计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200700061

二维人体姿态估计研究进展

冯晓月　宋　杰
东北大学软件学院　沈阳 110189
(cmnand@foxmail.com)

摘　要　人体姿态估计一直是计算机视觉领域的研究热点,随着人体姿态估计方法的性能和精度不断提升,目前可以广泛应用于人机交互、智能监控和人体活动分析等领域。人体姿态估计属于强应用相关的研究领域,现有研究成果均不同程度地涉及方法、模型和应用层面,亟待对其进行系统性归纳和总结。文中综述了大量二维人体姿态估计的研究成果,以供研究人员参考。具体包括:单人和多人姿态估计方法,基于 ResNet,Hourglass 和 HRNet 的姿态估计模型,以及姿态估计在人机交互和智能监控领域的应用。文中提出的关于移动设备中的人体姿态估计、拥挤场景下的人体姿态估计和装备人群的姿态估计等研究问题和研究思路,是现有研究的良好补充,为研究人员提供了广阔的研究空间。

关键词:人体姿态估计;关键点检测;神经网络;Hourglass;ResNet;HRNet
中图法分类号　TP311

Research Advance on 2D Human Pose Estimation

FENG Xiao-yue and SONG Jie
Software College,Northeastern University,Shenyang 110819,China

Abstract　Human pose estimation has always been a research hotspot in the field of computer vision. With the continuous improvement of the performance and accuracy of human pose estimation methods,it can be widely used in human-computer interaction,intelligent surveillance and human activity analysis,etc. In this paper,the methods,models and applications of two-dimensional human pose estimation are reviewed and analyzed,and the future research direction is prospected. The introduction of the method is divided into single person and multi-person pose estimation. In terms of the model,it mainly introduces the models based on ResNet,Hourglass and HRNet. In terms of the application,it mainly introduces the application in the field of human-computer interaction and intelligent surveillance. The research prospect is mainly aimed at the expansion of application scenarios. This paper summarizes the research results in recent years and sorts out the possible research directions.

Keywords　Human pose estimation,Key-point detection,Neural network,Hourglass,ResNet,HRNet

1　引言

人体姿态估计指从单个 RGB 图像中精确地估计出人体的位置以及检测骨骼关键点的位置。人体姿态估计一直是计算机视觉领域的研究热点,它是诸多计算机视觉任务的基础,如动作分类、异常行为检测,以及自动驾驶等。

人体姿态估计可分为二维姿态估计和三维姿态估计。在实际应用中,目前大多数图像是二维图像,三维图像相对较少,而三维姿态估计也可以利用二维姿态估计进行推理[1]。在进行三维姿态估计时,首先要进行二维姿态估计,因为基于检测的模型在二维的关键点检测中表现更好,而在三维空间下,基因非线性程度高,输出空间大,在三维的关键点检测中效果较差,所以二维姿态估计有着重要的研究价值。鉴于二维姿态估计的重要地位,本文主要总结其最新的研究进展。

人体姿态估计领域目前已经有许多优秀的研究成果。现有少量综述类研究报告总结了二维单人和多人姿态估计方法,研究人员可初步了解二维姿态估计的研究进展和主流方法[2],但缺少模型和应用层面的归纳。本文拟填补这一空白,从方法、模型及应用层面综述研究进展,以期研究人员能够系统、全面地了解现有方法及其应用领域。

本文第 2 节介绍了二维人体姿态估计的基本方法和现存挑战;第 3 节介绍了单人姿态估计和多人姿态估计的现有方法及其优缺点;第 4 节介绍了基于 ResNet、Hourglass、HRNet 和对抗网络的模型;第 5 节介绍了二维人体姿态估计在人机交互和智能安防等领域的应用;第 6 节介绍了人体姿态估计的未来研究趋势以及可能的研究方向;最后总结全文。

2　问题与挑战

姿态估计的评价指标大体上可以分为性能和精度两大类,前者指计算效率或达到某计算效率的算力大小,后者则为

到稿日期:2020-07-01　返修日期:2020-08-28　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61672143)
This work was supported by the National Natural Science Foundation of China (61672143).
通信作者:宋杰(songjie@mail.neu.edu.cn)

估计的精确度。在深度学习方法出现之前，传统方法主要研究单人二维姿态估计问题。传统方法基于图结构模型，其优点是能够估计任意人体姿态而无需构建人体姿态库。但由于其采用人工设定的 HOG 和 SHIFT 特征，无法充分利用图像信息，导致估计算法受制于图像中迥异的外观、视角、遮挡和固有几何模糊性，因此在部分场景中估计精度较低。此外，传统方法多采用深度图等数字图像来提取姿态特征。由于采集深度图像需要使用专业的采集设备，成本较高，且采集过程需要同步多个视角的深度摄像头，以减小遮挡问题带来的精度损失，导致姿态特征数据的获取过程复杂。综上所述，基于图结构模型的传统方法具有低精度和高成本的缺点。

深度学习理论的提出打破了传统方法的局限性。研究人员开始将深度神经网络应用于姿态估计。现有研究表明，神经网络初步解决了传统方法中“人体姿态特征提取难”这一问题，因此估计精度较传统方法有大幅度提升，且适用于更多的场景，由此引发了学界致力于基于深度学习的姿态估计方法的研究。随着估计精度的逐步提高，研究人员又将单人姿态估计方法推广到了多人姿态估计。

基于神经网络的多人姿态估计，无论是采用自上而下还是自下而上的方法，均面临着遮挡处理、检测性能和关键点跟踪等关键问题，这也是目前学界在该领域面临的挑战。首先是遮挡问题，包括人体姿态的图像中常见的因拥挤而导致的人体遮挡和复杂的背景干扰，增加了姿态估计的难度。其次是估计性能，为实现实时估计和实时判定，计算性能至关重要，或采用较小的算力满足性能要求，如移动设备，或在性能不变的前提下减少服务器的算力投入。然后是关键点跟踪问题。相比基于图像的姿态估计，基于视频的姿态估计还需完成姿态跟踪任务，以及针对视频场景中的每位目标人进行人体及关键点的跟踪。相比目标人整体跟踪，该目标人的关键点更有可能在视频中频繁移动，后者的跟踪难度大于前者。最后，一些具体应用领域也存在很多亟待研究的特定需求，如移动设备中的人体姿态估计、拥挤场景下的人体姿态估计和装备人群的姿态估计，目前这方面的研究尚存在空白。

本文将详细介绍基于深度学习的姿态估计的现有研究成果，这些成果在不同程度上解决了上述问题，并且在部分领域已得以应用。

3 估计方法

姿态估计分为单人姿态估计和多人姿态估计两大类，前者根据给定边界框的裁剪图像预测一个目标的所有人体关键点，后者需要进一步估计一幅图像中所有人的姿态。二维姿态估计方法的分类如图 1 所示。

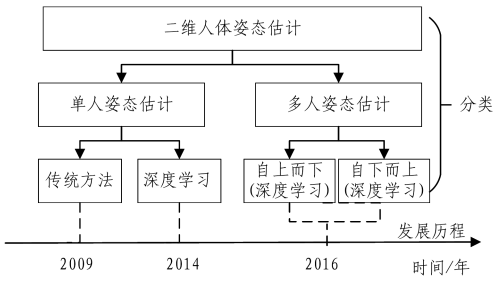

图 1 二维姿态估计方法分类

Fig. 1 Classification of two-dimensional human estimation methods

3.1 单人姿态估计

单人姿态估计指图像中只有待检测目标人，只需要检测该目标的所有关键点。从人工定义人体特征发展到基于深度特征，在方法层面上，单人姿态估计已趋于成熟，也为多人姿态估计的研究奠定了基础。

(1)传统方法

传统方法基于预先构建的人体姿态模板库，并通过特定的映射关系来实现图像特征信息到库中姿态的映射。事实上，图像中的人体姿态是千变万化的，而模板库只能囊括人体的部分姿态，因此该方法有较大的局限性。为了能够更加灵活且准确地定义人体姿态，学界提出了图结构模型[3]。进一步地，Felzenszwalb 等[4]将图结构归入统计学的框架，并假设各个部件之间服从树形结构，显著降低了计算复杂度，使图结构在人体姿态估计领域得以应用。基于此，基于图结构的人体姿态估计算法现已成为了主流算法。

图结构包含两个部分：1)部件模型(Part Model)，用于描述组成人体的各个部分；2)空间模型(Spatial Model)，用于描述各个部件之间的空间关系。大多数研究致力于寻找表达能力更强的部件模型和空间模型来提高姿态估计的精度。在部件模型方面，与人体姿态模板库方法相比，部件模型使用了表达能力更强的图像特征，如 HOG[5] 特征和 SIFT[6] 特征。例如，Ramanan 从图片中提取具有强表达能力的关节特征来解决关节匹配问题。图片中提取的关节特征如图 2 所示。此外，文献[7]将前景分割技术用于姿态估计以去除背景对估计精度的影响。文献[8]训练出更具判别性的部件检测器以提高姿态估计的精度。在空间模型方面，文献[9]提出非树形结构的模型，用于解决人体姿态估计中的自遮挡问题。文献[10]提出混合部件(Mixtures of Parts)模型，该模型能够表达更为复杂多样的空间约束。混合部件模型包含多种对姿态估计非常重要的信息，如部件的几何形变约束、部件的外观信息等。除了关注部件之间的空间约束，部分研究关注更大范围的约束。例如，用于上半身姿态估计的 Armlet[11] 和用于整体姿态估计的 Poselet[12]。

图 2 图片中提取的关节特征

Fig. 2 Joint features extracted from picture

综上所述，传统方法将人体关键点检测问题描述为图结构问题，并基于手工提取特征检测关键点位置[4,10,13-18]，但总体精度不高。近年来，随着卷积神经网络 (Convolutional Neural Networks，CNN) 的发展，姿态估计精度显著提高。

(2)基于深度学习的方法

2014 年，Toshev 等提出的 DeepPose[19] 首次引入 CNN 来解决姿态估计问题，并提出了一系列 CNN 姿态回归器(Pose Regressors)来处理姿态估计问题，提高了关节定位点

的精度。该方法易于实现，不需要显式地设计特征提取器和局部探测器，也不需要为关节建立拓扑结构，且相比传统方法，姿态估计的精度显著提高。DeepPose是人体姿态估计研究开始从传统方法转向深度学习方法的里程碑。类似地，Tompson等也采用深度神经网络(Deep Neural Network，DNN)和图形模型预测关键点热图(Heatmap)来提高关节定位的鲁棒性[20]。自此之后，人体姿态估计研究都基于神经网络。

众所周知，神经网络层数太深会出现梯度消失的问题。针对这一问题，Wei等[21]提出了一种多阶段估计方法——卷积姿态机(Convolutional Pose Machine，CPM)，即首先产生一个粗略的结果，然后在之后的阶段中不断重新确定结果。卷积姿态机对网络每个阶段的输出进行中间监控，进而解决梯度消失的问题。其中Belief Map提供易估计部位的上下文信息，为不易检测的部位(遮挡部位)提供了线索。

对于一张包含人体姿态的图像，目前普遍采用的检测方法是“从高分辨率处理至低分辨率”，或者“从低分辨率处理至高分辨率”，前者由细节至整体，后者由整体至细节。两种方法同样有效，且适用于不同的场景，大部分研究均选用其中之一，例如，全卷积网络[22]和整体嵌套架构[15]很好地支持了前者，但对后者的支持程度却不高。Hourglass网络[23]优雅地扩展了全卷积网络[24]和带残差学习的反卷积网络[25]，它更加均衡地支持这两种处理方法，因此得到了学术界的广泛认可。事实上，Hourglass网络较好地解决了基于神经网络的单人姿态检测，同时也可用于多人姿态估计的研究。

从应用角度来看，单人姿态检测最典型的应用是“人机交互”，而多人姿态估计则可以用于“人人交互”，如体育赛事、安防监控、工业领域、教育领域中都存在大量的多人互动场景。多人姿态估计的应用场景更加广泛，且逐渐成为了学术界研究的重点。

3.2 多人姿态估计

多人姿态估计比单人姿态估计更加复杂，不仅需要扫描整个图像以寻找可能的人体关节候选区域，还需要生成关节热图来进一步预测对应关键点的真实位置；同时，还需要根据当前身体关节和相邻层中每个身体部位之间的置信度，来确定哪些关节与当前关节属于同一个人。多人姿态估计方法主要分为自下而上和自上而下两类，表1列出了其中一些方法在COCO数据集上的性能。

表1 部分方法在COCO数据集上的性能比较

Table 1 Performance comparison of partial methods on COCO datasets

方法	AP	AP_{50}	AP_{75}	AP_M	AP_L
G-RMI	60.5	82.2	66.2	57.6	66.6
Mask R-CNN	63.1	87.3	68.7	57.8	71.4
RMPE	61.8	83.7	69.8	58.6	67.6
Simple Baselines	73.7	91.9	81.1	70.3	80.0
PAFs	61.8	84.9	67.5	57.1	68.2

(1)自上而下的方法

自上而下的方法指：先利用人体检测器检测出图像中的所有人，再使用单人姿态估计的方法估计每个人的姿态。这类方法虽然精度高但实时性较差，且会受到目标检测任务的制约。此外，由于其基于单人姿态估计的方法，在遇到遮挡问题时精度会下降。主流自上而下的方法有G-RMI[26]，CPM[21]，Mask R-CNN[27]，RMPE[28]和Simple Baselines[29]。

Papandreou等[26]提出的G-RMI方法首先使用Faster R-CNN预测可能包含人体目标的边界框的位置和大小，然后使用ResNet预测每个关键点的热图和热图上的点到地面真值位置的偏移量。为了结合这些输出，引入了一种新颖的热图-偏移聚合方法来获得精准的人体关键点，最终利用带偏移量的热图来获得关键点的最终位置。

Mask R-CNN[27]首先预测人体的包围盒(Bounding-box)，然后裁剪相应人体包围盒的特征图，检测人体关键点。Mask R-CNN的网络结构比较简单：在Faster R-CNN的基础上，在原本两个分支(分类和坐标回归)上增加了第三个分支进行图像的语义分割。Mask R-CNN的Mask生成和分类可以并行执行。DensePose借用了Mask R-CNN的架构，同时带有Feature Pyramid Network(FPN)的特征以及ROI-Align池化，DensePose用深度学习把2D图像坐标映射到3D表面上，再加以每秒多帧的速度处理密集坐标，最后实现动态人物的精确定位和姿态估计。

Fang等[28]提出了区域多人姿态检测(Regional Multi-person Pose Estimation，RMPE)框架，该框架由对称空间变换网络(Symmetric Spatial Transformer Network，SSTN)、参数化姿态非最大抑制(Non-Maximum-Suppression，NMS)和姿态引导区域框生成器(Pose-Guided Proposals Generator，PGPG)3部分组成。在单人姿势估计器(Single-Person Pose Estimation，SPPE)结构上添加SSTN，能够在不精准的区域框中提取到高质量的人体区域，并行的SPPE分支(SSTN)可以优化自身网络。使用NMS可以解决重复检测的问题，该结构使用了自创的姿态距离度量方法来比较姿态之间的相似度，用数据驱动的方法优化姿态距离参数，使用PGPG来强化训练数据。利用RMPE框架，AlphaPose对不精准的人类边界框进行精确的动作估计。AlphaPose是一款精准的多人姿态评估工具，它既可以在图片、视频或多图中进行姿态估计，也能在画面中对动作进行追踪。

Bing等[29]认为当下的姿态估计方法太过于复杂，提出了一种更简单的姿态估计和跟踪的基线方法(Simple Baselines)，其姿态估计是基于在骨干网络ResNet上添加一些去卷积层。Bing等认为此方法是从深度和低分辨率特征图估计热图最简单的方法，且效果出众，希望以此激发新的想法并提出简单的估计方法。

(2)自下而上的方法

自下而上的方法是先检测图像中所有的关键点，然后关联和分组属于不同人的部分，即关键点检测和关键点聚类。该方法的研究主要集中于探求优秀的关键点聚类算法。与自上而下的方法相比，自下而上的方法的精度普遍较差，关键点特征的提取难度大，且同样不能很好地解决遮挡问题，但实时性较好。主流的自下而上的方法有PAFs[30]、DeepCut[31]和Associative Embedding[32]。

Cao等[30]实现了将关节点快速连接到肢体躯干(Part Affinity Fields，PAFs)，并将预测到的关节点匹配到不同人的姿态中。该方法对全局上下文进行编码，使用相同序列预测

过程的两个分支。其中，一个分支使用 CNN 根据置信图进行关节点预测，另一个分支使用 CNN 获得每个关节点的 PAF，最后合并两个分支的学习结果。与 PAFs 高度相似的 OpenPose 是一种对多人身体、面部和手部形态进行实时评估的框架，同时提供 2D 和 3D 的多人关键点检测，还有针对估计具体区域参数的校准工具箱。

DeepCut[31]将图像中不同人的姿态估计问题抽象为一个整数线性规划（ILP）问题，并将部分检测候选对象划分为单人，然后将人体与标记的身体部位相结合，得到最终的姿态估计结果。DeeperCut[33]使用 ResNet[34]改进 DeepCut，并使用图像条件成对（Image-Conditioned Pairwise，ICP）将得到的大量候选区域节点压缩至一定数量，提高了姿态估计的精度。DeepCut 和 DeeperCut 的对比如图 3 所示[31,33]。

(a)DeepCut (b)DeeperCut

图 3 DeepCut 和 DeeperCut 对比

Fig. 3 Comparison between DeepCut and DeeperCut

Newell 等[32]提出了一种“检测和分组紧密结合”的关联嵌入（Associative Embedding）方法，即用标签将每个检测与同一组中的其他检测相关联。由于多人姿态系统必须检测所有人及其相应的关键点，因此文中将关联嵌入方法与堆叠 Hourglass 网络集成在一起，为每个身体关节生成和标记热图，然后将具有相似标签的身体关节分组到个人身上，得到最终的多人姿态估计。

无论是自上而下的方法还是自下而上的方法，决定其估计精度的关键都是神经网络分类模型。因此本文将在下一节重点讨论神经网络模型的研究。

4 神经网络

基于深度学习的姿态估计方法几乎都利用了神经网络分类器，神经网络结构很大程度上影响了分类精度。许多学者也致力于设计出更好的网络结构。目前学术界关注的网络模型主要基于 ResNet、Houglass、HRNet 和对抗网络。

4.1 ResNet

ResNet 的成功之处在于引入了残差模块。ResNet 设计了一种使用 Skip Connection 的残差结构，很好地解决了网络退化问题，从而提高了网络的深度，得到了表达能力更强的特征，具有更高的精确度。因此 ResNet 备受学者青睐，并以此为基础进行了大量扩展。本节介绍了部分基于 ResNet 的模型网络，表 2 对这些模型进行了简要总结。

表 2 基于 ResNet 的网络模型对比

Table 2 Comparison of network models based on ResNet

文献	网络	研究目标	基本思路	数据集
[35]	Dual Path Networks	权衡精度和计算成本	以一种前馈方式将网络的每一层连接，缓解了梯度消失问题，加强了特征的传播，大大减少了参数的数量	MS COCO，PoseTrack
[40]	Cascaded Pyramid Network	解决遮挡问题	通过 GlobalNet 和 RefineNet 两个阶段分别定位容易检测的关键点和难以检测的关键点	MS COCO
[41]	Real-time Convolutiona Networks	提高可靠性和速度	假设深度图像比 RGB 图像包含更少的结构且更容易处理，同时保留了人体检测和姿态推断所需的信息，从而允许使用更简单的网络来完成任务	DIH
[42]	End-to-End Feature Pyramid Network	实现稳定的实时速度和高精度	网络利用上下文信息产生更高质量的连接预测结果；增加了一个特征金字塔网络结构来适应不同的人类规模；增加了基于注意力的机制来掩盖不同比例尺特征图的最佳适用区域	COCO
[43]	Lightweight Pose Network	设计简单、轻便的网络	基于 Simple Baseline 的体系结构设计原则，其网络模型大小仅为 Simple Baseline (ResNet50) 的 9%，并且计算复杂度 (FLOPS) 仅为 11%	COCO

针对估计精度和计算成本之间权衡的问题，Ning 等[35]提出了双路径网络（Dual Path Networks，DPN），该网络用于关键点和肢体关联向量的回归，在速度和精度方面较 OpenPose[36]网络都有较大提高。双路径网络是一种混合网络[37]，它融合了 DenseNet[38]和 ResNeXt[39]的核心思想。ResNeXt 是广泛使用的 ResNet 的一个变体，DenseNet 的核心是其以一种前馈方式将每一层连接，缓解了梯度消失问题，加强了特征的传播，大大减少了参数的数量。

针对关键点被遮挡、关键点不可见以及背景复杂等问题，Chen 等[40]采用级联金字塔网络（Cascaded Pyramid Network，CPN），该网络包含两个部分：GlobalNet 和 RefineNet。GlobalNet 是一种特征金字塔网络，以 ResNet 为骨干网络，它可以成功地定位眼睛和手等容易检测的关键点，但无法准确检测被遮挡或不可见的关键点；RefineNet 则尝试通过集成来自 GlobalNet 的所有级别的特征表示，以及在线困难关键点挖掘（Online Hard Keypoint Mining）技术来处理“困难”关键点。与 CPN 相关的研究成果被广泛应用。一般来说，CPN 求解多人姿态估计问题时，首先采用自顶向下的流水线生成一组基于检测器的人体包围盒，然后用 CPN 对每个人体包围盒进行关键点定位。

为获得一种适用于人机交互（Human-Computer Interaction，HRI）场景的可靠、快速的多人姿态估计算法，Martiezt 等[41]提出了一种与深度图像相结合的 ResNet 扩展模型。他们假设深度图像比 RGB 图像包含更少的结构且更容易处理，同时保留了人体检测和姿态推断所需的信息，从而进一步简化了网络结构，提高了分类性能。该文在性能和精度之间权

衡，提高了姿态估计的实时性。同样地，为实现稳定的实时性和高精度，Luo 等[42]提出了一种基于 ResNet 的端到端姿态估计网络（End-to-End Feature Pyramid Network），该网络直接产生连接信息而不是中间信息，网络处理后唯一的计算是基于网络输出的节点连接，通过使用上下文信息，网络产生了更高质量的连接预测结果。为了适应不同的人数规模，该文增加了一个特征金字塔网络结构。此外，还增加了 Attention 机制，以掩盖不同比例尺特征图的最佳适用区域。这 3 方面的改进加快了网络的运算速度。

许多现有方法为了追求更高的估计精度，更倾向使用复杂网络结构或计算昂贵的网络模型，而忽略了算力成本。Zhang 等[43]基于 ResNet 提出了一种简单轻量的姿态网络。首先，采用深度卷积和注意力机制重新设计了一个轻量级瓶颈块；然后，根据 Simple Baseline[29]的体系结构设计原则，提出了一种基于轻量级块的姿态网络（Lightweight Pose Network，LPN）。网络 LPN-50 的模型大小（α-PARAMs）仅为 Simple Baseline（ResNet50）的 9%，并且计算复杂度（FLOPS）仅为 11%，而平均精度（AP）的差距只有 1.3。为了充分发挥 LPN 的潜力，得到更准确的预测结果，Zhang 等还提出了迭代训练策略和模型不可知的后处理函数 β-soft Argmax。

4.2 Hourglass

Hourglass[23]的出现是由“需要捕捉每一个尺度上的信息”这一需求驱动的。与许多产生像素级输出的卷积方法一样，Hourglass 网络能够汇集到一个非常低的分辨率，然后在多个分辨率上进行上采样并组合特征[22,27]。另一方面，Hourglass 与先前的设计不同，它采用更为对称的拓扑结构，允许将多个 Hourglass 模块端到端地放在一起来扩展单个网络。这使得在各个尺度上重复自下而上、自上而下的推断成为可能。同时，结合使用中间监督，重复的双向推理对网络的最终性能至关重要。图 4 给出了单个 Hourglass 模型的结构，图 5 给出了将多个 Hourglass 模型堆叠之后形成的模型结构。后续的许多网络模型都是以 Hourglass 为骨干网络进行设计的，本文介绍了部分基于 Hourglass 网络的模型，表 3 对这些模型进行了简要介绍。

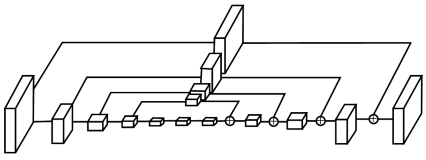

图 4 单个 Hourglass 网络

Fig. 4 Single Hourglass network

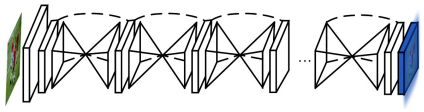

图 5 多个 Hourglass 网络堆叠

Fig. 5 Multiple Hourglass network stacks

为了研究学习特征金字塔，Yang 等[44]设计了一个以 Hourglass 为基本结构的金字塔残差模型（Pyramid Residual Module，PRMs），以增强深度卷积神经网络（Deep Convolutional Neural Networks，DCNNs）在尺度上的不变性。首先在多分支（Multi-branch）网络中以不同的子采样率获得多种特征，然后 PRMs 学习在输入不同特征尺度上的卷积滤波器。此外，Yang 等通过分析发现当前 MSR[11]和 Xavier[12]权重初始化方法不适用于多分支网络，并提出了新的权重初始化策略。

许多学者都致力于在降低网络复杂度的同时提高精度，文献[45-47]提供了可能的网络结构。Nie 等[45]提出了一种生成划分网络（Generative Partition Networks，GPN），它以 Hourglass 为骨干网络，具有低复杂度和高精度的联合检测和重组的优点。GPN 设计的生成模型有效地实现了鲁棒的人体检测和关节分割。此外，GPN 将人类姿态联合推理的推理过程归纳为一个图划分问题，并利用可靠的全局亲和线索对每个人的检测进行局部优化，从而降低了复杂性，提高了精度。随后，Nie 等[46]又提出了一种姿态分割网络（Pose Partition Networks，PPN），该网络同样以 Hourglass 为骨干网络，同时学习“联合检测器（Joint Detector）”和“稠密回归器（Dense Regressor）”，PPN 算法具有复杂度低、联合检测和分割精度高的优点。PPN 在特定的嵌入空间（由人的质心参数化）内对全局候选关节进行稠密回归，有效地生成了鲁棒的人体检测和关节分割。然后，PPN 通过对每个人的局部检测进行图形分割，利用可靠的全局亲和线索推断出人体关节配置，这样，PPN 降低了计算复杂度，提高了多人姿态估计的正确性。从减少网络参数的角度，Zhao 等[47]设计了一个 Lite Hourglass 网络，该网络使用混合卷积块来减少参数数目，同时保持较高的精度。混合卷积块用不同速率的扩展卷积来建立多上下文路径，这样不仅减少了参数的数量，而且扩大了感受野。在多人场景中，由于图像中人数众多且容易发生遮挡，难以将关节正确地分配给相应的个体。Luo 等[48-49]提出了一种提高图像多人姿态估计精度的有效方法，即将分形 Hourglass 网络与关节血缘模式匹配机制相结合，生成身体每个部位的关键点热图，并将其分配给相应的个体。首先，构建三层分形网络，对多人关节位置热图进行回归，该热图有助于增强图像区域的可接受性，捕获更多的关节局部上下文特征信息，从而生成关键点热图中间预测，优化人体关节回归结果。随后，提出了层次双向推理算法，即关节亲属关系模式匹配机制，来确定最佳匹配的关节对。迭代上述关节，逐层匹配过程，直到所有关节都分配给相应的个体。

在多人姿态估计中，由于左右关节的相似性，其类型识别一直是一个难题。一般意义上，可通过堆叠多个精化模块（Refinement Modules）以增加网络的感受野并捕获更多的全局上下文来解决这个问题。Zhao 等[50]利用多层上下文关联进行多人姿态估计的聚类特征聚合网络，网络由特征提取骨干网、关键点检测和标签嵌入组成。网络堆叠多个 Hourglass 模块，通过重复自上而下和自下而上的信息传递来提取多尺度特征，在多个密集的关键点簇的监督下叠加多个 Hourglass 模块，使网络的准确识别能力逐步增强。为了开发更多的上

下文信息，该网络同时学习互补的语义信息，使检测到的关键点受到一定的上下文约束。

表3 基于Hourglass网络的模型对比

Table 3 Comparison of network models based on Hourglass

文献	网络	研究目标	基本思路	数据集
[44]	Pyramid Residual Module	增强深度卷积神经网络在尺度上的不变性	学习在输入不同特征尺度上的卷积滤波器，特征在多分支网络中以不同的子采样率获得	MPII，LSP
[45]	Generative Partition Networks	降低复杂度，提高精度	将人类姿态联合推理的过程定义为一个图划分问题，并用可靠的全局AF-求偶线索对每个人的检测进行局部优化	MPII，WAF，PASCAL-Person-Part
[46]	Pose Partition Networks	降低复杂度，实现联合检测，提高分割精度	在特定的嵌入空间内对全局候选关节进行稠密回归，有效地生成鲁棒的人检测和关节分割。然后通过对每个人的局部检测进行图形分割，利用可靠的全局a-ffinity线索，推断出人体关节配置	MPII，WAF，PASCAL-Person-Part
[54]	Hierarchical Adversarial Network	提高对特殊部位的位置预测精度	网络的生成部分和判别部分被设计成编码层次结构第一阶段（父母）和层次结构第二阶段（子女）之间的空间关系	MPII，LSP，LIP
[48-49]	Multi-Layer Fractal Network，Combining fractal hourglass network	提高多人姿态估计精度	将分形Hourglass网络与关节血缘模式匹配机制相结合，生成身体每个部位的关键点热图，并将其分配给相应的个体	MPII，COCO
[47]	Lite Hourglass Network	减少网络参数以降低复杂性	混合卷积块用不同速率的扩展卷积建立多上下文路径，不仅减少了参数的数量，而且扩大了感受野	COCO
[50]	Cluster-wise learning network	提高关节预测的准确性	网络堆叠多个Hourglass模块，通过重复自上而下和自下而上的信息传递来提取多尺度特征，同时学习互补的语义信息以利用更多的上下文信息	MS COCO

4.3 HRNet

HRNet[51]是一种高分辨率网络。它通过并行连接高分辨率到低分辨率卷积来保持高分辨率表示，并通过重复跨并行卷积执行多尺度融合来增强高分辨率表示。在像素级分类、区域级分类和图像级分类中，证明了这些方法的有效性。

自下而上的多人姿态估计方法的关键步骤为热图预测和关键点分组。为提高热图预测的精度，Cheng等[52]基于HRNet提出了HigherHRNet，它是HRNet的简单扩展。HigherHRNet通过对HRNet输出的高分辨率特征图进行去卷积，生成高分辨率的特征图，对中小型人群的姿态估计更为精确。HigherHRNet建立高质量的多层次特征，进行多尺度姿态预测。在不使用多尺度测试的情况下，HigherHRNet的热图预测精度超过了COCO数据集上的所有自下而上的网络模型。

Zhang等[53]在HRNet的基础上提出了一种新型的注意模块，即标准化注意模块（De-Normalized Attention，DNA），以解决常规注意模块的特征衰减问题。该模块将原始HRNet扩展为空间、通道和分辨率的DNA，旨在评估不同位置、通道和分辨率特征的重要性，提高了网络的特征表征能力。同时在HRNet的每一层添加由高到低分辨率的由细到粗的连接，以增加网络拓扑的最大深度。此外，在HRNet的末尾修改关键点回归器，以准确预测关键点热图，该模型在不使用额外关键点训练数据的情况下达到了最高精度。

4.4 对抗网络

人体关节预测的正确性很大程度上影响了姿态估计的准确度，Wang等[54]提出生成性对抗网络（GAN）来提高人体姿态估计的精度。虽然GAN只能学习局部人体关节约束，但利用GAN框架中的注意机制可以学习人体关节的长期依赖关系，从而提高人体关节预测的正确性。该模型由生成网络G(Generator)和判别网络D(Discriminator)两个网络组成。G网络生成热图，显示身体所有关节的关键点在每个位置的置信度得分。D网络对预测的热图和Ground truth热图进行重建（Reconstruct），通过对抗训练区分真假热图。尽管GAN的出现提高了关节预测的正确性，但其对特殊部位的位置预测效果较差，如手腕和脚踝等变形较大的部位，以及颈部和臀部等变形较小的部位。为解决这个问题，Radwan等[55]提出了对抗深度神经网络，该网络显示了人体各部位的空间层次结构，且对抗网络的生成部分和判别部分用于对层次结构中第一阶段（父母）和第二阶段（子女）之间的空间关系进行编码。生成器和判别器网络为两个依次连接的组件，可以同时为人体部件存在的可能性，以及人体部件之间的关系编码。该网络同样具有良好的精度。

5 应用领域

人体姿态估计在人机交互、智能监控、虚拟现实以及运动分析等领域有着广阔的应用前景。本文主要介绍在人机交互和智能监控领域的应用。

5.1 人机交互

人机交互指人与机器的互动和交流，目的是让机器人能理解和模仿人的语言和行为，让人类能够更有效、更自然地与机器人互动。想要实现互动自然，传统的输入输出方式是远远不够的，人与人之间的互动交流在很大程度上依赖于语音和视觉，因此人机交互势必会沿着语音和视觉交互的方向发展。

姿态估计作为计算机视觉领域的研究热点，是人工智能及机器人研究领域的一个重要议题，同时也是实现人机交互的有效手段。例如，Tang等[56]提出了基于人体构造的姿态估计算法，该算法能快速准确地识别出绝大部分基本的动作并使机器人能实时跟随人体的简单运动。基于姿态估计的人机交互也可应用于日常生活中，例如，在我们日常使用的快手、抖音和微视等视频软件中，一些动作特效就是由姿态估计技术完成的，体感游戏也依赖于姿态估计技术，文献[57]提出了智能康复的概念，引入虚拟场景人机交互技术，推出康复游戏，激励患者的主动参与意识，从而增加训练时间，提高训练

强度和训练频率，最终提高训练效果。

在姿态估计的研究中，有许多研究成果具有实时性较高的优点，将这些实时性较高的姿态估计方法应用于人机交互可提高人与机器之间交流的流畅度，并使其形式更加多样。

5.2 智能监控

人体姿态估计目前最为广泛的应用是在智能监控中。随着人们安防意识的提升以及监控技术的日益成熟，智能监控系统的应用领域在不断扩大。智能监控与普通监控的区别主要在于其将人体姿态估计技术嵌入视频服务器中，运用算法估计、判断监控画面场景中的人体姿态，提取其中的关键信息，当出现异常行为时及时向用户发出警报。

智能监控可应用于校园、监狱、家居、医院等场景，例如，文献[58]中提到将智能化监控引入校园，智能化监控系统利用人体姿态估计监测学生的心理状态，有效防止校园暴力事件的发生；文献[59]提出利用智能监控对犯罪者的日常行为及特殊情况下的行为进行监控，有效保障狱警人员、在押人员和社会的安全；文献[60]将基于图像的人体姿态估计技术应用到老年公寓的视频监控系统上，可以及时发现并处理老年人摔倒的情况，以更好地保证老年人的安全。除上述应用领域，人体姿态估计在地铁、银行、宾馆、超市、社区、停车场等诸多社会公共安全领域发挥着越来越重要的作用。

5.3 其他应用

除了人机交互和智能监控，姿态估计还可应用于体育项目、队列评分、智能驾驶、厅店零售等。例如，在体育领域，文献[9]提及了建立人工智能教练系统来帮助运动员调整专业动作，为运动员提供个性化的运动训练体验；文献[61]研发了一套人体姿态自动评分系统，对队形的整齐度进行量化评估，以便客观、精准地评价队列训练的效果，指明队列训练提升的方向；文献[62]将人体头部与身体姿态估计方法应用在智能驾驶领域，用于判断行人注意力方向以及预测行人行走的轨迹，提前做出安全预警；文献[63]打造了基于人体姿态估计的用户偏好分析系统，该系统将用于厅店零售环节，通过对用户肢体形态进行分析，判断用户对商品的偏好及情绪，辅助厅店人员进行场景化接触营销和个性化推荐。姿态估计的应用领域十分广泛，除上述已知的应用以外，还有许多潜在的领域等待人们去探索。

6 研究展望

人体姿态估计在计算机视觉领域的重要性和应用前景已得到认可，但同时阻碍其应用的因素也逐渐显现出来。本节将就人体姿态估计面临的挑战和可能的研究方向加以概述。

(1)标准数据集的扩展。虽然目前数据量已经初具规模，例如 COCO 数据约有 6 万多的图片数据，但是大部分姿态都是正常姿态，如站立、走路等，对于一些特殊姿态，如摔倒、翻越等数据较少。“负样本”数据的收集成本很高，但对于某些领域却至关重要。研究人员可以通过生成数据的方法来扩充数据种类，如利用生成对抗网络 GAN “生成”尽量真实的图片，该网络以接受噪声并通过噪声生成图片的方式，同时存在一个判别网络来辨别生成图片的真实性，两个网络形成博弈，最终生成真实度高的图片。尽管上述方法在理论上可行，但仍未见成熟的研究报道。

(2)移动设备中的姿态估计。目前最佳的用户体验莫过于仅利用移动端完成姿态估计，如无需联网的智能摄像头。然而移动端的计算能力以及存储空间有限，因此需要降低网络复杂度和计算量来适应移动端。可能的解决思路是采用基于深度可分离的网络，即把标准卷积分解成深度卷积(Depth-wise Convolution)和逐点卷积(Point-wise Convolution)，深度卷积负责滤波，逐点卷积负责转换通道，以此降低参数量和计算量，缩短网络耗时，同时保持姿态估计的精度，从而达到更好的用户体验。

(3)拥挤场景下的姿态估计。实际应用场景中经常因人多而发生拥挤，在拥挤的场景中遮挡问题会极为严重，就目前的研究而言，被遮挡部分的关键点预测的准确率还不够高。针对遮挡问题可以引入除光学图像以外的成像方法，如红外线、无线信号等。无线信号的穿透力强，可准确找到遮挡部分的人体部位。在训练的同时采集图片和无线信号，并使用基于图片的神经网络来训练基于无线信号的神经网络。但就目前的研究水平而言，尚未有成功的研究报道，采用何种非光学辅助手段，还有待探究。

(4)装备人群的姿态估计。现有多人姿态估计技术已较为成熟，人体部分或是关节的标注数据广泛存在，且估计模型精度很高，但对于一些装备人群来说效果却不佳，例如特种兵、消防员、宇航员、潜水员、滑雪运动员等，原因是他们的装备改变了人体部分和关节的形态。提高此类人群的姿态估计精度需要增加标注，但由于装备种类繁多且差异较大，因此标注困难，并且一些特定装备也不允许标注。考虑在训练阶段提取非装备人群不同姿态特征的图像，为了提高计算效率，选择尽可能小的感兴趣区域(ROI)。将训练得到的特征迁移至深度神经网络，源领域数据与目标领域数据分布相同，从而可以在新的空间中更好地利用源领域已有的有标记数据样本进行分类训练，最终对目标领域的数据进行分类测试。基于迁移学习解决无标注姿态评估问题的研究尚属空白。

(5)在三维人体姿态估计中的应用。二维人体姿态估计的许多方法也可以应用于三维人体姿态估计，例如，我们可以直接用预训练好的二维姿态估计网络预测二维身体关节位置，将得到的二维坐标输入到三维姿态估计网络中，这样可减小模型在三维姿态估计上的学习压力，使得网络结构更加简单，达到实时性好、训练快速的效果。

结束语 人体姿态估计广泛应用于计算机视觉的相关领域，人们对姿态估计的研究也从未停止过。本文对近年来的相关研究进行了系统的梳理和分析，提出了目前姿态估计方法存在的问题和挑战，从单人和多人的角度分别介绍了姿态估计的方法，并且总结了从各个方面提升性能和精度的相关模型。最后，对人体姿态估计未来的发展趋势进行了展望。

参 考 文 献

[1] HEN C H, RAMANAN D. 3D Human Pose Estimation=2D Pose Estimation+Matching[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7035-7043.

[2] LI X H，LIU J F. A Review of the Research on Two-dimensional Human Posture Estimation[J]. Modern Computer，2019(22)：33-37.

[3] FISCHLER M A，ELSCHLAGER R A. The Representation and Matching of Pictorial Structures[J]. IEEE Transactions on Computers，1973，22(1)：67-92.

[4] ANDRILUKA M，ROTH S，SCHIELE B，et al. Pictorial structures revisited：People detection and articulated pose estimation[C]// 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2009：1014-1021.

[5] DALAL N，TRIGGS B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005：886-893.

[6] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision，2004，60(2)：91-110.

[7] NAGELI T，OBERHOLZER S，PLUSS S，et al. Flycon：real-time environment-independent multi-view human pose estimation with aerial vehicles[C]// International Conference on Computer Graphics and Interactive Techniques. 2019.

[8] ACHILLES F，ICHIM A E，COSKUN H，et al. Patient MoCap：Human Pose Estimation Under Blanket Occlusion for Hospital Monitoring Applications[C]// Medical Image Computing and Computer Assisted Intervention. 2016：491-499.

[9] WANG J，QIU K，PENG H，et al. AI Coach：Deep Human Pose Estimation and Analysis for Personalized Athletic Training Assistance[C]// ACM Multimedia. 2019：2228-2230.

[10] YANG Y，RAMANAN D. Articulated pose estimation with flexible mixtures-of-parts[C]// The 24th IEEE Conference on Computer Vision and Pattern Recognition. 2011：1385-1392.

[11] HE K，ZHANG X，REN S，et al. Delving Deep into Rectifiers：Surpassing Human-Level Performance on ImageNet Classification[C]// International Conference on Computer Vision. 2015：1026-1034.

[12] GLOROT X，BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]// International Conference on Artificial Intelligence and Statistics. 2010：249-256.

[13] SAPP B，TASKAR B. MODEC：Multimodal Decomposable Models for Human Pose Estimation[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. 2013：3674-3681.

[14] GKIOXARI G，ARBELAEZ P，BOURDEV L，et al. Articulated Pose Estimation Using Discriminative Armlet Classifiers[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. 2013：3342-3349.

[15] SAPP B，JORDAN C，TASKAR B，et al. Adaptive pose priors for pictorial structures[C]// Computer Vision and Pattern Recognition. 2010：422-429.

[16] Dantone M，Gall J，Leistner C，et al. Human Pose Estimation Using Body Parts Dependent Joint Regressors[C]// Computer Vision and Pattern Recognition. 2013：3041-3048.

[17] PISHCHULIN L，ANDRILUKA M，GEHLER P V，et al. Poselet Conditioned Pictorial Structures[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. 2013：588-595.

[18] JOHNSON S，EVERINGHAM M. Learning effective human pose estimation from inaccurate annotation[C]// The 24th IEEE Conference on Computer Vision and Pattern Recognition. 2011：1465-1472.

[19] TOSHEV A，SZEGEDY C. DeepPose：Human Pose Estimation via Deep Neural Networks[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014：1653-1660.

[20] TOMPSON J，JAIN A，LECUN Y，et al. Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation[C]// Neural Information Processing Systems. 2014：1799-1807.

[21] WEI S，RAMAKRISHNA V，KANADE T，et al. Convolutional Pose Machines[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016：4724-4732.

[22] LONG J，SHELHAMER E，DARRELL T，et al. Fully convolutional networks for semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2015：3431-3440.

[23] NEWELL A，YANG K，DENG J，et al. Stacked hourglass networks for human pose estimation[C]// European Conference on Computer Vision. 2016：483-499.

[24] LONG J，SHELHAMER E，DARRELL T，et al. Fully convolutional networks for semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2015：3431-3440.

[25] ZEILER M D，TAYLOR G W，FERGUS R，et al. Adaptive deconvolutional networks for mid and high level feature learning[C]// International Conference on Computer Vision. 2011：2018-2025.

[26] PAPANDREOU G，ZHU T，KANAZAWA N，et al. Towards Accurate Multi-person Pose Estimation in the Wild[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017：3711-3719.

[27] HE K，GKIOXARI G，DOLLAR P，et al. Mask R-CNN[C]// International Conference on Computer Vision. 2017：2980-2988.

[28] FANG H，XIE S，TAI Y，et al. RMPE：Regional Multi-person Pose Estimation[C]// International Conference on Computer Vision. 2017：2353-2362.

[29] XIAO B，WU H，WEI Y，et al. Simple Baselines for Human Pose Estimation and Tracking[C]// European Conference on Computer Vision. 2018：472-487.

[30] CAO Z，SIMON T，WEI S，et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017：1302-1310.

[31] PISHCHULIN L，INSAFUTDINOV E，TANG S，et al. DeepCut：Joint Subset Partition and Labeling for Multi Person Pose Estimation[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016：4929-4937.

[32] NEWELL A，HUANG Z，DENG J，et al. Associative Embedding：End-to-End Learning for Joint Detection and Grouping[C]// Neural Information Processing Systems. 2017：2277-2287.

[33] INSAFUTDINOV E，PISHCHULIN L，ANDRES B，et al. DeeperCut：A Deeper，Stronger，and Faster Multi-Person Pose Esti-

mation Model[C]// European Conference on Computer Vision. 2016:34-50.
[34] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
[35] NING G, HE Z. Dual Path Networks for Multi-Person Human Pose Estimation[J]. arXiv:1710.10192.
[36] CAO Z, SIMON T, WEI S, et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017:1302-1310.
[37] CHEN Y, LI J, XIAO H, et al. Dual Path Networks[J]. arXiv:1707.01629.
[38] HUANG G, LIU Z, DER MAATEN L V, et al. Densely Connected Convolutional Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017:2261-2269.
[39] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated Residual Transformations for Deep Neural Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017:5987-5995.
[40] CHEN Y, WANG Z, PENG Y, et al. Cascaded Pyramid Network for Multi-person Pose Estimation[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition. 2018:7103-7112.
[41] MARTIEZ G A, VILLAMIZAR M, CANÉVET O, et al. Real-time Convolutional Networks for Depth-based Human Pose Estimation[C]// Intelligent Robots and Systems. 2018:41-47.
[42] LUO D, DU S, IKENAGA T, et al. End-to-End Feature Pyramid Network for Real-Time Multi-Person Pose Estimation[C]// International Conference on Machine Vision. 2019:1-4.
[43] ZHANG Z, TANG J, WU G, et al. Simple and Lightweight Human Pose Estimation[J]. arXiv:1911.10346.
[44] YANG W, LI S, OUYANG W, et al. Learning Feature Pyramids for Human Pose Estimation[C]// International Conference on Computer Vision. 2017:1290-1299.
[45] NIE X, FENG J, XING J, et al. Generative Partition Networks for Multi-Person Pose Estimation[J]. arXiv:1705.07422.
[46] NIE X, FENG J, XING J, et al. Pose Partition Networks for Multi-person Pose Estimation[C]// European Conference on Computer Vision. 2018:705-720.
[47] ZHAO Y, LUO Z, QUAN C, et al. Lite Hourglass Network for Multi-person Pose Estimation[C]// MultiMedia Modeling - 26th International Conference. 2020:226-238.
[48] LUO Y, XU Z, LIU P, et al. Multi-Person Pose Estimation via Multi-Layer Fractal Network and Joints Kinship Pattern[J]. IEEE Transactions on Image Processing, 2019, 28(1):142-155.
[49] LUO Y, XU Z, LIU P, et al. Combining fractal hourglass network and skeleton joints pairwise affinity for multi-person pose estimation[J]. Multimedia Tools and Applications, 2019, 78(6):7341-7363.
[50] ZHAO Y, LUO Z W, QUAN C Q, et al. Cluster-wise Learning Network for Multi-person Pose Estimation[J]. Pattern Recognition, 2020, 98(2):107074.
[51] 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. 2019:5693-5703.
[52] CHENG B, XIAO B, WANG J, et al. Bottom-up Higher-Resolution Networks for Multi-Person Pose Estimation[J]. arXiv:1908.10357.
[53] ZHANG K, HE P, YAO P, et al. DNANet: De-Normalized Attention Based Multi-Resolution Network for Human Pose Estimation[J]. arXiv:1909.05090.
[54] WANG X, CAO Z, WANG R, et al. Improving Human Pose Estimation With Self-Attention Generative Adversarial Networks[J]. IEEE Access, 2019:119668-119680.
[55] RADWAN I, MOUSTAFA N, KEATING B, et al. Hierarchical Adversarial Network for Human Pose Estimation[J]. IEEE Access, 2019:103619-103628.
[56] TANG B, FAN Q R, SUN K X, et al. Application of Human Pose Recognition Algorithm in Visual Human-computerInteraction[J]. Computer Measurement and Control, 2019, 27(7):242-247.
[57] TANG X Y, SONG A G. Human Pose Estimation and Its Application in Rehabilitation Training Situational Interaction[J]. Journal of Instrumentation, 2018, 39(11):195-203.
[58] ZENG L Z. Intelligent Campus Management System[J]. Communication World, 2018(8):309-310.
[59] SONG X Y. Research and Implementation of Abnormal Behavior Identification Technology in Prison Intelligent Monitoring System[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2013.
[60] LI W. Design of Intelligent Video Surveillance System for Elderly Apartments[D]. Huaqiao: Huaqiao University, 2017.
[61] ZHOU P X. Design and Implementation of Queue Scoring System Based on Human Pose Estimation[D]. University of Electronic Science and Technology, 2019.
[62] XIA P. Pedestrian Pose Estimation for Active Safety of Intelligent Vehicle[D]. University of Electronic Science and Technology, 2019.
[63] LIU H J, ZHOU D M. User Preference Analysis System Based on Human Pose Estimation[J]. Tianjin science and technology, 2019, 46(4):53-56.

FENG Xiao-yue, born in 2000, undergraduate. Her main research interests include big data management and machine learning.

SONG Jie, born in 1980, Ph. D, professor. His main research interests include big data management, green computing and machine learning.