HRNet

Deep High-Resolution Representation Learning for Human Pose Estimation

2019 CVPR

Ke Sun[1,2*†] Bin Xiao[2*] Dong Liu[1] Jingdong Wang[2]
[1]University of Science and Technology of China [2]Microsoft Research Asia
{sunk,dongeliu}@ustc.edu.cn, {Bin.Xiao,jingdw}@microsoft.com

人体行为动作识别，人机交互，动画制作等

论文地址：https://arxiv.org/abs/1902.09212
推荐博文：https://blog.csdn.net/qq_37541097/article/details/124346626

HRNet

单一个体的姿态评估

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

MS COCO Dataset

HRNet

对于Human Pose Estimation任务，现在基于深度学习的方法主要有两种：

- 基于regressing的方式，即直接预测每个关键点的位置坐标。
- 基于heatmap的方式，即针对每个关键点预测一张热力图（预测出现在每个位置上的分数）。

HRNet

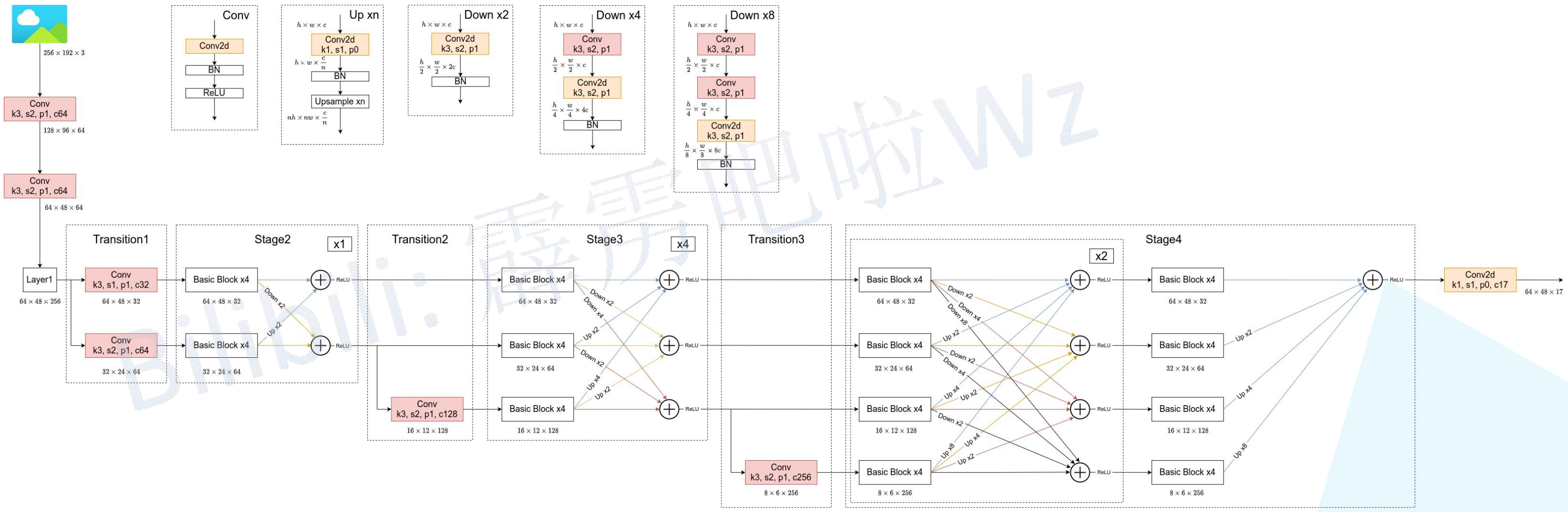
目录

- 1 HRNet网络结构
- 2 预测结果（heatmap）可视化
- 3 损失的计算
- 4 评价准则
- 5 其他
  - 5.1 数据增强
  - 5.2 注意输入图片比例

HRNet

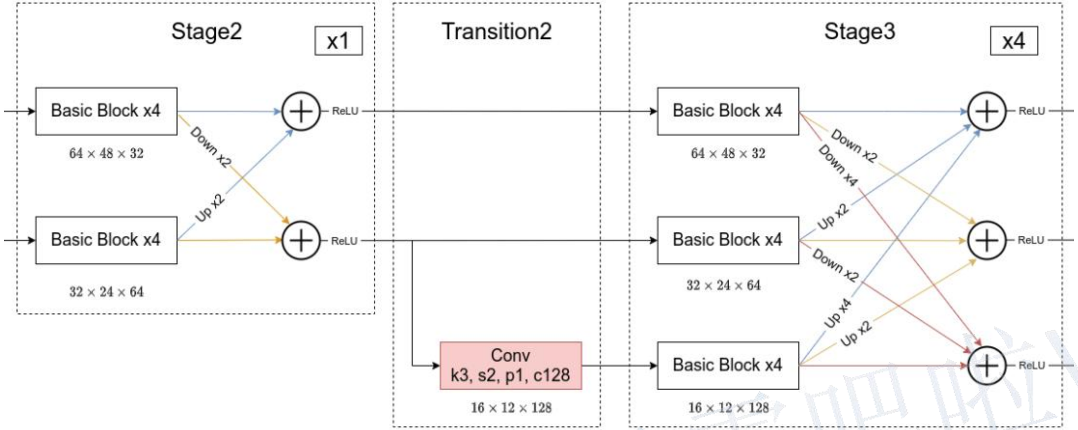
网络结构

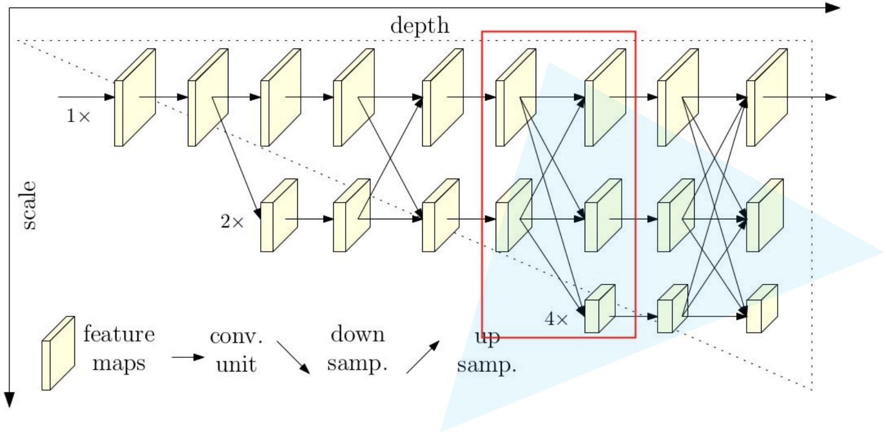
HRNet-W32网络结构简图



HRNet

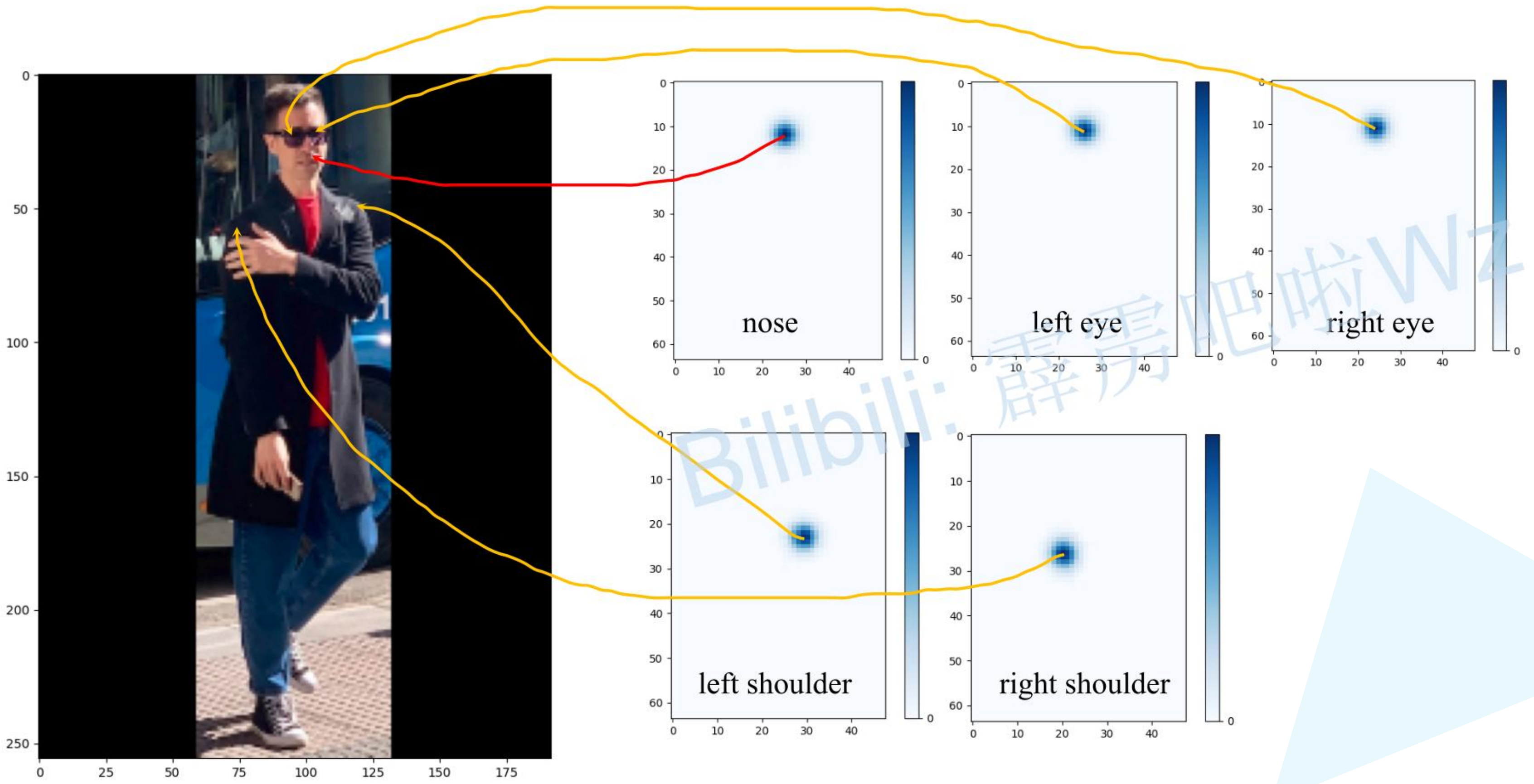
网络结构


Stage2
x1
Transition2
Stage3
x4
Basic Block x4
64 × 48 × 32
Basic Block x4
32 × 24 × 64
Down x2
Up x2
ReLU
Conv
k3, s2, p1, c128
16 × 12 × 128
Basic Block x4
64 × 48 × 32
Basic Block x4
32 × 24 × 64
Basic Block x4
16 × 12 × 128
Down x2
Down x4
Up x2
Down x2
Up x4
Up x2
ReLU


depth
scale
1×
2×
4×
feature maps
conv. unit
down samp.
up samp.

HRNet

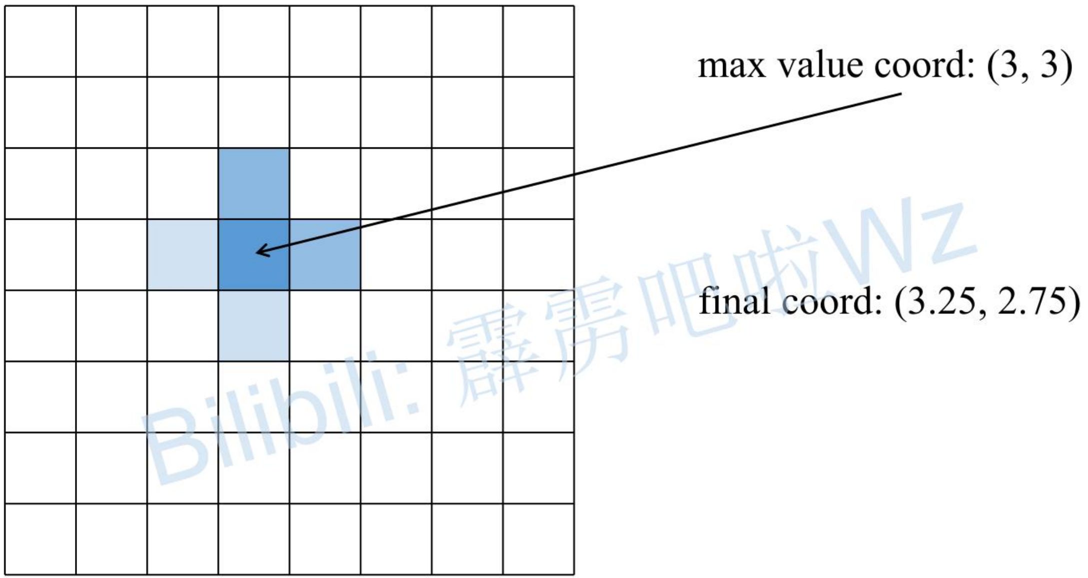
预测结果可视化



HRNet

预测结果可视化

Each keypoint location is predicted by adjusting the highest heatvalue location with a quarter offset in the direction from the highest response to the second highest response.



HRNet

损失的计算

均方误差Mean Squared Error

			1				

×

		0.1	0.2	0.1			
	0.1	0.3	0.5	0.3	0.1		
	0.2	0.5	1	0.5	0.2		
	0.1	0.3	0.5	0.3	0.1		
		0.1	0.2	0.1			

√

GT heatmap

HRNet

损失的计算

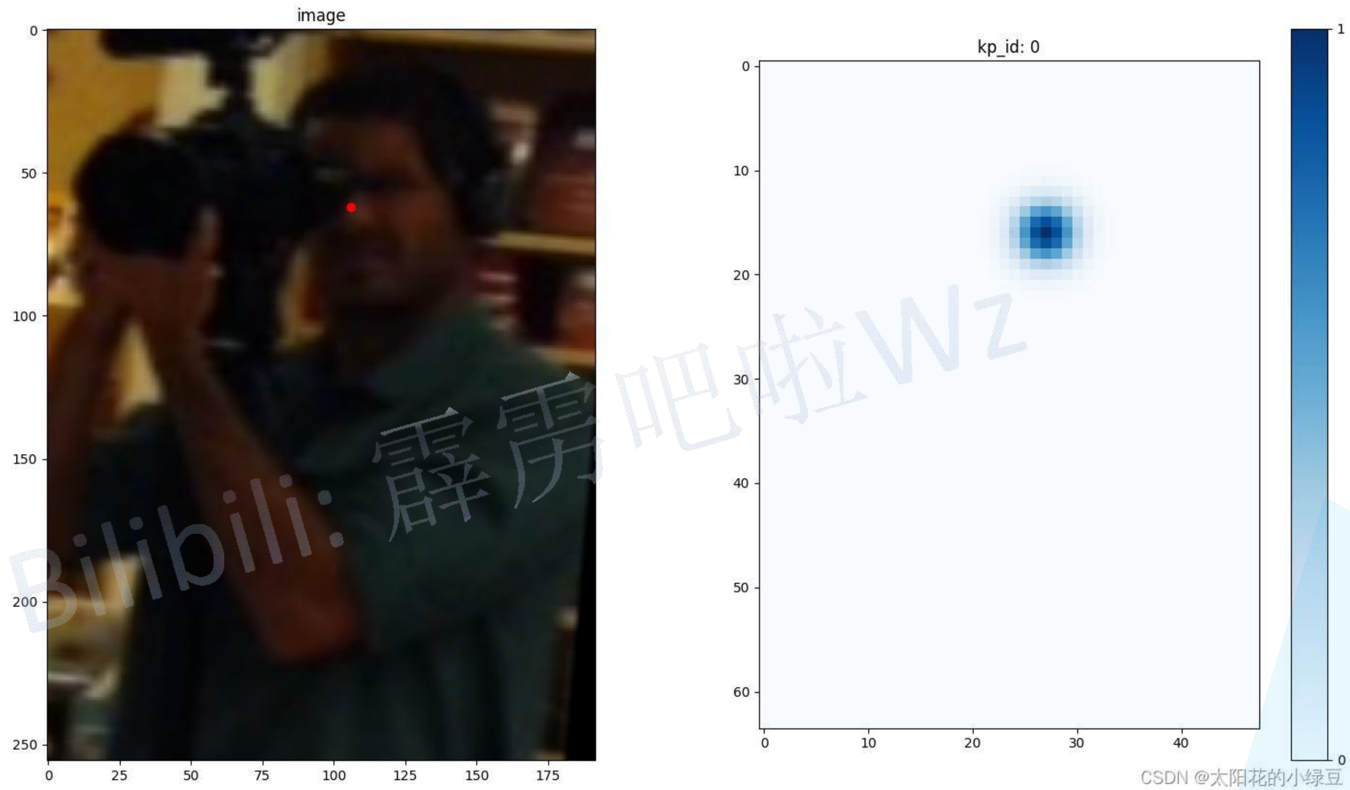


image
0
50
100
150
200
250
0
25
50
75
100
125
150
175
kp_id: 0
0
10
20
30
40
50
60
0
10
20
30
40
1
0
Bilibili: 霹雳吧啦Wz
CSDN @太阳花的小绿豆

HRNet

损失的计算

	0.2	0.3	0.1				
	0.1	0.9	0.4				
	0.1	0.2	0.2				

Predict heatmap

		0.1	0.2	0.1			
	0.1	0.3	0.5	0.3	0.1		
	0.2	0.5	1	0.5	0.2		
	0.1	0.3	0.5	0.3	0.1		
		0.1	0.2	0.1			

GT heatmap

HRNet

损失的计算

每个关键点所计算的损失采用不同的权重

["nose","left_eye","right_eye","left_ear","right_ear","left_shoulder","right_shoulder","left_elbow","right_elbow","left_wrist","right_wrist","left_hip","right_hip","left_knee","right_knee","left_ankle","right_ankle"]

[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.2, 1.2, 1.5, 1.5, 1.0, 1.0, 1.2, 1.2, 1.5, 1.5]

评价准则

在目标检测（Object Detection）任务中可以通过IoU（Intersection over Union）作为预测bbox和真实bbox之间的重合程度或相似程度。在关键点检测（Keypoint Detection）任务中一般用OKS（Object Keypoint Similarity）来表示预测keypoints与真实keypoints的相似程度，其值域在0到1之间，越靠近1表示相似度越高。

$$OKS = \frac{\sum_i [e^{-d_i^2/2s^2k_i^2} \cdot \delta(v_i > 0)]}{\sum_i [\delta(v_i > 0)]}$$

- i代表第i个关键点
- vi代表第i个关键点的可见性，这里的vi是由GT提供
- $\delta(x)$当x为True时值为1，x为False时值为0
- di为第i个预测关键点与对应GT之间的欧氏距离
- s为目标面积的平方根
- ki是用来控制关键点类别i的衰减常数

详情参考: https://cocodataset.org/#keypoints-eval

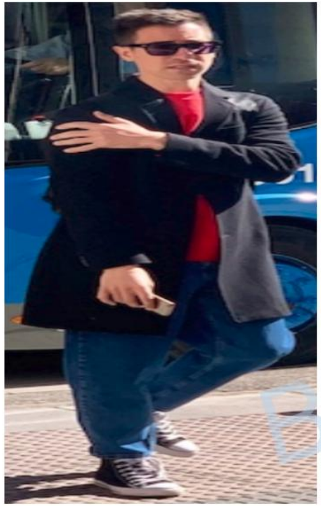
HRNet

数据增强

- 随机旋转（在 -45~45度之间）
- 随机缩放（在0.65到1.35之间）
- 随机水平翻转
- half body（有一定概率会对目标进行裁剪，只保留半身关键点，上半身或者下半身）

HRNet

注意输入图片比例



×

√

√