

基于卷积神经网络的 2D 人体姿态估计综述 *

乔 迪,曲 毅

(武警工程大学 信息工程学院,陕西 西安 710086)

摘要: 随着深度学习的快速发展,2D 人体姿态估计作为其他计算机视觉任务的研究基础,其检测速度和精度对后续应用落地具有实际意义。对近年来基于卷积神经网络的 2D 人体姿态估计的方法进行梳理介绍,将现有方法分为人体检测关节点回归融合算法和人体关节点检测聚类算法,同时对当前的主流数据集及其评价准则进行总结,最后对 2D 人体姿态估计当前所面临的困难以及未来的发展趋势做以阐述,为姿态估计相关研究提供一些参考。

关键词: 2D 人体姿态估计;卷积神经网络;人体关键点

中图分类号: TP391.4

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.201087

中文引用格式: 乔迪,曲毅. 基于卷积神经网络的 2D 人体姿态估计综述[J]. 电子技术应用, 2021, 47(6): 15–21.

英文引用格式: Qiao Yi, Qu Yi. Overview of 2D human pose estimation based on convolutional neural network[J]. Application of Electronic Technique, 2021, 47(6): 15–21.

Overview of 2D human pose estimation based on convolutional neural network

Qiao Yi, Qu Yi

(College of Information Engineering, Engineering University of PAP, Xi'an 710086, China)

Abstract: With the rapid development of deep learning, 2D human pose estimation is used as the research basis for other computer vision tasks, and its detection speed and accuracy have practical significance for subsequent applications. This paper introduces the methods of 2D human pose estimation based on convolutional neural networks in recent years. The existing methods are divided into human body detection combined with joint point regression algorithm and human body joint point detection clustering algorithm. At the same time, the current mainstream datasets and the evaluation criteria are summarized, and finally the current difficulties and future development trends of 2D human pose estimation are explained, which provides some references for related research on pose estimation.

Key words: 2D human pose estimation; convolutional neural network; keypoints of the human body

0 引言

传统的 2D 人体姿态估计是通过手工提取特征或建立人体模型,来设计 2D 人体部件检测器。WANG Y 等人^[1]提出基于多树模型的人体姿态估计,同时表征人体部件间的运动学约束关系和依赖关系,DANTONE M 等人^[2]提出建立依赖于身体部位的非线性联合回归器来预测关节位置。在特征的选取上,RAMANAN D 等人^[3]使用颜色直方图来提取部位的外观特征,SAPP B 等人^[4]利用级联的结构化模型来提取轮廓边缘和形状特征,YANG Y 等人^[5]使用 HOG 特征来建立人体各部位的混合模型。传统方法受图像背景、光照、遮挡等的影响较大,并且对于多维特征的选择主观性较强,不能很好地适应人体部件的复杂性和环境的变换,因此利用传统的基于部件模型的方法具有较大的局限性。

随着深度学习的发展,利用神经网络提取的特征具有更好的鲁棒性,基于深度学习的 2D 人体姿态估计利用卷积神经网络模拟人的视觉系统,通过多层网络,得到不同感受野下的丰富特征。TOSHEV A^[6]等人首次基于级联的深度神经网络搭建了人体关节点回归的网络框架,从而摆脱了传统方法无法充分利用图像整体信息的局限性,实现精准的关节点定位和姿态提取。作为计算机视觉的基本任务,人体姿态估计发展状况与虚拟现实、人机交互、智能视频监控和运动行为分析等研究领域都紧密相连。本文主要对基于深度学习的 2D 人体姿态估计方法进行梳理和总结,按视频图像中的人体数量,将人体姿态估计分为单人和多人场景下的检测,并分别阐述了人体姿态估计的常用方法,列举了测试基准数据集和评价准则。此外,本文还对人体姿态估计现阶段面临的问题进行分析,并对该领域未来发展的趋势进行展望。

1 2D 人体姿态估计

2D 人体姿态估计,是通过人体关节点的二维坐标

* 基金项目:国家自然科学基金(61801516);陕西省自然科学基础研究计划(2019JQ-238)

来描述人体骨架信息,基于深度学习的2D人体姿态估计是通过图像特征提取,进行关节点检测和骨架信息恢复,从复杂背景环境中对人体进行分割提取,当前姿态估计的目标在于提高实时检测速度和检测精度。根据检测对象和算法流程的不同,可将其分为人体检测关节点回归融合算法和人体关节点检测聚类算法。前者是利用目标检测算法将每个人从图像背景中识别出来,然后对单人姿态进行估计;后者是对图像中所有的关节点进行检测,而后对其进行聚类连接成相应个体,从而恢复人体姿态,具体分类如图1所示。

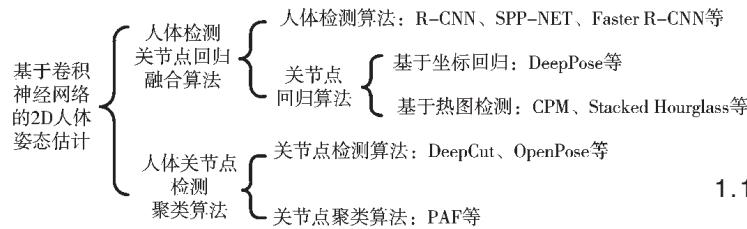


图1 2D人体姿态估计算法总结

1.1 人体检测关节点回归融合算法

人体检测关节点回归融合算法主要包含两个步骤:首先构建人体检测器,利用目标检测算法生成人体提议框(Human Proposals),而后在框内进行人体关节点回归。

1.1.1 人体检测算法

人体检测的任务是要从复杂环境背景中将所关心的部分提取出来,是人体姿态估计的初始化过程。经典的目标检测的算法可以分为两类:一类是先产生候选区域(Region Proposals)再利用卷积神经网络(CNN)分类的目标检测框架,这类算法是基于区域提取的R-CNN^[7]系列算法,如R-CNN^[7]、SPP-Net^[8]、Fast R-CNN^[9]、Faster R-CNN^[10];另一类是以YOLO算法为代表的将目标检测转换为回归问题的目标检测框架(YOLO、SSD),目标检测算法的好坏直接关系到后续单人姿态估计的准确度。当前人体姿态估计主要使用Faster R-CNN作为人体检测框架。

早在2000年PAPAGEORGIOU C等人^[11]就提出了早期的目标检测方法,常用无先验知识的滑窗遍历来进行候选区域的选择,由于检测区域多尺度变换,容易造成漏检和误检。随着深度卷积神经网络的发展,2014年,GIRSHICK R等人^[7]首次提出基于区域提取的R-CNN算法,利用selective search获得约2 000个候选区域,并对区域大小进行归一化,输入神经网络进行特征提取。针对R-CNN模型对候选区域尺寸的限制,HE K等人^[8]在此基础上设计了空间金字塔池化层(Spatial Pyramid Pooling, SPP),使得输入的候选框大小可以进行缩放。为了降低重复计算节省运算开支,GIRSHICK R等人^[9]在SPP-Net的基础上提出Fast R-CNN算法,用ROI Pooling层代替SPP层,将分类问题和边框回归问题进行了合并,极大加快了检测速度。DeepCut^[12]方法就是先利用算法

检测图像中的人体关节点,而后将关节点划分为不同的个人。为了从根本上解决selective search选择候选区域慢的问题,2016年REN S等人^[10]又提出了更快的Faster R-CNN算法,使用RPN(Region Proposals Network)网络将候选区域的选取融合到整个网络结构中,大幅提高了候选框的提取速度和精度。

虽然Faster R-CNN等算法取得了巨大进展,但是其实时性仍不能满足实际需求。随后出现的YOLO、SSD算法优越性逐步凸显。Redmon等人提出了YOLO方法,实现了端到端的目标检测,摒弃了之前先划分候选区域再分类识别的两步走方法,提出了一种单阶段的处理方法,将输入图片划分为7×7的网格,每个格子最多回归一种目标,通过边框回归进行目标区域的修正^[13],该方法舍弃检测精度换取检测速度的大幅提升。

1.1.2 关节点回归算法

通过人体检测得到包含一个人的bounding box,在单人关节点检测中对应输出(Ground Truth)有两种思路:一种是直接回归关节点位置坐标(x, y),即基于坐标直接回归;另一种输出的是heatmap, heatmap最大值对应关键点的位置, heatmap的Ground Truth是以关键点为中心的二维高斯分布,即基于热图回归。

(1)基于坐标回归

基于坐标回归的单人姿态估计是将人体各关节点的二维坐标作为Ground Truth,使输入图像经过端到端的网络直接映射得到每个关节点的位置坐标。这种从单个点直接回归的处理方法缺乏鲁棒性,姿态估计算法早期多采用这种算法,缺点是忽略了人体部件之间的结构信息,单独对每个关节点进行回归。

DeepPose^[6]基于AlexNet网络多阶段端到端回归人体关节点的二维坐标,无需借助人体模型和检测器,在初始阶段得到关节点的粗略位置,在其周围邻域切取小尺寸的子图像作为下一阶段的输入,后续阶段不断对其进行修正优化,具体结构如图2所示,这也奠定了基于坐

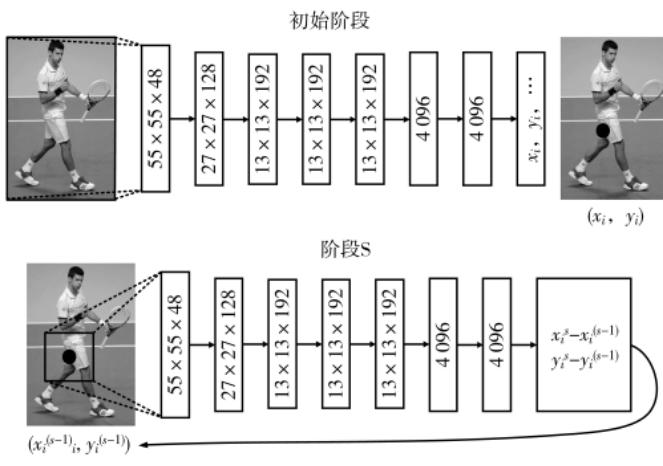


图2 DeepPose 结构图

标回归方法的基础。

由于直接从输入图像中预测关节点的位置是一个非线性问题,SUN X 等人^[14]为了增强结构信息,提出了结构感知的回归方法,使用骨骼来进行姿态表示,同时利用关节点的连接结构来定义损失函数,以便从 2D 姿态扩展到 3D 姿态。FAN X 等人^[15]基于生理图像提出一种双源深度卷积神经网络(DS-CNN),采用多尺度滑动窗口构建图像补丁(part patches)实现关节点的检测和定位。

(2) 基于热图检测

基于热图检测的单人姿态估计是用概率图 heatmap 来表示关节点的粗略位置,训练网络得到的是特征图(feature map)各像素的概率检测值。Heatmap 不仅可以反映像素属于各个关节点的概率分布,还提供了关节点本身和关节点之间的图像特征。基于热图的检测可以为关节点坐标的检测提供先验,使用更为广泛。

① 关节点关系的建模

TOMPSON J 等人^[16]提出采用热图的方式来回归关节点,每个关节点占据一个热图通道,其二维高斯分布以目标关节点位置为中心,针对得到的关节点有误检(false positive)的情况,作者提出建立空间模型(spatial model)进行优化,利用马尔科夫随机场来过滤异常关节点,也为多人场景中的关节点聚类提供了解决思路。PAPANDREOU G 等人^[17]在此基础上进行了改进提出了 G-RMI,基于全卷积 ResNet-101 的网络结构利用热图一偏移解码器将热图(heatmap)和对应的偏移量(offset)结合起来得到精确关节点的定位。为了更好地体现关节之间的依赖关系,YANG W 等人^[18]设计了一个可以灵活构建的树结构模型和循环模型的消息传递层,使得 DCNN 和部位检测器进行联合训练,但是此方法的计算量依然非常庞大。

② 主干网络结构的改进

人体姿态估计中应用的网络结构许多都是基于经典神经网络改进而来的,WEI S E 等人^[19]提出一种多阶段级联网络 CPM,使用顺序化卷积架构对可学习的空间纹理特征进行序列预测,不断去修正响应图,使用神经网络同时学习图片特征和空间信息。残差网络提出后,NEWELL A 等人^[20]提出一种结构对称的堆叠式沙漏(Stacked Hourglass)架构,使用剩余模块作为组成单元,如图 3 所示,使用多尺度特征来捕获各关节点的语义信息和空间位置信息,由于高分辨率是通过上采样获得的,因此会有潜在的信息缺失,四阶 Hourglass 子网络结构如图 4 所示。

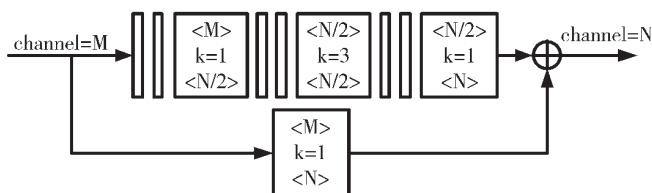


图 3 Residual 模块示意图

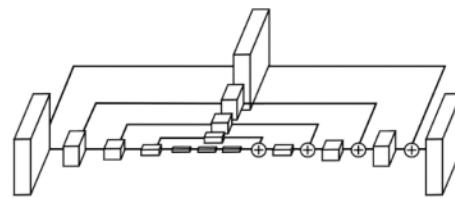


图 4 四阶 Hourglass 子网络

示。FANG H 等人^[21]提出的 RMPE 是基于对称空间变换网络(Symmetric Spatial Transformer Network, SSTN)构建了专门的人体检测框,提出了参数化姿态非最大值抑制(Parametric Pose NonMaximum-Suppression, p-Pose NMS)来消除冗余,提高了准确率和效率。BULAT A 等人^[22]使用了由两个深层子网组成的 CNN 级联结构,分别是基于 ResNet-152 的部位检测网络和基于 VGG 的回归网络。

CHEN Y 等人^[23]提出的 CPN,构建基于 FPN 和 Mask-RCNN 的人体检测器后,级联两个子阶段的网络模块 GlobalNet 和 RefineNet,第一阶段由特征金字塔网络 GlobalNet 负责所有关节点的检测重点提高明显部位的检测精度,第二阶段由 RefineNet 进行进一步修正。后续许多网络结构都是以此为基准模型进行改进的,例如 LI W 等人^[23]在此基础上提出了多阶段人体姿态估计网络 MSPN,将 CPN 的 GlobalNet 作为轻量级的单阶段模块进行堆叠,相邻阶段之间进行特征聚合以保证更高分辨率,MSPN 通过将前一个阶段的对应的两个特征图和当前阶段下采样得到的特征图相连接,避免由于反复下采样和上采样造成的信息丢失。SU K 等人^[24]提出 Channel Shuffle Module(CSM)来促进特征图间的跨通道融合,提出 Spatial Channel-wise Attention Residual Bottleneck(SCARB)利用注意力机制来学习空间和通道维度的关系。

Hourglass 是自上而下人体姿态估计的开山之作,在此基础上,KE L 等人^[25]基于 Hourglass 网络做了改进,使用多层 MSS-Net 进行多尺度监督来加强语义特征学习,而后利用 MSR-Net 处理多尺度信息保持全局一致性。为了提高精度有效应对遮挡问题,TANG W 等人^[26]在 Hourglass 网络的基础上提出了 DLCM 模型,利用深度神经网络学习人体的组成结构,可进行自上而下和自下而上的结构推理。

不同于上述两阶段的算法,为了提高检测效率,NIE X 等人^[27]提出将人体检测和关节点回归两者统一起来当做单阶段来处理的分层结构化姿态表示 SPR,直接预测根节点 root joint 作为人的中心位置,同时将所有关节点进行层次分割,从 root joint 开始预测相邻层级关节点的偏移量,具体如式(1)所示,从而减轻网络的预测压力。

$$(x_i^j, y_i^j) = (x_i^r, y_i^r) + \sum_{h \in H \setminus \{r\}} (\delta \tilde{x}_i^h, \delta \tilde{y}_i^h) \quad (1)$$

③ 多尺度特征提取

为了完善在深度卷积神经网络中特征金字塔的应用,YANG W 等人^[28]设计了金字塔残差模块(PRM)来克

服人体形态变换造成比例变化增强鲁棒性,首先使用不同缩放比例下采样得到不同尺度的特征图,而后再通过上采样将特征图恢复到相同大小融合形成多尺度特征,这一阶段姿态估计的方法大都是从高分辨率特征图下采样至低分辨率,再从低分辨率恢复至高分辨率,此过程中会损失部分细节信息,为了使姿态估计预测的关节点更加准确、空间更加精确,SUN K 等人^[29]提出了高分辨率的人体姿态估计网络 HRNet,通过在高分辨率特征主网络中逐渐并行加入低分辨率特征图子网络,将不同分辨率 feature map 之间的连接从串行改成了并行,在整个过程中一直保持高分辨率分支的存在,各并行网络之间相互交换信息,实现了不同网络间的多尺度融合和特征提取。

1.2 人体关节点检测聚类算法

自下而上的方法通常包括两个阶段,分别是关节点检测和关节点聚类。即首先对视频图像中所有的关节点进行检测,而后通过对其进行分组聚类,对不同人的关节点进行连接恢复人体姿态,该算法适用于多人场景,较之人体检测和关节点回归融合算法计算开销保持稳定,但是关节分组连接复杂度也会相对提高。代表算法有 DeepCut、Associative Embedding、DeeperCut、OpenPose、PIFPAF 等。

PISHCHULIN L 等人^[30]首次提出 DeepCut,利用 Fast R-CNN 来检测身体关节并进行相应类别标记,而后利用整数线性规划将这些类别划分给对应的人从而组合恢复出骨架信息。在此基础上,INSAFUTDINOV E 等人^[31]提出了 DeeperCut,采用残差网络 Resnet 来提高关节检测精度,使用 Image-Conditioned Pairwise Terms 来做优化,压缩候选区域的节点。为了提高实时检测效率,CAO Z 等人^[32]提出了基于 CPM 改进的 OpenPose 方法,可以实现单人和多人的人体动作和手指关节等的识别,利用 PAF 方法给每个肢体构造二维向量来编码肢体的位置和方向信息,从而提高部位之间的关联度。如图 5 所示,OpenPose 先通过 VGG-19 卷积神经网络获取图像特征,之后经历两个分支,上半支路预测关节点置信度,下半支路预测关节点亲和度向量完成关节点的匹配,最后利用匈牙利算法将同一个人的关节点进行聚类连接,恢复人体骨架信息,这一方法获得了 2016 年 COCO 关键点挑战赛的冠军。

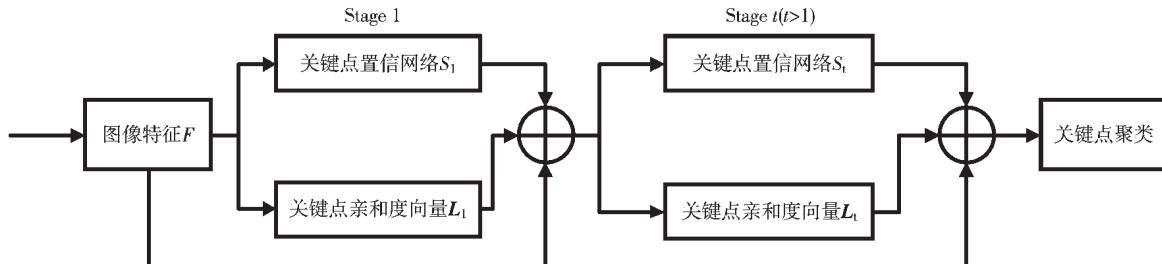


图 5 OpenPose 算法流程示意图

文献[33]提出 Associative Embedding 方法,通过使用标签值来编码部位之间的关联置信度,使得同一个人不同关节点对应的标签值尽可能相似,不同人的不同关节点对应的标签值相差尽可能大,通过划定阈值对标签值数据进行分割,从而将关节点分组到每个人体实例。PAPANDREOU G 等人^[34]提出 personlab 算法,可以同时完成人体姿态估计和人体实例分割任务。

2019 年,KREISS S 等人^[35]提出 PIFPAF 方法,这是一个基于复合场的人体姿态估计算法,使用部分强度场 PIF 来预测关节点热图,使用部分关联场 PAF 来确定关节点的连接,最后使用贪婪算法对部分强度场和部分关联场的信息进行处理,得到人体姿态估计的结果,相比 OpenPose 算法性能有了明显提升。2020 年,字节跳动 XIAO B 团队^[36]基于之前的 HRNet 和 Associative Embedding 提出了一种高分辨率网络 HigherHRNet,利用多分辨率训练和热图聚合的策略,使高分辨率网络能够预测具有尺度感知的热图,也是目前自下而上方法中性能最优秀的网络之一。

2 数据集和评价准则

2.1 目前常用的 2D 人体姿态估计数据集

2.1.1 单人姿态估计数据集

(1)LSP 数据集

LSP 数据集定义了 14 个人体关节点,分别为头顶、脖子、左右肩部、左右肘部、左右手腕、左右臀部、左右膝盖和左右脚踝。样本数在 2 000 左右,目前基本弃用。

(2)FLIC 数据集

FLIC 数据集是从电影中截取的场景片段,场景中包含多人的情况下,只标签了一个人的关节信息。共有 5 003 个样本,包括训练集 3 987 张、测试集 1 016 张,是单人人体关节点检测的数据集,通常在实验中作为第二数据集使用。

该数据集定义了 9 类人体关节点,包括左右眼睛、鼻子、左右肩部、左右手肘、左右手腕、左右臀部,关节点主要集中在上半身。

(3)MPII 数据集

MPII 数据集约有 2.5 万张图片,包含超过 4 万个具有关节点注释的人体,包括被遮挡部位关节点标注和头部方向。定义了 16 个人体关节点,分别为头顶、脖子、胸部、左右肩部、左右手肘、左右手腕、盆骨、左右臀部、左

右膝盖、左右脚踝。同时可作为多人姿态估计的数据集。

2.1.2 多人姿态估计数据集

在现有的多人姿态估计数据集中,常用的有MPII、COCO和AI Challenger数据集。

MSCOCO数据集是人体姿态识别领域应用的主流数据集,包含了20万张图片和25万个具有关节点注释的人体实例。定义了17个人体关节点,分别为左右眼睛、左右耳朵、鼻子、左右肩部、左右手肘、左右手腕、左右臀部、左右膝盖和左右脚踝。

AI Challenger数据集作为竞赛数据集使用,包含30万张图片,其中训练集21万张,验证集3万张,测试集A3万张,测试集B3万张,均对其进行了关节点标注。定义了14类人体关节点,分别为头部、脖子、左右肩膀、左右手肘、左右手腕、左右臀部、左右膝盖、左右脚踝。

2.2 评价准则

根据不同数据集的数据标签特点,2D人体姿态估计有多种评价准则,如表1所示。

表1 常见的2D人体姿态估计数据集及评价标准

数据集名称	关节点个数	样本数目	评价标准	使用情况
单人姿态估计	LSP	14	2K	PCP
	FLIC	9	5K	PCK
多人姿态估计	MPII	16	25K	PCKh
	MSCOCO	17	30W	OKS、AP、AR
	AI Challenger	14	30W	OKS、AP、AR

2.2.1 单人姿态估计评价准则

LSP数据集采用部位精度百分比PCP(Percentage of Correct Parts)来衡量定位的精确程度,如果预测的肢体两个端点都在Ground Truth阈值之内,则判定该定位正确有效。

MPII数据集中采用头部尺度关节点百分比PCKh(Percentage Correct Keypoints on head length)评价关节点检测的准确率,即计算出预测值和真实值的归一化距离小于设定阈值的比例,如式(2)所示:

$$PCKh@{\sigma}=\frac{1}{K} \sum_{k=1}^K \left(\frac{1}{N} \sum_{i=1}^N \left(\frac{\left| \left| y_k^i - \hat{y}_k^i \right| \right|^2}{\left| \left| y_{lhip}^i - \hat{y}_{rsho}^i \right| \right|^2} \leq \sigma \right) \right) \quad (2)$$

其中N表示样本数目,k表示第k个关节点, $\left| \left| y_k^i - \hat{y}_k^i \right| \right|^2$ 表示预测关节点和真实位置坐标之间的距离, $\left| \left| y_{lhip}^i - \hat{y}_{rsho}^i \right| \right|^2$ 表示人头部的最长距离,σ表示阈值,通常设为0.5。

区别于MPII,FLIC数据集使用PCK作为评价指标时用人体躯干大小代替头部框大小作归一化距离。

2.2.2 多人姿态估计评价准则

使用目标关键点相似度(Object Keypoint Similarity,

OKS)来评估预测值和真实值之间的差异,从而衡量模型的表现。OKS指标的计算公式如式(3):

$$OKS=\frac{\sum_i \left[\exp \left(-\frac{d_i^2}{2S^2k_i^2} \right) \delta(v_i>0) \right]}{\sum_i [\delta(v_i>0)]} \quad (3)$$

其中 d_i^2 是预测值和真实值的欧几里得距离的平方, S^2 是人体在图像中所占面积的平方, k_i^2 表示归一化因子是对数据集所有关节点标准差的平方, $v_i \in \{0, 1, 2\}$ 表示关键点的可见性。

AP表示所有的OKS的平均精度,AP50表示OKS阈值为0.5时的平均精度,APM表示小尺度目标的平均精度,APL表示尺度在阈值之上目标的平均精度。

在模型计算消耗比较方面,采用浮点计算量FLOPs作为指标,具体指的是卷积神经网络前向传播时所需要的乘法次数,用单张图片的测试速度作为评价测试速度的指标,同时用参数个数作为模型大小的评价指标。

3 难点与发展趋势

3.1 面临的问题

(1) 遮挡问题

由于真实生活中背景环境复杂,遮挡问题在一定程度上制约了模型性能的提升,容易造成漏检和误检。

(2) 实时性要求

如何降低网络的复杂度和计算开销是制约人体姿态估计算法落地的重要因素,随着图像视频中人数的增多,尤其是自上而下算法的时间开销将会明显增加。

3.2 未来发展的趋势

(1) 提高检测精度和效率

真正将人体姿态估计算法应用于无人驾驶、安全监控等实时性要求高的领域,还需要进一步简化网络结构设计轻量化模型,降低模型的计算开销,使其能够实现实时快速高精度检测。

(2) 移动端应用

在移动端设备部署人体姿态估计技术是一个方向,在不影响检测效率和准确率的情况下,与其他技术相结合拓展外围应用,增强现实真正做到人机交互。

4 结论

随着深度学习的发展卷积神经网络的出现,用于人体姿态估计的模型结构和算法性能也不断优化和提升,无人驾驶、智能监控等任务需求也在一定程度上促进了人体姿态估计工作的进行。本文对两大类2D人体姿态估计方法进行了介绍,对每个阶段具有代表性的模型特点进行了阐述,并介绍了人体姿态估计常用的数据集和评价准则。最后对人体姿态估计面临的问题和发展趋势进行了介绍,在未来具有广阔的应用前景。

参考文献

- [1] WANG Y, MORI G. Multiple tree models for occlusion and spatial constraints in human pose estimation[C]. European Conference on Computer Vision(ECCV), 2008: 710–724.
- [2] DANTONE M, GALL J, LEISTNER C, et al. Human pose estimation using body parts dependent joint regressors[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2013: 3041–3048.
- [3] RAMANAN D. Learning to parse images of articulated bodies[C]. Conference on Advances in Neural Information Processing Systems, 2006: 1129–1136.
- [4] SAPP B, TOSHEV A, TASKAR B. Cascaded models for articulated pose estimation[J]. European Conference on Computer Vision(ECCV), 2010: 406–420.
- [5] YANG Y, RAMANAN D. Articulated pose estimation with flexible mixtures-of-parts[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2011: 1385–1392.
- [6] TOSHEV A, SZEGEDY C. DeepPose: human pose estimation via deep neural networks[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2014: 1653–1660.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2014: 580–587.
- [8] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(9): 1904–16.
- [9] GIRSHICK R. Fast R-CNN[J]. IEEE International Conference on Computer Vision, 2015: 1440–1448.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.
- [11] PAPAGEORGIOU C, POGGIO T. A trainable system for object detection[J]. International Journal of Computer Vision, 2000, 38(1): 15–33..
- [12] CHEN Y, WANG Z, PENG Y, et al. Cascaded pyramid network for multi-person pose estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2018: 7103–7112.
- [13] 姚群力, 胡显, 雷宏. 深度卷积神经网络在目标检测中的研究进展[J]. 计算机工程与应用, 2018(1): 1–9.
- [14] SUN X, SHANG J, LIANG S, et al. Compositional human pose regression[J]. IEEE International Conference on Computer Vision(ICCV), 2017: 2602–2611.
- [15] FAN X, KANG Z, LIN Y, et al. Combining local appearance and holistic view: dual-source deep neural networks for human pose estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2015: 1347–1355.
- [16] TOMPSON J, JAIN A, LECUN Y, et al. Joint training of a convolutional network and a graphical model for human pose estimation[J]. Eprint Arxiv, 2014: 1799–1807.
- [17] PAPANDREOU G, ZHU T, KANAZAWA N, et al. Towards accurate multi-person pose estimation in the wild[J]. IEEE International Conference on Computer Vision(ICCV), 2017: 4904–4911.
- [18] YANG W, OUYANG W, LI H, et al. End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 3073–3082.
- [19] WEI S E, RAMAKRISHNA V, KANADE T, et al. Convolutional pose machines[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 4724–4732.
- [20] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2016: 483–499.
- [21] FANG H, XIE S, TAI Y, et al. RMPE: regional multi-person pose estimation[J]. IEEE International Conference on Computer Vision(ICCV), 2016: 2334–2343.
- [22] BULAT A, TZIMIROPOULOS G. Human pose estimation via convolutional part heatmap regression[C]. European Conference on Computer Vision(ECCV), 2016: 717–732.
- [23] LI W, WANG Z, YIN B, et al. Rethinking on multi-stage networks for human pose estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 1352–1363.
- [24] SU K, YU D, XU Z, et al. Multi-person pose estimation with enhanced channel-wise and spatial information[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 5674–5682.
- [25] KE L, CHANG M C, QI H, et al. Multi-scale structure-aware network for human pose estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 5674–5682.
- [26] TANG W, YU P, WU Y. Deeply learned compositional models for human pose estimation[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 190–206.
- [27] Nie Xuecheng, Feng Jiashi, Zhang Jianfeng, et al. Single-stage multi-person pose machines[J]. IEEE International Conference on Computer Vision(ICCV), 2019: 6951–6960.
- [28] YANG W, LI S, OUYANG W, et al. Learning feature pyramids for human pose estimation[J]. IEEE International Conference on Computer Vision(ICCV), 2017: 1281–1290.
- [29] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 5674–5682.

- (CVPR), 2019: 5693–5703.
- [30] PISHCHULIN L, INSAFUTDINOV E, TANG S, et al. DeepCut: joint subset partition and labeling for multi person pose estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 5667–5675.
- [31] INSAFUTDINOV E, PISHCHULIN L, ANDRES B, et al. DeeperCut: a deeper, stronger, and faster multi-person pose estimation mode[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2016: 34–50.
- [32] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: realtime multi-person 2D pose estimation using part affinity fields[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 7291–7299.
- [33] NEWELL A, HUANG Z, DENG J. Associative embedding: end-to-end learning for joint detection and grouping[C]. Advances in Neural Information Processing Systems, 2017: 2277–2287.
- [34] PAPANDREOU G, ZHU T, CHEN L C, et al. PersonLab:

person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 269–286.

- [35] KREISS S, BERTONI L, ALAHI A. PifPaf: composite fields for human pose estimation[C]. IEEE Computer Vision and Pattern Recognition(CVPR), 2019: 11977–11986.
- [36] CHENG B, XIAO B, WANG J, et al. HigherHRNet: scale-aware representation learning for bottom-up human pose estimation[C]. IEEE Computer Vision and Pattern Recognition(CVPR), 2020: 5386–5395.

(收稿日期: 2020-11-10)

作者简介:

乔逸(1996-), 女, 硕士, 主要研究方向: 计算机视觉。

曲毅(1976-), 男, 博士, 副教授, 主要研究方向: 雷达、深度学习。



扫码下载电子文档

(上接第 14 页)

- [11] LIU Y, YU F R, LI X, et al. Decentralized resource allocation for video transcoding and delivery in blockchain-based system with mobile edge computing[J]. IEEE Transactions on Vehicular Technology, 2019, 68(11): 11169–11185.
- [12] GUO F, YU F R, ZHANG H, et al. Adaptive resource allocation in future wireless networks with blockchain and mobile edge computing[J]. IEEE Transactions on Wireless Communications, 2020, 19(3): 1689–1703.
- [13] HAO Y, LI Y, DONG X, et al. Performance analysis of consensus algorithm in private blockchain[C]. 2018 IEEE Intelligent Vehicles Symposium(IV), Changshu, 2018: 280–285.

- [14] XU H, ZHANG L, LIU Y, et al. RAFT based wireless block-chain networks in the presence of malicious jamming[J]. IEEE Wireless Communications Letters, 2020, 9(6): 817–821.

- [15] WANG S, OUYANG L, YUAN Y, et al. Blockchain-enabled smart contracts: architecture, applications, and future trends[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 49(11): 2266–2277.

(收稿日期: 2021-05-20)

作者简介:

芒戈(1985-), 男, 硕士, 工程师, 主要研究方向: 频谱工程。



扫码下载电子文档



广告

欢迎订阅, 2021年度 《信息技术与网络安全》

中国科技期刊数据库来源期刊

主管单位: 中国电子信息产业集团有限公司 主办单位: 华北计算机系统工程研究所 (中国电子信息产业集团有限公司第六研究所)

月刊 定价: 26 元/期

编辑部电话: (010) 66608908 66608981

电 话 订 阅: (010) 82306084

邮 局 订 阅: 邮发代号 82-417



版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所