

基于改进深度学习的人体姿态识别方法研究

刘宇[†]

(清华大学 信息科学技术学院, 北京 100062)

摘要:针对现有2D多人人体姿态识别方法存在的耗时长、准确率低等问题,在对人体姿态识别方法进行分析的基础上,提出了一种用于2D多人人体姿态识别的改进复合场。引入空洞卷积模块降低参数数量的同时提高模型准确性,引入shuffleNet V2网络替换主干网ResNet提高模型识别速度。通过实验对所提方法的平均精确度、平均召回率和运行时间等进行分析。结果表明,与常规方法相比,所提方法对2D多人人体姿态识别具有较高的识别准确率和速度,1~8人的人体姿态平均识别时间为75ms。为计算机视觉的研究提供了一定的参考。

关键词:人体姿态;复合场;空洞卷积模块;shuffleNet V2网络;2D多人

中图分类号:TP391

文献标识码:A

Research on Human Posture Recognition Method Based on Improved Deep Learning

LIU Yu[†]

(School of Information Science and Technology, Tsinghua University, Beijing 100062, China)

Abstract: Based on the analysis of existing 2D multi human body pose recognition methods, an improved composite field for 2D multi human body pose recognition is proposed to address the issues of long time consumption and low accuracy. Introducing a hollow convolutional module to reduce the number of parameters while improving model accuracy, and introducing shuffleNet V2 network to replace the backbone network ResNet to improve model recognition speed. Analyze the average accuracy, average recall rate, and running time of the proposed method through experiments. The results show that compared with conventional methods, the proposed method has higher recognition accuracy and speed for 2D multi person human pose recognition, with an average recognition time of 75ms for 1-8 people. This provides a certain reference for the research of computer vision.

Key words: body posture; composite field; atrous convolution module; ShuffleNet V2 network; 2D multiple persons

视觉是人类感知世界的重要手段之一,而如何让计算机实现视觉感知,对图像目标信息进行识别,已成为当下非常热门的研究领域^[1]。如何从一张图像中识别和推断其中的2D人体姿态是一项重要的基础性工作,对其进行研究具有重要的实际意义^[2]。

目前,国内外已有很多学者在研究多人人体姿态识别方法,主要集中在两个方向:2D人体姿态识

别和3D人体姿态识别,都是基于深度学习的人体姿态识别方法^[3-5]。石跃祥等^[6]针对现有人体姿态识别方法速度受人数影响和检测性能受不同尺寸影响等问题,提出了一种改进的稠密卷积网络结构。所提方法在保持较高准确率的情况下,还兼顾了较快的速度。马鸿玥等^[7]针对现有人体姿态识别方法准确率低和目标的多尺度问题,提出了一种改进R-FCN算法。所提方法可以快速、准确地

收稿日期:2023-07-14

作者简介:刘宇(1986—),男,山东蓬莱人,硕士,工程师,研究方向:计算机视觉等。

[†]通信联系人,E-mail: waqc01@sohu.com

识别人体姿态,与常规方法相比,准确率提高明显。罗梦诗等^[8]针对现有人体姿态识别方法存在的目标丢失和遮挡导致的准确率较低等问题,提出一种改进级联金字塔模型。所提方法在目前遮挡下的人体姿态识别具有较高的精度,且降低了网络复杂度和参数量。宋玉琴等^[8]为提高不同人体尺度关键点的检测精度,提出了一种高分辨率表征关键点尺度变换网络。所提方法的平均检测精度较常规方法有明显提高,且网络参数量和浮点运算量也最小。虽然上述方法都可以实现人体姿态的识别,但上述方法在人体姿态识别中存在耗时长、准确率低等问题,需要进一步提高适应性。

在此基础上,提出了一种用于人体姿态识别的改进复合场,在引入空洞卷积模块降低参数数量的同时,提高了模型的准确性,引入 ShuffleNet V2 网络取代骨干网络 ResNet,提高模型识别速度。实验验证了该方法的可行性。

1 人体姿态识别概述

人体姿态识别是对图像中人体关键点进行检测,用于分析人们的行为和动作等信息^[9]。

如图 1 人体姿势识别结构。对于图像 I ,像素集合为 Z ,则像素点 $z \in Z$,且 $z = (x, y)$, x 和 y 为坐标信息。设人体关键点坐标集为 J ,总数为 N ,不同关键点 ID 为 i ,即 $i \in 1, 2, \dots, N$,则 $J = (J_1, J_2, \dots, J_N)$,其中 J_i 为图像中的人体第 i 个关键点位置。将图像输入网络,输出人体图像中 N 个关键点的特征图^[10]。

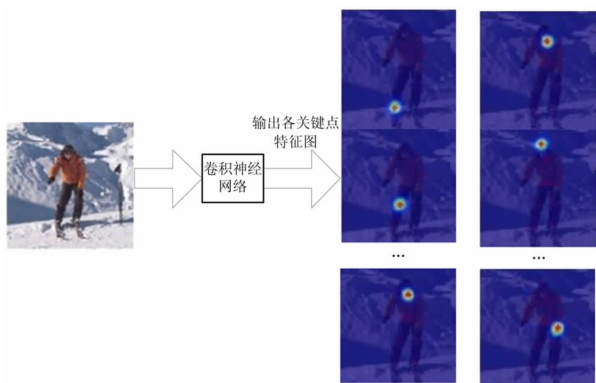


图 1 人体姿态识别结构

2 人体姿态识别

2.1 复合场人体姿态识别

复合场在人体姿态识别中的应用最早是由

Kreiss 等人^[11]提出,首先对所有关节进行检测,再通过分组算法获得不同姿态。复合场整体框图如图 2 所示,主要由编码网络和解码网络组成,编码网络是经过骨干网络,在由部分强度场(Part Intensity Field, PIF)和部分关联场(Part Association Field, PAF)子网获取 PIF 热图、PAF 热图^[12]。解码网络将热图进行集成,获取姿态识别结果。

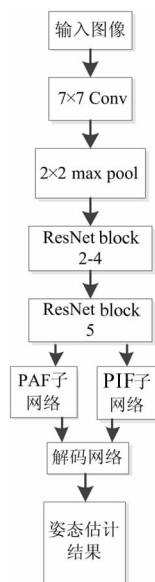


图 2 复合场整体框图

2.2 改进复合场人体姿态识别

基于复合场的人体姿态识别方法,在如遮挡、背景混淆等场景中,也存在一定的识别误差。本文通过引入空洞卷积模块,降低参数数量的同时提高模型准确性,引入 ShuffleNet V2 网络替换主干网 ResNet,提高模型预测速度^[13-14]。空洞卷积为新型卷积方式,在不增加训练参数数量的同时增加网络感受野^[15]。因此,在本方法中,为了实现网络预测精度的提高,引入空洞卷积模块^[16]。空洞卷积结构如图 3 所示。

假设输入图片大小为 225×225 ,卷积核的大小为 3×3 ,填充和步长均为 1,膨胀系数为 1 时,与普通卷积等同,即没有填充扩大普通的卷积核,而且此时的卷积核大小为 3×3 ,感受野为 9^[17]。膨胀系数设为 3,则感受野的大小为 49,权值为 0 的点不需要训练。根据式(1)和(2)可以计算出输出特征尺寸^[18]。

$$n = h + (h - 1) \times (d - 1) \quad (1)$$

$$c_{out} = \frac{c_{in} + 2p - h - (h - 1)(d - 1)}{s} + 1 \quad (2)$$

在式中, n 和 h 分别为空洞卷积填充后卷积核大小和普通卷积核大小; c_{in} 和 c_{out} 为空洞卷积输入和输出特征图的大小; p 为该层的填充值; s 为步长; d 为膨胀系数。

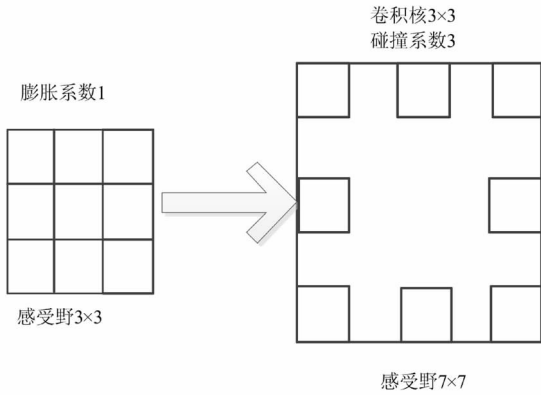


图3 空洞卷积结构

由式(1)和(2)可知,在卷积核为普通卷积核的情况下,输出特征图的大小为 225×225 ,空洞卷积核输出为 221×221 。因此,引入空洞卷积核特征图的大小基本上不会改变。网络参数量不变的情况下,几乎不丢失特征图的详细信息,同时能够增大感受野^[19]。

2.3 ShuffleNet V2 网络

ShuffleNet V2 是由旷视科技团队于 2018 年提出的轻量级网络架构,ShuffleNet V2 和 ShuffleNet V1 架构基本相同,唯一不同的是增加了 Conv5 这个 1×1 卷积层^[20-21]。为了提高复合场运行速度,将骨干网由 ResNet 网络替换为 ShuffleNet V2 网络。ShuffleNet V2 网络的基本结构如图 4 所示。

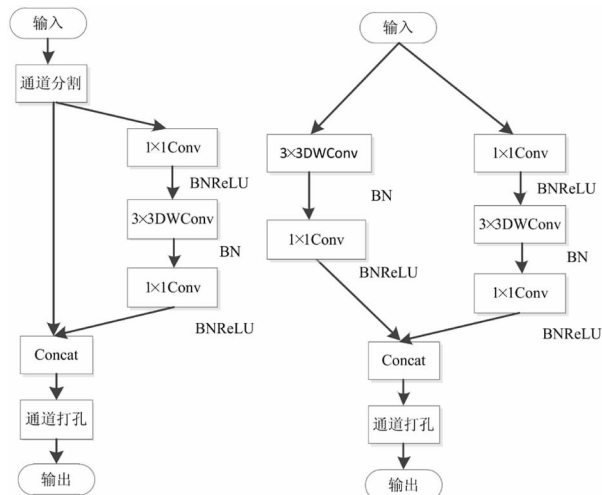


图4 ShuffleNet V2 网络两种基本结构框图

ShuffleNet V2 网络主要由卷积模块 1、最大池层、卷积模块 2-4、卷积模型 5、全局池层、全连接层等组成^[19]。

2.4 改进复合场识别网络

改进的复合场识别网络整体框图如图 5 所示,编码网络主要由卷积、骨干网络、空洞卷积模块和子网络组成^[22]。骨干网络采用 ShuffleNet V2 $\times 2$ 网络的卷积模块 2-4 完成特征提取工作,编码网络输出为热图,在通过解码网络将热图进行集成,获取姿态识别结果。

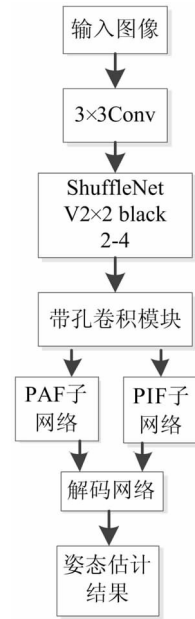


图5 改进复合场整体框图

3 实验结果与分析

3.1 实验参数

为了对文中所提方法的优越性进行验证,将文中所提方法与文献[23]方法、文献[24]方法和文献[25]进行比较分析。数据集为 MSCOCO,从准确率、召回率、速度三个方面对结果进行评估,系统参数如表 1 所示。实验参数如表 2 所示。

表1 系统参数

配置	参数
CPU	I7-13700K
GPU	NVIDIA Titan X(Pascal)12G \times 2
操作系统	Ubuntu 16.04
深度学习框架	PyTorch 1.5
Python 环境	Python 3.7.7

表 2 实验参数

参数	数值
输入图像大小	401×401
输入通道数	16
学习率	0.01
动量	0.9
批大小	16
迭代次数	150
优化器	SGN
损失函数	二元交叉熵、拉普拉斯、L1

本文选择了平均精确度、平均召回率和运行速度评价人体姿态识别方法模型。

(1)平均精确度

预测结果中实际包含的阳性样本数。只有两种可能的预测:正确预测的 TP 和错误预测的 FP。如式(3)所示。

$$P = \frac{TP}{TP + FP} \tag{3}$$

在 MSCOCO 数据集中,准确率的评价指标包含平均精度和跨尺度平均精度, AP (多个 IoU 阈值下检测的平均精确度)、AP₅₀ (IoU 阈值为 0.5 的 AP)、AP₇₅ (IoU 阈值为 0.75 的 AP)、AP_M (像素面积位于 32² ~ 96² 的检测对象 AP 值)、AP_L (像素面积大于 96² 的检测对象 AP 值)^[25]。

(2)平均召回率

所有正样本都有两种可能的预测结果。正确预测数 TP 和未预测数 FN,如式(4)所示。

$$R = \frac{TP}{TP + FN} \tag{4}$$

文中仅采用平均召回率 AR 进行评估。

(3)运行速度

除了上述指标外,速度也是评估模型性能的重要指标。本文选择处理图片耗时作为速度评价指标。

3.2 实验分析

为了验证本文所提姿态识别方法的优越性,对本文方法、文献[23]、文献[24]方法、文献[25]和复合场方法进行了实验对比分析,如表 3 所示不同方法的实验结果。

由表 3 可以看出,在人体姿态识别方法中,本文方法的准确率优于文献[23]、文献[24]、文献[25]和复合场的人体姿态识别方法。这是因为将 ShuffleNet V2×2 和空洞卷积模块引入复合场中,相比改进前准确率均有一定的提高,验证了加入空

洞卷积模块的有效性以及加入 ShuffleNet V 2×2 网络的优越性。

表 3 不同方法实验结果对比

指标	本文方法	文献[23]	文献[24]	文献[25]	复合场
AP	75.1	66.7	73.7	73.5	66.6
AP ₅₀	92.8	87.8	91.9	90.3	88.5
AP ₇₅	82.2	73.6	81.1	79.7	73.8
AP _M	72.3	63.5	70.3	69.3	64.2
AP _L	82.5	72.9	80.0	78.5	73.5
AR	81.4	72.6	79.0	78.3	73.2

对模型的运行速度进行分析,采用图片人数为 1~8 人,不同方法在不同图片人数时的预测总耗时如图 6 所示。

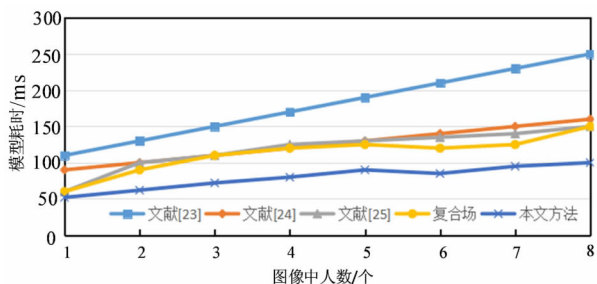


图 6 不同方法在不同图片人数时的总耗时

由图 6 可以看出,在人体姿态识别方法中,本文方法的总耗时优于文献[23]、文献[24]、文献[25]和复合场的人体姿态识别方法。本文方法性能较为稳定,在图片中人数变化时小幅波动。这是因为本文方法为自底向上方法,不需要对各人体框都进行一次识别。

3.3 结果可视化

图 7 为本文所提网络模型在分辨率较差环境下得到的部分实验结果。



图 7 本文方法在分辨率较差情况下的可视化

由图7可知,在分辨率较差的情况下,本文方法姿态识别效果较好,但也存在关键点漏检,在后续的实际应用中应不断完善和优化,以便提升本文方法在实际应用中的可行性。

图8为本文所提网络模型在复杂场景下得到的部分实验结果。



图8 本文方法在复杂场景下的可视化

由图8可知,在复杂场景下,本文方法姿态识别效果较好,这是因为网络使用了部分相关场和强度场,并引入了人体结构信息,因此在复杂场景中具有良好的姿态识别效果。

4 结论

提出了一种改进的复合场用于人体姿态识别,引入空洞卷积模块降低参数数量的同时提高模型准确性,引入 ShuffleNet V2 网络提高模型识别速度。结果表明,相比于常规方法,本文方法在准确率、召回率和运行速度上都是最优的,平均精确度为75.1%,平均召回率为81.4%,平均运行速度为75 ms左右,在复杂场景和低分辨率情况下仍具有较好的效果。但本文方法仍处于起步阶段,在复杂场景和低分辨率场景的性能还有待进一步提高。下一步将不断完善和优化本文所提人体姿态识别方法的性能。

参考文献

- [1] 赵鹏,刘杨,刘慧婷,等.基于深度卷积一递归神经网络的手绘草图识别方法[J].计算机辅助设计与图形学报,2018,30(2):217-224.
- [2] 赵莹.一种SVM的建筑施工项目安全预警系统研究[J].微型电网应用,2021,37(2):114-131.
- [3] 刘晶,钟力强,董娜.变电站巡检机器人视觉精确定位算法研究[J].工业仪表与自动化装置,2019,6(6):8-13.
- [4] 袁益琴,何国金,王桂周,等.背景差分与帧间差分相融合的遥感卫星视频运动车辆检测方法[J].中国科学院大学学报,2018,35(1):50-58.
- [5] 朱建宝,许志龙,孙玉玮,等.基于OpenPose人体姿态识别的变电站危险行为检测[J].自动化与仪表,2020,35(2):47-51.
- [6] 石跃祥,许湘麒.基于改进DenseNet网络的人体姿态估计[J].控制与决策,2021,36(5):1206-1212.
- [7] 马鸿羽,朴燕,鲁明阳.基于改进R-FCN与语义分割相结合的人体姿态估计[J].长春理工大学学报,2021,44(1):71-80.
- [8] 罗梦诗,徐杨,叶星鑫.基于轻量型高分辨率网络的被遮挡人体姿态估计[J].武汉大学学报(理学版),2021,67(5):403-410.
- [9] 宋玉琴,曾贺东,高师杰,等.基于改进高分辨表征的人体姿态估计算法[J].计算机工程与设计,2022,43(4):1045-1051.
- [10] 程禹,王晓华,王文杰,等.基于改进AKAZE算法的图像特征匹配方法[J].西安工程大学学报,2020,34(8):51-56.
- [11] KREISS S, BERTONI L, ALAHI A. Pifpaf: composite fields for human pose estimation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, 2019, 11977-11986.
- [12] 华夏,王新晴,马昭焯,等.基于递归神经网络的视频多目标检测技术[J].计算机应用研究,2020,37(2):615-620.
- [13] 高野.基于改进高斯混合模型的体育视频运动目标检测与跟踪[J].现代电子技术,2017,40(19):91-97.
- [14] 黄文婧,李华强,杨植雅,等.基于模糊聚类排序及状态均匀性的电网安全性风险评估[J].电测与仪表,2018,55(9):21-26.
- [15] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 40(4): 834-840.
- [16] 叶波.基于负载均衡度的云计算任务调度算法[J].东北电力大学学报,2019,39(1):88-95.
- [17] 谢黎,周华良,于同伟,等.一种智能变电站新型双网冗余设备及实现[J].电力系统保护与控制,2019,47(11):151-156.
- [18] 郭威强,胡立生.基于神经网络的三维重构研究[J].微型电网应用,2020,36(2):82-85.
- [19] 李成美,白宏阳,郭宏伟,等.一种改进光流法的运动目标检测及跟踪算法[J].现代电子技术,2018,39(5):249-256.
- [20] HUAN G, SONG Z, ZHANG S, et al. A fast marine sewage detection method for remote-sensing image[J]. Computational and Applied Mathematics, 2018, 37(4): 4544-4553.
- [21] 李红娟,郭向阳,刘宏建.随机动态规划和粒子群嵌套寻优的PHEV能量优化[J].机械设计与制造,2020,12(7):150-155.
- [22] KOCABAS M, KARAGOZ S, AKBAS E. Multiposenet: fast multi-person pose estimation using pose residual network[C]. Proceedings of the European Conference on Computer Vision (ECCV), Munich, 2018, 417-433.
- [23] 朱翠涛,李博.基于高分辨率网络的人体姿态估计[J].中南民族大学学报(自然科学版),2023,42(2):229-237.
- [24] 马皖宜,张德平.基于多尺度双注意力的人体姿态估计方法研究[J].计算机科学,2022,49(11):1-5.
- [25] 马赛,葛海波,何文昊,等.轻量高效的自底向上人体姿态估计算法研究[J].计算机工程与应用. <https://link.cnki.net/urlid/11.2127.TP.20230814.1802.022>.