

基于多特征融合的舞蹈动作识别技术研究

毕雪超

(西安航空职业技术学院 陕西 西安 710089)

摘要: 针对舞蹈动作存在复杂的姿态变化,导致舞蹈动作识别准确度低的问题。文中设计了一种多特征融合的舞蹈动作识别算法,该算法首先通过一个特征金字塔网络提取舞蹈动作的特征,再采用一个多特征融合模块来融合多路特征,以此提升算法对复杂姿态的估计,最终完成舞蹈动作识别。在数据集上的测试与分析结果表明,该算法可针对 Balletto 舞蹈视频数据库进行识别,且能够有效提升舞蹈动作识别准确度,从而实现对舞蹈演员的动作矫正功能。

关键词: 特征金字塔; 人体姿态估计; 多路径; 多特征融合; 动作识别

中图分类号: TP391.9

文献标识码: A

文章编号: 1674-6236(2020)18-0189-05

DOI: 10.14022/j.issn1674-6236.2020.18.042

Dance action recognition method based on multi feature fusion

BI Xue-chao

(Xi'an Vocational and Technical College of Aeronautics and Astronautics, Xi'an 710089, China)

Abstract: In view of the complex posture changes in dance movements, the accuracy of dance movement recognition is low. In this paper, we design a multi feature fusion algorithm for dance action recognition. Firstly, we extract the features of dance action through a feature pyramid network, and then use a multi feature fusion module to fuse multiple features, so as to improve the algorithm's estimation of complex posture, and finally complete the dance action recognition. The test and analysis results on the dataset show that the algorithm can identify the balletto dance video database, and can effectively improve the accuracy of dance action recognition, so as to achieve the action correction function of dancers.

Key words: feature pyramid; human posture estimation; multi-path; multi feature fusion; motion recognition

人体姿态估计是人体动作识别领域的一个关键技术,其原理是通过提取图像中的特征,识别人体姿态^[1-2]。这种技术可用于智能化舞蹈辅助训练,通过提取舞者图像的特征,获得舞者姿态骨架图。从而识别舞者舞蹈动作,对舞者姿势进行评估并矫正^[3]。

早期的人体姿态估计主要聚焦于人体轮廓特征或部件模型。如苏延超等人^[4]通过 Boosting 分类器提取边缘力场特征,设计了一种基于部件检测的人体姿态估计算法。韩贵金等人^[5]则提出一种结合直方图(Histogram of Oriented Gradients, HOG)与颜色特征的外观模型进行人体姿态估计。但由于人体姿态变化复杂,传统方法难以实现有效的姿态估计。所

以,基于深度学习的方法逐渐用于人体姿态估计。2015年,基于深度学习的人体姿态估计算法开始回归人体骨架热力图^[6]。2016年,密歇根大学的研究团队^[7]设计了一种沙漏状的神经网络结构用于提取多尺度特征,进行人体姿态估计。2017年,曹哲等人^[8]提出一种采用部分亲和域的方法去获得人体骨架图。此外,还有众多基于深度学习的人体姿态估计算法相继被提出。这些算法都可用于舞蹈动作识别,辅助舞者训练。舞者的动作变换迅速,姿态多变,人体骨架在2D图上尺度多样,这给舞者辅助训练智能化带来了挑战。

为此,文中设计了一种基于多特征融合的舞蹈动作识别算法,用于学习复杂多变的舞者动作识

收稿日期:2019-12-11 稿件编号:201912098

基金项目:陕西省高等教育工作委员会研究课题(2017FKT03);西航职院2018年度科研计划项目(18XHGZ-011)

作者简介:毕雪超(1988—),女,山东新泰人,硕士,讲师。研究方向:现代舞蹈教育信息化。

别。新模型在 Balletto 数据集上进行测试。

1 基于多特征融合的舞蹈动作识别

为了提升模型对复杂的舞者动作识别性能,本文设计了一个基于多特征融合的舞蹈动作识别算法。

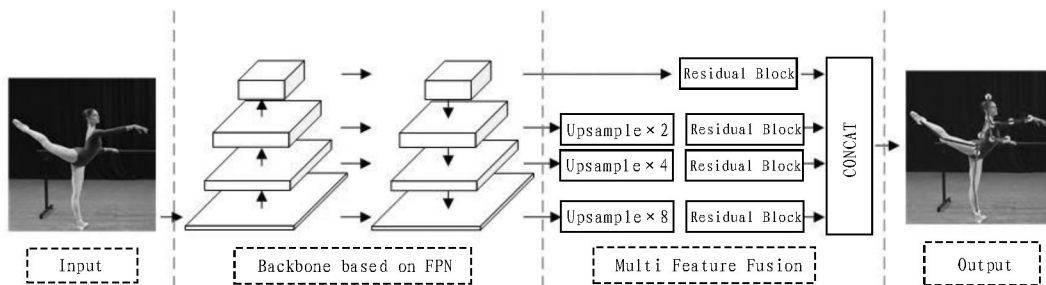


图1 基于多特征融合的舞蹈动作识别算法示意图

1.1 基于特征金字塔的主干网络

本文的主干网络与 ResNet^[11]类似,都包含5个尺度的特征提取阶段。把不同尺度卷积特征分别表示为 C1、C2、C3、C4、C5 五个阶段,每个阶段由若干个残差模块(Residual Block)组成,这些残差模块最终提取特征生成人体姿态关键点的热力图。模型的浅层特征图分辨率较大,随着网络的深入,特征图每经历一次下采样操作后,特征图缩小为原来的二分之一,分辨率也随之降低。因此,在 C1、C2、C3 的浅层特征具有较高的空间分辨率。但包含的语义信息不足,而在 C4、C5 的深层特征则相反。为了同时具有特征层的空间分辨率与语义信息,本文采用 FPN 结构以便进一步融合高底层特征信息。

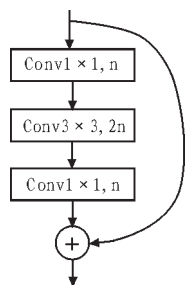


图2 残差模块示意图

基于 FPN 主干网络,如图3所示。该主干网络可以有效定位简单的可见关键点,却难以识别复杂环境的人体姿态关键点,如遮挡隐藏的关键点。对这类复杂关键点的定位通常需要更丰富的特征信息,为此文中设计了一个多特征融合模块。

1.2 多特征融合模块

基于 FPN 的主干网络用于识别简单关键点的估

新算法采用一个特征金字塔网络(Feature Pyramid Networks, FPN)^[9]进行特征提取,然后对于不同尺度特征再进行加深提取,最后将各特征上采样至原图大小进行特征融合。如图1所示,图中的 Residual Block^[10]表示残差模块。残差模块,如图2所示。

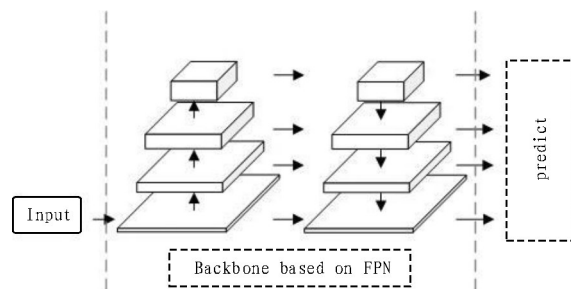


图3 基于FPN的主干网络示意图

计,多特征融合模块用于处理更为复杂关键点的估计,其结构如图4所示。为了获取更丰富的语义特征,本文采用残差模块来增强各层特征提取。同时为了获取更好的局部特征,文中通过上采样操作提升各阶段的特征分辨率。最终,来自 FPN 的各个特征进行 Concat 操作融合。

在训练过程中,FPN 提取特征并回归人体骨架关键点,简单的关键点在 FPN 阶段将基本完成。对于复杂的关键点,比如遮挡隐藏等情况,多特征融合模块会进一步加深学习来自 FPN 的各层特征并融合,最终回归人体骨架热力图。

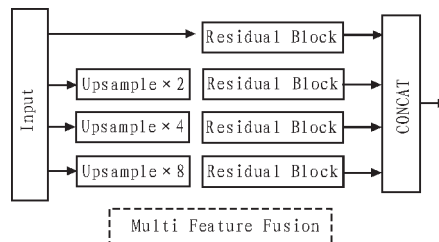


图4 多特征融合结构图

1.3 损失函数

人体姿态估计属于回归问题,回归问题中常见

的损失函数有 L1 损失函数和 L2 损失函数。本文的舞者动作识别采用回归舞者骨架关键点的方式,因此本文的算法采用 L2 范数优化欧式距离的损失函数,如式(1)所示。

$$L(\theta) = \frac{1}{2N} \sum_{i=1}^N \|F(X_i; \theta) - F_i\|_2^2 \quad (1)$$

其中, θ 表示待优化的舞者动作识别网络参数; N 是参与学习训练的舞者图片总数; X_i 表示当前学习的舞者图片样本 i ; F_i 表示第 i 幅舞者图像热力图的真实值; $F(X_i; \theta)$ 表示模型回归的舞者骨架关键点的热力图。

2 实验结果和数据分析

2.1 数据集

基于多特征融合的舞蹈动作识别算法将在 Balletto 舞蹈视频数据库上验证算法的有效性。Balletto 数据库来源于 250 个舞蹈动作视频进行分解,得到一些芭蕾舞动作的关键帧图片^[12-16]。图 5 选取部分图片作为展示。



图 5 Balletto 舞蹈视频数据库部分图片展示

2.2 模型的评价指标

本文采用 PCK (Percentage of Correct Keypoints) 对各关键点进行评估。PCK 的本质是通过设定阈值,计算人体关键点的预测值与真实值间的标准化距离小于阈值的比重,PCK 值越低算法准确度越高。

本文采用 AP、FLOPs 与 Param Size 对新算法的整体性能进行评估。其中 AP 用于评估算法的准确性,计算所有图片 OKS 分数大于设定阈值的比重。OKS 分数的计算公式,如式(2)所示。

$$OKS_p = \frac{\sum_i \{ -d_p^2 / 2S_p^2 \sigma_i^2 \} \delta(v_p i = 1)}{\sum_i \delta(v_p i = 1)} \quad (2)$$

P 表示在真实值中某个人 id; p^i 表示某个人的关键点 id; $v_p i = 1$ 表示这个关键点的可见性为 1 (即在图片上可见); S_p^2 表示这个人所占的面积大小平方根,根据真实值里人的检测框计算得到; σ_i 表示第 i 个关键点的归一化因子。

2.3 舞蹈数据库评估

基于多特征融合的舞蹈动作识别算法将在 Balletto 舞蹈视频数据库上进行测试,测试结果如图 6 所示。图 6 中(a)表示原始图像;(b)表示图中真实人体骨架关键点;(c)表示 hourglass 方法预测的人体骨架关键点;(d)表示基于 FPN 的算法预测的人体骨架关键点;(e)表示新算法预测的人体骨架关键点。从图 6 中的结果可以看出,在第 1、3、4 列简单的舞蹈动作中,3 种方法都基本能预测舞者的关键点。但是对于图 6 中第 2 列舞蹈动作,因为舞者姿态存在骨架尺度变化较大,基于 hourglass 的算法把左脚当作被右脚遮挡隐藏了,直接在右脚踝关键点处同时预测为左右脚踝。而基于 FPN 的算法显得更不理想,直接丢失了左脚踝关键点,本文算法在经过多特征融合,具备了更丰富的语义特征,成功预测出了左脚踝。但对于图 6 中第 5 列舞蹈动作,这个动作非常复杂,舞者四肢收缩,存在较大的骨架尺度变化,且各关键点存在不同程度上的遮挡。基于 hourglass 的算法与基于 FPN 的算法,因图中舞者左右腿交叉,误将左腿识别为右腿,右腿误识别为左腿。而本文算法虽然左脚踝同样没有成功预测,但是对于右腿做出了成功预测,且并没有因为左右腿交叉而出现膝盖关键点的误识别。

本文采用 PCK 指标对于新算法的各关键点进行了评估,评估结果如表 1 所示。从表 1 中可知,头部的预测中,3 种算法的 PCK 值都在 95% 以上。基于 hourglass 的算法 PCK 值最低为 95.1%,本算法 PCK 值最高达到了 96.2%。这说明,本方法在简单关键点的回归上也有一定地提高。在肩膀关键点中的预测也说明了算法对简单关键点的提高,本方法在肩膀关键点的 PCK 值相比基于 hourglass 的算法提高了 1.4%。但是对于手腕与脚踝关键点,基于 hourglass 的算法的 PCK 值分别为 79.8% 与 77.8%,基于 FPN 的算法的 PCK 值对于手腕略有提高,为 80.2%,但对于脚踝略有下降,为 77.4%。本算法加入了多特征融合模块,在加深特征提取的同时,融合了多路特征,使得提取特征语义更为丰富。本文算法对于手腕与



图6 模型的效果展示

脚踝的 PCK 值分别为 81.6% 与 78.5%，相比于 hourglass 的方法提高了 1.8% 与 0.7%，证明新算法有效提高了复杂关键点的预测。但另一方面，脚踝关键点只提高了 0.7%，说明人体在 2D 图中遮挡隐藏的关键点检测难度极大。

表1 模型各关键点准确度分析

方法	头部	肩膀	手肘	手腕	臀部	膝盖	脚踝	平均
hourglass	95.1	88.3	82.7	79.8	84.1	87.6	77.8	85.6
FPN	95.8	89.0	82.5	80.2	83.5	86.8	77.4	85.9
Our method	96.2	89.7	83.2	81.6	84.7	88.0	78.5	86.3

本文方法在 Balletto 舞蹈视频数据库上与 hourglass 等算法的实验结果对比，如表 2 所示。本文方法 AP 达到 70.3%，比 hourglass 提高了 2.4%，且新算法的 FLOPs 低于 hourglass。实验结果表明，本文方法在 Balletto 舞蹈视频数据库中性能优于 hourglass。

3 结 论

为了提升算法对复杂舞者动作的识别，文中设

表2 算法总体性能分析结果

方法	AP(OKS)	FLOPs	Param Size
hourglass	67.9	6.14G	23MB
基于 FPN	68.6	3.92G	92MB
Our method	70.3	5.62G	104MB

计了一种基于多特征融合的舞蹈动作识别算法。在 Balletto 舞蹈视频数据库的测试结果表明，该算法有效提升了模型对于遮挡、大尺度变化的舞蹈动作骨架关键点预测的准确度。在后续的研究中，将更加深入研究舞者复杂舞姿的动作识别问题，并期望建立一个性能更优良的智能舞蹈辅助训练系统。

参考文献:

- [1] Dang Q, Yin J, Wang B, et al. Deep learning based 2D human pose estimation: A survey[J]. Tsinghua Science & Technology, 2019, 24 (6): 663-676.
- [2] Andriluka M, Pishchulin L, Gehler P V, et al. 2D human pose estimation: new benchmark and state of the art analysis[C]. Boston: Computer Vision and Pattern Recognition, 2014.

- [3] 王健弘.基于视频的人体动作识别关键技术研究[D].南京:东南大学,2017.
- [4] 苏延超,艾海舟,劳世斌.图像和视频中基于部件检测器的人体姿态估计[J].电子与信息学报,2011,33(6):1413-1419.
- [5] 韩贵金,朱虹.基于HOG和颜色特征融合的人体姿态估计[J].模式识别与人工智能,2014,27(9):769-777.
- [6] Pfister T, CharlesJ, Zisserman A. Flowing ConvNets for human pose estimation in videos[J].IEEE Transactions on Computer Vision,2015(12):1913-1921.
- [7] Newell A, YangK, DengJ, et al. Stacked hourglass networks for human pose estimation[C].London: EuropeanConference on Computer Vision,2016.
- [8] Cao Z, SimonT, WeiS, et al.Realtime multi-person 2D pose estimation using part affinity fields[C].CA, USA:ComputerVision and Pattern Recognition, 2017.
- [9] Lin T, DollarP, GirshickR, et al. Feature pyramid networks for object detection[C]. CA, USA: Computer Vision and Pattern Recognition,2017.
- [10]He K, ZhangX, RenS, et al.Deep residual learning for image recognition[C].CA, USA:Computer Vision and Pattern Recognition,2017.
- [11]Xie S, Girshick R, Dollar P, et al. Aggregated residual transformations for deep neural networks [C]. CA, USA:Computer Vision and Pattern Recognition,2017.
- [12]韩丽,王露晨,张美超,等.基于特征向量匹配的舞蹈实时姿态分析方法[J].计算机应用研究,2016,33(12):3892-3896.
- [13]胡静.基于姿态的舞蹈动作识别[D].沈阳:辽宁大学,2017.
- [14]李红竹.舞蹈视频图像中动作识别方法研究[J].电视技术,2018,42(7):34-37,52.
- [15]邢占伟.基于多特征融合的舞蹈动作识别方法研究[D].沈阳:辽宁大学,2017.
- [16]唐彪,樊启润,孙开鑫,等.人体姿态识别算法在视觉人机交互中的应用[J].计算机测量与控制,2019,27(7):242-247.

(上接第184页)

- [13]雒僖,范九伦,于海燕,等.基于阴影集的截集式可能性C-均值聚类截集门限的选取[J].计算机科学,2019,46(8):249-254.
- [14]林嘉炜,祁云嵩,陈晓利,等.基于改进核可能性C均值类间极大化聚类算法[J].现代电子技术,2018,41(24):117-120,124.
- [15]李宇泊,李秦.对k均值算法和硬C-均值算法的对比分析[J].洛阳理工学院学报(自然科学版),2012,22(1):72-75,81.
- [16]储慧琳,赵博,张兴明.一种新的模糊C均值聚类算法的优化方法[J].信息工程大学学报,2011,12(3):347-351.
- [17]李明.大数据环境下无线终端智能通信信号采集系统设计[J].电子设计工程,2017,25(16):95-98,102.

(上接第188页)

- (4):528-537.
- [16]陈义,马云林.基于视觉的手势识别技术在车载主机上的应用[J].电子设计工程,2016,24(8):141-144.
- [17]文芳,康彩琴,陈立文,等.基于RGBD数据的静态手势识别[J].计算机与现代化,2018(1):74-77.
- [18]张义群,林培杰,程树英.基于FPGA的人脸实时检测与跟踪[J].有线电视技术,2016(2):33-36,42.
- [19]杨磊,隋云衡,姚立虎.结合肤色分割与手形匹配算法的静态手势检测[J].信息通信,2013(4):4-6.
- [20]梁永侦,潘斌,郭小明,等.基于LAB颜色空间的图像阴影检测与去除方法[J].计算机与现代化,2019(10):88-93.