

AI Department

TGIF

Talk is cheap? Show me the code...

2020-03-26 4:30PM

分享题目： 高分辨率深度神经网络 HRNet

分享者： 李艺如

AI Department

TGIF

发展背景

视觉识别主要包括三大类问题：图像层次（图像分类），区域层次（目标检测）和像素层次（比如图像分割、人体姿态估计和人脸对齐等）。最近几年，用于图像分类的卷积神经网络成为解决视觉识别问题的标准结构，比如图1所示的LeNet-5

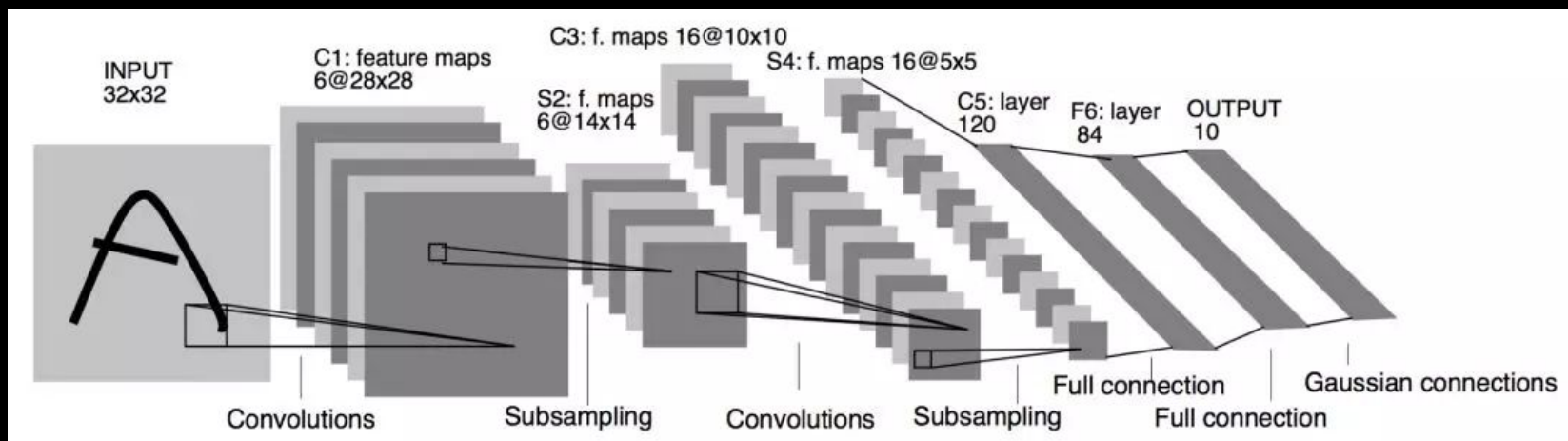


图1. 典型的卷积神经网络：LeNet-5。其它典型的卷积神经网络，如AlexNet、VGGNet、GoogleNet、ResNet、DenseNet等

AI Department

TGIF

发展背景

为了弥补空间精度的损失，研究者在分类卷积神经网络结构的基础上，通过引入上采样操作和/或组合空洞卷积减少降采样次数来提升表征的分辨率，典型的结构包括Hourglass、U-Net等

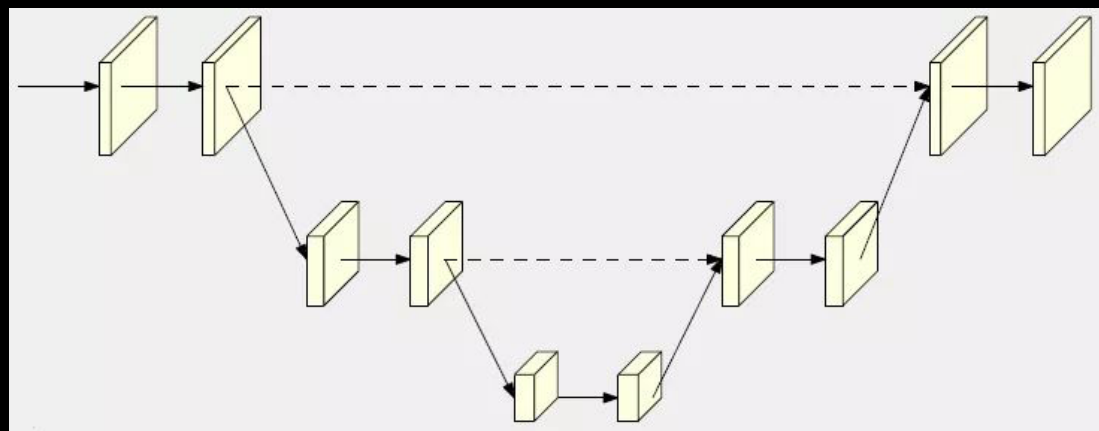


图2. 从低分辨率表征恢复高分辨率表征

AI Department

TGIF

高分辨率网络

先前的分类将分辨率从高到低的卷积串行连接，HRNet则是并行连接。HRNet的整个网络中始终保持高分辨率表征，逐步引入低分辨率卷积，并且将不同分辨率的卷积并行连接。

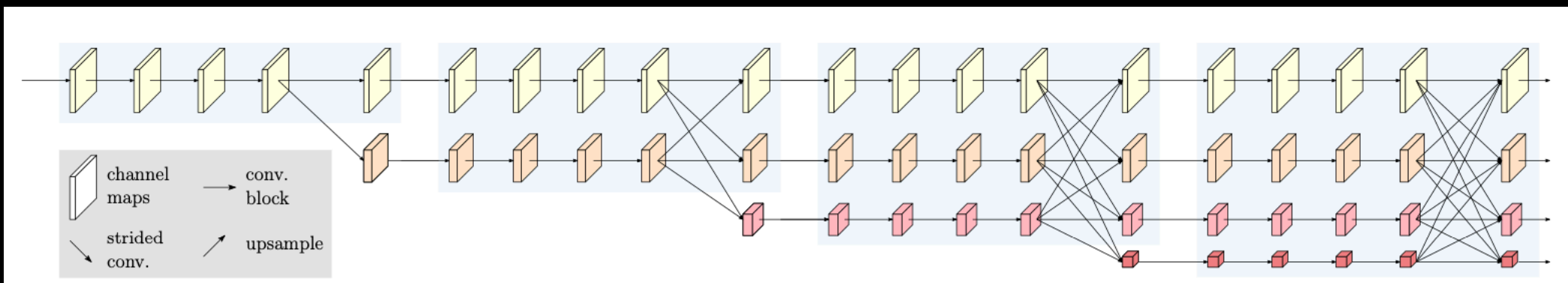


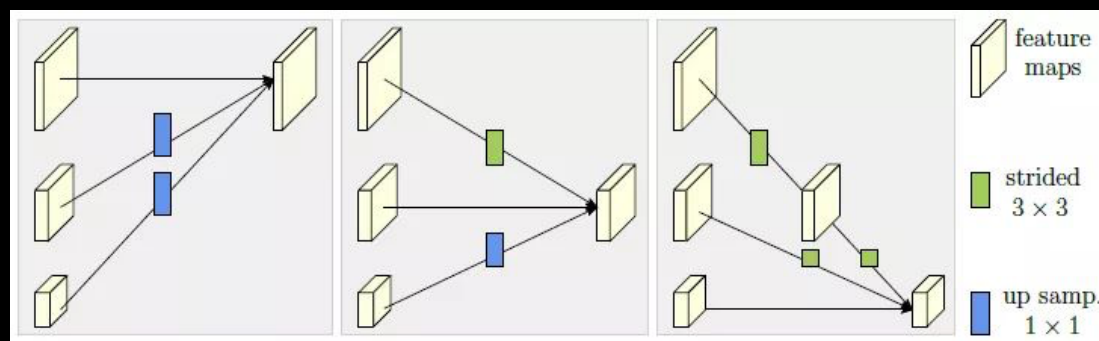
图3. 高分辨率网络 (High-Resolution Network, HRNet)

AI Department

TGIF

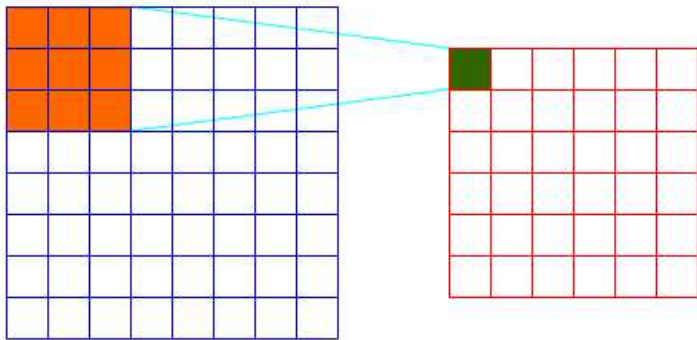
多分辨率表征信息交换

这里以三个分辨率输入和三个分辨率输出为例，如图4所示。每一个分辨率的输出表征都会融合三个分辨率输入的特征，以保证信息的充分利用和交互。将高分辨率特征降到低分辨率特征时，我们采用stride为2的3x3卷积；低分辨率特征到高分辨率特征时，先利用1x1卷积进行通道数的匹配，再利用最近邻插值的方式来提高分辨率。相同分辨率的表征则采用恒等映射的形式。



AI Department

TGIF



AI Department

TGIF

实验

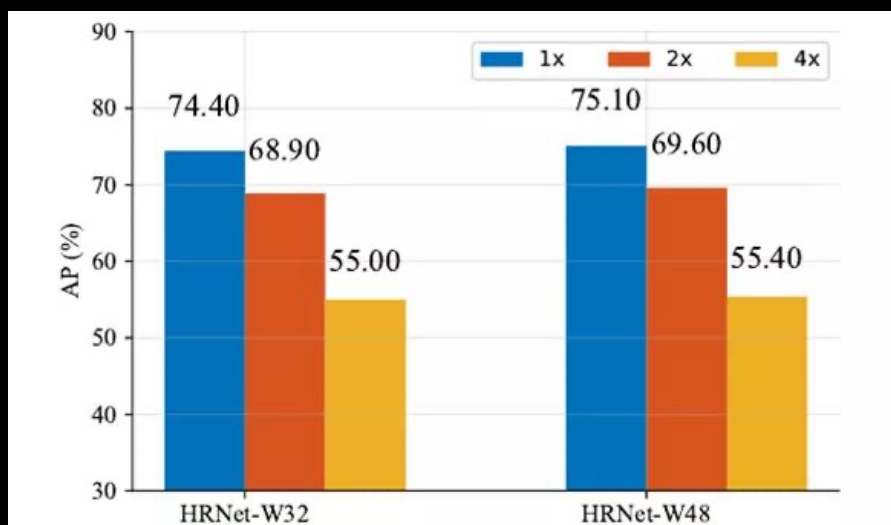
1. 在MS COCO数据集上进行关键点检测，来查看表征分辨率对性能的影响
2. 在MS COCO、PoseTrack等标准数据集中与最先进的方法进行公平对比，都取得了更好的性能。

AI Department

TGIF

1. 表征分辨率对性能的影响

HRNet可输出4种分辨率的表征（1x、2x、4x、以及8x），针对不同的网络输出分辨率在两组模型上做了对比实验。



网络输出分辨率对结果的影响，1x、2x和4x分辨率表征在人体姿态估计的性能。

AI Department

TGIF

2.在标准数据集上的性能

MS COCO数据集是关键点检测的最权威的数据集之一，我们在该数据上对我们的方法进行验证

method	Backbone	Input size	#Params	GFLOPs	AP	AP ⁵⁰	AP ⁷⁵	AP ^M	AP ^L	AR
Bottom-up: keypoint detection and grouping										
OpenPose [6]	-	-	-	-	61.8	84.9	67.5	57.1	68.2	66.5
Associative Embedding [39]	-	-	-	-	65.5	86.8	72.3	60.6	72.6	70.2
PersonLab [46]	-	-	-	-	68.7	89.0	75.4	64.1	75.5	75.4
MultiPoseNet [33]	-	-	-	-	69.6	86.3	76.6	65.0	76.3	73.5
Top-down: human detection and single-person keypoint detection										
Mask-RCNN [21]	ResNet-50-FPN	-	-	-	63.1	87.3	68.7	57.8	71.4	-
G-RMI [47]	ResNet-101	353×257	42.0M	57.0	64.9	85.5	71.3	62.3	70.0	69.7
Integral Pose Regression [60]	ResNet-101	256×256	45.0M	11.0	67.8	88.2	74.8	63.9	74.0	-
G-RMI + extra data [47]	ResNet-101	353×257	42.6M	57.0	68.5	87.1	75.5	65.8	73.3	73.3
CPN [11]	ResNet-Inception	384×288	-	-	72.1	91.4	80.0	68.7	77.2	78.5
RMPE [17]	PyraNet [77]	320×256	28.1M	26.7	72.3	89.2	79.1	68.0	78.6	-
CFN [25]	-	-	-	-	72.6	86.1	69.7	78.3	64.1	-
CPN(ensemble) [11]	ResNet-Inception	384×288	-	-	73.0	91.7	80.9	69.5	78.1	79.0
SimpleBaseline [72]	ResNet-152	384×288	68.6M	35.6	73.7	91.9	81.1	70.3	80.0	79.0
HRNet-W32	HRNet-W32	384×288	28.5M	16.0	74.9	92.5	82.8	71.3	80.9	80.1
HRNet-W48	HRNet-W48	384×288	63.6M	32.9	75.5	92.5	83.3	71.9	81.5	80.5
HRNet-W48 + extra data	HRNet-W48	384×288	63.6M	32.9	77.0	92.7	84.5	73.4	83.1	82.0

COCO test-dev上与最先进方法的性能比较

AI Department

TGIF

2.在标准数据集上的性能

Pose-Track是一个用于人体姿势估计和关节跟踪的大型基准数据集。在Pose-Track数据集的两个任务上进行了验证：（1）多帧人体姿态估计，可以利用其他帧的信息估计某帧的姿态；（2）多帧人体姿态跟踪，需要把不同帧间的同一个人的姿态关联起来。前者性能用mAP来评价，后者性能用MOTA来评价。

Entry	Additional training Data	mAP	MOTA
ML-LAB [84]	COCO + MPII-Pose	70.3	41.8
SOPT-PT [53]	COCO + MPII-Pose	58.2	42.0
BUTD2 [29]	COCO	59.2	50.6
MVIG [53]	COCO + MPII-Pose	63.2	50.7
PoseFlow [53]	COCO + MPII-Pose	63.0	51.0
ProTracker [19]	COCO	59.6	51.8
HMPT [53]	COCO + MPII-Pose	63.7	51.9
JointFlow [15]	COCO	63.6	53.1
STAF [53]	COCO + MPII-Pose	70.3	53.8
MIPAL [53]	COCO	68.8	54.5
FlowTrack [72]	COCO	74.6	57.8
HRNet-W48	COCO	74.9	57.9

在Pose-Track数据集上与最先进方法的性能比较

AI Department

TGIF

谢谢!

AI Department

TGIF

参考资料

Ke Sun, Bin Xiao, Dong Liu, Jingdong Wang: Deep High-Resolution Representation Learning for Human Pose Estimation. CVPR 2019

https://github.com/HRNet/HRNet-Image-Classification/blob/master/lib/models/cls_hrnet.py#L508

AI Department

TGIF