal and Li, Liang and Koger, Benjamin and Costelloe, Blair R and Couzin, Iain D},
  journal={Elife},
  volume={8},
  pages={e47994},
  year={2019},
  publisher={eLife Sciences Publications Limited}
}

```

Results on Grévy' s Zebra test set

## 33.34 COCO (ECCV' 2014)

### 33.34.1 Associative Embedding + Higherhrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↵Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↵Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↵(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↵Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

### 33.34.2 Associative Embedding + Mobilenetv2 on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 33.34.3 Associative Embedding + Hrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 33.34.4 Associative Embedding + Higherhrnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```



```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose_
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,_
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↵Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↵Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 33.34.5 Associative Embedding + Resnet on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↵recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 33.34.6 Associative Embedding + Hrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

### 33.34.7 Associative Embedding + Hourglass + Ae on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

### 33.34.8 Deeppose + Resnet on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.9 Deeppose + Resnet + Rle on Coco

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{li2021human,
  title={Human pose regression with residual log-likelihood estimation},
  author={Li, Jiefeng and Bian, Siyuan and Zeng, Ailing and Wang, Can and Pang, Bo
↪and Liu, Wentao and Lu, Cewu},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={11025--11034},
  year={2021}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.10 Dekr + Hrnet on Coco

```
@inproceedings{geng2021bottom,
  title={Bottom-up human pose estimation via disentangled keypoint regression},
  author={Geng, Zigang and Sun, Ke and Xiao, Bin and Zhang, Zhaoxiang and Wang,
↪Jingdong},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={14676--14686},
  year={2021}
}
```

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 without multi-scale test

Results on COCO val2017 with multi-scale test. 3 default scales ([2, 1, 0.5]) are used

\* these configs are generally used for evaluation. The training settings are identical to their single-scale counterparts.

The results of models provided by the authors on COCO val2017 using the same evaluation protocol

The discrepancy between these results and that shown in paper is attributed to the differences in implementation details in evaluation process.

### 33.34.11 Topdown Heatmap + Seresnet on Coco

```
@inproceedings{hu2018squeeze,
  title={Squeeze-and-excitation networks},
  author={Hu, Jie and Shen, Li and Sun, Gang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={7132--7141},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

(下页继续)

(续上页)

```

booktitle={European conference on computer vision},
pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that \* means without imagenet pre-training.

### 33.34.12 Topdown Heatmap + Resnetv1d on Coco

```

@inproceedings{he2019bag,
  title={Bag of tricks for image classification with convolutional neural networks},
  author={He, Tong and Zhang, Zhi and Zhang, Hang and Zhang, Zhongyue and Xie, Junyuan and Li, Mu},
  booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition},
  pages={558--567},
  year={2019}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.13 Topdown Heatmap + Hourglass on Coco

```

@inproceedings{newell2016stacked,
  title={Stacked hourglass networks for human pose estimation},
  author={Newell, Alejandro and Yang, Kaiyu and Deng, Jia},
  booktitle={European conference on computer vision},
  pages={483--499},
  year={2016},
}

```

(下页继续)

(续上页)

```
organization={Springer}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.14 Topdown Heatmap + RSN on Coco

```
@misc{cai2020learning,
  title={Learning Delicate Local Representations for Multi-Person Pose Estimation},
  author={Yuanhao Cai and Zhicheng Wang and Zhengxiong Luo and Binyi Yin and Angang
↪Du and Haoqian Wang and Xinyu Zhou and Erjin Zhou and Xiangyu Zhang and Jian Sun},
  year={2020},
  eprint={2003.04030},
  archivePrefix={arXiv},
  primaryClass={cs.CV}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset



### 33.34.15 Topdown Heatmap + Resnet + Fp16 on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@article{micikevicius2017mixed,
  title={Mixed precision training},
  author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
  journal={arXiv preprint arXiv:1710.03740},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a'}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.16 Topdown Heatmap + Mobilenetv2 on Coco

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.17 Topdown Heatmap + Shufflenetv1 on Coco

```
@inproceedings{zhang2018shufflenet,
  title={Shufflenet: An extremely efficient convolutional neural network for mobile
↪devices},
  author={Zhang, Xiangyu and Zhou, Xinyu and Lin, Mengxiao and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={6848--6856},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

(下页继续)

(续上页)

```

year={2014},
organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.18 Topdown Heatmap + MSPN on Coco

```

@article{li2019rethinking,
  title={Rethinking on Multi-Stage Networks for Human Pose Estimation},
  author={Li, Wenbo and Wang, Zhicheng and Yin, Binyi and Peng, Qixiang and Du,
↪Yuming and Xiao, Tianzi and Yu, Gang and Lu, Hongtao and Wei, Yichen and Sun, Jian},
  journal={arXiv preprint arXiv:1901.00148},
  year={2019}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.19 Topdown Heatmap + Hrnet + Fp16 on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{micikevicius2017mixed,
  title={Mixed precision training},

```

(下页继续)

(续上页)

```

    author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
    ↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
    ↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
    journal={arXiv preprint arXiv:1710.03740},
    year={2017}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
    ↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.20 Topdown Heatmap + Hrnet on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
    ↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
    ↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.21 Topdown Heatmap + Resnext on Coco

```
@inproceedings{xie2017aggregated,
  title={Aggregated residual transformations for deep neural networks},
  author={Xie, Saining and Girshick, Ross and Doll{\`a}r, Piotr and Tu, Zhuowen and
↪He, Kaiming},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={1492--1500},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.22 Topdown Heatmap + Resnet + Dark on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.23 Topdown Heatmap + VGG on Coco

```
@article{simonyan2014very,
  title={Very deep convolutional networks for large-scale image recognition},
  author={Simonyan, Karen and Zisserman, Andrew},
  journal={arXiv preprint arXiv:1409.1556},
  year={2014}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.24 Topdown Heatmap + Shufflenetv2 on Coco

```
@inproceedings{ma2018shufflenet,
  title={Shufflenet v2: Practical guidelines for efficient cnn architecture design},
  author={Ma, Ningning and Zhang, Xiangyu and Zheng, Hai-Tao and Sun, Jian},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={116--131},
  year={2018}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.25 Topdown Heatmap + Hrnet + Augmentation on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{buslaev2020albumentations,
  title={Albumentations: fast and flexible image augmentations},
  author={Buslaev, Alexander and Iglovikov, Vladimir I and Khvedchenya, Eugene and
↪Parinov, Alex and Druzhinin, Mikhail and Kalinin, Alexandr A},
  journal={Information},
  volume={11},
  number={2},
  pages={125},
  year={2020},
}
```

(下页继续)

(续上页)

```
publisher={Multidisciplinary Digital Publishing Institute}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.26 Topdown Heatmap + Swin on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{liu2021swin,
  title={Swin transformer: Hierarchical vision transformer using shifted windows},
  author={Liu, Ze and Lin, Yutong and Cao, Yue and Hu, Han and Wei, Yixuan and Zhang,
↪Zheng and Lin, Stephen and Guo, Baining},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={10012--10022},
  year={2021}
}
```

```
@inproceedings{lin2017feature,
  title={Feature pyramid networks for object detection},
  author={Lin, Tsung-Yi and Doll{'a}r, Piotr and Girshick, Ross and He, Kaiming and
↪Hariharan, Bharath and Belongie, Serge},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2117--2125},
}
```

(下页继续)



(续上页)

```

year={2017}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.27 Topdown Heatmap + Litehrnet on Coco

```

@inproceedings{Yulitehrnet21,
  title={Lite-HRNet: A Lightweight High-Resolution Network},
  author={Yu, Changqian and Xiao, Bin and Gao, Changxin and Yuan, Lu and Zhang, Lei
↪and Sang, Nong and Wang, Jingdong},
  booktitle={CVPR},
  year={2021}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.28 Topdown Heatmap + Snet on Coco

```
@inproceedings{liu2020improving,
  title={Improving Convolutional Networks with Self-Calibrated Convolutions},
  author={Liu, Jiang-Jiang and Hou, Qibin and Cheng, Ming-Ming and Wang, Changhu and
↪Feng, Jiashi},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={10096--10105},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.29 Topdown Heatmap + Hrformer on Coco

```
@article{yuan2021hrformer,
  title={HRFormer: High-Resolution Vision Transformer for Dense Predict},
  author={Yuan, Yuhui and Fu, Rao and Huang, Lang and Lin, Weihong and Zhang, Chao
↪and Chen, Xilin and Wang, Jingdong},
  journal={Advances in Neural Information Processing Systems},
  volume={34},
  year={2021}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
}
```

(下页继续)

(续上页)

```
organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.30 Topdown Heatmap + Resnest on Coco

```
@article{zhang2020resnest,
  title={ResNeSt: Split-Attention Networks},
  author={Zhang, Hang and Wu, Chongruo and Zhang, Zhongyue and Zhu, Yi and Zhang, Zhi-
↪and Lin, Haibin and Sun, Yue and He, Tong and Muller, Jonas and Manmatha, R. and Li,
↪Mu and Smola, Alexander},
  journal={arXiv preprint arXiv:2004.08955},
  year={2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.31 Topdown Heatmap + PVT on Coco

```
@inproceedings{wang2021pyramid,
  title={Pyramid vision transformer: A versatile backbone for dense prediction
↪without convolutions},
  author={Wang, Wenhai and Xie, Enze and Li, Xiang and Fan, Deng-Ping and Song,
↪Kaitao and Liang, Ding and Lu, Tong and Luo, Ping and Shao, Ling},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={568--578},
  year={2021}
}
```

```
@article{wang2022pvt,  
  title={PVT v2: Improved baselines with Pyramid Vision Transformer},  
  author={Wang, Wenhai and Xie, Enze and Li, Xiang and Fan, Deng-Ping and Song,   
↪Kaitao and Liang, Ding and Lu, Tong and Luo, Ping and Shao, Ling},  
  journal={Computational Visual Media},  
  pages={1--10},  
  year={2022},  
  publisher={Springer}  
}
```

```
@inproceedings{lin2014microsoft,  
  title={Microsoft coco: Common objects in context},  
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and   
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},  
  booktitle={European conference on computer vision},  
  pages={740--755},  
  year={2014},  
  organization={Springer}  
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.32 Topdown Heatmap + Resnet on Coco

```
@inproceedings{xiao2018simple,  
  title={Simple baselines for human pose estimation and tracking},  
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},  
  booktitle={Proceedings of the European conference on computer vision (ECCV)},  
  pages={466--481},  
  year={2018}  
}
```

```
@inproceedings{he2016deep,  
  title={Deep residual learning for image recognition},  
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern   
↪recognition},  
  pages={770--778},  
  year={2016}  
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.33 Topdown Heatmap + Alexnet on Coco

```
@inproceedings{krizhevsky2012imagenet,
  title={Imagenet classification with deep convolutional neural networks},
  author={Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E},
  booktitle={Advances in neural information processing systems},
  pages={1097--1105},
  year={2012}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.34 Topdown Heatmap + Hrnet + Udp on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
```

(下页继续)

(续上页)

```

    year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and_
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that, UDP also adopts the unbiased encoding/decoding algorithm of **DARK**.

### 33.34.35 Topdown Heatmap + Hrnet + Dark on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern_
↪Recognition},

```

(下页继续)

(续上页)

```

    pages={7093--7102},
    year={2020}
}

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.36 Topdown Heatmap + Vipnas on Coco

```

@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen
↪and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  year={2021}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.37 Topdown Heatmap + CPM on Coco

```
@inproceedings{wei2016convolutional,
  title={Convolutional pose machines},
  author={Wei, Shih-En and Ramakrishna, Varun and Kanade, Takeo and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↪Recognition},
  pages={4724--4732},
  year={2016}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 33.34.38 Cid + Hrnet on Coco

```
@InProceedings{Wang_2022_CVPR,
  author    = {Wang, Dongkai and Zhang, Shiliang},
  title     = {Contextual Instance Decoupling for Robust Multi-Person Pose
↪Estimation},
  booktitle = {Proceedings of the IEEE/CVF Conference on Computer Vision and
↪Pattern Recognition (CVPR)},
  month     = {June},
  year      = {2022},
  pages     = {11060-11068}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},

```

(下页继续)



(续上页)

```

year={2014},
organization={Springer}
}

```

Results on COCO val2017 without multi-scale test

### 33.34.39 Posewarper + Hrnet + Posetrack18 on Posetrack18

```

@inproceedings{NIPS2019_gberta,
title = {Learning Temporal Pose Estimation from Sparsely Labeled Videos},
author = {Bertasius, Gedas and Feichtenhofer, Christoph, and Tran, Du and Shi, Jianbo,
↪ and Torresani, Lorenzo},
booktitle = {Advances in Neural Information Processing Systems 33},
year = {2019},
}

```

```

@inproceedings{sun2019deep,
title={Deep high-resolution representation learning for human pose estimation},
author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
pages={5693--5703},
year={2019}
}

```

```

@inproceedings{andriluka2018posetrack,
title={Posetrack: A benchmark for human pose estimation and tracking},
author={Andriluka, Mykhaylo and Iqbal, Umar and Insafutdinov, Eldar and Pishchulin,
↪Leonid and Milan, Anton and Gall, Juergen and Schiele, Bernt},
booktitle={Proceedings of the IEEE Conference on Computer Vision and Pattern
↪Recognition},
pages={5167--5176},
year={2018}
}

```

```

@inproceedings{lin2014microsoft,
title={Microsoft coco: Common objects in context},
author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
booktitle={European conference on computer vision},
pages={740--755},
}

```

(下页继续)

(续上页)

```

year={2014},
organization={Springer}
}

```

Note that the training of PoseWarper can be split into two stages.

The first-stage is trained with the pre-trained [checkpoint](#) on COCO dataset, and the main backbone is fine-tuned on PoseTrack18 in a single-frame setting.

The second-stage is trained with the last [checkpoint](#) from the first stage, and the warping offsets are learned in a multi-frame setting while the backbone is frozen.

Results on PoseTrack2018 val with ground-truth bounding boxes

Results on PoseTrack2018 val with precomputed human bounding boxes from PoseWarper supplementary data files from [this link](#)<sup>1</sup>.

1 Please download the precomputed human bounding boxes on PoseTrack2018 val from `$PoseWarper_supp_files/posetrack18_precomputed_boxes/val_boxes.json` and place it here: `$mmpose/data/posetrack18/posetrack18_precomputed_boxes/val_boxes.json` to be consistent with the [config](#). Please refer to [DATA Preparation](#) for more detail about data preparation.

## 33.35 RHD (ICCV' 2017)

### 33.35.1 Deeppose + Resnet on Rhd2d

```

@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={1653--1660},
  year={2014}
}

```

```

@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={770--778},
  year={2016}
}

```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 33.35.2 Topdown Heatmap + Mobilenetv2 on Rhd2d

```
@inproceedings{sandler2018mobilenetv2,
  title={Mobilenetv2: Inverted residuals and linear bottlenecks},
  author={Sandler, Mark and Howard, Andrew and Zhu, Menglong and Zhmoginov, Andrey_
↪and Chen, Liang-Chieh},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={4510--4520},
  year={2018}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 33.35.3 Topdown Heatmap + Resnet on Rhd2d

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
}
```

(下页继续)

(续上页)

```
year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 33.35.4 Topdown Heatmap + Hrnetv2 + Udp on Rhd2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
```

(下页继续)

(续上页)

```

year = {2020}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}

```

Results on RHD test set

### 33.35.5 Topdown Heatmap + Hrnetv2 + Dark on Rhd2d

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}

```

Results on RHD test set

### 33.35.6 Topdown Heatmap + Hrnetv2 on Rhd2d

```
@article{WangSCJDZLMTWLX19,  
  title={Deep High-Resolution Representation Learning for Visual Recognition},  
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and  
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and  
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},  
  journal={TPAMI},  
  year={2019}  
}
```

```
@TechReport{zb2017hand,  
  author={Christian Zimmermann and Thomas Brox},  
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},  
  institution={arXiv:1705.01389},  
  year={2017},  
  note="https://arxiv.org/abs/1705.01389",  
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"  
}
```

Results on RHD test set

## 34.1 SoftWingloss (TIP' 2021)

### 34.1.1 Deeppose + Resnet + Softwingloss on WFLW

```
@inproceedings{toshev2014deeppose,  
  title={Deeppose: Human pose estimation via deep neural networks},  
  author={Toshev, Alexander and Szegedy, Christian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={1653--1660},  
  year={2014}  
}
```

```
@inproceedings{he2016deep,  
  title={Deep residual learning for image recognition},  
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},  
  booktitle={Proceedings of the IEEE conference on computer vision and pattern  
↪recognition},  
  pages={770--778},  
  year={2016}  
}
```

```
@article{lin2021structure,
  title={Structure-Coherent Deep Feature Learning for Robust Face Alignment},
  author={Lin, Chunze and Zhu, Beier and Wang, Quan and Liao, Renjie and Qian, Chen
↪and Lu, Jiwen and Zhou, Jie},
  journal={IEEE Transactions on Image Processing},
  year={2021},
  publisher={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 34.2 FP16 (ArXiv' 2017)

### 34.2.1 Topdown Heatmap + Resnet + Fp16 on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={770--778},
```

(下页继续)



(续上页)

```
year={2016}
}
```

```
@article{micikevicius2017mixed,
  title={Mixed precision training},
  author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
  journal={arXiv preprint arXiv:1710.03740},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 34.2.2 Topdown Heatmap + Hrnet + Fp16 on Coco

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@article{micikevicius2017mixed,
  title={Mixed precision training},
  author={Micikevicius, Paulius and Narang, Sharan and Alben, Jonah and Diamos,
↪Gregory and Elsen, Erich and Garcia, David and Ginsburg, Boris and Houston, Michael
↪and Kuchaiev, Oleksii and Venkatesh, Ganesh and others},
  journal={arXiv preprint arXiv:1710.03740},
}
```

(下页继续)

(续上页)

```

    year={2017}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 34.3 Albumentations (Information' 2020)

### 34.3.1 Topdown Heatmap + Hrnet + Augmentation on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@article{buslaev2020albumentations,
  title={Albumentations: fast and flexible image augmentations},
  author={Buslaev, Alexander and Iglovikov, Vladimir I and Khvedchenya, Eugene and
↪Parinov, Alex and Druzhinin, Mikhail and Kalinin, Alexandr A},
  journal={Information},
  volume={11},
  number={2},
  pages={125},
  year={2020},
  publisher={Multidisciplinary Digital Publishing Institute}
}

```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 34.4 AdaptiveWingloss (ICCV' 2019)

### 34.4.1 Topdown Heatmap + Hrnetv2 + Awing on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{wang2019adaptive,
  title={Adaptive wing loss for robust face alignment via heatmap regression},
  author={Wang, Xinyao and Bo, Liefeng and Fuxin, Li},
  booktitle={Proceedings of the IEEE/CVF international conference on computer vision},
  pages={6971--6981},
  year={2019}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 34.5 FPN (CVPR' 2017)

### 34.5.1 Topdown Heatmap + Swin on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{liu2021swin,
  title={Swin transformer: Hierarchical vision transformer using shifted windows},
  author={Liu, Ze and Lin, Yutong and Cao, Yue and Hu, Han and Wei, Yixuan and Zhang, ↵
↵Zheng and Lin, Stephen and Guo, Baining},
  booktitle={Proceedings of the IEEE/CVF International Conference on Computer Vision},
  pages={10012--10022},
  year={2021}
}
```

```
@inproceedings{lin2017feature,
  title={Feature pyramid networks for object detection},
  author={Lin, Tsung-Yi and Doll{'a}r, Piotr and Girshick, Ross and He, Kaiming and ↵
↵Hariharan, Bharath and Belongie, Serge},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern ↵
↵recognition},
  pages={2117--2125},
  year={2017}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and ↵
↵Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
}
```

(下页继续)

(续上页)

```
organization={Springer}
}
```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 34.6 UDP (CVPR' 2020)

### 34.6.1 Associative Embedding + Higherhrnet + Udp on Coco

```
@inproceedings{newell2017associative,
  title={Associative embedding: End-to-end learning for joint detection and grouping},
  author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
  booktitle={Advances in neural information processing systems},
  pages={2277--2287},
  year={2017}
}
```

```
@inproceedings{cheng2020higherhrnet,
  title={HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose↵
↵Estimation},
  author={Cheng, Bowen and Xiao, Bin and Wang, Jingdong and Shi, Honghui and Huang,↵
↵Thomas S and Zhang, Lei},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern↵
↵Recognition},
  pages={5386--5395},
  year={2020}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for↵
↵Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition↵
↵(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
```

(下页继续)

(续上页)

```

author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
booktitle={European conference on computer vision},
pages={740--755},
year={2014},
organization={Springer}
}

```

Results on COCO val2017 without multi-scale test

### 34.6.2 Associative Embedding + Hrnet + Udp on Coco

```

@inproceedings{newell2017associative,
title={Associative embedding: End-to-end learning for joint detection and grouping},
author={Newell, Alejandro and Huang, Zhiao and Deng, Jia},
booktitle={Advances in neural information processing systems},
pages={2277--2287},
year={2017}
}

```

```

@inproceedings{sun2019deep,
title={Deep high-resolution representation learning for human pose estimation},
author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
pages={5693--5703},
year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
month = {June},
year = {2020}
}

```

```

@inproceedings{lin2014microsoft,
title={Microsoft coco: Common objects in context},

```

(下页继续)

(续上页)

```

    author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
    booktitle={European conference on computer vision},
    pages={740--755},
    year={2014},
    organization={Springer}
}

```

Results on COCO val2017 without multi-scale test

### 34.6.3 Topdown Heatmap + Hrnet + Udp on Coco

```

@inproceedings{sun2019deep,
    title={Deep high-resolution representation learning for human pose estimation},
    author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
    booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
    pages={5693--5703},
    year={2019}
}

```

```

@InProceedings{Huang_2020_CVPR,
    author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
    title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for
↪Human Pose Estimation},
    booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition
↪(CVPR)},
    month = {June},
    year = {2020}
}

```

```

@inproceedings{lin2014microsoft,
    title={Microsoft coco: Common objects in context},
    author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\`a}r, Piotr and Zitnick, C Lawrence},
    booktitle={European conference on computer vision},
    pages={740--755},
    year={2014},
    organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

Note that, UDP also adopts the unbiased encoding/decoding algorithm of [DARK](#).

### 34.6.4 Topdown Heatmap + Hrnetv2 + Udp on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set



### 34.6.5 Topdown Heatmap + Hrnetv2 + Udp on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern_
↪Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 34.6.6 Topdown Heatmap + Hrnetv2 + Udp on Rh2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@InProceedings{Huang_2020_CVPR,
  author = {Huang, Junjie and Zhu, Zheng and Guo, Feng and Huang, Guan},
  title = {The Devil Is in the Details: Delving Into Unbiased Data Processing for_
↪Human Pose Estimation},
  booktitle = {The IEEE/CVF Conference on Computer Vision and Pattern Recognition_
↪(CVPR)},
  month = {June},
  year = {2020}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

## 34.7 Wingloss (CVPR' 2018)

### 34.7.1 Deeppose + Resnet + Wingloss on WFLW

```
@inproceedings{toshev2014deeppose,
  title={Deeppose: Human pose estimation via deep neural networks},
  author={Toshev, Alexander and Szegedy, Christian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={1653--1660},
  year={2014}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪recognition},
  pages={770--778},
  year={2016}
}
```

```
@inproceedings{feng2018wing,
  title={Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural_
↪ Networks},
  author={Feng, Zhen-Hua and Kittler, Josef and Awais, Muhammad and Huber, Patrik and_
↪ Wu, Xiao-Jun},
  booktitle={Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on},
  year={2018},
  pages = {2235-2245},
  organization={IEEE}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and_
↪ Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪ recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

## 34.8 DarkPose (CVPR' 2020)

### 34.8.1 Topdown Heatmap + Resnet + Dark on Coco

```
@inproceedings{xiao2018simple,
  title={Simple baselines for human pose estimation and tracking},
  author={Xiao, Bin and Wu, Haiping and Wei, Yichen},
  booktitle={Proceedings of the European conference on computer vision (ECCV)},
  pages={466--481},
  year={2018}
}
```

```
@inproceedings{he2016deep,
  title={Deep residual learning for image recognition},
  author={He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern_
↪ recognition},
```

(下页继续)

(续上页)

```

pages={770--778},
year={2016}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

## 34.8.2 Topdown Heatmap + Hrnet + Dark on Coco

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},

```

(下页继续)

(续上页)

```

year={2020}
}

```

```

@inproceedings{lin2014microsoft,
  title={Microsoft coco: Common objects in context},
  author={Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and
↪Perona, Pietro and Ramanan, Deva and Doll{\'a}r, Piotr and Zitnick, C Lawrence},
  booktitle={European conference on computer vision},
  pages={740--755},
  year={2014},
  organization={Springer}
}

```

Results on COCO val2017 with detector having human AP of 56.4 on COCO val2017 dataset

### 34.8.3 Topdown Heatmap + Hrnet + Dark on Mpii

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{andriluka14cvpr,
  author = {Mykhaylo Andriluka and Leonid Pishchulin and Peter Gehler and Schiele,
↪Bernt},
  title = {2D Human Pose Estimation: New Benchmark and State of the Art Analysis},
  booktitle = {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2014},
}

```

(下页继续)

(续上页)

```

    month = {June}
}

```

Results on MPII val set

### 34.8.4 Topdown Heatmap + Hrnetv2 + Dark on Aflw

```

@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{koestinger2011annotated,
  title={Annotated facial landmarks in the wild: A large-scale, real-world database
↵for facial landmark localization},
  author={Koestinger, Martin and Wohlhart, Paul and Roth, Peter M and Bischof, Horst},
  booktitle={2011 IEEE international conference on computer vision workshops (ICCV
↵workshops)},
  pages={2144--2151},
  year={2011},
  organization={IEEE}
}

```

Results on AFLW dataset

The model is trained on AFLW train and evaluated on AFLW full and frontal.

### 34.8.5 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_face

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↵ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Face val set

### 34.8.6 Topdown Heatmap + Hrnetv2 + Dark on WFLW

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{wu2018look,
  title={Look at boundary: A boundary-aware face alignment algorithm},
  author={Wu, Wayne and Qian, Chen and Yang, Shuo and Wang, Quan and Cai, Yici and
↪Zhou, Qiang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={2129--2138},
  year={2018}
}
```

Results on WFLW dataset

The model is trained on WFLW train.

### 34.8.7 Topdown Heatmap + Hrnetv2 + Dark + Coco + Wholebody on Coco\_wholebody\_hand

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```



```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}
```

Results on COCO-WholeBody-Hand val set

### 34.8.8 Topdown Heatmap + Hrnetv2 + Dark on Onehand10k

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
        Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
        Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@article{wang2018mask,
  title={Mask-pose cascaded cnn for 2d hand pose estimation from single color image},
  author={Wang, Yangang and Peng, Cong and Liu, Yebin},
  journal={IEEE Transactions on Circuits and Systems for Video Technology},
  volume={29},
  number={11},
  pages={3258--3268},
  year={2018},
  publisher={IEEE}
}
```

Results on OneHand10K val set

### 34.8.9 Topdown Heatmap + Hrnetv2 + Dark on Panoptic2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{simon2017hand,
  title={Hand keypoint detection in single images using multiview bootstrapping},
  author={Simon, Tomas and Joo, Hanbyul and Matthews, Iain and Sheikh, Yaser},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern
↵Recognition},
  pages={1145--1153},
  year={2017}
}
```

Results on CMU Panoptic (MPII+NZSL val set)

### 34.8.10 Topdown Heatmap + Hrnetv2 + Dark on Rhd2d

```
@article{WangSCJDZLMTWLX19,
  title={Deep High-Resolution Representation Learning for Visual Recognition},
  author={Jingdong Wang and Ke Sun and Tianheng Cheng and
    Borui Jiang and Chaorui Deng and Yang Zhao and Dong Liu and Yadong Mu and
    Mingkui Tan and Xinggang Wang and Wenyu Liu and Bin Xiao},
  journal={TPAMI},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@TechReport{zb2017hand,
  author={Christian Zimmermann and Thomas Brox},
  title={Learning to Estimate 3D Hand Pose from Single RGB Images},
  institution={arXiv:1705.01389},
  year={2017},
  note="https://arxiv.org/abs/1705.01389",
  url="https://lmb.informatik.uni-freiburg.de/projects/hand3d/"
}
```

Results on RHD test set

### 34.8.11 Topdown Heatmap + Vipnas + Dark on Coco-Wholebody

```
@article{xu2021vipnas,
  title={ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search},
  author={Xu, Lumin and Guan, Yingda and Jin, Sheng and Liu, Wentao and Qian, Chen
↵and Luo, Ping and Ouyang, Wanli and Wang, Xiaogang},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↵recognition},
  year={2021}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↵Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
```

(下页继续)

(续上页)

```

author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

### 34.8.12 Topdown Heatmap + Hrnet + Dark on Coco-Wholebody

```

@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪ recognition},
  pages={5693--5703},
  year={2019}
}

```

```

@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪ Recognition},
  pages={7093--7102},
  year={2020}
}

```

```

@inproceedings{jin2020whole,
  title={Whole-Body Human Pose Estimation in the Wild},
  author={Jin, Sheng and Xu, Lumin and Xu, Jin and Wang, Can and Liu, Wentao and Qian,
↪ Chen and Ouyang, Wanli and Luo, Ping},
  booktitle={Proceedings of the European Conference on Computer Vision (ECCV)},
  year={2020}
}

```

Results on COCO-WholeBody v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on COCO-WholeBody dataset. We find this will lead to better performance.

### 34.8.13 Topdown Heatmap + Hrnet + Dark on Halpe

```
@inproceedings{sun2019deep,
  title={Deep high-resolution representation learning for human pose estimation},
  author={Sun, Ke and Xiao, Bin and Liu, Dong and Wang, Jingdong},
  booktitle={Proceedings of the IEEE conference on computer vision and pattern
↪recognition},
  pages={5693--5703},
  year={2019}
}
```

```
@inproceedings{zhang2020distribution,
  title={Distribution-aware coordinate representation for human pose estimation},
  author={Zhang, Feng and Zhu, Xiatian and Dai, Hanbin and Ye, Mao and Zhu, Ce},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
↪Recognition},
  pages={7093--7102},
  year={2020}
}
```

```
@inproceedings{li2020pastanet,
  title={PaStaNet: Toward Human Activity Knowledge Engine},
  author={Li, Yong-Lu and Xu, Liang and Liu, Xinpeng and Huang, Xijie and Xu, Yue and
↪Wang, Shiyi and Fang, Hao-Shu and Ma, Ze and Chen, Mingyang and Lu, Cewu},
  booktitle={CVPR},
  year={2020}
}
```

Results on Halpe v1.0 val with detector having human AP of 56.4 on COCO val2017 dataset

Note: + means the model is first pre-trained on original COCO dataset, and then fine-tuned on Halpe dataset. We find this will lead to better performance.



---

## 教程 0: 模型配置文件

---

我们使用 `python` 文件作为配置文件，将模块化设计和继承设计结合到配置系统中，便于进行各种实验。您可以在 `$MMPose/configs` 下找到所有提供的配置。如果要检查配置文件，您可以运行 `python tools/analysis/print_config.py /PATH/TO/CONFIG` 来查看完整的配置。

- 通过脚本参数修改配置
- 配置文件命名约定
  - 配置系统
- 常见问题
  - 在配置中使用中间变量

### 35.1 通过脚本参数修改配置

当使用 “`tools/train.py`” 或 “`tools/test.py`” 提交作业时，您可以指定 `--cfg-options` 来修改配置。

- 更新配置字典链的键值。

可以按照原始配置文件中字典的键的顺序指定配置选项。例如，`--cfg-options model.backbone.norm_eval=False` 将主干网络中的所有 BN 模块更改为 `train` 模式。

- 更新配置列表内部的键值。

一些配置字典在配置文件中会形成一个列表。例如，训练流水线 `data.train.pipeline` 通常是一个列表。例如，`[dict(type='LoadImageFromFile'), dict(type='TopDownRandomFlip',`

`flip_prob=0.5)`, ...]。如果要流水线中的 `'flip_prob=0.5'` 更改为 `'flip_prob=0.0'`, 您可以这样指定 `--cfg-options data.train.pipeline.1.flip_prob=0.0`。

- 更新列表 / 元组的值。

如果要更新的值是列表或元组, 例如, 配置文件通常设置为 `workflow=[('train', 1)]`。如果您想更改这个键, 您可以这样指定 `--cfg-options workflow="[(train, 1), (val, 1)]"`。请注意, 引号”是必要的, 以支持列表 / 元组数据类型, 并且指定值的引号内 **不允许**有空格。

## 35.2 配置文件命名约定

我们按照下面的样式命名配置文件。建议贡献者也遵循同样的风格。

```
configs/{topic}/{task}/{algorithm}/{dataset}/{backbone}_{model_setting}_{dataset}_
→[input_size]_[technique].py
```

{xxx} 是必填字段, [yyy] 是可选字段。

- {topic}: 主题类型, 如 `body`, `face`, `hand`, `animal` 等。
- {task}: 任务类型, `[2d | 3d]_[kpt | mesh]_[sview | mview]_[rgb | rgbd]_[img | vid]`。任务类型从 5 个维度定义: (1) 二维或三维姿态估计; (2) 姿态表示形式: 关键点 (`kpt`)、网格 (`mesh`) 或密集姿态 (`dense`); (3) 单视图 (`sview`) 或多视图 (`mview`); (4) RGB 或 RGBD; 以及 (5) 图像 (`img`) 或视频 (`vid`)。例如, `2d_kpt_sview_rgb_img`, `3d_kpt_sview_rgb_vid`, 等等。
- {algorithm}: 算法类型, 例如, `associative_embedding`, `deeppose` 等。
- {dataset}: 数据集名称, 例如, `coco` 等。
- {backbone}: 主干网络类型, 例如, `res50` (`ResNet-50`) 等。
- [model\_setting]: 对某些模型的特定设置。
- [input\_size]: 模型的输入大小。
- [technique]: 一些特定的技术, 包括损失函数, 数据增强, 训练技巧等, 例如, `wingloss`, `udp`, `fp16` 等。

### 35.2.1 配置系统

- 基于热图的二维自顶向下的人体姿态估计实例

为了帮助用户对完整的配置结构和配置系统中的模块有一个基本的了解, 我们下面对配置文件 `'https://github.com/open-mmlab/mmpose/tree/e1ec589884235bee875c89102170439a991f8450/configs/top\_down/resnet/coco/res50'` 作简要的注释。有关每个模块中每个参数的更详细用法和替代方法, 请参阅 API 文档。



```

# 运行设置
log_level = 'INFO' # 日志记录级别
load_from = None # 从给定路径加载预训练模型
resume_from = None # 从给定路径恢复模型权重文件, 将从保存模型权重文件时的轮次开始继续训练
dist_params = dict(backend='nccl') # 设置分布式训练的参数, 也可以设置端口
workflow = [('train', 1)] # 运行程序的工作流。[('train', 1)] 表示只有一个工作流, 名为
    ↳ 'train' 的工作流执行一次
checkpoint_config = dict( # 设置模型权重文件钩子的配置, 请参阅 https://github.com/open-
    ↳ mmlab/mmcv/blob/master/mmcv/runner/hooks/checkpoint.py 的实现
    interval=10) # 保存模型权重文件的间隔
evaluation = dict( # 训练期间评估的配置
    interval=10, # 执行评估的间隔
    metric='mAP', # 采用的评价指标
    key_indicator='AP') # 将 `AP` 设置为关键指标以保存最佳模型权重文件
# 优化器
optimizer = dict(
    # 用于构建优化器的配置, 支持 (1). PyTorch 中的所有优化器,
    # 其参数也与 PyTorch 中的相同. (2). 自定义的优化器
    # 它们通过 `constructor` 构建, 可参阅 "tutorials/4_new_modules.md"
    # 的实现。
    type='Adam', # 优化器的类型, 可参阅 https://github.com/open-mmlab/mmcv/blob/
    ↳ master/mmcv/runner/optimizer/default\_constructor.py#L13 获取更多细节
    lr=5e-4, # 学习率, 参数的详细用法见 PyTorch 文档
)
optimizer_config = dict(grad_clip=None) # 不限制梯度的范围
# 学习率调整策略
lr_config = dict( # 用于注册 LrUpdater 钩子的学习率调度器的配置
    policy='step', # 调整策略, 还支持 CosineAnnealing, Cyclic, 等等, 请参阅 https://
    ↳ github.com/open-mmlab/mmcv/blob/master/mmcv/runner/hooks/lr\_updater.py#L9 获取支
    持的 LrUpdater 细节
    warmup='linear', # 使用的预热类型, 它可以是 None (不使用预热), 'constant', 'linear'
    ↳ 或者 'exp'.
    warmup_iters=500, # 预热的迭代次数或者轮数
    warmup_ratio=0.001, # 预热开始时使用的学习率, 等于预热比 (warmup_ratio) * 初始学习率
    step=[170, 200]) # 降低学习率的步数
total_epochs = 210 # 训练模型的总轮数
log_config = dict( # 注册日志记录器钩子的配置
    interval=50, # 打印日志的间隔
    hooks=[
        dict(type='TextLoggerHook'), # 用来记录训练过程的日志记录器
        # dict(type='TensorboardLoggerHook') # 也支持 Tensorboard 日志记录器
    ])
channel_cfg = dict(

```

(下页继续)

(续上页)

```

num_output_channels=17, # 关键点头部的输出通道数
dataset_joints=17, # 数据集的关节数
dataset_channel=[ # 数据集支持的通道数
    [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16],
],
inference_channel=[ # 输出通道数
    [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]
])

# 模型设置
model = dict( # 模型的配置
    type='TopDown', # 模型的类型
    pretrained='torchvision://resnet50', # 预训练模型的 url / 网址
    backbone=dict( # 主干网络的字典
        type='ResNet', # 主干网络的名称
        depth=50), # ResNet 模型的深度
    keypoint_head=dict( # 关键点头部的字典
        type='TopdownHeatmapSimpleHead', # 关键点头部的名称
        in_channels=2048, # 关键点头部的输入通道数
        out_channels=channel_cfg['num_output_channels'], # 关键点头部的输出通道数
        loss_keypoint=dict( # 关键点损失函数的字典
            type='JointsMSELoss', # 关键点损失函数的名称
            use_target_weight=True)), # 在损失计算中是否考虑目标权重
    train_cfg=dict(), # 训练超参数的配置
    test_cfg=dict( # 测试超参数的配置
        flip_test=True, # 推断时是否使用翻转测试
        post_process='default', # 使用“默认”(default)后处理方法。
        shift_heatmap=True, # 移动并对齐翻转的热图以获得更高的性能
        modulate_kernel=11)) # 用于调制的高斯核大小。仅用于 "post_process='unbiased'"

data_cfg = dict(
    image_size=[192, 256], # 模型输入分辨率的大小
    heatmap_size=[48, 64], # 输出热图的大小
    num_output_channels=channel_cfg['num_output_channels'], # 输出通道数
    num_joints=channel_cfg['dataset_joints'], # 关节点数量
    dataset_channel=channel_cfg['dataset_channel'], # 数据集支持的通道数
    inference_channel=channel_cfg['inference_channel'], # 输出通道数
    soft_nms=False, # 推理过程中是否执行 soft_nms
    nms_thr=1.0, # 非极大抑制阈值
    oks_thr=0.9, # nms 期间 oks (对象关键点相似性) 得分阈值
    vis_thr=0.2, # 关键点可见性阈值
    use_gt_bbox=False, # 测试时是否使用人工标注的边界框
    det_bbox_thr=0.0, # 检测到的边界框分数的阈值。当 'use_gt_bbox=True' 时使用

```

(下页继续)

(续上页)

```

bbox_file='data/coco/person_detection_results/' # 边界框检测文件的路径
'COCO_val2017_detections_AP_H_56_person.json',
)

train_pipeline = [
    dict(type='LoadImageFromFile'), # 从文件加载图像
    dict(type='TopDownRandomFlip', # 执行随机翻转增强
          flip_prob=0.5), # 执行翻转的概率
    dict(
        type='TopDownHalfBodyTransform', # TopDownHalfBodyTransform 数据增强的配置
        num_joints_half_body=8, # 执行半身变换的阈值
        prob_half_body=0.3), # 执行翻转的概率
    dict(
        type='TopDownGetRandomScaleRotation', #
        TopDownGetRandomScaleRotation 的配置
        rot_factor=40, # 旋转到 `[-2*rot_factor, 2*rot_factor]`.
        scale_factor=0.5), # 缩放到 `[1-scale_factor, 1+scale_factor]`.
    dict(type='TopDownAffine', # 对图像进行仿射变换形成输入
          use_udp=False), # 不使用无偏数据处理
    dict(type='ToTensor'), # 将其他类型转换为张量类型流水线
    dict(
        type='NormalizeTensor', # 标准化输入张量
        mean=[0.485, 0.456, 0.406], # 要标准化的不同通道的平均值
        std=[0.229, 0.224, 0.225]), # 要标准化的不同通道的标准差
    dict(type='TopDownGenerateTarget', # 生成热图目标。支持不同的编码类型
          sigma=2), # 热图高斯的 Sigma
    dict(
        type='Collect', # 收集决定数据中哪些键应该传递到检测器的流水线
        keys=['img', 'target', 'target_weight'], # 输入键
        meta_keys=[ # 输入的元键
            'image_file', 'joints_3d', 'joints_3d_visible', 'center', 'scale',
            'rotation', 'bbox_score', 'flip_pairs'
        ]),
]

val_pipeline = [
    dict(type='LoadImageFromFile'), # 从文件加载图像
    dict(type='TopDownAffine'), # 对图像进行仿射变换形成输入
    dict(type='ToTensor'), # ToTensor 的配置
    dict(
        type='NormalizeTensor',
        mean=[0.485, 0.456, 0.406], # 要标准化的不同通道的平均值
        std=[0.229, 0.224, 0.225]), # 要标准化的不同通道的标准差

```

(下页继续)

```

dict(
    type='Collect', # 收集决定数据中哪些键应该传递到检测器的流水线
    keys=['img'], # 输入键
    meta_keys=[ # 输入的元键
        'image_file', 'center', 'scale', 'rotation', 'bbox_score',
        'flip_pairs'
    ]),
]

test_pipeline = val_pipeline

data_root = 'data/coco' # 数据集的配置
data = dict(
    samples_per_gpu=64, # 训练期间每个 GPU 的 Batch size
    workers_per_gpu=2, # 每个 GPU 预取数据的 worker 个数
    val_dataloader=dict(samples_per_gpu=32), # 验证期间每个 GPU 的 Batch size
    test_dataloader=dict(samples_per_gpu=32), # 测试期间每个 GPU 的 Batch size
    train=dict( # 训练数据集的配置
        type='TopDownCocoDataset', # 数据集的名称
        ann_file=f'{data_root}/annotations/person_keypoints_train2017.json', # 标注
        文件的路径
        img_prefix=f'{data_root}/train2017/',
        data_cfg=data_cfg,
        pipeline=train_pipeline),
    val=dict( # 验证数据集的配置
        type='TopDownCocoDataset', # 数据集的名称
        ann_file=f'{data_root}/annotations/person_keypoints_val2017.json', # 标注
        文件的路径
        img_prefix=f'{data_root}/val2017/',
        data_cfg=data_cfg,
        pipeline=val_pipeline),
    test=dict( # 测试数据集的配置
        type='TopDownCocoDataset', # 数据集的名称
        ann_file=f'{data_root}/annotations/person_keypoints_val2017.json', # 标注
        文件的路径
        img_prefix=f'{data_root}/val2017/',
        data_cfg=data_cfg,
        pipeline=val_pipeline),
)

```

## 35.3 常见问题

### 35.3.1 在配置中使用中间变量

配置文件中使用了一些中间变量，如 `train_pipeline/val_pipeline/test_pipeline` 等。

例如，我们首先要定义 `train_pipeline/val_pipeline/test_pipeline`，然后将它们传递到 `data` 中。因此，`train_pipeline/val_pipeline/test_pipeline` 是中间变量。



---

### 教程 1：如何微调模型

---

在 COCO 数据集上进行预训练，然后在其他数据集（如 COCO-WholeBody 数据集）上进行微调，往往可以提升模型的效果。本教程介绍如何使用模型库中的预训练模型，并在其他数据集上进行微调。

- 概要
- 修改 *Head*
- 修改数据集
- 修改训练策略
- 使用预训练模型

### 36.1 概要

对新数据集上的模型微调需要两个步骤：

1. 支持新数据集。详情参见教程 2：如何增加新数据集
2. 修改配置文件。这部分将在本教程中做具体讨论。

例如，如果想要在自定义数据集上，微调 COCO 预训练的模型，则需要修改配置文件中网络头、数据集、训练策略、预训练模型四个部分。

## 36.2 修改网络头

如果自定义数据集的关键点个数，与 COCO 不同，则需要相应修改 `keypoint_head` 中的 `out_channels` 参数。网络头 (head) 的最后一层的预训练参数不会被载入，而其他层的参数都会被正常载入。例如，COCO-WholeBody 拥有 133 个关键点，因此需要把 17 (COCO 数据集的关键点数目) 改为 133。

```
channel_cfg = dict(
    num_output_channels=133, # 从 17 改为 133
    dataset_joints=133, # 从 17 改为 133
    dataset_channel=[
        list(range(133)), # 从 17 改为 133
    ],
    inference_channel=list(range(133))) # 从 17 改为 133

# model settings
model = dict(
    type='TopDown',
    pretrained='https://download.openmmlab.com/mmpose/'
    'pretrain_models/hrnet_w48-8ef0771d.pth',
    backbone=dict(
        type='HRNet',
        in_channels=3,
        extra=dict(
            stage1=dict(
                num_modules=1,
                num_branches=1,
                block='BOTTLENECK',
                num_blocks=(4, ),
                num_channels=(64, )),
            stage2=dict(
                num_modules=1,
                num_branches=2,
                block='BASIC',
                num_blocks=(4, 4),
                num_channels=(48, 96)),
            stage3=dict(
                num_modules=4,
                num_branches=3,
                block='BASIC',
                num_blocks=(4, 4, 4),
                num_channels=(48, 96, 192)),
            stage4=dict(
                num_modules=3,
                num_branches=4,
```

(下页继续)



(续上页)

```

        block='BASIC',
        num_blocks=(4, 4, 4, 4),
        num_channels=(48, 96, 192, 384))),
    ),
    keypoint_head=dict(
        type='TopdownHeatmapSimpleHead',
        in_channels=48,
        out_channels=channel_cfg['num_output_channels'], # 已对应修改
        num_deconv_layers=0,
        extra=dict(final_conv_kernel=1, ),
        loss_keypoint=dict(type='JointsMSELoss', use_target_weight=True)),
    train_cfg=dict(),
    test_cfg=dict(
        flip_test=True,
        post_process='unbiased',
        shift_heatmap=True,
        modulate_kernel=17))

```

其中, pretrained='https://download.openmmlab.com/mmpose/pretrain\_models/hrnet\_w48-8ef0771d.pth' 表示采用 ImageNet 预训练的权重, 初始化主干网络 (backbone)。不过, pretrained 只会初始化主干网络 (backbone), 而不会初始化网络头 (head)。因此, 我们模型微调时的预训练权重一般通过 load\_from 指定, 而不是使用 pretrained 指定。

## 36.3 支持自己的数据集

MMPose 支持十余种不同的数据集, 包括 COCO, COCO-WholeBody, MPII, MPII-TRB 等数据集。用户可将自定义数据集转换为已有数据集格式, 并修改如下字段。

```

data_root = 'data/coco'
data = dict(
    samples_per_gpu=32,
    workers_per_gpu=2,
    val_dataloader=dict(samples_per_gpu=32),
    test_dataloader=dict(samples_per_gpu=32),
    train=dict(
        type='TopDownCocoWholeBodyDataset', # 对应修改数据集名称
        ann_file=f'{data_root}/annotations/coco_wholebody_train_v1.0.json', # 修改数据集
        标签路径
        img_prefix=f'{data_root}/train2017/',
        data_cfg=data_cfg,
        pipeline=train_pipeline),
    val=dict(

```

(下页继续)

(续上页)

```

type='TopDownCocoWholeBodyDataset', # 对应修改数据集名称
ann_file=f'{data_root}/annotations/coco_wholebody_val_v1.0.json', # 修改数据集标
签路径
img_prefix=f'{data_root}/val2017/',
data_cfg=data_cfg,
pipeline=val_pipeline),
test=dict(
type='TopDownCocoWholeBodyDataset', # 对应修改数据集名称
ann_file=f'{data_root}/annotations/coco_wholebody_val_v1.0.json', # 修改数据集标
签路径
img_prefix=f'{data_root}/val2017/',
data_cfg=data_cfg,
pipeline=val_pipeline)
)

```

## 36.4 修改训练策略

通常情况下，微调模型时设置较小的学习率和训练轮数，即可取得较好效果。

```

# 优化器
optimizer = dict(
    type='Adam',
    lr=5e-4, # 可以适当减小
)
optimizer_config = dict(grad_clip=None)
# 学习策略
lr_config = dict(
    policy='step',
    warmup='linear',
    warmup_iters=500,
    warmup_ratio=0.001,
    step=[170, 200]) # 可以适当减小
total_epochs = 210 # 可以适当减小

```

## 36.5 使用预训练模型

网络设置中的 `pretrained`, 仅会在主干网络模型上加载预训练参数。若要载入整个网络的预训练参数, 需要通过 `load_from` 指定模型文件路径或模型链接。

```
# 将预训练模型用于整个 HRNet 网络
load_from = 'https://download.openmmlab.com/mmpose/top_down/hrnet/hrnet_w48_coco_
→384x288_dark-741844ba_20200812.pth' # 模型路径可以在 model zoo 中找到
```



## 教程 2: 增加新的数据集

## 37.1 将数据集转化为 COCO 格式

我们首先需要将自定义数据集，转换为 COCO 数据集格式。

COCO 数据集格式的 json 标注文件有以下关键字：

```
'images': [  
  {  
    'file_name': '000000001268.jpg',  
    'height': 427,  
    'width': 640,  
    'id': 1268  
  },  
  ...  
],  
'annotations': [  
  {  
    'segmentation': [[426.36,  
      ...  
      424.34,  
      223.3]],  
    'keypoints': [0,0,0,  
      0,0,0,  
      0,0,0,
```

(下页继续)

(续上页)

```

        427, 220, 2,
        443, 222, 2,
        414, 228, 2,
        449, 232, 2,
        408, 248, 1,
        454, 261, 2,
        0, 0, 0,
        0, 0, 0,
        411, 287, 2,
        431, 287, 2,
        0, 0, 0,
        458, 265, 2,
        0, 0, 0,
        466, 300, 1],
    'num_keypoints': 10,
    'area': 3894.5826,
    'iscrowd': 0,
    'image_id': 1268,
    'bbox': [402.34, 205.02, 65.26, 88.45],
    'category_id': 1,
    'id': 215218
},
...
],
'categories': [
    {'id': 1, 'name': 'person'},
]

```

Json 文件中必须包含以下三个关键字：

- images：包含图片信息的列表，提供图片的 file\_name, height, width 和 id 等信息。
- annotations：包含实例标注的列表。
- categories：包含类别名称（'person'）和对应的 ID (1)。

## 37.2 为自定义数据集创建 dataset\_info 数据集配置文件

在如下位置，添加一个数据集配置文件。

```
configs/_base_/datasets/custom.py
```

数据集配置文件的样例如下：

keypoint\_info 包含每个关键点的信息，其中：

1. name: 代表关键点的名称。一个数据集的每个关键点，名称必须唯一。
2. id: 关键点的标识号。
3. color: ([B, G, R]) 用于可视化关键点。
4. type: 分为 ‘upper’ 和 ‘lower’ 两种，用于数据增强。
5. swap: 表示与当前关键点，“镜像对称”的关键点名称。

skeleton\_info 包含关键点之间的连接关系，主要用于可视化。

joint\_weights 可以为不同的关键点设置不同的损失权重，用于训练。

sigmas 用于计算 OKS 得分，具体内容请参考 [keypoints-eval](#)。

```
dataset_info = dict(
    dataset_name='coco',
    paper_info=dict(
        author='Lin, Tsung-Yi and Maire, Michael and '
        'Belongie, Serge and Hays, James and '
        'Perona, Pietro and Ramanan, Deva and '
        r'Doll{\'a}r, Piotr and Zitnick, C Lawrence',
        title='Microsoft coco: Common objects in context',
        container='European conference on computer vision',
        year='2014',
        homepage='http://cocodataset.org/',
    ),
    keypoint_info={
        0:
            dict(name='nose', id=0, color=[51, 153, 255], type='upper', swap=''),
        1:
            dict(
                name='left_eye',
                id=1,
                color=[51, 153, 255],
                type='upper',
                swap='right_eye'),
        2:
            dict(
                name='right_eye',
                id=2,
                color=[51, 153, 255],
                type='upper',
                swap='left_eye'),
        3:
            dict(
                name='left_ear',
```

(下页继续)

(续上页)

```
        id=3,
        color=[51, 153, 255],
        type='upper',
        swap='right_ear'),
4:
dict(
    name='right_ear',
    id=4,
    color=[51, 153, 255],
    type='upper',
    swap='left_ear'),
5:
dict(
    name='left_shoulder',
    id=5,
    color=[0, 255, 0],
    type='upper',
    swap='right_shoulder'),
6:
dict(
    name='right_shoulder',
    id=6,
    color=[255, 128, 0],
    type='upper',
    swap='left_shoulder'),
7:
dict(
    name='left_elbow',
    id=7,
    color=[0, 255, 0],
    type='upper',
    swap='right_elbow'),
8:
dict(
    name='right_elbow',
    id=8,
    color=[255, 128, 0],
    type='upper',
    swap='left_elbow'),
9:
dict(
    name='left_wrist',
    id=9,
```

(下页继续)



(续上页)

```
        color=[0, 255, 0],
        type='upper',
        swap='right_wrist'),
10:
dict(
    name='right_wrist',
    id=10,
    color=[255, 128, 0],
    type='upper',
    swap='left_wrist'),
11:
dict(
    name='left_hip',
    id=11,
    color=[0, 255, 0],
    type='lower',
    swap='right_hip'),
12:
dict(
    name='right_hip',
    id=12,
    color=[255, 128, 0],
    type='lower',
    swap='left_hip'),
13:
dict(
    name='left_knee',
    id=13,
    color=[0, 255, 0],
    type='lower',
    swap='right_knee'),
14:
dict(
    name='right_knee',
    id=14,
    color=[255, 128, 0],
    type='lower',
    swap='left_knee'),
15:
dict(
    name='left_ankle',
    id=15,
    color=[0, 255, 0],
```

(下页继续)

(续上页)

```

        type='lower',
        swap='right_ankle'),
16:
dict(
    name='right_ankle',
    id=16,
    color=[255, 128, 0],
    type='lower',
    swap='left_ankle')
},
skeleton_info={
    0:
    dict(link=('left_ankle', 'left_knee'), id=0, color=[0, 255, 0]),
    1:
    dict(link=('left_knee', 'left_hip'), id=1, color=[0, 255, 0]),
    2:
    dict(link=('right_ankle', 'right_knee'), id=2, color=[255, 128, 0]),
    3:
    dict(link=('right_knee', 'right_hip'), id=3, color=[255, 128, 0]),
    4:
    dict(link=('left_hip', 'right_hip'), id=4, color=[51, 153, 255]),
    5:
    dict(link=('left_shoulder', 'left_hip'), id=5, color=[51, 153, 255]),
    6:
    dict(link=('right_shoulder', 'right_hip'), id=6, color=[51, 153, 255]),
    7:
    dict(
        link=('left_shoulder', 'right_shoulder'),
        id=7,
        color=[51, 153, 255]),
    8:
    dict(link=('left_shoulder', 'left_elbow'), id=8, color=[0, 255, 0]),
    9:
    dict(
        link=('right_shoulder', 'right_elbow'), id=9, color=[255, 128, 0]),
    10:
    dict(link=('left_elbow', 'left_wrist'), id=10, color=[0, 255, 0]),
    11:
    dict(link=('right_elbow', 'right_wrist'), id=11, color=[255, 128, 0]),
    12:
    dict(link=('left_eye', 'right_eye'), id=12, color=[51, 153, 255]),
    13:
    dict(link=('nose', 'left_eye'), id=13, color=[51, 153, 255]),

```

(下页继续)

(续上页)

```

14:
dict(link=('nose', 'right_eye'), id=14, color=[51, 153, 255]),
15:
dict(link=('left_eye', 'left_ear'), id=15, color=[51, 153, 255]),
16:
dict(link=('right_eye', 'right_ear'), id=16, color=[51, 153, 255]),
17:
dict(link=('left_ear', 'left_shoulder'), id=17, color=[51, 153, 255]),
18:
dict(
    link=('right_ear', 'right_shoulder'), id=18, color=[51, 153, 255])
},
joint_weights=[
    1., 1., 1., 1., 1., 1., 1., 1.2, 1.2, 1.5, 1.5, 1., 1., 1.2, 1.2, 1.5,
    1.5
],
sigmas=[
    0.026, 0.025, 0.025, 0.035, 0.035, 0.079, 0.079, 0.072, 0.072, 0.062,
    0.062, 0.107, 0.107, 0.087, 0.087, 0.089, 0.089
]
)

```

## 37.3 创建自定义数据集类

1. 首先在 `mmpose/datasets/datasets` 文件夹创建一个包，比如命名为 `custom`。
2. 定义数据集类，并且注册这个类。

```

@DATASETS.register_module(name='MyCustomDataset')
class MyCustomDataset(SomeOtherBaseClassAsPerYourNeed):

```

3. 为你的自定义类别创建 `mmpose/datasets/datasets/custom/__init__.py`
4. 更新 `mmpose/datasets/__init__.py`

## 37.4 创建和修改训练配置文件

创建和修改训练配置文件，来使用你的自定义数据集。

在 `configs/my_custom_config.py` 中，修改如下几行。

```

...
# dataset settings

```

(下页继续)

(续上页)

```
dataset_type = 'MyCustomDataset'
...
data = dict(
    samples_per_gpu=2,
    workers_per_gpu=2,
    train=dict(
        type=dataset_type,
        ann_file='path/to/your/train/json',
        img_prefix='path/to/your/train/img',
        ...),
    val=dict(
        type=dataset_type,
        ann_file='path/to/your/val/json',
        img_prefix='path/to/your/val/img',
        ...),
    test=dict(
        type=dataset_type,
        ann_file='path/to/your/test/json',
        img_prefix='path/to/your/test/img',
        ...))
...
```

---

## 教程 3: 自定义数据前处理流水线

---

### 38.1 设计数据前处理流水线

参照惯例，MMPose 使用 Dataset 和 DataLoader 实现多进程数据加载。Dataset 返回一个字典，作为模型的输入。由于姿态估计任务的数据大小不一定相同（图片大小，边界框大小等），MMPose 使用 MMCV 中的 DataContainer 收集和分配不同大小的数据。详情可见[此处](#)。

数据前处理流水线和数据集是相互独立的。通常，数据集定义如何处理标注文件，而数据前处理流水线将原始数据处理成网络输入。数据前处理流水线包含一系列操作。每个操作都输入一个字典 (dict)，新增/更新/删除相关字段，最终输出更新后的字典作为下一个操作的输入。

数据前处理流水线的操作可以被分类为数据加载、预处理、格式化和生成监督等（后文将详细介绍）。

这里以 Simple Baseline (ResNet50) 的数据前处理流水线为例：

```
train_pipeline = [  
    dict(type='LoadImageFromFile'),  
    dict(type='TopDownRandomFlip', flip_prob=0.5),  
    dict(type='TopDownHalfBodyTransform', num_joints_half_body=8, prob_half_body=0.3),  
    dict(type='TopDownGetRandomScaleRotation', rot_factor=40, scale_factor=0.5),  
    dict(type='TopDownAffine'),  
    dict(type='ToTensor'),  
    dict(  
        type='NormalizeTensor',  
        mean=[0.485, 0.456, 0.406],  
        std=[0.229, 0.224, 0.225]),  
]
```

(下页继续)

(续上页)

```

dict(type='TopDownGenerateTarget', sigma=2),
dict(
    type='Collect',
    keys=['img', 'target', 'target_weight'],
    meta_keys=[
        'image_file', 'joints_3d', 'joints_3d_visible', 'center', 'scale',
        'rotation', 'bbox_score', 'flip_pairs'
    ]),
]

val_pipeline = [
    dict(type='LoadImageFromFile'),
    dict(type='TopDownAffine'),
    dict(type='ToTensor'),
    dict(
        type='NormalizeTensor',
        mean=[0.485, 0.456, 0.406],
        std=[0.229, 0.224, 0.225]),
    dict(
        type='Collect',
        keys=['img'],
        meta_keys=[
            'image_file', 'center', 'scale', 'rotation', 'bbox_score',
            'flip_pairs'
        ]),
]

```

下面列出每个操作新增/更新/删除的相关字典字段。

### 38.1.1 数据加载

LoadImageFromFile

- 新增: img, img\_file

### 38.1.2 预处理

TopDownRandomFlip

- 更新: img, joints\_3d, joints\_3d\_visible, center

TopDownHalfBodyTransform

- 更新: center, scale

TopDownGetRandomScaleRotation

- 更新: scale, rotation

TopDownAffine

- 更新: img, joints\_3d, joints\_3d\_visible

NormalizeTensor

- 更新: img

### 38.1.3 生成监督

TopDownGenerateTarget

- 新增: target, target\_weight

### 38.1.4 格式化

ToTensor

- 更新: 'img'

Collect

- 新增: img\_meta (其包含的字段由 meta\_keys 指定)
- 删除: 除了 keys 指定以外的所有字段

## 38.2 扩展和使用自定义流水线

1. 将一个新的处理流水线操作写入任一文件中，例如 my\_pipeline.py。它以一个字典作为输入，并返回一个更新后的字典。

```
from mmpose.datasets import PIPELINES

@PIPELINES.register_module()
class MyTransform:

    def __call__(self, results):
        results['dummy'] = True
        return results
```

2. 导入定义好的新类。

```
from .my_pipeline import MyTransform
```

3. 在配置文件中使用它。

```
train_pipeline = [
    dict(type='LoadImageFromFile'),
    dict(type='TopDownRandomFlip', flip_prob=0.5),
    dict(type='TopDownHalfBodyTransform', num_joints_half_body=8, prob_half_body=0.3),
    dict(type='TopDownGetRandomScaleRotation', rot_factor=40, scale_factor=0.5),
    dict(type='TopDownAffine'),
    dict(type='MyTransform'),
    dict(type='ToTensor'),
    dict(
        type='NormalizeTensor',
        mean=[0.485, 0.456, 0.406],
        std=[0.229, 0.224, 0.225]),
    dict(type='TopDownGenerateTarget', sigma=2),
    dict(
        type='Collect',
        keys=['img', 'target', 'target_weight'],
        meta_keys=[
            'image_file', 'joints_3d', 'joints_3d_visible', 'center', 'scale',
            'rotation', 'bbox_score', 'flip_pairs'
        ]),
]
```



## 教程 4: 增加新的模块

### 39.1 自定义优化器

在本教程中，我们将介绍如何为项目定制优化器。假设想要添加一个名为 `MyOptimizer` 的优化器，它有 `a`, `b` 和 `c` 三个参数。那么首先需要在文件中实现该优化器，例如 `mmpose/core/optimizer/my_optimizer.py`:

```
from mmcv.runner import OPTIMIZERS
from torch.optim import Optimizer

@OPTIMIZERS.register_module()
class MyOptimizer(Optimizer):

    def __init__(self, a, b, c)
```

然后需要将其添加到 `mmpose/core/optimizer/__init__.py` 中，从而让注册器可以找到这个新的优化器并添加它：

```
from .my_optimizer import MyOptimizer
```

之后，可以在配置文件的 `optimizer` 字段中使用 `MyOptimizer`。在配置中，优化器由 `optimizer` 字段所定义，如下所示：

```
optimizer = dict(type='SGD', lr=0.02, momentum=0.9, weight_decay=0.0001)
```

若要使用自己新定义的优化器，可以将字段修改为：

```
optimizer = dict(type='MyOptimizer', a=a_value, b=b_value, c=c_value)
```

我们已经支持使用 PyTorch 实现的所有优化器，只需要更改配置文件的 optimizer 字段。例如：若用户想要使用 ADAM 优化器，只需要做出如下修改，虽然这会造成网络效果下降。

```
optimizer = dict(type='Adam', lr=0.0003, weight_decay=0.0001)
```

用户可以直接根据 PyTorch API 文档 对参数进行设置。

## 39.2 自定义优化器构造器

某些模型可能对不同层的参数有特定的优化设置，例如 BatchNorm 层的权值衰减。用户可以通过自定义优化器构造函数来进行这些细粒度的参数调整。

```
from mmcv.utils import build_from_cfg

from mmcv.runner import OPTIMIZER_BUILDERS, OPTIMIZERS
from mmpose.utils import get_root_logger
from .cocktail_optimizer import CocktailOptimizer

@OPTIMIZER_BUILDERS.register_module()
class CocktailOptimizerConstructor:

    def __init__(self, optimizer_cfg, paramwise_cfg=None):

    def __call__(self, model):

        return my_optimizer
```

## 39.3 开发新组件

MMPose 将模型组件分为 3 种基础模型：

- 检测器 (detector)：整个检测器模型流水线，通常包含一个主干网络 (backbone) 和关键点头 (keypoint\_head)。
- 主干网络 (backbone)：通常为一个用于提取特征的 FCN 网络，例如 ResNet，HRNet。
- 关键点头 (keypoint\_head)：用于姿势估计的组件，通常包括一系列反卷积层。

1. 创建一个新文件 `mmpose/models/backbones/my_model.py`。

```
import torch.nn as nn

from ..builder import BACKBONES

@BACKBONES.register_module()
class MyModel(nn.Module):

    def __init__(self, arg1, arg2):
        pass

    def forward(self, x): # should return a tuple
        pass

    def init_weights(self, pretrained=None):
        pass
```

2. 在 `mmpose/models/backbones/__init__.py` 中导入新的主干网络。

```
from .my_model import MyModel
```

3. 创建一个新文件 `mmpose/models/keypoint_heads/my_head.py`。

用户可以通过继承 `nn.Module` 编写一个新的关键点头，并重写 `init_weights(self)` 和 `forward(self, x)` 方法。

```
from ..builder import HEADS

@HEADS.register_module()
class MyHead(nn.Module):

    def __init__(self, arg1, arg2):
        pass
```

(下页继续)

(续上页)

```
def forward(self, x):
    pass

def init_weights(self):
    pass
```

4. 在 `mmpose/models/keypoint_heads/__init__.py` 中导入新的关键点头

```
from .my_head import MyHead
```

5. 在配置文件中使用它。

对于自顶向下的 2D 姿态估计模型，我们将模型类型设置为 `TopDown`。

```
model = dict(
    type='TopDown',
    backbone=dict(
        type='MyModel',
        arg1=xxx,
        arg2=xxx),
    keypoint_head=dict(
        type='MyHead',
        arg1=xxx,
        arg2=xxx))
```

### 39.3.1 添加新的损失函数

假设用户想要为关键点估计添加一个名为 `MyLoss` 的新损失函数。为了添加一个新的损失函数，用户需要在 `mmpose/models/losses/my_loss.py` 下实现该函数。其中，装饰器 `weighted_loss` 使损失函数能够为每个元素加权。

```
import torch
import torch.nn as nn

from mmpose.models import LOSSES

def my_loss(pred, target):
    assert pred.size() == target.size() and target.numel() > 0
    loss = torch.abs(pred - target)
    loss = torch.mean(loss)
    return loss

@LOSSES.register_module()
```

(下页继续)

(续上页)

```

class MyLoss(nn.Module):

    def __init__(self, use_target_weight=False):
        super(MyLoss, self).__init__()
        self.criterion = my_loss()
        self.use_target_weight = use_target_weight

    def forward(self, output, target, target_weight):
        batch_size = output.size(0)
        num_joints = output.size(1)

        heatmaps_pred = output.reshape(
            (batch_size, num_joints, -1)).split(1, 1)
        heatmaps_gt = target.reshape((batch_size, num_joints, -1)).split(1, 1)

        loss = 0.

        for idx in range(num_joints):
            heatmap_pred = heatmaps_pred[idx].squeeze(1)
            heatmap_gt = heatmaps_gt[idx].squeeze(1)
            if self.use_target_weight:
                loss += self.criterion(
                    heatmap_pred * target_weight[:, idx],
                    heatmap_gt * target_weight[:, idx])
            else:
                loss += self.criterion(heatmap_pred, heatmap_gt)

        return loss / num_joints

```

之后，用户需要把它添加进 `mmpose/models/losses/__init__.py`。

```

from .my_loss import MyLoss, my_loss

```

若要使用新的损失函数，可以修改模型中的 `loss_keypoint` 字段。

```

loss_keypoint=dict(type='MyLoss', use_target_weight=False)

```



---

### 教程 5：如何导出模型为 onnx 格式

---

开放式神经网络交换格式（Open Neural Network Exchange，即 **ONNX**）是各种框架共用的一种模型交换格式，AI 开发人员可以方便将模型部署到所需的框架之中。

- 支持的模型
- 如何使用
  - 准备工作

#### 40.1 支持的模型

MMPose 支持将训练好的各种 Pytorch 模型导出为 ONNX 格式。支持的模型包括但不限于：

- ResNet
- HRNet
- HigherHRNet

## 40.2 如何使用

用户可以使用这里的 [脚本](#) 来导出 ONNX 格式。

### 40.2.1 准备工作

首先，安装 onnx

```
pip install onnx onnxruntime
```

MMPose 提供了一个 python 脚本，将 MMPose 训练的 pytorch 模型导出到 ONNX。

```
python tools/deployment/pytorch2onnx.py ${CONFIG_FILE} ${CHECKPOINT_FILE} [--shape $
↪ ${SHAPE}] \
    [--verify] [--show] [--output-file ${OUTPUT_FILE}] [--is-localizer] [--opset-
↪ version ${VERSION}]
```

可选参数：

- `--shape`: 模型输入张量的形状。对于 2D 关键点检测模型（如 HRNet），输入形状应当为 `$batch $channel $height $width` (例如, 1 3 256 192);
- `--verify`: 是否对导出模型进行验证，验证项包括是否可运行，数值是否正确等。如果没有手动指定，默认为 `False`。
- `--show`: 是否打印导出模型的结构。如果没有手动指定，默认为 `False`。
- `--output-file`: 导出的 onnx 模型名。如果没有手动指定，默认为 `tmp.onnx`。
- `--opset-version`: 决定 onnx 的执行版本，MMPose 推荐用户使用高版本（例如 11 版本）的 onnx 以确保稳定性。如果没有手动指定，默认为 11。

如果发现提供的模型权重文件没有被成功导出，或者存在精度损失，可以在本 [repo](#) 下提出问题（issue）。



---

## 教程 6: 自定义运行配置

---

在这篇教程中，我们将会介绍如何在您的项目中自定义优化方法、训练策略、工作流和钩子。

- 自定义优化方法
  - 使用 PyTorch 支持的优化器
  - 使用自己实现的优化器
    - \* 1. 定义一个新优化器
    - \* 2. 注册这个优化器
    - \* 3. 在配置文件中指定优化器
  - 自定义优化器构造器
  - 更多设置
- 自定义训练策略
- 自定义工作流
- 自定义钩子
  - 使用自己实现的钩子
    - \* 1. 定义一个新的钩子
    - \* 2. 注册这个新的钩子
    - \* 3. 修改配置文件
  - 使用 MMCV 中的钩子

- 修改默认的运行钩子
  - \* 模型权重文件配置
  - \* 日志配置
  - \* 测试配置

## 41.1 自定义优化方法

### 41.1.1 使用 PyTorch 支持的优化器

我们现已支持 PyTorch 自带的所有优化器。若要使用这些优化器，用户只需在配置文件中修改 `optimizer` 这一项。比如说，若您想使用 Adam 优化器，可以对配置文件做如下修改

```
optimizer = dict(type='Adam', lr=0.0003, weight_decay=0.0001)
```

若要修改模型的学习率，用户只需在配置文件中修改优化器的 `lr` 参数。优化器各参数的设置可参考 PyTorch 的 [API 文档](#)。

例如，用户想要使用在 PyTorch 中配置为 `torch.optim.Adam(params, lr=0.001, betas=(0.9, 0.999), eps=1e-08, weight_decay=0, amsgrad=False)` 的 Adam 优化器，可按照以下形式修改配置文件。

```
optimizer = dict(type='Adam', lr=0.001, betas=(0.9, 0.999), eps=1e-08, weight_decay=0,
  ↳ amsgrad=False)
```

### 41.1.2 使用自己实现的优化器

#### 1. 定义一个新优化器

如果您想添加一个新的优化器，名字叫 `MyOptimizer`，参数包括 `a`、`b`、`c`，可以按照以下步骤定义该优化器。

首先，创建一个新目录 `mmpose/core/optimizer`。然后，在新文件 `mmpose/core/optimizer/my_optimizer.py` 中实现该优化器：

```
from .builder import OPTIMIZERS
from torch.optim import Optimizer

@OPTIMIZERS.register_module()
class MyOptimizer(Optimizer):
```

(下页继续)

(续上页)

```
def __init__(self, a, b, c):
```

## 2. 注册这个优化器

新优化器必须先导入主命名空间才能被成功调用。有两种实现方式。

- 修改 `mmpose/core/optimizer/__init__.py` 来导入

新定义的优化器得在 `mmpose/core/optimizer/__init__.py` 中被导入，注册器才能发现并添加它。

```
from .my_optimizer import MyOptimizer
```

- 在配置文件中使用 `custom_imports` 手动导入

```
custom_imports = dict(imports=['mmpose.core.optimizers.my_optimizer'], allow_failed_
↪ imports=False)
```

在程序运行之初，库 `mmpose.core.optimizer.my_optimizer` 将会被导入。此时类 `MyOptimizer` 会自动注册。注意只有包含类 `MyOptimizer` 的库才能被导入。`mmpose.core.optimizer.my_optimizer.MyOptimizer` 不可以被直接导入。

## 3. 在配置文件中指定优化器

在新优化器 `MyOptimizer` 注册之后，它可以在配置文件中通过 `optimizer` 调用。在配置文件中，优化器通过 `optimizer` 以如下方式指定：

```
optimizer = dict(type='SGD', lr=0.02, momentum=0.9, weight_decay=0.0001)
```

如果要使用自己实现的新优化器 `MyOptimizer`，可以进行如下修改：

```
optimizer = dict(type='MyOptimizer', a=a_value, b=b_value, c=c_value)
```

### 41.1.3 自定义优化器构造器

有些模型可能需要在优化器里对一些特别参数进行设置，例如批归一化层的权重衰减系数。用户可以通过自定义优化器构造器来实现这些精细参数的调整。

```
from mmcv.utils import build_from_cfg

from mmcv.runner.optimizer import OPTIMIZER_BUILDERS, OPTIMIZERS
```

(下页继续)

(续上页)

```
from mmpose.utils import get_root_logger
from .my_optimizer import MyOptimizer

@OPTIMIZER_BUILDERS.register_module()
class MyOptimizerConstructor:

    def __init__(self, optimizer_cfg, paramwise_cfg=None):
        pass

    def __call__(self, model):

        return my_optimizer
```

这里是默认优化器构造器的实现。它还可以用作新的优化器构造器的模板。

#### 41.1.4 更多设置

有些优化器没有实现的功能可以通过优化器构造器（例如对不同权重设置不同学习率）或者钩子实现。我们列出了一些用于稳定、加速训练的常用设置。欢迎通过 [PR](#)、[issue](#) 提出更多这样的设置。

- **使用梯度截断来稳定训练：**有些模型需要梯度截断来使梯度数值保持在某个范围，以让训练过程更加稳定。例如：

```
optimizer_config = dict(grad_clip=dict(max_norm=35, norm_type=2))
```

- **使用动量策略加速模型收敛**我们支持根据学习率来修改模型动量的动量调度器。它可以让模型收敛更快。动量调度器通常和学习率调度器一起使用。例如 3D 检测中使用下面的配置来加速收敛。更多细节可以参考 [CyclicLrUpdater](#) 和 [CyclicMomentumUpdater](#) 的实现。

```
lr_config = dict(
    policy='cyclic',
    target_ratio=(10, 1e-4),
    cyclic_times=1,
    step_ratio_up=0.4,
)
momentum_config = dict(
    policy='cyclic',
    target_ratio=(0.85 / 0.95, 1),
    cyclic_times=1,
    step_ratio_up=0.4,
)
```

## 41.2 自定义训练策略

我们默认使用的学习率变化策略为阶梯式衰减策略，即 MMCV 中的 `StepLRHook`。此外，我们还支持很多学习率变化策略，例如余弦退火策略 `CosineAnnealing` 和多项式策略 `Poly`。其调用方式如下

- 多项式策略:

```
lr_config = dict(policy='poly', power=0.9, min_lr=1e-4, by_epoch=False)
```

- 余弦退火策略:

```
lr_config = dict(
    policy='CosineAnnealing',
    warmup='linear',
    warmup_iters=1000,
    warmup_ratio=1.0 / 10,
    min_lr_ratio=1e-5)
```

## 41.3 自定义工作流

我们推荐用户在每轮训练结束后对模型进行评估，即采用 `EpochEvalHook` 工作流。不过很多用户仍采用 `val` 工作流。

工作流是一个由（阶段，轮数）构成的列表，它规定了程序运行中不同阶段的顺序和轮数。默认的工作流为

```
workflow = [('train', 1)]
```

即“训练 1 轮”。有时候用户可能想要计算模型在验证集上的某些指标（例如损失、准确率）。此时可将工作流设定为

```
[('train', 1), ('val', 1)]
```

即 1 轮训练后进行 1 轮验证，两者交替进行。

**注解:**

1. 进行验证时，模型权重不会发生变化。
2. 配置文件中，参数 `total_epochs` 只控制训练轮数，不影响验证工作流
3. 工作流 `[('train', 1), ('val', 1)]` 和 `[('train', 1)]` 不会改变 `EpochEvalHook` 的行为。因为 `EpochEvalHook` 只在 `after_train_epoch` 中被调用。而验证工作流只会影响被 `after_val_epoch` 调用的钩子。因此，工作流 `[('train', 1), ('val', 1)]` 与 `[('train', 1)]` 唯一的差别就是运行程序会在每轮训练后计算模型在验证集上的损失。

## 41.4 自定义钩子

### 41.4.1 使用自己实现的钩子

#### 1. 定义一个新的钩子

下面的例子展示了如何定义一个新的钩子并将其用于训练。

```
from mmcv.runner import HOOKS, Hook

@HOOKS.register_module()
class MyHook(Hook):

    def __init__(self, a, b):
        pass

    def before_run(self, runner):
        pass

    def after_run(self, runner):
        pass

    def before_epoch(self, runner):
        pass

    def after_epoch(self, runner):
        pass

    def before_iter(self, runner):
        pass

    def after_iter(self, runner):
        pass
```

用户需要根据钩子的实际用途定义该钩子在 `before_run`、`after_run`、`before_epoch`、`after_epoch`、`before_iter` 以及 `after_iter` 中的行为。

## 2. 注册这个新的钩子

定义好钩子 MyHook 之后，我们需要将其导入。假设 MyHook 在文件 `mmpose/core/utils/my_hook.py` 中定义，则有两种方式可以导入：

- 通过修改 `mmpose/core/utils/__init__.py` 进行导入。

新定义的模块需要被导入到 `mmpose/core/utils/__init__.py` 才能被注册器找到并添加：

```
from .my_hook import MyHook
```

- 在配置文件中 使用 `custom_imports` 手动导入

```
custom_imports = dict(imports=['mmpose.core.utils.my_hook'], allow_failed_
→ imports=False)
```

## 3. 修改配置文件

```
custom_hooks = [
    dict(type='MyHook', a=a_value, b=b_value)
]
```

用户可以通过将钩子的参数 `priority` 设置为 'NORMAL' 或 'HIGHEST' 来设定它的优先级

```
custom_hooks = [
    dict(type='MyHook', a=a_value, b=b_value, priority='NORMAL')
]
```

钩子在注册时，其优先级默认为 NORMAL。

### 41.4.2 使用 MMCV 中的钩子

用户可以直接修改配置文件来调用 MMCV 中已实现的钩子

```
mmcv_hooks = [
    dict(type='MMCVHook', a=a_value, b=b_value, priority='NORMAL')
]
```

### 41.4.3 修改默认的运行钩子

有部分常用钩子没有通过 `custom_hooks` 注册。在导入 `MMCV` 时，它们会自动注册。这些钩子包括：

- `log_config`
- `checkpoint_config`
- `evaluation`
- `lr_config`
- `optimizer_config`
- `momentum_config`

这些钩子中，只有日志钩子的优先级为 `VERY_LOW`，其他钩子的优先级都是 `NORMAL`。前面的教程已经讲述了如何修改 `optimizer_config`、`momentum_config`、`lr_config`。这里我们介绍如何修改 `log_config`、`checkpoint_config`、`evaluation`。

#### 模型权重文件配置

`MMCV` 的运行程序会使用 `checkpoint_config` 来初始化 `CheckpointHook`。

```
checkpoint_config = dict(interval=1)
```

用户可以通过设置 `max_keep_ckpts` 来保存有限的模型权重文件；通过设置 `save_optimizer` 以决定是否保存优化器的状态。[这份文档](#)介绍了更多参数的细节。

#### 日志配置

日志配置 `log_config` 可以设置多个日志钩子，并且可以设定记录间隔。目前 `MMCV` 支持的日志钩子包括 `WandbLoggerHook`、`MlflowLoggerHook`、`TensorboardLoggerHook`。[这份文档](#)介绍了更多日志钩子的使用细节。

```
log_config = dict(  
    interval=50,  
    hooks=[  
        dict(type='TextLoggerHook'),  
        dict(type='TensorboardLoggerHook')  
    ])
```



## 测试配置

测试配置 `evaluation` 可以用来初始化 `EvalHook`。除了参数 `interval`, 其他参数 (例如 `metric`) 会被传递给 `dataset.evaluate()`。

```
evaluation = dict(interval=1, metric='mAP')
```



---

### 教程 7：摄像头应用接口

---

MMPose 中提供了一组摄像头应用接口（Webcam API），用以快速搭建简单的姿态估计应用。在这篇教程中，我们将介绍摄像头应用接口的功能和使用方法。用户也可以在 [接口文档](#) 中查询详细信息。

- 摄像头应用接口概览
- 摄像头应用示例
  - 运行程序
  - 配置文件
    - \* 缓存器配置
    - \* 热键配置
  - 应用程序概览
- 通过定义新节点扩展功能
  - 定义一般的节点类：以目标检测为例
    - \* 继承节点基类 *Node*
    - \* 定义 `__init__()` 方法
    - \* 定义 `process()` 方法
    - \* 定义 `bypass()` 方法
  - 定义可视化节点类：以文字显示为例
    - \* 继承可视化节点基类 *BaseVisualizerNode*

\* 实现 *draw()* 方法

## 42.1 摄像头应用接口概览

摄像头应用接口主要由以下模块组成（如图 1 所示）：

1. **执行器** (WebcamExecutor)：应用程序的主体，提供了启动程序、读取视频、显示输出等基本功能。此外，在执行器中会根据配置构建若干功能模块，分别执行不同的任务，如模型推理、数据处理、逻辑判断、图像绘制等。这些模块会在不同线程中异步运行。执行器在读入视频数据后，会控制数据在各个功能模块间的流动，最终将处理完毕的视频数据显示在屏幕上。与执行器相关的概念还有：
  1. **配置文件** (Config)：包括执行器的基本参数，功能模块组成和模块间的逻辑关系。与 OpenMMLab 中常用的配置文件形式类似，我们使用 python 文件作为配置文件；
  2. **启动脚本**（如 `mmpose/demo/webcam_demo.py`）：读取配置文件，构建执行器类实例，并调用执行器接口启动应用程序。
2. **节点** (Node)：应用程序中的功能模块。一个节点通常用于实现一个基本功能，如 `DetectorNode` 用于实现目标检测，`ObjectVisualizerNode` 用于绘制图像中物体的检测框和关键点，`RecorderNode` 用于将视频输出保存到本地文件等。目前已经提供的节点根据功能可以分为模型节点（**Model Node**）、可视化节点（**Visualizer Node**）和辅助节点（**Helper Node**）。用户也可以根据需求，通过继承节点基类 `Node` 实现自定义的节点。
3. **公共组件** (Utils)：提供了一些公共的底层模块，其中比较重要的有：
  1. **消息类** (Message)：节点输入、输出的基本数据结构，其中可以包括图像、模型推理结果、文本以及用户自定义的数据内容；
  2. **消息缓存器** (Buffer)：用于节点间的数据通信。由于各个节点是异步运行的，需要将上游节点的输出缓存下来，等待下游节点进行读取。换言之，节点之间并不直接进行数据交换，而是从指定的消息缓存器读入数据，并将输出到其他消息缓存器；
  3. **事件** (Event)：用于执行器与节点间或不同节点间的事件通信。与数据流通过配置好的路线依次经过节点不同，事件可以立即被任意节点响应。例如，当用户按下键盘按键时，执行器会将该事件广播给所有节点，对应节点可以立即做出响应，从而实现与用户的互动。

## 42.2 摄像头应用示例

在了解摄像头应用接口的基本组成后，我们通过一个简单的例子，来介绍如何使用这组接口搭建应用程序。

### 42.2.1 运行程序

通过以下指令，可以启动示例程序。它的作用是打开摄像头，将读取到的视频显示在屏幕上，同时将其保存在本地文件中：

```
# python demo/webcam_demo.py --config CONFIG_PATH [--debug]
python demo/webcam_demo.py --config demo/webcam_cfg/test_camera.py
```

### 42.2.2 配置文件

这个示例的功能模块如下所示：

```
# 摄像头应用配置
executor_cfg = dict(
    name='Test Webcam', # 应用名称
    camera_id=0, # 摄像头 ID (也可以用本地文件路径作为视频输入)
    camera_max_fps=30, # 读取视频的最大帧率
    nodes=[
        # MonitorNode 用于显示系统和节点信息
        dict(
            type='MonitorNode', # 节点类型
            name='monitor', # 节点名称
            enable_key='m', # 开/关快捷键
            enable=False, # 初始开/关状态
            input_buffer='_frame_', # 输入缓存
            output_buffer='display'), # 输出缓存
        # RecorderNode 用于将视频保存到本地文件
        dict(
            type='RecorderNode', # 节点类型
            name='recorder', # 节点名称
            out_video_file='webcam_output.mp4', # 保存视频的路径
            input_buffer='display', # 输入缓存
            output_buffer='_display_') # 输出缓存
    ])
```

可以看到，配置文件包含一个叫做 `executor_cfg` 的字典，其中的内容包括基本参数（如 `name`, `camera_id` 等，可以参考 [WebcamExecutor 文档](#)）和节点配置信息（`nodes`）。节点配置信息是一个列表，其中每个元素是一个节点的参数字典。例如，在该示例中包括 2 个节点，类型分别是 `MonitorNode` 和 `RecorderNode`。关于节点的功能和参数，可以参考 [节点文档](#)。

## 缓存器配置

在节点配置中有一类特殊参数——输入输出缓存器。这些配置定义了节点之间的上下游逻辑关系。比如这个例子中，MonitorNode 会从名为 "\_frame\_" 的缓存器中读取数据，并将输出存放到名为 "display" 的缓存器中；而 RecorderNode 则从缓存器 "display" 中读取数据，并将输出存放到缓存器 "\_display\_" 中。

在配置文件中，用户可以任意指定缓存器的名字，执行器会自动构建缓存器，并将缓存器与对应的节点建立关联。需要注意的是，以下 3 个是保留的缓存器名字，对应用于执行器与节点间数据交换的特殊缓存器：

- "\_input\_"：存放执行器读入的视频帧，通常用于模型推理
- "\_frame\_"：存放执行器读入的视频帧（与 "\_input\_" 相同），通常用于可视化输出
- "\_display\_"：存放经过所有节点处理后的输出结果，用于执行器在屏幕上的显示

在应用程序中，所有的缓存器是由执行器中的 **缓存管理器（BufferManager）** 进行管理（可参考 [缓存管理器文档](#)）。

## 热键配置

在程序运行时，部分节点（如 MonitorNode）可以通过键盘按键，实时控制是否生效。这类节点有以下参数：

- enable\_key：指定开/关节点功能的热键
- enable：指定节点开/关的初始状态

热键响应是通过事件机制实现的。应用程序中的事件，由执行器中的 **事件管理器（EventManager）** 进行管理（可参考 [事件管理器文档](#)）。节点在初始化时，可以注册与自己相关的事件，之后就可以在运行过程中触发、等待或清除这些事件。

### 42.2.3 应用程序概览

在了解执行器、节点、缓存器、事件等基本概念后，我们可以用图 2 概括一个基于摄像头应用接口开发的应用程序的基本结构。

## 42.3 通过定义新节点扩展功能

用户可以通过定义新的节点类，来扩展摄像头应用接口的功能。我们通过一些节点类实例，介绍自定义节点类的方法。

### 42.3.1 定义一般的节点类：以目标检测为例

我们以实现目标检测功能为例，介绍实现节点类的一般步骤和注意事项。

#### 继承节点基类 Node

在定义节点类时，需要继承基类 `Node`，并用注册器 `NODES` 注册新的节点类，使其可以通过配置参数构建实例。

```
from mmpose.apis.webcam.nodes import Node, NODES

@NODES.register_module()
class DetectorNode(Node):
    ...
```

#### 定义 `__init__()` 方法

我们为 `DetectorNode` 类定义以下的初始化方法，代码如下：

```
def __init__(self,
             name: str,
             model_config: str,
             model_checkpoint: str,
             input_buffer: str,
             output_buffer: Union[str, List[str]],
             enable_key: Optional[Union[str, int]] = None,
             enable: bool = True,
             device: str = 'cuda:0',
             bbox_thr: float = 0.5):

    # 初始化基类
    super().__init__(name=name, enable_key=enable_key, enable=enable)

    # 初始化节点参数
    self.model_config = get_config_path(model_config, 'mmdet')
    self.model_checkpoint = model_checkpoint
    self.device = device.lower()
    self.bbox_thr = bbox_thr

    self.model = init_detector(
        self.model_config, self.model_checkpoint, device=self.device)

    # 注册输入、输出缓存器
```

(下页继续)

(续上页)

```
self.register_input_buffer(input_buffer, 'input', trigger=True) # 设为触发器
self.register_output_buffer(output_buffer)
```

可以看出，初始化方法中一般会执行以下操作：

1. **初始化基类**：通常需要 `name`、`enable_key` 和 `enable` 等参数；
2. **初始化节点参数**：根据需要初始化子类的参数，如在本例中，对 `model_config`、`device`、`bbox_thr` 等参数进行了初始化，并调用 `MMDetection` 中的 `init_detector` API 加载了检测模型；
3. **注册缓存器**：节点收到配置参数中指定的输入、输出缓存器名称后，需要在初始化时对相应的缓存器进行注册，从而能在程序运行时自动从缓存器中存取数据。具体的方法是：1. 对每个输入缓存器，调用基类的 `register_input_buffer()` 方法进行注册，将缓存器名（即来自配置文件的 `input_buffer` 参数）对应到一个输入名（即 `"input"`）。完成注册后，就可以在运行时通过输入名访问对应缓存器的数据（详见下一节“定义 `process()` 方法”）；2. 调用基类的 `register_output_buffer()` 对所有的输出缓存器进行注册。完成注册后，节点在运行时每次的输出会被自动存入所有的输出缓存器（每个缓存器会存入一份输出的深拷贝）。

当节点有多个输入时，由于输入数据是异步到达的，需要指定至少一个输出缓存器为触发器（Trigger）。当所有被设为触发器的输入缓存器都有数据到达时，会触发节点执行一次操作（即执行一次 `process()` 方法）。当节点只有一个输入时，也应在注册时将其显示设置为触发器。

## 定义 `process()` 方法

节点类的 `process()` 方法定义了节点的行为。我们在 `DetectorNode` 的 `process()` 方法中实现目标检测模型的推理过程：

```
def process(self, input_msgs):

    # 根据输入名 "input", 从输入缓存器获取数据
    input_msg = input_msgs['input']

    # 从输入数据中获取视频帧图像
    img = input_msg.get_image()

    # 使用 MMDetection API 进行检测模型推理
    preds = inference_detector(self.model, img)
    objects = self._post_process(preds)

    # 将目标检测结果存入数据
    input_msg.update_objects(objects)

    # 返回节点处理结果
    return input_msg
```



这段代码主要完成了以下操作：

1. **读取输入数据**：process() 方法的参数 input\_msgs 中包含所有已注册输入缓存器的数据，可以通过输入名（如 "input"）获得对应缓存器的数据；
2. **解析输入数据**：缓存器中的数据通常为“帧信息”（FrameMessage，详见文档）。节点可以从中获取视频帧的图像，模型推理结果等信息。
3. **处理输入数据**：这里使用 MMDetection 中的 inference\_detector() API 检测视频帧中的物体，并进行后处理（略）。
4. **返回结果**：将模型推理得到的视频帧中的物体信息，通过 update\_objects() 方法添加进 input\_msg 中，并将其返回。该结果会被自动发送到 DetectorNode 的所有输出缓存器，供下游节点读取。

### 定义 bypass() 方法

由于允许通过热键开关 DetectorNode 的功能，因此需要实现 bypass() 方法，定义该节点在处于关闭状态时的行为。bypass() 方法与 process() 方法的接口完全相同。DetectorNode 在关闭时不需要对输入做任何处理，因此对 bypass() 实现如下：

```
def bypass(self, input_msgs):
    return input_msgs['input']
```

## 42.3.2 定义可视化节点类：以文字显示为例

可视化节点是一类特殊的节点，它们的功能是对视频帧的图像进行编辑。为了方便拓展可视化功能，我们为可视化节点提供了更简单的抽象接口。这里以实现文字显示为例，介绍实现可视化节点的一般步骤和注意事项。

### 继承可视化节点基类 BaseVisualizerNode

可视化节点基类 BaseVisualizerNode 继承自 Node 类，并对 process() 方法进行了进一步封装，暴露 draw() 接口供子类实现具体的可视化功能。与一般的节点类类似，可视化节点类需要继承 BaseVisualizerNode 并注册进 NODES。

```
from mmpose.apis.webcam.nodes import BaseVisualizerNode, NODES

@NODES.register_module()
class NoticeBoardNode(BaseVisualizerNode):
    ...
```

节点初始化方法的实现方式与一般节点类似，请参考定义 \_\_init\_\_() 方法。需要注意的是，可视化节点必须注册唯一的输入缓存器，对应于输入名 "input"。

## 实现 draw() 方法

可视化节点类的 draw() 方法用于绘制对帧图像的更新。draw() 方法有 1 个输入参数 input\_msg，为来自 input 缓存器的数据；draw() 方法的返回值应为一幅图像（np.ndarray 类型），该图像将被用于更新 input\_msg 中的图像。节点的输出即为更新图像后的 input\_msg。

NoticeBoardNode 的 draw() 方法实现如下：

```
def draw(self, input_msg: FrameMessage) -> np.ndarray:
    # 获取帧图像
    img = input_msg.get_image()

    # 创建画布图像
    canvas = np.full(img.shape, self.background_color, dtype=img.dtype)

    # 逐行将文字绘制在画布图像上
    x = self.x_offset
    y = self.y_offset
    max_len = max([len(line) for line in self.content_lines])

    def _put_line(line=''):
        nonlocal y
        cv2.putText(canvas, line, (x, y), cv2.FONT_HERSHEY_DUPLEX,
                      self.text_scale, self.text_color, 1)
        y += self.y_delta

    for line in self.content_lines:
        _put_line(line)

    # 将画布图像的有效区域叠加在帧图像上
    x1 = max(0, self.x_offset)
    x2 = min(img.shape[1], int(x + max_len * self.text_scale * 20))
    y1 = max(0, self.y_offset - self.y_delta)
    y2 = min(img.shape[0], y)

    src1 = canvas[y1:y2, x1:x2]
    src2 = img[y1:y2, x1:x2]
    img[y1:y2, x1:x2] = cv2.addWeighted(src1, 0.5, src2, 0.5, 0)

    # 返回绘制结果
    return img
```

## CHAPTER 43

---

### 常用工具

---

内容建设中……

- 常见问题
  - 安装
  - 开发
  - 数据
  - 训练
  - 评测
  - 推理
  - 部署



我们在这里列出了一些常见问题及其相应的解决方案。如果您发现任何常见问题并有方法帮助解决，欢迎随时丰富列表。如果这里的内容没有涵盖您的问题，请按照[提问模板](#)在 [GitHub](#) 上提出问题，并补充模板中需要的信息。

### 44.1 安装

- MMCV 与 MMPose 的兼容问题。如 “AssertionError: MMCV==xxx is used but incompatible. Please install mmcv>=xxx, <=xxx.”

这里列举了各版本 MMPose 对 MMCV 版本的依赖，请选择合适的 MMCV 版本来避免安装和使用中的问题。

- 无法安装 `xtcocotools`

1. 尝试使用 `pip` 手动安装：`pip install xtcocotools`.
2. 如果第一步无法安装，尝试从[源码](#)安装：

```
git clone https://github.com/jin-s13/xtcocoapi
cd xtcocoapi
python setup.py install
```

- 报错: `No matching distribution found for xtcocotools>=1.6`

1. 安装 `cython`: `pip install cython`.

## 2. 从源码 安装 xtcocotools :

```
git clone https://github.com/jin-s13/xtcocoapi
cd xtcocoapi
python setup.py install
```

### • 报错:” No module named ‘mmcv.ops’” ; “No module named ‘mmcv.\_ext’”

1. 如果您已经安装了 mmcv, 您需要运行 `pip uninstall mmcv` 来卸载已经安装的 mmcv。如果您在本地同时安装了 mmcv 和 mmcv-full, `ModuleNotFoundError` 将会抛出。
2. 按照[安装指南](#)安装 mmcv-full。

## 44.2 开发

- 如果对源码进行了改动, 需要重新安装以使改动生效吗?

如果您遵照[最佳实践](#)的指引, 从源码安装 mmpose, 那么任何本地修改都不需要重新安装即可生效。

- 如何在多个 MMPose 版本下进行开发?

通常来说, 我们推荐通过不同虚拟环境来管理多个开发目录下的 MMPose。但如果您希望在不同目录 (如 mmpose-0.26.0, mmpose-0.25.0 等) 使用同一个环境进行开发, 我们提供的训练和测试 shell 脚本会自动使用当前目录的 mmpose, 其他 Python 脚本则可以在命令前添加 `PYTHONPATH=`pwd`` 来使用当前目录的代码。

反过来, 如果您希望 shell 脚本使用环境中安装的 MMPose, 而不是当前目录的, 则可以去掉 shell 脚本中如下一行代码:

```
PYTHONPATH="$(dirname $0)/..":$PYTHONPATH
```

- 加载模型过程中:” Unexpected keys in source state dict” when loading pre-trained weights

在姿态估计模型中不使用预训练模型中的某些层是正常的。在 ImageNet 预训练的分类网络和姿态估计网络可能具有不同的架构 (例如, 没有分类头)。因此, 在源模型权重文件 (source state dict) 中确实会出现一些预期之外的键。

- 怎么使用经过训练的模型作为主干网络的预训练

如果要对整个网络 (主干网络 + 头部网络) 使用预训练模型, 请参考教程文档: [使用预训练模型](#), 配置文件中的 `load_from` 字段指明了预训练模型的链接。

如果要使用主干网进行预训练, 可以将配置文件中主干网络的 “pretrained” 值改为模型权重文件的路径或者 URL。训练时, 将忽略意外的键值。

- 怎么实时地可视化训练的准确率/损失函数曲线?

在 `log_config` 中使用 `TensorboardLoggerHook`, 如:

```
log_config=dict(interval=20, hooks=[dict(type='TensorboardLoggerHook')])
```

您还可以参考教程文档:[自定义运行配置](#) 以及配置文件的例子。

- 没有打印日志信息

使用更小的日志打印间隔。例如, 将这个配置文件的 `interval=50` 改为 `interval=1`。

- 微调模型的时候怎么固定主干网络多个阶段的网络参数?

您可以参考这个函数: `def _freeze_stages()` 以及这个参数: `frozen_stages`。如果使用分布式训练或者测试, 请在配置文件中设置 `find_unused_parameters = True`。

## 44.3 数据

- 怎么将 2D 关键点数据集转换为 COCO 格式?

您可以参考这个[转换工具](#)来准备您的数据。这是一个关于 COCO 格式 json 文件的示例。COCO 格式的 json 文件需要这些字段信息: “categories”, “annotations” and “images”. “categories” 包括了数据集的一些基本信息, 如类别名称和关键点名称。” images” 包含了图片级别的信息, 需要这些字段的信息:” id”, “file\_name”, “height”, “width”. 其他字段是可选的。注: “id” 可以是不连续或者没有排序好的 (如 1000, 40, 352, 333 …)。

“annotations” 包含了实例级别的信息, 需要这些字段的信息:” image\_id”, “id”, “keypoints”, “num\_keypoints”, “bbox”, “iscrowd”, “area”, “category\_id”. 其他字段是可选的。注: (1) “num\_keypoints” 表示可见关键点的数量. (2) 默认情况下, 请设置 “iscrowd: 0”. (3) “area” 可以通过标记的边界框信息计算得到: ( $\text{area} = w * h$ ). (4) 简单地设置 “category\_id: 1” 即可. (5) “annotations” 中的 “image\_id” 应该和 “images” 中的 “id” 匹配。

- 如果数据集没有人工标注的边界框信息怎么办?

我们可以认为一个人的边界框是恰好包围所有关键点的最小框。

- 如果数据集没有 segmentation 信息怎么办?

设置人体的 area 为边界框的面积即可。在评测的时候, 请按照这个[例子](#)设置 `use_area=False`。

- COCO\_val2017\_detections\_AP\_H\_56\_person.json 是什么文件? 可以不使用它来训练姿态估计的模型吗?

COCO\_val2017\_detections\_AP\_H\_56\_person.json 包含了在 COCO 验证集上检测到的人体边界框, 是使用 FasterRCNN 生成的。您可以使用真实标记的边界框来评测模型, 设置 `use_gt_bbox=True` 和 `bbox_file=''` 即可。或者您可以使用检测到的边界框来评测模型的泛化性, 只要设置 `use_gt_bbox=False` 和 `bbox_file='COCO_val2017_detections_AP_H_56_person.json'` 即可。

## 44.4 训练

- 报错: `RuntimeError: Address already in use`

设置环境变量 `MASTER_PORT=XXX`。例如 `MASTER_PORT=29517 GPUS=16 GPUS_PER_NODE=8 CPUS_PER_TASK=2 ./tools/slurm_train.sh Test res50 configs/body/2D_Kpt_SV_RGB_Img/topdown_hm/coco/res50_coco_256x192.py work_dirs/res50_coco_256x192`。

## 44.5 评测

- 怎么在 **MPII** 测试集上运行评测？

因为我们没有 **MPII** 测试集上的真实标注信息，我们不能在本地评测。如果您获得在测试集上的评测结果，根据 **MPII** 指南，您需要通过邮件上传这个文件 `pred.mat`（在测试过程中生成）到官方的服务器。

- 对于自顶向下的 **2D** 姿态估计方法，为什么预测的关键点坐标可以超出边界框？

**MMPose** 没有直接使用边界框来裁剪图片。边界框首先会被转换至中心和尺度，尺度会乘上一个系数 (1.25) 来包含一些图片的上下文信息。如果图片的宽高比和模型的输入 (通常是 192/256) 不一致，会调整这个边界框的大小。详细可参考[代码](#)。

## 44.6 推理

- 怎么在 **CPU** 上运行 **MMPose** ？

运行示例的时候设置: `--device=cpu`。

- 怎么加快推理速度？

对于自顶向下的模型，可以尝试修改配置文件。例如：

1. 在 `topdown-res50` 中设置 `flip_test=False`。
2. 在 `topdown-res50` 中设置 `post_process='default'`。
3. 使用更快的人体边界框检测器，可参考 **MMDetection**。

对于自底向上的模型，也可以尝试修改配置文件。例如：

1. 在 `AE-res50` 中设置 `flip_test=False`。
2. 在 `AE-res50` 中设置 `adjust=False`。
3. 在 `AE-res50` 中设置 `refine=False`。
4. 在 `AE-res50` 中使用更小的输入图片尺寸。



## 44.7 部署

- 为什么用 MMPose 转换导出的 onnx 模型在转移到其他框架如 TensorRT 时会抛出错误？

目前，我们只能确保 MMPose 中的模型与 onnx 兼容。但是您的目标部署框架可能不支持 onnx 中的某些操作，例如[这个问题](#)中提到的 TensorRT。

请注意，MMPose 中的 pytorch2onnx 将不再维护，未来将不再保留。我们将在 [MMDeploy](#) 中支持所有 OpenMMLab 代码库的模型部署，包括 MMPose。您可以在其[文档](#)中找到有关受支持模型和用户指南的详细信息，并提出问题以请求支持您要使用的模型。



`mmpose.apis.collect_multi_frames` (*video*, *frame\_id*, *indices*, *online=False*)

Collect multi frames from the video.

**参数**

- **video** (*mmcv.VideoReader*) –A VideoReader of the input video file.
- **frame\_id** (*int*) –index of the current frame
- **indices** (*list(int)*) –index offsets of the frames to collect
- **online** (*bool*) –inference mode, if set to True, can not use future frame information.

**返回** multi frames collected from the input video file.

**返回类型** `list(ndarray)`

`mmpose.apis.extract_pose_sequence` (*pose\_results*, *frame\_idx*, *causal*, *seq\_len*, *step=1*)

Extract the target frame from 2D pose results, and pad the sequence to a fixed length.

**参数**

- **pose\_results** (*list[list[dict]]*) –Multi-frame pose detection results stored in a nested list. Each element of the outer list is the pose detection results of a single frame, and each element of the inner list is the pose information of one person, which contains:
  - keypoints (*ndarray[K, 2 or 3]*): x, y, [score]
  - track\_id (*int*): unique id of each person, required when `with_track_id==True`.
  - bbox (*((4, ) or (5, ))*): left, right, top, bottom, [score]

- **frame\_idx** (*int*) –The index of the frame in the original video.
- **causal** (*bool*) –If True, the target frame is the last frame in a sequence. Otherwise, the target frame is in the middle of a sequence.
- **seq\_len** (*int*) –The number of frames in the input sequence.
- **step** (*int*) –Step size to extract frames from the video.

返回 Multi-frame pose detection results stored in a nested list with a length of seq\_len.

返回类型 list[list[dict]]

```
mmpose.apis.get_track_id(results, results_last, next_id, min_keypoints=3, use_oks=False, tracking_thr=0.3,  
                        use_one_euro=False, fps=None, sigmas=None)
```

Get track id for each person instance on the current frame.

参数

- **results** (*list[dict]*) –The bbox & pose results of the current frame (bbox\_result, pose\_result).
- **results\_last** (*list[dict]*, *optional*) –The bbox & pose & track\_id info of the last frame (bbox\_result, pose\_result, track\_id). None is equivalent to an empty result list. Default: None
- **next\_id** (*int*) –The track id for the new person instance.
- **min\_keypoints** (*int*) –Minimum number of keypoints recognized as person. 0 means no minimum threshold required. Default: 3.
- **use\_oks** (*bool*) –Flag to using oks tracking. default: False.
- **tracking\_thr** (*float*) –The threshold for tracking.
- **use\_one\_euro** (*bool*) –Option to use one-euro-filter. default: False.
- **fps** (*optional*) –Parameters that d\_cutoff when one-euro-filter is used as a video input
- **sigmas** (*np.ndarray*) –Standard deviation of keypoint labelling. It is necessary for oks\_iou tracking (*use\_oks==True*). It will be use sigmas of COCO as default if it is set to None. Default is None.

返回

- results (list[dict]): The bbox & pose & track\_id info of the current frame (bbox\_result, pose\_result, track\_id).
- next\_id (int): The track id for the new person instance.

返回类型 tuple

---

```
mmpose.apis.inference_bottom_up_pose_model(model, img_or_path, dataset='BottomUpCocoDataset',
                                             dataset_info=None, pose_nms_thr=0.9,
                                             return_heatmap=False, outputs=None)
```

Inference a single image with a bottom-up pose model.

---

注解:

- num\_people: P
  - num\_keypoints: K
  - bbox height: H
  - bbox width: W
- 

参数

- **model** (*nn.Module*) –The loaded pose model.
- **img\_or\_path** (*str* | *np.ndarray*) –Image filename or loaded image.
- **dataset** (*str*) –Dataset name, e.g. ‘BottomUpCocoDataset’ . It is deprecated. Please use **dataset\_info** instead.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **pose\_nms\_thr** (*float*) –retain oks overlap < pose\_nms\_thr, default: 0.9.
- **return\_heatmap** (*bool*) –Flag to return heatmap, default: False.
- **outputs** (*list(str)* | *tuple(str)*) –Names of layers whose outputs need to be returned, default: None.

返回

- **pose\_results** (*list[*np.ndarray*]*): The predicted pose info. The length of the list is the number of people (P). Each item in the list is a ndarray, containing each person’s pose (*np.ndarray*[Kx3]): x, y, score.
- **returned\_outputs** (*list[dict[*np.ndarray*[N, K, H, W] | torch.Tensor[N, K, H, W]]]*): Output feature maps from layers specified in *outputs*. Includes ‘heatmap’ if *return\_heatmap* is True.

返回类型 *tuple*

```
mmpose.apis.inference_interhand_3d_model(model, img_or_path, det_results, bbox_thr=None,
                                           format='xywh', dataset='InterHand3DDataset')
```

Inference a single image with a list of hand bounding boxes.

---

注解:

---

- num\_bboxes: N
  - num\_keypoints: K
- 

### 参数

- **model** (*nn.Module*) –The loaded pose model.
- **img\_or\_path** (*str* | *np.ndarray*) –Image filename or loaded image.
- **det\_results** (*list[dict]*) –The 2D bbox sequences stored in a list. Each element of the list is the bbox of one person, whose shape is (*ndarray*[4 or 5]), containing 4 box coordinates (and score).
- **dataset** (*str*) –Dataset name.
- **format** –bbox format ( 'xyxy' | 'xywh' ). Default: 'xywh' . 'xyxy' means (left, top, right, bottom), 'xywh' means (left, top, width, height).

**返回** 3D pose inference results. Each element is the result of an instance, which contains the predicted 3D keypoints with shape (*ndarray*[K,3]). If there is no valid instance, an empty list will be returned.

**返回类型** *list[dict]*

```
mmpose.apis.inference_mesh_model(model, img_or_path, det_results, bbox_thr=None, format='xywh',  
                                  dataset='MeshH36MDataset')
```

Inference a single image with a list of bounding boxes.

---

### 注解:

- num\_bboxes: N
  - num\_keypoints: K
  - num\_vertices: V
  - num\_faces: F
- 

### 参数

- **model** (*nn.Module*) –The loaded pose model.
- **img\_or\_path** (*str* | *np.ndarray*) –Image filename or loaded image.
- **det\_results** (*list[dict]*) –The 2D bbox sequences stored in a list. Each element of the list is the bbox of one person. “bbox” (*ndarray*[4 or 5]): The person bounding box, which contains 4 box coordinates (and score).
- **bbox\_thr** (*float* | *None*) –Threshold for bounding boxes. Only bboxes with higher scores will be fed into the pose detector. If *bbox\_thr* is *None*, all boxes will be used.

- **format** (*str*) – bbox format ( ‘xyxy’ | ‘xywh’ ). Default: ‘xywh’ .
  - ‘xyxy’ means (left, top, right, bottom),
  - ‘xywh’ means (left, top, width, height).
- **dataset** (*str*) – Dataset name.

## 返回

3D pose inference results. Each element is the result of an instance, which contains:

- ‘bbox’ (ndarray[4]): instance bounding bbox
- ‘center’ (ndarray[2]): bbox center
- ‘scale’ (ndarray[2]): bbox scale
- ‘keypoints\_3d’ (ndarray[K,3]): predicted 3D keypoints
- ‘camera’ (ndarray[3]): camera parameters
- ‘vertices’ (ndarray[V, 3]): predicted 3D vertices
- ‘faces’ (ndarray[F, 3]): mesh faces

If there is no valid instance, an empty list will be returned.

返回类型 list[dict]

```
mmpose.apis.inference_pose_lifter_model(model, pose_results_2d, dataset=None,
                                         dataset_info=None, with_track_id=True,
                                         image_size=None, norm_pose_2d=False)
```

Inference 3D pose from 2D pose sequences using a pose lifter model.

## 参数

- **model** (*nn.Module*) – The loaded pose lifter model
- **pose\_results\_2d** (*list[list[dict]]*) – The 2D pose sequences stored in a nested list. Each element of the outer list is the 2D pose results of a single frame, and each element of the inner list is the 2D pose of one person, which contains:
  - “keypoints” (ndarray[K, 2 or 3]): x, y, [score]
  - “track\_id” (int)
- **dataset** (*str*) – Dataset name, e.g. ‘Body3DH36MDataset’
- **with\_track\_id** – If True, the element in pose\_results\_2d is expected to contain “track\_id”, which will be used to gather the pose sequence of a person from multiple frames. Otherwise, the pose results in each frame are expected to have a consistent number and order of identities. Default is True.

- **image\_size** (*tuple/list*) –image width, image height. If None, image size will not be contained in dict data.
- **norm\_pose\_2d** (*bool*) –If True, scale the bbox (along with the 2D pose) to the average bbox scale of the dataset, and move the bbox (along with the 2D pose) to the average bbox center of the dataset.

### 返回

3D pose inference results. Each element is the result of an instance, which contains:

- " keypoints\_3d" (ndarray[K, 3]): predicted 3D keypoints
- " keypoints" (ndarray[K, 2 or 3]): from the last frame in `pose_results_2d`.
- " track\_id" (int): from the last frame in `pose_results_2d`. If there is no valid instance, an empty list will be returned.

返回类型 list[dict]

```
mmpose.apis.inference_top_down_pose_model (model, imgs_or_paths, person_results=None,  
                                             bbox_thr=None, format='xywh',  
                                             dataset='TopDownCocoDataset', dataset_info=None,  
                                             return_heatmap=False, outputs=None)
```

Inference a single image with a list of person bounding boxes. Support single-frame and multi-frame inference setting.

---

### 注解:

- num\_frames: F
  - num\_people: P
  - num\_keypoints: K
  - bbox height: H
  - bbox width: W
- 

### 参数

- **model** (*nn.Module*) –The loaded pose model.
- **imgs\_or\_paths** (*str | np.ndarray | list(str) | list(np.ndarray)*) –Image filename(s) or loaded image(s).
- **person\_results** (*list(dict), optional*) –a list of detected persons that contains bbox and/or track\_id:
  - **bbox** (4, ) or (5, ): The person bounding box, which contains 4 box coordinates (and score).



- **track\_id** (int): The unique id for each human instance. If not provided, a dummy person result with a bbox covering the entire image will be used. Default: None.
- **bbox\_thr** (float | None) –Threshold for bounding boxes. Only bboxes with higher scores will be fed into the pose detector. If bbox\_thr is None, all boxes will be used.
- **format** (str) –bbox format ( ‘xyxy’ | ‘xywh’ ). Default: ‘xywh’ .
  - xyxy means (left, top, right, bottom),
  - xywh means (left, top, width, height).
- **dataset** (str) –Dataset name, e.g. ‘TopDownCocoDataset’ . It is deprecated. Please use dataset\_info instead.
- **dataset\_info** (DatasetInfo) –A class containing all dataset info.
- **return\_heatmap** (bool) –Flag to return heatmap, default: False
- **outputs** (list(str) | tuple(str)) –Names of layers whose outputs need to be returned. Default: None.

#### 返回

- pose\_results (list[dict]): The bbox & pose info. Each item in the list is a dictionary, containing the bbox: (left, top, right, bottom, [score]) and the pose (ndarray[Kx3]): x, y, score.
- returned\_outputs (list[dict[np.ndarray[N, K, H, W] | torch.Tensor[N, K, H, W]]): Output feature maps from layers specified in *outputs*. Includes ‘heatmap’ if *return\_heatmap* is True.

#### 返回类型 tuple

`mmpose.apis.init_pose_model (config, checkpoint=None, device='cuda:0')`

Initialize a pose model from config file.

#### 参数

- **config** (str or `mmcv.Config`) –Config file path or the config object.
- **checkpoint** (str, optional) –Checkpoint path. If left as None, the model will not load any weights.

**返回** The constructed detector.

#### 返回类型 nn.Module

`mmpose.apis.init_random_seed (seed=None, device='cuda')`

Initialize random seed.

If the seed is not set, the seed will be automatically randomized, and then broadcast to all processes to prevent some potential bugs.

#### 参数

- **seed** (int, Optional) –The seed. Default to None.

- **device** (*str*) –The device where the seed will be put on. Default to ‘cuda’.

返回 Seed to be used.

返回类型 `int`

`mmpose.apis.multi_gpu_test(model, data_loader, tmpdir=None, gpu_collect=False)`

Test model with multiple gpus.

This method tests model with multiple gpus and collects the results under two different modes: gpu and cpu modes. By setting ‘gpu\_collect=True’ it encodes results to gpu tensors and use gpu communication for results collection. On cpu mode it saves the results on different gpus to ‘tmpdir’ and collects them by the rank 0 worker.

参数

- **model** (*nn.Module*) –Model to be tested.
- **data\_loader** (*nn.DataLoader*) –Pytorch data loader.
- **tmpdir** (*str*) –Path of directory to save the temporary results from different gpus under cpu mode.
- **gpu\_collect** (*bool*) –Option to use either gpu or cpu to collect results.

返回 The prediction results.

返回类型 `list`

`mmpose.apis.process_mmdet_results(mmdet_results, cat_id=1)`

Process mmdet results, and return a list of bboxes.

参数

- **mmdet\_results** (*list/tuple*) –mmdet results.
- **cat\_id** (*int*) –category id (default: 1 for human)

返回 a list of detected bounding boxes

返回类型 `person_results (list)`

`mmpose.apis.single_gpu_test(model, data_loader)`

Test model with a single gpu.

This method tests model with a single gpu and displays test progress bar.

参数

- **model** (*nn.Module*) –Model to be tested.
- **data\_loader** (*nn.DataLoader*) –Pytorch data loader.

返回 The prediction results.

返回类型 `list`

`mmpose.apis.train_model` (*model, dataset, cfg, distributed=False, validate=False, timestamp=None, meta=None*)

Train model entry function.

#### 参数

- **model** (*nn.Module*) –The model to be trained.
- **dataset** (*Dataset*) –Train dataset.
- **cfg** (*dict*) –The config dict for training.
- **distributed** (*bool*) –Whether to use distributed training. Default: False.
- **validate** (*bool*) –Whether to do evaluation. Default: False.
- **timestamp** (*str | None*) –Local time for runner. Default: None.
- **meta** (*dict | None*) –Meta dict to record some important information. Default: None

`mmpose.apis.vis_3d_mesh_result` (*model, result, img=None, show=False, out\_file=None*)

Visualize the 3D mesh estimation results.

#### 参数

- **model** (*nn.Module*) –The loaded model.
- **result** (*list[dict]*) –3D mesh estimation results.

`mmpose.apis.vis_3d_pose_result` (*model, result, img=None, dataset='Body3DH36MDataset', dataset\_info=None, kpt\_score\_thr=0.3, radius=8, thickness=2, vis\_height=400, num\_instances=-1, axis\_azimuth=70, show=False, out\_file=None*)

Visualize the 3D pose estimation results.

#### 参数

- **model** (*nn.Module*) –The loaded model.
- **result** (*list[dict]*) –

`mmpose.apis.vis_pose_result` (*model, img, result, radius=4, thickness=1, kpt\_score\_thr=0.3, bbox\_color='green', dataset='TopDownCocoDataset', dataset\_info=None, show=False, out\_file=None*)

Visualize the detection results on the image.

#### 参数

- **model** (*nn.Module*) –The loaded detector.
- **img** (*str | np.ndarray*) –Image filename or loaded image.
- **result** (*list[dict]*) –The results to draw over *img* (bbox\_result, pose\_result).
- **radius** (*int*) –Radius of circles.

- **thickness** (*int*) –Thickness of lines.
- **kpt\_score\_thr** (*float*) –The threshold to visualize the keypoints.
- **skeleton** (*list[tuple()]*) –Default None.
- **show** (*bool*) –Whether to show the image. Default True.
- **out\_file** (*str/None*) –The filename of the output visualization image.

```
mmpose.apis.vis_pose_tracking_result(model, img, result, radius=4, thickness=1, kpt_score_thr=0.3,  
                                     dataset='TopDownCocoDataset', dataset_info=None,  
                                     show=False, out_file=None)
```

Visualize the pose tracking results on the image.

#### 参数

- **model** (*nn.Module*) –The loaded detector.
- **img** (*str | np.ndarray*) –Image filename or loaded image.
- **result** (*list[dict]*) –The results to draw over *img* (bbox\_result, pose\_result).
- **radius** (*int*) –Radius of circles.
- **thickness** (*int*) –Thickness of lines.
- **kpt\_score\_thr** (*float*) –The threshold to visualize the keypoints.
- **skeleton** (*list[tuple]*) –Default None.
- **show** (*bool*) –Whether to show the image. Default True.
- **out\_file** (*str/None*) –The filename of the output visualization image.

---

mmpose.apis.webcam

---

### **MMPose Webcam API: Tools to build simple interactive webcam applications and demos**

- *Executor*
- *Nodes*
  - *Base Nodes*
  - *Model Nodes*
  - *Visualizer Nodes*
  - *Helper Nodes*
- *Utils*
  - *Buffer and Message*
  - *Pose*
  - *Event*
  - *Misc*

## 46.1 Executor

---

*WebcamExecutor*

The interface to build and execute webcam applications from configs.

---

### 46.1.1 mmpose.apis.webcam.WebcamExecutor

```
class mmpose.apis.webcam.WebcamExecutor (nodes: List[Dict], name: str = 'MMPose Webcam App',
                                          camera_id: Union[int, str] = 0, camera_max_fps: int =
                                          30, camera_frame_shape: Optional[Tuple[int, int]] =
                                          None, synchronous: bool = False, buffer_sizes:
                                          Optional[Dict[str, int]] = None)
```

The interface to build and execute webcam applications from configs.

#### 参数

- **nodes** (*list[dict]*) –Node configs. See `webcam.nodes.Node` for details
- **name** (*str*) –Executor name. Default: 'MMPose Webcam App' .
- **camera\_id** (*int | str*) –The camera ID (usually the ID of the default camera is 0). Alternatively a file path or a URL can be given to load from a video or image file.
- **camera\_frame\_shape** (*tuple, optional*) –Set the frame shape of the camera in (width, height). If not given, the default frame shape will be used. This argument is only valid when using a camera as the input source. Default: None
- **camera\_max\_fps** (*int*) –Video reading maximum FPS. Default: 30
- **buffer\_sizes** (*dict, optional*) –A dict to specify buffer sizes. The key is the buffer name and the value is the buffer size. Default: None

#### Example::

```
>>> cfg = dict(
>>>     name='Test Webcam',
>>>     camera_id=0,
>>>     camera_max_fps=30,
>>>     nodes=[
>>>         dict(
>>>             type='MonitorNode',
>>>             name='monitor',
>>>             enable_key='m',
>>>             enable=False,
>>>             input_buffer='_frame_',
```

(下页继续)

(续上页)

```
>>>         output_buffer='display'),
>>>     dict(
>>>         type='RecorderNode',
>>>         name='recorder',
>>>         out_video_file='webcam_output.mp4',
>>>         input_buffer='display',
>>>         output_buffer='_display_')
>>>     ])
```

```
>>> executor = WebcamExecutor(**cfg)
```

`__init__` (nodes: List[Dict], name: str = 'MMPose Webcam App', camera\_id: Union[int, str] = 0, camera\_max\_fps: int = 30, camera\_frame\_shape: Optional[Tuple[int, int]] = None, synchronous: bool = False, buffer\_sizes: Optional[Dict[str, int]] = None)

Initialize self. See help(type(self)) for accurate signature.

Methods

<code>__init__</code> (nodes[, name, camera_id, ...])	Initialize self.
<code>run</code> ()	Start the executor.

46.2 Nodes

46.2.1 Base Nodes

<code>Node</code>	Base class for node, which is the interface of basic function module.
<code>BaseVisualizerNode</code>	Base class for nodes whose function is to create visual effects, like visualizing model predictions, showing graphics or showing text messages.

### `mmpose.apis.webcam.nodes.Node`

```
class mmpose.apis.webcam.nodes.Node (name: str, enable_key: Optional[Union[str, int]] = None,  
                                     max_fps: int = 30, input_check_interval: float = 0.01, enable:  
                                     bool = True, daemon: bool = False, multi_input: bool = False)
```

Base class for node, which is the interface of basic function module.

`Node` inherits `threading.Thread`. All subclasses should override following methods:

- `process()`
- `bypass()` (optional)

#### 参数

- **name** (*str*) –The node name (also thread name)
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: `None`
- **max\_fps** (*int*) –Maximum FPS of the node. This is to avoid the node running unrestrictedly and causing large resource consuming. Default: 30
- **input\_check\_interval** (*float*) –Minimum interval (in millisecond) between checking if input is ready. Default: 0.001
- **enable** (*bool*) –Default enable/disable status. Default: `True`
- **daemon** (*bool*) –Whether node is a daemon. Default: `True`
- **multi\_input** (*bool*) –Whether load all messages in buffer. If `False`, only one message will be loaded each time. Default: `False`

### `mmpose.apis.webcam.nodes.BaseVisualizerNode`

```
class mmpose.apis.webcam.nodes.BaseVisualizerNode (name: str, input_buffer: str, output_buffer:  
                                                    Union[str, List[str]], enable_key:  
                                                    Optional[Union[str, int]] = None, enable:  
                                                    bool = True)
```

Base class for nodes whose function is to create visual effects, like visualizing model predictions, showing graphics or showing text messages.

All subclass should implement the method `draw()`.

#### 参数

- **name** (*str*) –The node name (also thread name)



- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str | list*) –The name(s) of the output buffer(s).
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: `None`
- **enable** (*bool*) –Default enable/disable status. Default: `True`

46.2.2 Model Nodes

<i>DetectorNode</i>	Detect objects from the frame image using MMDetection model.
<i>TopDownPoseEstimatorNode</i>	Perform top-down pose estimation using MMPose model.
<i>PoseTrackerNode</i>	Perform object detection and top-down pose estimation.

mmpose.apis.webcam.nodes.DetectorNode

```
class mmpose.apis.webcam.nodes.DetectorNode (name: str, model_config: str, model_checkpoint: str,
                                             input_buffer: str, output_buffer: Union[str, List[str]],
                                             enable_key: Optional[Union[str, int]] = None,
                                             enable: bool = True, device: str = 'cuda:0', bbox_thr:
                                             float = 0.5, multi_input: bool = False)
```

Detect objects from the frame image using MMDetection model.

Note that MMDetection is required for this node. Please refer to [MMDetection documentation](#) for the installation guide.

参数

- **name** (*str*) –The node name (also thread name)
- **model\_cfg** (*str*) –The model config file
- **model\_checkpoint** (*str*) –The model checkpoint file
- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node

behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: None

- **enable** (*bool*) –Default enable/disable status. Default: True
- **device** (*str*) –Specify the device to hold model weights and inference the model. Default: 'cuda:0'
- **bbox\_thr** (*float*) –Set a threshold to filter out objects with low bbox scores. Default: 0.5
- **multi\_input** (*bool*) –Whether load all frames in input buffer. If True, all frames in buffer will be loaded and stacked. The latest frame is used to detect objects of interest. Default: False

**Example::**

```
>>> cfg = dict(  
...     type='DetectorNode',  
...     name='detector',  
...     model_config='demo/mmdetection_cfg/'  
...     'ssdlite_mobilenetv2_scratch_600e_coco.py',  
...     model_checkpoint='https://download.openmmlab.com/'  
...     '/mmdetection/v2.0/ssd/'  
...     'ssdlite_mobilenetv2_scratch_600e_coco/ssdlite_mobilenetv2_'  
...     'scratch_600e_coco_20210629_110627-974d9307.pth',  
...     # `__input__` is an executor-reserved buffer  
...     input_buffer='__input__',  
...     output_buffer='det_result')
```

```
>>> from mmpose.apis.webcam.nodes import NODES  
>>> node = NODES.build(cfg)
```

**mmpose.apis.webcam.nodes.TopDownPoseEstimatorNode**

```
class mmpose.apis.webcam.nodes.TopDownPoseEstimatorNode (name: str, model_config: str,
                                                         model_checkpoint: str,
                                                         input_buffer: str, output_buffer:
                                                         Union[str, List[str]], enable_key:
                                                         Optional[Union[str, int]] = None,
                                                         enable: bool = True, device: str =
                                                         'cuda:0', class_ids:
                                                         Optional[List[int]] = None, labels:
                                                         Optional[List[str]] = None,
                                                         bbox_thr: float = 0.5, smooth:
                                                         bool = False, smooth_filter_cfg:
                                                         str = 'con-
                                                         figs/_base_/filters/one_euro.py')
```

Perform top-down pose estimation using MMPose model.

The node should be placed after an object detection node.

**参数**

- **name** (*str*) –The node name (also thread name)
- **model\_cfg** (*str*) –The model config file
- **model\_checkpoint** (*str*) –The model checkpoint file
- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: `None`
- **enable** (*bool*) –Default enable/disable status. Default: `True`
- **device** (*str*) –Specify the device to hold model weights and inference the model. Default: `'cuda:0'`
- **class\_ids** (*list[int], optional*) –Specify the object category indices to apply pose estimation. If both `class_ids` and `labels` are given, `labels` will be ignored. If neither is given, pose estimation will be applied for all objects. Default: `None`
- **labels** (*list[str], optional*) –Specify the object category names to apply pose estimation. See also `class_ids`. Default: `None`
- **bbox\_thr** (*float*) –Set a threshold to filter out objects with low bbox scores. Default: `0.5`

- **smooth** (*bool*) –If set to True, a Smoother will be used to refine the pose estimation result. Default: True
- **smooth\_filter\_cfg** (*str*) –The filter config path to build the smoother. Only valid when smooth==True. By default, OneEuro filter will be used.

**Example::**

```
>>> cfg = dict(
...     type='TopDownPoseEstimatorNode',
...     name='human pose estimator',
...     model_config='configs/wholebody/2d_kpt_sview_rgb_img/'
...     'topdown_heatmap/coco-wholebody/'
...     'vipnas_mbv3_coco_wholebody_256x192_dark.py',
...     model_checkpoint='https://download.openmmlab.com/mmpose/'
...     'top_down/vipnas/vipnas_mbv3_coco_wholebody_256x192_dark'
...     '-e2158108_20211205.pth',
...     labels=['person'],
...     smooth=True,
...     input_buffer='det_result',
...     output_buffer='human_pose')
```

```
>>> from mmpose.apis.webcam.nodes import NODES
>>> node = NODES.build(cfg)
```

### mmpose.apis.webcam.nodes.PoseTrackerNode

```
class mmpose.apis.webcam.nodes.PoseTrackerNode (name: str, det_model_config: str,
                                                det_model_checkpoint: str, pose_model_config:
                                                str, pose_model_checkpoint: str, input_buffer:
                                                str, output_buffer: Union[str, List[str]],
                                                enable_key: Optional[Union[str, int]] = None,
                                                enable: bool = True, device: str = 'cuda:0',
                                                det_interval: int = 1, class_ids: Optional[List] =
                                                None, labels: Optional[List] = None, bbox_thr:
                                                float = 0.5, kpt2bbox_cfg: Optional[dict] =
                                                None, smooth: bool = False, smooth_filter_cfg:
                                                str = 'configs/_base_/filters/one_euro.py')
```

Perform object detection and top-down pose estimation. Only detect objects every few frames, and use the pose estimation results to track the object at interval.

Note that MMDetection is required for this node. Please refer to [MMDetection documentation](#) for the installation guide.

## 参数

- **name** (*str*) –The node name (also thread name)
- **det\_model\_cfg** (*str*) –The config file of the detection model
- **det\_model\_checkpoint** (*str*) –The checkpoint file of the detection model
- **pose\_model\_cfg** (*str*) –The config file of the pose estimation model
- **pose\_model\_checkpoint** (*str*) –The checkpoint file of the pose estimation model
- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: None
- **enable** (*bool*) –Default enable/disable status. Default: True
- **device** (*str*) –Specify the device to hold model weights and inference the model. Default: 'cuda:0'
- **det\_interval** (*int*) –Set the detection interval in frames. For example, `det_interval==10` means inference the detection model every 10 frames. Default: 1
- **class\_ids** (*list[int], optional*) –Specify the object category indices to apply pose estimation. If both `class_ids` and `labels` are given, `labels` will be ignored. If neither is given, pose estimation will be applied for all objects. Default: None
- **labels** (*list[str], optional*) –Specify the object category names to apply pose estimation. See also `class_ids`. Default: None
- **bbox\_thr** (*float*) –Set a threshold to filter out objects with low bbox scores. Default: 0.5
- **kpt2bbox\_cfg** (*dict, optional*) –Configure the process to get object bbox from its keypoints during tracking. Specifically, the bbox is obtained from the minimal outer rectangle of the keyponits with following configurable arguments: 'scale', the coefficient to expand the keypoint outer rectangle, defaults to 1.5; 'kpt\_thr': a threshold to filter out low-scored keypoint, defaults to 0.3. See `self.default_kpt2bbox_cfg` for details
- **smooth** (*bool*) –If set to True, a Smoother will be used to refine the pose estimation result. Default: True
- **smooth\_filter\_cfg** (*str*) –The filter config path to build the smoother. Only valid when `smooth==True`. Default to use an OneEuro filter

Example::

```

>>> cfg = dict(
...     type='PoseTrackerNode',
...     name='pose_tracker',
...     det_model_config='demo/mmdetection_cfg/'
...     'ssdlite_mobilenetv2_scratch_600e_coco.py',
...     det_model_checkpoint='https://download.openmmlab.com'
...     '/mmdetection/v2.0/ssd/'
...     'ssdlite_mobilenetv2_scratch_600e_coco/ssdlite_mobilenetv2_'
...     'scratch_600e_coco_20210629_110627-974d9307.pth',
...     pose_model_config='configs/wholebody/2d_kpt_sview_rgb_img/'
...     'topdown_heatmap/coco-wholebody/'
...     'vipnas_mbv3_coco_wholebody_256x192_dark.py',
...     pose_model_checkpoint='https://download.openmmlab.com/mmpose/'
...     'top_down/vipnas/vipnas_mbv3_coco_wholebody_256x192_dark'
...     '-e2158108_20211205.pth',
...     det_interval=10,
...     labels=['person'],
...     smooth=True,
...     device='cuda:0',
...     # `_input` is an executor-reserved buffer
...     input_buffer='_input_',
...     output_buffer='human_pose')

```

```

>>> from mmpose.apis.webcam.nodes import NODES
>>> node = NODES.build(cfg)

```

### 46.2.3 Visualizer Nodes

<i>ObjectVisualizerNode</i>	Visualize the bounding box and keypoints of objects.
<i>NoticeBoardNode</i>	Show text messages in the frame.
<i>SunglassesEffectNode</i>	Apply sunglasses effect (draw sunglasses at the facial area) to the objects with eye keypoints in the frame.
<i>BigeyeEffectNode</i>	Apply big-eye effect to the objects with eye keypoints in the frame.

**mmpose.apis.webcam.nodes.ObjectVisualizerNode**

```
class mmpose.apis.webcam.nodes.ObjectVisualizerNode (name: str, input_buffer: str,
                                                    output_buffer: Union[str, List[str]],
                                                    enable_key: Optional[Union[str, int]] =
None, enable: bool = True, show_bbox:
bool = True, show_keypoint: bool =
True, must_have_keypoint: bool =
False, kpt_thr: float = 0.3, radius: int =
4, thickness: int = 2, bbox_color:
Optional[Union[str, Tuple, Dict]] =
None)
```

Visualize the bounding box and keypoints of objects.

**参数**

- **name** (*str*) –The node name (also thread name)
- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: `None`
- **enable** (*bool*) –Default enable/disable status. Default: `True`
- **show\_bbox** (*bool*) –Set `True` to show the bboxes of detection objects. Default: `True`
- **show\_keypoint** (*bool*) –Set `True` to show the pose estimation results. Default: `True`
- **must\_have\_bbox** (*bool*) –Only show objects with keypoints. Default: `False`
- **kpt\_thr** (*float*) –The threshold of keypoint score. Default: `0.3`
- **radius** (*int*) –The radius of keypoint. Default: `4`
- **thickness** (*int*) –The thickness of skeleton. Default: `2`
- **bbox\_color** (*str/tuple/dict*) –The color of bboxes. If a single color is given (a str like ‘green’ or a BGR tuple like (0, 255, 0)), it will be used for all bboxes. If a dict is given, it will be used as a map from class labels to bbox colors. If not given, a default color map will be used. Default: `None`

**Example::**

```
>>> cfg = dict(
...     type='ObjectVisualizerNode',
...     name='object visualizer',
...     enable_key='v',
...     enable=True,
...     show_bbox=True,
...     must_have_keypoint=False,
...     show_keypoint=True,
...     input_buffer='frame',
...     output_buffer='vis')
```

```
>>> from mmpose.apis.webcam.nodes import NODES
>>> node = NODES.build(cfg)
```

### mmpose.apis.webcam.nodes.NoticeBoardNode

```
class mmpose.apis.webcam.nodes.NoticeBoardNode (name: str, input_buffer: str, output_buffer:
                                                Union[str, List[str]], enable_key:
                                                Optional[Union[str, int]] = None, enable: bool
                                                = True, content_lines: Optional[List[str]] =
                                                None, x_offset: int = 20, y_offset: int = 20,
                                                y_delta: int = 15, text_color: Union[str,
                                                Tuple[int, int, int]] = 'black', background_color:
                                                Union[str, Tuple[int, int, int]] = (255, 183, 0),
                                                text_scale: float = 0.4)
```

Show text messages in the frame.

#### 参数

- **name** (*str*) –The node name (also thread name)
- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: None
- **enable** (*bool*) –Default enable/disable status. Default: True
- **content\_lines** (*list[str], optional*) –The lines of text message to show in the frame. If not given, a default message will be shown. Default: None



- **x\_offset** (*int*) –The position of the notice board' s left border in pixels. Default: 20
- **y\_offset** (*int*) –The position of the notice board' s top border in pixels. Default: 20
- **y\_delta** (*int*) –The line height in pixels. Default: 15
- **text\_color** (*str/tuple*) –The font color represented in a color name or a BGR tuple. Default: 'black'
- **backbround\_color** (*str/tuple*) –The background color represented in a color name or a BGR tuple. Default: (255, 183, 0)
- **text\_scale** (*float*) –The font scale factor that is multiplied by the base size. Default: 0.4

**Example::**

```
>>> cfg = dict(
...     type='NoticeBoardNode',
...     name='instruction',
...     enable_key='h',
...     enable=True,
...     input_buffer='vis_bigeye',
...     output_buffer='vis_notice',
...     content_lines=[
...         'This is a demo for pose visualization and simple image '
...         'effects. Have fun!', '', 'Hot-keys:',
...         '"v": Pose estimation result visualization',
...         '"s": Sunglasses effect B-)', '"b": Big-eye effect 0_0',
...         '"h": Show help information',
...         '"m": Show diagnostic information', '"q": Exit'
...     ],
... )
```

```
>>> from mmpose.apis.webcam.nodes import NODES
>>> node = NODES.build(cfg)
```

**mmpose.apis.webcam.nodes.SunglassesEffectNode**

```
class mmpose.apis.webcam.nodes.SunglassesEffectNode (name: str, input_buffer: str,
                                                    output_buffer: Union[str, List[str]],
                                                    enable_key: Optional[Union[str, int]] =
None, enable: bool = True, kpt_thr:
float = 0.5, resource_img_path:
Optional[str] = None)
```

Apply sunglasses effect (draw sunglasses at the facial area)to the objects with eye keypoints in the frame.

### 参数

- **name** (*str*) –The node name (also thread name)
  - **input\_buffer** (*str*) –The name of the input buffer
  - **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
  - **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note:
    1. If **enable\_key** is set, the **bypass method need to be** overridden to define the node behavior when disabled
    2. Some hot-key has been use for particular use. For example: 'q', 'Q' and 27 are used for quit
- Default: None
- **enable** (*bool*) –Default enable/disable status. Default: True.
  - **kpt\_thr** (*float*) –The score threshold of valid keypoints. Default: 0.5
  - **resource\_img\_path** (*str, optional*) –The resource image path or url. The image should be a pair of sunglasses with white background. If not specified, the url of a default image will be used. See `SunglassesNode.default_resource_img_path`. Default: None

### Example::

```
>>> cfg = dict(  
...     type='SunglassesEffectNode',  
...     name='sunglasses',  
...     enable_key='s',  
...     enable=False,  
...     input_buffer='vis',  
...     output_buffer='vis_sunglasses')
```

```
>>> from mmpose.apis.webcam.nodes import NODES  
>>> node = NODES.build(cfg)
```

**mmpose.apis.webcam.nodes.BigeyeEffectNode**

```
class mmpose.apis.webcam.nodes.BigeyeEffectNode (name: str, input_buffer: str, output_buffer: Union[str, List[str]], enable_key: Optional[Union[str, int]] = None, enable: bool = True, kpt_thr: float = 0.5)
```

Apply big-eye effect to the objects with eye keypoints in the frame.

**参数**

- **name** (*str*) –The node name (also thread name)
- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: `None`
- **enable** (*bool*) –Default enable/disable status. Default: `True`
- **kpt\_thr** (*float*) –The score threshold of valid keypoints. Default: `0.5`

**Example::**

```
>>> cfg = dict(
...     type='SunglassesEffectNode',
...     name='sunglasses',
...     enable_key='s',
...     enable=False,
...     input_buffer='vis',
...     output_buffer='vis_sunglasses')
```

```
>>> from mmpose.apis.webcam.nodes import NODES
>>> node = NODES.build(cfg)
```

## 46.2.4 Helper Nodes

<i>ObjectAssignerNode</i>	Assign the object information to the frame message.
<i>MonitorNode</i>	Show diagnostic information.
<i>RecorderNode</i>	Record the video frames into a local file.

### mmpose.apis.webcam.nodes.ObjectAssignerNode

**class** mmpose.apis.webcam.nodes.**ObjectAssignerNode** (*name: str, frame\_buffer: str, object\_buffer: str, output\_buffer: Union[str, List[str]]*)

Assign the object information to the frame message.

*ObjectAssignerNode* enables asynchronous processing of model inference and video I/O, so the video will be captured and displayed smoothly regardless of the model inference speed. Specifically, *ObjectAssignerNode* takes messages from both model branch and video I/O branch as its input, indicated as “object message” and “frame message” respectively. When an object message arrives it will update the latest object information; and when a frame message arrives, it will be assigned with the latest object information and output.

Specially, if the webcam executor is set to synchronous mode, the behavior of *ObjectAssignerNode* will be different: When an object message arrives, it will trigger an output of itself; and the frame messages will be ignored.

#### 参数

- **name** (*str*) –The node name (also thread name)
- **frame\_buffer** (*str*) –Buffer name for frame messages
- **object\_buffer** (*str*) –Buffer name for object messages
- **output\_buffer** (*str*) –The name(s) of the output buffer(s)

#### Example::

```
>>> cfg =dict(
...     type='ObjectAssignerNode',
...     name='object assigner',
...     frame_buffer='_frame_',
...     # `_frame_` is an executor-reserved buffer
...     object_buffer='animal_pose',
...     output_buffer='frame')
```

```
>>> from mmpose.apis.webcam.nodes import NODES
>>> node = NODES.build(cfg)
```

**mmpose.apis.webcam.nodes.MonitorNode**

```
class mmpose.apis.webcam.nodes.MonitorNode (name: str, input_buffer: str, output_buffer: Union[str,
                                             List[str]], enable_key: Optional[Union[str, int]] =
                                             None, enable: bool = False, x_offset=20, y_offset=20,
                                             y_delta=15, text_color='black',
                                             background_color=(255, 183, 0), text_scale=0.4,
                                             ignore_items: Optional[List[str]] = None)
```

Show diagnostic information.

**参数**

- **name** (*str*) –The node name (also thread name)
- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
- **enable\_key** (*str/int, optional*) –Set a hot-key to toggle enable/disable of the node. If an int value is given, it will be treated as an ascii code of a key. Please note: (1) If `enable_key` is set, the `bypass()` method need to be overridden to define the node behavior when disabled; (2) Some hot-keys are reserved for particular use. For example: ‘q’, ‘Q’ and 27 are used for exiting. Default: `None`
- **enable** (*bool*) –Default enable/disable status. Default: `True`
- **x\_offset** (*int*) –The position of the text box’s left border in pixels. Default: 20
- **y\_offset** (*int*) –The position of the text box’s top border in pixels. Default: 20
- **y\_delta** (*int*) –The line height in pixels. Default: 15
- **text\_color** (*str/tuple*) –The font color represented in a color name or a BGR tuple. Default: ‘black’
- **backbround\_color** (*str/tuple*) –The background color represented in a color name or a BGR tuple. Default: (255, 183, 0)
- **text\_scale** (*float*) –The font scale factor that is multiplied by the base size. Default: 0.4
- **ignore\_items** (*list[str], optional*) –Specify the node information items that will not be shown. See `MonitorNode._default_ignore_items` for the default setting.

**Example::**

```
>>> cfg = dict(
...     type='MonitorNode',
...     name='monitor',
```

(下页继续)

(续上页)

```
...     enable_key='m',
...     enable=False,
...     input_buffer='vis_notice',
...     output_buffer='display')
```

```
>>> from mmpose.apis.webcam.nodes import NODES
>>> node = NODES.build(cfg)
```

## mmpose.apis.webcam.nodes.RecorderNode

```
class mmpose.apis.webcam.nodes.RecorderNode (name: str, input_buffer: str, output_buffer:
                                             Union[str, List[str]], out_video_file: str,
                                             out_video_fps: int = 30, out_video_codec: str =
                                             'mp4v', buffer_size: int = 30, enable: bool = True)
```

Record the video frames into a local file.

*RecorderNode* uses OpenCV backend to record the video. Recording is performed in a separate thread to avoid blocking the data stream. A buffer queue is used to cached the arrived frame images.

### 参数

- **name** (*str*) –The node name (also thread name)
- **input\_buffer** (*str*) –The name of the input buffer
- **output\_buffer** (*str/list*) –The name(s) of the output buffer(s)
- **out\_video\_file** (*str*) –The path of the output video file
- **out\_video\_fps** (*int*) –The frame rate of the output video. Default: 30
- **out\_video\_codec** (*str*) –The codec of the output video. Default: 'mp4v'
- **buffer\_size** (*int*) –Size of the buffer queue that caches the arrived frame images.
- **enable** (*bool*) –Default enable/disable status. Default: True.

### Example::

```
>>> cfg = dict(
...     type='RecorderNode',
...     name='recorder',
...     out_video_file='webcam_demo.mp4',
...     input_buffer='display',
...     output_buffer='_display_'
...     # `_display_` is an executor-reserved buffer
... )
```

```
>>> from mmpose.apis.webcam.nodes import NODES
>>> node = NODES.build(cfg)
```

## 46.3 Utils

### 46.3.1 Buffer and Message

<i>BufferManager</i>	A helper class to manage multiple buffers.
<i>Message</i>	Message base class.
<i>FrameMessage</i>	The message to store information of a video frame.
<i>VideoEndingMessage</i>	The special message to indicate the ending of the input video.

#### mmpose.apis.webcam.utils.BufferManager

**class** mmpose.apis.webcam.utils.**BufferManager** (*buffer\_type*: type = <class 'mmpose.apis.webcam.utils.buffer.Buffer'>, *buffers*: Optional[Dict] = None)

A helper class to manage multiple buffers.

参数

- **buffer\_type** (*type*) –The class to build buffer instances. Default: mmpose.apis.webcam.utils.buffer.Buffer.
- **buffers** (*dict*, *optional*) –Create *BufferManager* from existing buffers. Each item should a buffer name and the buffer. If not given, an empty buffer manager will be create. Default: None

**\_\_init\_\_** (*buffer\_type*: type = <class 'mmpose.apis.webcam.utils.buffer.Buffer'>, *buffers*: Optional[Dict] = None)

Initialize self. See help(type(self)) for accurate signature.

## Methods

<code>__init__([buffer_type, buffers])</code>	Initialize self.
<code>get(name[, block, timeout])</code>	Remove an return an item from the specified buffer.
<code>get_info()</code>	Returns the information of all buffers in the manager.
<code>get_sub_manager(buffer_names)</code>	Return a <i>BufferManager</i> instance that covers a subset of the buffers in the parent.
<code>is_empty(name)</code>	Check if a buffer is empty.
<code>is_full(name)</code>	Check if a buffer is full.
<code>put(name, item[, block, timeout])</code>	Put an item into specified buffer.
<code>put_force(name, item)</code>	Force to put an item into specified buffer.
<code>register_buffer(name[, maxsize])</code>	Register a buffer.

## mmpose.apis.webcam.utils.Message

**class** mmpose.apis.webcam.utils.**Message** (*msg: str = "", data: Optional[Dict] = None*)

Message base class.

All message class should inherit this class. The basic use of a Message instance is to carry a piece of text message (`self.msg`) and a dict that stores structured data (`self.data`), e.g. frame image, model prediction, et al.

A message may also hold route information, which is composed of information of all nodes the message has passed through.

### 参数

- **msg** (*str*) –The text message.
- **data** (*dict, optional*) –The structured data.

**\_\_init\_\_** (*msg: str = "", data: Optional[Dict] = None*)

Initialize self. See `help(type(self))` for accurate signature.

## Methods

<code>__init__(msg, data)</code>	Initialize self.
<code>get_route_info()</code>	
<code>merge_route_info(route_info)</code>	Merge the given route information into the original one of the message.
<code>set_route_info(route_info)</code>	Directly set the entire route information.

下页继续



表 9 - 续上页

<code>update_route_info([node, node_name, ...])</code>	Append new node information to the route information.
--	---

**mmpose.apis.webcam.utils.FrameMessage**

**class** mmpose.apis.webcam.utils.**FrameMessage** (*img*)

The message to store information of a video frame.

`__init__` (*img*)

Initialize self. See help(type(self)) for accurate signature.

**Methods**

<code>__init__(img)</code>	Initialize self.
<code>get_image()</code>	Get the frame image.
<code>get_objects([obj_filter])</code>	Get object information from the frame data.
<code>get_route_info()</code>	
<code>merge_route_info(route_info)</code>	Merge the given route information into the original one of the message.
<code>set_image(img)</code>	Set the frame image to the message.
<code>set_objects(objects)</code>	Set the object information.
<code>set_route_info(route_info)</code>	Directly set the entire route information.
<code>update_objects(objects)</code>	Update object information.
<code>update_route_info([node, node_name, ...])</code>	Append new node information to the route information.

**mmpose.apis.webcam.utils.VideoEndingMessage**

**class** mmpose.apis.webcam.utils.**VideoEndingMessage** (*msg: str = "*, *data: Optional[Dict] = None*)

The special message to indicate the ending of the input video.

`__init__` (*msg: str = "*, *data: Optional[Dict] = None*)

Initialize self. See help(type(self)) for accurate signature.

## Methods

<code>__init__([msg, data])</code>	Initialize self.
<code>get_route_info()</code>	
<code>merge_route_info(route_info)</code>	Merge the given route information into the original one of the message.
<code>set_route_info(route_info)</code>	Directly set the entire route information.
<code>update_route_info([node, node_name, ...])</code>	Append new node information to the route information.

## 46.3.2 Pose

<code>get_eye_keypoint_ids</code>	A helper function to get the keypoint indices of left and right eyes from the model config.
<code>get_face_keypoint_ids</code>	A helper function to get the keypoint indices of the face from the model config.
<code>get_hand_keypoint_ids</code>	A helper function to get the keypoint indices of left and right hand from the model config.
<code>get_mouth_keypoint_ids</code>	A helper function to get the mouth keypoint index from the model config.
<code>get_wrist_keypoint_ids</code>	A helper function to get the keypoint indices of left and right wrists from the model config.

### mmpose.apis.webcam.utils.get\_eye\_keypoint\_ids

`mmpose.apis.webcam.utils.get_eye_keypoint_ids(model_cfg: mmcv.utils.config.Config) → Tuple[int, int]`

A helper function to get the keypoint indices of left and right eyes from the model config.

**参数** `model_cfg` (*Config*) –mmpose model config

**返回** The keypoint indices of left eye and right eye.

**返回类型** `tuple[int, int]`

### **mmpose.apis.webcam.utils.get\_face\_keypoint\_ids**

`mmpose.apis.webcam.utils.get_face_keypoint_ids(model_cfg: mmpcv.utils.config.Config) → List`

A helper function to get the keypoint indices of the face from the model config.

**参数** `model_cfg` (*Config*) –pose model config.

**返回** face keypoint indices. The length depends on the dataset.

**返回类型** list[int]

### **mmpose.apis.webcam.utils.get\_hand\_keypoint\_ids**

`mmpose.apis.webcam.utils.get_hand_keypoint_ids(model_cfg: mmpcv.utils.config.Config) → List[int]`

A helper function to get the keypoint indices of left and right hand from the model config.

**参数** `model_cfg` (*Config*) –pose model config.

**返回** hand keypoint indices. The length depends on the dataset.

**返回类型** list[int]

### **mmpose.apis.webcam.utils.get\_mouth\_keypoint\_ids**

`mmpose.apis.webcam.utils.get_mouth_keypoint_ids(model_cfg: mmpcv.utils.config.Config) → int`

A helper function to get the mouth keypoint index from the model config.

**参数** `model_cfg` (*Config*) –pose model config.

**返回** The mouth keypoint index

**返回类型** int

### **mmpose.apis.webcam.utils.get\_wrist\_keypoint\_ids**

`mmpose.apis.webcam.utils.get_wrist_keypoint_ids(model_cfg: mmpcv.utils.config.Config) → Tuple[int, int]`

A helper function to get the keypoint indices of left and right wrists from the model config.

**参数** `model_cfg` (*Config*) –pose model config.

**返回** The keypoint indices of left and right wrists.

**返回类型** tuple[int, int]

### 46.3.3 Event

---

*EventManager*

A helper class to manage events.

---

#### mmpose.apis.webcam.utils.EventManager

**class** mmpose.apis.webcam.utils.**EventManager**

A helper class to manage events.

*EventManager* provides interfaces to register, set, clear and check events by name.

**\_\_init\_\_()**

Initialize self. See help(type(self)) for accurate signature.

#### Methods

<b>__init__()</b>	Initialize self.
<b>clear([event_name, is_keyboard])</b>	Reset the internal flag of an event to False.
<b>is_set([event_name, is_keyboard])</b>	Check whether the internal flag of an event is True.
<b>register_event(event_name[, is_keyboard])</b>	Register an event.
<b>set(event_name[, is_keyboard])</b>	Set the internal flag of an event to True.
<b>wait([event_name, is_keyboard, timeout])</b>	Block until the internal flag of an event is True.
<b>wait_and_handle([event_name, is_keyboard])</b>	Context manager that blocks until an event is set True and then goes into the context.

### 46.3.4 Misc

<i>copy_and_paste</i>	Copy the image region and paste to the background.
<i>screen_matting</i>	Get screen matting mask.
<i>expand_and_clamp</i>	Expand the bbox and clip it to fit the image shape.
<i>limit_max_fps</i>	A context manager to limit maximum frequency of entering the context.
<i>is_image_file</i>	Check if a path is an image file by its extension.
<i>get_cached_file_path</i>	Loads the Torch serialized object at the given URL.
<i>load_image_from_disk_or_url</i>	Load an image file, from disk or url.
<i>get_config_path</i>	Get config path from an OpenMMLab codebase.

### mmpose.apis.webcam.utils.copy\_and\_paste

`mmpose.apis.webcam.utils.copy_and_paste` (*img: numpy.ndarray, background\_img: numpy.ndarray, mask: numpy.ndarray, bbox: Optional[List] = None, effect\_region: Tuple = (0.2, 0.2, 0.8, 0.8), min\_size: Tuple = (20, 20)*) → `numpy.ndarray`

Copy the image region and paste to the background.

#### 参数

- **img** (`np.ndarray`) –Image data.
- **background\_img** (`np.ndarray`) –Background image data.
- **mask** (`ndarray`) –instance segmentation result.
- **bbox** (`list, optional`) –instance bbox in (x1, y1, x2, y2). If not given, the bbox will be obtained by `_find_bbox()`. Default: `None`
- **effect\_region** (`tuple`) –The region to apply mask, the coordinates are normalized (x1, y1, x2, y2). Default: (0.2, 0.2, 0.8, 0.8)
- **min\_size** (`tuple`) –The minimum region size (w, h) in pixels. Default: (20, 20)

返回 The background with pasted image region.

返回类型 `np.ndarray`

### mmpose.apis.webcam.utils.screen\_matting

`mmpose.apis.webcam.utils.screen_matting` (*img: numpy.ndarray, color\_low: Optional[Tuple] = None, color\_high: Optional[Tuple] = None, color: Optional[str] = None*) → `numpy.ndarray`

Get screen matting mask.

#### 参数

- **img** (`np.ndarray`) –Image data.
- **color\_low** (`tuple`) –Lower limit (b, g, r).
- **color\_high** (`tuple`) –Higher limit (b, g, r).
- **color** (`str`) –Support colors include:
  - 'green' or 'g'
  - 'blue' or 'b'
  - 'black' or 'k'
  - 'white' or 'w'

返回 A mask with the same shape of the input image. The value is 0 at the pixels in the matting color range, and 1 everywhere else.

返回类型 `np.ndarray`

### `mmpose.apis.webcam.utils.expand_and_clamp`

`mmpose.apis.webcam.utils.expand_and_clamp` (*box*: List, *im\_shape*: Tuple, *scale*: float = 1.25) → List  
Expand the bbox and clip it to fit the image shape.

参数

- **box** (*list*) –x1, y1, x2, y2
- **im\_shape** (*tuple*) –image shape (h, w, c)
- **scale** (*float*) –expand ratio

返回 x1, y1, x2, y2

返回类型 list

### `mmpose.apis.webcam.utils.limit_max_fps`

`mmpose.apis.webcam.utils.limit_max_fps` (*fps*: float)  
A context manager to limit maximum frequency of entering the context.

参数 **fps** (*float*) –The maximum frequency of entering the context

Example::

```
>>> from mmpose.apis.webcam.utils import limit_max_fps
>>> import cv2
```

```
>>> while True:
...     with limit_max_fps(20):
...         cv2.imshow(img)  # display image at most 20 fps
```

### `mmpose.apis.webcam.utils.is_image_file`

`mmpose.apis.webcam.utils.is_image_file` (*path*: str) → bool  
Check if a path is an image file by its extension.

参数 **path** (*str*) –The image path.

返回 Whether the path is an image file.

返回类型 bool

**mmpose.apis.webcam.utils.get\_cached\_file\_path**

`mmpose.apis.webcam.utils.get_cached_file_path(url: str, save_dir: Optional[str] = None, progress: bool = True, check_hash: bool = False, file_name: Optional[str] = None) → str`

Loads the Torch serialized object at the given URL.

If downloaded file is a zip file, it will be automatically decompressed

If the object is already present in `model_dir`, it's deserialized and returned. The default value of `model_dir` is `<hub_dir>/checkpoints` where `hub_dir` is the directory returned by `get_dir()`.

**参数**

- **url** (*str*) –URL of the object to download
- **save\_dir** (*str, optional*) –directory in which to save the object
- **progress** (*bool*) –whether or not to display a progress bar to stderr. Default: True
- **check\_hash** (*bool*) –If True, the filename part of the URL should follow the naming convention `filename-<sha256>.ext` where `<sha256>` is the first eight or more digits of the SHA256 hash of the contents of the file. The hash is used to ensure unique names and to verify the contents of the file. Default: False
- **file\_name** (*str, optional*) –name for the downloaded file. Filename from `url` will be used if not set. Default: None.

**返回** The path to the cached file.

**返回类型** str

**mmpose.apis.webcam.utils.load\_image\_from\_disk\_or\_url**

`mmpose.apis.webcam.utils.load_image_from_disk_or_url(filename: str, readFlag: int = 1) → numpy.ndarray`

Load an image file, from disk or url.

**参数**

- **filename** (*str*) –file name on the disk or url link
- **readFlag** (*int*) –readFlag for imdecode. Default: cv2.IMREAD\_COLOR

**返回** A loaded image

**返回类型** np.ndarray

**mmpose.apis.webcam.utils.get\_config\_path**

`mmpose.apis.webcam.utils.get_config_path(path: str, module_name: str)`

Get config path from an OpenMMLab codebase.

If the path is an existing file, it will be directly returned. If the file doesn't exist, it will be searched in the 'configs' folder of the specified module.

**参数**

- **path** (*str*) –the path of the config file
- **module\_name** (*str*) –The module name of an OpenMMLab codebase

**返回** The config file path.

**返回类型** str

**Example::**

```
>>> path = 'configs/_base_/filters/one_euro.py'
>>> get_config_path(path, 'mmpose')
'/home/mmpose/configs/_base_/filters/one_euro.py'
```



## 47.1 evaluation

```
class mmpose.core.evaluation.DistEvalHook (dataloader, start=None, interval=1, by_epoch=True,
                                             save_best=None, rule=None, test_fn=None,
                                             greater_keys=['acc', 'ap', 'ar', 'pck', 'auc', '3dpck',
                                                           'p-3dpck', '3dauc', 'p-3dauc', 'pcp'], less_keys=['loss',
                                                           'epe', 'nme', 'mpjpe', 'p-mpjpe', 'n-mpjpe'],
                                             broadcast_bn_buffer=True, tmpdir=None,
                                             gpu_collect=False, **eval_kwargs)
```

```
class mmpose.core.evaluation.EvalHook (dataloader, start=None, interval=1, by_epoch=True,
                                          save_best=None, rule=None, test_fn=None,
                                          greater_keys=['acc', 'ap', 'ar', 'pck', 'auc', '3dpck', 'p-3dpck',
                                                        '3dauc', 'p-3dauc', 'pcp'], less_keys=['loss', 'epe', 'nme',
                                                        'mpjpe', 'p-mpjpe', 'n-mpjpe'], **eval_kwargs)
```

```
mmpose.core.evaluation.aggregate_scale (feature_maps_list, align_corners=False,
                                           project2image=True, size_projected=None,
                                           aggregate_scale='average')
```

Aggregate multi-scale outputs.

---

**注解:** batch size: N keypoints num : K heatmap width: W heatmap height: H

---

**参数**

- **feature\_maps\_list** (*list[Tensor]*) –Aggregated feature maps.
- **project2image** (*bool*) –Option to resize to base scale.
- **size\_projected** (*list[int, int]*) –Base size of heatmaps [w, h].
- **align\_corners** (*bool*) –Align corners when performing interpolation.
- **aggregate\_scale** (*str*) –Methods to aggregate multi-scale feature maps. Options: ‘average’ , ‘unsqueeze\_concat’ .
  - ‘average’ : Get the average of the feature maps.
  - ‘unsqueeze\_concat’ : **Concatenate the feature maps along new axis.** Default: ‘average’.

**返回** Aggregated feature maps.

**返回类型** Tensor

```
mmpose.core.evaluation.aggregate_stage_flip(feature_maps, feature_maps_flip, index=-1,
                                             project2image=True, size_projected=None,
                                             align_corners=False, aggregate_stage='concat',
                                             aggregate_flip='average')
```

Inference the model to get multi-stage outputs (heatmaps & tags), and resize them to base sizes.

**参数**

- **feature\_maps** (*list[Tensor]*) –feature\_maps can be heatmaps, tags, and pafs.
- **feature\_maps\_flip** (*list[Tensor] | None*) –flipped feature\_maps. feature maps can be heatmaps, tags, and pafs.
- **project2image** (*bool*) –Option to resize to base scale.
- **size\_projected** (*list[int, int]*) –Base size of heatmaps [w, h].
- **align\_corners** (*bool*) –Align corners when performing interpolation.
- **aggregate\_stage** (*str*) –Methods to aggregate multi-stage feature maps. Options: ‘concat’ , ‘average’ . Default: ‘concat’.
  - ‘concat’ : Concatenate the original and the flipped feature maps.
  - ‘average’ : **Get the average of the original and the flipped** feature maps.
- **aggregate\_flip** (*str*) –Methods to aggregate the original and the flipped feature maps. Options: ‘concat’ , ‘average’ , ‘none’ . Default: ‘average’.
  - ‘concat’ : Concatenate the original and the flipped feature maps.
  - ‘average’ : **Get the average of the original and the flipped** feature maps..

– 'none' : no flipped feature maps.

返回 Aggregated feature maps with shape [NxKxWxH].

返回类型 list[*Tensor*]

`mmpose.core.evaluation.compute_similarity_transform(source_points, target_points)`

Computes a similarity transform (sR, t) that takes a set of 3D points `source_points` (N x 3) closest to a set of 3D points `target_points`, where R is an 3x3 rotation matrix, t 3x1 translation, s scale. And return the transformed 3D points `source_points_hat` (N x 3). i.e. solves the orthogonal Procrustes problem.

---

注解: Points number: N

---

参数

- **source\_points** (*np.ndarray*) –Source point set with shape [N, 3].
- **target\_points** (*np.ndarray*) –Target point set with shape [N, 3].

返回 Transformed source point set with shape [N, 3].

返回类型 *np.ndarray*

`mmpose.core.evaluation.flip_feature_maps(feature_maps, flip_index=None)`

Flip the feature maps and swap the channels.

参数

- **feature\_maps** (*list[*Tensor*]*) –Feature maps.
- **flip\_index** (*list[int] | None*) –Channel-flip indexes. If None, do not flip channels.

返回 Flipped feature\_maps.

返回类型 list[*Tensor*]

`mmpose.core.evaluation.get_group_preds(grouped_joints, center, scale, heatmap_size, use_udp=False)`

Transform the grouped joints back to the image.

参数

- **grouped\_joints** (*list*) –Grouped person joints.
- **center** (*np.ndarray[2, ]*) –Center of the bounding box (x, y).
- **scale** (*np.ndarray[2, ]*) –Scale of the bounding box wrt [width, height].
- **heatmap\_size** (*np.ndarray[2, ]*) –Size of the destination heatmaps.
- **use\_udp** (*bool*) –Unbiased data processing. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR' 2020).

返回 List of the pose result for each person.

返回类型 list

`mmpose.core.evaluation.keypoint_3d_auc(pred, gt, mask, alignment='none')`

Calculate the Area Under the Curve (3DAUC) computed for a range of 3DPCK thresholds.

Paper ref: [Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision](#) 3DV' 2017. .  
This implementation is derived from `mpii_compute_3d_pck.m`, which is provided as part of the MPI-INF-3DHP test data release.

---

注解: batch\_size: N num\_keypoints: K keypoint\_dims: C

---

### 参数

- **pred** (`np.ndarray[N, K, C]`) –Predicted keypoint location.
- **gt** (`np.ndarray[N, K, C]`) –Groundtruth keypoint location.
- **mask** (`np.ndarray[N, K]`) –Visibility of the target. False for invisible joints, and True for visible. Invisible joints will be ignored for accuracy calculation.
- **alignment** (`str, optional`) –method to align the prediction with the groundtruth. Supported options are:
  - 'none': no alignment will be applied
  - 'scale': align in the least-square sense in scale
  - '**procrustes**': align in the least-square sense in scale, rotation and translation.

返回 AUC computed for a range of 3DPCK thresholds.

返回类型 auc

`mmpose.core.evaluation.keypoint_3d_pck(pred, gt, mask, alignment='none', threshold=0.15)`

Calculate the Percentage of Correct Keypoints (3DPCK) w. or w/o rigid alignment.

Paper ref: [Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision](#) 3DV' 2017. .

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - keypoint\_dims: C
-

### 参数

- **pred** (*np.ndarray*[*N*, *K*, *C*]) –Predicted keypoint location.
- **gt** (*np.ndarray*[*N*, *K*, *C*]) –Groundtruth keypoint location.
- **mask** (*np.ndarray*[*N*, *K*]) –Visibility of the target. False for invisible joints, and True for visible. Invisible joints will be ignored for accuracy calculation.
- **alignment** (*str*, *optional*) –method to align the prediction with the groundtruth. Supported options are:
  - 'none': no alignment will be applied
  - 'scale': align in the least-square sense in scale
  - '**procrustes**': align in the least-square sense in scale, rotation and translation.
- **threshold** –If L2 distance between the prediction and the groundtruth is less then threshold, the predicted result is considered as correct. Default: 0.15 (m).

返回 percentage of correct keypoints.

返回类型 pck

`mmpose.core.evaluation.keypoint_auc(pred, gt, mask, normalize, num_step=20)`

Calculate the pose accuracy of PCK for each individual keypoint and the averaged accuracy across all keypoints for coordinates.

### 注解:

- batch\_size: N
- num\_keypoints: K

### 参数

- **pred** (*np.ndarray*[*N*, *K*, 2]) –Predicted keypoint location.
- **gt** (*np.ndarray*[*N*, *K*, 2]) –Groundtruth keypoint location.
- **mask** (*np.ndarray*[*N*, *K*]) –Visibility of the target. False for invisible joints, and True for visible. Invisible joints will be ignored for accuracy calculation.
- **normalize** (*float*) –Normalization factor.

返回 Area under curve.

返回类型 float

`mmpose.core.evaluation.keypoint_epe(pred, gt, mask)`

Calculate the end-point error.

---

注解:

- `batch_size`: N
  - `num_keypoints`: K
- 

#### 参数

- **`pred`** (`np.ndarray[N, K, 2]`) –Predicted keypoint location.
- **`gt`** (`np.ndarray[N, K, 2]`) –Groundtruth keypoint location.
- **`mask`** (`np.ndarray[N, K]`) –Visibility of the target. False for invisible joints, and True for visible. Invisible joints will be ignored for accuracy calculation.

返回 Average end-point error.

返回类型 float

`mmpose.core.evaluation.keypoint_mpjpe(pred, gt, mask, alignment='none')`

Calculate the mean per-joint position error (MPJPE) and the error after rigid alignment with the ground truth (P-MPJPE).

---

注解:

- `batch_size`: N
  - `num_keypoints`: K
  - `keypoint_dims`: C
- 

#### 参数

- **`pred`** (`np.ndarray`) –Predicted keypoint location with shape [N, K, C].
- **`gt`** (`np.ndarray`) –Groundtruth keypoint location with shape [N, K, C].
- **`mask`** (`np.ndarray`) –Visibility of the target with shape [N, K]. False for invisible joints, and True for visible. Invisible joints will be ignored for accuracy calculation.
- **`alignment`** (`str, optional`) –method to align the prediction with the groundtruth. Supported options are:
  - `'none'`: no alignment will be applied
  - `'scale'`: align in the least-square sense in scale
  - **`'procrustes'`: align in the least-square sense in** scale, rotation and translation.

## 返回

A tuple containing joint position errors

- (float | np.ndarray): mean per-joint position error (mpjpe).
- (float | np.ndarray): mpjpe after rigid alignment with the ground truth (p-mpjpe).

返回类型 tuple

`mmpose.core.evaluation.keypoint_pck_accuracy(pred, gt, mask, thr, normalize)`

Calculate the pose accuracy of PCK for each individual keypoint and the averaged accuracy across all keypoints for coordinates.

**注解:** PCK metric measures accuracy of the localization of the body joints. The distances between predicted positions and the ground-truth ones are typically normalized by the bounding box size. The threshold (thr) of the normalized distance is commonly set as 0.05, 0.1 or 0.2 etc.

- batch\_size: N
- num\_keypoints: K

## 参数

- **pred** (`np.ndarray[N, K, 2]`) – Predicted keypoint location.
- **gt** (`np.ndarray[N, K, 2]`) – Groundtruth keypoint location.
- **mask** (`np.ndarray[N, K]`) – Visibility of the target. False for invisible joints, and True for visible. Invisible joints will be ignored for accuracy calculation.
- **thr** (`float`) – Threshold of PCK calculation.
- **normalize** (`np.ndarray[N, 2]`) – Normalization factor for H&W.

## 返回

A tuple containing keypoint accuracy.

- acc (`np.ndarray[K]`): Accuracy of each keypoint.
- avg\_acc (`float`): Averaged accuracy across all keypoints.
- cnt (`int`): Number of valid keypoints.

返回类型 tuple

```
mmpose.core.evaluation.keypoints_from_heatmaps(heatmaps, center, scale, unbiased=False,
                                                post_process='default', kernel=11,
                                                valid_radius_factor=0.0546875,
                                                use_udp=False,
                                                target_type='GaussianHeatmap')
```

Get final keypoint predictions from heatmaps and transform them back to the image.

---

#### 注解:

- batch size: N
  - num keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

#### 参数

- **heatmaps** (*np.ndarray*[N, K, H, W]) –model predicted heatmaps.
- **center** (*np.ndarray*[N, 2]) –Center of the bounding box (x, y).
- **scale** (*np.ndarray*[N, 2]) –Scale of the bounding box wrt height/width.
- **post\_process** (*str/None*) –Choice of methods to post-process heatmaps. Currently supported: None, ‘default’ , ‘unbiased’ , ‘megvii’ .
- **unbiased** (*bool*) –Option to use unbiased decoding. Mutually exclusive with megvii. Note: this arg is deprecated and unbiased=True can be replaced by post\_process=‘unbiased’ Paper ref: Zhang et al. Distribution-Aware Coordinate Representation for Human Pose Estimation (CVPR 2020).
- **kernel** (*int*) –Gaussian kernel size (K) for modulation, which should match the heatmap gaussian sigma when training. K=17 for sigma=3 and k=11 for sigma=2.
- **valid\_radius\_factor** (*float*) –The radius factor of the positive area in classification heatmap for UDP.
- **use\_udp** (*bool*) –Use unbiased data processing.
- **target\_type** (*str*) –‘GaussianHeatmap’ or ‘CombinedTarget’ . GaussianHeatmap: Classification target with gaussian distribution. CombinedTarget: The combination of classification target (response map) and regression target (offset map). Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

#### 返回

A tuple containing keypoint predictions and scores.



- preds (np.ndarray[N, K, 2]): Predicted keypoint location in images.
- maxvals (np.ndarray[N, K, 1]): Scores (confidence) of the keypoints.

返回类型 tuple

`mmpose.core.evaluation.keypoints_from_heatmaps3d(heatmaps, center, scale)`

Get final keypoint predictions from 3d heatmaps and transform them back to the image.

---

注解:

- batch size: N
  - num keypoints: K
  - heatmap depth size: D
  - heatmap height: H
  - heatmap width: W
- 

参数

- **heatmaps** (np.ndarray[N, K, D, H, W]) –model predicted heatmaps.
- **center** (np.ndarray[N, 2]) –Center of the bounding box (x, y).
- **scale** (np.ndarray[N, 2]) –Scale of the bounding box wrt height/width.

返回

A tuple containing keypoint predictions and scores.

- preds (np.ndarray[N, K, 3]): Predicted 3d keypoint location in images.
- maxvals (np.ndarray[N, K, 1]): Scores (confidence) of the keypoints.

返回类型 tuple

`mmpose.core.evaluation.keypoints_from_regression(regression_preds, center, scale, img_size)`

Get final keypoint predictions from regression vectors and transform them back to the image.

---

注解:

- batch\_size: N
  - num\_keypoints: K
- 

参数

- **regression\_preds** (np.ndarray[N, K, 2]) –model prediction.

- **center** (*np.ndarray*[*N*, 2]) –Center of the bounding box (x, y).
- **scale** (*np.ndarray*[*N*, 2]) –Scale of the bounding box wrt height/width.
- **img\_size** (*list* (*img\_width*, *img\_height*)) –model input image size.

返回

- preds (*np.ndarray*[*N*, *K*, 2]): Predicted keypoint location in images.
- maxvals (*np.ndarray*[*N*, *K*, 1]): Scores (confidence) of the keypoints.

返回类型 tuple

`mmpose.core.evaluation.multilabel_classification_accuracy(pred, gt, mask, thr=0.5)`

Get multi-label classification accuracy.

---

注解:

- batch size: N
  - label number: L
- 

参数

- **pred** (*np.ndarray*[*N*, *L*, 2]) –model predicted labels.
- **gt** (*np.ndarray*[*N*, *L*, 2]) –ground-truth labels.
- **mask** (*np.ndarray*[*N*, 1] or *np.ndarray*[*N*, *L*]) –reliability of
- **labels.** (*ground-truth*) –

返回 multi-label classification accuracy.

返回类型 float

`mmpose.core.evaluation.pose_pck_accuracy(output, target, mask, thr=0.05, normalize=None)`

Calculate the pose accuracy of PCK for each individual keypoint and the averaged accuracy across all keypoints from heatmaps.

---

注解: PCK metric measures accuracy of the localization of the body joints. The distances between predicted positions and the ground-truth ones are typically normalized by the bounding box size. The threshold (thr) of the normalized distance is commonly set as 0.05, 0.1 or 0.2 etc.

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **output** (`np.ndarray[N, K, H, W]`) –Model output heatmaps.
- **target** (`np.ndarray[N, K, H, W]`) –Groundtruth heatmaps.
- **mask** (`np.ndarray[N, K]`) –Visibility of the target. False for invisible joints, and True for visible. Invisible joints will be ignored for accuracy calculation.
- **thr** (`float`) –Threshold of PCK calculation. Default 0.05.
- **normalize** (`np.ndarray[N, 2]`) –Normalization factor for H&W.

### 返回

A tuple containing keypoint accuracy.

- `np.ndarray[K]`: Accuracy of each keypoint.
- `float`: Averaged accuracy across all keypoints.
- `int`: Number of valid keypoints.

返回类型 `tuple`

`mmpose.core.evaluation.post_dark_udp(coords, batch_heatmaps, kernel=3)`

DARK post-processing. Implemented by udp. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020). Zhang et al. Distribution-Aware Coordinate Representation for Human Pose Estimation (CVPR 2020).

### 注解:

- batch size: B
- num keypoints: K
- num persons: N
- height of heatmaps: H
- width of heatmaps: W

B=1 for bottom\_up paradigm where all persons share the same heatmap. B=N for top\_down paradigm where each person has its own heatmaps.

### 参数

- **coords** (`np.ndarray[N, K, 2]`) –Initial coordinates of human pose.
- **batch\_heatmaps** (`np.ndarray[B, K, H, W]`) –batch\_heatmaps

- **kernel** (*int*) –Gaussian kernel size (K) for modulation.

返回 Refined coordinates.

返回类型 `np.ndarray([N, K, 2])`

`mmpose.core.evaluation.split_ae_outputs` (*outputs, num\_joints, with\_heatmaps, with\_ae, select\_output\_index*)

Split multi-stage outputs into heatmaps & tags.

参数

- **outputs** (*list(Tensor)*) –Outputs of network
- **num\_joints** (*int*) –Number of joints
- **with\_heatmaps** (*list[bool]*) –Option to output heatmaps for different stages.
- **with\_ae** (*list[bool]*) –Option to output ae tags for different stages.
- **select\_output\_index** (*list[int]*) –Output keep the selected index

返回

A tuple containing multi-stage outputs.

- `list[Tensor]`: multi-stage heatmaps.
- `list[Tensor]`: multi-stage tags.

返回类型 `tuple`

## 47.2 fp16

**class** `mmpose.core.fp16.Fp16OptimizerHook` (*grad\_clip=None, coalesce=True, bucket\_size\_mb=-1, loss\_scale=512.0, distributed=True*)

FP16 optimizer hook.

The steps of fp16 optimizer is as follows. 1. Scale the loss value. 2. BP in the fp16 model. 2. Copy gradients from fp16 model to fp32 weights. 3. Update fp32 weights. 4. Copy updated parameters from fp32 weights to fp16 model.

Refer to <https://arxiv.org/abs/1710.03740> for more details.

参数 **loss\_scale** (*float*) –Scale factor multiplied with loss.

**after\_train\_iter** (*runner*)

Backward optimization steps for Mixed Precision Training.

1. Scale the loss by a scale factor.
2. Backward the loss to obtain the gradients (fp16).

3. Copy gradients from the model to the fp32 weight copy.
4. Scale the gradients back and update the fp32 weight copy.
5. Copy back the params from fp32 weight copy to the fp16 model.

参数 **runner** (`mmcv.Runner`) –The underlines training runner.

**before\_run** (*runner*)

Preparing steps before Mixed Precision Training.

1. Make a master copy of fp32 weights for optimization.
2. Convert the main model from fp32 to fp16.

参数 **runner** (`mmcv.Runner`) –The underlines training runner.

**static copy\_grads\_to\_fp32** (*fp16\_net, fp32\_weights*)

Copy gradients from fp16 model to fp32 weight copy.

**static copy\_params\_to\_fp16** (*fp16\_net, fp32\_weights*)

Copy updated params from fp32 weight copy to fp16 model.

`mmpose.core.fp16.auto_fp16` (*apply\_to=None, out\_fp32=False*)

Decorator to enable fp16 training automatically.

This decorator is useful when you write custom modules and want to support mixed precision training. If inputs arguments are fp32 tensors, they will be converted to fp16 automatically. Arguments other than fp32 tensors are ignored.

参数

- **apply\_to** (*Iterable, optional*) –The argument names to be converted. *None* indicates all arguments.
- **out\_fp32** (*bool*) –Whether to convert the output back to fp32.

示例

```
>>> import torch.nn as nn
>>> class MyModule1(nn.Module):
>>>
>>>     # Convert x and y to fp16
>>>     @auto_fp16()
>>>     def forward(self, x, y):
>>>         pass
```

```

>>> import torch.nn as nn
>>> class MyModule2(nn.Module):
>>>
>>>     # convert pred to fp16
>>>     @auto_fp16(apply_to=('pred', ))
>>>     def do_something(self, pred, others):
>>>         pass

```

`mmpose.core.fp16.cast_tensor_type(inputs, src_type, dst_type)`

Recursively convert Tensor in inputs from `src_type` to `dst_type`.

#### 参数

- **inputs** –Inputs that to be casted.
- **src\_type** (`torch.dtype`) –Source type.
- **dst\_type** (`torch.dtype`) –Destination type.

**返回** The same type with inputs, but all contained Tensors have been cast.

`mmpose.core.fp16.force_fp32(apply_to=None, out_fp16=False)`

Decorator to convert input arguments to fp32 in force.

This decorator is useful when you write custom modules and want to support mixed precision training. If there are some inputs that must be processed in fp32 mode, then this decorator can handle it. If inputs arguments are fp16 tensors, they will be converted to fp32 automatically. Arguments other than fp16 tensors are ignored.

#### 参数

- **apply\_to** (`Iterable`, *optional*) –The argument names to be converted. *None* indicates all arguments.
- **out\_fp16** (`bool`) –Whether to convert the output back to fp16.

#### 示例

```

>>> import torch.nn as nn
>>> class MyModule1(nn.Module):
>>>
>>>     # Convert x and y to fp32
>>>     @force_fp32()
>>>     def loss(self, x, y):
>>>         pass

```

```

>>> import torch.nn as nn
>>> class MyModule2(nn.Module):

```

(下页继续)

(续上页)

```

>>>
>>>     # convert pred to fp32
>>>     @force_fp32(apply_to=('pred', ))
>>>     def post_process(self, pred, others):
>>>         pass

```

`mmpose.core.fp16.wrap_fp16_model(model)`

Wrap the FP32 model to FP16.

1. Convert FP32 model to FP16.
2. Remain some necessary layers to be FP32, e.g., normalization layers.

**参数** `model` (`nn.Module`) – Model in FP32.

## 47.3 utils

**class** `mmpose.core.utils.ModelSetEpochHook`

The hook that tells model the current epoch in training.

**class** `mmpose.core.utils.WeightNormClipHook` (`max_norm=1.0`, `module_param_names='weight'`)

Apply weight norm clip regularization.

The module's parameter will be clip to a given maximum norm before each forward pass.

**参数**

- **max\_norm** (`float`) – The maximum norm of the parameter.
- **module\_param\_names** (`str/list`) – The parameter name (or name list) to apply weight norm clip.

**hook** (`module`, `_input`)

Hook function.

**property** `hook_type`

Hook type Subclasses should overwrite this function to return a string value in.

{`forward`, `forward_pre`, `backward`}

`mmpose.core.utils.allreduce_grads` (`params`, `coalesce=True`, `bucket_size_mb=-1`)

Allreduce gradients.

**参数**

- **params** (`list[torch.Parameters]`) – List of parameters of a model
- **coalesce** (`bool`, `optional`) – Whether allreduce parameters as a whole. Default: True.

- **bucket\_size\_mb** (*int*, *optional*) –Size of bucket, the unit is MB. Default: -1.

`mmpose.core.utils.sync_random_seed (seed=None, device='cuda')`

Make sure different ranks share the same seed.

All workers must call this function, otherwise it will deadlock. This method is generally used in *DistributedSampler*, because the seed should be identical across all processes in the distributed group. In distributed sampling, different ranks should sample non-overlapped data in the dataset. Therefore, this function is used to make sure that each rank shuffles the data indices in the same order based on the same seed. Then different ranks could use different indices to select non-overlapped data from the same data list. :param seed: The seed. Default to None. :type seed: int, Optional :param device: The device where the seed will be put on.

Default to ‘cuda’ .

返回 Seed to be used.

返回类型 int

## 47.4 post\_processing

**class** `mmpose.core.post_processing.Smoother` (*filter\_cfg*: *Union[Dict, str]*, *keypoint\_dim*: *int* = 2, *keypoint\_key*: *str* = 'keypoints')

Smoother to apply temporal smoothing on pose estimation results with a filter.

---

**注解:** T: The temporal length of the pose sequence K: The keypoint number of each target C: The keypoint coordinate dimension

---

### 参数

- **filter\_cfg** (*dict* | *str*) –The filter config. See example config files in *configs/\_base\_/filters/* for details. Alternatively a config file path can be accepted and the config will be loaded.
- **keypoint\_dim** (*int*) –The keypoint coordinate dimension, which is also indicated as C. Default: 2
- **keypoint\_key** (*str*) –The dict key of the keypoints in the pose results. Default: ‘keypoints’



## 示例

```

>>> import numpy as np
>>> # Build dummy pose result
>>> results = []
>>> for t in range(10):
>>>     results_t = []
>>>     for track_id in range(2):
>>>         result = {
>>>             'track_id': track_id,
>>>             'keypoints': np.random.rand(17, 3)
>>>         }
>>>         results_t.append(result)
>>>     results.append(results_t)
>>> # Example 1: Smooth multi-frame pose results offline.
>>> filter_cfg = dict(type='GaussianFilter', window_size=3)
>>> smoother = Smoother(filter_cfg, keypoint_dim=2)
>>> smoothed_results = smoother.smooth(results)
>>> # Example 2: Smooth pose results online frame-by-frame
>>> filter_cfg = dict(type='GaussianFilter', window_size=3)
>>> smoother = Smoother(filter_cfg, keypoint_dim=2)
>>> for result_t in results:
>>>     smoothed_result_t = smoother.smooth(result_t)

```

**smooth** (results)

Apply temporal smoothing on pose estimation sequences.

**参数 results** (`list[dict]` | `list[list[dict]]`) – The pose results of a single frame (non-nested list) or multiple frames (nested list). The result of each target is a dict, which should contains:

- **track\_id** (optional, Any): The track ID of the target
- **keypoints** (`np.ndarray`): The keypoint coordinates in [K, C]

**返回** Temporal smoothed pose results, which has the same data structure as the input's.

**返回类型** (`list[dict]` | `list[list[dict]]`)

`mmpose.core.post_processing.affine_transform` (`pt`, `trans_mat`)

Apply an affine transformation to the points.

**参数**

- **pt** (`np.ndarray`) – a 2 dimensional point to be transformed
- **trans\_mat** (`np.ndarray`) – 2x3 matrix of an affine transform

**返回** Transformed points.

返回类型 `np.ndarray`

`mmpose.core.post_processing.flip_back(output_flipped, flip_pairs, target_type='GaussianHeatmap')`

Flip the flipped heatmaps back to the original form.

---

注解:

- `batch_size`: N
  - `num_keypoints`: K
  - `heatmap height`: H
  - `heatmap width`: W
- 

参数

- **`output_flipped`** (`np.ndarray[N, K, H, W]`) –The output heatmaps obtained from the flipped images.
- **`flip_pairs`** (`list[tuple()]`) –Pairs of keypoints which are mirrored (for example, left ear –right ear).
- **`target_type`** (`str`) –GaussianHeatmap or CombinedTarget

返回 heatmaps that flipped back to the original image

返回类型 `np.ndarray`

`mmpose.core.post_processing.flipplr_joints(joints_3d, joints_3d_visible, img_width, flip_pairs)`

Flip human joints horizontally.

---

注解:

- `num_keypoints`: K
- 

参数

- **`joints_3d`** (`np.ndarray([K, 3])`) –Coordinates of keypoints.
- **`joints_3d_visible`** (`np.ndarray([K, 1])`) –Visibility of keypoints.
- **`img_width`** (`int`) –Image width.
- **`flip_pairs`** (`list[tuple]`) –Pairs of keypoints which are mirrored (for example, left ear and right ear).

返回

Flipped human joints.

- `joints_3d_flipped` (`np.ndarray([K, 3])`): Flipped joints.
- `joints_3d_visible_flipped` (`np.ndarray([K, 1])`): Joint visibility.

返回类型 `tuple`

```
mmpose.core.post_processing.fliplr_regression(regression, flip_pairs, center_mode='static',
                                              center_x=0.5, center_index=0)
```

Flip human joints horizontally.

---

注解:

- `batch_size`: N
  - `num_keypoint`: K
- 

参数

- **`regression`** (`np.ndarray([..., K, C])`) –Coordinates of keypoints, where K is the joint number and C is the dimension. Example shapes are:
  - [N, K, C]: a batch of keypoints where N is the batch size.
  - [N, T, K, C]: a batch of pose sequences, where T is the frame number.
- **`flip_pairs`** (`list[tuple()]`) –Pairs of keypoints which are mirrored (for example, left ear –right ear).
- **`center_mode`** (`str`) –The mode to set the center location on the x-axis to flip around. Options are:
  - `static`: use a static x value (see `center_x` also)
  - `root`: use a root joint (see `center_index` also)
- **`center_x`** (`float`) –Set the x-axis location of the flip center. Only used when `center_mode=static`.
- **`center_index`** (`int`) –Set the index of the root joint, whose x location will be used as the flip center. Only used when `center_mode=root`.

返回 Flipped joints.

返回类型 `np.ndarray([..., K, C])`

```
mmpose.core.post_processing.get_affine_transform(center, scale, rot, output_size, shift=(0.0,
                                                                                   0.0), inv=False)
```

Get the affine transform matrix, given the center/scale/rot/output\_size.

参数

- **center** (*np.ndarray*[2, ]) –Center of the bounding box (x, y).
- **scale** (*np.ndarray*[2, ]) –Scale of the bounding box wrt [width, height].
- **rot** (*float*) –Rotation angle (degree).
- **output\_size** (*np.ndarray*[2, ] | *list*(2, )) –Size of the destination heatmaps.
- **shift** (0–100%) –Shift translation ratio wrt the width/height. Default (0., 0.).
- **inv** (*bool*) –Option to inverse the affine transform direction. (inv=False: src->dst or inv=True: dst->src)

返回 The transform matrix.

返回类型 *np.ndarray*

`mmpose.core.post_processing.get_warp_matrix(theta, size_input, size_dst, size_target)`

Calculate the transformation matrix under the constraint of unbiased. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

参数

- **theta** (*float*) –Rotation angle in degrees.
- **size\_input** (*np.ndarray*) –Size of input image [w, h].
- **size\_dst** (*np.ndarray*) –Size of output image [w, h].
- **size\_target** (*np.ndarray*) –Size of ROI in input plane [w, h].

返回 A matrix for transformation.

返回类型 *np.ndarray*

`mmpose.core.post_processing.nearby_joints_nms(kpts_db, dist_thr, num_nearby_joints_thr=None, score_per_joint=False, max_dets=-1)`

Nearby joints NMS implementations.

参数

- **kpts\_db** (*list*[*dict*]) –keypoints and scores.
- **dist\_thr** (*float*) –threshold for judging whether two joints are close.
- **num\_nearby\_joints\_thr** (*int*) –threshold for judging whether two instances are close.
- **max\_dets** (*int*) –max number of detections to keep.
- **score\_per\_joint** (*bool*) –the input scores (in kpts\_db) are per joint scores.

返回 indexes to keep.

返回类型 *np.ndarray*

`mmpose.core.post_processing.oks_iou(g, d, a_g, a_d, sigmas=None, vis_thr=None)`

Calculate oks ious.

**参数**

- **g** –Ground truth keypoints.
- **d** –Detected keypoints.
- **a\_g** –Area of the ground truth object.
- **a\_d** –Area of the detected object.
- **sigmas** –standard deviation of keypoint labelling.
- **vis\_thr** –threshold of the keypoint visibility.

**返回** The oks ious.

**返回类型** list

```
mmpose.core.post_processing.oks_nms(kpts_db, thr, sigmas=None, vis_thr=None,
                                     score_per_joint=False)
```

OXS NMS implementations.

**参数**

- **kpts\_db** –keypoints.
- **thr** –Retain overlap < thr.
- **sigmas** –standard deviation of keypoint labelling.
- **vis\_thr** –threshold of the keypoint visibility.
- **score\_per\_joint** –the input scores (in kpts\_db) are per joint scores

**返回** indexes to keep.

**返回类型** np.ndarray

```
mmpose.core.post_processing.rotate_point(pt, angle_rad)
```

Rotate a point by an angle.

**参数**

- **pt** (*list[float]*) –2 dimensional point to be rotated
- **angle\_rad** (*float*) –rotation angle by radian

**返回** Rotated point.

**返回类型** list[float]

```
mmpose.core.post_processing.soft_oks_nms(kpts_db, thr, max_dets=20, sigmas=None, vis_thr=None,
                                           score_per_joint=False)
```

Soft OXS NMS implementations.

**参数**

- **kpts\_db** –keypoints and scores.

- **thr** –retain oks overlap < thr.
- **max\_dets** –max number of detections to keep.
- **sigmas** –Keypoint labelling uncertainty.
- **score\_per\_joint** –the input scores (in kpts\_db) are per joint scores

返回 indexes to keep.

返回类型 np.ndarray

`mmpose.core.post_processing.transform_preds` (*coords, center, scale, output\_size, use\_udp=False*)

Get final keypoint predictions from heatmaps and apply scaling and translation to map them back to the image.

---

注解: num\_keypoints: K

---

#### 参数

- **coords** (*np.ndarray[K, ndims]*) –
  - If ndims=2, corrrds are predicted keypoint location.
  - If ndims=4, corrrds are composed of (x, y, scores, tags)
  - If ndims=5, corrrds are composed of (x, y, scores, tags, flipped\_tags)
- **center** (*np.ndarray[2, ]*) –Center of the bounding box (x, y).
- **scale** (*np.ndarray[2, ]*) –Scale of the bounding box wrt [width, height].
- **output\_size** (*np.ndarray[2, ] | list(2, )*) –Size of the destination heatmaps.
- **use\_udp** (*bool*) –Use unbiased data processing

返回 Predicted coordinates in the images.

返回类型 np.ndarray

`mmpose.core.post_processing.warp_affine_joints` (*joints, mat*)

Apply affine transformation defined by the transform matrix on the joints.

#### 参数

- **joints** (*np.ndarray[..., 2]*) –Origin coordinate of joints.
- **mat** (*np.ndarray[3, 2]*) –The affine matrix.

返回 Result coordinate of joints.

返回类型 np.ndarray[..., 2]

## 48.1 backbones

**class** mmpose.models.backbones.**AlexNet** (*num\_classes=-1*)

AlexNet backbone.

The input for AlexNet is a 224x224 RGB image.

**参数 num\_classes** (*int*) –number of classes for classification. The default value is -1, which uses the backbone as a feature extractor without the top classifier.

**forward** (*x*)

Forward function.

**参数 x** (*Tensor | tuple[Tensor]*) –x could be a torch.Tensor or a tuple of torch.Tensor, containing input data for forward computation.

**class** mmpose.models.backbones.**CPM** (*in\_channels, out\_channels, feat\_channels=128,*  
*middle\_channels=32, num\_stages=6, norm\_cfg={'requires\_grad':*  
*True, 'type': 'BN'})*

CPM backbone.

Convolutional Pose Machines. More details can be found in the [paper](#) .

**参数**

- **in\_channels** (*int*) –The input channels of the CPM.
- **out\_channels** (*int*) –The output channels of the CPM.

- **feat\_channels** (*int*) –Feature channel of each CPM stage.
- **middle\_channels** (*int*) –Feature channel of conv after the middle stage.
- **num\_stages** (*int*) –Number of stages.
- **norm\_cfg** (*dict*) –Dictionary to construct and config norm layer.

### 示例

```
>>> from mmpose.models import CPM
>>> import torch
>>> self = CPM(3, 17)
>>> self.eval()
>>> inputs = torch.rand(1, 3, 368, 368)
>>> level_outputs = self.forward(inputs)
>>> for level_output in level_outputs:
...     print(tuple(level_output.shape))
(1, 17, 46, 46)
(1, 17, 46, 46)
(1, 17, 46, 46)
(1, 17, 46, 46)
(1, 17, 46, 46)
(1, 17, 46, 46)
(1, 17, 46, 46)
```

### **forward** (*x*)

Model forward function.

### **init\_weights** (*pretrained=None*)

Initialize the weights in backbone.

**参数** **pretrained** (*str*, *optional*) –Path to pre-trained weights. Defaults to None.

```
class mmpose.models.backbones.HRFormer (extra, in_channels=3, conv_cfg=None,
                                         norm_cfg={'requires_grad': True, 'type': 'BN'},
                                         transformer_norm_cfg={'eps': 1e-06, 'type': 'LN'},
                                         norm_eval=False, with_cp=False, zero_init_residual=False,
                                         frozen_stages=-1)
```

HRFormer backbone.

This backbone is the implementation of **HRFormer**: [High-Resolution Transformer for Dense Prediction](#).

### 参数

- **extra** (*dict*) –Detailed configuration for each stage of HRNet. There must be 4 stages, the configuration for each stage must have 5 keys:
  - **num\_modules** (*int*): The number of HRModule in this stage.



- `num_branches` (int): The number of branches in the HRModule.
- `block` (str): The type of block.
- **`num_blocks` (tuple): The number of blocks in each branch.** The length must be equal to `num_branches`.
- **`num_channels` (tuple): The number of channels in each branch.** The length must be equal to `num_branches`.
- **`in_channels` (int)** -Number of input image channels. Normally 3.
- **`conv_cfg` (dict)** -Dictionary to construct and config conv layer. Default: None.
- **`norm_cfg` (dict)** -Config of norm layer. Use *SyncBN* by default.
- **`transformer_norm_cfg` (dict)** -Config of transformer norm layer. Use *LN* by default.
- **`norm_eval` (bool)** -Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **`zero_init_residual` (bool)** -Whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: False.
- **`frozen_stages` (int)** -Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.

### 示例

```
>>> from mmpose.models import HRFormer
>>> import torch
>>> extra = dict(
>>>     stage1=dict(
>>>         num_modules=1,
>>>         num_branches=1,
>>>         block='BOTTLENECK',
>>>         num_blocks=(2, ),
>>>         num_channels=(64, )),
>>>     stage2=dict(
>>>         num_modules=1,
>>>         num_branches=2,
>>>         block='HRFORMER',
>>>         window_sizes=(7, 7),
>>>         num_heads=(1, 2),
>>>         mlp_ratios=(4, 4),
>>>         num_blocks=(2, 2),
>>>         num_channels=(32, 64)),
>>>     stage3=dict(
```

(下页继续)

(续上页)

```

>>>     num_modules=4,
>>>     num_branches=3,
>>>     block='HRFORMER',
>>>     window_sizes=(7, 7, 7),
>>>     num_heads=(1, 2, 4),
>>>     mlp_ratios=(4, 4, 4),
>>>     num_blocks=(2, 2, 2),
>>>     num_channels=(32, 64, 128)),
>>>     stage4=dict(
>>>         num_modules=2,
>>>         num_branches=4,
>>>         block='HRFORMER',
>>>         window_sizes=(7, 7, 7, 7),
>>>         num_heads=(1, 2, 4, 8),
>>>         mlp_ratios=(4, 4, 4, 4),
>>>         num_blocks=(2, 2, 2, 2),
>>>         num_channels=(32, 64, 128, 256)))
>>> self = HRFormer(extra, in_channels=1)
>>> self.eval()
>>> inputs = torch.rand(1, 1, 32, 32)
>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 32, 8, 8)
(1, 64, 4, 4)
(1, 128, 2, 2)
(1, 256, 1, 1)

```

```

class mmpose.models.backbones.HRNet (extra, in_channels=3, conv_cfg=None, norm_cfg={'type': 'BN'},
                                     norm_eval=False, with_cp=False, zero_init_residual=False,
                                     frozen_stages=-1)

```

HRNet backbone.

High-Resolution Representations for Labeling Pixels and Regions

#### 参数

- **extra** (*dict*) –detailed configuration for each stage of HRNet.
- **in\_channels** (*int*) –Number of input image channels. Default: 3.
- **conv\_cfg** (*dict*) –dictionary to construct and config conv layer.
- **norm\_cfg** (*dict*) –dictionary to construct and config norm layer.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False

- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed.
- **zero\_init\_residual** (*bool*) –whether to use zero init for last norm layer in resblocks to let them behave as identity.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.

### 示例

```
>>> from mmpose.models import HRNet
>>> import torch
>>> extra = dict(
>>>     stage1=dict(
>>>         num_modules=1,
>>>         num_branches=1,
>>>         block='BOTTLENECK',
>>>         num_blocks=(4, ),
>>>         num_channels=(64, )),
>>>     stage2=dict(
>>>         num_modules=1,
>>>         num_branches=2,
>>>         block='BASIC',
>>>         num_blocks=(4, 4),
>>>         num_channels=(32, 64)),
>>>     stage3=dict(
>>>         num_modules=4,
>>>         num_branches=3,
>>>         block='BASIC',
>>>         num_blocks=(4, 4, 4),
>>>         num_channels=(32, 64, 128)),
>>>     stage4=dict(
>>>         num_modules=3,
>>>         num_branches=4,
>>>         block='BASIC',
>>>         num_blocks=(4, 4, 4, 4),
>>>         num_channels=(32, 64, 128, 256)))
>>> self = HRNet(extra, in_channels=1)
>>> self.eval()
>>> inputs = torch.rand(1, 1, 32, 32)
>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 32, 8, 8)
```

**forward** (*x*)

Forward function.

**init\_weights** (*pretrained=None*)

Initialize the weights in backbone.

参数 **pretrained** (*str, optional*) –Path to pre-trained weights. Defaults to None.

**property norm1**

the normalization layer named “norm1”

Type nn.Module

**property norm2**

the normalization layer named “norm2”

Type nn.Module

**train** (*mode=True*)

Convert the model into training mode.

```
class mmpose.models.backbones.HourglassAENet (downsample_times=4, num_stacks=1,  
                                              out_channels=34, stage_channels=(256, 384, 512,  
                                              640, 768), feat_channels=256,  
                                              norm_cfg={'requires_grad': True, 'type': 'BN'})
```

Hourglass-AE Network proposed by Newell et al.

Associative Embedding: End-to-End Learning for Joint Detection and Grouping.

More details can be found in the [paper](#) .

参数

- **downsample\_times** (*int*) –Downsample times in a HourglassModule.
- **num\_stacks** (*int*) –Number of HourglassModule modules stacked, 1 for Hourglass-52, 2 for Hourglass-104.
- **stage\_channels** (*list[int]*) –Feature channel of each sub-module in a Hourglass-Module.
- **stage\_blocks** (*list[int]*) –Number of sub-modules stacked in a HourglassModule.
- **feat\_channels** (*int*) –Feature channel of conv after a HourglassModule.
- **norm\_cfg** (*dict*) –Dictionary to construct and config norm layer.

## 示例

```
>>> from mmpose.models import HourglassAENet
>>> import torch
>>> self = HourglassAENet()
>>> self.eval()
>>> inputs = torch.rand(1, 3, 512, 512)
>>> level_outputs = self.forward(inputs)
>>> for level_output in level_outputs:
...     print(tuple(level_output.shape))
(1, 34, 128, 128)
```

**forward** (*x*)

Model forward function.

**init\_weights** (*pretrained=None*)

Initialize the weights in backbone.

参数 **pretrained** (*str, optional*) –Path to pre-trained weights. Defaults to None.

```
class mmpose.models.backbones.HourglassNet (downsample_times=5, num_stacks=2,
                                             stage_channels=(256, 256, 384, 384, 384, 512),
                                             stage_blocks=(2, 2, 2, 2, 2, 4), feat_channel=256,
                                             norm_cfg={'requires_grad': True, 'type': 'BN'})
```

HourglassNet backbone.

Stacked Hourglass Networks for Human Pose Estimation. More details can be found in the [paper](#).

## 参数

- **downsample\_times** (*int*) –Downsample times in a HourglassModule.
- **num\_stacks** (*int*) –Number of HourglassModule modules stacked, 1 for Hourglass-52, 2 for Hourglass-104.
- **stage\_channels** (*list[int]*) –Feature channel of each sub-module in a HourglassModule.
- **stage\_blocks** (*list[int]*) –Number of sub-modules stacked in a HourglassModule.
- **feat\_channel** (*int*) –Feature channel of conv after a HourglassModule.
- **norm\_cfg** (*dict*) –Dictionary to construct and config norm layer.

## 示例

```
>>> from mmpose.models import HourglassNet
>>> import torch
>>> self = HourglassNet()
>>> self.eval()
>>> inputs = torch.rand(1, 3, 511, 511)
>>> level_outputs = self.forward(inputs)
>>> for level_output in level_outputs:
...     print(tuple(level_output.shape))
(1, 256, 128, 128)
(1, 256, 128, 128)
```

**forward** (*x*)

Model forward function.

**init\_weights** (*pretrained=None*)

Initialize the weights in backbone.

参数 **pretrained** (*str, optional*) –Path to pre-trained weights. Defaults to None.**class** mmpose.models.backbones.**I3D** (*in\_channels=3, expansion=1.0*)

I3D backbone.

Please refer to the [paper](#) for details.

Args: **in\_channels** (*int*): Input channels of the backbone, which is decided  
on the input modality.

**expansion** (*float*): The multiplier of **in\_channels** and **out\_channels**. Default: 1.**forward** (*x*)

Forward function.

参数 **x** (*Tensor | tuple[Tensor]*) –*x* could be a torch.Tensor or a tuple of torch.Tensor,  
containing input data for forward computation.

**class** mmpose.models.backbones.**LiteHRNet** (*extra, in\_channels=3, conv\_cfg=None, norm\_cfg={'type':  
'BN'}, norm\_eval=False, with\_cp=False*)

Lite-HRNet backbone.

Lite-HRNet: A Lightweight High-Resolution Network.

Code adapted from '<https://github.com/HRNet/Lite-HRNet>' .

## 参数

- **extra** (*dict*) –detailed configuration for each stage of HRNet.
- **in\_channels** (*int*) –Number of input image channels. Default: 3.

- **conv\_cfg** (*dict*) –dictionary to construct and config conv layer.
- **norm\_cfg** (*dict*) –dictionary to construct and config norm layer.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed.

### 示例

```
>>> from mmpose.models import LiteHRNet
>>> import torch
>>> extra=dict(
>>>     stem=dict(stem_channels=32, out_channels=32, expand_ratio=1),
>>>     num_stages=3,
>>>     stages_spec=dict(
>>>         num_modules=(2, 4, 2),
>>>         num_branches=(2, 3, 4),
>>>         num_blocks=(2, 2, 2),
>>>         module_type=('LITE', 'LITE', 'LITE'),
>>>         with_fuse=(True, True, True),
>>>         reduce_ratios=(8, 8, 8),
>>>         num_channels=(
>>>             (40, 80),
>>>             (40, 80, 160),
>>>             (40, 80, 160, 320),
>>>         )),
>>>     with_head=False)
>>> self = LiteHRNet(extra, in_channels=1)
>>> self.eval()
>>> inputs = torch.rand(1, 1, 32, 32)
>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 40, 8, 8)
```

### **forward** (*x*)

Forward function.

### **init\_weights** (*pretrained=None*)

Initialize the weights in backbone.

**参数 pretrained** (*str, optional*) –Path to pre-trained weights. Defaults to None.

**train** (*mode=True*)

Convert the model into training mode.

```
class mmpose.models.backbones.MSPN (unit_channels=256, num_stages=4, num_units=4,  
                                     num_blocks=[2, 2, 2, 2], norm_cfg={'type': 'BN'},  
                                     res_top_channels=64)
```

MSPN backbone. Paper ref: Li et al. “Rethinking on Multi-Stage Networks for Human Pose Estimation” (CVPR 2020).

### 参数

- **unit\_channels** (*int*) –Number of Channels in an upsample unit. Default: 256
- **num\_stages** (*int*) –Number of stages in a multi-stage MSPN. Default: 4
- **num\_units** (*int*) –Number of downsample/upsample units in a single-stage network. Default: 4 Note: Make sure num\_units == len(self.num\_blocks)
- **num\_blocks** (*list*) –Number of bottlenecks in each downsample unit. Default: [2, 2, 2, 2]
- **norm\_cfg** (*dict*) –dictionary to construct and config norm layer. Default: dict(type=' BN' )
- **res\_top\_channels** (*int*) –Number of channels of feature from ResNetTop. Default: 64.

### 示例

```
>>> from mmpose.models import MSPN
>>> import torch
>>> self = MSPN(num_stages=2,num_units=2,num_blocks=[2,2])
>>> self.eval()
>>> inputs = torch.rand(1, 3, 511, 511)
>>> level_outputs = self.forward(inputs)
>>> for level_output in level_outputs:
...     for feature in level_output:
...         print(tuple(feature.shape))
...
(1, 256, 64, 64)
(1, 256, 128, 128)
(1, 256, 64, 64)
(1, 256, 128, 128)
```

**forward** (*x*)

Model forward function.



**init\_weights** (*pretrained=None*)

Initialize model weights.

```
class mmpose.models.backbones.MobileNetV2 (widen_factor=1.0, out_indices=(7), frozen_stages=-1,  
                                             conv_cfg=None, norm_cfg={'type': 'BN'},  
                                             act_cfg={'type': 'ReLU6'}, norm_eval=False,  
                                             with_cp=False)
```

MobileNetV2 backbone.

#### 参数

- **widen\_factor** (*float*) –Width multiplier, multiply number of channels in each layer by this amount. Default: 1.0.
- **out\_indices** (*None or Sequence[int]*) –Output from which stages. Default: (7, ).
- **frozen\_stages** (*int*) –Stages to be frozen (all param fixed). Default: -1, which means not freezing any parameters.
- **conv\_cfg** (*dict*) –Config dict for convolution layer. Default: None, which means using conv2d.
- **norm\_cfg** (*dict*) –Config dict for normalization layer. Default: dict(type='BN').
- **act\_cfg** (*dict*) –Config dict for activation layer. Default: dict(type='ReLU6').
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.

**forward** (*x*)

Forward function.

参数 **x** (*Tensor | tuple[Tensor]*) –x could be a torch.Tensor or a tuple of torch.Tensor, containing input data for forward computation.

**init\_weights** (*pretrained=None*)

Init backbone weights.

参数 **pretrained** (*str | None*) –If pretrained is a string, then it initializes backbone weights by loading the pretrained checkpoint. If pretrained is None, then it follows default initializer or customized initializer in subclasses.

**make\_layer** (*out\_channels, num\_blocks, stride, expand\_ratio*)

Stack InvertedResidual blocks to build a layer for MobileNetV2.

#### 参数

- **out\_channels** (*int*) –out\_channels of block.

- **num\_blocks** (*int*) –number of blocks.
- **stride** (*int*) –stride of the first block. Default: 1
- **expand\_ratio** (*int*) –Expand the number of channels of the hidden layer in InvertedResidual by this ratio. Default: 6.

**train** (*mode=True*)

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

**参数** **mode** (*bool*) –whether to set training mode (True) or evaluation mode (False). Default: True.

**返回** self

**返回类型** Module

```
class mmpose.models.backbones.MobileNetV3 (arch='small', conv_cfg=None, norm_cfg={'type': 'BN'},  
                                           out_indices=(-1), frozen_stages=-1, norm_eval=False,  
                                           with_cp=False)
```

MobileNetV3 backbone.

**参数**

- **arch** (*str*) –Architecture of mobilnetv3, from {small, big}. Default: small.
- **conv\_cfg** (*dict*) –Config dict for convolution layer. Default: None, which means using conv2d.
- **norm\_cfg** (*dict*) –Config dict for normalization layer. Default: dict(type=' BN' ).
- **out\_indices** (*None or Sequence[int]*) –Output from which stages. Default: (-1, ), which means output tensors from final stage.
- **frozen\_stages** (*int*) –Stages to be frozen (all param fixed). Default: -1, which means not freezing any parameters.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.

**forward** (*x*)

Forward function.

**参数** **x** (*Tensor | tuple[Tensor]*) –x could be a torch.Tensor or a tuple of torch.Tensor, containing input data for forward computation.

**init\_weights** (*pretrained=None*)

Init backbone weights.

**参数 pretrained** (*str* / *None*)—If pretrained is a string, then it initializes backbone weights by loading the pretrained checkpoint. If pretrained is None, then it follows default initializer or customized initializer in subclasses.

**train** (*mode=True*)

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

**参数 mode** (*bool*)—whether to set training mode (True) or evaluation mode (False). Default: True.

**返回** self

**返回类型** Module

```
class mmpose.models.backbones.PyramidVisionTransformer (pretrain_img_size=224,
                                                         in_channels=3, embed_dims=64,
                                                         num_stages=4, num_layers=[3, 4,
                                                         6, 3], num_heads=[1, 2, 5, 8],
                                                         patch_sizes=[4, 2, 2, 2], strides=[4,
                                                         2, 2, 2], paddings=[0, 0, 0, 0],
                                                         sr_ratios=[8, 4, 2, 1],
                                                         out_indices=(0, 1, 2, 3),
                                                         mlp_ratios=[8, 8, 4, 4],
                                                         qkv_bias=True, drop_rate=0.0,
                                                         attn_drop_rate=0.0,
                                                         drop_path_rate=0.1,
                                                         use_abs_pos_embed=True,
                                                         norm_after_stage=False,
                                                         use_conv_ffn=False,
                                                         act_cfg={'type': 'GELU'},
                                                         norm_cfg={'eps': 1e-06, 'type':
                                                         'LN'}, pretrained=None,
                                                         convert_weights=True,
                                                         init_cfg=None)
```

Pyramid Vision Transformer (PVT)

Implementation of Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions.

**参数**

- **pretrain\_img\_size** (*int* / *tuple[int]*)—The size of input image when pretrain. Defaults: 224.

- **in\_channels** (*int*) –Number of input channels. Default: 3.
- **embed\_dims** (*int*) –Embedding dimension. Default: 64.
- **num\_stags** (*int*) –The num of stages. Default: 4.
- **num\_layers** (*Sequence[int]*) –The layer number of each transformer encode layer. Default: [3, 4, 6, 3].
- **num\_heads** (*Sequence[int]*) –The attention heads of each transformer encode layer. Default: [1, 2, 5, 8].
- **patch\_sizes** (*Sequence[int]*) –The patch\_size of each patch embedding. Default: [4, 2, 2, 2].
- **strides** (*Sequence[int]*) –The stride of each patch embedding. Default: [4, 2, 2, 2].
- **padding**s (*Sequence[int]*) –The padding of each patch embedding. Default: [0, 0, 0, 0].
- **sr\_ratios** (*Sequence[int]*) –The spatial reduction rate of each transformer encode layer. Default: [8, 4, 2, 1].
- **out\_indices** (*Sequence[int] | int*) –Output from which stages. Default: (0, 1, 2, 3).
- **mlp\_ratios** (*Sequence[int]*) –The ratio of the mlp hidden dim to the embedding dim of each transformer encode layer. Default: [8, 8, 4, 4].
- **qkv\_bias** (*bool*) –Enable bias for qkv if True. Default: True.
- **drop\_rate** (*float*) –Probability of an element to be zeroed. Default 0.0.
- **attn\_drop\_rate** (*float*) –The drop out rate for attention layer. Default 0.0.
- **drop\_path\_rate** (*float*) –stochastic depth rate. Default 0.1.
- **use\_abs\_pos\_embed** (*bool*) –If True, add absolute position embedding to the patch embedding. Defaults: True.
- **use\_conv\_ffn** (*bool*) –If True, use Convolutional FFN to replace FFN. Default: False.
- **act\_cfg** (*dict*) –The activation config for FFNs. Default: dict(type=' GELU' ).
- **norm\_cfg** (*dict*) –Config dict for normalization layer. Default: dict(type=' LN' ).
- **pretrained** (*str, optional*) –model pretrained path. Default: None.
- **convert\_weights** (*bool*) –The flag indicates whether the pre-trained model is from the original repo. We may need to convert some keys to make it compatible. Default: True.
- **init\_cfg** (*dict or list[dict], optional*) –Initialization config dict. Default: None.

**forward** (*x*)

Defines the computation performed at every call.

Should be overridden by all subclasses.

---

**注解:** Although the recipe for forward pass needs to be defined within this function, one should call the `Module` instance afterwards instead of this since the former takes care of running the registered hooks while the latter silently ignores them.

---

**init\_weights** (*pretrained=None*)

Initialize the weights.

**class** `mmpose.models.backbones.PyramidVisionTransformerV2` (*\*\*kwargs*)

Implementation of PVTv2: Improved Baselines with Pyramid Vision Transformer.

**class** `mmpose.models.backbones.RSN` (*unit\_channels=256, num\_stages=4, num\_units=4, num\_blocks=[2, 2, 2], num\_steps=4, norm\_cfg={'type': 'BN'}, res\_top\_channels=64, expand\_times=26*)

Residual Steps Network backbone. Paper ref: Cai et al. “Learning Delicate Local Representations for Multi-Person Pose Estimation” (ECCV 2020).

#### 参数

- **unit\_channels** (*int*) –Number of Channels in an upsample unit. Default: 256
- **num\_stages** (*int*) –Number of stages in a multi-stage RSN. Default: 4
- **num\_units** (*int*) –NUmber of downsample/upsample units in a single-stage RSN. Default: 4 Note: Make sure `num_units == len(self.num_blocks)`
- **num\_blocks** (*list*) –Number of RSBs (Residual Steps Block) in each downsample unit. Default: [2, 2, 2, 2]
- **num\_steps** (*int*) –Number of steps in a RSB. Default:4
- **norm\_cfg** (*dict*) –dictionary to construct and config norm layer. Default: `dict(type=' BN' )`
- **res\_top\_channels** (*int*) –Number of channels of feature from ResNet\_top. Default: 64.
- **expand\_times** (*int*) –Times by which the in\_channels are expanded in RSB. Default:26.

## 示例

```
>>> from mmpose.models import RSN
>>> import torch
>>> self = RSN(num_stages=2, num_units=2, num_blocks=[2, 2])
>>> self.eval()
>>> inputs = torch.rand(1, 3, 511, 511)
>>> level_outputs = self.forward(inputs)
>>> for level_output in level_outputs:
...     for feature in level_output:
...         print(tuple(feature.shape))
...
(1, 256, 64, 64)
(1, 256, 128, 128)
(1, 256, 64, 64)
(1, 256, 128, 128)
```

**forward** (*x*)

Model forward function.

**init\_weights** (*pretrained=None*)

Initialize model weights.

```
class mmpose.models.backbones.RegNet (arch, in_channels=3, stem_channels=32, base_channels=32,
                                         strides=(2, 2, 2, 2), dilations=(1, 1, 1, 1), out_indices=(3),
                                         style='pytorch', deep_stem=False, avg_down=False,
                                         frozen_stages=- 1, conv_cfg=None, norm_cfg={'requires_grad':
                                         True, 'type': 'BN'}, norm_eval=False, with_cp=False,
                                         zero_init_residual=True)
```

RegNet backbone.

More details can be found in [paper](#) .

## 参数

- **arch** (*dict*) –The parameter of RegNets. - w0 (int): initial width - wa (float): slope of width - wm (float): quantization parameter to quantize the width - depth (int): depth of the backbone - group\_w (int): width of group - bot\_mul (float): bottleneck ratio, i.e. expansion of bottleneck.
- **strides** (*Sequence[int]*) –Strides of the first block of each stage.
- **base\_channels** (*int*) –Base channels after stem layer.
- **in\_channels** (*int*) –Number of input image channels. Default: 3.
- **dilations** (*Sequence[int]*) –Dilation of each stage.
- **out\_indices** (*Sequence[int]*) –Output from which stages.

- **style** (*str*) –*pytorch* or *caffe*. If set to “pytorch”, the stride-two layer is the 3x3 conv layer, otherwise the stride-two layer is the first 1x1 conv layer. Default: “pytorch”.
- **frozen\_stages** (*int*) –Stages to be frozen (all param fixed). -1 means not freezing any parameters. Default: -1.
- **norm\_cfg** (*dict*) –dictionary to construct and config norm layer. Default: dict(type='BN', requires\_grad=True).
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **zero\_init\_residual** (*bool*) –whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: True.

### 示例

```
>>> from mmpose.models import RegNet
>>> import torch
>>> self = RegNet(
    arch=dict(
        w0=88,
        wa=26.31,
        wm=2.25,
        group_w=48,
        depth=25,
        bot_mul=1.0),
    out_indices=(0, 1, 2, 3))
>>> self.eval()
>>> inputs = torch.rand(1, 3, 32, 32)
>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 96, 8, 8)
(1, 192, 4, 4)
(1, 432, 2, 2)
(1, 1008, 1, 1)
```

**adjust\_width\_group** (*widths, bottleneck\_ratio, groups*)

Adjusts the compatibility of widths and groups.

### 参数

- **widths** (*list[int]*) –Width of each stage.

- **bottleneck\_ratio** (*float*) –Bottleneck ratio.
- **groups** (*int*) –number of groups in each stage

返回 The adjusted widths and groups of each stage.

返回类型 tuple(list)

**forward** (*x*)

Forward function.

**static generate\_regnet** (*initial\_width, width\_slope, width\_parameter, depth, divisor=8*)

Generates per block width from RegNet parameters.

参数

- **initial\_width** (*[int]*) –Initial width of the backbone
- **width\_slope** (*[float]*) –Slope of the quantized linear function
- **width\_parameter** (*[int]*) –Parameter used to quantize the width.
- **depth** (*[int]*) –Depth of the backbone.
- **divisor** (*int, optional*) –The divisor of channels. Defaults to 8.

返回

return a list of widths of each stage and the number of stages

返回类型 list, int

**get\_stages\_from\_blocks** (*widths*)

Gets widths/stage\_blocks of network at each stage.

参数 **widths** (*list[int]*) –Width in each stage.

返回 width and depth of each stage

返回类型 tuple(list)

**static quantize\_float** (*number, divisor*)

Converts a float to closest non-zero int divisible by divisor.

参数

- **number** (*int*) –Original number to be quantized.
- **divisor** (*int*) –Divisor used to quantize the number.

返回 quantized number that is divisible by divisor.

返回类型 int

**class** mmpose.models.backbones.**ResNeSt** (*depth, groups=1, width\_per\_group=4, radix=2, reduction\_factor=4, avg\_down\_stride=True, \*\*kwargs*)

ResNeSt backbone.



Please refer to the [paper](#) for details.

### 参数

- **depth** (*int*) –Network depth, from {50, 101, 152, 200}.
- **groups** (*int*) –Groups of conv2 in Bottleneck. Default: 32.
- **width\_per\_group** (*int*) –Width per group of conv2 in Bottleneck. Default: 4.
- **radix** (*int*) –Radix of SpltAtConv2d. Default: 2
- **reduction\_factor** (*int*) –Reduction factor of SplitAttentionConv2d. Default: 4.
- **avg\_down\_stride** (*bool*) –Whether to use average pool for stride in Bottleneck. Default: True.
- **in\_channels** (*int*) –Number of input image channels. Default: 3.
- **stem\_channels** (*int*) –Output channels of the stem layer. Default: 64.
- **num\_stages** (*int*) –Stages of the network. Default: 4.
- **strides** (*Sequence[int]*) –Strides of the first block of each stage. Default: (1, 2, 2, 2).
- **dilations** (*Sequence[int]*) –Dilation of each stage. Default: (1, 1, 1, 1).
- **out\_indices** (*Sequence[int]*) –Output from which stages. If only one stage is specified, a single tensor (feature map) is returned, otherwise multiple stages are specified, a tuple of tensors will be returned. Default: (3, ).
- **style** (*str*) –*pytorch* or *caffe*. If set to “pytorch”, the stride-two layer is the 3x3 conv layer, otherwise the stride-two layer is the first 1x1 conv layer.
- **deep\_stem** (*bool*) –Replace 7x7 conv in input stem with 3 3x3 conv. Default: False.
- **avg\_down** (*bool*) –Use AvgPool instead of stride conv when downsampling in the bottleneck. Default: False.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.
- **conv\_cfg** (*dict* | *None*) –The config dict for conv layers. Default: None.
- **norm\_cfg** (*dict*) –The config dict for norm layers.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **zero\_init\_residual** (*bool*) –Whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: True.

**make\_res\_layer** (*\*\*kwargs*)

Make a ResLayer.

**class** mmpose.models.backbones.**ResNeXt** (*depth, groups=32, width\_per\_group=4, \*\*kwargs*)

ResNeXt backbone.

Please refer to the [paper](#) for details.

#### 参数

- **depth** (*int*) –Network depth, from {50, 101, 152}.
- **groups** (*int*) –Groups of conv2 in Bottleneck. Default: 32.
- **width\_per\_group** (*int*) –Width per group of conv2 in Bottleneck. Default: 4.
- **in\_channels** (*int*) –Number of input image channels. Default: 3.
- **stem\_channels** (*int*) –Output channels of the stem layer. Default: 64.
- **num\_stages** (*int*) –Stages of the network. Default: 4.
- **strides** (*Sequence[int]*) –Strides of the first block of each stage. Default: (1, 2, 2, 2).
- **dilations** (*Sequence[int]*) –Dilation of each stage. Default: (1, 1, 1, 1).
- **out\_indices** (*Sequence[int]*) –Output from which stages. If only one stage is specified, a single tensor (feature map) is returned, otherwise multiple stages are specified, a tuple of tensors will be returned. Default: (3, ).
- **style** (*str*) –*pytorch* or *caffe*. If set to “pytorch”, the stride-two layer is the 3x3 conv layer, otherwise the stride-two layer is the first 1x1 conv layer.
- **deep\_stem** (*bool*) –Replace 7x7 conv in input stem with 3 3x3 conv. Default: False.
- **avg\_down** (*bool*) –Use AvgPool instead of stride conv when downsampling in the bottleneck. Default: False.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.
- **conv\_cfg** (*dict | None*) –The config dict for conv layers. Default: None.
- **norm\_cfg** (*dict*) –The config dict for norm layers.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **zero\_init\_residual** –Whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: True.

**make\_res\_layer** (*\*\*kwargs*)

Make a ResLayer.

```
class mmpose.models.backbones.ResNet (depth, in_channels=3, stem_channels=64, base_channels=64,
                                         expansion=None, num_stages=4, strides=(1, 2, 2, 2),
                                         dilations=(1, 1, 1, 1), out_indices=(3), style='pytorch',
                                         deep_stem=False, avg_down=False, frozen_stages=-1,
                                         conv_cfg=None, norm_cfg={'requires_grad': True, 'type':
                                         'BN'}, norm_eval=False, with_cp=False,
                                         zero_init_residual=True)
```

ResNet backbone.

Please refer to the [paper](#) for details.

### 参数

- **depth** (*int*) –Network depth, from {18, 34, 50, 101, 152}.
- **in\_channels** (*int*) –Number of input image channels. Default: 3.
- **stem\_channels** (*int*) –Output channels of the stem layer. Default: 64.
- **base\_channels** (*int*) –Middle channels of the first stage. Default: 64.
- **num\_stages** (*int*) –Stages of the network. Default: 4.
- **strides** (*Sequence[int]*) –Strides of the first block of each stage. Default: (1, 2, 2, 2).
- **dilations** (*Sequence[int]*) –Dilation of each stage. Default: (1, 1, 1, 1).
- **out\_indices** (*Sequence[int]*) –Output from which stages. If only one stage is specified, a single tensor (feature map) is returned, otherwise multiple stages are specified, a tuple of tensors will be returned. Default: (3, ).
- **style** (*str*) –*pytorch* or *caffe*. If set to “pytorch”, the stride-two layer is the 3x3 conv layer, otherwise the stride-two layer is the first 1x1 conv layer.
- **deep\_stem** (*bool*) –Replace 7x7 conv in input stem with 3 3x3 conv. Default: False.
- **avg\_down** (*bool*) –Use AvgPool instead of stride conv when downsampling in the bottleneck. Default: False.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.
- **conv\_cfg** (*dict* | *None*) –The config dict for conv layers. Default: None.
- **norm\_cfg** (*dict*) –The config dict for norm layers.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.

- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **zero\_init\_residual** (*bool*) –Whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: True.

### 示例

```
>>> from mmpose.models import ResNet
>>> import torch
>>> self = ResNet(depth=18, out_indices=(0, 1, 2, 3))
>>> self.eval()
>>> inputs = torch.rand(1, 3, 32, 32)
>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 64, 8, 8)
(1, 128, 4, 4)
(1, 256, 2, 2)
(1, 512, 1, 1)
```

**forward** (*x*)

Forward function.

**init\_weights** (*pretrained=None*)

Initialize the weights in backbone.

参数 **pretrained** (*str, optional*) –Path to pre-trained weights. Defaults to None.

**make\_res\_layer** (*\*\*kwargs*)

Make a ResLayer.

property **norm1**

the normalization layer named “norm1”

Type `nn.Module`

**train** (*mode=True*)

Convert the model into training mode.

**class** `mmpose.models.backbones.ResNetV1d` (*\*\*kwargs*)

ResNetV1d variant described in [Bag of Tricks](#).

Compared with default ResNet(ResNetV1b), ResNetV1d replaces the 7x7 conv in the input stem with three 3x3 convs. And in the downsampling block, a 2x2 avg\_pool with stride 2 is added before conv, whose stride is changed to 1.

**class** mmpose.models.backbones.SCNet (*depth*, *\*\*kwargs*)

SCNet backbone.

Improving Convolutional Networks with Self-Calibrated Convolutions, Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Changhu Wang, Jiashi Feng, IEEE CVPR, 2020. <http://mftp.mmcheng.net/Papers/20cvprSCNet.pdf>

### 参数

- **depth** (*int*) –Depth of scnet, from {50, 101}.
- **in\_channels** (*int*) –Number of input image channels. Normally 3.
- **base\_channels** (*int*) –Number of base channels of hidden layer.
- **num\_stages** (*int*) –SCNet stages, normally 4.
- **strides** (*Sequence[int]*) –Strides of the first block of each stage.
- **dilations** (*Sequence[int]*) –Dilation of each stage.
- **out\_indices** (*Sequence[int]*) –Output from which stages.
- **style** (*str*) –*pytorch* or *caffe*. If set to “pytorch”, the stride-two layer is the 3x3 conv layer, otherwise the stride-two layer is the first 1x1 conv layer.
- **deep\_stem** (*bool*) –Replace 7x7 conv in input stem with 3 3x3 conv
- **avg\_down** (*bool*) –Use AvgPool instead of stride conv when downsampling in the bottle-neck.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters.
- **norm\_cfg** (*dict*) –Dictionary to construct and config norm layer.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed.
- **zero\_init\_residual** (*bool*) –Whether to use zero init for last norm layer in resblocks to let them behave as identity.

### 示例

```
>>> from mmpose.models import SCNet
>>> import torch
>>> self = SCNet(depth=50, out_indices=(0, 1, 2, 3))
>>> self.eval()
>>> inputs = torch.rand(1, 3, 224, 224)
```

(下页继续)

(续上页)

```

>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 256, 56, 56)
(1, 512, 28, 28)
(1, 1024, 14, 14)
(1, 2048, 7, 7)

```

**class** mmpose.models.backbones.**SEResNeXt** (*depth, groups=32, width\_per\_group=4, \*\*kwargs*)  
 SEResNeXt backbone.

Please refer to the [paper](#) for details.

### 参数

- **depth** (*int*) –Network depth, from {50, 101, 152}.
- **groups** (*int*) –Groups of conv2 in Bottleneck. Default: 32.
- **width\_per\_group** (*int*) –Width per group of conv2 in Bottleneck. Default: 4.
- **se\_ratio** (*int*) –Squeeze ratio in SELayer. Default: 16.
- **in\_channels** (*int*) –Number of input image channels. Default: 3.
- **stem\_channels** (*int*) –Output channels of the stem layer. Default: 64.
- **num\_stages** (*int*) –Stages of the network. Default: 4.
- **strides** (*Sequence[int]*) –Strides of the first block of each stage. Default: (1, 2, 2, 2).
- **dilations** (*Sequence[int]*) –Dilation of each stage. Default: (1, 1, 1, 1).
- **out\_indices** (*Sequence[int]*) –Output from which stages. If only one stage is specified, a single tensor (feature map) is returned, otherwise multiple stages are specified, a tuple of tensors will be returned. Default: (3, ).
- **style** (*str*) –*pytorch* or *caffe*. If set to “pytorch”, the stride-two layer is the 3x3 conv layer, otherwise the stride-two layer is the first 1x1 conv layer.
- **deep\_stem** (*bool*) –Replace 7x7 conv in input stem with 3 3x3 conv. Default: False.
- **avg\_down** (*bool*) –Use AvgPool instead of stride conv when downsampling in the bottleneck. Default: False.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.
- **conv\_cfg** (*dict | None*) –The config dict for conv layers. Default: None.
- **norm\_cfg** (*dict*) –The config dict for norm layers.

- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **zero\_init\_residual** (*bool*) –Whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: True.

### 示例

```
>>> from mmpose.models import SEResNeXt
>>> import torch
>>> self = SEResNet(depth=50, out_indices=(0, 1, 2, 3))
>>> self.eval()
>>> inputs = torch.rand(1, 3, 224, 224)
>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 256, 56, 56)
(1, 512, 28, 28)
(1, 1024, 14, 14)
(1, 2048, 7, 7)
```

**make\_res\_layer** (*\*\*kwargs*)

Make a ResLayer.

**class** mmpose.models.backbones.**SEResNet** (*depth, se\_ratio=16, \*\*kwargs*)

SEResNet backbone.

Please refer to the [paper](#) for details.

### 参数

- **depth** (*int*) –Network depth, from {50, 101, 152}.
- **se\_ratio** (*int*) –Squeeze ratio in SELayer. Default: 16.
- **in\_channels** (*int*) –Number of input image channels. Default: 3.
- **stem\_channels** (*int*) –Output channels of the stem layer. Default: 64.
- **num\_stages** (*int*) –Stages of the network. Default: 4.
- **strides** (*Sequence[int]*) –Strides of the first block of each stage. Default: (1, 2, 2, 2).
- **dilations** (*Sequence[int]*) –Dilation of each stage. Default: (1, 1, 1, 1).

- **out\_indices** (*Sequence[int]*) –Output from which stages. If only one stage is specified, a single tensor (feature map) is returned, otherwise multiple stages are specified, a tuple of tensors will be returned. Default: (3, ).
- **style** (*str*) –*pytorch* or *caffe*. If set to “pytorch”, the stride-two layer is the 3x3 conv layer, otherwise the stride-two layer is the first 1x1 conv layer.
- **deep\_stem** (*bool*) –Replace 7x7 conv in input stem with 3 3x3 conv. Default: False.
- **avg\_down** (*bool*) –Use AvgPool instead of stride conv when downsampling in the bottle-neck. Default: False.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.
- **conv\_cfg** (*dict* | *None*) –The config dict for conv layers. Default: None.
- **norm\_cfg** (*dict*) –The config dict for norm layers.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **zero\_init\_residual** (*bool*) –Whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: True.

### 示例

```
>>> from mmpose.models import SEResNet
>>> import torch
>>> self = SEResNet(depth=50, out_indices=(0, 1, 2, 3))
>>> self.eval()
>>> inputs = torch.rand(1, 3, 224, 224)
>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 256, 56, 56)
(1, 512, 28, 28)
(1, 1024, 14, 14)
(1, 2048, 7, 7)
```

**make\_res\_layer** (*\*\*kwargs*)  
Make a ResLayer.



```
class mmpose.models.backbones.ShuffleNetV1 (groups=3, widen_factor=1.0, out_indices=(2),
                                             frozen_stages=-1, conv_cfg=None, norm_cfg={ 'type':
                                             'BN'}, act_cfg={ 'type': 'ReLU'}, norm_eval=False,
                                             with_cp=False)
```

ShuffleNetV1 backbone.

#### 参数

- **groups** (*int, optional*) –The number of groups to be used in grouped 1x1 convolutions in each ShuffleUnit. Default: 3.
- **widen\_factor** (*float, optional*) –Width multiplier - adjusts the number of channels in each layer by this amount. Default: 1.0.
- **out\_indices** (*Sequence[int]*) –Output from which stages. Default: (2, )
- **frozen\_stages** (*int*) –Stages to be frozen (all param fixed). Default: -1, which means not freezing any parameters.
- **conv\_cfg** (*dict*) –Config dict for convolution layer. Default: None, which means using conv2d.
- **norm\_cfg** (*dict*) –Config dict for normalization layer. Default: dict(type=' BN' ).
- **act\_cfg** (*dict*) –Config dict for activation layer. Default: dict(type=' ReLU' ).
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.

**forward** (*x*)

Forward function.

**参数** **x** (*Tensor | tuple[Tensor]*) –x could be a torch.Tensor or a tuple of torch.Tensor, containing input data for forward computation.

**init\_weights** (*pretrained=None*)

Init backbone weights.

**参数** **pretrained** (*str | None*) –If pretrained is a string, then it initializes backbone weights by loading the pretrained checkpoint. If pretrained is None, then it follows default initializer or customized initializer in subclasses.

**make\_layer** (*out\_channels, num\_blocks, first\_block=False*)

Stack ShuffleUnit blocks to make a layer.

#### 参数

- **out\_channels** (*int*) –out\_channels of the block.
- **num\_blocks** (*int*) –Number of blocks.

- **first\_block** (*bool, optional*) –Whether is the first ShuffleUnit of a sequential ShuffleUnits. Default: False, which means using the grouped 1x1 convolution.

**train** (*mode=True*)

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

**参数** **mode** (*bool*) –whether to set training mode (True) or evaluation mode (False). Default: True.

**返回** *self*

**返回类型** Module

**class** mmpose.models.backbones.**ShuffleNetV2** (*widen\_factor=1.0, out\_indices=(3), frozen\_stages=-1, conv\_cfg=None, norm\_cfg={'type': 'BN'}, act\_cfg={'type': 'ReLU'}, norm\_eval=False, with\_cp=False*)

ShuffleNetV2 backbone.

**参数**

- **widen\_factor** (*float*) –Width multiplier - adjusts the number of channels in each layer by this amount. Default: 1.0.
- **out\_indices** (*Sequence[int]*) –Output from which stages. Default: (0, 1, 2, 3).
- **frozen\_stages** (*int*) –Stages to be frozen (all param fixed). Default: -1, which means not freezing any parameters.
- **conv\_cfg** (*dict*) –Config dict for convolution layer. Default: None, which means using conv2d.
- **norm\_cfg** (*dict*) –Config dict for normalization layer. Default: dict(type=' BN' ).
- **act\_cfg** (*dict*) –Config dict for activation layer. Default: dict(type=' ReLU' ).
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.

**forward** (*x*)

Forward function.

**参数** **x** (*Tensor | tuple[Tensor]*) –x could be a torch.Tensor or a tuple of torch.Tensor, containing input data for forward computation.

**init\_weights** (*pretrained=None*)

Init backbone weights.

参数 **pretrained** (*str* | *None*) – If pretrained is a string, then it initializes backbone weights by loading the pretrained checkpoint. If pretrained is None, then it follows default initializer or customized initializer in subclasses.

**train** (*mode=True*)

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

参数 **mode** (*bool*) – whether to set training mode (*True*) or evaluation mode (*False*). Default: *True*.

返回 *self*

返回类型 *Module*

```
class mmpose.models.backbones.SwinTransformer (pretrain_img_size=224, in_channels=3,
                                              embed_dims=96, patch_size=4, window_size=7,
                                              mlp_ratio=4, depths=(2, 2, 6, 2), num_heads=(3,
                                              6, 12, 24), strides=(4, 2, 2, 2), out_indices=(0, 1,
                                              2, 3), qkv_bias=True, qk_scale=None,
                                              patch_norm=True, drop_rate=0.0,
                                              attn_drop_rate=0.0, drop_path_rate=0.1,
                                              use_abs_pos_embed=False, act_cfg={'type':
                                              'GELU'}, norm_cfg={'type': 'LN'},
                                              with_cp=False, convert_weights=False,
                                              frozen_stages=-1)
```

Swin Transformer A PyTorch implement of : *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows* -

<https://arxiv.org/abs/2103.14030>

Inspiration from <https://github.com/microsoft/Swin-Transformer>

参数

- **pretrain\_img\_size** (*int* | *tuple[int]*) – The size of input image when pretrain. Defaults: 224.
- **in\_channels** (*int*) – The num of input channels. Defaults: 3.
- **embed\_dims** (*int*) – The feature dimension. Default: 96.
- **patch\_size** (*int* | *tuple[int]*) – Patch size. Default: 4.
- **window\_size** (*int*) – Window size. Default: 7.
- **mlp\_ratio** (*int*) – Ratio of mlp hidden dim to embedding dim. Default: 4.
- **depths** (*tuple[int]*) – Depths of each Swin Transformer stage. Default: (2, 2, 6, 2).

- **num\_heads** (*tuple[int]*) –Parallel attention heads of each Swin Transformer stage. Default: (3, 6, 12, 24).
- **strides** (*tuple[int]*) –The patch merging or patch embedding stride of each Swin Transformer stage. (In swin, we set kernel size equal to stride.) Default: (4, 2, 2, 2).
- **out\_indices** (*tuple[int]*) –Output from which stages. Default: (0, 1, 2, 3).
- **qkv\_bias** (*bool, optional*) –If True, add a learnable bias to query, key, value. Default: True
- **qk\_scale** (*float | None, optional*) –Override default qk scale of head\_dim \*\* -0.5 if set. Default: None.
- **patch\_norm** (*bool*) –If add a norm layer for patch embed and patch merging. Default: True.
- **drop\_rate** (*float*) –Dropout rate. Defaults: 0.
- **attn\_drop\_rate** (*float*) –Attention dropout rate. Default: 0.
- **drop\_path\_rate** (*float*) –Stochastic depth rate. Defaults: 0.1.
- **use\_abs\_pos\_embed** (*bool*) –If True, add absolute position embedding to the patch embedding. Defaults: False.
- **act\_cfg** (*dict*) –Config dict for activation layer. Default: dict(type='LN').
- **norm\_cfg** (*dict*) –Config dict for normalization layer at output of backbone. Defaults: dict(type='LN').
- **with\_cp** (*bool, optional*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **pretrained** (*str, optional*) –model pretrained path. Default: None.
- **convert\_weights** (*bool*) –The flag indicates whether the pre-trained model is from the original repo. We may need to convert some keys to make it compatible. Default: False.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). Default: -1 (-1 means not freezing any parameters).

**forward** (*x*)

Forward function.

参数 **x** (*Tensor | tuple[Tensor]*) –x could be a torch.Tensor or a tuple of torch.Tensor, containing input data for forward computation.

**init\_weights** (*pretrained=None*)

Initialize the weights in backbone.

参数 **pretrained** (*str, optional*) –Path to pre-trained weights. Defaults to None.

**train** (*mode=True*)

Convert the model into training mode while keep layers freezed.

```
class mmpose.models.backbones.TCFormer (in_channels=3, embed_dims=[64, 128, 256, 512],
                                         num_heads=[1, 2, 4, 8], mlp_ratios=[4, 4, 4, 4],
                                         qkv_bias=True, qk_scale=None, drop_rate=0.0,
                                         attn_drop_rate=0.0, drop_path_rate=0.0, norm_cfg={'eps':
                                         1e-06, 'type': 'LN'}, num_layers=[3, 4, 6, 3], sr_ratios=[8,
                                         4, 2, 1], num_stages=4, pretrained=None, k=5,
                                         sample_ratios=[0.25, 0.25, 0.25], return_map=False,
                                         convert_weights=True)
```

Token Clustering Transformer (TCFormer)

Implementation of *Not All Tokens Are Equal: Human-centric Visual Analysis via Token Clustering Transformer*  
<<https://arxiv.org/abs/2204.08680>>

Args: *in\_channels* (int): Number of input channels. Default: 3. *embed\_dims* (list[int]): Embedding dimension. Default:

[64, 128, 256, 512].

**num\_heads** (Sequence[int]): The attention heads of each transformer encode layer. Default: [1, 2, 5, 8].

**mlp\_ratios** (Sequence[int]): The ratio of the mlp hidden dim to the embedding dim of each transformer block.

*qkv\_bias* (bool): Enable bias for qkv if True. Default: True. *qk\_scale* (float | None, optional): Override default qk scale of

head\_dim \*\* -0.5 if set. Default: None.

**drop\_rate** (float): Probability of an element to be zeroed. Default 0.0.

**attn\_drop\_rate** (float): The drop out rate for attention layer. Default 0.0.

*drop\_path\_rate* (float): stochastic depth rate. Default 0. *norm\_cfg* (dict): Config dict for normalization layer.

Default: dict(type='LN', eps=1e-6).

**num\_layers** (Sequence[int]): The layer number of each transformer encode layer. Default: [3, 4, 6, 3].

**sr\_ratios** (Sequence[int]): The spatial reduction rate of each transformer block. Default: [8, 4, 2, 1].

`num_stages` (int): The num of stages. Default: 4. `pretrained` (str, optional): model pretrained path. Default: None. `k` (int): number of the nearest neighbor used for local density. `sample_ratios` (list[float]): The sample ratios of CTM modules.

Default: [0.25, 0.25, 0.25]

**`return_map` (bool):** If True, transfer dynamic tokens to feature map at last. Default: False

**`convert_weights` (bool):** The flag indicates whether the pre-trained model is from the original repo. We may need to convert some keys to make it compatible. Default: True.

**`forward`** (*x*)

Defines the computation performed at every call.

Should be overridden by all subclasses.

---

**注解:** Although the recipe for forward pass needs to be defined within this function, one should call the `Module` instance afterwards instead of this since the former takes care of running the registered hooks while the latter silently ignores them.

---

```
class mmpose.models.backbones.TCN(in_channels, stem_channels=1024, num_blocks=2, kernel_sizes=(3,
3, 3), dropout=0.25, causal=False, residual=True,
                                use_stride_conv=False, conv_cfg={'type': 'Conv1d'},
                                norm_cfg={'type': 'BN1d'}, max_norm=None)
```

TCN backbone.

Temporal Convolutional Networks. More details can be found in the [paper](#).

### 参数

- **`in_channels`** (*int*) –Number of input channels, which equals to `num_keypoints * num_features`.
- **`stem_channels`** (*int*) –Number of feature channels. Default: 1024.
- **`num_blocks`** (*int*) –Number of basic temporal convolutional blocks. Default: 2.
- **`kernel_sizes`** (*Sequence[int]*) –Sizes of the convolving kernel of each basic block. Default: (3, 3, 3).
- **`dropout`** (*float*) –Dropout rate. Default: 0.25.
- **`causal`** (*bool*) –Use causal convolutions instead of symmetric convolutions (for real-time applications). Default: False.
- **`residual`** (*bool*) –Use residual connection. Default: True.
- **`use_stride_conv`** (*bool*) –Use TCN backbone optimized for single-frame batching, i.e. where batches have input length = receptive field, and output length = 1. This implementation

replaces dilated convolutions with strided convolutions to avoid generating unused intermediate results. The weights are interchangeable with the reference implementation. Default: False

- **conv\_cfg** (*dict*) –dictionary to construct and config conv layer. Default: dict(type='Conv1d').
- **norm\_cfg** (*dict*) –dictionary to construct and config norm layer. Default: dict(type='BN1d').
- **max\_norm** (*float/None*) –if not None, the weight of convolution layers will be clipped to have a maximum norm of max\_norm.

### 示例

```
>>> from mmpose.models import TCN
>>> import torch
>>> self = TCN(in_channels=34)
>>> self.eval()
>>> inputs = torch.rand(1, 34, 243)
>>> level_outputs = self.forward(inputs)
>>> for level_out in level_outputs:
...     print(tuple(level_out.shape))
(1, 1024, 235)
(1, 1024, 217)
```

**forward** (*x*)

Forward function.

**init\_weights** (*pretrained=None*)

Initialize the weights.

**class** mmpose.models.backbones.V2VNet (*input\_channels, output\_channels, mid\_channels=32*)  
V2VNet.

Please refer to the *paper* <<https://arxiv.org/abs/1711.07399>> for details.

### 参数

- **input\_channels** (*int*) –Number of channels of the input feature volume.
- **output\_channels** (*int*) –Number of channels of the output volume.
- **mid\_channels** (*int*) –Input and output channels of the encoder-decoder block.

**forward** (*x*)

Forward function.

```
class mmpose.models.backbones.VGG(depth, num_classes=- 1, num_stages=5, dilations=(1, 1, 1, 1, 1),  
                                   out_indices=None, frozen_stages=- 1, conv_cfg=None,  
                                   norm_cfg=None, act_cfg={'type': 'ReLU'}, norm_eval=False,  
                                   ceil_mode=False, with_last_pool=True)
```

VGG backbone.

#### 参数

- **depth** (*int*) –Depth of vgg, from {11, 13, 16, 19}.
- **with\_norm** (*bool*) –Use BatchNorm or not.
- **num\_classes** (*int*) –number of classes for classification.
- **num\_stages** (*int*) –VGG stages, normally 5.
- **dilations** (*Sequence[int]*) –Dilation of each stage.
- **out\_indices** (*Sequence[int]*) –Output from which stages. If only one stage is specified, a single tensor (feature map) is returned, otherwise multiple stages are specified, a tuple of tensors will be returned. When it is None, the default behavior depends on whether num\_classes is specified. If num\_classes <= 0, the default value is (4, ), outputting the last feature map before classifier. If num\_classes > 0, the default value is (5, ), outputting the classification score. Default: None.
- **frozen\_stages** (*int*) –Stages to be frozen (all param fixed). -1 means not freezing any parameters.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **ceil\_mode** (*bool*) –Whether to use ceil\_mode of MaxPool. Default: False.
- **with\_last\_pool** (*bool*) –Whether to keep the last pooling before classifier. Default: True.

**forward** (*x*)

Forward function.

**参数** **x** (*Tensor | tuple[Tensor]*) –x could be a torch.Tensor or a tuple of torch.Tensor, containing input data for forward computation.

**init\_weights** (*pretrained=None*)

Init backbone weights.

**参数** **pretrained** (*str | None*) –If pretrained is a string, then it initializes backbone weights by loading the pretrained checkpoint. If pretrained is None, then it follows default initializer or customized initializer in subclasses.

**train** (*mode=True*)

Sets the module in training mode.



This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

**参数** `mode` (*bool*) – whether to set training mode (`True`) or evaluation mode (`False`). Default: `True`.

**返回** `self`

**返回类型** `Module`

```
class mmpose.models.backbones.ViPNAS_MobileNetV3 (wid=[16, 16, 24, 40, 80, 112, 160],
                                                    expan=[None, 1, 5, 4, 5, 5, 6], dep=[None,
1, 4, 4, 4, 4, 4], ks=[3, 3, 7, 7, 5, 7, 5],
                                                    group=[None, 8, 120, 20, 100, 280, 240],
                                                    att=[None, True, True, False, True, True,
True], stride=[2, 1, 2, 2, 2, 1, 2],
                                                    act=['HSwish', 'ReLU', 'ReLU', 'ReLU',
'HSwish', 'HSwish', 'HSwish'],
                                                    conv_cfg=None, norm_cfg={'type': 'BN'},
                                                    frozen_stages=- 1, norm_eval=False,
                                                    with_cp=False)
```

ViPNAS\_MobileNetV3 backbone.

“ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search” More details can be found in the [paper](#).

#### 参数

- **wid** (*list(int)*) – Searched width config for each stage.
- **expan** (*list(int)*) – Searched expansion ratio config for each stage.
- **dep** (*list(int)*) – Searched depth config for each stage.
- **ks** (*list(int)*) – Searched kernel size config for each stage.
- **group** (*list(int)*) – Searched group number config for each stage.
- **att** (*list(bool)*) – Searched attention config for each stage.
- **stride** (*list(int)*) – Stride config for each stage.
- **act** (*list(dict)*) – Activation config for each stage.
- **conv\_cfg** (*dict*) – Config dict for convolution layer. Default: `None`, which means using `conv2d`.
- **norm\_cfg** (*dict*) – Config dict for normalization layer. Default: `dict(type='BN')`.
- **frozen\_stages** (*int*) – Stages to be frozen (all param fixed). Default: `-1`, which means not freezing any parameters.

- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.

**forward** (*x*)

Forward function.

**参数** **x** (*Tensor | tuple[Tensor]*) –x could be a torch.Tensor or a tuple of torch.Tensor, containing input data for forward computation.

**init\_weights** (*pretrained=None*)

Init backbone weights.

**参数** **pretrained** (*str | None*) –If pretrained is a string, then it initializes backbone weights by loading the pretrained checkpoint. If pretrained is None, then it follows default initializer or customized initializer in subclasses.

**train** (*mode=True*)

Sets the module in training mode.

This has any effect only on certain modules. See documentations of particular modules for details of their behaviors in training/evaluation mode, if they are affected, e.g. Dropout, BatchNorm, etc.

**参数** **mode** (*bool*) –whether to set training mode (True) or evaluation mode (False). Default: True.

**返回** self

**返回类型** Module

```
class mmpose.models.backbones.ViPNAS_ResNet (depth, in_channels=3, num_stages=4, strides=(1, 2,
2, 2), dilations=(1, 1, 1, 1), out_indices=(3),
style='pytorch', deep_stem=False, avg_down=False,
frozen_stages=-1, conv_cfg=None,
norm_cfg={'requires_grad': True, 'type': 'BN'},
norm_eval=False, with_cp=False,
zero_init_residual=True, wid=[48, 80, 160, 304,
608], expan=[None, 1, 1, 1, 1], dep=[None, 4, 6, 7,
3], ks=[7, 3, 5, 5, 5], group=[None, 16, 16, 16, 16],
att=[None, True, False, True, True])
```

ViPNAS\_ResNet backbone.

“ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search” More details can be found in the [paper](#).

**参数**

- **depth** (*int*) –Network depth, from {18, 34, 50, 101, 152}.

- **in\_channels** (*int*) –Number of input image channels. Default: 3.
- **num\_stages** (*int*) –Stages of the network. Default: 4.
- **strides** (*Sequence[int]*) –Strides of the first block of each stage. Default: (1, 2, 2, 2).
- **dilations** (*Sequence[int]*) –Dilation of each stage. Default: (1, 1, 1, 1).
- **out\_indices** (*Sequence[int]*) –Output from which stages. If only one stage is specified, a single tensor (feature map) is returned, otherwise multiple stages are specified, a tuple of tensors will be returned. Default: (3, ).
- **style** (*str*) –*pytorch* or *caffe*. If set to “pytorch”, the stride-two layer is the 3x3 conv layer, otherwise the stride-two layer is the first 1x1 conv layer.
- **deep\_stem** (*bool*) –Replace 7x7 conv in input stem with 3 3x3 conv. Default: False.
- **avg\_down** (*bool*) –Use AvgPool instead of stride conv when downsampling in the bottleneck. Default: False.
- **frozen\_stages** (*int*) –Stages to be frozen (stop grad and set eval mode). -1 means not freezing any parameters. Default: -1.
- **conv\_cfg** (*dict | None*) –The config dict for conv layers. Default: None.
- **norm\_cfg** (*dict*) –The config dict for norm layers.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: False.
- **with\_cp** (*bool*) –Use checkpoint or not. Using checkpoint will save some memory while slowing down the training speed. Default: False.
- **zero\_init\_residual** (*bool*) –Whether to use zero init for last norm layer in resblocks to let them behave as identity. Default: True.
- **wid** (*list(int)*) –Searched width config for each stage.
- **expan** (*list(int)*) –Searched expansion ratio config for each stage.
- **dep** (*list(int)*) –Searched depth config for each stage.
- **ks** (*list(int)*) –Searched kernel size config for each stage.
- **group** (*list(int)*) –Searched group number config for each stage.
- **att** (*list(bool)*) –Searched attention config for each stage.

**forward** (*x*)

Forward function.

**init\_weights** (*pretrained=None*)

Initialize model weights.

```
make_res_layer (**kwargs)
    Make a ViPNAS ResLayer.

property norm1
    the normalization layer named "norm1"

    Type nn.Module

train (mode=True)
    Convert the model into training mode.
```

## 48.2 necks

```
class mmpose.models.necks.FPN (in_channels, out_channels, num_outs, start_level=0, end_level=-1,
                                add_extra_convs=False, relu_before_extra_convs=False,
                                no_norm_on_lateral=False, conv_cfg=None, norm_cfg=None,
                                act_cfg=None, upsample_cfg={'mode': 'nearest'})
```

Feature Pyramid Network.

This is an implementation of paper [Feature Pyramid Networks for Object Detection](#).

### 参数

- **in\_channels** (*list[int]*) –Number of input channels per scale.
- **out\_channels** (*int*) –Number of output channels (used at each scale).
- **num\_outs** (*int*) –Number of output scales.
- **start\_level** (*int*) –Index of the start input backbone level used to build the feature pyramid. Default: 0.
- **end\_level** (*int*) –Index of the end input backbone level (exclusive) to build the feature pyramid. Default: -1, which means the last level.
- **add\_extra\_convs** (*bool | str*) –If bool, it decides whether to add conv layers on top of the original feature maps. Default to False. If True, it is equivalent to *add\_extra\_convs='on\_input'* . If str, it specifies the source feature map of the extra convs. Only the following options are allowed
  - 'on\_input' : Last feat map of neck inputs (i.e. backbone feature).
  - 'on\_lateral' : Last feature map after lateral convs.
  - 'on\_output' : The last output feature map after fpn convs.
- **relu\_before\_extra\_convs** (*bool*) –Whether to apply relu before the extra conv. Default: False.
- **no\_norm\_on\_lateral** (*bool*) –Whether to apply norm on lateral. Default: False.

- **conv\_cfg** (*dict*) –Config dict for convolution layer. Default: None.
- **norm\_cfg** (*dict*) –Config dict for normalization layer. Default: None.
- **act\_cfg** (*dict*) –Config dict for activation layer in ConvModule. Default: None.
- **upsample\_cfg** (*dict*) –Config dict for interpolate layer. Default: dict(mode='nearest').

### 示例

```
>>> import torch
>>> in_channels = [2, 3, 5, 7]
>>> scales = [340, 170, 84, 43]
>>> inputs = [torch.rand(1, c, s, s)
...           for c, s in zip(in_channels, scales)]
>>> self = FPN(in_channels, 11, len(in_channels)).eval()
>>> outputs = self.forward(inputs)
>>> for i in range(len(outputs)):
...     print(f'outputs[{i}].shape = {outputs[i].shape}')
outputs[0].shape = torch.Size([1, 11, 340, 340])
outputs[1].shape = torch.Size([1, 11, 170, 170])
outputs[2].shape = torch.Size([1, 11, 84, 84])
outputs[3].shape = torch.Size([1, 11, 43, 43])
```

**forward** (*inputs*)

Forward function.

**init\_weights** ()

Initialize model weights.

**class** mmpose.models.necks.**GlobalAveragePooling**

Global Average Pooling neck.

Note that we use *view* to remove extra channel after pooling. We do not use *squeeze* as it will also remove the batch dimension when the tensor has a batch dimension of size 1, which can lead to unexpected errors.

**forward** (*inputs*)

Defines the computation performed at every call.

Should be overridden by all subclasses.

---

**注解:** Although the recipe for forward pass needs to be defined within this function, one should call the Module instance afterwards instead of this since the former takes care of running the registered hooks while the latter silently ignores them.

---

```
class mmpose.models.necks.MTA (in_channels=[64, 128, 256, 512], out_channels=128, num_outs=4,
                                start_level=0, end_level=- 1, add_extra_convs=False,
                                relu_before_extra_convs=False, no_norm_on_lateral=False,
                                conv_cfg=None, norm_cfg=None, act_cfg=None, num_heads=[2, 2, 2,
                                2], mlp_ratios=[4, 4, 4, 4], sr_ratios=[8, 4, 2, 1], qkv_bias=True,
                                qk_scale=None, drop_rate=0.0, attn_drop_rate=0.0, drop_path_rate=0.0,
                                transformer_norm_cfg={'type': 'LN'}, use_sr_conv=False)
```

Multi-stage Token feature Aggregation (MTA) module in TCFormer.

### 参数

- **in\_channels** (*list[int]*) –Number of input channels per stage. Default: [64, 128, 256, 512].
- **out\_channels** (*int*) –Number of output channels (used at each scale).
- **num\_outs** (*int*) –Number of output scales. Default: 4.
- **start\_level** (*int*) –Index of the start input backbone level used to build the feature pyramid. Default: 0.
- **end\_level** (*int*) –Index of the end input backbone level (exclusive) to build the feature pyramid. Default: -1, which means the last level.
- **add\_extra\_convs** (*bool | str*) –If bool, it decides whether to add conv layers on top of the original feature maps. Default to False. If True, it is equivalent to `add_extra_convs='on_input'`. If str, it specifies the source feature map of the extra convs. Only the following options are allowed - 'on\_input' : Last feat map of neck inputs (i.e. backbone feature). - 'on\_output' : The last output feature map after fpn convs.
- **relu\_before\_extra\_convs** (*bool*) –Whether to apply relu before the extra conv. Default: False.
- **no\_norm\_on\_lateral** (*bool*) –Whether to apply norm on lateral. Default: False.
- **conv\_cfg** (*dict*) –Config dict for convolution layer. Default: None.
- **norm\_cfg** (*dict*) –Config dict for normalization layer. Default: None.
- **act\_cfg** (*dict*) –Config dict for activation layer in ConvModule.
- **num\_heads** (*Sequence[int]*) –The attention heads of each transformer block. Default: [2, 2, 2, 2].
- **mlp\_ratios** (*Sequence[int]*) –The ratio of the mlp hidden dim to the embedding dim of each transformer block.
- **sr\_ratios** (*Sequence[int]*) –The spatial reduction rate of each transformer block. Default: [8, 4, 2, 1].
- **qkv\_bias** (*bool*) –Enable bias for qkv if True. Default: True.

- **qk\_scale** (*float | None, optional*) –Override default qk scale of head\_dim \*\* -0.5 if set. Default: None.
- **drop\_rate** (*float*) –Probability of an element to be zeroed. Default 0.0.
- **attn\_drop\_rate** (*float*) –The drop out rate for attention layer. Default 0.0.
- **drop\_path\_rate** (*float*) –stochastic depth rate. Default 0.
- **transformer\_norm\_cfg** (*dict*) –Config dict for normalization layer in transformer blocks. Default: dict(type='LN' ).
- **use\_sr\_conv** (*bool*) –If True, use a conv layer for spatial reduction. If False, use a pooling process for spatial reduction. Defaults: False.

**forward** (*inputs*)

Forward function.

**init\_weights** ()

Initialize the weights.

```
class mmpose.models.necks.PoseWarperNeck (in_channels, out_channels, inner_channels,
                                          deform_groups=17, dilations=(3, 6, 12, 18, 24),
                                          trans_conv_kernel=1, res_blocks_cfg=None,
                                          offsets_kernel=3, deform_conv_kernel=3, in_index=0,
                                          input_transform=None, freeze_trans_layer=True,
                                          norm_eval=False, im2col_step=80)
```

PoseWarper neck.

“Learning temporal pose estimation from sparsely-labeled videos” .

### 参数

- **in\_channels** (*int*) –Number of input channels from backbone
- **out\_channels** (*int*) –Number of output channels
- **inner\_channels** (*int*) –Number of intermediate channels of the res block
- **deform\_groups** (*int*) –Number of groups in the deformable conv
- **dilations** (*list | tuple*) –different dilations of the offset conv layers
- **trans\_conv\_kernel** (*int*) –the kernel of the trans conv layer, which is used to get heatmap from the output of backbone. Default: 1
- **res\_blocks\_cfg** (*dict | None*) –config of residual blocks. If None, use the default values. If not None, it should contain the following keys:
  - **block** (str): the type of residual block, Default: ‘BASIC’ .
  - **num\_blocks** (int): the number of blocks, Default: 20.
- **offsets\_kernel** (*int*) –the kernel of offset conv layer.

- **deform\_conv\_kernel** (*int*) –the kernel of defomrable conv layer.
- **in\_index** (*int* | *Sequence*[*int*]) –Input feature index. Default: 0
- **input\_transform** (*str* | *None*) –Transformation type of input features. Options: ‘resize\_concat’ , ‘multiple\_select’ , *None*. Default: *None*.
  - ‘resize\_concat’ : Multiple feature maps will be resize to the same size as first one and than concat together. Usually used in FCN head of HRNet.
  - ‘multiple\_select’ : Multiple feature maps will be bundle into a list and passed into decode head.
  - *None*: Only one select feature map is allowed.
- **freeze\_trans\_layer** (*bool*) –Whether to freeze the transition layer (stop grad and set eval mode). Default: *True*.
- **norm\_eval** (*bool*) –Whether to set norm layers to eval mode, namely, freeze running stats (mean and var). Note: Effect on Batch Norm and its variants only. Default: *False*.
- **im2col\_step** (*int*) –the argument *im2col\_step* in deformable conv, Default: 80.

**forward** (*inputs*, *frame\_weight*)

Defines the computation performed at every call.

Should be overridden by all subclasses.

---

**注解:** Although the recipe for forward pass needs to be defined within this function, one should call the *Module* instance afterwards instead of this since the former takes care of running the registered hooks while the latter silently ignores them.

---

**train** (*mode=True*)

Convert the model into training mode.

## 48.3 detectors

```
class mmpose.models.detectors.AssociativeEmbedding (backbone, keypoint_head=None,  
                                                    train_cfg=None, test_cfg=None,  
                                                    pretrained=None, loss_pose=None)
```

Associative embedding pose detectors.

**参数**

- **backbone** (*dict*) –Backbone modules to extract feature.
- **keypoint\_head** (*dict*) –Keypoint head to process feature.



- **train\_cfg** (*dict*) –Config for training. Default: None.
- **test\_cfg** (*dict*) –Config for testing. Default: None.
- **pretrained** (*str*) –Path to the pretrained models.
- **loss\_pose** (*None*) –Deprecated arguments. Please use `loss_keypoint` for heads instead.

**forward** (*img=None, targets=None, masks=None, joints=None, img metas=None, return\_loss=True, return\_heatmap=False, \*\*kwargs*)

Calls either `forward_train` or `forward_test` depending on whether `return_loss` is `True`.

---

#### 注解:

- `batch_size`: N
  - `num_keypoints`: K
  - `num_img_channel`: C
  - `img_width`: imgW
  - `img_height`: imgH
  - `heatmaps weight`: W
  - `heatmaps height`: H
  - `max_num_people`: M
- 

#### 参数

- **img** (*torch.Tensor[N, C, imgH, imgW]*) –Input image.
- **targets** (*list(torch.Tensor[N, K, H, W])*) –Multi-scale target heatmaps.
- **masks** (*list(torch.Tensor[N, H, W])*) –Masks of multi-scale target heatmaps
- **joints** (*list(torch.Tensor[N, M, K, 2])*) –Joints of multi-scale target heatmaps for ae loss
- **img\_metas** (*dict*) –Information about val & test. By default it includes:
  - "image\_file" : image path
  - "aug\_data" : input
  - "test\_scale\_factor" : test scale factor
  - "base\_size" : base size of input
  - "center" : center of image
  - "scale" : scale of image

- "flip\_index" : flip index of keypoints
- **loss** (*return*) -return\_loss=True for training, return\_loss=False for validation & test.
- **return\_heatmap** (*bool*) -Option to return heatmap.

返回 if 'return\_loss' is true, then return losses. Otherwise, return predicted poses, scores, image paths and heatmaps.

返回类型 dict/tuple

**forward\_dummy** (*img*)

Used for computing network FLOPs.

See `tools/get_flops.py`.

参数 **img** (*torch.Tensor*) -Input image.

返回 Outputs.

返回类型 Tensor

**forward\_test** (*img, img metas, return\_heatmap=False, \*\*kwargs*)

Inference the bottom-up model.

---

注解:

- Batchsize: N (currently support batchsize = 1)
  - num\_img\_channel: C
  - img\_width: imgW
  - img\_height: imgH
- 

参数

- **flip\_index** (*List (int)*) -
- **aug\_data** (*List (Tensor [NxCximgHximgW])*) -Multi-scale image
- **test\_scale\_factor** (*List (float)*) -Multi-scale factor
- **base\_size** (*Tuple (int)*) -Base size of image when scale is 1
- **center** (*np.ndarray*) -center of image
- **scale** (*np.ndarray*) -the scale of image

**forward\_train** (*img, targets, masks, joints, img metas, \*\*kwargs*)

Forward the bottom-up model and calculate the loss.

---

**注解:** batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH  
heatmaps weight: W heatmaps height: H max\_num\_people: M

---

### 参数

- **img** (*torch.Tensor*[N, C, imgH, imgW]) –Input image.
- **targets** (*List* (*torch.Tensor*[N, K, H, W])) –Multi-scale target heatmaps.
- **masks** (*List* (*torch.Tensor*[N, H, W])) –Masks of multi-scale target heatmaps
- **joints** (*List* (*torch.Tensor*[N, M, K, 2])) –Joints of multi-scale target heatmaps for ae loss
- **img metas** (*dict*) –Information about val&test By default this includes: - “image\_file” : image path - “aug\_data” : input - “test\_scale\_factor” : test scale factor - “base\_size” : base size of input - “center” : center of image - “scale” : scale of image - “flip\_index” : flip index of keypoints

**返回** The total loss for bottom-up

**返回类型** dict

**init\_weights** (*pretrained=None*)

Weight initialization for model.

**show\_result** (*img, result, skeleton=None, kpt\_score\_thr=0.3, bbox\_color=None, pose\_kpt\_color=None, pose\_link\_color=None, radius=4, thickness=1, font\_scale=0.5, win\_name="", show=False, show\_keypoint\_weight=False, wait\_time=0, out\_file=None*)

Draw result over img.

### 参数

- **img** (*str or Tensor*) –The image to be displayed.
- **result** (*list[dict]*) –The results to draw over *img* (bbox\_result, pose\_result).
- **skeleton** (*list[list]*) –The connection of keypoints. skeleton is 0-based indexing.
- **kpt\_score\_thr** (*float, optional*) –Minimum score of keypoints to be shown. Default: 0.3.
- **pose\_kpt\_color** (*np.array[Nx3]*) –Color of N keypoints. If None, do not draw keypoints.
- **pose\_link\_color** (*np.array[Mx3]*) –Color of M links. If None, do not draw links.
- **radius** (*int*) –Radius of circles.
- **thickness** (*int*) –Thickness of lines.

- **font\_scale** (*float*) –Font scales of texts.
- **win\_name** (*str*) –The window name.
- **show** (*bool*) –Whether to show the image. Default: False.
- **show\_keypoint\_weight** (*bool*) –Whether to change the transparency using the predicted confidence scores of keypoints.
- **wait\_time** (*int*) –Value of waitKey param. Default: 0.
- **out\_file** (*str or None*) –The filename to write the image. Default: None.

返回 Visualized image only if not *show* or *out\_file*

返回类型 Tensor

**property with\_keypoint**

Check if has keypoint\_head.

**class** mmpose.models.detectors.**CID** (*backbone, keypoint\_head=None, train\_cfg=None, test\_cfg=None, pretrained=None, loss\_pose=None*)

Contextual Instance Decouple for Multi-Person Pose Estimation.

参数

- **backbone** (*dict*) –Backbone modules to extract feature.
- **keypoint\_head** (*dict*) –Keypoint head to process feature.
- **train\_cfg** (*dict*) –Config for training. Default: None.
- **test\_cfg** (*dict*) –Config for testing. Default: None.
- **pretrained** (*str*) –Path to the pretrained models.
- **loss\_pose** (*None*) –Deprecated arguments. Please use *loss\_keypoint* for heads instead.

**forward** (*img=None, multi\_heatmap=None, multi\_mask=None, instance\_coord=None, instance\_heatmap=None, instance\_mask=None, instance\_valid=None, img metas=None, return\_loss=True, return\_heatmap=False, \*\*kwargs*)

Calls either *forward\_train* or *forward\_test* depending on whether *return\_loss* is True.

---

注解:

- *batch\_size*: N
- *num\_keypoints*: K
- *num\_img\_channel*: C
- *img\_width*: imgW
- *img\_height*: imgH

- heatmaps weight: W
- heatmaps height: H
- max\_num\_people: M

### 参数

- **img** (*torch.Tensor*[N, C, imgH, imgW]) –Input image.
- **multi\_heatmap** (*torch.Tensor*[N, C, H, W]) –Multi-person heatmaps
- **multi\_mask** (*torch.Tensor*[N, 1, H, W]) –Multi-person heatmap mask
- **instance\_coord** (*torch.Tensor*[N, M, 2]) –Instance center coord
- **instance\_heatmap** (*torch.Tensor*[N, M, C, H, W]) –Single person heatmap for each instance
- **instance\_mask** (*torch.Tensor*[N, M, C, 1, 1]) –Single person heatmap mask
- **instance\_valid** (*torch.Tensor*[N, M]) –Bool mask to indicate the existence of each person
- **img metas** (*dict*) –Information about val & test. By default it includes:
  - "image\_file" : image path
  - "aug\_data" : input
  - "test\_scale\_factor" : test scale factor
  - "base\_size" : base size of input
  - "center" : center of image
  - "scale" : scale of image
  - "flip\_index" : flip index of keypoints
- **loss** (*return*) –return\_loss=True for training, return\_loss=False for validation & test.
- **return\_heatmap** (*bool*) –Option to return heatmap.

**返回** if 'return\_loss' is true, then return losses. Otherwise, return predicted poses, scores, image paths and heatmaps.

**返回类型** dicttuple

**forward\_dummy** (*img*)

Used for computing network FLOPs.

See `tools/get_flops.py`.

参数 **img** (*torch.Tensor*) –Input image.

返回 Outputs.

返回类型 Tensor

**forward\_test** (*img, img metas, return\_heatmap=False, \*\*kwargs*)

Inference the bottom-up model.

---

注解:

- Batchsize: N (currently support batchsize = 1)
  - num\_img\_channel: C
  - img\_width: imgW
  - img\_height: imgH
- 

参数

- **flip\_index** (*List (int)*) –
- **aug\_data** (*List (Tensor [NxCximgHximgW])*) –Multi-scale image
- **test\_scale\_factor** (*List (float)*) –Multi-scale factor
- **base\_size** (*Tuple (int)*) –Base size of image when scale is 1
- **center** (*np.ndarray*) –center of image
- **scale** (*np.ndarray*) –the scale of image

**forward\_train** (*img, multi\_heatmap, multi\_mask, instance\_coord, instance\_heatmap, instance\_mask, instance\_valid, img metas, \*\*kwargs*)

Forward CID model and calculate the loss.

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH  
heatmaps weight: W heatmaps height: H max\_num\_people: M

---

参数

- **img** (*torch.Tensor [N, C, imgH, imgW]*) –Input image.
- **multi\_heatmap** (*torch.Tensor [N, C, H, W]*) –Multi-person heatmaps
- **multi\_mask** (*torch.Tensor [N, 1, H, W]*) –Multi-person heatmap mask
- **instance\_coord** (*torch.Tensor [N, M, 2]*) –Instance center coord

- **instance\_heatmap** (*torch.Tensor*[*N, M, C, H, W*]) –Single person heatmap for each instance
- **instance\_mask** (*torch.Tensor*[*N, M, C, 1, 1*]) –Single person heatmap mask
- **instance\_valid** (*torch.Tensor*[*N, M*]) –Bool mask to indicate the existence of each person
- **img metas** (*dict*) –Information about val&test By default this includes: - “image\_file” : image path - “aug\_data” : input - “test\_scale\_factor” : test scale factor - “base\_size” : base size of input - “center” : center of image - “scale” : scale of image - “flip\_index” : flip index of keypoints

返回 The total loss for bottom-up

返回类型 dict

**init\_weights** (*pretrained=None*)

Weight initialization for model.

**show\_result** (*img, result, skeleton=None, kpt\_score\_thr=0.3, bbox\_color=None, pose\_kpt\_color=None, pose\_link\_color=None, radius=4, thickness=1, font\_scale=0.5, win\_name="", show=False, show\_keypoint\_weight=False, wait\_time=0, out\_file=None*)

Draw result over *img*.

#### 参数

- **img** (*str or Tensor*) –The image to be displayed.
- **result** (*list[dict]*) –The results to draw over *img* (bbox\_result, pose\_result).
- **skeleton** (*list[list]*) –The connection of keypoints. skeleton is 0-based indexing.
- **kpt\_score\_thr** (*float, optional*) –Minimum score of keypoints to be shown. Default: 0.3.
- **pose\_kpt\_color** (*np.array[Nx3]*) –Color of N keypoints. If None, do not draw keypoints.
- **pose\_link\_color** (*np.array[Mx3]*) –Color of M links. If None, do not draw links.
- **radius** (*int*) –Radius of circles.
- **thickness** (*int*) –Thickness of lines.
- **font\_scale** (*float*) –Font scales of texts.
- **win\_name** (*str*) –The window name.
- **show** (*bool*) –Whether to show the image. Default: False.
- **show\_keypoint\_weight** (*bool*) –Whether to change the transparency using the predicted confidence scores of keypoints.

- **wait\_time** (*int*) –Value of waitKey param. Default: 0.
- **out\_file** (*str* or *None*) –The filename to write the image. Default: None.

返回 Visualized image only if not *show* or *out\_file*

返回类型 Tensor

**property with\_keypoint**

Check if has keypoint\_head.

```
class mmpose.models.detectors.DetectAndRegress (backbone, human_detector, pose_regressor,  
                                              train_cfg=None, test_cfg=None,  
                                              pretrained=None, freeze_2d=True)
```

DetectAndRegress approach for multiview human pose detection.

参数

- **backbone** (*ConfigDict*) –Dictionary to construct the 2D pose detector
- **human\_detector** (*ConfigDict*) –dictionary to construct human detector
- **pose\_regressor** (*ConfigDict*) –dictionary to construct pose regressor
- **train\_cfg** (*ConfigDict*) –Config for training. Default: None.
- **test\_cfg** (*ConfigDict*) –Config for testing. Default: None.
- **pretrained** (*str*) –Path to the pretrained 2D model. Default: None.
- **freeze\_2d** (*bool*) –Whether to freeze the 2D model in training. Default: True.

```
forward (img=None, img metas=None, return_loss=True, targets=None, masks=None, targets_3d=None,  
        input_heatmaps=None, **kwargs)
```

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH feature\_maps width: W feature\_maps height: H volume\_length: cubeL volume\_width: cubeW volume\_height: cubeH

---

参数

- **img** (*list (torch.Tensor [Nx C x imgH x imgW])*) –Multi-camera input images to the 2D model.
- **img\_metas** (*list (dict)*) –Information about image, 3D groundtruth and camera parameters.
- **return\_loss** –Option to return loss. *return loss=True* for training, *return loss=False* for validation & test.



- **targets** (*list(torch.Tensor[NxKxHxW])*) –Multi-camera target feature\_maps of the 2D model.
- **masks** (*list(torch.Tensor[NxHxW])*) –Multi-camera masks of the input to the 2D model.
- **targets\_3d** (*torch.Tensor[NxcubeLxcubeWxcubeH]*) –Ground-truth 3D heatmap of human centers.
- **input\_heatmaps** (*list(torch.Tensor[NxKxHxW])*) –  
Multi-camera feature\_maps when the 2D model is not available. Default: None.
- **\*\*kwargs** –

返回

if ‘return\_loss’ is true, then return losses. Otherwise, return predicted poses, human centers and sample\_id

返回类型 dict

**forward\_dummy** (*img, input\_heatmaps=None, num\_candidates=5*)

Used for computing network FLOPs.

**forward\_test** (*img, img metas, input\_heatmaps=None*)

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH feature\_maps width: W feature\_maps height: H volume\_length: cubeL volume\_width: cubeW volume\_height: cubeH

---

参数

- **img** (*list(torch.Tensor[NxCximgHximgW])*) –Multi-camera input images to the 2D model.
- **img\_metas** (*list(dict)*) –Information about image, 3D groundtruth and camera parameters.
- **input\_heatmaps** (*list(torch.Tensor[NxKxHxW])*) –  
Multi-camera feature\_maps when the 2D model is not available. Default: None.

返回 predicted poses, human centers and sample\_id

返回类型 dict

**forward\_train** (*img, img\_metas, targets=None, masks=None, targets\_3d=None, input\_heatmaps=None*)

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH feature\_maps width: W feature\_maps height: H volume\_length: cubeL volume\_width: cubeW volume\_height: cubeH

---

### 参数

- **img** (*list(torch.Tensor[NxCximgHximgW])*) –Multi-camera input images to the 2D model.
- **img metas** (*list(dict)*) –Information about image, 3D groundtruth and camera parameters.
- **targets** (*list(torch.Tensor[NxKxHxW])*) –Multi-camera target feature\_maps of the 2D model.
- **masks** (*list(torch.Tensor[NxHxW])*) –Multi-camera masks of the input to the 2D model.
- **targets\_3d** (*torch.Tensor[NxcubeLxcubeWxcubeH]*) –Ground-truth 3D heatmap of human centers.
- **input\_heatmaps** (*list(torch.Tensor[NxKxHxW])*) –  
Multi-camera feature\_maps when the 2D model is not available. Default: None.

返回 losses.

返回类型 dict

**show\_result** (*img, img\_metas, visualize\_2d=False, input\_heatmaps=None, dataset\_info=None, radius=4, thickness=2, out\_dir=None, show=False*)

Visualize the results.

**train** (*mode=True*)

Sets the module in training mode. :param mode: whether to set training mode (True)

or evaluation mode (False). Default: True.

返回 self

返回类型 Module

**train\_step** (*data\_batch, optimizer, \*\*kwargs*)

The iteration step during training.

This method defines an iteration step during training, except for the back propagation and optimizer updating, which are done in an optimizer hook. Note that in some complicated cases or models, the whole process including back propagation and optimizer updating is also defined in this method, such as GAN.

### 参数

- **data\_batch** (*dict*) –The output of dataloader.
- **optimizer** (*torch.optim.Optimizer | dict*) –The optimizer of runner is passed to `train_step()`. This argument is unused and reserved.

### 返回

It should contain at least 3 keys: **loss**, **log\_vars**, **num\_samples**. **loss** is a tensor for back propagation, which can be a weighted sum of multiple losses. **log\_vars** contains all the variables to be sent to the logger. **num\_samples** indicates the batch size (when the model is DDP, it means the batch size on each GPU), which is used for averaging the logs.

返回类型 `dict`

```
class mmpose.models.detectors.DisentangledKeypointRegressor (backbone,
                                                             keypoint_head=None,
                                                             train_cfg=None,
                                                             test_cfg=None,
                                                             pretrained=None)
```

Disentangled keypoint regression pose detector.

### 参数

- **backbone** (*dict*) –Backbone modules to extract feature.
- **keypoint\_head** (*dict*) –Keypoint head to process feature.
- **train\_cfg** (*dict*) –Config for training. Default: None.
- **test\_cfg** (*dict*) –Config for testing. Default: None.
- **pretrained** (*str*) –Path to the pretrained models.

**forward** (*img=None, heatmaps=None, masks=None, offsets=None, offset\_weights=None, img metas=None, return\_loss=True, return\_heatmap=False, \*\*kwargs*)

Calls either `forward_train` or `forward_test` depending on whether `return_loss` is True.

### 注解:

- `batch_size`: N
- `num_keypoints`: K
- `num_img_channel`: C
- `img_width`: imgW
- `img_height`: imgH
- `heatmaps weight`: W

- heatmaps height: H
  - max\_num\_people: M
- 

### 参数

- **img** (*torch.Tensor*[N, C, imgH, imgW]) –# input image.
- **targets** (*list* (*torch.Tensor*[N, K, H, W])) –Multi-scale target heatmaps.
- **masks** (*list* (*torch.Tensor*[N, H, W])) –Masks of multi-scale target heatmaps
- **joints** (*list* (*torch.Tensor*[N, M, K, 2])) –Joints of multi-scale target heatmaps for ae loss
- **img metas** (*dict*) –Information about val & test. By default it includes:
  - "image\_file" : image path
  - "aug\_data" : # input
  - "test\_scale\_factor" : test scale factor
  - "base\_size" : base size of # input
  - "center" : center of image
  - "scale" : scale of image
  - "flip\_index" : flip index of keypoints
- **loss** (*return*) –return\_loss=True for training, return\_loss=False for validation & test.
- **return\_heatmap** (*bool*) –Option to return heatmap.

返回 if 'return\_loss' is true, then return losses. Otherwise, return predicted poses, scores, image paths and heatmaps.

返回类型 dicttuple

### **forward\_dummy** (*img*)

Used for computing network FLOPs.

See `tools/get_flops.py`.

参数 **img** (*torch.Tensor*) –# input image.

返回 Outputs.

返回类型 Tensor

### **forward\_test** (*img, img\_metas, return\_heatmap=False, \*\*kwargs*)

Inference the one-stage model.

---

注解:

- Batchsize: N (currently support batchsize = 1)
  - num\_img\_channel: C
  - img\_width: imgW
  - img\_height: imgH
- 

参数

- **flip\_index** (*List (int)*) –
- **aug\_data** (*List (Tensor [NxCximgHximgW])*) –Multi-scale image
- **num\_joints** (*int*) –Number of joints of an instance. **test\_scale\_factor** (*List(float)*): Multi-scale factor
- **base\_size** (*Tuple (int)*) –Base size of image when scale is 1
- **image\_size** (*int*) –Short edge of images when scale is 1
- **heatmap\_size** (*int*) –Short edge of outputs when scale is 1
- **center** (*np.ndarray*) –center of image
- **scale** (*np.ndarray*) –the scale of image
- **skeleton** (*List (List (int))*) –Links of joints

**forward\_train** (*img, heatmaps, masks, offsets, offset\_weights, img metas, \*\*kwargs*)

Forward the bottom-up model and calculate the loss.

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH  
heatmaps weight: W heatmaps height: H max\_num\_people: M

---

参数

- **img** (*torch.Tensor [N, C, imgH, imgW]*) –# input image.
- **targets** (*List (torch.Tensor [N, K, H, W])*) –Multi-scale target heatmaps.
- **masks** (*List (torch.Tensor [N, H, W])*) –Masks of multi-scale target heatmaps
- **joints** (*List (torch.Tensor [N, M, K, 2])*) –Joints of multi-scale target heatmaps for ae loss
- **img\_metas** (*dict*) –Information about val&test By default this includes: - “image\_file” : image path - “aug\_data” : # input - “test\_scale\_factor” : test scale factor - “base\_size” :

base size of # input - “center” : center of image - “scale” : scale of image - “flip\_index” : flip index of keypoints

返回 The total loss for bottom-up

返回类型 dict

**init\_weights** (*pretrained=None*)

Weight initialization for model.

**show\_result** (*img, result, skeleton=None, kpt\_score\_thr=0.3, bbox\_color=None, pose\_kpt\_color=None, pose\_link\_color=None, radius=4, thickness=1, font\_scale=0.5, win\_name="", show=False, show\_keypoint\_weight=False, wait\_time=0, out\_file=None*)

Draw result over *img*.

### 参数

- **img** (*str or Tensor*) –The image to be displayed.
- **result** (*list[dict]*) –The results to draw over *img* (bbox\_result, pose\_result).
- **skeleton** (*list[list]*) –The connection of keypoints. skeleton is 0-based indexing.
- **kpt\_score\_thr** (*float, optional*) –Minimum score of keypoints to be shown. Default: 0.3.
- **pose\_kpt\_color** (*np.array[Nx3]*) –Color of N keypoints. If None, do not draw keypoints.
- **pose\_link\_color** (*np.array[Mx3]*) –Color of M links. If None, do not draw links.
- **radius** (*int*) –Radius of circles.
- **thickness** (*int*) –Thickness of lines.
- **font\_scale** (*float*) –Font scales of texts.
- **win\_name** (*str*) –The window name.
- **show** (*bool*) –Whether to show the image. Default: False.
- **show\_keypoint\_weight** (*bool*) –Whether to change the transparency using the predicted confidence scores of keypoints.
- **wait\_time** (*int*) –Value of waitKey param. Default: 0.
- **out\_file** (*str or None*) –The filename to write the image. Default: None.

返回 Visualized image only if not *show* or *out\_file*

返回类型 Tensor

**property with\_keypoint**

Check if has keypoint\_head.

```
class mmpose.models.detectors.GestureRecognizer (backbone, neck=None, cls_head=None,  
train_cfg=None, test_cfg=None,  
modality='rgb', pretrained=None)
```

Hand gesture recognizer.

#### 参数

- **backbone** (*dict*) –Backbone modules to extract feature.
- **neck** (*dict*) –Neck Modules to process feature.
- **cls\_head** (*dict*) –Classification head to process feature.
- **train\_cfg** (*dict*) –Config for training. Default: None.
- **test\_cfg** (*dict*) –Config for testing. Default: None.
- **modality** (*str or list or tuple*) –Data modality. Default: None.
- **pretrained** (*str*) –Path to the pretrained models.

**forward** (*video, label=None, img metas=None, return\_loss=True, \*\*kwargs*)

Calls either forward\_train or forward\_test depending on whether return\_loss=True. Note this setting will change the expected inputs.

#### Note:

- batch\_size: N
- num\_vid\_channel: C (Default: 3)
- video height: vidH
- video width: vidW
- video length: vidL

**Args:** video (list[torch.Tensor[NxCxvidLxvidHxvidW]]): Input videos. label (torch.Tensor[N]): Category label of videos. img\_metas (list(dict)): Information about data.

By default this includes: - “fps: video frame rate - “modality” : modality of input videos

**return\_loss (bool):** Option to return loss. *return loss=True* for training, *return loss=False* for validation & test.

**Returns:** dicttuple: if *return loss* is true, then return losses. Otherwise, return predicted gestures for clips with a certain length. .

**forward\_test** (*video, label, img\_metas, \*\*kwargs*)

Defines the computation performed at every call when testing.

**forward\_train** (*video, label, img\_metas, \*\*kwargs*)

Defines the computation performed at every call when training.

**init\_weights** (*pretrained=None*)

Weight initialization for model.

**set\_train\_epoch** (*epoch: int*)

set the training epoch of heads to support customized behaviour.

**show\_result** (*video, result, \*\*kwargs*)

Visualize the results.

**class** mmpose.models.detectors.**Interhand3D** (*backbone, neck=None, keypoint\_head=None, train\_cfg=None, test\_cfg=None, pretrained=None, loss\_pose=None*)

Top-down interhand 3D pose detector of paper ref: Gyeongsik Moon.

“InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image” .  
A child class of TopDown detector.

**forward** (*img, target=None, target\_weight=None, img metas=None, return\_loss=True, \*\*kwargs*)

Calls either forward\_train or forward\_test depending on whether return\_loss=True. Note this setting will change the expected inputs. When *return\_loss=True*, *img* and *img\_meta* are single-nested (i.e. Tensor and List[dict]), and when *return\_loss=False*, *img* and *img\_meta* should be double nested (i.e. list[Tensor], list[list[dict]]), with the outer list indicating test time augmentations.

---

#### 注解:

- batch\_size: N
  - num\_keypoints: K
  - num\_img\_channel: C (Default: 3)
  - img height: imgH
  - img width: imgW
  - heatmaps height: H
  - heatmaps weight: W
- 

#### 参数

- **img** (*torch.Tensor[NxCximgHximgW]*) –Input images.
- **target** (*list[torch.Tensor]*) –Target heatmaps, relative hand
- **depth and hand type.** (*root*) –
- **target\_weight** (*list[torch.Tensor]*) –Weights for target
- **heatmaps** –



- **hand root depth and hand type.** (*relative*) –
- **img metas** (*list(dict)*) – Information about data augmentation By default this includes:
  - "image\_file": path to the image file
  - "center": center of the bbox
  - "scale": scale of the bbox
  - "rotation": rotation of the bbox
  - "bbox\_score": score of bbox
  - "heatmap3d\_depth\_bound": depth bound of hand keypoint 3D heatmap
  - "root\_depth\_bound": depth bound of relative root depth 1D heatmap
- **return\_loss** (*bool*) – Option to return loss. *return loss=True* for training, *return loss=False* for validation & test.

**返回** if *return loss* is true, then return losses. Otherwise, return predicted poses, boxes, image paths, heatmaps, relative hand root depth and hand type.

**返回类型** dicttuple

**forward\_test** (*img, img\_metas, \*\*kwargs*)

Defines the computation performed at every call when testing.

**show\_result** (*result, img=None, skeleton=None, kpt\_score\_thr=0.3, radius=8, bbox\_color='green', thickness=2, pose\_kpt\_color=None, pose\_link\_color=None, vis\_height=400, num\_instances=-1, axis\_azimuth=-115, win\_name="", show=False, wait\_time=0, out\_file=None*)

Visualize 3D pose estimation results.

**参数**

- **result** (*list(dict)*) – The pose estimation results containing:
  - "keypoints\_3d" ([K,4]): 3D keypoints
  - "keypoints" ([K,3] or [T,K,3]): **Optional for visualizing** 2D inputs. If a sequence is given, only the last frame will be used for visualization
  - "bbox" ([4,] or [T,4]): **Optional for visualizing** 2D inputs
  - "title" (str): title for the subplot
- **img** (*str or Tensor*) – Optional. The image to visualize 2D inputs on.
- **skeleton** (*list of [idx\_i, idx\_j]*) – Skeleton described by a list of links, each is a pair of joint indices.
- **kpt\_score\_thr** (*float, optional*) – Minimum score of keypoints to be shown. Default: 0.3.

- **radius** (*int*) –Radius of circles.
- **bbox\_color** (*str* or *tuple* or *Color*) –Color of bbox lines.
- **thickness** (*int*) –Thickness of lines.
- **pose\_kpt\_color** (*np.array[Nx3]* `) –Color of N keypoints. If None, do not draw keypoints.
- **pose\_link\_color** (*np.array[Mx3]*) –Color of M limbs. If None, do not draw limbs.
- **vis\_height** (*int*) –The image height of the visualization. The width will be  $N \times \text{vis\_height}$  depending on the number of visualized items.
- **num\_instances** (*int*) –Number of instances to be shown in 3D. If smaller than 0, all the instances in the pose\_result will be shown. Otherwise, pad or truncate the pose\_result to a length of num\_instances.
- **axis\_azimuth** (*float*) –axis azimuth angle for 3D visualizations.
- **win\_name** (*str*) –The window name.
- **show** (*bool*) –Whether to show the image. Default: False.
- **wait\_time** (*int*) –Value of waitKey param. Default: 0.
- **out\_file** (*str* or *None*) –The filename to write the image. Default: None.

返回 Visualized img, only if not *show* or *out\_file*.

返回类型 Tensor

```
class mmpose.models.detectors.MultiTask (backbone, heads, necks=None, head2neck=None, pretrained=None)
```

Multi-task detectors.

#### 参数

- **backbone** (*dict*) –Backbone modules to extract feature.
- **heads** (*list[dict]*) –heads to output predictions.
- **necks** (*list[dict]* | *None*) –necks to process feature.
- **(dict{int (head2neck) –int})**: head index to neck index.
- **pretrained** (*str*) –Path to the pretrained models.

**forward** (*img*, *target=None*, *target\_weight=None*, *img metas=None*, *return\_loss=True*, *\*\*kwargs*)

Calls either forward\_train or forward\_test depending on whether return\_loss=True. Note this setting will change the expected inputs. When *return\_loss=True*, *img* and *img\_meta* are single-nested (i.e. Tensor and List[dict]), and when *return\_loss=False*, *img* and *img\_meta* should be double nested (i.e. List[Tensor], List[List[dict]]), with the outer list indicating test time augmentations.

注解:

- batch\_size: N
- num\_keypoints: K
- num\_img\_channel: C (Default: 3)
- img height: imgH
- img weight: imgW
- heatmaps height: H
- heatmaps weight: W

参数

- **img** (*torch.Tensor*[N, C, imgH, imgW]) –Input images.
- **target** (*list*[*torch.Tensor*]) –Targets.
- **target\_weight** (*List*[*torch.Tensor*]) –Weights.
- **img metas** (*list*(*dict*)) –Information about data augmentation By default this includes:
  - " image\_file": path to the image file
  - " center" : center of the bbox
  - " scale" : scale of the bbox
  - " rotation" : rotation of the bbox
  - " bbox\_score" : score of bbox
- **return\_loss** (*bool*) –Option to *return loss*. *return loss=True* for training, *return loss=False* for validation & test.

**返回** if *return loss* is true, then return losses. Otherwise, return predicted poses, boxes, image paths and heatmaps.

**返回类型** dicttuple

**forward\_dummy** (*img*)

Used for computing network FLOPs.

See `tools/get_flops.py`.

**参数** **img** (*torch.Tensor*) –Input image.

**返回** Outputs.

返回类型 `list[Tensor]`

**forward\_test** (*img, img metas, \*\*kwargs*)

Defines the computation performed at every call when testing.

**forward\_train** (*img, target, target\_weight, img metas, \*\*kwargs*)

Defines the computation performed at every call when training.

**init\_weights** (*pretrained=None*)

Weight initialization for model.

**property with\_necks**

Check if has keypoint\_head.

**class** `mmpose.models.detectors.ParametricMesh` (*backbone, mesh\_head, smpl, disc=None, loss\_gan=None, loss\_mesh=None, train\_cfg=None, test\_cfg=None, pretrained=None*)

Model-based 3D human mesh detector. Take a single color image as input and output 3D joints, SMPL parameters and camera parameters.

#### 参数

- **backbone** (*dict*) –Backbone modules to extract feature.
- **mesh\_head** (*dict*) –Mesh head to process feature.
- **smpl** (*dict*) –Config for SMPL model.
- **disc** (*dict*) –Discriminator for SMPL parameters. Default: None.
- **loss\_gan** (*dict*) –Config for adversarial loss. Default: None.
- **loss\_mesh** (*dict*) –Config for mesh loss. Default: None.
- **train\_cfg** (*dict*) –Config for training. Default: None.
- **test\_cfg** (*dict*) –Config for testing. Default: None.
- **pretrained** (*str*) –Path to the pretrained models.

**forward** (*img, img metas=None, return\_loss=False, \*\*kwargs*)

Forward function.

Calls either `forward_train` or `forward_test` depending on whether `return_loss=True`.

---

#### 注解:

- `batch_size`: N
- `num_img_channel`: C (Default: 3)
- `img height`: imgH
- `img width`: imgW

**参数**

- **img** (*torch.Tensor*[*N* x *C* x *imgH* x *imgW*]) –Input images.
- **img metas** (*list(dict)*) –Information about data augmentation By default this includes:
  - " image\_file": path to the image file
  - " center" : center of the bbox
  - " scale" : scale of the bbox
  - " rotation" : rotation of the bbox
  - " bbox\_score" : score of bbox
- **return\_loss** (*bool*) –Option to *return loss*. *return loss=True* for training, *return loss=False* for validation & test.

**返回** Return predicted 3D joints, SMPL parameters, boxes and image paths.

**forward\_dummy** (*img*)

Used for computing network FLOPs.

See `tools/get_flops.py`.

**参数** **img** (*torch.Tensor*) –Input image.

**返回** Outputs.

**返回类型** Tensor

**forward\_test** (*img, img\_metas, return\_vertices=False, return\_faces=False, \*\*kwargs*)

Defines the computation performed at every call when testing.

**forward\_train** (*\*args, \*\*kwargs*)

Forward function for training.

For ParametricMesh, we do not use this interface.

**get\_3d\_joints\_from\_mesh** (*vertices*)

Get 3D joints from 3D mesh using predefined joints regressor.

**init\_weights** (*pretrained=None*)

Weight initialization for model.

**show\_result** (*result, img, show=False, out\_file=None, win\_name="", wait\_time=0, bbox\_color='green', mesh\_color=(76, 76, 204), \*\*kwargs*)

Visualize 3D mesh estimation results.

**参数**

- **result** (*list[dict]*) –The mesh estimation results containing:
  - "bbox" (*ndarray[4]*): instance bounding bbox
  - "center" (*ndarray[2]*): bbox center
  - "scale" (*ndarray[2]*): bbox scale
  - "keypoints\_3d" (*ndarray[K,3]*): predicted 3D keypoints
  - "camera" (*ndarray[3]*): camera parameters
  - "vertices" (*ndarray[V, 3]*): predicted 3D vertices
  - "faces" (*ndarray[F, 3]*): mesh faces
- **img** (*str or Tensor*) –Optional. The image to visualize 2D inputs on.
- **win\_name** (*str*) –The window name.
- **show** (*bool*) –Whether to show the image. Default: False.
- **wait\_time** (*int*) –Value of waitKey param. Default: 0.
- **out\_file** (*str or None*) –The filename to write the image. Default: None.
- **bbox\_color** (*str or tuple or Color*) –Color of bbox lines.
- **mesh\_color** (*str or tuple or Color*) –Color of mesh surface.

返回 Visualized img, only if not *show* or *out\_file*.

返回类型 *ndarray*

**train\_step** (*data\_batch, optimizer, \*\*kwargs*)

Train step function.

In this function, the detector will finish the train step following the pipeline:

1. get fake and real SMPL parameters
2. optimize discriminator (if have)
3. optimize generator

If *self.train\_cfg.disc\_step > 1*, the train step will contain multiple iterations for optimizing discriminator with different input data and only one iteration for optimizing generator after *disc\_step* iterations for discriminator.

参数

- **data\_batch** (*torch.Tensor*) –Batch of data as input.
- **optimizer** (*dict[torch.optim.Optimizer]*) –Dict with optimizers for generator and discriminator (if have).

返回 Dict with loss, information for logger, the number of samples.

返回类型 *outputs (dict)*

**val\_step** (*data\_batch*, *\*\*kwargs*)

Forward function for evaluation.

**参数** **data\_batch** (*dict*) –Contain data for forward.

**返回** Contain the results from model.

**返回类型** *dict*

**class** `mmpose.models.detectors.PoseLifter` (*backbone*, *neck=None*, *keypoint\_head=None*,  
*traj\_backbone=None*, *traj\_neck=None*, *traj\_head=None*,  
*loss\_semi=None*, *train\_cfg=None*, *test\_cfg=None*,  
*pretrained=None*)

Pose lifter that lifts 2D pose to 3D pose.

The basic model is a pose model that predicts root-relative pose. If *traj\_head* is not *None*, a trajectory model that predicts absolute root joint position is also built.

**参数**

- **backbone** (*dict*) –Config for the backbone of pose model.
- **neck** (*dict/None*) –Config for the neck of pose model.
- **keypoint\_head** (*dict/None*) –Config for the head of pose model.
- **traj\_backbone** (*dict/None*) –Config for the backbone of trajectory model. If *traj\_backbone* is *None* and *traj\_head* is not *None*, trajectory model will share backbone with pose model.
- **traj\_neck** (*dict/None*) –Config for the neck of trajectory model.
- **traj\_head** (*dict/None*) –Config for the head of trajectory model.
- **loss\_semi** (*dict/None*) –Config for semi-supervision loss.
- **train\_cfg** (*dict/None*) –Config for keypoint head during training.
- **test\_cfg** (*dict/None*) –Config for keypoint head during testing.
- **pretrained** (*str/None*) –Path to pretrained weights.

**forward** (*input*, *target=None*, *target\_weight=None*, *metas=None*, *return\_loss=True*, *\*\*kwargs*)

Calls either `forward_train` or `forward_test` depending on whether *return\_loss=True*.

---

**注解:**

- *batch\_size*: N
- *num\_input\_keypoints*: Ki
- *input\_keypoint\_dim*: Ci
- *input\_sequence\_len*: Ti

- `num_output_keypoints`: Ko
  - `output_keypoint_dim`: Co
  - `input_sequence_len`: To
- 

### 参数

- **input** (*torch.Tensor*[NxKixCixTi]) –Input keypoint coordinates.
- **target** (*torch.Tensor*[NxKoxCoxTo]) –Output keypoint coordinates. Defaults to None.
- **target\_weight** (*torch.Tensor*[NxKox1]) –Weights across different joint types. Defaults to None.
- **metas** (*list(dict)*) –Information about data augmentation
- **return\_loss** (*bool*) –Option to *return loss*. *return loss=True* for training, *return loss=False* for validation & test.

返回 If *reutrn\_loss* is true, return losses. Otherwise return predicted poses.

返回类型 dict|Tensor

**forward\_dummy** (*input*)

Used for computing network FLOPs. See `tools/get_flops.py`.

参数 **input** (*torch.Tensor*) –Input pose

返回 Model output

返回类型 Tensor

**forward\_test** (*input, metas, \*\*kwargs*)

Defines the computation performed at every call when training.

**forward\_train** (*input, target, target\_weight, metas, \*\*kwargs*)

Defines the computation performed at every call when training.

**init\_weights** (*pretrained=None*)

Weight initialization for model.

**show\_result** (*result, img=None, skeleton=None, pose\_kpt\_color=None, pose\_link\_color=None, radius=8, thickness=2, vis\_height=400, num\_instances=-1, axis\_azimuth=70, win\_name="", show=False, wait\_time=0, out\_file=None*)

Visualize 3D pose estimation results.

### 参数

- **result** (*list(dict)*) –The pose estimation results containing:
  - "keypoints\_3d" ([K,4]): 3D keypoints



- “keypoints” ([K,3] or [T,K,3]): Optional for visualizing 2D inputs. If a sequence is given, only the last frame will be used for visualization
- “bbox” ([4,] or [T,4]): Optional for visualizing 2D inputs
- “title” (str): title for the subplot
- **img** (str or Tensor) –Optional. The image to visualize 2D inputs on.
- **skeleton** (list of [idx\_i, idx\_j]) –Skeleton described by a list of links, each is a pair of joint indices.
- **pose\_kpt\_color** (np.array[Nx3]) –Color of N keypoints. If None, do not draw keypoints.
- **pose\_link\_color** (np.array[Mx3]) –Color of M links. If None, do not draw links.
- **radius** (int) –Radius of circles.
- **thickness** (int) –Thickness of lines.
- **vis\_height** (int) –The image height of the visualization. The width will be N\*vis\_height depending on the number of visualized items.
- **num\_instances** (int) –Number of instances to be shown in 3D. If smaller than 0, all the instances in the result will be shown. Otherwise, pad or truncate the result to a length of num\_instances.
- **axis\_azimuth** (float) –axis azimuth angle for 3D visualizations.
- **win\_name** (str) –The window name.
- **show** (bool) –Whether to directly show the visualization.
- **wait\_time** (int) –Value of waitKey param. Default: 0.
- **out\_file** (str or None) –The filename to write the image. Default: None.

返回 Visualized img, only if not *show* or *out\_file*.

返回类型 Tensor

**property with\_keypoint**

Check if has keypoint\_head.

**property with\_neck**

Check if has keypoint\_neck.

**property with\_traj**

Check if has trajectory\_head.

**property with\_traj\_backbone**

Check if has trajectory\_backbone.

**property with\_traj\_neck**

Check if has trajectory\_neck.

```
class mmpose.models.detectors.PoseWarper (backbone, neck=None, keypoint_head=None,  
                                           train_cfg=None, test_cfg=None, pretrained=None,  
                                           loss_pose=None, concat_tensors=True)
```

Top-down pose detectors for multi-frame settings for video inputs.

“[Learning temporal pose estimation from sparsely-labeled videos](#)” .

A child class of TopDown detector. The main difference between PoseWarper and TopDown lies in that the former takes a list of tensors as input image while the latter takes a single tensor as input image in forward method.

**参数**

- **backbone** (*dict*) –Backbone modules to extract features.
- **neck** (*dict*) –intermediate modules to transform features.
- **keypoint\_head** (*dict*) –Keypoint head to process feature.
- **train\_cfg** (*dict*) –Config for training. Default: None.
- **test\_cfg** (*dict*) –Config for testing. Default: None.
- **pretrained** (*str*) –Path to the pretrained models.
- **loss\_pose** (*None*) –Deprecated arguments. Please use *loss\_keypoint* for heads instead.
- **concat\_tensors** (*bool*) –Whether to concat the tensors on the batch dim, which can speed up, Default: True

```
forward (img, target=None, target_weight=None, img metas=None, return_loss=True, return_heatmap=False,  
         **kwargs)
```

Calls either forward\_train or forward\_test depending on whether return\_loss=True. Note this setting will change the expected inputs. When *return\_loss=True*, *img* and *img\_meta* are single-nested (i.e. Tensor and List[dict]), and when *return\_loss=False*, *img* and *img\_meta* should be double nested (i.e. List[Tensor], List[List[dict]]), with the outer list indicating test time augmentations.

---

**注解:**

- number of frames: F
- batch\_size: N
- num\_keypoints: K
- num\_img\_channel: C (Default: 3)
- img height: imgH
- img width: imgW

- heatmaps height: H
- heatmaps weight: W

### 参数

- **imgs** (*list* [*F*, *torch.Tensor* [*N*, *C*, *imgH*, *imgW*]]) –multiple input frames
- **target** (*torch.Tensor* [*N*, *K*, *H*, *W*]) –Target heatmaps for one frame.
- **target\_weight** (*torch.Tensor* [*N*, *K*, *1*]) –Weights across different joint types.
- **img metas** (*list* (*dict*)) –Information about data augmentation By default this includes:
  - " image\_file": paths to multiple video frames
  - " center" : center of the bbox
  - " scale" : scale of the bbox
  - " rotation" : rotation of the bbox
  - " bbox\_score" : score of bbox
- **return\_loss** (*bool*) –Option to *return loss*. *return loss=True* for training, *return loss=False* for validation & test.
- **return\_heatmap** (*bool*) –Option to return heatmap.

**返回** if *return loss* is true, then return losses. Otherwise, return predicted poses, boxes, image paths and heatmaps.

**返回类型** dicttuple

### **forward\_dummy** (*img*)

Used for computing network FLOPs.

See `tools/get_flops.py`.

**参数** **img** (*torch.Tensor* [*N*, *C*, *imgH*, *imgW*], or *list/tuple of tensors*)  
–multiple input frames, *N* >= 2.

**返回** Output heatmaps.

**返回类型** Tensor

### **forward\_test** (*imgs*, *img\_metas*, *return\_heatmap=False*, *\*\*kwargs*)

Defines the computation performed at every call when testing.

### **forward\_train** (*imgs*, *target*, *target\_weight*, *img\_metas*, *\*\*kwargs*)

Defines the computation performed at every call when training.

```
class mmpose.models.detectors.TopDown (backbone, neck=None, keypoint_head=None,  
                                         train_cfg=None, test_cfg=None, pretrained=None,  
                                         loss_pose=None)
```

Top-down pose detectors.

#### 参数

- **backbone** (*dict*) –Backbone modules to extract feature.
- **keypoint\_head** (*dict*) –Keypoint head to process feature.
- **train\_cfg** (*dict*) –Config for training. Default: None.
- **test\_cfg** (*dict*) –Config for testing. Default: None.
- **pretrained** (*str*) –Path to the pretrained models.
- **loss\_pose** (*None*) –Deprecated arguments. Please use *loss\_keypoint* for heads instead.

```
forward (img, target=None, target_weight=None, img metas=None, return_loss=True, return_heatmap=False,  
         **kwargs)
```

Calls either `forward_train` or `forward_test` depending on whether `return_loss=True`. Note this setting will change the expected inputs. When *return\_loss=True*, *img* and *img\_meta* are single-nested (i.e. Tensor and List[dict]), and when *return\_loss=False*, *img* and *img\_meta* should be double nested (i.e. List[Tensor], List[List[dict]]), with the outer list indicating test time augmentations.

---

#### 注解:

- batch\_size: N
- num\_keypoints: K
- num\_img\_channel: C (Default: 3)
- img height: imgH
- img width: imgW
- heatmaps height: H
- heatmaps weight: W

---

#### 参数

- **img** (*torch.Tensor[NxCximgHximgW]*) –Input images.
- **target** (*torch.Tensor[NxKxHxW]*) –Target heatmaps.
- **target\_weight** (*torch.Tensor[NxKx1]*) –Weights across different joint types.

- **img metas** (*list(dict)*) –Information about data augmentation By default this includes:
  - " image\_file: path to the image file
  - " center" : center of the bbox
  - " scale" : scale of the bbox
  - " rotation" : rotation of the bbox
  - " bbox\_score" : score of bbox
- **return\_loss** (*bool*) –Option to *return loss*. *return loss=True* for training, *return loss=False* for validation & test.
- **return\_heatmap** (*bool*) –Option to return heatmap.

**返回** if *return loss* is true, then return losses. Otherwise, return predicted poses, boxes, image paths and heatmaps.

**返回类型** dicttuple

**forward\_dummy** (*img*)

Used for computing network FLOPs.

See `tools/get_flops.py`.

**参数** **img** (*torch.Tensor*) –Input image.

**返回** Output heatmaps.

**返回类型** Tensor

**forward\_test** (*img, img\_metas, return\_heatmap=False, \*\*kwargs*)

Defines the computation performed at every call when testing.

**forward\_train** (*img, target, target\_weight, img\_metas, \*\*kwargs*)

Defines the computation performed at every call when training.

**init\_weights** (*pretrained=None*)

Weight initialization for model.

**show\_result** (*img, result, skeleton=None, kpt\_score\_thr=0.3, bbox\_color='green', pose\_kpt\_color=None, pose\_link\_color=None, text\_color='white', radius=4, thickness=1, font\_scale=0.5, bbox\_thickness=1, win\_name="", show=False, show\_keypoint\_weight=False, wait\_time=0, out\_file=None*)

Draw *result* over *img*.

**参数**

- **img** (*str or Tensor*) –The image to be displayed.
- **result** (*list[dict]*) –The results to draw over *img* (bbox\_result, pose\_result).

- **skeleton** (*list[list]*) –The connection of keypoints. skeleton is 0-based indexing.
- **kpt\_score\_thr** (*float, optional*) –Minimum score of keypoints to be shown. Default: 0.3.
- **bbox\_color** (*str or tuple or Color*) –Color of bbox lines.
- **pose\_kpt\_color** (*np.array[Nx3]*) –Color of N keypoints. If None, do not draw keypoints.
- **pose\_link\_color** (*np.array[Mx3]*) –Color of M links. If None, do not draw links.
- **text\_color** (*str or tuple or Color*) –Color of texts.
- **radius** (*int*) –Radius of circles.
- **thickness** (*int*) –Thickness of lines.
- **font\_scale** (*float*) –Font scales of texts.
- **win\_name** (*str*) –The window name.
- **show** (*bool*) –Whether to show the image. Default: False.
- **show\_keypoint\_weight** (*bool*) –Whether to change the transparency using the predicted confidence scores of keypoints.
- **wait\_time** (*int*) –Value of waitKey param. Default: 0.
- **out\_file** (*str or None*) –The filename to write the image. Default: None.

返回 Visualized img, only if not *show* or *out\_file*.

返回类型 Tensor

**property with\_keypoint**

Check if has keypoint\_head.

**property with\_neck**

Check if has neck.

```
class mmpose.models.detectors.VoxelCenterDetector (image_size, heatmap_size, space_size,  
                                                    cube_size, space_center, center_net,  
                                                    center_head, train_cfg=None,  
                                                    test_cfg=None)
```

Detect human center by 3D CNN on voxels.

Please refer to the *paper* <<https://arxiv.org/abs/2004.06239>> for details. :param image\_size: input size of the 2D model. :type image\_size: list :param heatmap\_size: output size of the 2D model. :type heatmap\_size: list :param space\_size: Size of the 3D space. :type space\_size: list :param cube\_size: Size of the input volume to the 3D CNN. :type cube\_size: list :param space\_center: Coordinate of the center of the 3D space. :type space\_center: list :param

center\_net: Dictionary to construct the center net. :type center\_net: ConfigDict :param center\_head: Dictionary to construct the center head. :type center\_head: ConfigDict :param train\_cfg: Config for training. Default: None. :type train\_cfg: ConfigDict :param test\_cfg: Config for testing. Default: None. :type test\_cfg: ConfigDict

**assign2gt** (*center\_candidates, gt\_centers, gt\_num\_persons*)

“Assign gt id to each valid human center candidate.

**forward** (*img, img metas, return\_loss=True, feature\_maps=None, targets\_3d=None*)

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH  
heatmaps width: W heatmaps height: H

---

### 参数

- **img** (*list(torch.Tensor[NxCximgHximgW])*) –Multi-camera input images to the 2D model.
- **img\_metas** (*list(dict)*) –Information about image, 3D groundtruth and camera parameters.
- **return\_loss** –Option to *return loss*. *return loss=True* for training, *return loss=False* for validation & test.
- **targets\_3d** (*torch.Tensor[NxcubeLxcubeWxcubeH]*) –Ground-truth 3D heatmap of human centers.
- **feature\_maps** (*list(torch.Tensor[NxKxHxW])*) –Multi-camera feature\_maps.

### 返回

if ‘return\_loss’ is true, then return losses. Otherwise, return predicted poses

返回类型 dict

**forward\_dummy** (*feature\_maps*)

Used for computing network FLOPs.

**forward\_test** (*img, img metas, feature\_maps=None*)

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH  
heatmaps width: W heatmaps height: H

---

### 参数

- **img** (*list (torch.Tensor [NxCximgHximgW])*) –Multi-camera input images to the 2D model.
- **img metas** (*list (dict)*) –Information about image, 3D groundtruth and camera parameters.
- **feature\_maps** (*list (torch.Tensor [NxKxHxW])*) –Multi-camera feature\_maps.

返回 human centers

**forward\_train** (*img, img\_metas, feature\_maps=None, targets\_3d=None, return\_preds=False*)

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH  
heatmaps width: W heatmaps height: H

---

#### 参数

- **img** (*list (torch.Tensor [NxCximgHximgW])*) –Multi-camera input images to the 2D model.
- **img metas** (*list (dict)*) –Information about image, 3D groundtruth and camera parameters.
- **targets\_3d** (*torch.Tensor [NxcubeLxcubeWxcubeH]*) –Ground-truth 3D heatmap of human centers.
- **feature\_maps** (*list (torch.Tensor [NxKxHxW])*) –Multi-camera feature\_maps.
- **return\_preds** (*bool*) –Whether to return prediction results

#### 返回

if ‘return\_pred’ is true, then return losses and human centers. Otherwise, return losses only

返回类型 dict

**show\_result** (*\*\*kwargs*)

Visualize the results.

**class** mmpose.models.detectors.**VoxelSinglePose** (*image\_size, heatmap\_size, sub\_space\_size, sub\_cube\_size, num\_joints, pose\_net, pose\_head, train\_cfg=None, test\_cfg=None*)

VoxelPose Please refer to the paper <<https://arxiv.org/abs/2004.06239>> for details.

#### 参数



- **image\_size** (*list*) –input size of the 2D model.
- **heatmap\_size** (*list*) –output size of the 2D model.
- **sub\_space\_size** (*list*) –Size of the cuboid human proposal.
- **sub\_cube\_size** (*list*) –Size of the input volume to the pose net.
- **pose\_net** (*ConfigDict*) –Dictionary to construct the pose net.
- **pose\_head** (*ConfigDict*) –Dictionary to construct the pose head.
- **train\_cfg** (*ConfigDict*) –Config for training. Default: None.
- **test\_cfg** (*ConfigDict*) –Config for testing. Default: None.

**forward** (*img, img metas, return\_loss=True, feature\_maps=None, human\_candidates=None, \*\*kwargs*)

---

注解: batch\_size: N num\_keypoints: K num\_img\_channel: C img\_width: imgW img\_height: imgH feature\_maps width: W feature\_maps height: H volume\_length: cubeL volume\_width: cubeW volume\_height: cubeH

---

### 参数

- **img** (*list (torch.Tensor [NxCximgHximgW])*) –Multi-camera input images to the 2D model.
- **feature\_maps** (*list (torch.Tensor [NxCxHxW])*) –Multi-camera input feature\_maps.
- **img\_metas** (*list (dict)*) –Information about image, 3D groundtruth and camera parameters.
- **human\_candidates** (*torch.Tensor [NxPx5]*) –Human candidates.
- **return\_loss** –Option to *return loss*. *return loss=True* for training, *return loss=False* for validation & test.

**forward\_dummy** (*feature\_maps, num\_candidates=5*)

Used for computing network FLOPs.

**forward\_test** (*img, img metas, feature\_maps=None, human\_candidates=None, \*\*kwargs*)

Defines the computation performed at training. .. note:

```
batch_size: N
num_keypoints: K
num_img_channel: C
img_width: imgW
```

(下页继续)

(续上页)

```

img_height: imgH
feature_maps width: W
feature_maps height: H
volume_length: cubeL
volume_width: cubeW
volume_height: cubeH

```

### 参数

- **img** (*list(torch.Tensor[NxCximgHximgW])*) –Multi-camera input images to the 2D model.
- **feature\_maps** (*list(torch.Tensor[NxCxHxW])*) –Multi-camera input feature\_maps.
- **img metas** (*list(dict)*) –Information about image, 3D groundtruth and camera parameters.
- **human\_candidates** (*torch.Tensor[NxPx5]*) –Human candidates.

**返回** predicted poses, human centers and sample\_id

**返回类型** dict

**forward\_train** (*img, img\_metas, feature\_maps=None, human\_candidates=None, return\_preds=False, \*\*kwargs*)

Defines the computation performed at training. .. note:

```

batch_size: N
num_keypoints: K
num_img_channel: C
img_width: imgW
img_height: imgH
feature_maps width: W
feature_maps height: H
volume_length: cubeL
volume_width: cubeW
volume_height: cubeH

```

### 参数

- **img** (*list(torch.Tensor[NxCximgHximgW])*) –Multi-camera input images to the 2D model.
- **feature\_maps** (*list(torch.Tensor[NxCxHxW])*) –Multi-camera input feature\_maps.

- **img metas** (*list(dict)*) –Information about image, 3D groundtruth and camera parameters.
- **human\_candidates** (*torch.Tensor[NxPx5]*) –Human candidates.
- **return\_preds** (*bool*) –Whether to return prediction results

返回 losses.

返回类型 dict

**show\_result** (*\*\*kwargs*)

Visualize the results.

## 48.4 heads

```
class mmpose.models.heads.AEHigherResolutionHead (in_channels, num_joints,
                                                    tag_per_joint=True, extra=None,
                                                    num_deconv_layers=1,
                                                    num_deconv_filters=(32),
                                                    num_deconv_kernels=(4),
                                                    num_basic_blocks=4, cat_output=None,
                                                    with_ae_loss=None, loss_keypoint=None)
```

Associative embedding with higher resolution head. paper ref: Bowen Cheng et al. “HigherHRNet: Scale-Aware Representation Learning for Bottom- Up Human Pose Estimation” .

参数

- **in\_channels** (*int*) –Number of input channels.
- **num\_joints** (*int*) –Number of joints
- **tag\_per\_joint** (*bool*) –If tag\_per\_joint is True, the dimension of tags equals to num\_joints, else the dimension of tags is 1. Default: True
- **extra** (*dict*) –Configs for extra conv layers. Default: None
- **num\_deconv\_layers** (*int*) –Number of deconv layers. num\_deconv\_layers should  $\geq 0$ . Note that 0 means no deconv layers.
- **num\_deconv\_filters** (*list/tuple*) –Number of filters. If num\_deconv\_layers  $> 0$ , the length of
- **num\_deconv\_kernels** (*list/tuple*) –Kernel sizes.
- **cat\_output** (*list[bool]*) –Option to concat outputs.
- **with\_ae\_loss** (*list[bool]*) –Option to use ae loss.
- **loss\_keypoint** (*dict*) –Config for loss. Default: None.

**forward** (*x*)

Forward function.

**get\_loss** (*outputs, targets, masks, joints*)

Calculate bottom-up keypoint loss.

---

注解:

- **batch\_size**: N
  - **num\_keypoints**: K
  - **num\_outputs**: O
  - **heatmaps height**: H
  - **heatmaps weight**: W
- 

参数

- **outputs** (*list (torch.Tensor[N, K, H, W])*) –Multi-scale output heatmaps.
- **targets** (*List (torch.Tensor[N, K, H, W])*) –Multi-scale target heatmaps.
- **masks** (*List (torch.Tensor[N, H, W])*) –Masks of multi-scale target heatmaps
- **joints** (*List (torch.Tensor[N, M, K, 2])*) –Joints of multi-scale target heatmaps for ae loss

**init\_weights** ()

Initialize model weights.

```
class mmpose.models.heads.AEMultiStageHead (in_channels, out_channels, num_stages=1,  
                                           num_deconv_layers=3, num_deconv_filters=(256,  
                                           256, 256), num_deconv_kernels=(4, 4, 4),  
                                           extra=None, loss_keypoint=None)
```

Associative embedding multi-stage head. paper ref: Alejandro Newell et al. “Associative Embedding: End-to-end Learning for Joint Detection and Grouping”

参数

- **in\_channels** (*int*) –Number of input channels.
- **out\_channels** (*int*) –Number of output channels.
- **num\_deconv\_layers** (*int*) –Number of deconv layers. `num_deconv_layers` should  $\geq 0$ . Note that 0 means no deconv layers.

- **num\_deconv\_filters** (*list/tuple*) –Number of filters. If num\_deconv\_layers > 0, the length of
- **num\_deconv\_kernels** (*list/tuple*) –Kernel sizes.
- **loss\_keypoint** (*dict*) –Config for loss. Default: None.

**forward** (*x*)

Forward function.

返回 a list of heatmaps from multiple stages.

返回类型 out (list[*Tensor*])

**get\_loss** (*output, targets, masks, joints*)

Calculate bottom-up keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*List (torch.Tensor [NxKxHxW])*) –Output heatmaps.
- **targets** (*List (List (torch.Tensor [NxKxHxW]))*) –Multi-stage and multi-scale target heatmaps.
- **masks** (*List (List (torch.Tensor [NxHxW]))*) –Masks of multi-stage and multi-scale target heatmaps
- **joints** (*List (List (torch.Tensor [NxMxKx2]))*) –Joints of multi-stage multi-scale target heatmaps for ae loss

**init\_weights** ()

Initialize model weights.

```
class mmpose.models.heads.AESimpleHead (in_channels, num_joints, num_deconv_layers=3,  
                                         num_deconv_filters=(256, 256, 256),  
                                         num_deconv_kernels=(4, 4, 4), tag_per_joint=True,  
                                         with_ae_loss=None, extra=None, loss_keypoint=None)
```

Associative embedding simple head. paper ref: Alejandro Newell et al. “Associative Embedding: End-to-end Learning for Joint Detection and Grouping”

### 参数

- **in\_channels** (*int*) –Number of input channels.
- **num\_joints** (*int*) –Number of joints.
- **num\_deconv\_layers** (*int*) –Number of deconv layers. num\_deconv\_layers should  $\geq 0$ . Note that 0 means no deconv layers.
- **num\_deconv\_filters** (*list/tuple*) –Number of filters. If num\_deconv\_layers  $> 0$ , the length of
- **num\_deconv\_kernels** (*list/tuple*) –Kernel sizes.
- **tag\_per\_joint** (*bool*) –If tag\_per\_joint is True, the dimension of tags equals to num\_joints, else the dimension of tags is 1. Default: True
- **with\_ae\_loss** (*list[bool]*) –Option to use ae loss or not.
- **loss\_keypoint** (*dict*) –Config for loss. Default: None.

**get\_loss** (*outputs, targets, masks, joints*)

Calculate bottom-up keypoint loss.

---

### 注解:

- batch\_size: N
  - num\_keypoints: K
  - num\_outputs: O
  - heatmaps height: H
  - heatmaps weight: W
- 

### 参数

- **outputs** (*list(torch.Tensor[N, K, H, W])*) –Multi-scale output heatmaps.
- **targets** (*List(torch.Tensor[N, K, H, W])*) –Multi-scale target heatmaps.
- **masks** (*List(torch.Tensor[N, H, W])*) –Masks of multi-scale target heatmaps
- **joints** (*List(torch.Tensor[N, M, K, 2])*) –Joints of multi-scale target heatmaps for ae loss

```
class mmpose.models.heads.CIDHead(in_channels, gfd_channels, num_joints, multi_hm_loss_factor=1.0,
                                   single_hm_loss_factor=4.0, contrastive_loss_factor=1.0,
                                   max_train_instances=200, prior_prob=0.01)
```

CID head. paper ref: Dongkai Wang et al. “Contextual Instance Decouple for Robust Multi-Person Pose Estimation” .

#### 参数

- **in\_channels** (*int*) –Number of input channels.
- **gfd\_channels** (*int*) –Number of instance feature map channels
- **num\_joints** (*int*) –Number of joints
- **multi\_hm\_loss\_factor** (*float*) –loss weight for multi-person heatmap
- **single\_hm\_loss\_factor** (*float*) –loss weight for single person heatmap
- **contrastive\_loss\_factor** (*float*) –loss weight for contrastive loss
- **max\_train\_instances** (*int*) –limit the number of instances
- **training to avoid** (*during*) –
- **prior\_prob** (*float*) –focal loss bias initialization

```
forward (features, forward_info=None)
```

Forward function.

```
init_weights ()
```

Initialize model weights.

```
class mmpose.models.heads.CuboidCenterHead(space_size, space_center, cube_size, max_num=10,
                                             max_pool_kernel=3)
```

Get results from the 3D human center heatmap. In this module, human 3D centers are local maximums obtained from the 3D heatmap via NMS (max- pooling).

#### 参数

- **space\_size** (*list [3]*) –The size of the 3D space.
- **cube\_size** (*list [3]*) –The size of the heatmap volume.
- **space\_center** (*list [3]*) –The coordinate of space center.
- **max\_num** (*int*) –Maximum of human center detections.
- **max\_pool\_kernel** (*int*) –Kernel size of the max-pool kernel in nms.

```
forward (heatmap_volumes)
```

参数 **heatmap\_volumes** (*torch.Tensor (NXLXWXH)*) –3D human center heatmaps predicted by the network.

返回 Coordinates of human centers.

返回类型 human\_centers (torch.Tensor(NXPX5))

```
class mmpose.models.heads.CuboidPoseHead(beta)
```

```
forward(heatmap_volumes, grid_coordinates)
```

#### 参数

- **heatmap\_volumes** (torch.Tensor(NxKxLxWxH)) –3D human pose heatmaps predicted by the network.
- **grid\_coordinates** (torch.Tensor(Nx(LxWxH)x3)) –Coordinates of the grids in the heatmap volumes.

返回 Coordinates of human poses.

返回类型 human\_poses (torch.Tensor(NxKx3))

```
class mmpose.models.heads.DEKRHead(in_channels, num_joints, num_heatmap_filters=32,
                                   num_offset_filters_per_joint=15, in_index=0,
                                   input_transform=None, num_deconv_layers=0,
                                   num_deconv_filters=None, num_deconv_kernels=None,
                                   extra={'final_conv_kernel': 0}, align_corners=False,
                                   heatmap_loss=None, offset_loss=None)
```

DisEntangled Keypoint Regression head. “Bottom-up human pose estimation via disentangled keypoint regression”, CVPR’ 2021.

#### 参数

- **in\_channels** (int) –Number of input channels.
- **num\_joints** (int) –Number of joints.
- **num\_heatmap\_filters** (int) –Number of filters for heatmap branch.
- **num\_offset\_filters\_per\_joint** (int) –Number of filters for each joint.
- **in\_index** (int|Sequence[int]) –Input feature index. Default: 0
- **input\_transform** (str|None) –Transformation type of input features. Options: ‘resize\_concat’, ‘multiple\_select’, None. Default: None.
  - ‘resize\_concat’: Multiple feature maps will be resized to the same size as the first one and then concat together. Usually used in FCN head of HRNet.
  - ‘multiple\_select’: Multiple feature maps will be bundle into a list and passed into decode head.
  - None: Only one select feature map is allowed.



- **align\_corners** (*bool*) –align\_corners argument of F.interpolate. Default: False.
- **heatmap\_loss** (*dict*) –Config for heatmap loss. Default: None.
- **offset\_loss** (*dict*) –Config for offset loss. Default: None.

**forward** (*x*)

Forward function.

**get\_loss** (*outputs, heatmaps, masks, offsets, offset\_weights*)

Calculate the dekr loss.

---

注解:

- batch\_size: N
  - num\_channels: C
  - num\_joints: K
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **outputs** (*List(torch.Tensor[N, C, H, W])*) –Multi-scale outputs.
- **heatmaps** (*List(torch.Tensor[N, K+1, H, W])*) –Multi-scale heatmap targets.
- **masks** (*List(torch.Tensor[N, K+1, H, W])*) –Weights of multi-scale heatmap targets.
- **offsets** (*List(torch.Tensor[N, K\*2, H, W])*) –Multi-scale offset targets.
- **offset\_weights** (*List(torch.Tensor[N, K\*2, H, W])*) –Weights of multi-scale offset targets.

**init\_weights** ()

Initialize model weights.

```
class mmpose.models.heads.DeconvHead (in_channels=3, out_channels=17, num_deconv_layers=3,  
                                         num_deconv_filters=(256, 256, 256), num_deconv_kernels=(4,  
                                         4, 4), extra=None, in_index=0, input_transform=None,  
                                         align_corners=False, loss_keypoint=None)
```

Simple deconv head.

参数

- **in\_channels** (*int*) –Number of input channels.

- **out\_channels** (*int*) –Number of output channels.
- **num\_deconv\_layers** (*int*) –Number of deconv layers. num\_deconv\_layers should  $\geq 0$ . Note that 0 means no deconv layers.
- **num\_deconv\_filters** (*list/tuple*) –Number of filters. If num\_deconv\_layers  $> 0$ , the length of
- **num\_deconv\_kernels** (*list/tuple*) –Kernel sizes.
- **in\_index** (*int/Sequence[int]*) –Input feature index. Default: 0
- **input\_transform** (*str/None*) –Transformation type of input features. Options: ‘resize\_concat’, ‘multiple\_select’, None. Default: None.
  - ‘**resize\_concat**’: Multiple feature maps will be resized to the same size as the first one and then concat together. Usually used in FCN head of HRNet.
  - ‘**multiple\_select**’: Multiple feature maps will be bundle into a list and passed into decode head.
  - None: Only one select feature map is allowed.
- **align\_corners** (*bool*) –align\_corners argument of F.interpolate. Default: False.
- **loss\_keypoint** (*dict*) –Config for loss. Default: None.

**forward** (*x*)

Forward function.

**get\_loss** (*outputs, targets, masks*)

Calculate bottom-up masked mse loss.

---

注解:

- batch\_size: N
  - num\_channels: C
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **outputs** (*List(torch.Tensor[N, C, H, W])*) –Multi-scale outputs.
- **targets** (*List(torch.Tensor[N, C, H, W])*) –Multi-scale targets.
- **masks** (*List(torch.Tensor[N, H, W])*) –Masks of multi-scale targets.

**init\_weights()**

Initialize model weights.

```
class mmpose.models.heads.DeepposeRegressionHead(in_channels, num_joints,
                                                loss_keypoint=None, out_sigma=False,
                                                train_cfg=None, test_cfg=None)
```

Deeppose regression head with fully connected layers.

“DeepPose: Human Pose Estimation via Deep Neural Networks” .

参数

- **in\_channels** (*int*) –Number of input channels
- **num\_joints** (*int*) –Number of joints
- **loss\_keypoint** (*dict*) –Config for keypoint loss. Default: None.
- **out\_sigma** (*bool*) –Predict the sigma (the viriance of the joint location) together with the joint location. Default: False

**decode** (*img metas*, *output*, *\*\*kwargs*)

Decode the keypoints from output regression.

参数

- **img\_metas** (*list(dict)*) –Information about data augmentation By default this includes:
  - ” image\_file”: path to the image file
  - ” center” : center of the bbox
  - ” scale” : scale of the bbox
  - ” rotation” : rotation of the bbox
  - ” bbox\_score” : score of bbox
- **output** (*np.ndarray[N, K, >=2]*) –predicted regression vector.
- **kwargs** –dict contains ‘img\_size’ . img\_size (tuple(img\_width, img\_height)): input image size.

**forward** (*x*)

Forward function.

**get\_accuracy** (*output*, *target*, *target\_weight*)

Calculate accuracy for top-down keypoint loss.

---

注解:

- batch\_size: N

- **num\_keypoints**: K
- 

#### 参数

- **output** (*torch.Tensor*[N, K, 2 or 4]) –Output keypoints.
- **target** (*torch.Tensor*[N, K, 2]) –Target keypoints.
- **target\_weight** (*torch.Tensor*[N, K, 2]) –Weights across different joint types.

**get\_loss** (*output, target, target\_weight*)

Calculate top-down keypoint loss.

---

#### 注解:

- **batch\_size**: N
  - **num\_keypoints**: K
- 

#### 参数

- **output** (*torch.Tensor*[N, K, 2 or 4]) –Output keypoints.
- **target** (*torch.Tensor*[N, K, 2]) –Target keypoints.
- **target\_weight** (*torch.Tensor*[N, K, 2]) –Weights across different joint types.

**inference\_model** (*x, flip\_pairs=None*)

Inference function.

返回 Output regression.

返回类型 *output\_regression* (*np.ndarray*)

#### 参数

- **x** (*torch.Tensor*[N, K, 2]) –Input features.
- **flip\_pairs** (*None* | *list[tuple()]*) –Pairs of keypoints which are mirrored.

**class** *mmpose.models.heads.HMRMeshHead* (*in\_channels, smpl\_mean\_params=None, n\_iter=3*)

SMPL parameters regressor head of simple baseline. “End-to-end Recovery of Human Shape and Pose”, CVPR’ 2018.

#### 参数

- **in\_channels** (*int*) –Number of input channels
- **smpl\_mean\_params** (*str*) –The file name of the mean SMPL parameters

- **n\_iter** (*int*) –The iterations of estimating delta parameters

**forward** (*x*)

Forward function.

*x* is the image feature map and is expected to be in shape (batch size x channel number x height x width)

**init\_weights** ()

Initialize model weights.

```
class mmpose.models.heads.Interhand3DHead (keypoint_head_cfg, root_head_cfg,
                                             hand_type_head_cfg, loss_keypoint=None,
                                             loss_root_depth=None, loss_hand_type=None,
                                             train_cfg=None, test_cfg=None)
```

Interhand 3D head of paper ref: Gyeongsik Moon. “InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image” .

#### 参数

- **keypoint\_head\_cfg** (*dict*) –Configs of Heatmap3DHead for hand keypoint estimation.
- **root\_head\_cfg** (*dict*) –Configs of Heatmap1DHead for relative hand root depth estimation.
- **hand\_type\_head\_cfg** (*dict*) –Configs of MultilabelClassificationHead for hand type classification.
- **loss\_keypoint** (*dict*) –Config for keypoint loss. Default: None.
- **loss\_root\_depth** (*dict*) –Config for relative root depth loss. Default: None.
- **loss\_hand\_type** (*dict*) –Config for hand type classification loss. Default: None.

**decode** (*img metas, output, \*\*kwargs*)

Decode hand keypoint, relative root depth and hand type.

#### 参数

- **img\_metas** (*list (dict)*) –Information about data augmentation By default this includes:
  - ” image\_file: path to the image file
  - ” center” : center of the bbox
  - ” scale” : scale of the bbox
  - ” rotation” : rotation of the bbox
  - ” bbox\_score” : score of bbox
  - ” heatmap3d\_depth\_bound” : depth bound of hand keypoint 3D heatmap

– “**root\_depth\_bound**” : depth bound of relative root depth 1D heatmap

- **output** (*list* [*np.ndarray*]) –model predicted 3D heatmaps, relative root depth and hand type.

**forward** (*x*)

Forward function.

**get\_accuracy** (*output*, *target*, *target\_weight*)

Calculate accuracy for hand type.

参数

- **output** (*list* [*Tensor*]) –a list of outputs from multiple heads.
- **target** (*list* [*Tensor*]) –a list of targets for multiple heads.
- **target\_weight** (*list* [*Tensor*]) –a list of targets weight for multiple heads.

**get\_loss** (*output*, *target*, *target\_weight*)

Calculate loss for hand keypoint heatmaps, relative root depth and hand type.

参数

- **output** (*list* [*Tensor*]) –a list of outputs from multiple heads.
- **target** (*list* [*Tensor*]) –a list of targets for multiple heads.
- **target\_weight** (*list* [*Tensor*]) –a list of targets weight for multiple heads.

**inference\_model** (*x*, *flip\_pairs=None*)

Inference function.

返回 list of output hand keypoint heatmaps, relative root depth and hand type.

返回类型 *output* (*list* [*np.ndarray*])

参数

- **x** (*torch.Tensor* [*N, K, H, W*]) –Input features.
- **flip\_pairs** (*None* | *list* [*tuple* ()]) –Pairs of keypoints which are mirrored.

```
class mmpose.models.heads.MultiModalSSAHead (num_classes, modality, in_channels=1024,
                                             avg_pool_kernel=(1, 7, 7), dropout_prob=0.0,
                                             train_cfg=None, test_cfg=None, **kwargs)
```

Spatial-temporal Semantic Alignment Head proposed in “Improving the performance of unimodal dynamic hand-gesture recognition with multimodal training” ,

Please refer to the [paper](#) for details.

参数

- **num\_classes** (*int*) –number of classes.
- **modality** (*list* [*str*]) –modalities of input videos for backbone.

- **in\_channels** (*int*) –number of channels of feature maps. Default: 1024
- **avg\_pool\_kernel** (*tuple[int]*) –kernel size of pooling layer. Default: (1, 7, 7)
- **dropout\_prob** (*float*) –probability to use dropout on input feature map. Default: 0
- **train\_cfg** (*dict*) –training config.
- **test\_cfg** (*dict*) –testing config.

**forward** (*x, img metas*)

Forward function.

**get\_accuracy** (*logits, label, img metas*)

Compute the accuracy of predicted gesture.

---

注解:

- batch\_size: N
  - number of classes: nC
  - logit length: L
- 

参数

- **logits** (*list[NxnCxL]*) –predicted logits for each modality.
- **label** (*list(dict)*) –Category label.
- **img\_metas** (*list(dict)*) –Information about data. By default this includes: -  
“fps: video frame rate - “modality” : modality of input videos

返回 computed accuracy for each modality.

返回类型 dict[str, torch.tensor]

**get\_loss** (*logits, label, fmaps=None*)

Compute the Cross Entropy loss and SSA loss.

---

注解:

- batch\_size: N
- number of classes: nC
- feature map channel: C
- feature map height: H
- feature map width: W
- feature map length: L

- logit length:  $L_g$
- 

#### 参数

- **logits** (*list* [ $N \times n \times C \times L_g$ ]) –predicted logits for each modality.
- **label** (*list* (*dict*)) –Category label.
- **fmaps** (*list* [*torch.Tensor* [ $N \times C \times L \times H \times W$ ]]) –feature maps for each modality.

返回 computed losses.

返回类型 dict[str, torch.tensor]

**init\_weights** ()

Initialize model weights.

**set\_train\_epoch** (*epoch: int*)

set the epoch to control the activation of SSA loss.

```
class mmpose.models.heads.TemporalRegressionHead (in_channels, num_joints, max_norm=None,  
                                                    loss_keypoint=None, is_trajectory=False,  
                                                    train_cfg=None, test_cfg=None)
```

Regression head of VideoPose3D.

“3D human pose estimation in video with temporal convolutions and semi-supervised training” , CVPR’ 2019.

#### 参数

- **in\_channels** (*int*) –Number of input channels
- **num\_joints** (*int*) –Number of joints
- **loss\_keypoint** (*dict*) –Config for keypoint loss. Default: None.
- **max\_norm** (*float | None*) –if not None, the weight of convolution layers will be clipped to have a maximum norm of max\_norm.
- **is\_trajectory** (*bool*) –If the model only predicts root joint position, then this arg should be set to True. In this case, traj\_loss will be calculated. Otherwise, it should be set to False. Default: False.

**decode** (*metas, output*)

Decode the keypoints from output regression.

#### 参数

- **metas** (*list* (*dict*)) –Information about data augmentation. By default this includes:
  - ” target\_image\_path” : path to the image file



- **output** (*np.ndarray*[*N*, *K*, 3]) –predicted regression vector.
- **metas** –Information about data augmentation including:
  - **target\_image\_path** (str): Optional, path to the image file
  - **target\_mean** (float): **Optional, normalization parameter of** the target pose.
  - **target\_std** (float): **Optional, normalization parameter of the** target pose.
  - **root\_position** (*np.ndarray*[3,1]): **Optional, global** position of the root joint.
  - **root\_index** (*torch.ndarray*[1,]): **Optional, original index of** the root joint before root-centering.

**forward** (*x*)

Forward function.

**get\_accuracy** (*output*, *target*, *target\_weight*, *metas*)

Calculate accuracy for keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
- 

参数

- **output** (*torch.Tensor*[*N*, *K*, 3]) –Output keypoints.
- **target** (*torch.Tensor*[*N*, *K*, 3]) –Target keypoints.
- **target\_weight** (*torch.Tensor*[*N*, *K*, 3]) –Weights across different joint types.
- **metas** (*list(dict)*) –Information about data augmentation including:
  - **target\_image\_path** (str): Optional, path to the image file
  - **target\_mean** (float): **Optional, normalization parameter of** the target pose.
  - **target\_std** (float): **Optional, normalization parameter of the** target pose.
  - **root\_position** (*np.ndarray*[3,1]): **Optional, global** position of the root joint.
  - **root\_index** (*torch.ndarray*[1,]): **Optional, original index of** the root joint before root-centering.

**get\_loss** (*output*, *target*, *target\_weight*)

Calculate keypoint loss.

---

注解:

- `batch_size`: N
  - `num_keypoints`: K
- 

参数

- **output** (*torch.Tensor*[N, K, 3]) –Output keypoints.
- **target** (*torch.Tensor*[N, K, 3]) –Target keypoints.
- **target\_weight** (*torch.Tensor*[N, K, 3]) –Weights across different joint types. If `self.is_trajectory` is True and `target_weight` is None, `target_weight` will be set inversely proportional to joint depth.

**inference\_model** (*x*, *flip\_pairs=None*)

Inference function.

返回 Output regression.

返回类型 `output_regression` (`np.ndarray`)

参数

- **x** (*torch.Tensor*[N, K, 2]) –Input features.
- **flip\_pairs** (*None* | *list[tuple()]*) –Pairs of keypoints which are mirrored.

**init\_weights** ()

Initialize the weights.

**class** `mmpose.models.heads.TopdownHeatmapBaseHead`

Base class for top-down heatmap heads.

All top-down heatmap heads should subclass it. All subclass should overwrite:

Methods:`get_loss`, supporting to calculate loss. Methods:`get_accuracy`, supporting to calculate accuracy. Methods:`forward`, supporting to forward model. Methods:`inference_model`, supporting to inference model.

**decode** (*img metas*, *output*, *\*\*kwargs*)

Decode keypoints from heatmaps.

参数

- **img\_metas** (*list(dict)*) –Information about data augmentation By default this includes:
  - " `image_file`: path to the image file
  - " `center`" : center of the bbox

- "scale" : scale of the bbox
- "rotation" : rotation of the bbox
- "bbox\_score" : score of bbox

• **output** (*np.ndarray*[*N*, *K*, *H*, *W*]) –model predicted heatmaps.

**abstract forward** (*\*\*kwargs*)

Forward function.

**abstract get\_accuracy** (*\*\*kwargs*)

Gets the accuracy.

**abstract get\_loss** (*\*\*kwargs*)

Gets the loss.

**abstract inference\_model** (*\*\*kwargs*)

Inference function.

```
class mmpose.models.heads.TopdownHeatmapMSMUHead(out_shape, unit_channels=256,
                                                  out_channels=17, num_stages=4,
                                                  num_units=4, use_prm=False,
                                                  norm_cfg={'type': 'BN'},
                                                  loss_keypoint=None, train_cfg=None,
                                                  test_cfg=None)
```

Heads for multi-stage multi-unit heads used in Multi-Stage Pose estimation Network (MSPN), and Residual Steps Networks (RSN).

#### 参数

- **unit\_channels** (*int*) –Number of input channels.
- **out\_channels** (*int*) –Number of output channels.
- **out\_shape** (*tuple*) –Shape of the output heatmap.
- **num\_stages** (*int*) –Number of stages.
- **num\_units** (*int*) –Number of units in each stage.
- **use\_prm** (*bool*) –Whether to use pose refine machine (PRM). Default: False.
- **norm\_cfg** (*dict*) –dictionary to construct and config norm layer. Default: dict(type='BN' )
- **loss\_keypoint** (*dict*) –Config for keypoint loss. Default: None.

**forward** (*x*)

Forward function.

#### 返回

a list of heatmaps from multiple stages and units.

返回类型 out (list[Tensor])

**get\_accuracy** (*output, target, target\_weight*)

Calculate accuracy for top-down keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*torch.Tensor [N, K, H, W]*) –Output heatmaps.
- **target** (*torch.Tensor [N, K, H, W]*) –Target heatmaps.
- **target\_weight** (*torch.Tensor [N, K, 1]*) –Weights across different joint types.

**get\_loss** (*output, target, target\_weight*)

Calculate top-down keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - num\_outputs: O
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*torch.Tensor [N, O, K, H, W]*) –Output heatmaps.
- **target** (*torch.Tensor [N, O, K, H, W]*) –Target heatmaps.
- **target\_weight** (*torch.Tensor [N, O, K, 1]*) –Weights across different joint types.

**inference\_model** (*x, flip\_pairs=None*)

Inference function.

---

返回 Output heatmaps.

返回类型 `output_heatmap` (`np.ndarray`)

参数

- **x** (`list[torch.Tensor[N, K, H, W]]`) –Input features.
- **flip\_pairs** (`None` | `list[tuple]`) –Pairs of keypoints which are mirrored.

**init\_weights** ()

Initialize model weights.

```
class mmpose.models.heads.TopdownHeatmapMultiStageHead (in_channels=512,
                                                         out_channels=17, num_stages=1,
                                                         num_deconv_layers=3,
                                                         num_deconv_filters=(256, 256,
                                                         256), num_deconv_kernels=(4, 4,
                                                         4), extra=None,
                                                         loss_keypoint=None,
                                                         train_cfg=None, test_cfg=None)
```

Top-down heatmap multi-stage head.

TopdownHeatmapMultiStageHead is consisted of multiple branches, each of which has `num_deconv_layers` ( $\geq 0$ ) number of deconv layers and a simple conv2d layer.

参数

- **in\_channels** (`int`) –Number of input channels.
- **out\_channels** (`int`) –Number of output channels.
- **num\_stages** (`int`) –Number of stages.
- **num\_deconv\_layers** (`int`) –Number of deconv layers. `num_deconv_layers` should  $\geq 0$ . Note that 0 means no deconv layers.
- **num\_deconv\_filters** (`list/tuple`) –Number of filters. If `num_deconv_layers`  $> 0$ , the length of
- **num\_deconv\_kernels** (`list/tuple`) –Kernel sizes.
- **loss\_keypoint** (`dict`) –Config for keypoint loss. Default: None.

**forward** (*x*)

Forward function.

返回 a list of heatmaps from multiple stages.

返回类型 `out` (`list[Tensor]`)

**get\_accuracy** (*output, target, target\_weight*)

Calculate accuracy for top-down keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*torch.Tensor* [N, K, H, W]) –Output heatmaps.
- **target** (*torch.Tensor* [N, K, H, W]) –Target heatmaps.
- **target\_weight** (*torch.Tensor* [N, K, 1]) –Weights across different joint types.

**get\_loss** (*output, target, target\_weight*)

Calculate top-down keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - num\_outputs: O
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*torch.Tensor* [N, K, H, W]) –Output heatmaps.
- **target** (*torch.Tensor* [N, K, H, W]) –Target heatmaps.
- **target\_weight** (*torch.Tensor* [N, K, 1]) –Weights across different joint types.

**inference\_model** (*x, flip\_pairs=None*)

Inference function.

返回 Output heatmaps.

返回类型 output\_heatmap (np.ndarray)

**参数**

- **x** (*List[torch.Tensor[NxKxHxW]]*) –Input features.
- **flip\_pairs** (*None | list[tuple()]*) –Pairs of keypoints which are mirrored.

**init\_weights()**

Initialize model weights.

```
class mmpose.models.heads.TopdownHeatmapSimpleHead(in_channels, out_channels,
                                                    num_deconv_layers=3,
                                                    num_deconv_filters=(256, 256, 256),
                                                    num_deconv_kernels=(4, 4, 4),
                                                    extra=None, in_index=0,
                                                    input_transform=None,
                                                    align_corners=False,
                                                    loss_keypoint=None, train_cfg=None,
                                                    test_cfg=None)
```

Top-down heatmap simple head. paper ref: Bin Xiao et al. Simple Baselines for Human Pose Estimation and Tracking.

TopdownHeatmapSimpleHead is consisted of ( $\geq 0$ ) number of deconv layers and a simple conv2d layer.

**参数**

- **in\_channels** (*int*) –Number of input channels
- **out\_channels** (*int*) –Number of output channels
- **num\_deconv\_layers** (*int*) –Number of deconv layers. num\_deconv\_layers should  $\geq 0$ . Note that 0 means no deconv layers.
- **num\_deconv\_filters** (*list/tuple*) –Number of filters. If num\_deconv\_layers  $> 0$ , the length of
- **num\_deconv\_kernels** (*list/tuple*) –Kernel sizes.
- **in\_index** (*int/Sequence[int]*) –Input feature index. Default: 0
- **input\_transform** (*str/None*) –Transformation type of input features. Options: 'resize\_concat', 'multiple\_select', None. Default: None.
  - 'resize\_concat' : Multiple feature maps will be resized to the same size as the first one and then concat together. Usually used in FCN head of HRNet.
  - 'multiple\_select' : Multiple feature maps will be bundle into a list and passed into decode head.
  - None: Only one select feature map is allowed.
- **align\_corners** (*bool*) –align\_corners argument of F.interpolate. Default: False.
- **loss\_keypoint** (*dict*) –Config for keypoint loss. Default: None.

**forward** (*x*)

Forward function.

**get\_accuracy** (*output, target, target\_weight*)

Calculate accuracy for top-down keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*torch.Tensor [N, K, H, W]*) –Output heatmaps.
- **target** (*torch.Tensor [N, K, H, W]*) –Target heatmaps.
- **target\_weight** (*torch.Tensor [N, K, 1]*) –Weights across different joint types.

**get\_loss** (*output, target, target\_weight*)

Calculate top-down keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*torch.Tensor [N, K, H, W]*) –Output heatmaps.
- **target** (*torch.Tensor [N, K, H, W]*) –Target heatmaps.
- **target\_weight** (*torch.Tensor [N, K, 1]*) –Weights across different joint types.

**inference\_model** (*x, flip\_pairs=None*)

Inference function.



返回 Output heatmaps.

返回类型 `output_heatmap` (`np.ndarray`)

参数

- **x** (`torch.Tensor[N, K, H, W]`) –Input features.
- **flip\_pairs** (`None` | `list[tuple]`) –Pairs of keypoints which are mirrored.

**init\_weights()**

Initialize model weights.

```
class mmpose.models.heads.ViPNASHeatmapSimpleHead(in_channels, out_channels,
                                                    num_deconv_layers=3,
                                                    num_deconv_filters=(144, 144, 144),
                                                    num_deconv_kernels=(4, 4, 4),
                                                    num_deconv_groups=(16, 16, 16),
                                                    extra=None, in_index=0,
                                                    input_transform=None,
                                                    align_corners=False, loss_keypoint=None,
                                                    train_cfg=None, test_cfg=None)
```

ViPNAS heatmap simple head.

ViPNAS: Efficient Video Pose Estimation via Neural Architecture Search. More details can be found in the [paper](#).

TopdownHeatmapSimpleHead is consisted of ( $\geq 0$ ) number of deconv layers and a simple conv2d layer.

参数

- **in\_channels** (`int`) –Number of input channels
- **out\_channels** (`int`) –Number of output channels
- **num\_deconv\_layers** (`int`) –Number of deconv layers. `num_deconv_layers` should  $\geq 0$ . Note that 0 means no deconv layers.
- **num\_deconv\_filters** (`list/tuple`) –Number of filters. If `num_deconv_layers`  $> 0$ , the length of
- **num\_deconv\_kernels** (`list/tuple`) –Kernel sizes.
- **num\_deconv\_groups** (`list/tuple`) –Group number.
- **in\_index** (`int/Sequence[int]`) –Input feature index. Default: -1
- **input\_transform** (`str/None`) –Transformation type of input features. Options: 'resize\_concat', 'multiple\_select', None. Default: None.
  - 'resize\_concat' : Multiple feature maps will be resize to the same size as first one and then concat together. Usually used in FCN head of HRNet.

- 'multiple\_select' : Multiple feature maps will be bundle into a list and passed into decode head.
- None: Only one select feature map is allowed.
- **align\_corners** (*bool*) –align\_corners argument of F.interpolate. Default: False.
- **loss\_keypoint** (*dict*) –Config for keypoint loss. Default: None.

**forward** (*x*)

Forward function.

**get\_accuracy** (*output, target, target\_weight*)

Calculate accuracy for top-down keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*torch.Tensor [N, K, H, W]*) –Output heatmaps.
- **target** (*torch.Tensor [N, K, H, W]*) –Target heatmaps.
- **target\_weight** (*torch.Tensor [N, K, 1]*) –Weights across different joint types.

**get\_loss** (*output, target, target\_weight*)

Calculate top-down keypoint loss.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmaps height: H
  - heatmaps weight: W
- 

参数

- **output** (*torch.Tensor [N, K, H, W]*) –Output heatmaps.

- **target** (*torch.Tensor* [N, K, H, W]) –Target heatmaps.
- **target\_weight** (*torch.Tensor* [N, K, 1]) –Weights across different joint types.

**inference\_model** (*x*, *flip\_pairs=None*)

Inference function.

返回 Output heatmaps.

返回类型 output\_heatmap (np.ndarray)

参数

- **x** (*torch.Tensor* [N, K, H, W]) –Input features.
- **flip\_pairs** (*None* | *list[tuple]*) –Pairs of keypoints which are mirrored.

**init\_weights** ()

Initialize model weights.

## 48.5 losses

**class** mmpose.models.losses.**AELoss** (*loss\_type*)

Associative Embedding loss.

Associative Embedding: End-to-End Learning for Joint Detection and Grouping.

**forward** (*tags*, *joints*)

Accumulate the tag loss for each image in the batch.

---

注解:

- batch\_size: N
  - heatmaps weight: W
  - heatmaps height: H
  - max\_num\_people: M
  - num\_keypoints: K
- 

参数

- **tags** (*torch.Tensor* [N, KxHxW, 1]) –tag channels of output.
- **joints** (*torch.Tensor* [N, M, K, 2]) –joints information.

**singleTagLoss** (*pred\_tag, joints*)

Associative embedding loss for one image.

---

注解:

- heatmaps weight: W
  - heatmaps height: H
  - max\_num\_people: M
  - num\_keypoints: K
- 

参数

- **pred\_tag** (*torch.Tensor [KxHxW, 1]*) –tag of output for one image.
- **joints** (*torch.Tensor [M, K, 2]*) –joints information for one image.

**class** mmpose.models.losses.**AdaptiveWingLoss** (*alpha=2.1, omega=14, epsilon=1, theta=0.5, use\_target\_weight=False, loss\_weight=1.0*)

Adaptive wing loss. paper ref: ‘Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression’ Wang et al. ICCV’ 2019.

参数

- **alpha** (*float*), **omega** (*float*), **epsilon** (*float*), **theta** (*float*) –are hyper-parameters.
- **use\_target\_weight** (*bool*) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **loss\_weight** (*float*) –Weight of the loss. Default: 1.0.

**criterion** (*pred, target*)

Criterion of wingloss.

---

注解: batch\_size: N num\_keypoints: K

---

参数

- **pred** (*torch.Tensor [NxKxHxW]*) –Predicted heatmaps.
- **target** (*torch.Tensor [NxKxHxW]*) –Target heatmaps.

**forward** (*output, target, target\_weight*)

Forward function.

---

注解: batch\_size: N num\_keypoints: K

---

#### 参数

- **output** (*torch.Tensor* [N x K x H x W]) –Output heatmaps.
- **target** (*torch.Tensor* [N x K x H x W]) –Target heatmaps.
- **target\_weight** (*torch.Tensor* [N x K x 1]) –Weights across different joint types.

**class** mmpose.models.losses.**BCELoss** (*use\_target\_weight=False, loss\_weight=1.0*)

Binary Cross Entropy loss.

**forward** (*output, target, target\_weight=None*)

Forward function.

---

注解:

- batch\_size: N
  - num\_labels: K
- 

#### 参数

- **output** (*torch.Tensor* [N, K]) –Output classification.
- **target** (*torch.Tensor* [N, K]) –Target classification.
- **target\_weight** (*torch.Tensor* [N, K] or *torch.Tensor* [N]) –Weights across different labels.

**class** mmpose.models.losses.**BoneLoss** (*joint\_parents, use\_target\_weight=False, loss\_weight=1.0*)

Bone length loss.

#### 参数

- **joint\_parents** (*list*) –Indices of each joint's parent joint.
- **use\_target\_weight** (*bool*) –Option to use weighted bone loss. Different bone types may have different target weights.
- **loss\_weight** (*float*) –Weight of the loss. Default: 1.0.

**forward** (*output, target, target\_weight=None*)

Forward function.

---

注解:

- `batch_size`: N
  - `num_keypoints`: K
  - `dimension of keypoints`: D (D=2 or D=3)
- 

参数

- **output** (*torch.Tensor*[N, K, D]) –Output regression.
- **target** (*torch.Tensor*[N, K, D]) –Target regression.
- **target\_weight** (*torch.Tensor*[N, K-1]) –Weights across different bone types.

**class** mmpose.models.losses.**FocalHeatmapLoss** (*alpha=2, beta=4*)

**forward** (*pred, gt, mask=None*)

Modified focal loss.

Exactly the same as CornerNet. Runs faster and costs a little bit more memory :param *pred*: :type *pred*: batch x c x h x w :param *gt\_regr*: :type *gt\_regr*: batch x c x h x w

**class** mmpose.models.losses.**GANLoss** (*gan\_type, real\_label\_val=1.0, fake\_label\_val=0.0, loss\_weight=1.0*)

Define GAN loss.

参数

- **gan\_type** (*str*) –Support ‘vanilla’ , ‘lsgan’ , ‘wgan’ , ‘hinge’ .
- **real\_label\_val** (*float*) –The value for real label. Default: 1.0.
- **fake\_label\_val** (*float*) –The value for fake label. Default: 0.0.
- **loss\_weight** (*float*) –Loss weight. Default: 1.0. Note that `loss_weight` is only for generators; and it is always 1.0 for discriminators.

**forward** (*input, target\_is\_real, is\_disc=False*)

参数

- **input** (*Tensor*) –The input for the loss module, i.e., the network prediction.
- **target\_is\_real** (*bool*) –Whether the target is real or fake.
- **is\_disc** (*bool*) –Whether the loss for discriminators or not. Default: False.

返回 GAN loss value.

返回类型 Tensor

**get\_target\_label** (*input*, *target\_is\_real*)

Get target label.

参数

- **input** (*Tensor*) –Input tensor.
- **target\_is\_real** (*bool*) –Whether the target is real or fake.

返回 Target tensor. Return bool for wgan, otherwise, return Tensor.

返回类型 (bool | Tensor)

**class** mmpose.models.losses.**HeatmapLoss** (*supervise\_empty=True*)

Accumulate the heatmap loss for each image in the batch.

参数 **supervise\_empty** (*bool*) –Whether to supervise empty channels.

**forward** (*pred*, *gt*, *mask*)

Forward function.

---

注解:

- batch\_size: N
  - heatmaps weight: W
  - heatmaps height: H
  - max\_num\_people: M
  - num\_keypoints: K
- 

参数

- **pred** (*torch.Tensor* [N, K, H, W]) –heatmap of output.
- **gt** (*torch.Tensor* [N, K, H, W]) –target heatmap.
- **mask** (*torch.Tensor* [N, H, W]) –mask of target.

**class** mmpose.models.losses.**JointsMSELoss** (*use\_target\_weight=False*, *loss\_weight=1.0*)

MSE loss for heatmaps.

参数

- **use\_target\_weight** (*bool*) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **loss\_weight** (*float*) –Weight of the loss. Default: 1.0.

**forward** (*output, target, target\_weight*)

Forward function.

**class** mmpose.models.losses.**JointsOHKMMSELoss** (*use\_target\_weight=False, topk=8, loss\_weight=1.0*)

MSE loss with online hard keypoint mining.

#### 参数

- **use\_target\_weight** (*bool*) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **topk** (*int*) –Only top k joint losses are kept.
- **loss\_weight** (*float*) –Weight of the loss. Default: 1.0.

**forward** (*output, target, target\_weight*)

Forward function.

**class** mmpose.models.losses.**L1Loss** (*use\_target\_weight=False, loss\_weight=1.0*)

L1Loss loss .

**forward** (*output, target, target\_weight=None*)

Forward function.

---

#### 注解:

- batch\_size: N
  - num\_keypoints: K
- 

#### 参数

- **output** (*torch.Tensor[N, K, 2]*) –Output regression.
- **target** (*torch.Tensor[N, K, 2]*) –Target regression.
- **target\_weight** (*torch.Tensor[N, K, 2]*) –Weights across different joint types.

**class** mmpose.models.losses.**MPJPELoss** (*use\_target\_weight=False, loss\_weight=1.0*)

MPJPE (Mean Per Joint Position Error) loss.

#### 参数

- **use\_target\_weight** (*bool*) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **loss\_weight** (*float*) –Weight of the loss. Default: 1.0.



**forward** (*output, target, target\_weight=None*)

Forward function.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - dimension of keypoints: D (D=2 or D=3)
- 

参数

- **output** (*torch.Tensor[N, K, D]*) –Output regression.
- **target** (*torch.Tensor[N, K, D]*) –Target regression.
- **target\_weight** (*torch.Tensor[N, K, D]*) –Weights across different joint types.

**class** `mmpose.models.losses.MSELoss` (*use\_target\_weight=False, loss\_weight=1.0*)

MSE loss for coordinate regression.

**forward** (*output, target, target\_weight=None*)

Forward function.

---

注解:

- batch\_size: N
  - num\_keypoints: K
- 

参数

- **output** (*torch.Tensor[N, K, 2]*) –Output regression.
- **target** (*torch.Tensor[N, K, 2]*) –Target regression.
- **target\_weight** (*torch.Tensor[N, K, 2]*) –Weights across different joint types.

**class** `mmpose.models.losses.MeshLoss` (*joints\_2d\_loss\_weight, joints\_3d\_loss\_weight, vertex\_loss\_weight, smpl\_pose\_loss\_weight, smpl\_beta\_loss\_weight, img\_res, focal\_length=5000*)

Mix loss for 3D human mesh. It is composed of loss on 2D joints, 3D joints, mesh vertices and smpl parameters (if any).

参数

- **joints\_2d\_loss\_weight** (*float*) –Weight for loss on 2D joints.
- **joints\_3d\_loss\_weight** (*float*) –Weight for loss on 3D joints.
- **vertex\_loss\_weight** (*float*) –Weight for loss on 3D vertices.
- **smpl\_pose\_loss\_weight** (*float*) –Weight for loss on SMPL pose parameters.
- **smpl\_beta\_loss\_weight** (*float*) –Weight for loss on SMPL shape parameters.
- **img\_res** (*int*) –Input image resolution.
- **focal\_length** (*float*) –Focal length of camera model. Default=5000.

**forward** (*output, target*)

Forward function.

#### 参数

- **output** (*dict*) –dict of network predicted results. Keys: ‘vertices’ , ‘joints\_3d’ , ‘camera’ , ‘pose’ (optional), ‘beta’ (optional)
- **target** (*dict*) –dict of ground-truth labels. Keys: ‘vertices’ , ‘joints\_3d’ , ‘joints\_3d\_visible’ , ‘joints\_2d’ , ‘joints\_2d\_visible’ , ‘pose’ , ‘beta’ , ‘has\_smpl’

返回 dict of losses.

返回类型 dict

**joints\_2d\_loss** (*pred\_joints\_2d, gt\_joints\_2d, joints\_2d\_visible*)

Compute 2D reprojection loss on the joints.

The loss is weighted by joints\_2d\_visible.

**joints\_3d\_loss** (*pred\_joints\_3d, gt\_joints\_3d, joints\_3d\_visible*)

Compute 3D joints loss for the examples that 3D joint annotations are available.

The loss is weighted by joints\_3d\_visible.

**project\_points** (*points\_3d, camera*)

Perform orthographic projection of 3D points using the camera parameters, return projected 2D points in image plane.

---

#### 注解:

- batch size: B
  - point number: N
- 

#### 参数

- **points\_3d** (*Tensor ([B, N, 3])*) –3D points.

- **camera** (*Tensor* ([B, 3])) –camera parameters with the 3 channel as (scale, translation\_x, translation\_y)

返回 projected 2D points in image space.

返回类型 *Tensor* ([B, N, 2])

**smpl\_losses** (*pred\_rotmat, pred\_betas, gt\_pose, gt\_betas, has\_smpl*)

Compute SMPL parameters loss for the examples that SMPL parameter annotations are available.

The loss is weighted by has\_smpl.

**vertex\_loss** (*pred\_vertices, gt\_vertices, has\_smpl*)

Compute 3D vertex loss for the examples that 3D human mesh annotations are available.

The loss is weighted by the has\_smpl.

**class** mmpose.models.losses.**MultiLossFactory** (*num\_joints, num\_stages, ae\_loss\_type, with\_ae\_loss, push\_loss\_factor, pull\_loss\_factor, with\_heatmaps\_loss, heatmaps\_loss\_factor, supervise\_empty=True*)

Loss for bottom-up models.

参数

- **num\_joints** (*int*) –Number of keypoints.
- **num\_stages** (*int*) –Number of stages.
- **ae\_loss\_type** (*str*) –Type of ae loss.
- **with\_ae\_loss** (*list[bool]*) –Use ae loss or not in multi-heatmap.
- **push\_loss\_factor** (*list[float]*) –Parameter of push loss in multi-heatmap.
- **pull\_loss\_factor** (*list[float]*) –Parameter of pull loss in multi-heatmap.
- **with\_heatmap\_loss** (*list[bool]*) –Use heatmap loss or not in multi-heatmap.
- **heatmaps\_loss\_factor** (*list[float]*) –Parameter of heatmap loss in multi-heatmap.
- **supervise\_empty** (*bool*) –Whether to supervise empty channels.

**forward** (*outputs, heatmaps, masks, joints*)

Forward function to calculate losses.

注解:

- batch\_size: N
- heatmaps weight: W
- heatmaps height: H

- `max_num_people`: M
  - `num_keypoints`: K
  - `output_channel`: C C=2K if use ae loss else K
- 

#### 参数

- **outputs** (*list (torch.Tensor[N, C, H, W])*) –outputs of stages.
- **heatmaps** (*list (torch.Tensor[N, K, H, W])*) –target of heatmaps.
- **masks** (*list (torch.Tensor[N, H, W])*) –masks of heatmaps.
- **joints** (*list (torch.Tensor[N, M, K, 2])*) –joints of ae loss.

```
class mmpose.models.losses.RLELoss (use_target_weight=False, size_average=True, residual=True,  
                                     q_dis='laplace')
```

RLE Loss.

Human Pose Regression With Residual Log-Likelihood Estimation arXiv:.

Code is modified from [the official implementation](#).

#### 参数

- **use\_target\_weight** (*bool*) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **size\_average** (*bool*) –Option to average the loss by the batch\_size.
- **residual** (*bool*) –Option to add L1 loss and let the flow learn the residual error distribution.
- **q\_dis** (*string*) –Option for the identity Q(error) distribution, Options: “laplace” or “gaussian”

```
forward (output, target, target_weight=None)
```

Forward function.

---

#### 注解:

- `batch_size`: N
  - `num_keypoints`: K
  - dimension of keypoints: D (D=2 or D=3)
- 

#### 参数

- **output** (`torch.Tensor[N, K, D*2]`) –Output regression, including coords and sigmas.
- **target** (`torch.Tensor[N, K, D]`) –Target regression.
- **target\_weight** (`torch.Tensor[N, K, D]`) –Weights across different joint types.

**class** `mmpose.models.losses.SemiSupervisionLoss` (`joint_parents`, `projection_loss_weight=1.0`,  
`bone_loss_weight=1.0`, `warmup_iterations=0`)

Semi-supervision loss for unlabeled data. It is composed of projection loss and bone loss.

Paper ref: *3D human pose estimation in video with temporal convolutions and semi-supervised training* Dario Pavlo et al. CVPR' 2019.

#### 参数

- **joint\_parents** (`list`) –Indices of each joint's parent joint.
- **projection\_loss\_weight** (`float`) –Weight for projection loss.
- **bone\_loss\_weight** (`float`) –Weight for bone loss.
- **warmup\_iterations** (`int`) –Number of warmup iterations. In the first `warmup_iterations` iterations, the model is trained only on labeled data, and semi-supervision loss will be 0. This is a workaround since currently we cannot access epoch number in loss functions. Note that the iteration number in an epoch can be changed due to different GPU numbers in multi-GPU settings. So please set this parameter carefully.  
`warmup_iterations = dataset_size // samples_per_gpu // gpu_num * warmup_epochs`

**forward** (`output`, `target`)

Defines the computation performed at every call.

Should be overridden by all subclasses.

---

**注解:** Although the recipe for forward pass needs to be defined within this function, one should call the `Module` instance afterwards instead of this since the former takes care of running the registered hooks while the latter silently ignores them.

---

**static project\_joints** (`x`, `intrinsic`)

Project 3D joint coordinates to 2D image plane using camera intrinsic parameters.

#### 参数

- **x** (`torch.Tensor[N, K, 3]`) –3D joint coordinates.
- **intrinsic** (`torch.Tensor[N, 4] | torch.Tensor[N, 9]`) –Camera intrinsic: f (2), c (2), k (3), p (2).

```
class mmpose.models.losses.SmoothL1Loss (use_target_weight=False, loss_weight=1.0)
```

SmoothL1Loss loss.

#### 参数

- **use\_target\_weight** (*bool*) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **loss\_weight** (*float*) –Weight of the loss. Default: 1.0.

```
forward (output, target, target_weight=None)
```

Forward function.

---

#### 注解:

- batch\_size: N
  - num\_keypoints: K
  - dimension of keypoints: D (D=2 or D=3)
- 

#### 参数

- **output** (*torch.Tensor*[N, K, D]) –Output regression.
- **target** (*torch.Tensor*[N, K, D]) –Target regression.
- **target\_weight** (*torch.Tensor*[N, K, D]) –Weights across different joint types.

```
class mmpose.models.losses.SoftWeightSmoothL1Loss (use_target_weight=False,  
                                                    supervise_empty=True, beta=1.0,  
                                                    loss_weight=1.0)
```

Smooth L1 loss with soft weight for regression.

#### 参数

- **use\_target\_weight** (*bool*) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **supervise\_empty** (*bool*) –Whether to supervise the output with zero weight.
- **beta** (*float*) –Specifies the threshold at which to change between L1 and L2 loss.
- **loss\_weight** (*float*) –Weight of the loss. Default: 1.0.

```
forward (output, target, target_weight=None)
```

Forward function.

---

#### 注解:

- `batch_size`: N
- `num_keypoints`: K
- `dimension of keypoints`: D (D=2 or D=3)

### 参数

- **output** (`torch.Tensor[N, K, D]`) –Output regression.
- **target** (`torch.Tensor[N, K, D]`) –Target regression.
- **target\_weight** (`torch.Tensor[N, K, D]`) –Weights across different joint types.

**static smooth\_l1\_loss** (`input, target, reduction='none', beta=1.0`)

Re-implement `torch.nn.functional.smooth_l1_loss` with `beta` to support `pytorch <= 1.6`.

**class** `mmpose.models.losses.SoftWingLoss` (`omega1=2.0, omega2=20.0, epsilon=0.5, use_target_weight=False, loss_weight=1.0`)

Soft Wing Loss ‘Structure-Coherent Deep Feature Learning for Robust Face Alignment’ Lin et al. TIP’ 2021.

**loss =**

1.  $|x|$ , if  $|x| < \text{omega1}$
2.  $\text{omega2} * \ln(1 + |x| / \text{epsilon}) + B$ , if  $|x| \geq \text{omega1}$

### 参数

- **omega1** (`float`) –The first threshold.
- **omega2** (`float`) –The second threshold.
- **epsilon** (`float`) –Also referred to as curvature.
- **use\_target\_weight** (`bool`) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **loss\_weight** (`float`) –Weight of the loss. Default: 1.0.

**criterion** (`pred, target`)

Criterion of wingloss.

注解: `batch_size`: N `num_keypoints`: K `dimension of keypoints`: D (D=2 or D=3)

### 参数

- **pred** (`torch.Tensor[N, K, D]`) –Output regression.

- **target** (`torch.Tensor[N, K, D]`) – Target regression.

**forward** (*output, target, target\_weight=None*)

Forward function.

**注解:** batch\_size: N num\_keypoints: K dimension of keypoints: D (D=2 or D=3)

### 参数

- **output** (*torch.Tensor*[*N*, *K*, *D*]) –Output regression.
- **target** (*torch.Tensor*[*N*, *K*, *D*]) –Target regression.
- **target\_weight** (*torch.Tensor*[*N*, *K*, *D*]) –Weights across different joint types.

```
class mmpose.models.losses.WingLoss(omega=10.0, epsilon=2.0, use_target_weight=False,  
                                     loss_weight=1.0)
```

Wing Loss. paper ref: ‘Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural Networks’  
Feng et al. CVPR’ 2018.

### 参数

- **omega** (*float*) –Also referred to as width.
- **epsilon** (*float*) –Also referred to as curvature.
- **use\_target\_weight** (*bool*) –Option to use weighted MSE loss. Different joint types may have different target weights.
- **loss\_weight** (*float*) –Weight of the loss. Default: 1.0.

**criterion** (*pred, target*)

Criterion of wingloss.

**注解:**

- batch\_size: N
- num\_keypoints: K
- dimension of keypoints: D (D=2 or D=3)

### 参数

- **pred**(*torch.Tensor*[*N*, *K*, *D*])–Output regression.
- **target**(*torch.Tensor*[*N*, *K*, *D*])–Target regression.



**forward** (*output, target, target\_weight=None*)

Forward function.

---

注解:

- `batch_size`: N
  - `num_keypoints`: K
  - `dimension of keypoints`: D (D=2 or D=3)
- 

参数

- **output** (*torch.Tensor[N, K, D]*) –Output regression.
- **target** (*torch.Tensor[N, K, D]*) –Target regression.
- **target\_weight** (*torch.Tensor[N, K, D]*) –Weights across different joint types.

## 48.6 misc



```
class mmpose.datasets.AnimalATRWDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                           dataset_info=None, test_mode=False)
```

ATRW dataset for animal pose estimation.

“ATRW: A Benchmark for Amur Tiger Re-identification in the Wild” ACM MM’ 2020. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

ATRW keypoint indexes:

```
0: "left_ear",  
1: "right_ear",  
2: "nose",  
3: "right_shoulder",  
4: "right_front_paw",  
5: "left_shoulder",  
6: "left_front_paw",  
7: "right_hip",  
8: "right_knee",  
9: "right_back_paw",  
10: "left_hip",  
11: "left_knee",  
12: "left_back_paw",
```

(下页继续)

(续上页)

```
13: "tail",
14: "center"
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='mAP', \*\*kwargs*)

Evaluate coco keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (*list[str]*): For example, [ 'data/coco/val2017/000000393226.jpg' ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap
  - bbox\_id (*list(int)*).
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: 'mAP' .

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.AnimalFlyDataset (ann_file, img_prefix, data_cfg, pipeline, dataset_info=None,  
                                         test_mode=False)
```

AnimalFlyDataset for animal pose estimation.

“Fast animal pose estimation using deep neural networks” Nature methods’ 2019. More details can be found in the [paper](#).

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

Vinegar Fly keypoint indexes:

```
0: "head",
1: "eyeL",
2: "eyeR",
3: "neck",
4: "thorax",
5: "abdomen",
6: "forelegR1",
7: "forelegR2",
8: "forelegR3",
9: "forelegR4",
10: "midlegR1",
11: "midlegR2",
12: "midlegR3",
13: "midlegR4",
14: "hindlegR1",
15: "hindlegR2",
16: "hindlegR3",
17: "hindlegR4",
18: "forelegL1",
19: "forelegL2",
20: "forelegL3",
21: "forelegL4",
22: "midlegL1",
23: "midlegL2",
24: "midlegL3",
25: "midlegL4",
26: "hindlegL1",
27: "hindlegL2",
28: "hindlegL3",
29: "hindlegL4",
```

(下页继续)

(续上页)

```
30: "wingL",
31: "wingR"
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='PCK'*, *\*\*kwargs*)

Evaluate Fly keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (*list[str]*): For example, [ 'Test/source/0.jpg' ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str*) –Path of directory to save the results.
- **metric** (*str | list[str]*) –Metric to be performed. Options: 'PCK', 'AUC', 'EPE'.

**返回** Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.AnimalHorse10Dataset (ann_file, img_prefix, data_cfg, pipeline,  
                                             dataset_info=None, test_mode=False)
```

AnimalHorse10Dataset for animal pose estimation.

“Pretraining boosts out-of-domain robustness for pose estimation” WACV’ 2021. More details can be found in the [paper](#).

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

Horse-10 keypoint indexes:

```
0: 'Nose',  
1: 'Eye',  
2: 'Nearknee',  
3: 'Nearfrontfetlock',  
4: 'Nearfrontfoot',  
5: 'Offknee',  
6: 'Offfrontfetlock',  
7: 'Offfrontfoot',  
8: 'Shoulder',  
9: 'Midshoulder',  
10: 'Elbow',  
11: 'Girth',  
12: 'Wither',  
13: 'Nearhindhock',  
14: 'Nearhindfetlock',  
15: 'Nearhindfoot',  
16: 'Hip',  
17: 'Stifle',  
18: 'Offhindhock',  
19: 'Offhindfetlock',  
20: 'Offhindfoot',  
21: 'Ischium'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.

- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCK', \*\*kwargs*)

Evaluate horse-10 keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (`np.ndarray[N,K,3]`): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (`np.ndarray[N,6]`): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (`list[str]`): For example, [ 'Test/source/0.jpg' ]
  - output\_heatmap (`np.ndarray[N, K, H, W]`): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: 'PCK', 'NME'

**返回** Evaluation results for evaluation metric.

**返回类型** dict

**class** `mmpose.datasets.AnimalLocustDataset` (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

AnimalLocustDataset for animal pose estimation.

“DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning” Elife’ 2019. More details can be found in the paper.

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

Desert Locust keypoint indexes:



```
0: "head",
1: "neck",
2: "thorax",
3: "abdomen1",
4: "abdomen2",
5: "anttipL",
6: "antbaseL",
7: "eyeL",
8: "forelegL1",
9: "forelegL2",
10: "forelegL3",
11: "forelegL4",
12: "midlegL1",
13: "midlegL2",
14: "midlegL3",
15: "midlegL4",
16: "hindlegL1",
17: "hindlegL2",
18: "hindlegL3",
19: "hindlegL4",
20: "anttipR",
21: "antbaseR",
22: "eyeR",
23: "forelegR1",
24: "forelegR2",
25: "forelegR3",
26: "forelegR4",
27: "midlegR1",
28: "midlegR2",
29: "midlegR3",
30: "midlegR4",
31: "hindlegR1",
32: "hindlegR2",
33: "hindlegR3",
34: "hindlegR4"
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.

- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCK', \*\*kwargs*)

Evaluate Fly keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (`np.ndarray[N,K,3]`): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (`np.ndarray[N,6]`): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (`list[str]`): For example, [ 'Test/source/0.jpg' ]
  - output\_heatmap (`np.ndarray[N, K, H, W]`): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: 'PCK', 'AUC', 'EPE' .

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** `mmpose.datasets.AnimalMacaqueDataset` (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

MacaquePose dataset for animal pose estimation.

“MacaquePose: A novel ‘in the wild’ macaque monkey pose dataset for markerless motion capture” bioRxiv 2020. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

Macaque keypoint indexes:

```

0: 'nose',
1: 'left_eye',
2: 'right_eye',
3: 'left_ear',
4: 'right_ear',
5: 'left_shoulder',
6: 'right_shoulder',
7: 'left_elbow',
8: 'right_elbow',
9: 'left_wrist',
10: 'right_wrist',
11: 'left_hip',
12: 'right_hip',
13: 'left_knee',
14: 'right_knee',
15: 'left_ankle',
16: 'right_ankle'

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='mAP', \*\*kwargs*)

Evaluate coco keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

**注解:** batch\_size: N num\_keypoints: K heatmap height: H heatmap width: W

---

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]

- `image_paths` (list[str]): For example, [ 'data/coco/val2017/000000393226.jpg' ]
- `heatmap` (np.ndarray[N, K, H, W]): model output heatmap
- `bbox_id` (list(int)).
- **res\_folder** (*str*, *optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str* | *list[str]*) –Metric to be performed. Defaults: 'mAP' .

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.AnimalPoseDataset (ann_file, img_prefix, data_cfg, pipeline,
                                         dataset_info=None, test_mode=False)
```

Animal-Pose dataset for animal pose estimation.

“Cross-domain Adaptation For Animal Pose Estimation” ICCV’ 2019 More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

Animal-Pose keypoint indexes:

```
0: 'L_Eye',
1: 'R_Eye',
2: 'L_EarBase',
3: 'R_EarBase',
4: 'Nose',
5: 'Throat',
6: 'TailBase',
7: 'Withers',
8: 'L_F_Elbow',
9: 'R_F_Elbow',
10: 'L_B_Elbow',
11: 'R_B_Elbow',
12: 'L_F_Knee',
13: 'R_F_Knee',
14: 'L_B_Knee',
15: 'R_B_Knee',
16: 'L_F_Paw',
17: 'R_F_Paw',
18: 'L_B_Paw',
19: 'R_B_Paw'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.

- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='mAP', \*\*kwargs*)

Evaluate coco keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (`np.ndarray[N,K,3]`): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (`np.ndarray[N,6]`): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (`list[str]`): For example, [ 'data/coco/val2017/000000393226.jpg' ]
  - heatmap (`np.ndarray[N, K, H, W]`): model output heatmap
  - bbox\_id (`list(int)`).
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: 'mAP' .

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** `mmpose.datasets.AnimalZebraDataset` (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

AnimalZebraDataset for animal pose estimation.

“DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning” Elife’ 2019. More details can be found in the paper.

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

Desert Locust keypoint indexes:

```
0: "snout",
1: "head",
2: "neck",
3: "forelegL1",
4: "forelegR1",
5: "hindlegL1",
6: "hindlegR1",
7: "tailbase",
8: "tailtip"
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='PCK'*, *\*\*kwargs*)

Evaluate Fly keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

### 注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:

- `preds` (`np.ndarray[N,K,3]`): The first two dimensions are coordinates, score is the third dimension of the array.
- `boxes` (`np.ndarray[N,6]`): [`center[0]`, `center[1]`, `scale[0]`, `scale[1]`, `area`, `score`]
- `image_paths` (`list[str]`): For example, [ 'Test/source/0.jpg' ]
- `output_heatmap` (`np.ndarray[N, K, H, W]`): model outputs.
- **`res_folder`** (`str`, *optional*) – The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **`metric`** (`str` | `list[str]`) – Metric to be performed. Options: 'PCK', 'AUC', 'EPE'.

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** `mmpose.datasets.Body3DH36MDataset` (`ann_file`, `img_prefix`, `data_cfg`, `pipeline`,  
`dataset_info=None`, `test_mode=False`)

Human3.6M dataset for 3D human pose estimation.

“Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments” , TPAMI’2014. More details can be found in the [paper](#).

Human3.6M keypoint indexes:

```
0: 'root (pelvis)',
1: 'right_hip',
2: 'right_knee',
3: 'right_foot',
4: 'left_hip',
5: 'left_knee',
6: 'left_foot',
7: 'spine',
8: 'thorax',
9: 'neck_base',
10: 'head',
11: 'left_shoulder',
12: 'left_elbow',
13: 'left_wrist',
14: 'right_shoulder',
15: 'right_elbow',
16: 'right_wrist'
```

参数

- **`ann_file`** (`str`) – Path to the annotation file.

- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**build\_sample\_indices** ()

Split original videos into sequences and build frame indices.

This method overrides the default one in the base class.

**evaluate** (*results, res\_folder=None, metric='mpjpe', \*\*kwargs*)

Evaluate keypoint results.

**get\_camera\_param** (*imgname*)

Get camera parameters of a frame by its image name.

**load\_annotations** ()

Load data annotation.

**load\_config** (*data\_cfg*)

Initialize dataset attributes according to the config.

Override this method to set dataset specific attributes.

```
class mmpose.datasets.Body3DMviewDirectCampusDataset (ann_file, img_prefix, data_cfg,  
                                                    pipeline, dataset_info=None,  
                                                    test_mode=False)
```

Campus dataset for direct multi-view human pose estimation.

*3D Pictorial Structures for Multiple Human Pose Estimation* CVPR' 2014 More details can be found in the paper  
<<http://campar.in.tum.de/pub/belagiannis2014cvpr/belagiannis2014cvpr.pdf>>

The dataset loads both 2D and 3D annotations as well as camera parameters. It is worth mentioning that when training multi-view 3D pose models, due to the limited and incomplete annotations of this dataset, we may not use this dataset to train the model. Instead, we use the 2D pose estimator trained on COCO, and use independent 3D human poses from the CMU Panoptic dataset to train the 3D model. For testing, we first estimate 2D poses and generate 2D heatmaps for this dataset as the input to 3D model.

Campus keypoint indices:

```
'Right-Ankle': 0,  
'Right-Knee': 1,  
'Right-Hip': 2,  
'Left-Hip': 3,  
'Left-Knee': 4,
```

(下页继续)



(续上页)

```
'Left-Ankle': 5,
'Right-Wrist': 6,
'Right-Elbow': 7,
'Right-Shoulder': 8,
'Left-Shoulder': 9,
'Left-Elbow': 10,
'Left-Wrist': 11,
'Bottom-Head': 12,
'Top-Head': 13,
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**static coco2campus3D** (*coco\_pose*)

transform coco order(our method output) 3d pose to campus dataset order with interpolation.

参数 **coco\_pose** –np.array with shape 17x3

Returns: 3D pose in campus order with shape 14x3

**evaluate** (*results, res\_folder=None, metric='pcp', recall\_threshold=500, alpha\_error=0.5, \*\*kwargs*)

### 参数

- **results** (*list[dict]*) –Testing results containing the following items: - pose\_3d (np.ndarray): predicted 3D human pose - sample\_id (np.ndarray): sample id of a frame.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: ‘pcp’ .
- **recall\_threshold** –threshold for calculating recall.
- **alpha\_error** –coefficient when calculating error for correct parts.
- **\*\*kwargs** –

Returns:

**static get\_new\_center** (*center\_list*)

Generate new center or select from the center list randomly.

The probability and the parameters related to coordinates can also be tuned, just make sure that the center is within the given 3D space.

**isvalid** (*new\_center, bbox, bbox\_list*)

Check if the new person bbox are valid, which need to satisfies:

1. the center is visible in at least 2 views, and
2. have a sufficiently small iou with all other person bboxes.

**load\_config** (*data\_cfg*)

Initialize dataset attributes according to the config.

Override this method to set dataset specific attributes.

**class** mmpose.datasets.**Body3DMviewDirectPanopticDataset** (*ann\_file, img\_prefix, data\_cfg,*  
*pipeline, dataset\_info=None,*  
*test\_mode=False*)

Panoptic dataset for direct multi-view human pose estimation.

Panoptic Studio: A Massively Multiview System for Social Motion Capture’ ICCV’ 2015 More details can be found in the ‘paper’ .

The dataset loads both 2D and 3D annotations as well as camera parameters.

Panoptic keypoint indexes:

```
'neck': 0,  
'nose': 1,  
'mid-hip': 2,  
'l-shoulder': 3,  
'l-elbow': 4,  
'l-wrist': 5,  
'l-hip': 6,  
'l-knee': 7,  
'l-ankle': 8,  
'r-shoulder': 9,  
'r-elbow': 10,  
'r-wrist': 11,  
'r-hip': 12,  
'r-knee': 13,  
'r-ankle': 14,  
'l-eye': 15,  
'l-ear': 16,  
'r-eye': 17,  
'r-ear': 18,
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='mpjpe', \*\*kwargs*)

### 参数

- **results** (*list[dict]*) –Testing results containing the following items: - pose\_3d (np.ndarray): predicted 3D human pose - sample\_id (np.ndarray): sample id of a frame.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: ‘mpjpe’ .
- **\*\*kwargs** –

Returns:

**load\_config** (*data\_cfg*)

Initialize dataset attributes according to the config.

Override this method to set dataset specific attributes.

**class** mmpose.datasets.**Body3DMviewDirectShelfDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

Shelf dataset for direct multi-view human pose estimation.

*3D Pictorial Structures for Multiple Human Pose Estimation’ CVPR’ 2014 More details can be found in the paper <<http://campar.in.tum.de/pub/belagiannis2014cvpr/belagiannis2014cvpr.pdf>>*

The dataset loads both 2D and 3D annotations as well as camera parameters. It is worth mentioning that when training multi-view 3D pose models, due to the limited and incomplete annotations of this dataset, we may not use this dataset to train the model. Instead, we use the 2D pose estimator trained on COCO, and use independent 3D human poses from the CMU Panoptic dataset to train the 3D model. For testing, we first estimate 2D poses and generate 2D heatmaps for this dataset as the input to 3D model.

Shelf keypoint indices:

```

'Right-Ankle': 0,
'Right-Knee': 1,
'Right-Hip': 2,
'Left-Hip': 3,
'Left-Knee': 4,
'Left-Ankle': 5,
'Right-Wrist': 6,
'Right-Elbow': 7,
'Right-Shoulder': 8,
'Left-Shoulder': 9,
'Left-Elbow': 10,
'Left-Wrist': 11,
'Bottom-Head': 12,
'Top-Head': 13,

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**static coco2shelf3D** (*coco\_pose*, *alpha=0.75*)

transform coco order(our method output) 3d pose to shelf dataset order with interpolation.

参数 **coco\_pose** –np.array with shape 17x3

Returns: 3D pose in shelf order with shape 14x3

**evaluate** (*results*, *res\_folder=None*, *metric='pcp'*, *recall\_threshold=500*, *alpha\_error=0.5*, *alpha\_head=0.75*,  
\*\**kwargs*)

### 参数

- **results** (*list[dict]*) –Testing results containing the following items: - pose\_3d (np.ndarray): predicted 3D human pose - sample\_id (np.ndarray): sample id of a frame.
- **res\_folder** (*str*, *optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: ‘pcp’ .
- **recall\_threshold** –threshold for calculating recall.

- **alpha\_error** –coefficient when calculating correct parts.
- **alpha\_head** –coefficient for computing head keypoints position when converting coco poses to shelf poses
- **\*\*kwargs** –

Returns:

**static get\_new\_center** (*center\_list*)

Generate new center or select from the center list randomly.

The probability and the parameters related to coordinates can also be tuned, just make sure that the center is within the given 3D space.

**static isvalid** (*bbox, bbox\_list*)

Check if the new person bbox are valid, which need to satisfies:

have a sufficiently small iou with all other person bboxes.

**load\_config** (*data\_cfg*)

Initialize dataset attributes according to the config.

Override this method to set dataset specific attributes.

**class** mmpose.datasets.**BottomUpAicDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

Aic dataset for bottom-up pose estimation.

“AI Challenger : A Large-scale Dataset for Going Deeper in Image Understanding” , arXiv’ 2017. More details can be found in the [paper](#)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

AIC keypoint indexes:

```
0: "right_shoulder",
1: "right_elbow",
2: "right_wrist",
3: "left_shoulder",
4: "left_elbow",
5: "left_wrist",
6: "right_hip",
7: "right_knee",
8: "right_ankle",
9: "left_hip",
10: "left_knee",
11: "left_ankle",
12: "head_top",
13: "neck"
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.BottomUpCocoDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                           dataset_info=None, test_mode=False)
```

COCO dataset for bottom-up pose estimation.

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

COCO keypoint indexes:

```
0: 'nose',  
1: 'left_eye',  
2: 'right_eye',  
3: 'left_ear',  
4: 'right_ear',  
5: 'left_shoulder',  
6: 'right_shoulder',  
7: 'left_elbow',  
8: 'right_elbow',  
9: 'left_wrist',  
10: 'right_wrist',  
11: 'left_hip',  
12: 'right_hip',  
13: 'left_knee',  
14: 'right_knee',  
15: 'left_ankle',  
16: 'right_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.

- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='mAP', \*\*kwargs*)

Evaluate coco keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

注解:

- num\_people: P
  - num\_keypoints: K
- 

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*list[np.ndarray(P, K, 3+tag\_num)]*): Pose predictions for all people in images.
  - scores (*list[P]*): List of person scores.
  - image\_path (*list[str]*): For example, [ 'coco/images/ val2017/000000397133.jpg' ]
  - heatmap (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: 'mAP' .

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** mmpose.datasets.**BottomUpCocoWholeBodyDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

CocoWholeBodyDataset dataset for bottom-up pose estimation.

Whole-Body Human Pose Estimation in the Wild' , ECCV' 2020. More details can be found in the 'paper' .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

In total, we have 133 keypoints for wholebody pose estimation.

COCO-WholeBody keypoint indexes:

```
0-16: 17 body keypoints,
17-22: 6 foot keypoints,
```

(下页继续)

(续上页)

```
23-90: 68 face keypoints,
91-132: 42 hand keypoints
```

**参数**

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.BottomUpCrowdPoseDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                                dataset_info=None, test_mode=False)
```

CrowdPose dataset for bottom-up pose estimation.

“CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark” , CVPR’ 2019. More details can be found in the [paper](#).

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

CrowdPose keypoint indexes:

```
0: 'left_shoulder',
1: 'right_shoulder',
2: 'left_elbow',
3: 'right_elbow',
4: 'left_wrist',
5: 'right_wrist',
6: 'left_hip',
7: 'right_hip',
8: 'left_knee',
9: 'right_knee',
10: 'left_ankle',
11: 'right_ankle',
12: 'top_head',
13: 'neck'
```

**参数**

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.



- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.BottomUpMhpDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                           dataset_info=None, test_mode=False)
```

MHPv2.0 dataset for top-down pose estimation.

“Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and A New Benchmark for Multi-Human Parsing” , ACM MM’ 2018. More details can be found in the [paper](#)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

MHP keypoint indexes:

```
0: "right ankle",
1: "right knee",
2: "right hip",
3: "left hip",
4: "left knee",
5: "left ankle",
6: "pelvis",
7: "thorax",
8: "upper neck",
9: "head top",
10: "right wrist",
11: "right elbow",
12: "right shoulder",
13: "left shoulder",
14: "left elbow",
15: "left wrist",
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**class** mmpose.datasets.**Compose** (*transforms*)

Compose a data pipeline with a sequence of transforms.

**参数** **transforms** (*list[dict | callable]*) –Either config dicts of transforms or transform objects.

**class** mmpose.datasets.**DeepFashionDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, subset="", dataset\_info=None, test\_mode=False*)

DeepFashion dataset (full-body clothes) for fashion landmark detection.

“DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations” , CVPR’ 2016.  
“Fashion Landmark Detection in the Wild” , ECCV’ 2016.

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

The dataset contains 3 categories for full-body, upper-body and lower-body.

Fashion landmark indexes for upper-body clothes:

```
0: 'left collar',
1: 'right collar',
2: 'left sleeve',
3: 'right sleeve',
4: 'left hem',
5: 'right hem'
```

Fashion landmark indexes for lower-body clothes:

```
0: 'left waistline',
1: 'right waistline',
2: 'left hem',
3: 'right hem'
```

Fashion landmark indexes for full-body clothes:

```
0: 'left collar',
1: 'right collar',
2: 'left sleeve',
3: 'right sleeve',
4: 'left waistline',
5: 'right waistline',
6: 'left hem',
7: 'right hem'
```

**参数**

- **ann\_file** (*str*) –Path to the annotation file.

- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCK', \*\*kwargs*)

Evaluate freihand keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (`np.ndarray[N,K,3]`): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (`np.ndarray[N,6]`): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (`list[str]`): For example, [ 'img\_00000001.jpg' ]
  - output\_heatmap (`np.ndarray[N, K, H, W]`): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: 'PCK', 'AUC', 'EPE'.

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** `mmpose.datasets.DistributedSampler` (*dataset, num\_replicas=None, rank=None, shuffle=True, seed=0*)

`DistributedSampler` inheriting from `torch.utils.data.DistributedSampler`.

In pytorch of lower versions, there is no `shuffle` argument. This child class will port one to `DistributedSampler`.

```
class mmpose.datasets.Face300WDataset (ann_file, img_prefix, data_cfg, pipeline, dataset_info=None,
                                         test_mode=False)
```

Face300W dataset for top-down face keypoint localization.

“300 faces In-the-wild challenge: Database and results” , Image and Vision Computing (IMAVIS) 2019.

The dataset loads raw images and apply specified transforms to return a dict containing the image tensors and other information.

The landmark annotations follow the 68 points mark-up. The definition can be found in <https://ibug.doc.ic.ac.uk/resources/300-W/>.

#### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
evaluate (results, res_folder=None, metric='NME', **kwargs)
```

Evaluate freihand keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

#### 注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

#### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[1,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[1,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_path (*list[str]*): For example, [ ‘300W/ibug/ image\_018.jpg’ ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.

- **res\_folder** (*str*, *optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str* | *list[str]*) –Metric to be performed. Options: ‘NME’ .

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.FaceAFLWDataset (ann_file, img_prefix, data_cfg, pipeline, dataset_info=None,  
                                         test_mode=False)
```

Face AFLW dataset for top-down face keypoint localization.

“Annotated Facial Landmarks in the Wild: A Large-scale, Real-world Database for Facial Landmark Localization”  
. In Proc. First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies, 2011.

The dataset loads raw images and apply specified transforms to return a dict containing the image tensors and other information.

The landmark annotations follow the 19 points mark-up. The definition can be found in  
<https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/aflw/>

#### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
evaluate (results, res_folder=None, metric='NME', **kwargs)
```

Evaluate freihand keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

#### 注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

#### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[1,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[1,6]*): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_path (*list[str]*): For example, [ ‘ aflw/images/flickr/ 0/image00002.jpg’ ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: ‘NME’ .

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.FaceCOFWDataset (ann_file, img_prefix, data_cfg, pipeline, dataset_info=None,
                                         test_mode=False)
```

Face COFW dataset for top-down face keypoint localization.

“Robust face landmark estimation under occlusion” , ICCV’ 2013.

The dataset loads raw images and apply specified transforms to return a dict containing the image tensors and other information.

The landmark annotations follow the 29 points mark-up. The definition can be found in <http://www.vision.caltech.edu/xpburgos/ICCV13/>.

#### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
evaluate (results, res_folder=None, metric='NME', **kwargs)
```

Evaluate freihand keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

#### 注解:

- batch\_size: N

- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[1,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[1,6]*): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_path (*list[str]*): For example, [ ‘cofw/images/ 000001.jpg’ ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: ‘NME’ .

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** mmpose.datasets.**FaceCocoWholeBodyDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

CocoWholeBodyDataset for face keypoint localization.

Whole-Body Human Pose Estimation in the Wild’ , ECCV’ 2020. More details can be found in the ‘paper’ .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

The face landmark annotations follow the 68 points mark-up.

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='NME'*, *\*\*kwargs*)

Evaluate COCO-WholeBody Face keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

注解:

- `batch_size`: N
- `num_keypoints`: K
- `heatmap height`: H
- `heatmap width`: W

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - `preds` (`np.ndarray[1,K,3]`): The first two dimensions are coordinates, score is the third dimension of the array.
  - `boxes` (`np.ndarray[1,6]`): [`center[0]`, `center[1]`, `scale[0]`, `scale[1]`, `area`, `score`]
  - `image_path` (`list[str]`): For example, [ `'coco/train2017/000000000009.jpg'` ]
  - `output_heatmap` (`np.ndarray[N, K, H, W]`): model outputs.
- **res\_folder** (*str*, *optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: `None`.
- **metric** (*str* | *list[str]*) –Metric to be performed. Options: `'NME'` .

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** `mmpose.datasets.FaceWFLWDataset` (*ann\_file*, *img\_prefix*, *data\_cfg*, *pipeline*, *dataset\_info=None*, *test\_mode=False*)

Face WFLW dataset for top-down face keypoint localization.

“Look at Boundary: A Boundary-Aware Face Alignment Algorithm” , CVPR’ 2018.

The dataset loads raw images and apply specified transforms to return a dict containing the image tensors and other information.

The landmark annotations follow the 98 points mark-up. The definition can be found in <https://wywu.github.io/projects/LAB/WFLW.html>.

参数

- **ann\_file** (*str*) –Path to the annotation file.



- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='NME', \*\*kwargs*)

Evaluate freihand keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[1,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[1,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_path (*list[str]*): For example, [ 'wflw/images/0-Parade/0\_Parade\_marchingband\_1\_1015.jpg' ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: 'NME' .

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** `mmpose.datasets.FreiHandDataset` (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

FreiHand dataset for top-down hand pose estimation.

“FreiHAND: A Dataset for Markerless Capture of Hand Pose and Shape from Single RGB Images” , ICCV’ 2019. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

FreiHand keypoint indexes:

```
0: 'wrist',
1: 'thumb1',
2: 'thumb2',
3: 'thumb3',
4: 'thumb4',
5: 'forefinger1',
6: 'forefinger2',
7: 'forefinger3',
8: 'forefinger4',
9: 'middle_finger1',
10: 'middle_finger2',
11: 'middle_finger3',
12: 'middle_finger4',
13: 'ring_finger1',
14: 'ring_finger2',
15: 'ring_finger3',
16: 'ring_finger4',
17: 'pinky_finger1',
18: 'pinky_finger2',
19: 'pinky_finger3',
20: 'pinky_finger4'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='PCK'*, *\*\*kwargs*)

Evaluate freihand keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_paths (*list[str]*): For example, [ ‘training/rgb/ 00031426.jpg’ ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: ‘PCK’, ‘AUC’, ‘EPE’ .

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** mmpose.datasets.**HandCocoWholeBodyDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

CocoWholeBodyDataset for top-down hand pose estimation.

“Whole-Body Human Pose Estimation in the Wild” , ECCV’ 2020. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

COCO-WholeBody Hand keypoint indexes:

```
0: 'wrist',
1: 'thumb1',
2: 'thumb2',
3: 'thumb3',
4: 'thumb4',
5: 'forefinger1',
```

(下页继续)

(续上页)

```

6: 'forefinger2',
7: 'forefinger3',
8: 'forefinger4',
9: 'middle_finger1',
10: 'middle_finger2',
11: 'middle_finger3',
12: 'middle_finger4',
13: 'ring_finger1',
14: 'ring_finger2',
15: 'ring_finger3',
16: 'ring_finger4',
17: 'pinky_finger1',
18: 'pinky_finger2',
19: 'pinky_finger3',
20: 'pinky_finger4'

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCK', \*\*kwargs*)

Evaluate COCO-WholeBody Hand keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:

- `preds` (`np.ndarray[N,K,3]`): The first two dimensions are coordinates, score is the third dimension of the array.
- `boxes` (`np.ndarray[N,6]`): [`center[0]`, `center[1]`, `scale[0]`, `scale[1]`, `area`, `score`]
- `image_paths` (`list[str]`): For example, [ 'Test/source/0.jpg' ]
- `output_heatmap` (`np.ndarray[N, K, H, W]`): model outputs.
- **`res_folder`** (`str`, *optional*) – The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **`metric`** (`str` | `list[str]`) – Metric to be performed. Options: 'PCK', 'AUC', 'EPE'.

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.InterHand2DDataset (ann_file, camera_file, joint_file, img_prefix, data_cfg,  
pipeline, dataset_info=None, test_mode=False)
```

InterHand2.6M 2D dataset for top-down hand pose estimation.

“InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image” , ECCV’ 2020. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

InterHand2.6M keypoint indexes:

```
0: 'thumb4',
1: 'thumb3',
2: 'thumb2',
3: 'thumb1',
4: 'forefinger4',
5: 'forefinger3',
6: 'forefinger2',
7: 'forefinger1',
8: 'middle_finger4',
9: 'middle_finger3',
10: 'middle_finger2',
11: 'middle_finger1',
12: 'ring_finger4',
13: 'ring_finger3',
14: 'ring_finger2',
15: 'ring_finger1',
16: 'pinky_finger4',
17: 'pinky_finger3',
```

(下页继续)

(续上页)

```

18: 'pinky_finger2',
19: 'pinky_finger1',
20: 'wrist'

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **camera\_file** (*str*) –Path to the camera file.
- **joint\_file** (*str*) –Path to the joint file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*str*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCK', \*\*kwargs*)

Evaluate interhand2d keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (*list[str]*): For example, [ ‘Capture12/0390\_dh\_touchROM/cam410209/image62434.jpg’ ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.

- **res\_folder**(*str*, *optional*) – The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric**(*str* | *list[str]*) – Metric to be performed. Options: ‘PCK’, ‘AUC’, ‘EPE’.

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.InterHand3DDataset (ann_file, camera_file, joint_file, img_prefix, data_cfg,
                                           pipeline, use_gt_root_depth=True,
                                           rootnet_result_file=None, dataset_info=None,
                                           test_mode=False)
```

InterHand2.6M 3D dataset for top-down hand pose estimation.

“InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image”, ECCV’ 2020. More details can be found in the [paper](#).

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

InterHand2.6M keypoint indexes:

```
0: 'r_thumb4',
1: 'r_thumb3',
2: 'r_thumb2',
3: 'r_thumb1',
4: 'r_index4',
5: 'r_index3',
6: 'r_index2',
7: 'r_index1',
8: 'r_middle4',
9: 'r_middle3',
10: 'r_middle2',
11: 'r_middle1',
12: 'r_ring4',
13: 'r_ring3',
14: 'r_ring2',
15: 'r_ring1',
16: 'r_pinky4',
17: 'r_pinky3',
18: 'r_pinky2',
19: 'r_pinky1',
20: 'r_wrist',
21: 'l_thumb4',
22: 'l_thumb3',
```

(下页继续)

(续上页)

```

23: 'l_thumb2',
24: 'l_thumb1',
25: 'l_index4',
26: 'l_index3',
27: 'l_index2',
28: 'l_index1',
29: 'l_middle4',
30: 'l_middle3',
31: 'l_middle2',
32: 'l_middle1',
33: 'l_ring4',
34: 'l_ring3',
35: 'l_ring2',
36: 'l_ring1',
37: 'l_pinky4',
38: 'l_pinky3',
39: 'l_pinky2',
40: 'l_pinky1',
41: 'l_wrist'

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **camera\_file** (*str*) –Path to the camera file.
- **joint\_file** (*str*) –Path to the joint file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **use\_gt\_root\_depth** (*bool*) –Using the ground truth depth of the wrist or given depth from rootnet\_result\_file.
- **rootnet\_result\_file** (*str*) –Path to the wrist depth file.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*str*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='MPJPE'*, *\*\*kwargs*)

Evaluate interhand2d keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

注解:



- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (np.ndarray[N,K,3]): The first two dimensions are coordinates, score is the third dimension of the array.
  - hand\_type (np.ndarray[N, 4]): The first two dimensions are hand type, scores is the last two dimensions.
  - rel\_root\_depth (np.ndarray[N]): The relative depth of left wrist and right wrist.
  - boxes (np.ndarray[N,6]): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_paths (list[str]): For example, [ ‘Capture6/0012\_aokay\_upright/cam410061/image4996.jpg’ ]
  - output\_heatmap (np.ndarray[N, K, H, W]): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: ‘MRRPE’ , ‘MPJPE’ , ‘Handedness\_acc’ .

**返回** Evaluation results for evaluation metric.

**返回类型** dict

**class** mmpose.datasets.**MeshAdversarialDataset** (*train\_dataset, adversarial\_dataset*)

Mix Dataset for the adversarial training in 3D human mesh estimation task.

The dataset combines data from two datasets and return a dict containing data from two datasets.

### 参数

- **train\_dataset** (*Dataset*) –Dataset for 3D human mesh estimation.
- **adversarial\_dataset** (*Dataset*) –Dataset for adversarial learning, provides real SMPL parameters.

**class** mmpose.datasets.**MeshH36MDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, test\_mode=False*)

Human3.6M Dataset for 3D human mesh estimation. It inherits all function from MeshBaseDataset and has its own evaluate function.

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

#### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*outputs, res\_folder, metric='joint\_error', logger=None*)

Evaluate 3D keypoint results.

**class** mmpose.datasets.**MeshMixDataset** (*configs, partition*)

Mix Dataset for 3D human mesh estimation.

The dataset combines data from multiple datasets (MeshBaseDataset) and sample the data from different datasets with the provided proportions. The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

#### 参数

- **configs** (*list*) –List of configs for multiple datasets.
- **partition** (*list*) –Sample proportion of multiple datasets. The length of partition should be same with that of configs. The elements of it should be non-negative and is not necessary summing up to one.

#### 示例

```
>>> from mmpose.datasets import MeshMixDataset
>>> data_cfg = dict(
>>>     image_size=[256, 256],
>>>     iuv_size=[64, 64],
>>>     num_joints=24,
>>>     use_IUV=True,
>>>     uv_type='BF')
>>>
>>> mix_dataset = MeshMixDataset(
>>>     configs=[
>>>         dict(
>>>             ann_file='tests/data/h36m/test_h36m.npz',
>>>             img_prefix='tests/data/h36m',
>>>             data_cfg=data_cfg,
```

(下页继续)

(续上页)

```

>>>         pipeline=[]),
>>>         dict(
>>>             ann_file='tests/data/h36m/test_h36m.npz',
>>>             img_prefix='tests/data/h36m',
>>>             data_cfg=data_cfg,
>>>             pipeline=[]),
>>>     ],
>>>     partition=[0.6, 0.4])

```

**class** mmpose.datasets.**MoshDataset** (*ann\_file, pipeline, test\_mode=False*)

Mosh Dataset for the adversarial training in 3D human mesh estimation task.

The dataset return a dict containing real-world SMPL parameters.

#### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**class** mmpose.datasets.**NVGestureDataset** (*ann\_file, vid\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

NVGesture dataset for gesture recognition.

“Online Detection and Classification of Dynamic Hand Gestures With Recurrent 3D Convolutional Neural Network” , Conference on Computer Vision and Pattern Recognition (CVPR) 2016.

The dataset loads raw videos and apply specified transforms to return a dict containing the image tensors and other information.

#### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **vid\_prefix** (*str*) –Path to a directory where videos are held.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='AP', \*\*kwargs*)

Evaluate nvgesture recognition results. The gesture prediction results will be saved in `${res_folder}/result_gesture.json`.

---

注解:

- batch\_size: N
- heatmap length: L

### 参数

- **results** (*dict*) –Testing results containing the following items: - logits (*dict[str, torch.tensor[N,25,L]]*): For each item, the key represents the modality of input video, while the value represents the prediction of gesture. Three dimensions represent batch, category and temporal length, respectively. - label (*np.ndarray[N]*): [center[0], center[1], scale[0], scale[1],area, score]
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: ‘AP’ .

**返回** Evaluation results for evaluation metric.

**返回类型** dict

```
class mmpose.datasets.OneHand10KDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                         dataset_info=None, test_mode=False)
```

OneHand10K dataset for top-down hand pose estimation.

“Mask-pose Cascaded CNN for 2D Hand Pose Estimation from Single Color Images” , TCSVT’ 2019. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

OneHand10K keypoint indexes:

```
0: 'wrist',
1: 'thumb1',
2: 'thumb2',
3: 'thumb3',
4: 'thumb4',
5: 'forefinger1',
6: 'forefinger2',
7: 'forefinger3',
8: 'forefinger4',
9: 'middle_finger1',
10: 'middle_finger2',
11: 'middle_finger3',
12: 'middle_finger4',
13: 'ring_finger1',
```

(下页继续)

(续上页)

```

14: 'ring_finger2',
15: 'ring_finger3',
16: 'ring_finger4',
17: 'pinky_finger1',
18: 'pinky_finger2',
19: 'pinky_finger3',
20: 'pinky_finger4'

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCK', \*\*kwargs*)

Evaluate onehand10k keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (*list[str]*): For example, [ 'Test/source/0.jpg' ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.

- **res\_folder**(*str*, *optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric**(*str* | *list[str]*) –Metric to be performed. Options: ‘PCK’, ‘AUC’, ‘EPE’.

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** mmpose.datasets.**PanopticDataset** (*ann\_file*, *img\_prefix*, *data\_cfg*, *pipeline*, *dataset\_info=None*, *test\_mode=False*)

Panoptic dataset for top-down hand pose estimation.

“Hand Keypoint Detection in Single Images using Multiview Bootstrapping”, CVPR’ 2017. More details can be found in the [paper](#).

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

Panoptic keypoint indexes:

```
0: 'wrist',
1: 'thumb1',
2: 'thumb2',
3: 'thumb3',
4: 'thumb4',
5: 'forefinger1',
6: 'forefinger2',
7: 'forefinger3',
8: 'forefinger4',
9: 'middle_finger1',
10: 'middle_finger2',
11: 'middle_finger3',
12: 'middle_finger4',
13: 'ring_finger1',
14: 'ring_finger2',
15: 'ring_finger3',
16: 'ring_finger4',
17: 'pinky_finger1',
18: 'pinky_finger2',
19: 'pinky_finger3',
20: 'pinky_finger4'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.

- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCKh', \*\*kwargs*)

Evaluate panoptic keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (*list[str]*): For example, [ 'hand\_labels/ manual\_test/000648952\_02\_1.jpg' ]
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: 'PCKh', 'AUC', 'EPE'.

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** `mmpose.datasets.TopDownAicDataset` (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

AicDataset dataset for top-down pose estimation.

“AI Challenger : A Large-scale Dataset for Going Deeper in Image Understanding” , arXiv’ 2017. More details can be found in the [paper](#)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

AIC keypoint indexes:

```
0: "right_shoulder",
1: "right_elbow",
2: "right_wrist",
3: "left_shoulder",
4: "left_elbow",
5: "left_wrist",
6: "right_hip",
7: "right_knee",
8: "right_ankle",
9: "left_hip",
10: "left_knee",
11: "left_ankle",
12: "head_top",
13: "neck"
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.TopDownCocoDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                           dataset_info=None, test_mode=False)
```

CocoDataset dataset for top-down pose estimation.

“Microsoft COCO: Common Objects in Context” , ECCV’ 2014. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

COCO keypoint indexes:



```

0: 'nose',
1: 'left_eye',
2: 'right_eye',
3: 'left_ear',
4: 'right_ear',
5: 'left_shoulder',
6: 'right_shoulder',
7: 'left_elbow',
8: 'right_elbow',
9: 'left_wrist',
10: 'right_wrist',
11: 'left_hip',
12: 'right_hip',
13: 'left_knee',
14: 'right_knee',
15: 'left_ankle',
16: 'right_ankle'

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='mAP', \*\*kwargs*)

Evaluate coco keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_paths (*list[str]*): For example, [ ‘data/coco/val2017 /0000000393226.jpg’ ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap
  - bbox\_id (*list(int)*).
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: ‘mAP’ .

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** mmpose.datasets.**TopDownCocoWholeBodyDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

CocoWholeBodyDataset dataset for top-down pose estimation.

“Whole-Body Human Pose Estimation in the Wild” , ECCV’ 2020. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

COCO-WholeBody keypoint indexes:

```
0-16: 17 body keypoints,
17-22: 6 foot keypoints,
23-90: 68 face keypoints,
91-132: 42 hand keypoints
```

In total, we have 133 keypoints for wholebody pose estimation.

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.TopDownCrowdPoseDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                              dataset_info=None, test_mode=False)
```

CrowdPoseDataset dataset for top-down pose estimation.

“CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark” , CVPR’ 2019. More details can be found in the [paper](#).

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

CrowdPose keypoint indexes:

```
0: 'left_shoulder',  
1: 'right_shoulder',  
2: 'left_elbow',  
3: 'right_elbow',  
4: 'left_wrist',  
5: 'right_wrist',  
6: 'left_hip',  
7: 'right_hip',  
8: 'left_knee',  
9: 'right_knee',  
10: 'left_ankle',  
11: 'right_ankle',  
12: 'top_head',  
13: 'neck'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.TopDownFreiHandDataset (*args, **kwargs)
```

Deprecated TopDownFreiHandDataset.

```
evaluate (cfg, preds, output_dir, *args, **kwargs)
```

Evaluate keypoint results.

```
class mmpose.datasets.TopDownH36MDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                              dataset_info=None, test_mode=False)
```

Human3.6M dataset for top-down 2D pose estimation.

“Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments” , TPAMI’2014. More details can be found in the [paper](#).

Human3.6M keypoint indexes:

```
0: 'root (pelvis)',
1: 'right_hip',
2: 'right_knee',
3: 'right_foot',
4: 'left_hip',
5: 'left_knee',
6: 'left_foot',
7: 'spine',
8: 'thorax',
9: 'neck_base',
10: 'head',
11: 'left_shoulder',
12: 'left_elbow',
13: 'left_wrist',
14: 'right_shoulder',
15: 'right_elbow',
16: 'right_wrist'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCK', \*\*kwargs*)

Evaluate human3.6m 2d keypoint results. The pose prediction results will be saved in `{res_folder}/result_keypoints.json`.

---

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) – Testing results containing the following items:
  - **preds** (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - **boxes** (*np.ndarray[N,6]*): [**center[0]**, **center[1]**, **scale[0]**, **scale[1]**, **area**, **score**]
  - **image\_paths** (*list[str]*): For example, [ **‘data/coco/val2017/000000393226.jpg’** ]
  - **heatmap** (*np.ndarray[N, K, H, W]*): model output heatmap
  - **bbox\_id** (*list(int)*).
- **res\_folder** (*str, optional*) – The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) – Metric to be performed. Defaults: **‘PCK’** .

**返回** Evaluation results for evaluation metric.

**返回类型** dict

```
class mmpose.datasets.TopDownJhmdbDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                           dataset_info=None, test_mode=False)
```

JhmdbDataset dataset for top-down pose estimation.

“Towards understanding action recognition” , ICCV’ 2013. More details can be found in the [paper](#)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

sub-JHMDB keypoint indexes:

```
0: "neck",
1: "belly",
2: "head",
3: "right_shoulder",
4: "left_shoulder",
5: "right_hip",
6: "left_hip",
7: "right_elbow",
8: "left_elbow",
9: "right_knee",
10: "left_knee",
11: "right_wrist",
12: "left_wrist",
```

(下页继续)

(续上页)

```
13: "right_ankle",
14: "left_ankle"
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='PCK'*, *\*\*kwargs*)

Evaluate onhand10k keypoint results. The pose prediction results will be saved in *\${res\_folder}/result\_keypoints.json*.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_path (*list[str]*)
  - output\_heatmap (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str*, *optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Options: ‘PCK’, ‘tPCK’. PCK means normalized by the bounding boxes, while tPCK means normalized by the torso size.

返回 Evaluation results for evaluation metric.

返回类型 dict

**class** mmpose.datasets.**TopDownMhpDataset** (*ann\_file, img\_prefix, data\_cfg, pipeline, dataset\_info=None, test\_mode=False*)

MHPv2.0 dataset for top-down pose estimation.

“Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and A New Benchmark for Multi-Human Parsing” , ACM MM’ 2018. More details can be found in the [paper](#)

Note that, the evaluation metric used here is mAP (adapted from COCO), which may be different from the official evaluation codes. ‘<https://github.com/ZhaoJ9014/Multi-Human-Parsing/tree/master/>’ ‘Evaluation/Multi-Human-Pose’ Please be cautious if you use the results in papers.

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

MHP keypoint indexes:

```
0: "right ankle",
1: "right knee",
2: "right hip",
3: "left hip",
4: "left knee",
5: "left ankle",
6: "pelvis",
7: "thorax",
8: "upper neck",
9: "head top",
10: "right wrist",
11: "right elbow",
12: "right shoulder",
13: "left shoulder",
14: "left elbow",
15: "left wrist",
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.TopDownMpiiDataset (ann_file, img_prefix, data_cfg, pipeline,  
                                         dataset_info=None, test_mode=False)
```

MPII Dataset for top-down pose estimation.

“2D Human Pose Estimation: New Benchmark and State of the Art Analysis” ,CVPR’ 2014. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

MPII keypoint indexes:

```
0: 'right_ankle'  
1: 'right_knee',  
2: 'right_hip',  
3: 'left_hip',  
4: 'left_knee',  
5: 'left_ankle',  
6: 'pelvis',  
7: 'thorax',  
8: 'upper_neck',  
9: 'head_top',  
10: 'right_wrist',  
11: 'right_elbow',  
12: 'right_shoulder',  
13: 'left_shoulder',  
14: 'left_elbow',  
15: 'left_wrist'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='PCKh'*, *\*\*kwargs*)

Evaluate PCKh for MPII dataset. Adapted from <https://github.com/leoxiaobin/deep-high-resolution-net.pytorch> Copyright (c) Microsoft, under the MIT License.

---

注解:



- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_paths (*list[str]*): For example, [ ‘/val2017/000000 397133.jpg’ ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap.
- **res\_folder** (*str, optional*) –The folder to save the testing results. Default: None.
- **metric** (*str | list[str]*) –Metrics to be performed. Defaults: ‘PCKh’ .

返回 PCKh for each joint

返回类型 dict

```
class mmpose.datasets.TopDownMpiiTrbDataset (ann_file, img_prefix, data_cfg, pipeline,
                                              dataset_info=None, test_mode=False)
```

MPII-TRB Dataset dataset for top-down pose estimation.

“TRB: A Novel Triplet Representation for Understanding 2D Human Body” , ICCV’ 2019. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

MPII-TRB keypoint indexes:

```
0: 'left_shoulder'
1: 'right_shoulder'
2: 'left_elbow'
3: 'right_elbow'
4: 'left_wrist'
5: 'right_wrist'
6: 'left_hip'
7: 'right_hip'
8: 'left_knee'
```

(下页继续)

(续上页)

```
9: 'right_knee'
10: 'left_ankle'
11: 'right_ankle'
12: 'head'
13: 'neck'

14: 'right_neck'
15: 'left_neck'
16: 'medial_right_shoulder'
17: 'lateral_right_shoulder'
18: 'medial_right_bow'
19: 'lateral_right_bow'
20: 'medial_right_wrist'
21: 'lateral_right_wrist'
22: 'medial_left_shoulder'
23: 'lateral_left_shoulder'
24: 'medial_left_bow'
25: 'lateral_left_bow'
26: 'medial_left_wrist'
27: 'lateral_left_wrist'
28: 'medial_right_hip'
29: 'lateral_right_hip'
30: 'medial_right_knee'
31: 'lateral_right_knee'
32: 'medial_right_ankle'
33: 'lateral_right_ankle'
34: 'medial_left_hip'
35: 'lateral_left_hip'
36: 'medial_left_knee'
37: 'lateral_left_knee'
38: 'medial_left_ankle'
39: 'lateral_left_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='PCKh'*, *\*\*kwargs*)

Evaluate PCKh for MPII-TRB dataset.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_paths (*list[str]*): For example, [ ‘/val2017/ 000000397133.jpg’ ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap.
  - bbox\_ids (*list[str]*): For example, [ ‘27407’ ].
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metrics to be performed. Defaults: ‘PCKh’ .

返回 PCKh for each joint

返回类型 dict

**class** mmpose.datasets.**TopDownOCHumanDataset** (*ann\_file*, *img\_prefix*, *data\_cfg*, *pipeline*,  
*dataset\_info=None*, *test\_mode=False*)

OCHuman dataset for top-down pose estimation.

“Pose2Seg: Detection Free Human Instance Segmentation” , CVPR’ 2019. More details can be found in the [paper](#) .

“Occluded Human (OCHuman)” dataset contains 8110 heavily occluded human instances within 4731 images. OCHuman dataset is designed for validation and testing. To evaluate on OCHuman, the model should be trained on COCO training set, and then test the robustness of the model to occlusion using OCHuman.

OCHuman keypoint indexes (same as COCO):

```
0: 'nose',
1: 'left_eye',
2: 'right_eye',
3: 'left_ear',
4: 'right_ear',
5: 'left_shoulder',
6: 'right_shoulder',
7: 'left_elbow',
8: 'right_elbow',
9: 'left_wrist',
10: 'right_wrist',
11: 'left_hip',
12: 'right_hip',
13: 'left_knee',
14: 'right_knee',
15: 'left_ankle',
16: 'right_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**class** mmpose.datasets.**TopDownOneHand10KDataset** (*\*args, \*\*kwargs*)

Deprecated TopDownOneHand10KDataset.

**evaluate** (*cfg, preds, output\_dir, \*args, \*\*kwargs*)

Evaluate keypoint results.

**class** mmpose.datasets.**TopDownPanopticDataset** (*\*args, \*\*kwargs*)

Deprecated TopDownPanopticDataset.

**evaluate** (*cfg, preds, output\_dir, \*args, \*\*kwargs*)

Evaluate keypoint results.

**class** mmpose.datasets.**TopDownPoseTrack18Dataset** (*ann\_file, img\_prefix, data\_cfg, pipeline,*  
*dataset\_info=None, test\_mode=False*)

PoseTrack18 dataset for top-down pose estimation.

“Posetrack: A benchmark for human pose estimation and tracking” , CVPR’ 2018. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

PoseTrack2018 keypoint indexes:

```
0: 'nose',
1: 'head_bottom',
2: 'head_top',
3: 'left_ear',
4: 'right_ear',
5: 'left_shoulder',
6: 'right_shoulder',
7: 'left_elbow',
8: 'right_elbow',
9: 'left_wrist',
10: 'right_wrist',
11: 'left_hip',
12: 'right_hip',
13: 'left_knee',
14: 'right_knee',
15: 'left_ankle',
16: 'right_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='mAP'*, *\*\*kwargs*)

Evaluate posetrack keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

### 注解:

- num\_keypoints: K

## 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_paths (*list[str]*): For example, [ ‘val/010016\_mpii\_test /000024.jpg’ ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap.
  - bbox\_id (*list(int)*)
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: ‘mAP’ .

**返回** Evaluation results for evaluation metric.

**返回类型** dict

```
class mmpose.datasets.TopDownPoseTrack18VideoDataset (ann_file, img_prefix, data_cfg,  
pipeline, dataset_info=None,  
test_mode=False, ph_fill_len=6)
```

PoseTrack18 dataset for top-down pose estimation.

“Posetrack: A benchmark for human pose estimation and tracking” , CVPR’ 2018. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

PoseTrack2018 keypoint indexes:

```
0: 'nose',  
1: 'head_bottom',  
2: 'head_top',  
3: 'left_ear',  
4: 'right_ear',  
5: 'left_shoulder',  
6: 'right_shoulder',  
7: 'left_elbow',  
8: 'right_elbow',  
9: 'left_wrist',  
10: 'right_wrist',  
11: 'left_hip',  
12: 'right_hip',  
13: 'left_knee',
```

(下页继续)

(续上页)

```

14: 'right_knee',
15: 'left_ankle',
16: 'right_ankle'

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where videos/images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.
- **ph\_fill\_len** (*int*) –The length of the placeholder to fill in the image filenames, default: 6 in PoseTrack18.

**evaluate** (*results, res\_folder=None, metric='mAP', \*\*kwargs*)

Evaluate posetrack keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

### 注解:

- num\_keypoints: K

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (*list[str]*): For example, [ 'val/010016\_mpii\_test/000024.jpg' ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap.
  - bbox\_id (*list(int)*)
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: 'mAP' .

**返回** Evaluation results for evaluation metric.

返回类型 dict

```
mmpose.datasets.build_dataloader (dataset, samples_per_gpu, workers_per_gpu, num_gpus=1,  
                                   dist=True, shuffle=True, seed=None, drop_last=True,  
                                   pin_memory=True, **kwargs)
```

Build PyTorch DataLoader.

In distributed training, each GPU/process has a dataloader. In non-distributed training, there is only one dataloader for all GPUs.

参数

- **dataset** (*Dataset*) –A PyTorch dataset.
- **samples\_per\_gpu** (*int*) –Number of training samples on each GPU, i.e., batch size of each GPU.
- **workers\_per\_gpu** (*int*) –How many subprocesses to use for data loading for each GPU.
- **num\_gpus** (*int*) –Number of GPUs. Only used in non-distributed training.
- **dist** (*bool*) –Distributed training/test or not. Default: True.
- **shuffle** (*bool*) –Whether to shuffle the data at every epoch. Default: True.
- **drop\_last** (*bool*) –Whether to drop the last incomplete batch in epoch. Default: True
- **pin\_memory** (*bool*) –Whether to use pin\_memory in DataLoader. Default: True
- **kwargs** –any keyword argument to be used to initialize DataLoader

返回 A PyTorch dataloader.

返回类型 DataLoader

```
mmpose.datasets.build_dataset (cfg, default_args=None)
```

Build a dataset from config dict.

参数

- **cfg** (*dict*) –Config dict. It should at least contain the key “type” .
- **default\_args** (*dict, optional*) –Default initialization arguments. Default: None.

返回 The constructed dataset.

返回类型 Dataset



## 49.1 datasets

```
class mmpose.datasets.datasets.top_down.TopDownAicDataset (ann_file, img_prefix, data_cfg,  
                                                         pipeline, dataset_info=None,  
                                                         test_mode=False)
```

AicDataset dataset for top-down pose estimation.

“AI Challenger : A Large-scale Dataset for Going Deeper in Image Understanding” , arXiv’ 2017. More details can be found in the [paper](#)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

AIC keypoint indexes:

```
0: "right_shoulder",  
1: "right_elbow",  
2: "right_wrist",  
3: "left_shoulder",  
4: "left_elbow",  
5: "left_wrist",  
6: "right_hip",  
7: "right_knee",  
8: "right_ankle",  
9: "left_hip",  
10: "left_knee",  
11: "left_ankle",  
12: "head_top",  
13: "neck"
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.top_down.TopDownCocoDataset (ann_file, img_prefix,  
data_cfg, pipeline,  
dataset_info=None,  
test_mode=False)
```

CocoDataset dataset for top-down pose estimation.

“Microsoft COCO: Common Objects in Context” , ECCV’ 2014. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

COCO keypoint indexes:

```
0: 'nose',  
1: 'left_eye',  
2: 'right_eye',  
3: 'left_ear',  
4: 'right_ear',  
5: 'left_shoulder',  
6: 'right_shoulder',  
7: 'left_elbow',  
8: 'right_elbow',  
9: 'left_wrist',  
10: 'right_wrist',  
11: 'left_hip',  
12: 'right_hip',  
13: 'left_knee',  
14: 'right_knee',  
15: 'left_ankle',  
16: 'right_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='mAP', \*\*kwargs*)

Evaluate coco keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_paths (*list[str]*): For example, [ ‘data/coco/val2017/000000393226.jpg’ ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap
  - bbox\_id (*list(int)*).
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: ‘mAP’ .

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.datasets.top_down.TopDownCocoWholeBodyDataset (ann_file,
                                                                    img_prefix,
                                                                    data_cfg,
                                                                    pipeline,
                                                                    dataset_info=None,
                                                                    test_mode=False)
```

CocoWholeBodyDataset dataset for top-down pose estimation.

“Whole-Body Human Pose Estimation in the Wild” , ECCV’ 2020. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

COCO-WholeBody keypoint indexes:

```
0-16: 17 body keypoints,
17-22: 6 foot keypoints,
23-90: 68 face keypoints,
91-132: 42 hand keypoints
```

In total, we have 133 keypoints for wholebody pose estimation.

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.top_down.TopDownCrowdPoseDataset (ann_file, img_prefix,
                                                                data_cfg, pipeline,
                                                                dataset_info=None,
                                                                test_mode=False)
```

CrowdPoseDataset dataset for top-down pose estimation.

“CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark” , CVPR’ 2019. More details can be found in the [paper](#).

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

CrowdPose keypoint indexes:

```
0: 'left_shoulder',
1: 'right_shoulder',
2: 'left_elbow',
3: 'right_elbow',
4: 'left_wrist',
5: 'right_wrist',
6: 'left_hip',
7: 'right_hip',
8: 'left_knee',
9: 'right_knee',
10: 'left_ankle',
11: 'right_ankle',
```

(下页继续)

(续上页)

```
12: 'top_head',
13: 'neck'
```

**参数**

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.top_down.TopDownH36MDataset (ann_file, img_prefix,  
                                                             data_cfg, pipeline,  
                                                             dataset_info=None,  
                                                             test_mode=False)
```

Human3.6M dataset for top-down 2D pose estimation.

“Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments” , TPAMI’2014. More details can be found in the [paper](#).

Human3.6M keypoint indexes:

```
0: 'root (pelvis)',
1: 'right_hip',
2: 'right_knee',
3: 'right_foot',
4: 'left_hip',
5: 'left_knee',
6: 'left_foot',
7: 'spine',
8: 'thorax',
9: 'neck_base',
10: 'head',
11: 'left_shoulder',
12: 'left_elbow',
13: 'left_wrist',
14: 'right_shoulder',
15: 'right_elbow',
16: 'right_wrist'
```

**参数**

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='PCK'*, *\*\*kwargs*)

Evaluate human3.6m 2d keypoint results. The pose prediction results will be saved in *#{res\_folder}/result\_keypoints.json*.

---

注解:

- batch\_size: N
  - num\_keypoints: K
  - heatmap height: H
  - heatmap width: W
- 

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - **preds** (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - **boxes** (*np.ndarray[N,6]*): [*center[0]*, *center[1]*, *scale[0]*, *scale[1]*, *area*, *score*]
  - **image\_paths** (*list[str]*): For example, [ *'data/coco/val2017/000000393226.jpg'* ]
  - **heatmap** (*np.ndarray[N, K, H, W]*): model output heatmap
  - **bbox\_id** (*list(int)*).
- **res\_folder** (*str*, *optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: *'PCK'* .

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.datasets.top_down.TopDownHalpeDataset (ann_file, img_prefix,
                                                             data_cfg, pipeline,
                                                             dataset_info=None,
                                                             test_mode=False)
```

HalpeDataset for top-down pose estimation.

[‘https://github.com/Fang-Haoshu/Halpe-FullBody’](https://github.com/Fang-Haoshu/Halpe-FullBody)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

Halpe keypoint indexes:

```
0-19: 20 body keypoints,
20-25: 6 foot keypoints,
26-93: 68 face keypoints,
94-135: 42 hand keypoints
```

In total, we have 136 keypoints for wholebody pose estimation.

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.top_down.TopDownJhmdbDataset (ann_file, img_prefix,
                                                             data_cfg, pipeline,
                                                             dataset_info=None,
                                                             test_mode=False)
```

JhmdbDataset dataset for top-down pose estimation.

“Towards understanding action recognition” , ICCV’ 2013. More details can be found in the [paper](#)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

sub-JHMDB keypoint indexes:

```
0: "neck",
1: "belly",
2: "head",
```

(下页继续)

(续上页)

```

3: "right_shoulder",
4: "left_shoulder",
5: "right_hip",
6: "left_hip",
7: "right_elbow",
8: "left_elbow",
9: "right_knee",
10: "left_knee",
11: "right_wrist",
12: "left_wrist",
13: "right_ankle",
14: "left_ankle"

```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='PCK'*, *\*\*kwargs*)

Evaluate onehand10k keypoint results. The pose prediction results will be saved in *\${res\_folder}/result\_keypoints.json*.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]



- `image_path` (`list[str]`)
- `output_heatmap` (`np.ndarray[N, K, H, W]`): model outputs.
- **`res_folder`** (`str`, *optional*) – The folder to save the testing results. If not specified, a temp folder will be created. Default: `None`.
- **`metric`** (`str` | `list[str]`) – Metric to be performed. Options: ‘PCK’, ‘tPCK’. PCK means normalized by the bounding boxes, while tPCK means normalized by the torso size.

返回 Evaluation results for evaluation metric.

返回类型 `dict`

```
class mmpose.datasets.datasets.top_down.TopDownMhpDataset (ann_file, img_prefix, data_cfg,
                                                         pipeline, dataset_info=None,
                                                         test_mode=False)
```

MHPv2.0 dataset for top-down pose estimation.

“Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and A New Benchmark for Multi-Human Parsing”, ACM MM’ 2018. More details can be found in the [paper](#)

Note that, the evaluation metric used here is mAP (adapted from COCO), which may be different from the official evaluation codes. ‘<https://github.com/ZhaoJ9014/Multi-Human-Parsing/tree/master/>’ ‘Evaluation/Multi-Human-Pose’ Please be cautious if you use the results in papers.

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

MHP keypoint indexes:

```
0: "right ankle",
1: "right knee",
2: "right hip",
3: "left hip",
4: "left knee",
5: "left ankle",
6: "pelvis",
7: "thorax",
8: "upper neck",
9: "head top",
10: "right wrist",
11: "right elbow",
12: "right shoulder",
13: "left shoulder",
14: "left elbow",
15: "left wrist",
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.top_down.TopDownMpiiDataset (ann_file, img_prefix,  
                                                            data_cfg, pipeline,  
                                                            dataset_info=None,  
                                                            test_mode=False)
```

MPII Dataset for top-down pose estimation.

“2D Human Pose Estimation: New Benchmark and State of the Art Analysis” ,CVPR’ 2014. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

MPII keypoint indexes:

```
0: 'right_ankle'  
1: 'right_knee',  
2: 'right_hip',  
3: 'left_hip',  
4: 'left_knee',  
5: 'left_ankle',  
6: 'pelvis',  
7: 'thorax',  
8: 'upper_neck',  
9: 'head_top',  
10: 'right_wrist',  
11: 'right_elbow',  
12: 'right_shoulder',  
13: 'left_shoulder',  
14: 'left_elbow',  
15: 'left_wrist'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.

- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCKh', \*\*kwargs*)

Evaluate PCKh for MPII dataset. Adapted from <https://github.com/leoxiaobin/deep-high-resolution-net.pytorch> Copyright (c) Microsoft, under the MIT License.

注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (*list[str]*): For example, [ ‘/val2017/0000000 397133.jpg’ ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap.
- **res\_folder** (*str, optional*) –The folder to save the testing results. Default: None.
- **metric** (*str | list[str]*) –Metrics to be performed. Defaults: ‘PCKh’ .

返回 PCKh for each joint

返回类型 dict

```
class mmpose.datasets.datasets.top_down.TopDownMpiiTrbDataset (ann_file, img_prefix,  
                                                             data_cfg, pipeline,  
                                                             dataset_info=None,  
                                                             test_mode=False)
```

MPII-TRB Dataset dataset for top-down pose estimation.

“TRB: A Novel Triplet Representation for Understanding 2D Human Body” , ICCV’ 2019. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

MPII-TRB keypoint indexes:

```
0: 'left_shoulder'
1: 'right_shoulder'
2: 'left_elbow'
3: 'right_elbow'
4: 'left_wrist'
5: 'right_wrist'
6: 'left_hip'
7: 'right_hip'
8: 'left_knee'
9: 'right_knee'
10: 'left_ankle'
11: 'right_ankle'
12: 'head'
13: 'neck'

14: 'right_neck'
15: 'left_neck'
16: 'medial_right_shoulder'
17: 'lateral_right_shoulder'
18: 'medial_right_bow'
19: 'lateral_right_bow'
20: 'medial_right_wrist'
21: 'lateral_right_wrist'
22: 'medial_left_shoulder'
23: 'lateral_left_shoulder'
24: 'medial_left_bow'
25: 'lateral_left_bow'
26: 'medial_left_wrist'
27: 'lateral_left_wrist'
28: 'medial_right_hip'
29: 'lateral_right_hip'
30: 'medial_right_knee'
31: 'lateral_right_knee'
32: 'medial_right_ankle'
33: 'lateral_right_ankle'
34: 'medial_left_hip'
35: 'lateral_left_hip'
36: 'medial_left_knee'
37: 'lateral_left_knee'
38: 'medial_left_ankle'
```

(下页继续)

(续上页)

```
39: 'lateral_left_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results, res\_folder=None, metric='PCKh', \*\*kwargs*)

Evaluate PCKh for MPII-TRB dataset.

### 注解:

- batch\_size: N
- num\_keypoints: K
- heatmap height: H
- heatmap width: W

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (*np.ndarray[N,6]*): [center[0], center[1], scale[0], scale[1], area, score]
  - image\_paths (*list[str]*): For example, [ ‘/val2017/ 000000397133.jpg’ ]
  - heatmap (*np.ndarray[N, K, H, W]*): model output heatmap.
  - bbox\_ids (*list[str]*): For example, [ ‘27407’ ].
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metrics to be performed. Defaults: ‘PCKh’ .

返回 PCKh for each joint

返回类型 dict

```
class mmpose.datasets.datasets.top_down.TopDownOCHumanDataset (ann_file, img_prefix,  
                                                                data_cfg, pipeline,  
                                                                dataset_info=None,  
                                                                test_mode=False)
```

OCHuman dataset for top-down pose estimation.

“Pose2Seg: Detection Free Human Instance Segmentation” , CVPR’ 2019. More details can be found in the [paper](#) .

“Occluded Human (OCHuman)” dataset contains 8110 heavily occluded human instances within 4731 images. OCHuman dataset is designed for validation and testing. To evaluate on OCHuman, the model should be trained on COCO training set, and then test the robustness of the model to occlusion using OCHuman.

OCHuman keypoint indexes (same as COCO):

```
0: 'nose',  
1: 'left_eye',  
2: 'right_eye',  
3: 'left_ear',  
4: 'right_ear',  
5: 'left_shoulder',  
6: 'right_shoulder',  
7: 'left_elbow',  
8: 'right_elbow',  
9: 'left_wrist',  
10: 'right_wrist',  
11: 'left_hip',  
12: 'right_hip',  
13: 'left_knee',  
14: 'right_knee',  
15: 'left_ankle',  
16: 'right_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.top_down.TopDownPoseTrack18Dataset (ann_file,
                                                                    img_prefix,
                                                                    data_cfg, pipeline,
                                                                    dataset_info=None,
                                                                    test_mode=False)
```

PoseTrack18 dataset for top-down pose estimation.

“Posetrack: A benchmark for human pose estimation and tracking” , CVPR’ 2018. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

PoseTrack2018 keypoint indexes:

```
0: 'nose',
1: 'head_bottom',
2: 'head_top',
3: 'left_ear',
4: 'right_ear',
5: 'left_shoulder',
6: 'right_shoulder',
7: 'left_elbow',
8: 'right_elbow',
9: 'left_wrist',
10: 'right_wrist',
11: 'left_hip',
12: 'right_hip',
13: 'left_knee',
14: 'right_knee',
15: 'left_ankle',
16: 'right_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder*=None, *metric*='mAP', *\*\*kwargs*)

Evaluate posetrack keypoint results. The pose prediction results will be saved in \${res\_folder}/

```
result_keypoints.json.
```

注解:

- num\_keypoints: K

### 参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (np.ndarray[N,K,3]): The first two dimensions are coordinates, score is the third dimension of the array.
  - boxes (np.ndarray[N,6]): [center[0], center[1], scale[0], scale[1],area, score]
  - image\_paths (list[str]): For example, [ ‘val/010016\_mpii\_test /000024.jpg’ ]
  - heatmap (np.ndarray[N, K, H, W]): model output heatmap.
  - bbox\_id (list(int))
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: None.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: ‘mAP’ .

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.datasets.top_down.TopDownPoseTrack18VideoDataset (ann_file,
                                                                    img_prefix,
                                                                    data_cfg,
                                                                    pipeline,
                                                                    dataset_info=None,
                                                                    test_mode=False,
                                                                    ph_fill_len=6)
```

PoseTrack18 dataset for top-down pose estimation.

“Posetrack: A benchmark for human pose estimation and tracking” , CVPR’ 2018. More details can be found in the [paper](#) .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

PoseTrack2018 keypoint indexes:

```
0: 'nose',
1: 'head_bottom',
```

(下页继续)



(续上页)

```

2: 'head_top',
3: 'left_ear',
4: 'right_ear',
5: 'left_shoulder',
6: 'right_shoulder',
7: 'left_elbow',
8: 'right_elbow',
9: 'left_wrist',
10: 'right_wrist',
11: 'left_hip',
12: 'right_hip',
13: 'left_knee',
14: 'right_knee',
15: 'left_ankle',
16: 'right_ankle'

```

**参数**

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where videos/images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.
- **ph\_fill\_len** (*int*) –The length of the placeholder to fill in the image filenames, default: 6 in PoseTrack18.

**evaluate** (*results*, *res\_folder=None*, *metric='mAP'*, *\*\*kwargs*)

Evaluate posetrack keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

**注解:**

- num\_keypoints: K

**参数**

- **results** (*list[dict]*) –Testing results containing the following items:
  - preds (*np.ndarray[N,K,3]*): The first two dimensions are coordinates, score is the third dimension of the array.

- `boxes` (`np.ndarray[N,6]`): [`center[0]`, `center[1]`, `scale[0]`, `scale[1]`, `area`, `score`]
- `image_paths` (`list[str]`): For example, [ `'val/010016_mpii_test/000024.jpg'` ]
- `heatmap` (`np.ndarray[N, K, H, W]`): model output heatmap.
- `bbox_id` (`list(int)`)
- **`res_folder`** (`str`, *optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: `None`.
- **`metric`** (`str` | `list[str]`) –Metric to be performed. Defaults: `'mAP'` .

返回 Evaluation results for evaluation metric.

返回类型 `dict`

```
class mmpose.datasets.datasets.bottom_up.BottomUpAicDataset (ann_file, img_prefix,
                                                             data_cfg, pipeline,
                                                             dataset_info=None,
                                                             test_mode=False)
```

Aic dataset for bottom-up pose estimation.

“AI Challenger : A Large-scale Dataset for Going Deeper in Image Understanding” , arXiv’ 2017. More details can be found in the [paper](#)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

AIC keypoint indexes:

```
0: "right_shoulder",
1: "right_elbow",
2: "right_wrist",
3: "left_shoulder",
4: "left_elbow",
5: "left_wrist",
6: "right_hip",
7: "right_knee",
8: "right_ankle",
9: "left_hip",
10: "left_knee",
11: "left_ankle",
12: "head_top",
13: "neck"
```

### 参数

- **`ann_file`** (`str`) –Path to the annotation file.
- **`img_prefix`** (`str`) –Path to a directory where images are held. Default: `None`.

- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.bottom_up.BottomUpCocoDataset (ann_file, img_prefix,  
                                                             data_cfg, pipeline,  
                                                             dataset_info=None,  
                                                             test_mode=False)
```

COCO dataset for bottom-up pose estimation.

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

COCO keypoint indexes:

```
0: 'nose',  
1: 'left_eye',  
2: 'right_eye',  
3: 'left_ear',  
4: 'right_ear',  
5: 'left_shoulder',  
6: 'right_shoulder',  
7: 'left_elbow',  
8: 'right_elbow',  
9: 'left_wrist',  
10: 'right_wrist',  
11: 'left_hip',  
12: 'right_hip',  
13: 'left_knee',  
14: 'right_knee',  
15: 'left_ankle',  
16: 'right_ankle'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

**evaluate** (*results*, *res\_folder=None*, *metric='mAP'*, *\*\*kwargs*)

Evaluate coco keypoint results. The pose prediction results will be saved in `${res_folder}/result_keypoints.json`.

注解:

- `num_people`: P
- `num_keypoints`: K

参数

- **results** (*list[dict]*) –Testing results containing the following items:
  - `preds` (*list[np.ndarray(P, K, 3+tag\_num)]*): Pose predictions for all people in images.
  - `scores` (*list[P]*): List of person scores.
  - `image_path` (*list[str]*): For example, [ `'coco/images/ val2017/000000397133.jpg'` ]
  - `heatmap` (*np.ndarray[N, K, H, W]*): model outputs.
- **res\_folder** (*str, optional*) –The folder to save the testing results. If not specified, a temp folder will be created. Default: `None`.
- **metric** (*str | list[str]*) –Metric to be performed. Defaults: `'mAP'` .

返回 Evaluation results for evaluation metric.

返回类型 dict

```
class mmpose.datasets.datasets.bottom_up.BottomUpCocoWholeBodyDataset (ann_file,
                                                                    img_prefix,
                                                                    data_cfg,
                                                                    pipeline,
                                                                    dataset_info=None,
                                                                    test_mode=False)
```

CocoWholeBodyDataset dataset for bottom-up pose estimation.

Whole-Body Human Pose Estimation in the Wild’ , ECCV’ 2020. More details can be found in the ‘paper’ .

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

In total, we have 133 keypoints for wholebody pose estimation.

COCO-WholeBody keypoint indexes:

```
0-16: 17 body keypoints,
17-22: 6 foot keypoints,
```

(下页继续)

(续上页)

```
23-90: 68 face keypoints,
91-132: 42 hand keypoints
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.bottom_up.BottomUpCrowdPoseDataset (ann_file,
                                                                    img_prefix,
                                                                    data_cfg, pipeline,
                                                                    dataset_info=None,
                                                                    test_mode=False)
```

CrowdPose dataset for bottom-up pose estimation.

“CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark” , CVPR’ 2019. More details can be found in the [paper](#).

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

CrowdPose keypoint indexes:

```
0: 'left_shoulder',
1: 'right_shoulder',
2: 'left_elbow',
3: 'right_elbow',
4: 'left_wrist',
5: 'right_wrist',
6: 'left_hip',
7: 'right_hip',
8: 'left_knee',
9: 'right_knee',
10: 'left_ankle',
11: 'right_ankle',
12: 'top_head',
13: 'neck'
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config
- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

```
class mmpose.datasets.datasets.bottom_up.BottomUpMhpDataset (ann_file, img_prefix,  
                                                         data_cfg, pipeline,  
                                                         dataset_info=None,  
                                                         test_mode=False)
```

MHPv2.0 dataset for top-down pose estimation.

“Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and A New Benchmark for Multi-Human Parsing” , ACM MM’ 2018. More details can be found in the [paper](#)

The dataset loads raw features and apply specified transforms to return a dict containing the image tensors and other information.

MHP keypoint indexes:

```
0: "right ankle",
1: "right knee",
2: "right hip",
3: "left hip",
4: "left knee",
5: "left ankle",
6: "pelvis",
7: "thorax",
8: "upper neck",
9: "head top",
10: "right wrist",
11: "right elbow",
12: "right shoulder",
13: "left shoulder",
14: "left elbow",
15: "left wrist",
```

### 参数

- **ann\_file** (*str*) –Path to the annotation file.
- **img\_prefix** (*str*) –Path to a directory where images are held. Default: None.
- **data\_cfg** (*dict*) –config

- **pipeline** (*list[dict | callable]*) –A sequence of data transforms.
- **dataset\_info** (*DatasetInfo*) –A class containing all dataset info.
- **test\_mode** (*bool*) –Store True when building test or validation dataset. Default: False.

## 49.2 pipelines

```
class mmpose.datasets.pipelines.loading.LoadImageFromFile (to_float32=False,
                                                         color_type='color',
                                                         channel_order='rgb',
                                                         file_client_args={ 'backend':
                                                         'disk'})
```

Loading image(s) from file.

Required key: “image\_file” .

Added key: “img” .

### 参数

- **to\_float32** (*bool*) –Whether to convert the loaded image to a float32 numpy array. If set to False, the loaded image is an uint8 array. Defaults to False.
- **color\_type** (*str*) –Flags specifying the color type of a loaded image, candidates are ‘color’ , ‘grayscale’ and ‘unchanged’ .
- **channel\_order** (*str*) –Order of channel, candidates are ‘bgr’ and ‘rgb’ .
- **file\_client\_args** (*dict*) –Arguments to instantiate a FileClient. See `mmpcv.fileio.FileClient` for details. Defaults to `dict (backend='disk')` .

```
class mmpose.datasets.pipelines.loading.LoadVideoFromFile (to_float32=False,
                                                         file_client_args={ 'backend':
                                                         'disk'})
```

Loading video(s) from file.

Required key: “video\_file” .

Added key: “video” .

### 参数

- **to\_float32** (*bool*) –Whether to convert the loaded video to a float32 numpy array. If set to False, the loaded video is an uint8 array. Defaults to False.
- **file\_client\_args** (*dict*) –Arguments to instantiate a FileClient. See `mmpcv.fileio.FileClient` for details. Defaults to `dict (backend='disk')` .

```
class mmpose.datasets.pipelines.shared_transform.Albumentation (transforms,  
                                                             keymap=None)
```

Albumentation augmentation (pixel-level transforms only). Adds custom pixel-level transformations from Albumentations library. Please visit <https://albumentations.readthedocs.io> to get more information.

Note: we only support pixel-level transforms. Please visit <https://github.com/albumentations-team/albumentations#pixel-level-transforms> to get more information about pixel-level transforms.

An example of transforms is as followed:

```
[
    dict(
        type='RandomBrightnessContrast',
        brightness_limit=[0.1, 0.3],
        contrast_limit=[0.1, 0.3],
        p=0.2),
    dict(type='ChannelShuffle', p=0.1),
    dict(
        type='OneOf',
        transforms=[
            dict(type='Blur', blur_limit=3, p=1.0),
            dict(type='MedianBlur', blur_limit=3, p=1.0)
        ],
        p=0.1),
]
```

### 参数

- **transforms** (*list[dict]*) – A list of Albumentation transformations
- **keymap** (*dict*) – Contains { ‘input key’ : ‘albumentation-style key’ }, e.g., { ‘img’ : ‘image’ }.

**albu\_builder** (*cfg*)

Import a module from albumentations.

It resembles some of `build_from_cfg()` logic.

**参数** **cfg** (*dict*) – Config dict. It should at least contain the key “type” .

**返回** The constructed object.

**返回类型** obj

**static mapper** (*d, keymap*)

Dictionary mapper.

Renames keys according to keymap provided.

**参数**



- **d** (*dict*) –old dict
- **keymap** (*dict*) –{ ‘old\_key’ : ‘new\_key’ }

返回 new dict.

返回类型 dict

```
class mmpose.datasets.pipelines.shared_transform.Collect (keys, meta_keys,  
                                                    meta_name='img_metas')
```

Collect data from the loader relevant to the specific task.

This keeps the items in *keys* as it is, and collect items in *meta\_keys* into a meta item called *meta\_name*. This is usually the last stage of the data loader pipeline. For example, when *keys*=‘imgs’, *meta\_keys*=(‘filename’, ‘label’, ‘original\_shape’), *meta\_name*=‘img\_metas’, the results will be a dict with keys ‘imgs’ and ‘img\_metas’, where ‘img\_metas’ is a DataContainer of another dict with keys ‘filename’, ‘label’, ‘original\_shape’.

#### 参数

- **keys** (*Sequence[str|tuple]*) –Required keys to be collected. If a tuple (key, key\_new) is given as an element, the item retrieved by key will be renamed as key\_new in collected data.
- **meta\_name** (*str*) –The name of the key that contains meta information. This key is always populated. Default: “img\_metas” .
- **meta\_keys** (*Sequence[str|tuple]*) –Keys that are collected under meta\_name. The contents of the *meta\_name* dictionary depends on *meta\_keys*.

```
class mmpose.datasets.pipelines.shared_transform.Compose (transforms)
```

Compose a data pipeline with a sequence of transforms.

**参数 transforms** (*list[dict | callable]*) –Either config dicts of transforms or transform objects.

```
class mmpose.datasets.pipelines.shared_transform.MultiItemProcess (pipeline)
```

Process each item and merge multi-item results to lists.

**参数 pipeline** (*dict*) –Dictionary to construct pipeline for a single item.

```
class mmpose.datasets.pipelines.shared_transform.MultitaskGatherTarget (pipeline_list,  
                                                    pipeline_indices=None,  
                                                    keys=('target',  
                                                    'tar-  
                                                    get_weight'))
```

Gather the targets for multitask heads.

#### 参数

- **pipeline\_list** (*list[list]*) –List of pipelines for all heads.
- **pipeline\_indices** (*list[int]*) –Pipeline index of each head.

**class** mmpose.datasets.pipelines.shared\_transform.**NormalizeTensor** (*mean, std*)

Normalize the Tensor image (CxHxW), with mean and std.

Required key: 'img' . Modifies key: 'img' .

#### 参数

- **mean** (*list[float]*) –Mean values of 3 channels.
- **std** (*list[float]*) –Std values of 3 channels.

**class** mmpose.datasets.pipelines.shared\_transform.**PhotometricDistortion** (*brightness\_delta=32,*

*contrast\_range=(0.5, 1.5),*  
*saturation\_range=(0.5, 1.5),*  
*hue\_delta=18*)

Apply photometric distortion to image sequentially, every transformation is applied with a probability of 0.5. The position of random contrast is in second or second to last.

1. random brightness
2. random contrast (mode 0)
3. convert color from BGR to HSV
4. random saturation
5. random hue
6. convert color from HSV to BGR
7. random contrast (mode 1)
8. randomly swap channels

#### 参数

- **brightness\_delta** (*int*) –delta of brightness.
- **contrast\_range** (*tuple*) –range of contrast.
- **saturation\_range** (*tuple*) –range of saturation.
- **hue\_delta** (*int*) –delta of hue.

**brightness** (*img*)

Brightness distortion.

**contrast** (*img*)

Contrast distortion.

**convert** (*img*, *alpha=1*, *beta=0*)

Multiple with alpha and add beta with clip.

**class** mmpose.datasets.pipelines.shared\_transform.**RenameKeys** (*key\_pairs*)

Rename the keys.

**参数** **key\_pairs** (*Sequence[tuple]*) –Required keys to be renamed. If a tuple (key\_src, key\_tgt) is given as an element, the item retrieved by key\_src will be renamed as key\_tgt.

**class** mmpose.datasets.pipelines.shared\_transform.**ToTensor** (*device='cpu'*)

Transform image to Tensor.

Required key: 'img' . Modifies key: 'img' .

**参数** **results** (*dict*) –contain all information about training.

**class** mmpose.datasets.pipelines.top\_down\_transform.**TopDownAffine** (*use\_udp=False*)

Affine transform the image to make input.

Required key: 'img' , 'joints\_3d' , 'joints\_3d\_visible' , 'ann\_info' , 'scale' , 'rotation' and 'center' .

Modified key: 'img' , 'joints\_3d' , and 'joints\_3d\_visible' .

**参数** **use\_udp** (*bool*) –To use unbiased data processing. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

**class** mmpose.datasets.pipelines.top\_down\_transform.**TopDownGenerateTarget** (*sigma=2*,

*ker-*

*nel=(11,*

*11),*

*valid\_radius\_factor=0.054*

*tar-*

*get\_type='GaussianHeatmap'*

*encod-*

*ing='MSRA',*

*unbi-*

*ased\_encoding=False*)

Generate the target heatmap.

Required key: 'joints\_3d' , 'joints\_3d\_visible' , 'ann\_info' .

Modified key: 'target' , and 'target\_weight' .

**参数**

- **sigma** –Sigma of heatmap gaussian for 'MSRA' approach.
- **kernel** –Kernel of heatmap gaussian for 'Megvii' approach.
- **encoding** (*str*) –Approach to generate target heatmaps. Currently supported approaches: 'MSRA' , 'Megvii' , 'UDP' . Default: 'MSRA'

- **unbiased\_encoding** (*bool*) –Option to use unbiased encoding methods. Paper ref: Zhang et al. Distribution-Aware Coordinate Representation for Human Pose Estimation (CVPR 2020).
- **keypoint\_pose\_distance** –Keypoint pose distance for UDP. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).
- **target\_type** (*str*) –supported targets: ‘GaussianHeatmap’, ‘CombinedTarget’. Default: ‘GaussianHeatmap’. CombinedTarget: The combination of classification target (response map) and regression target (offset map). Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

**class**

`mmpose.datasets.pipelines.top_down_transform.TopDownGenerateTargetRegression`  
Generate the target regression vector (coordinates).

Required key: ‘joints\_3d’, ‘joints\_3d\_visible’, ‘ann\_info’. Modified key: ‘target’, and ‘target\_weight’.

**class** `mmpose.datasets.pipelines.top_down_transform.TopDownGetBboxCenterScale` (*padding:*  
*float*  
*=*  
*1.25*)

Convert bbox from [x, y, w, h] to center and scale.

The center is the coordinates of the bbox center, and the scale is the bbox width and height normalized by a scale factor.

Required key: ‘bbox’, ‘ann\_info’

Modifies key: ‘center’, ‘scale’

**参数 padding** (*float*) –bbox padding scale that will be multilied to scale. Default: 1.25

**class** `mmpose.datasets.pipelines.top_down_transform.TopDownGetRandomScaleRotation` (*rot\_factor=40*  
*scale\_factor=(*  
*rot\_prob=0.6)*)

Data augmentation with random scaling & rotating.

Required key: ‘scale’.

Modifies key: ‘scale’ and ‘rotation’.

**参数**

- **rot\_factor** (*int*) –Rotating to  $[-2 * \text{rot\_factor}, 2 * \text{rot\_factor}]$ .
- **scale\_factor** (*float*) –Scaling to  $[1 - \text{scale\_factor}, 1 + \text{scale\_factor}]$ .
- **rot\_prob** (*float*) –Probability of random rotation.

**class** mmpose.datasets.pipelines.top\_down\_transform.**TopDownHalfBodyTransform** (*num\_joints\_half\_body*  
*prob\_half\_body=0.3*)

Data augmentation with half-body transform. Keep only the upper body or the lower body at random.

Required key: 'joints\_3d', 'joints\_3d\_visible', and 'ann\_info'.

Modifies key: 'scale' and 'center'.

#### 参数

- **num\_joints\_half\_body** (*int*) –Threshold of performing half-body transform. If the body has fewer number of joints ( $< \text{num\_joints\_half\_body}$ ), ignore this step.
- **prob\_half\_body** (*float*) –Probability of half-body transform.

**static half\_body\_transform** (*cfg, joints\_3d, joints\_3d\_visible*)

Get center&scale for half-body transform.

**class** mmpose.datasets.pipelines.top\_down\_transform.**TopDownRandomFlip** (*flip\_prob=0.5*)

Data augmentation with random image flip.

Required key: 'img', 'joints\_3d', 'joints\_3d\_visible', 'center' and 'ann\_info'.

Modifies key: 'img', 'joints\_3d', 'joints\_3d\_visible', 'center' and 'flipped'.

#### 参数

- **flip** (*bool*) –Option to perform random flip.
- **flip\_prob** (*float*) –Probability of flip.

**class** mmpose.datasets.pipelines.top\_down\_transform.**TopDownRandomShiftBboxCenter** (*shift\_factor:*  
*float*  
*=*  
*0.16,*  
*prob:*  
*float*  
*=*  
*0.3*)

Random shift the bbox center.

Required key: 'center', 'scale'

Modifies key: 'center'

#### 参数

- **shift\_factor** (*float*) –The factor to control the shift range, which is  $\text{scale} * \text{pixel\_std} * \text{scale\_factor}$ . Default: 0.16
- **prob** (*float*) –Probability of applying random shift. Default: 0.3

```
class mmpose.datasets.pipelines.bottom_up_transform.BottomUpGenerateHeatmapTarget (sigma,  
                                                                           bg_weight=1.0,  
                                                                           gen_center_heatmap=True,  
                                                                           use_udp=False)
```

Generate multi-scale heatmap target for bottom-up.

Required key: 'joints' , 'mask' and 'center' .

Modifies key: 'target' , 'heatmaps' and 'masks' .

#### 参数

- **sigma** (*int or tuple*) –Sigma of heatmap Gaussian. If sigma is a tuple, the first item should be the sigma of keypoints and the second item should be the sigma of center.
- **bg\_weight** (*float*) –Weight for background. Default: 1.0.
- **gen\_center\_heatmap** (*bool*) –Whether to generate heatmaps for instance centers. Default: False.
- **use\_udp** (*bool*) –To use unbiased data processing. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

```
class mmpose.datasets.pipelines.bottom_up_transform.BottomUpGenerateOffsetTarget (radius=4)  
Generate multi-scale offset target for bottom-up.
```

Required key: 'center' , 'joints' and 'area' .

Modifies key: 'offsets' , 'offset\_weights'.

参数 **radius** (*int*) –Radius of labeled area for each instance.

```
class mmpose.datasets.pipelines.bottom_up_transform.BottomUpGeneratePAFTarget (limb_width,  
                                                                           skeleton=True,  
                                                                           use_udp=False)
```

Generate multi-scale heatmaps and part affinity fields (PAF) target for bottom-up. Paper ref: Cao et al. Realtime Multi-Person 2D Human Pose Estimation using Part Affinity Fields (CVPR 2017).

参数 **limb\_width** (*int*) –Limb width of part affinity fields

```
class mmpose.datasets.pipelines.bottom_up_transform.BottomUpGenerateTarget (sigma,  
                                                                           max_num_people=1,  
                                                                           use_udp=False)
```

Generate multi-scale heatmap target for associate embedding.

#### 参数

- **sigma** (*int*) –Sigma of heatmap Gaussian
- **max\_num\_people** (*int*) –Maximum number of people in an image

- **use\_udp** (*bool*) –To use unbiased data processing. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

```
class mmpose.datasets.pipelines.bottom_up_transform.BottomUpGetImgSize (test_scale_factor,
                                                                    current_scale=1,
                                                                    base_length=64,
                                                                    use_udp=False)
```

Get multi-scale image sizes for bottom-up, including base\_size and test\_scale\_factor. Keep the ratio and the image is resized to `results[ 'ann_info' ][ 'image_size' ]×current_scale`.

#### 参数

- **test\_scale\_factor** (*List[float]*) –Multi scale
- **current\_scale** (*int*) –default 1
- **base\_length** (*int*) –The width and height should be multiples of base\_length. Default: 64.
- **use\_udp** (*bool*) –To use unbiased data processing. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

```
class mmpose.datasets.pipelines.bottom_up_transform.BottomUpRandomAffine (rot_factor,
                                                                    scale_factor,
                                                                    scale_type,
                                                                    trans_factor,
                                                                    use_udp=False)
```

Data augmentation with random scaling & rotating.

#### 参数

- **rot\_factor** (*int*) –Rotating to [-rotation\_factor, rotation\_factor]
- **scale\_factor** (*float*) –Scaling to [1-scale\_factor, 1+scale\_factor]
- **scale\_type** –wrt long or short length of the image.
- **trans\_factor** –Translation factor.
- **use\_udp** (*bool*) –To use unbiased data processing. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

```
class mmpose.datasets.pipelines.bottom_up_transform.BottomUpRandomFlip (flip_prob=0.5)
```

Data augmentation with random image flip for bottom-up.

参数 **flip\_prob** (*float*) –Probability of flip.

```
class mmpose.datasets.pipelines.bottom_up_transform.BottomUpResizeAlign (transforms,  
                                                                    base_length=64,  
                                                                    use_udp=False)
```

Resize multi-scale size and align transform for bottom-up.

**参数**

- **transforms** (*List*) –ToTensor & Normalize
- **base\_length** (*int*) –The width and height should be multiples of base\_length. Default: 64.
- **use\_udp** (*bool*) –To use unbiased data processing. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).

```
class mmpose.datasets.pipelines.bottom_up_transform.CIDGenerateTarget (max_num_people)  
Generate target for CID training.
```

**参数** **max\_num\_people** (*int*) –Maximum number of people in an image

```
class mmpose.datasets.pipelines.bottom_up_transform.GetKeypointCenterArea (minimal_area=32)  
Copmute center and area from keypointns for each instance.
```

Required key: ‘joints’ .

Modifies key: ‘center’ and ‘area’ .

**参数** **minimal\_area** (*float*) –Minimum of allowed area. Instance with smaller area will be ignored in training. Default: 32.

```
class mmpose.datasets.pipelines.bottom_up_transform.HeatmapGenerator (output_size,  
                                                                    num_joints,  
                                                                    sigma=- 1,  
                                                                    use_udp=False)
```

Generate heatmaps for bottom-up models.

**参数**

- **num\_joints** (*int*) –Number of keypoints
- **output\_size** (*np.ndarray*) –Size (w, h) of feature map
- **sigma** (*int*) –Sigma of the heatmaps.
- **use\_udp** (*bool*) –To use unbiased data processing. Paper ref: Huang et al. The Devil is in the Details: Delving into Unbiased Data Processing for Human Pose Estimation (CVPR 2020).



```
class mmpose.datasets.pipelines.bottom_up_transform.JointsEncoder (max_num_people,  
                                                                num_joints,  
                                                                output_size,  
                                                                tag_per_joint)
```

Encodes the visible joints into (coordinates, score); The coordinate of one joint and its score are of *int* type.

(idx \* output\_size\*\*2 + y \* output\_size + x, 1) or (0, 0).

#### 参数

- **max\_num\_people** (*int*) –Max number of people in an image
- **num\_joints** (*int*) –Number of keypoints
- **output\_size** (*np.ndarray*) –Size (w, h) of feature map
- **tag\_per\_joint** (*bool*) –Option to use one tag map per joint.

```
class mmpose.datasets.pipelines.bottom_up_transform.OffsetGenerator (output_size,  
                                                                num_joints,  
                                                                radius=4)
```

Generate offset maps for bottom-up models.

#### 参数

- **num\_joints** (*int*) –Number of keypoints
- **output\_size** (*np.ndarray*) –Size (w, h) of feature map
- **radius** (*int*) –Radius of area assigned with valid offset

```
class mmpose.datasets.pipelines.bottom_up_transform.PAFGenerator (output_size,  
                                                                limb_width,  
                                                                skeleton)
```

Generate part affinity fields.

#### 参数

- **output\_size** (*np.ndarray*) –Size (w, h) of feature map.
- **limb\_width** (*int*) –Limb width of part affinity fields.
- **skeleton** (*list[list]*) –connections of joints.

```
class mmpose.datasets.pipelines.mesh_transform.IUVToTensor
```

Transform IUV image to part index mask and uv coordinates image. The 3 channels of IUV image means: part index, u coordinates, v coordinates.

Required key: 'iuv', 'ann\_info'. Modifies key: 'part\_index', 'uv\_coordinates'.

参数 **results** (*dict*) –contain all information about training.

```
class mmpose.datasets.pipelines.mesh_transform.LoadIUVFromFile (to_float32=False)
```

Loading IUV image from file.

**class** mmpose.datasets.pipelines.mesh\_transform.**MeshAffine**

Affine transform the image to get input image. Affine transform the 2D keypoints, 3D keypoints and IUV image too.

Required keys: 'img', 'joints\_2d', 'joints\_2d\_visible', 'joints\_3d', 'joints\_3d\_visible', 'pose', 'iuv', 'ann\_info', 'scale', 'rotation' and 'center'. Modifies key: 'img', 'joints\_2d', 'joints\_2d\_visible', 'joints\_3d', 'pose', 'iuv'.

**class** mmpose.datasets.pipelines.mesh\_transform.**MeshGetRandomScaleRotation** (*rot\_factor=30*,  
*scale\_factor=0.25*,  
*rot\_prob=0.6*)

Data augmentation with random scaling & rotating.

Required key: 'scale'. Modifies key: 'scale' and 'rotation'.

**参数**

- **rot\_factor** (*int*) –Rotating to  $[-2*\text{rot\_factor}, 2*\text{rot\_factor}]$ .
- **scale\_factor** (*float*) –Scaling to  $[1-\text{scale\_factor}, 1+\text{scale\_factor}]$ .
- **rot\_prob** (*float*) –Probability of random rotation.

**class** mmpose.datasets.pipelines.mesh\_transform.**MeshRandomChannelNoise** (*noise\_factor=0.4*)

Data augmentation with random channel noise.

Required keys: 'img' Modifies key: 'img'

**参数 noise\_factor** (*float*) –Multiply each channel with a factor between “[1-scale\_factor, 1+scale\_factor]”

**class** mmpose.datasets.pipelines.mesh\_transform.**MeshRandomFlip** (*flip\_prob=0.5*)

Data augmentation with random image flip.

Required keys: 'img', 'joints\_2d', 'joints\_2d\_visible', 'joints\_3d', 'joints\_3d\_visible', 'center', 'pose', 'iuv' and 'ann\_info'. Modifies key: 'img', 'joints\_2d', 'joints\_2d\_visible', 'joints\_3d', 'joints\_3d\_visible', 'center', 'pose', 'iuv'.

**参数 flip\_prob** (*float*) –Probability of flip.

**class** mmpose.datasets.pipelines.pose3d\_transform.**AffineJoints** (*item='joints'*,  
*visible\_item=None*)

Apply affine transformation to joints coordinates.

**参数**

- **item** (*str*) –The name of the joints to apply affine.
- **visible\_item** (*str*) –The name of the visibility item.

**Required keys:** item, visible\_item(optional)

**Modified keys:** item, visible\_item(optional)

```
class mmpose.datasets.pipelines.pose3d_transform.CameraProjection (item, mode, out-  
put_name=None,  
cam-  
era_type='SimpleCamera',  
cam-  
era_param=None)
```

Apply camera projection to joint coordinates.

#### 参数

- **item** (*str*) –The name of the pose to apply camera projection.
- **mode** (*str*) –The type of camera projection, supported options are
  - world\_to\_camera
  - world\_to\_pixel
  - camera\_to\_world
  - camera\_to\_pixel
- **output\_name** (*str/None*) –The name of the projected pose. If None (default) is given, the projected pose will be stored in place.
- **camera\_type** (*str*) –The camera class name (should be registered in CAMERA).
- **camera\_param** (*dict/None*) –The camera parameter dict. See the camera class definition for more details. If None is given, the camera parameter will be obtained during processing of each data sample with the key “camera\_param” .

Required keys:

- item
- camera\_param (if camera parameters are not given in initialization)

**Modified keys:** output\_name

```
class mmpose.datasets.pipelines.pose3d_transform.CollectCameraIntrinsics (camera_param=None,  
need_distortion=True)
```

Store camera intrinsics in a 1-dim array, including f, c, k, p.

#### 参数

- **camera\_param** (*dict/None*) –The camera parameter dict. See the camera class definition for more details. If None is given, the camera parameter will be obtained during processing of each data sample with the key “camera\_param” .
- **need\_distortion** (*bool*) –Whether need distortion parameters k and p. Default: True.

**Required keys:** camera\_param (if camera parameters are not given in initialization)

**Modified keys:** intrinsics

```
class mmpose.datasets.pipelines.pose3d_transform.Generate3DHeatmapTarget (sigma=2,  
                                                                           joint_indices=None,  
                                                                           max_bound=1.0)
```

Generate the target 3d heatmap.

Required keys: 'joints\_3d', 'joints\_3d\_visible', 'ann\_info'. Modified keys: 'target', and 'target\_weight'

.

#### 参数

- **sigma** –Sigma of heatmap gaussian.
- **joint\_indices** (*list*) –Indices of joints used for heatmap generation. If None (default) is given, all joints will be used.
- **max\_bound** (*float*) –The maximal value of heatmap.

```
class mmpose.datasets.pipelines.pose3d_transform.GenerateInputHeatmaps (item='joints',  
                                                                           visi-  
                                                                           ble_item=None,  
                                                                           ob-  
                                                                           scured=0.0,  
                                                                           from_pred=True,  
                                                                           sigma=3,  
                                                                           scale=None,  
                                                                           base_size=96,  
                                                                           tar-  
                                                                           get_type='gaussian',  
                                                                           heatmap_cfg=None)
```

Generate 2D input heatmaps for multi-camera heatmaps when the 2D model is not available.

Required keys: 'joints' Modified keys: 'input\_heatmaps'

#### 参数

- **sigma** (*int*) –Sigma of heatmap gaussian (mm).
- **base\_size** (*int*) –the base size of human
- **target\_type** (*str*) –type of target heatmap, only support 'gaussian' now

```
class mmpose.datasets.pipelines.pose3d_transform.GenerateVoxel3DHeatmapTarget (sigma=200.0,  
                                                                           joint_indices=None)
```

Generate the target 3d heatmap.

Required keys: 'joints\_3d', 'joints\_3d\_visible', 'ann\_info\_3d'. Modified keys: 'target', and 'target\_weight'.

### 参数

- **sigma** –Sigma of heatmap gaussian (mm).
- **joint\_indices** (*list*) –Indices of joints used for heatmap generation. If None (default) is given, all joints will be used.

```
class mmpose.datasets.pipelines.pose3d_transform.GetRootCenteredPose (item,  
                                                                    root_index,  
                                                                    visi-  
                                                                    ble_item=None,  
                                                                    re-  
                                                                    move_root=False,  
                                                                    root_name=None)
```

Zero-center the pose around a given root joint. Optionally, the root joint can be removed from the original pose and stored as a separate item.

Note that the root-centered joints may no longer align with some annotation information (e.g. flip\_pairs, num\_joints, inference\_channel, etc.) due to the removal of the root joint.

### 参数

- **item** (*str*) –The name of the pose to apply root-centering.
- **root\_index** (*int*) –Root joint index in the pose.
- **visible\_item** (*str*) –The name of the visibility item.
- **remove\_root** (*bool*) –If true, remove the root joint from the pose
- **root\_name** (*str*) –Optional. If not none, it will be used as the key to store the root position separated from the original pose.

**Required keys:** item

**Modified keys:** item, visible\_item, root\_name

```
class mmpose.datasets.pipelines.pose3d_transform.ImageCoordinateNormalization (item,  
                                                                    norm_camera=False,  
                                                                    cam-  
                                                                    era_param=None)
```

Normalize the 2D joint coordinate with image width and height. Range [0, w] is mapped to [-1, 1], while preserving the aspect ratio.

### 参数

- **item** (*str/list[str]*) –The name of the pose to normalize.

- **norm\_camera** (*bool*) –Whether to normalize camera intrinsics. Default: False.
- **camera\_param** (*dict | None*) –The camera parameter dict. See the camera class definition for more details. If None is given, the camera parameter will be obtained during processing of each data sample with the key “camera\_param” .

**Required keys:** item

**Modified keys:** item (, camera\_param)

```
class mmpose.datasets.pipelines.pose3d_transform.NormalizeJointCoordinate (item,  
                                                                    mean=None,  
                                                                    std=None,  
                                                                    norm_param_file=None)
```

Normalize the joint coordinate with given mean and std.

**参数**

- **item** (*str*) –The name of the pose to normalize.
- **mean** (*array*) –Mean values of joint coordinates in shape [K, C].
- **std** (*array*) –Std values of joint coordinates in shape [K, C].
- **norm\_param\_file** (*str*) –Optionally load a dict containing *mean* and *std* from a file using *mmcv.load*.

**Required keys:** item

**Modified keys:** item

```
class mmpose.datasets.pipelines.pose3d_transform.PoseSequenceToTensor (item)
```

Convert pose sequence from numpy array to Tensor.

The original pose sequence should have a shape of [T,K,C] or [K,C], where T is the sequence length, K and C are keypoint number and dimension. The converted pose sequence will have a shape of [KxC, T].

**参数** **item** (*str*) –The name of the pose sequence

**Required keys:** item

**Modified keys:** item

```
class mmpose.datasets.pipelines.pose3d_transform.RelativeJointRandomFlip (item,
                                                                    flip_cfg,
                                                                    visi-
                                                                    ble_item=None,
                                                                    flip_prob=0.5,
                                                                    flip_camera=False,
                                                                    cam-
                                                                    era_param=None)
```

Data augmentation with random horizontal joint flip around a root joint.

#### 参数

- **item** (*str*/*list*[*str*]) –The name of the pose to flip.
- **flip\_cfg** (*dict*/*list*[*dict*]) –Configurations of the `flipr_regression` function. It should contain the following arguments:
  - **center\_mode**: The mode to set the center location on the x-axis to flip around.
  - **center\_x** or **center\_index**: Set the x-axis location or the root joint' s index to define the flip center.

Please refer to the docstring of the `flipr_regression` function for more details.

- **visible\_item** (*str*/*list*[*str*]) –The name of the visibility item which will be flipped accordingly along with the pose.
- **flip\_prob** (*float*) –Probability of flip.
- **flip\_camera** (*bool*) –Whether to flip horizontal distortion coefficients.
- **camera\_param** (*dict*/*None*) –The camera parameter dict. See the camera class definition for more details. If *None* is given, the camera parameter will be obtained during processing of each data sample with the key “camera\_param” .

**Required keys:** item

**Modified keys:** item (, camera\_param)

## 49.3 samplers

```
class mmpose.datasets.samplers.DistributedSampler (dataset, num_replicas=None, rank=None,
                                                    shuffle=True, seed=0)
```

DistributedSampler inheriting from `torch.utils.data.DistributedSampler`.

In pytorch of lower versions, there is no `shuffle` argument. This child class will port one to DistributedSampler.





**class** mmpose.utils.StopWatch(window=1)

A helper class to measure FPS and detailed time consuming of each phase in a video processing loop or similar scenarios.

**参数** **window** (*int*) – The sliding window size to calculate the running average of the time consuming.

#### 示例

```
>>> from mmpose.utils import StopWatch
>>> import time
>>> stop_watch = StopWatch(window=10)
>>> with stop_watch.timeit('total'):
>>>     time.sleep(0.1)
>>>     # 'timeit' support nested use
>>>     with stop_watch.timeit('phase1'):
>>>         time.sleep(0.1)
>>>     with stop_watch.timeit('phase2'):
>>>         time.sleep(0.2)
>>>     time.sleep(0.2)
>>> report = stop_watch.report()
```

**report** (*key=None*)

Report timing information.

**返回** The key is the timer name and the value is the corresponding average time consuming.

**返回类型** dict

**report\_strings()**

Report timing information in texture strings.

**返回** Each element is the information string of a timed event, in format of ‘{timer\_name}: {time\_in\_ms}’. Specially, if timer\_name is ‘\_FPS\_’, the result will be converted to fps.

**返回类型** list(str)

**timeit** (timer\_name='\_FPS\_')

Timing a code snippet with an assigned name.

**参数** **timer\_name** (str) –The unique name of the interested code snippet to handle multiple timers and generate reports. Note that ‘\_FPS\_’ is a special key that the measurement will be in *fps* instead of *millisecond*. Also see *report* and *report\_strings*. Default: ‘\_FPS\_’.

---

**注解:** This function should always be used in a *with* statement, as shown in the example.

---

`mmpose.utils.get_root_logger(log_file=None, log_level=20)`

Use *get\_logger* method in *mmcv* to get the root logger.

The logger will be initialized if it has not been initialized. By default a *StreamHandler* will be added. If *log\_file* is specified, a *FileHandler* will also be added. The name of the root logger is the top-level package name, e.g., “mmpose”.

**参数**

- **log\_file** (str | None) –The log filename. If specified, a *FileHandler* will be added to the root logger.
- **log\_level** (int) –The root logger level. Note that only the process of rank 0 is affected, while other processes will set the level to “Error” and be silent most of the time.

**返回** The root logger.

**返回类型** logging.Logger

`mmpose.utils.setup_multi_processes(cfg)`

Setup multi-processing environment variables.

## CHAPTER 51

---

### Indices and tables

---

- `genindex`
- `search`



## m

`mmpose.apis`, 631  
`mmpose.core.evaluation`, 669  
`mmpose.core.fp16`, 680  
`mmpose.core.post_processing`, 684  
`mmpose.core.utils`, 683  
`mmpose.datasets`, 807  
`mmpose.datasets.pipelines`, 891  
`mmpose.datasets.pipelines.bottom_up_transform`,  
897  
`mmpose.datasets.pipelines.loading`, 891  
`mmpose.datasets.pipelines.mesh_transform`,  
901  
`mmpose.datasets.pipelines.pose3d_transform`,  
902  
`mmpose.datasets.pipelines.shared_transform`,  
891  
`mmpose.datasets.pipelines.top_down_transform`,  
895  
`mmpose.models.backbones`, 691  
`mmpose.models.detectors`, 732  
`mmpose.models.heads`, 767  
`mmpose.models.losses`, 791  
`mmpose.models.misc`, 805  
`mmpose.models.necks`, 728  
`mmpose.utils`, 909



## 符号

`__init__()` (*mmpose.apis.webcam.WebcamExecutor* 方法), 643

`__init__()` (*mmpose.apis.webcam.utils.BufferManager* 方法), 659

`__init__()` (*mmpose.apis.webcam.utils.EventManager* 方法), 664

`__init__()` (*mmpose.apis.webcam.utils.FrameMessage* 方法), 661

`__init__()` (*mmpose.apis.webcam.utils.Message* 方法), 660

`__init__()` (*mmpose.apis.webcam.utils.VideoEndingMessage* 方法), 661

## A

*AdaptiveWingLoss* (*mmpose.models.losses* 中的类), 792

*adjust\_width\_group()* (*mmpose.models.backbones.RegNet* 方法), 707

*AEHigherResolutionHead* (*mmpose.models.heads* 中的类), 767

*AELoss* (*mmpose.models.losses* 中的类), 791

*AEMultiStageHead* (*mmpose.models.heads* 中的类), 768

*AESimpleHead* (*mmpose.models.heads* 中的类), 769

*affine\_transform()* (在 *mmpose.core.post\_processing* 模块中), 685

*AffineJoints* (*mmpose.datasets.pipelines.pose3d\_transform* 中的类), 902

*after\_train\_iter()* (*mmpose.core.fp16.Fp16OptimizerHook* 方法), 680

*aggregate\_scale()* (在 *mmpose.core.evaluation* 模块中), 669

*aggregate\_stage\_flip()* (在 *mmpose.core.evaluation* 模块中), 670

*albu\_builder()* (*mmpose.datasets.pipelines.shared\_transform.Albumentation* 方法), 892

*Albumentation* (*mmpose.datasets.pipelines.shared\_transform* 中的类), 891

*AlexNet* (*mmpose.models.backbones* 中的类), 691

*allreduce\_grads()* (在 *mmpose.core.utils* 模块中), 683

*AnimalATRWDataset* (*mmpose.datasets* 中的类), 807

*AnimalFlyDataset* (*mmpose.datasets* 中的类), 809

*AnimalHorse10Dataset* (*mmpose.datasets* 中的类), 811

*AnimalLocustDataset* (*mmpose.datasets* 中的类), 812

*AnimalMacaqueDataset* (*mmpose.datasets* 中的类), 814

*AnimalPoseDataset* (*mmpose.datasets* 中的类), 816

*AnimalZebraDataset* (*mmpose.datasets* 中的类), 817

*assign2gt()* (*mmpose.models.detectors.VoxelCenterDetector* 方法), 763

*AssociativeEmbedding* (*mmpose.models.detectors*

中的类), 732  
 auto\_fp16() (在 *mmpose.core.fp16* 模块中), 681

## B

BaseVisualizerNode (*mmpose.apis.webcam.nodes* 中的类), 644  
 BCELoss (*mmpose.models.losses* 中的类), 793  
 before\_run() (*mmpose.core.fp16.Fp16OptimizerHook* 方法), 681  
 BigeyeEffectNode (*mmpose.apis.webcam.nodes* 中的类), 655  
 Body3DH36MDataset (*mmpose.datasets* 中的类), 819  
 Body3DMviewDirectCampusDataset (*mmpose.datasets* 中的类), 820  
 Body3DMviewDirectPanopticDataset (*mmpose.datasets* 中的类), 822  
 Body3DMviewDirectShelfDataset (*mmpose.datasets* 中的类), 823  
 BoneLoss (*mmpose.models.losses* 中的类), 793  
 BottomUpAicDataset (*mmpose.datasets* 中的类), 825  
 BottomUpCocoDataset (*mmpose.datasets* 中的类), 826  
 BottomUpCocoWholeBodyDataset (*mmpose.datasets* 中的类), 827  
 BottomUpCrowdPoseDataset (*mmpose.datasets* 中的类), 828  
 BottomUpGenerateHeatmapTarget (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 897  
 BottomUpGenerateOffsetTarget (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 898  
 BottomUpGeneratePAFTarget (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 898  
 BottomUpGenerateTarget (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 898  
 BottomUpGetImgSize (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 899

BottomUpMhpDataset (*mmpose.datasets* 中的类), 829  
 BottomUpRandomAffine (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 899  
 BottomUpRandomFlip (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 899  
 BottomUpResizeAlign (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 899  
 brightness() (*mmpose.datasets.pipelines.shared\_transform.PhotometricDistortion* 方法), 894  
 BufferManager (*mmpose.apis.webcam.utils* 中的类), 659  
 build\_dataloader() (在 *mmpose.datasets* 模块中), 868  
 build\_dataset() (在 *mmpose.datasets* 模块中), 868  
 build\_sample\_indices() (*mmpose.datasets.Body3DH36MDataset* 方法), 820

## C

CameraProjection (*mmpose.datasets.pipelines.pose3d\_transform* 中的类), 903  
 cast\_tensor\_type() (在 *mmpose.core.fp16* 模块中), 682  
 CID (*mmpose.models.detectors* 中的类), 736  
 CIDGenerateTarget (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 900  
 CIDHead (*mmpose.models.heads* 中的类), 770  
 coco2campus3D() (*mmpose.datasets.Body3DMviewDirectCampusDataset* 静态方法), 821  
 coco2shelf3D() (*mmpose.datasets.Body3DMviewDirectShelfDataset* 静态方法), 824  
 Collect (*mmpose.datasets.pipelines.shared\_transform* 中的类), 893



- `collect_multi_frames()` (在 `mmpose.apis` 模块中), 631
- `CollectCameraIntrinsics` (`mmpose.datasets.pipelines.pose3d_transform` 中的类), 903
- `Compose` (`mmpose.datasets` 中的类), 829
- `Compose` (`mmpose.datasets.pipelines.shared_transform` 中的类), 893
- `compute_similarity_transform()` (在 `mmpose.core.evaluation` 模块中), 671
- `contrast()` (`mmpose.datasets.pipelines.shared_transform.PhotometricDisortio` 方法), 894
- `convert()` (`mmpose.datasets.pipelines.shared_transform.PhotometricDisortio` 方法), 894
- `copy_and_paste()` (在 `mmpose.apis.webcam.utils` 模块中), 665
- `copy_grads_to_fp32()` (`mmpose.core.fp16.Fp16OptimizerHook` 静态方法), 681
- `copy_params_to_fp16()` (`mmpose.core.fp16.Fp16OptimizerHook` 静态方法), 681
- `CPM` (`mmpose.models.backbones` 中的类), 691
- `criterion()` (`mmpose.models.losses.AdaptiveWingLoss` 方法), 792
- `criterion()` (`mmpose.models.losses.SoftWingLoss` 方法), 803
- `criterion()` (`mmpose.models.losses.WingLoss` 方法), 804
- `CuboidCenterHead` (`mmpose.models.heads` 中的类), 771
- `CuboidPoseHead` (`mmpose.models.heads` 中的类), 772
- ## D
- `decode()` (`mmpose.models.heads.DeepposeRegressionHead` 方法), 775
- `decode()` (`mmpose.models.heads.Interhand3DHead` 方法), 777
- `decode()` (`mmpose.models.heads.TemporalRegressionHead` 方法), 780
- `decode()` (`mmpose.models.heads.TopdownHeatmapBaseHead` 方法), 782
- `DeconvHead` (`mmpose.models.heads` 中的类), 773
- `DeepFashionDataset` (`mmpose.datasets` 中的类), 830
- `DeepposeRegressionHead` (`mmpose.models.heads` 中的类), 775
- `DEKRHead` (`mmpose.models.heads` 中的类), 772
- `DetectAndRegress` (`mmpose.models.detectors` 中的类), 740
- `DetectorNode` (`mmpose.apis.webcam.nodes` 中的类), 645
- `DynamicCythonKeypointRegressor` (`mmpose.models.detectors` 中的类), 743
- `DynamicCythonKeypointRegressor` (`mmpose.core.evaluation` 中的类), 669
- `DistributedSampler` (`mmpose.datasets` 中的类), 831
- ## E
- `EvalHook` (`mmpose.core.evaluation` 中的类), 669
- `evaluate()` (`mmpose.datasets.AnimalATRWDataset` 方法), 808
- `evaluate()` (`mmpose.datasets.AnimalFlyDataset` 方法), 810
- `evaluate()` (`mmpose.datasets.AnimalHorse10Dataset` 方法), 812
- `evaluate()` (`mmpose.datasets.AnimalLocustDataset` 方法), 814
- `evaluate()` (`mmpose.datasets.AnimalMacaqueDataset` 方法), 815
- `evaluate()` (`mmpose.datasets.AnimalPoseDataset` 方法), 817
- `evaluate()` (`mmpose.datasets.AnimalZebraDataset` 方法), 818
- `evaluate()` (`mmpose.datasets.Body3DH36MDataset` 方法), 820
- `evaluate()` (`mmpose.datasets.Body3DMviewDirectCampusDataset` 方法), 821
- `evaluate()` (`mmpose.datasets.Body3DMviewDirectPanopticDataset` 方法), 823
- `evaluate()` (`mmpose.datasets.Body3DMviewDirectShelfDataset` 方法), 824
- `evaluate()` (`mmpose.datasets.BottomUpCocoDataset` 方法), 827

- `evaluate()` (`mmpose.datasets.DeepFashionDataset` 方法), 831
- `evaluate()` (`mmpose.datasets.Face300WDataset` 方法), 832
- `evaluate()` (`mmpose.datasets.FaceAFLWDataset` 方法), 833
- `evaluate()` (`mmpose.datasets.FaceCocoWholeBodyDataset` 方法), 835
- `evaluate()` (`mmpose.datasets.FaceCOFWDataset` 方法), 834
- `evaluate()` (`mmpose.datasets.FaceWFLWDataset` 方法), 837
- `evaluate()` (`mmpose.datasets.FreiHandDataset` 方法), 838
- `evaluate()` (`mmpose.datasets.HandCocoWholeBodyDataset` 方法), 840
- `evaluate()` (`mmpose.datasets.InterHand2DDataset` 方法), 842
- `evaluate()` (`mmpose.datasets.InterHand3DDataset` 方法), 844
- `evaluate()` (`mmpose.datasets.MeshH36MDataset` 方法), 846
- `evaluate()` (`mmpose.datasets.NVGestureDataset` 方法), 847
- `evaluate()` (`mmpose.datasets.OneHand10KDataset` 方法), 849
- `evaluate()` (`mmpose.datasets.PanopticDataset` 方法), 851
- `evaluate()` (`mmpose.datasets.TopDownCocoDataset` 方法), 853
- `evaluate()` (`mmpose.datasets.TopDownFreiHandDataset` 方法), 855
- `evaluate()` (`mmpose.datasets.TopDownH36MDataset` 方法), 856
- `evaluate()` (`mmpose.datasets.TopDownJhmdbDataset` 方法), 858
- `evaluate()` (`mmpose.datasets.TopDownMpiiDataset` 方法), 860
- `evaluate()` (`mmpose.datasets.TopDownMpiiTrbDataset` 方法), 862
- `evaluate()` (`mmpose.datasets.TopDownOneHand10KDataset` 方法), 864
- `evaluate()` (`mmpose.datasets.TopDownPanopticDataset` 方法), 864
- `evaluate()` (`mmpose.datasets.TopDownPoseTrack18Dataset` 方法), 865
- `evaluate()` (`mmpose.datasets.TopDownPoseTrack18VideoDataset` 方法), 867
- `EventManager` (`mmpose.apis.webcam.utils` 中的类), 664
- `expand_and_clamp()` (在 `mmpose.apis.webcam.utils` 模块中), 666
- `extract_pose_sequence()` (在 `mmpose.apis` 模块中), 631
- ## F
- `Face300WDataset` (`mmpose.datasets` 中的类), 831
- `FaceAFLWDataset` (`mmpose.datasets` 中的类), 833
- `FaceCocoWholeBodyDataset` (`mmpose.datasets` 中的类), 835
- `FaceCOFWDataset` (`mmpose.datasets` 中的类), 834
- `FaceWFLWDataset` (`mmpose.datasets` 中的类), 836
- `flip_back()` (在 `mmpose.core.post_processing` 模块中), 686
- `flip_feature_maps()` (在 `mmpose.core.evaluation` 模块中), 671
- `fliplr_joints()` (在 `mmpose.core.post_processing` 模块中), 686
- `fliplr_regression()` (在 `mmpose.core.post_processing` 模块中), 687
- `FocalHeatmapLoss` (`mmpose.models.losses` 中的类), 794
- `force_fp32()` (在 `mmpose.core.fp16` 模块中), 682
- `forward()` (`mmpose.models.backbones.AlexNet` 方法), 691
- `forward()` (`mmpose.models.backbones.CPM` 方法), 692
- `forward()` (`mmpose.models.backbones.HourglassAENet` 方法), 697
- `forward()` (`mmpose.models.backbones.HourglassNet` 方法), 698
- `forward()` (`mmpose.models.backbones.HRNet` 方法), 695
- `forward()` (`mmpose.models.backbones.I3D` 方法), 698
- `forward()` (`mmpose.models.backbones.LiteHRNet` 方法), 699

- 法), 699
- `forward()` (`mmpose.models.backbones.MobileNetV2` 方法), 701
- `forward()` (`mmpose.models.backbones.MobileNetV3` 方法), 702
- `forward()` (`mmpose.models.backbones.MSPN` 方法), 700
- `forward()` (`mmpose.models.backbones.PyramidVisionTransformer` 方法), 704
- `forward()` (`mmpose.models.backbones.RegNet` 方法), 708
- `forward()` (`mmpose.models.backbones.ResNet` 方法), 712
- `forward()` (`mmpose.models.backbones.RSN` 方法), 706
- `forward()` (`mmpose.models.backbones.ShuffleNetV1` 方法), 717
- `forward()` (`mmpose.models.backbones.ShuffleNetV2` 方法), 718
- `forward()` (`mmpose.models.backbones.SwinTransformer` 方法), 720
- `forward()` (`mmpose.models.backbones.TCFormer` 方法), 722
- `forward()` (`mmpose.models.backbones.TCN` 方法), 723
- `forward()` (`mmpose.models.backbones.V2VNet` 方法), 723
- `forward()` (`mmpose.models.backbones.VGG` 方法), 724
- `forward()` (`mmpose.models.backbones.ViPNAS_MobileNetV3` 方法), 726
- `forward()` (`mmpose.models.backbones.ViPNAS_ResNet` 方法), 727
- `forward()` (`mmpose.models.detectors.AssociativeEmbedding` 方法), 733
- `forward()` (`mmpose.models.detectors.CID` 方法), 736
- `forward()` (`mmpose.models.detectors.DetectAndRegress` 方法), 740
- `forward()` (`mmpose.models.detectors.DisentangledKeypointRegressor` 方法), 743
- `forward()` (`mmpose.models.detectors.GestureRecognizer` 方法), 747
- `forward()` (`mmpose.models.detectors.Interhand3D` 方法), 748
- `forward()` (`mmpose.models.detectors.MultiTask` 方法), 750
- `forward()` (`mmpose.models.detectors.ParametricMesh` 方法), 752
- `forward()` (`mmpose.models.detectors.PoseLifter` 方法), 755
- `forward()` (`mmpose.models.detectors.PoseWarper` 方法), 758
- `forward()` (`mmpose.models.detectors.TopDown` 方法), 760
- `forward()` (`mmpose.models.detectors.VoxelCenterDetector` 方法), 763
- `forward()` (`mmpose.models.detectors.VoxelSinglePose` 方法), 765
- `forward()` (`mmpose.models.heads.AEHigherResolutionHead` 方法), 767
- `forward()` (`mmpose.models.heads.AEMultiStageHead` 方法), 769
- `forward()` (`mmpose.models.heads.CIDHead` 方法), 771
- `forward()` (`mmpose.models.heads.CuboidCenterHead` 方法), 771
- `forward()` (`mmpose.models.heads.CuboidPoseHead` 方法), 772
- `forward()` (`mmpose.models.heads.DeconvHead` 方法), 774
- `forward()` (`mmpose.models.heads.DeepPoseRegressionHead` 方法), 775
- `forward()` (`mmpose.models.heads.DEKRHead` 方法), 773
- `forward()` (`mmpose.models.heads.HMRMeshHead` 方法), 777
- `forward()` (`mmpose.models.heads.Interhand3DHead` 方法), 778
- `forward()` (`mmpose.models.heads.MultiModalSSAHead` 方法), 779
- `forward()` (`mmpose.models.heads.TemporalRegressionHead` 方法), 781
- `forward()` (`mmpose.models.heads.TopdownHeatmapBaseHead` 方法), 783
- `forward()` (`mmpose.models.heads.TopdownHeatmapMSMUHead` 方法), 783
- `forward()` (`mmpose.models.heads.TopdownHeatmapMultiStageHead` 方法), 785

<code>forward()</code> ( <code>mmpose.models.heads.TopdownHeatmapSimpleHead</code> 方法), 788	<code>pose.models.detectors.AssociativeEmbedding</code> 方法), 734
<code>forward()</code> ( <code>mmpose.models.heads.ViPNASHeatmapSimpleHead</code> 方法), 790	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.CID</code> 方法), 737
<code>forward()</code> ( <code>mmpose.models.losses.AdaptiveWingLoss</code> 方法), 792	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.DetectAndRegress</code> 方法), 741
<code>forward()</code> ( <code>mmpose.models.losses.AELoss</code> 方法), 791	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.DisentangledKeypointRegressor</code> 方法), 744
<code>forward()</code> ( <code>mmpose.models.losses.BCELoss</code> 方法), 793	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.MultiTask</code> 方法), 751
<code>forward()</code> ( <code>mmpose.models.losses.BoneLoss</code> 方法), 793	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.ParametricMesh</code> 方法), 753
<code>forward()</code> ( <code>mmpose.models.losses.FocalHeatmapLoss</code> 方法), 794	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.PoseLifter</code> 方法), 756
<code>forward()</code> ( <code>mmpose.models.losses.GANLoss</code> 方法), 794	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.PoseWarper</code> 方法), 759
<code>forward()</code> ( <code>mmpose.models.losses.HeatmapLoss</code> 方法), 795	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.TopDown</code> 方法), 761
<code>forward()</code> ( <code>mmpose.models.losses.JointsMSELoss</code> 方法), 796	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.VoxelCenterDetector</code> 方法), 763
<code>forward()</code> ( <code>mmpose.models.losses.JointsOHKMMSELoss</code> 方法), 796	<code>forward_dummy()</code> ( <code>mmpose.models.detectors.VoxelSinglePose</code> 方法), 765
<code>forward()</code> ( <code>mmpose.models.losses.L1Loss</code> 方法), 796	<code>forward_test()</code> ( <code>mmpose.models.detectors.AssociativeEmbedding</code> 方法), 734
<code>forward()</code> ( <code>mmpose.models.losses.MeshLoss</code> 方法), 798	<code>forward_test()</code> ( <code>mmpose.models.detectors.CID</code> 方法), 738
<code>forward()</code> ( <code>mmpose.models.losses.MPJPELoss</code> 方法), 796	<code>forward_test()</code> ( <code>mmpose.models.detectors.DetectAndRegress</code> 方法), 741
<code>forward()</code> ( <code>mmpose.models.losses.MSELoss</code> 方法), 797	<code>forward_test()</code> ( <code>mmpose.models.detectors.DisentangledKeypointRegressor</code> 方法), 744
<code>forward()</code> ( <code>mmpose.models.losses.MultiLossFactory</code> 方法), 799	<code>forward_test()</code> ( <code>mmpose.models.detectors.GestureRecognizer</code> 方法), 747
<code>forward()</code> ( <code>mmpose.models.losses.RLELoss</code> 方法), 800	
<code>forward()</code> ( <code>mmpose.models.losses.SemiSupervisionLoss</code> 方法), 801	
<code>forward()</code> ( <code>mmpose.models.losses.SmoothL1Loss</code> 方法), 802	
<code>forward()</code> ( <code>mmpose.models.losses.SoftWeightSmoothL1Loss</code> 方法), 802	
<code>forward()</code> ( <code>mmpose.models.losses.SoftWingLoss</code> 方法), 804	
<code>forward()</code> ( <code>mmpose.models.losses.WingLoss</code> 方法), 804	
<code>forward()</code> ( <code>mmpose.models.necks.FPN</code> 方法), 729	
<code>forward()</code> ( <code>mmpose.models.necks.GlobalAveragePooling</code> 方法), 729	
<code>forward()</code> ( <code>mmpose.models.necks.MTA</code> 方法), 731	
<code>forward()</code> ( <code>mmpose.models.necks.PoseWarperNeck</code> 方法), 732	
<code>forward_dummy()</code> ( <code>mmpose.models.detectors.AssociativeEmbedding</code> 方法), 734	

<code>forward_test()</code> ( <code>mm-pose.models.detectors.Interhand3D</code> 方法), 749	<code>forward_train()</code> ( <code>mm-pose.models.detectors.PoseWarper</code> 方法), 759
<code>forward_test()</code> ( <code>mmpose.models.detectors.MultiTask</code> 方法), 752	<code>forward_train()</code> ( <code>mm-pose.models.detectors.TopDown</code> 方法), 761
<code>forward_test()</code> ( <code>mm-pose.models.detectors.ParametricMesh</code> 方法), 753	<code>forward_train()</code> ( <code>mm-pose.models.detectors.VoxelCenterDetector</code> 方法), 764
<code>forward_test()</code> ( <code>mmpose.models.detectors.PoseLifter</code> 方法), 756	<code>forward_train()</code> ( <code>mm-pose.models.detectors.VoxelSinglePose</code> 方法), 766
<code>forward_test()</code> ( <code>mm-pose.models.detectors.PoseWarper</code> 方法), 759	<code>Fp16OptimizerHook</code> ( <code>mmpose.core.fp16</code> 中的类), 680
<code>forward_test()</code> ( <code>mmpose.models.detectors.TopDown</code> 方法), 761	<code>FPN</code> ( <code>mmpose.models.necks</code> 中的类), 728
<code>forward_test()</code> ( <code>mm-pose.models.detectors.VoxelCenterDetector</code> 方法), 763	<code>FrameMessage</code> ( <code>mmpose.apis.webcam.utils</code> 中的类), 661
<code>forward_test()</code> ( <code>mm-pose.models.detectors.VoxelSinglePose</code> 方法), 765	<b>G</b>
<code>forward_train()</code> ( <code>mm-pose.models.detectors.AssociativeEmbedding</code> 方法), 734	<code>GANLoss</code> ( <code>mmpose.models.losses</code> 中的类), 794
<code>forward_train()</code> ( <code>mmpose.models.detectors.CID</code> 方法), 738	<code>Generate3DHeatmapTarget</code> ( <code>mm-pose.datasets.pipelines.pose3d_transform</code> 中的类), 904
<code>forward_train()</code> ( <code>mm-pose.models.detectors.DetectAndRegress</code> 方法), 741	<code>generate_regnet()</code> ( <code>mm-pose.models.backbones.RegNet</code> 静态方法), 708
<code>forward_train()</code> ( <code>mm-pose.models.detectors.DisentangledKeypointRegressor</code> 方法), 745	<code>GenerateInputHeatmaps</code> ( <code>mm-pose.datasets.pipelines.pose3d_transform</code> 中的类), 904
<code>forward_train()</code> ( <code>mm-pose.models.detectors.GestureRecognizer</code> 方法), 747	<code>GenerateVoxel3DHeatmapTarget</code> ( <code>mm-pose.datasets.pipelines.pose3d_transform</code> 中的类), 904
<code>forward_train()</code> ( <code>mm-pose.models.detectors.MultiTask</code> 方法), 752	<code>GestureRecognizer</code> ( <code>mmpose.models.detectors</code> 中的类), 746
<code>forward_train()</code> ( <code>mm-pose.models.detectors.ParametricMesh</code> 方法), 753	<code>get_3d_joints_from_mesh()</code> ( <code>mm-pose.models.detectors.ParametricMesh</code> 方法), 753
<code>forward_train()</code> ( <code>mm-pose.models.detectors.PoseLifter</code> 方法), 756	<code>get_accuracy()</code> ( <code>mm-pose.models.heads.DeepPoseRegressionHead</code> 方法), 775
	<code>get_accuracy()</code> ( <code>mm-pose.models.heads.Interhand3DHead</code> 方法),

- 778
- `get_accuracy()` (`mm-pose.models.heads.MultiModalSSAHead` 方法), 779
- `get_accuracy()` (`mm-pose.models.heads.TemporalRegressionHead` 方法), 781
- `get_accuracy()` (`mm-pose.models.heads.TopdownHeatmapBaseHead` 方法), 783
- `get_accuracy()` (`mm-pose.models.heads.TopdownHeatmapMSMUHead` 方法), 784
- `get_accuracy()` (`mm-pose.models.heads.TopdownHeatmapMultiStageHead` 方法), 785
- `get_accuracy()` (`mm-pose.models.heads.TopdownHeatmapSimpleHead` 方法), 788
- `get_accuracy()` (`mm-pose.models.heads.ViPNASHeatmapSimpleHead` 方法), 790
- `get_affine_transform()` (在 `mm-pose.core.post_processing` 模块中), 687
- `get_cached_file_path()` (在 `mm-pose.apis.webcam.utils` 模块中), 667
- `get_camera_param()` (`mm-pose.datasets.Body3DH36MDataset` 方法), 820
- `get_config_path()` (在 `mm-pose.apis.webcam.utils` 模块中), 668
- `get_eye_keypoint_ids()` (在 `mm-pose.apis.webcam.utils` 模块中), 662
- `get_face_keypoint_ids()` (在 `mm-pose.apis.webcam.utils` 模块中), 663
- `get_group_preds()` (在 `mm-pose.core.evaluation` 模块中), 671
- `get_hand_keypoint_ids()` (在 `mm-pose.apis.webcam.utils` 模块中), 663
- `get_loss()` (`mm-pose.models.heads.AEHigherResolutionHead` 方法), 768
- `get_loss()` (`mm-pose.models.heads.AEMultiStageHead` 方法), 769
- `get_loss()` (`mm-pose.models.heads.AESimpleHead` 方法), 770
- `get_loss()` (`mm-pose.models.heads.DeconvHead` 方法), 774
- `get_loss()` (`mm-pose.models.heads.DeepPoseRegressionHead` 方法), 776
- `get_loss()` (`mm-pose.models.heads.DEKRHead` 方法), 773
- `get_loss()` (`mm-pose.models.heads.Interhand3DHead` 方法), 778
- `get_loss()` (`mm-pose.models.heads.MultiModalSSAHead` 方法), 779
- `get_loss()` (`mm-pose.models.heads.TemporalRegressionHead` 方法), 781
- `get_loss()` (`mm-pose.models.heads.TopdownHeatmapBaseHead` 方法), 783
- `get_loss()` (`mm-pose.models.heads.TopdownHeatmapMSMUHead` 方法), 784
- `get_loss()` (`mm-pose.models.heads.TopdownHeatmapMultiStageHead` 方法), 786
- `get_loss()` (`mm-pose.models.heads.TopdownHeatmapSimpleHead` 方法), 788
- `get_loss()` (`mm-pose.models.heads.ViPNASHeatmapSimpleHead` 方法), 790
- `get_mouth_keypoint_ids()` (在 `mm-pose.apis.webcam.utils` 模块中), 663
- `get_new_center()` (`mm-pose.datasets.Body3DMviewDirectCampusDataset` 静态方法), 821
- `get_new_center()` (`mm-pose.datasets.Body3DMviewDirectShelfDataset` 静态方法), 825
- `get_root_logger()` (在 `mm-pose.utils` 模块中), 910
- `get_stages_from_blocks()` (`mm-pose.models.backbones.RegNet` 方法), 708
- `get_target_label()` (`mm-pose.models.losses.GANLoss` 方法), 795
- `get_track_id()` (在 `mm-pose.apis` 模块中), 632
- `get_warp_matrix()` (在 `mm-pose.core.post_processing` 模块中), 688
- `get_wrist_keypoint_ids()` (在 `mm-`



- pose.apis.webcam.utils* 模块中), 663
- GetKeypointCenterArea (mm-  
*pose.datasets.pipelines.bottom\_up\_transform*  
中的类), 900
- GetRootCenteredPose (mm-  
*pose.datasets.pipelines.pose3d\_transform* 中  
的类), 905
- GlobalAveragePooling (*mmpose.models.necks* 中  
的类), 729
- ## H
- half\_body\_transform() (mm-  
*pose.datasets.pipelines.top\_down\_transform.TopDownHalfBodyTransform*  
静态方法), 897
- HandCocoWholeBodyDataset (*mmpose.datasets* 中  
的类), 839
- HeatmapGenerator (mm-  
*pose.datasets.pipelines.bottom\_up\_transform*  
中的类), 900
- HeatmapLoss (*mmpose.models.losses* 中的类), 795
- HMRMeshHead (*mmpose.models.heads* 中的类), 776
- hook() (*mmpose.core.utils.WeightNormClipHook* 方法),  
683
- hook\_type (*mmpose.core.utils.WeightNormClipHook*  
property), 683
- HourglassAENet (*mmpose.models.backbones* 中的  
类), 696
- HourglassNet (*mmpose.models.backbones* 中的类),  
697
- HRFormer (*mmpose.models.backbones* 中的类), 692
- HRNet (*mmpose.models.backbones* 中的类), 694
- ## I
- I3D (*mmpose.models.backbones* 中的类), 698
- ImageCoordinateNormalization (mm-  
*pose.datasets.pipelines.pose3d\_transform* 中  
的类), 905
- inference\_bottom\_up\_pose\_model() (在 mm-  
*pose.apis* 模块中), 632
- inference\_interhand\_3d\_model() (在 mm-  
*pose.apis* 模块中), 633
- inference\_mesh\_model() (在 *mmpose.apis* 模块  
中), 634
- inference\_model() (mm-  
*pose.models.heads.DeepposeRegressionHead*  
方法), 776
- inference\_model() (mm-  
*pose.models.heads.Interhand3DHead* 方法),  
778
- inference\_model() (mm-  
*pose.models.heads.TemporalRegressionHead*  
方法), 782
- inference\_model() (mm-  
*pose.models.heads.TopdownHeatmapBaseHead*  
方法), 783
- inference\_model() (mm-  
*pose.models.heads.TopdownHeatmapMSMUHead*  
方法), 784
- inference\_model() (mm-  
*pose.models.heads.TopdownHeatmapMultiStageHead*  
方法), 786
- inference\_model() (mm-  
*pose.models.heads.TopdownHeatmapSimpleHead*  
方法), 788
- inference\_model() (mm-  
*pose.models.heads.ViPNASHeatmapSimpleHead*  
方法), 791
- inference\_pose\_lifter\_model() (在 mm-  
*pose.apis* 模块中), 635
- inference\_top\_down\_pose\_model() (在 mm-  
*pose.apis* 模块中), 636
- init\_pose\_model() (在 *mmpose.apis* 模块中), 637
- init\_random\_seed() (在 *mmpose.apis* 模块中), 637
- init\_weights() (*mmpose.models.backbones.CPM* 方  
法), 692
- init\_weights() (mm-  
*pose.models.backbones.HourglassAENet* 方  
法), 697
- init\_weights() (mm-  
*pose.models.backbones.HourglassNet* 方法),  
698
- init\_weights() (*mmpose.models.backbones.HRNet*  
方法), 696

<code>init_weights()</code> <i>pose.models.backbones.LiteHRNet</i> 699	( <i>mm-</i> 方法),	<code>init_weights()</code> <i>pose.models.detectors.DisentangledKeypointRegressor</i> 方法), 746	( <i>mm-</i>
<code>init_weights()</code> <i>pose.models.backbones.MobileNetV2</i> 701	( <i>mm-</i> 方法),	<code>init_weights()</code> <i>pose.models.detectors.GestureRecognizer</i> 方法), 747	( <i>mm-</i>
<code>init_weights()</code> <i>pose.models.backbones.MobileNetV3</i> 702	( <i>mm-</i> 方法),	<code>init_weights()</code> ( <i>mmpose.models.detectors.MultiTask</i> 方法), 752	
<code>init_weights()</code> ( <i>mmpose.models.backbones.MSPN</i> 方法), 700		<code>init_weights()</code> ( <i>pose.models.detectors.ParametricMesh</i> 方法), 753	( <i>mm-</i>
<code>init_weights()</code> <i>pose.models.backbones.PyramidVisionTransformer</i> 方法), 705	( <i>mm-</i> 方法),	<code>init_weights()</code> ( <i>mmpose.models.detectors.PoseLifter</i> 方法), 756	
<code>init_weights()</code> ( <i>mmpose.models.backbones.ResNet</i> 方法), 712		<code>init_weights()</code> ( <i>mmpose.models.detectors.TopDown</i> 方法), 761	
<code>init_weights()</code> ( <i>mmpose.models.backbones.RSN</i> 方 法), 706		<code>init_weights()</code> <i>pose.models.heads.AEHigherResolutionHead</i> 方法), 768	( <i>mm-</i>
<code>init_weights()</code> <i>pose.models.backbones.ShuffleNetV1</i> 717	( <i>mm-</i> 方法),	<code>init_weights()</code> <i>pose.models.heads.AEMultiStageHead</i> 方法), 769	( <i>mm-</i>
<code>init_weights()</code> <i>pose.models.backbones.ShuffleNetV2</i> 718	( <i>mm-</i> 方法),	<code>init_weights()</code> ( <i>mmpose.models.heads.CIDHead</i> 方 法), 771	
<code>init_weights()</code> <i>pose.models.backbones.SwinTransformer</i> 方法), 720	( <i>mm-</i> 方	<code>init_weights()</code> ( <i>mmpose.models.heads.DeconvHead</i> 方法), 774	
<code>init_weights()</code> ( <i>mmpose.models.backbones.TCN</i> 方 法), 723		<code>init_weights()</code> ( <i>mmpose.models.heads.DEKRHead</i> 方法), 773	
<code>init_weights()</code> ( <i>mmpose.models.backbones.VGG</i> 方 法), 724		<code>init_weights()</code> <i>pose.models.heads.HMRMeshHead</i> 方法), 777	( <i>mm-</i>
<code>init_weights()</code> <i>pose.models.backbones.ViPNAS_MobileNetV3</i> 方法), 726	( <i>mm-</i> 方	<code>init_weights()</code> <i>pose.models.heads.MultiModalSSAHead</i> 方法), 780	( <i>mm-</i>
<code>init_weights()</code> <i>pose.models.backbones.ViPNAS_ResNet</i> 方法), 727	( <i>mm-</i> 方	<code>init_weights()</code> <i>pose.models.heads.TemporalRegressionHead</i> 方法), 782	( <i>mm-</i>
<code>init_weights()</code> <i>pose.models.detectors.AssociativeEmbedding</i> 方法), 735	( <i>mm-</i> 方	<code>init_weights()</code> <i>pose.models.heads.TopdownHeatmapMSMUHead</i> 方法), 785	( <i>mm-</i>
<code>init_weights()</code> ( <i>mmpose.models.detectors.CID</i> 方 法), 739		<code>init_weights()</code> <i>pose.models.heads.TopdownHeatmapMultiStageHead</i> 方法), 787	( <i>mm-</i>



<code>init_weights()</code> ( <i>mm-pose.models.heads.TopdownHeatmapSimpleHead</i> 方法), 789	<code>keypoint_3d_pck()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 672
<code>init_weights()</code> ( <i>mm-pose.models.heads.ViPNASHeatmapSimpleHead</i> 方法), 791	<code>keypoint_auc()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 673
<code>init_weights()</code> ( <i>mm-pose.models.necks.FPN</i> 方法), 729	<code>keypoint_epe()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 673
<code>init_weights()</code> ( <i>mm-pose.models.necks.MTA</i> 方法), 731	<code>keypoint_mpjpe()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 674
<code>InterHand2DDataset</code> ( <i>mm-pose.datasets</i> 中的类), 841	<code>keypoint_pck_accuracy()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 675
<code>Interhand3D</code> ( <i>mm-pose.models.detectors</i> 中的类), 748	<code>keypoints_from_heatmaps()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 675
<code>InterHand3DDataset</code> ( <i>mm-pose.datasets</i> 中的类), 843	<code>keypoints_from_heatmaps3d()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 677
<code>Interhand3DHead</code> ( <i>mm-pose.models.heads</i> 中的类), 777	<code>keypoints_from_regression()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 677
<code>is_image_file()</code> (在 <i>mm-pose.apis.webcam.utils</i> 模块中), 666	<b>L</b>
<code>isvalid()</code> ( <i>mm-pose.datasets.Body3DMviewDirectCampusDataset</i> 方法), 822	
<code>isvalid()</code> ( <i>mm-pose.datasets.Body3DMviewDirectShelfDataset</i> 静态方法), 825	<code>L1Loss</code> ( <i>mm-pose.models.losses</i> 中的类), 796
<code>IUVToTensor</code> ( <i>mm-pose.datasets.pipelines.mesh_transform</i> 中的类), 901	<code>Limit_max_fps()</code> (在 <i>mm-pose.apis.webcam.utils</i> 模块中), 666
<b>J</b>	<code>LiteHRNet</code> ( <i>mm-pose.models.backbones</i> 中的类), 698
	<code>load_annotations()</code> ( <i>mm-pose.datasets.Body3DH36MDataset</i> 方法), 820
<code>joints_2d_loss()</code> ( <i>mm-pose.models.losses.MeshLoss</i> 方法), 798	<code>load_config()</code> ( <i>mm-pose.datasets.Body3DH36MDataset</i> 方法), 820
<code>joints_3d_loss()</code> ( <i>mm-pose.models.losses.MeshLoss</i> 方法), 798	<code>load_config()</code> ( <i>mm-pose.datasets.Body3DMviewDirectCampusDataset</i> 方法), 822
<code>JointsEncoder</code> ( <i>mm-pose.datasets.pipelines.bottom_up_transform</i> 中的类), 900	<code>load_config()</code> ( <i>mm-pose.datasets.Body3DMviewDirectPanopticDataset</i> 方法), 823
<code>JointsMSELoss</code> ( <i>mm-pose.models.losses</i> 中的类), 795	<code>load_config()</code> ( <i>mm-pose.datasets.Body3DMviewDirectShelfDataset</i> 方法), 825
<code>JointsOHKMMSELoss</code> ( <i>mm-pose.models.losses</i> 中的类), 796	<code>load_image_from_disk_or_url()</code> (在 <i>mm-pose.apis.webcam.utils</i> 模块中), 667
<b>K</b>	<code>LoadImageFromFile</code> ( <i>mm-pose.datasets.pipelines.loading</i> 中的类), 891
	<code>LoadIUVEFromFile</code> ( <i>mm-</i>
<code>keypoint_3d_auc()</code> (在 <i>mm-pose.core.evaluation</i> 模块中), 672	

<code>pose.datasets.pipelines.mesh_transform</code> 中的类), 901	<code>pose.datasets.pipelines.mesh_transform</code> 中的类), 902
<code>LoadVideoFromFile</code> ( <code>mm-pose.datasets.pipelines.loading</code> 中的类), 891	<code>Message</code> ( <code>mmpose.apis.webcam.utils</code> 中的类), 660
<b>M</b>	<code>mmpose.apis</code> 模块, 631
<code>make_layer()</code> ( <code>mm-pose.models.backbones.MobileNetV2</code> 方法), 701	<code>mmpose.core.evaluation</code> 模块, 669
<code>make_layer()</code> ( <code>mm-pose.models.backbones.ShuffleNetV1</code> 方法), 717	<code>mmpose.core.fp16</code> 模块, 680
<code>make_res_layer()</code> ( <code>mm-pose.models.backbones.ResNeSt</code> 方法), 709	<code>mmpose.core.post_processing</code> 模块, 684
<code>make_res_layer()</code> ( <code>mm-pose.models.backbones.ResNet</code> 方法), 712	<code>mmpose.core.utils</code> 模块, 683
<code>make_res_layer()</code> ( <code>mm-pose.models.backbones.ResNeXt</code> 方法), 710	<code>mmpose.datasets</code> 模块, 807
<code>make_res_layer()</code> ( <code>mm-pose.models.backbones.SEResNet</code> 方法), 716	<code>mmpose.datasets.pipelines</code> 模块, 891
<code>make_res_layer()</code> ( <code>mm-pose.models.backbones.SEResNeXt</code> 方法), 715	<code>mmpose.datasets.pipelines.bottom_up_transform</code> 模块, 897
<code>make_res_layer()</code> ( <code>mm-pose.models.backbones.ViPNAS_ResNet</code> 方法), 727	<code>mmpose.datasets.pipelines.loading</code> 模块, 891
<code>mapper()</code> ( <code>mmpose.datasets.pipelines.shared_transform.Albumenation</code> 静态方法), 892	<code>mmpose.datasets.pipelines.mesh_transform</code> 模块, 901
<code>MeshAdversarialDataset</code> ( <code>mmpose.datasets</code> 中的类), 845	<code>mmpose.datasets.pipelines.pose3d_transform</code> 模块, 902
<code>MeshAffine</code> ( <code>mmpose.datasets.pipelines.mesh_transform</code> 中的类), 901	<code>mmpose.datasets.pipelines.shared_transform</code> 模块, 891
<code>MeshGetRandomScaleRotation</code> ( <code>mm-pose.datasets.pipelines.mesh_transform</code> 中的类), 902	<code>mmpose.datasets.pipelines.top_down_transform</code> 模块, 895
<code>MeshH36MDataset</code> ( <code>mmpose.datasets</code> 中的类), 845	<code>mmpose.models.backbones</code> 模块, 691
<code>MeshLoss</code> ( <code>mmpose.models.losses</code> 中的类), 797	<code>mmpose.models.detectors</code> 模块, 732
<code>MeshMixDataset</code> ( <code>mmpose.datasets</code> 中的类), 846	<code>mmpose.models.heads</code> 模块, 767
<code>MeshRandomChannelNoise</code> ( <code>mm-pose.datasets.pipelines.mesh_transform</code> 中的类), 902	<code>mmpose.models.losses</code> 模块, 791
<code>MeshRandomFlip</code> ( <code>mm-</code>	<code>mmpose.models.misc</code> 模块, 805
	<code>mmpose.models.necks</code> 模块, 728
	<code>mmpose.utils</code>

- 模块, 909
- MobileNetV2 (*mmpose.models.backbones* 中的类), 701
- MobileNetV3 (*mmpose.models.backbones* 中的类), 702
- ModelSetEpochHook (*mmpose.core.utils* 中的类), 683
- MonitorNode (*mmpose.apis.webcam.nodes* 中的类), 657
- MoshDataset (*mmpose.datasets* 中的类), 847
- MPJPELoss (*mmpose.models.losses* 中的类), 796
- MSELoss (*mmpose.models.losses* 中的类), 797
- MSPN (*mmpose.models.backbones* 中的类), 700
- MTA (*mmpose.models.necks* 中的类), 729
- multi\_gpu\_test() (在 *mmpose.apis* 模块中), 638
- MultiItemProcess (*mmpose.datasets.pipelines.shared\_transform* 中的类), 893
- multilabel\_classification\_accuracy() (在 *mmpose.core.evaluation* 模块中), 678
- MultiLossFactory (*mmpose.models.losses* 中的类), 799
- MultiModalSSAHead (*mmpose.models.heads* 中的类), 778
- MultiTask (*mmpose.models.detectors* 中的类), 750
- MultitaskGatherTarget (*mmpose.datasets.pipelines.shared\_transform* 中的类), 893
- ## N
- nearby\_joints\_nms() (在 *mmpose.core.post\_processing* 模块中), 688
- Node (*mmpose.apis.webcam.nodes* 中的类), 644
- norm1 (*mmpose.models.backbones.HRNet* property), 696
- norm1 (*mmpose.models.backbones.ResNet* property), 712
- norm1 (*mmpose.models.backbones.ViPNAS\_ResNet* property), 728
- norm2 (*mmpose.models.backbones.HRNet* property), 696
- NormalizeJointCoordinate (*mmpose.datasets.pipelines.pose3d\_transform* 中的类), 906
- NormalizeTensor (*mmpose.datasets.pipelines.shared\_transform* 中的类), 893
- NoticeBoardNode (*mmpose.apis.webcam.nodes* 中的类), 652
- NVGestureDataset (*mmpose.datasets* 中的类), 847
- ## O
- ObjectAssignerNode (*mmpose.apis.webcam.nodes* 中的类), 656
- ObjectVisualizerNode (*mmpose.apis.webcam.nodes* 中的类), 651
- OffsetGenerator (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 901
- oks\_iou() (在 *mmpose.core.post\_processing* 模块中), 688
- oks\_nms() (在 *mmpose.core.post\_processing* 模块中), 689
- OneHand10KDataset (*mmpose.datasets* 中的类), 848
- ## P
- PAFGenerator (*mmpose.datasets.pipelines.bottom\_up\_transform* 中的类), 901
- PanopticDataset (*mmpose.datasets* 中的类), 850
- ParametricMesh (*mmpose.models.detectors* 中的类), 752
- PhotometricDistortion (*mmpose.datasets.pipelines.shared\_transform* 中的类), 894
- pose\_pck\_accuracy() (在 *mmpose.core.evaluation* 模块中), 678
- PoseLifter (*mmpose.models.detectors* 中的类), 755
- PoseSequenceToTensor (*mmpose.datasets.pipelines.pose3d\_transform* 中的类), 906
- PoseTrackerNode (*mmpose.apis.webcam.nodes* 中的类), 648
- PoseWarper (*mmpose.models.detectors* 中的类), 758
- PoseWarperNeck (*mmpose.models.necks* 中的类), 731
- post\_dark\_udp() (在 *mmpose.core.evaluation* 模块中), 679
- process\_mmdet\_results() (在 *mmpose.apis* 模块中), 638

- `project_joints()` (*mm-pose.models.losses.SemiSupervisionLoss* 静态方法), 801
- `project_points()` (*mm-pose.models.losses.MeshLoss* 方法), 798
- `PyramidVisionTransformer` (*mm-pose.models.backbones* 中的类), 703
- `PyramidVisionTransformerV2` (*mm-pose.models.backbones* 中的类), 705
- ## Q
- `quantize_float()` (*mm-pose.models.backbones.RegNet* 静态方法), 708
- ## R
- `RecorderNode` (*mm-pose.apis.webcam.nodes* 中的类), 658
- `RegNet` (*mm-pose.models.backbones* 中的类), 706
- `RelativeJointRandomFlip` (*mm-pose.datasets.pipelines.pose3d\_transform* 中的类), 906
- `RenameKeys` (*mm-pose.datasets.pipelines.shared\_transform* 中的类), 895
- `report()` (*mm-pose.utils.StopWatch* 方法), 909
- `report_strings()` (*mm-pose.utils.StopWatch* 方法), 910
- `ResNeSt` (*mm-pose.models.backbones* 中的类), 708
- `ResNet` (*mm-pose.models.backbones* 中的类), 711
- `ResNetV1d` (*mm-pose.models.backbones* 中的类), 712
- `ResNeXt` (*mm-pose.models.backbones* 中的类), 710
- `RLELoss` (*mm-pose.models.losses* 中的类), 800
- `rotate_point()` (在 *mm-pose.core.post\_processing* 模块中), 689
- `RSN` (*mm-pose.models.backbones* 中的类), 705
- ## S
- `SCNet` (*mm-pose.models.backbones* 中的类), 712
- `screen_matting()` (在 *mm-pose.apis.webcam.utils* 模块中), 665
- `SemiSupervisionLoss` (*mm-pose.models.losses* 中的类), 801
- `SEResNet` (*mm-pose.models.backbones* 中的类), 715
- `SEResNeXt` (*mm-pose.models.backbones* 中的类), 714
- `set_train_epoch()` (*mm-pose.models.detectors.GestureRecognizer* 方法), 748
- `set_train_epoch()` (*mm-pose.models.heads.MultiModalSSAHead* 方法), 780
- `setup_multi_processes()` (在 *mm-pose.utils* 模块中), 910
- `show_result()` (*mm-pose.models.detectors.AssociativeEmbedding* 方法), 735
- `show_result()` (*mm-pose.models.detectors.CID* 方法), 739
- `show_result()` (*mm-pose.models.detectors.DetectAndRegress* 方法), 742
- `show_result()` (*mm-pose.models.detectors.DisentangledKeypointRegressor* 方法), 746
- `show_result()` (*mm-pose.models.detectors.GestureRecognizer* 方法), 748
- `show_result()` (*mm-pose.models.detectors.Interhand3D* 方法), 749
- `show_result()` (*mm-pose.models.detectors.ParametricMesh* 方法), 753
- `show_result()` (*mm-pose.models.detectors.PoseLifter* 方法), 756
- `show_result()` (*mm-pose.models.detectors.TopDown* 方法), 761
- `show_result()` (*mm-pose.models.detectors.VoxelCenterDetector* 方法), 764
- `show_result()` (*mm-pose.models.detectors.VoxelSinglePose* 方法), 767
- `ShuffleNetV1` (*mm-pose.models.backbones* 中的类), 716

- ShuffleNetV2 (*mmpose.models.backbones* 中的类), 718
- single\_gpu\_test() (在 *mmpose.apis* 模块中), 638
- singleTagLoss() (*mmpose.models.losses.AELoss* 方法), 791
- smooth() (*mmpose.core.post\_processing.Smoother* 方法), 685
- smooth\_l1\_loss() (*mmpose.models.losses.SoftWeightSmoothL1Loss* 静态方法), 803
- Smoother (*mmpose.core.post\_processing* 中的类), 684
- SmoothL1Loss (*mmpose.models.losses* 中的类), 801
- smpl\_loss() (*mmpose.models.losses.MeshLoss* 方法), 799
- soft\_oks\_nms() (在 *mmpose.core.post\_processing* 模块中), 689
- SoftWeightSmoothL1Loss (*mmpose.models.losses* 中的类), 802
- SoftWingLoss (*mmpose.models.losses* 中的类), 803
- split\_ae\_outputs() (在 *mmpose.core.evaluation* 模块中), 680
- StopWatch (*mmpose.utils* 中的类), 909
- SunglassesEffectNode (*mmpose.apis.webcam.nodes* 中的类), 653
- SwinTransformer (*mmpose.models.backbones* 中的类), 719
- sync\_random\_seed() (在 *mmpose.core.utils* 模块中), 684
- T**
- TCFormer (*mmpose.models.backbones* 中的类), 721
- TCN (*mmpose.models.backbones* 中的类), 722
- TemporalRegressionHead (*mmpose.models.heads* 中的类), 780
- timeit() (*mmpose.utils.StopWatch* 方法), 910
- TopDown (*mmpose.models.detectors* 中的类), 759
- TopDownAffine (*mmpose.datasets.pipelines.top\_down\_transform* 中的类), 895
- TopDownAicDataset (*mmpose.datasets* 中的类), 851
- TopDownCocoDataset (*mmpose.datasets* 中的类), 852
- TopDownCocoWholeBodyDataset (*mmpose.datasets* 中的类), 854
- TopDownCrowdPoseDataset (*mmpose.datasets* 中的类), 854
- TopDownFreiHandDataset (*mmpose.datasets* 中的类), 855
- TopDownGenerateTarget (*mmpose.datasets.pipelines.top\_down\_transform* 中的类), 895
- TopDownGenerateTargetRegression (*mmpose.datasets.pipelines.top\_down\_transform* 中的类), 896
- TopDownGetBboxCenterScale (*mmpose.datasets.pipelines.top\_down\_transform* 中的类), 896
- TopDownGetRandomScaleRotation (*mmpose.datasets.pipelines.top\_down\_transform* 中的类), 896
- TopDownH36MDataset (*mmpose.datasets* 中的类), 855
- TopDownHalfBodyTransform (*mmpose.datasets.pipelines.top\_down\_transform* 中的类), 897
- TopdownHeatmapBaseHead (*mmpose.models.heads* 中的类), 782
- TopdownHeatmapMSMUHead (*mmpose.models.heads* 中的类), 783
- TopdownHeatmapMultiStageHead (*mmpose.models.heads* 中的类), 785
- TopdownHeatmapSimpleHead (*mmpose.models.heads* 中的类), 787
- TopDownJhmdbDataset (*mmpose.datasets* 中的类), 857
- TopDownMhpDataset (*mmpose.datasets* 中的类), 859
- TopDownMpiiDataset (*mmpose.datasets* 中的类), 860
- TopDownMpiiTrbDataset (*mmpose.datasets* 中的类), 861
- TopDownOCHumanDataset (*mmpose.datasets* 中的类), 863
- TopDownOneHand10KDataset (*mmpose.datasets* 中的类), 864

- TopDownPanopticDataset (*mmpose.datasets* 中的类), 864
- TopDownPoseEstimatorNode (*mmpose.apis.webcam.nodes* 中的类), 647
- TopDownPoseTrack18Dataset (*mmpose.datasets* 中的类), 864
- TopDownPoseTrack18VideoDataset (*mmpose.datasets* 中的类), 866
- TopDownRandomFlip (*mmpose.datasets.pipelines.top\_down\_transform* 中的类), 897
- TopDownRandomShiftBboxCenter (*mmpose.datasets.pipelines.top\_down\_transform* 中的类), 897
- ToTensor (*mmpose.datasets.pipelines.shared\_transform* 中的类), 895
- train() (*mmpose.models.backbones.HRNet* 方法), 696
- train() (*mmpose.models.backbones.LiteHRNet* 方法), 699
- train() (*mmpose.models.backbones.MobileNetV2* 方法), 702
- train() (*mmpose.models.backbones.MobileNetV3* 方法), 703
- train() (*mmpose.models.backbones.ResNet* 方法), 712
- train() (*mmpose.models.backbones.ShuffleNetV1* 方法), 718
- train() (*mmpose.models.backbones.ShuffleNetV2* 方法), 719
- train() (*mmpose.models.backbones.SwinTransformer* 方法), 720
- train() (*mmpose.models.backbones.VGG* 方法), 724
- train() (*mmpose.models.backbones.ViPNAS\_MobileNetV3* 方法), 726
- train() (*mmpose.models.backbones.ViPNAS\_ResNet* 方法), 728
- train() (*mmpose.models.detectors.DetectAndRegress* 方法), 742
- train() (*mmpose.models.necks.PoseWarperNeck* 方法), 732
- train\_model() (在 *mmpose.apis* 模块中), 638
- train\_step() (*mmpose.models.detectors.DetectAndRegress* 方法), 742
- train\_step() (*mmpose.models.detectors.ParametricMesh* 方法), 754
- transform\_preds() (在 *mmpose.core.post\_processing* 模块中), 690
- ## V
- V2VNet (*mmpose.models.backbones* 中的类), 723
- val\_step() (*mmpose.models.detectors.ParametricMesh* 方法), 754
- vertex\_loss() (*mmpose.models.losses.MeshLoss* 方法), 799
- VGG (*mmpose.models.backbones* 中的类), 723
- VideoEndingMessage (*mmpose.apis.webcam.utils* 中的类), 661
- ViPNAS\_MobileNetV3 (*mmpose.models.backbones* 中的类), 725
- ViPNAS\_ResNet (*mmpose.models.backbones* 中的类), 726
- ViPNASHeatmapSimpleHead (*mmpose.models.heads* 中的类), 789
- vis\_3d\_mesh\_result() (在 *mmpose.apis* 模块中), 639
- vis\_3d\_pose\_result() (在 *mmpose.apis* 模块中), 639
- vis\_pose\_result() (在 *mmpose.apis* 模块中), 639
- vis\_pose\_tracking\_result() (在 *mmpose.apis* 模块中), 640
- VoxelCenterDetector (*mmpose.models.detectors* 中的类), 762
- VoxelSinglePose (*mmpose.models.detectors* 中的类), 764
- ## W
- warp\_affine\_joints() (在 *mmpose.core.post\_processing* 模块中), 690
- WebcamExecutor (*mmpose.apis.webcam* 中的类), 642
- WeightNormClipHook (*mmpose.core.utils* 中的类), 683
- WingLoss (*mmpose.models.losses* 中的类), 804



`with_keypoint` (`mm-pose.models.detectors.AssociativeEmbedding` property), 736  
`with_keypoint` (`mmpose.models.detectors.CID` property), 740  
`with_keypoint` (`mm-pose.models.detectors.DisentangledKeypointRegressor` property), 746  
`with_keypoint` (`mmpose.models.detectors.PoseLifter` property), 757  
`with_keypoint` (`mmpose.models.detectors.TopDown` property), 762  
`with_neck` (`mmpose.models.detectors.PoseLifter` property), 757  
`with_neck` (`mmpose.models.detectors.TopDown` property), 762  
`with_necks` (`mmpose.models.detectors.MultiTask` property), 752  
`with_traj` (`mmpose.models.detectors.PoseLifter` property), 757  
`with_traj_backbone` (`mm-pose.models.detectors.PoseLifter` property), 757  
`with_traj_neck` (`mmpose.models.detectors.PoseLifter` property), 757  
`wrap_fp16_model()` (在 `mmpose.core.fp16` 模块中), 683



模块

`mmpose.apis`, 631  
`mmpose.core.evaluation`, 669  
`mmpose.core.fp16`, 680  
`mmpose.core.post_processing`, 684  
`mmpose.core.utils`, 683  
`mmpose.datasets`, 807  
`mmpose.datasets.pipelines`, 891  
`mmpose.datasets.pipelines.bottom_up_transform`, 897  
`mmpose.datasets.pipelines.loading`, 891  
`mmpose.datasets.pipelines.mesh_transform`,

`mmpose.datasets.pipelines.pose3d_transform`, 902  
`mmpose.datasets.pipelines.shared_transform`, 891  
`mmpose.datasets.pipelines.top_down_transform`, 895  
`mmpose.models.backbones`, 691  
`mmpose.models.detectors`, 732  
`mmpose.models.heads`, 767  
`mmpose.models.losses`, 791  
`mmpose.models.misc`, 805  
`mmpose.models.necks`, 728  
`mmpose.utils`, 909