

DOI:10.19651/j.cnki.emt.2107904

基于 OpenPose 改进的轻量化人体动作识别模型*

李一凡 袁龙健 王 瑞

(上海大学通信与信息工程学院 上海 200444)

摘要: 本文聚焦于自下而上的人体姿态估计网络 OpenPose 模型参数量大的缺点,对 OpenPose 模型的特征提取网络和预测网络分别进行改进,实现轻量化模型的目标。使用参数量更少、准确度更高的 ResNet18 网络替代了原模型中的 VGG19 网络,并且在不损失过多识别精度的前提下,以深度可分离卷积替代了预测网络中的部分卷积核,以此来降低网络结构的参数量。接着通过神经网络对人体动作进行分类,在传统的非线性网络中加入了线性模块提高了网络的记忆能力和泛化能力。结果表明,轻量化 OpenPose 模型的运行帧数比原先提高了 9%~16%,动作识别网络经过 3 000 次迭代训练后,站立、坐着、走、坐下和起立的识别精度达到了 0.877、0.835、0.793、0.815 和 0.808。最后,将整体识别网络应用于真实场景下,根据结果表明,该方法可以在嵌入式设备中正常运行,且识别效果较好。

关键词: 人体姿态估计;OpenPose;轻量化模型;深度可分离卷积;神经网络

中图分类号: TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Improved lightweight human action recognition model based on OpenPose

Li Yifan Yuan Longjian Wang Rui

(College of Communication and Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: This article focuses on the shortcomings of the bottom-up human pose estimation network OpenPose model with large parameters, and improves the feature extraction network and prediction network of the OpenPose model to achieve the goal of lightweight model. This article uses the ResNet18 network which has fewer parameters and higher accuracy to replace the VGG19 network in the original model. In order to reduce the amount of parameters of the network structure, we replace part of the convolution kernel in the prediction network with the deep separable convolution without losing too much recognition accuracy. Then, the human body actions are classified through the artificial neural network, and the linear module is added to the traditional nonlinear network to improve the memory and generalization ability of the network. The results show that the FPS of the lightweight OpenPose model has increased by 9% to 16% compared to the original. After 3 000 iterations of the network training, the recognition accuracy of standing, sitting, walking, sitting and standing up reach 0.877, 0.835, 0.793, 0.815 and 0.808, respectively. Finally, the recognition network is applied to a real scene. According to the results, it is shown that the method runs normally in embedded devices and performs well.

Keywords: human pose estimation;OpenPose;lightweight model;deep separable convolution;artificial neural network

0 引 言

近些年来,随着人工智能^[1]的飞速发展,图像模式识别技术^[2]在人们的日常生活中发挥着举足轻重的作用。人体动作识别是模式识别领域中的一个重要研究方向,可广泛应用于人机交互、辅助驾驶等场景。人体动作识别过程可以分为人体姿态估计和动作分类两个部分。

人体姿态估计^[3-4]是对给定图像序列中的人体姿态信

息进行预测。目前大量的研究旨在对图像序列中人体的各个部分进行检测,并完成人体骨骼关键点的定位和跟踪。

人体姿态估计的方法大致可以分为传统方法和基于深度学习^[5]的方法。传统的人体姿态估计方法是基于模板匹配的几何算法^[6],该方法通过多维度建模的方式建立人体各个部位的模板图像并比较这些模板与目标图像之间的特征差异,以此来判断目标图像中各个关键点之间的空间关系。近些年来随着深度学习的快速发展,开始将深度学习

收稿日期:2021-09-17

* 基金项目:国家自然科学基金(61771299)项目资助

方法应用于人体姿态估计的研究中。

人体姿态估计需要同时考虑到多组人体关键点的检测和连接问题,根据处理层次性将多人姿态估计的检测方法分为了“自上而下”的检测方法和“自下而上”的检测方法。“自上而下”的方法^[7]需要借助人体的检测器来获取与人数相当的人体框图,再以单个人体为最小单位分别进行检测,该方法的运行时间与人数成正比并且十分依赖于人体检测器的性能。因此,大部分的研究将目标集中于提升人体检测器的性能。比如, Iqbal 等^[8]提出了卷积姿态机生成初始姿态,然后应用整数线性规划来得到最终的姿态。Pishchulin 等^[9]采用基于 Fast-CNN 的人体检测器,首先检测出所有的人体候选部位,然后将每个部位标记为对应类别,并用整数线性规划将这些部位组装成完整的骨架。“自下而上”的检测方法则是先对人体骨骼关键点进行预测再拼接的过程,比如 Nie 等^[10]提出了一种姿态划分网络来进行关节检测 and 密集回归。Cao 等^[11]所提出的 OpenPose 模型,该模型使用的是一种名为部分亲和力场的方法,该方法可以对肢体的位置和方向进行编码,以这种方式对关键点之间进行正确连接。

动作分类模型是对人体姿态估计模型输出的人体关键点坐标通过人工神经网络进行深度融合并分类的网络模

型,以完成实时检测人体动作的任务。

当前的人体姿态估计模型存在参数量过大的局限,如果应用在实际场景中,则会出现模型运算耗时久、画面帧数过低等问题,因此需将检测模型轻量化以满足目前移动嵌入式终端的算力。本文在 OpenPose 模型的基础上加以改造,对原始特征提取网络进行改进并以深度可分离卷积的方式代替部分传统卷积得到轻量化模型,并在 COCO2017 数据集上验证本模型在减少计算参数的基础上还具有一定的准确性,接着通过该模型得到的 18 个关键点坐标信息建立动作分类人工神经网络,并在自建数据集中加以训练和验证。

1 基准模型结构分析

OpenPose^[11]是以卷积神经网络为基础,对其进行改进,然后在监督学习环境下所实现的实时多人人体关键点检测的网络模型。其原始网络主要架构如图 1 所示。其结构可以分为两大部分,首先通过传统卷积神经网络 VGG19^[12](前 10 层)进行特征提取得到特征图 F ,接着将其输入双分支多阶段网络,该网络的上支路用于预测部分亲和力(part affinity fields, PAF),它能够记录关键点之间的位置信息和方向信息;下支路用于预测部分置信度图(part confidence map, PCM),它的作用是表征关键点的位置。

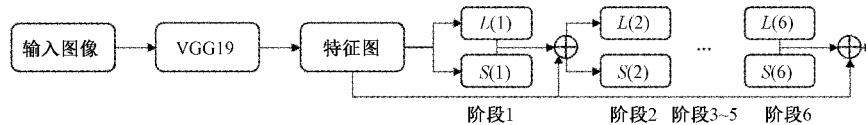


图 1 OpenPose 网络主要架构

图 2 为部分亲和力预测网络和置信度预测网络的内部

结构,该网络采用了多个阶段提取关键点之间的语义信息。

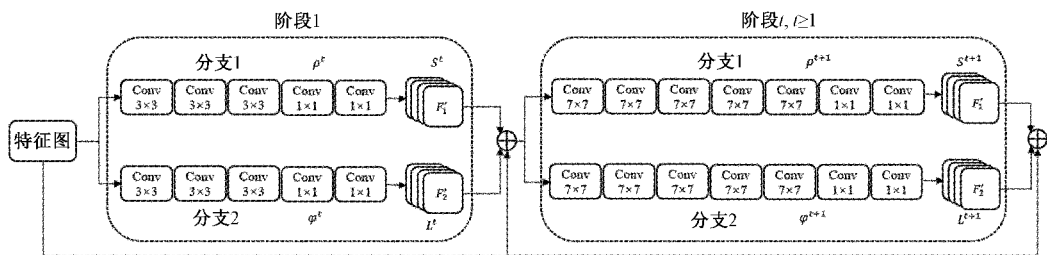


图 2 预测网络内部结构

除第 1 阶段外,其他阶段均采用了多个能够获取大感受野的 7×7 卷积核,并且在每个阶段的末尾都会连接来自两个子网络的预测值 S^l 、 L^l 和初始特征图 F 并作为下一阶段的输入,目的是更好地融合深层次特征信息,并且不会忽略浅层特征。具体表达如式(1)所示。

$$\begin{aligned} S^l &= \rho^l(F, S^{l-1}, L^{l-1}), \quad l \geq 2 \\ L^l &= \phi^l(F, S^{l-1}, L^{l-1}), \quad l \geq 2 \end{aligned} \quad (1)$$

2 算法设计

2.1 特征提取网络

OpenPose 原模型采用 VGG19 卷积神经网络进行特

征提取。深度过浅的卷积神经网络对特征提取的效果不佳,但对于 VGG 等没有残差结构的网络而言,随着深度的增加,会影响网络的收敛速度,反而会影响检测效果。因此,本文将原 VGG19 的结构更改为 ResNet18^[13] 结构如图 3 所示。与前者相比,ResNet18 的参数量更少(参数量约为 VGG19 的 1/5);其次,ResNet18 是一类拥有残差结构的深层卷积神经网络,在结构上不再是简单的卷积核堆叠,而是利用残差结构的理念解决了深层网络中梯度消失和梯度爆炸的退化问题。

ResNet18 网络分为 4 个残差层从不同尺度进行特征提取,如图 4 所示。每层由两个 BasicBlock 残差块所构成,

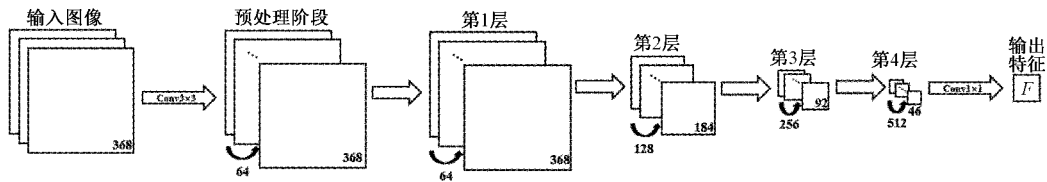


图 3 ResNet18 网络结构

在卷积神经网络深度加深的同时,使用跳级连接的方式创建两条捷径,并在卷积核之间添加了 ReLU 激活函数保证网络的非线性,在训练的过程中不断融合主路径和卷积路径的特征。考虑到原模型网络结构中的实际情况,在第 4 个残差模块后紧跟一个 1×1 的卷积核来匹配预测网络的输入尺寸。

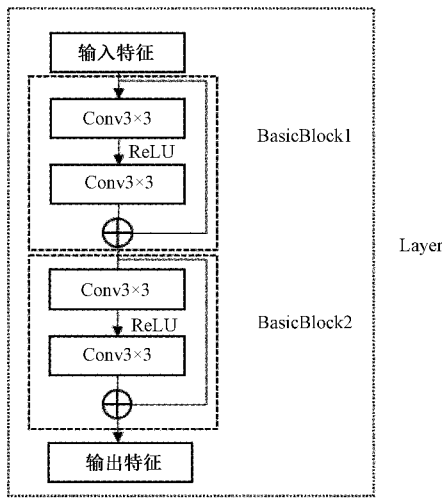


图 4 ResNet18 残差层结构

2.2 支线网络

OpenPose 内部网络结构如图 2 所示,两个预测支路由多个卷积核串联而成,并且在第 2~6 阶段的预层网络中大量使用了 7×7 卷积核。虽然大卷积核可以获得更大的感受野,但是同时也会造成很大的计算量。因此本文采用了多个小卷积核代替大卷积核的思想,如图 5 和 6 所示,使用 3 个串联的 3×3 卷积核代替 1 个 7×7 卷积核,每 3 个卷积核之间增加 1 条由 1×1 卷积核构成的跳跃连接,用于解决网络结构加深导致梯度消失的问题。

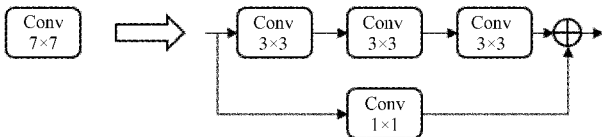


图 5 改进卷积核结构

同时为了进一步减少网络的参数量,每一个 3×3 卷积核使用的是深度可分离卷积。深度可分离卷积来源于在 2017 年提出的移动端轻量级网络 MobileNet^[14],这种卷积计算方法非常适用于移动端和嵌入式设备,它相比于传

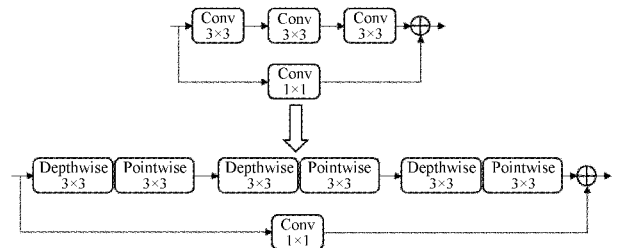


图 6 使用深度可分离卷积改进传统卷积核

统卷积运算的参数量更小,计算效率更高。

深度可分离卷积可以分为深度卷积 (depthwise convolution, DW) 和点阵卷积 (pointwise convolution, PW) 两部分,结构如图 7 所示,它通过拆分空间维度和通道维度之间的关联性,将传统卷积分为两个步骤。

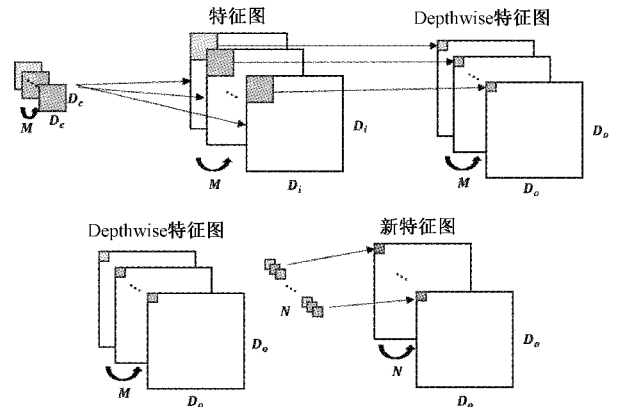


图 7 深度可分离卷积计算过程

若将输入特征图的尺寸表示为 $D_i \times D_i \times M$,卷积核的尺寸为 $D_c \times D_c \times M$,输出特征图的尺寸为 $D_o \times D_o \times N$ 。对于传统卷积运算,每个卷积核与每层输入特征图对应进行卷积,因此参数的计算公式如式(2)所示。

$$W_c = (D_c \times D_c \times M) \times N \quad (2)$$

而对于深度卷积而言,多层的卷积核分别对输入特征进行卷积,得到与输入特征数量相同的新特征图,参数量如式(3)所示。

$$W_{dw} = (D_c \times D_c \times 1) \times M \quad (3)$$

点阵卷积则是在上一步的基础上使用一定量的 1×1 卷积核对中间特征进行融合,最终得到输入特征图,参数量如式(4)所示。

$$W_{pw} = (1 \times 1 \times M) \times N \quad (4)$$

综上,传统卷积运算和深度可分离卷积的总参数量的比值如式(5)所示。

$$\mu = \frac{1}{N} + \frac{1}{D^2} \quad (5)$$

由于被替换的传统卷积核的尺寸都为 3×3 ,因此双分支预测结构的卷积运算量将会缩减为原来的 $1/9$ 。

2.3 人体动作分类算法

人体动作形式可以理解为单位时间内的姿态变化,因此本文人体动作分类算法主要是对已经识别好的人体关键点在一定时间内的 2D 坐标平面位置变化进行分类。具体的方式是使用人工神经网络(artificial neural network, ANN)^[15]对动作输出结果进行分类,输入数据为一些重要关键点在 0.5 s 内的位置变化。

人体动作主要可以分为站立、坐着、走、坐下、起立这 5 种基本状态。如果将识别所得的所有 18 个人体关键点(如图 8 所示)的位置变化全部作为输入数据,会产生许多冗余输入影响网络运行速度并且会直接影响分类效果。因为在判断这些基本动作状态时,并不需要关注到所有关键点的位置变化,例如进行动作判断时并不需要结合头部五个关键点的位置变化,鼻子和颈部关键点的可以代表头部关键点的位置变化。因此,本文舍弃了头部和手臂的部分关键点数据,使得网络更加注重于学习身体部位关键点的位置变化信息。

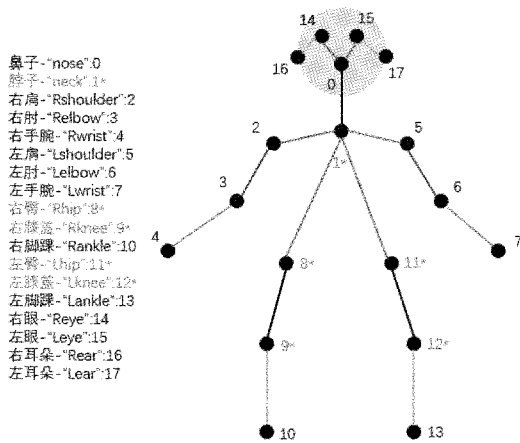


图 8 OpenPose 网络输出的人体 18 个关键点

2.4 人工神经网络

人工神经网络是一类拥有多个隐藏层和激活函数的非线性网络。因为其拥有很大的“深度”,因此也被称为深度神经网络(deep neural network, DNN)。

本文参考了 Google 的 Wide and Deep^[16]模型对人工神经网络进行了改进,在传统网络的非线性系统中加入了线性模块,具体网络结构如图 9 所示。左边部分便是线性模块,它体现了输入参数与网络输出结果之间的直接关系,这样的改进所训练出的分类网络拥有线性模块的记忆能力和非线性模块的泛化能力,并达到一定的权衡。最后

输出经过 Softmax 归一化指数函数得到 5 种基本动作状态的概率。

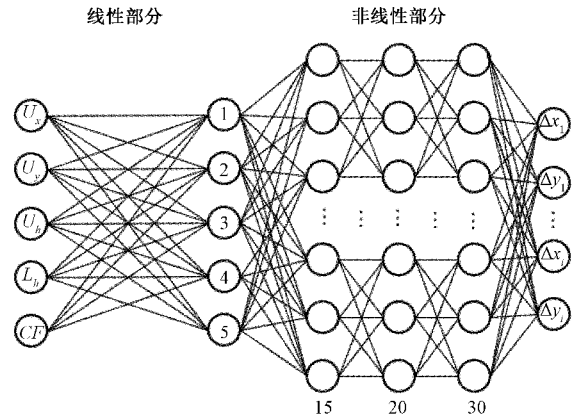


图 9 改进人工神经网络结构

非线性模块能够学习连续性数值的特征,是对相关性进行的传递。根据前文所述选取了鼻子、颈部、左右肩、左右臀部、左右膝和左右脚踝这 10 个关键点在 0.5 s 内的纵横坐标变化值作为输入数据,让网络去学习重要关键点之间的位置变化对人体输出动作的影响,因此非线性模块的输入有 20 个。本文设置了 3 个隐藏层进行非线性分离,根据隐藏层节点数量一般规则,将 3 个隐藏层分别设定为 30、20、15 个神经元,其中每个神经元上都有独立的权重值 ω 和偏置值 b ,神经元间的传递关系如式(6)所示。

$$y_j^{h+1} = \sum_i \omega_i^h x_i^h + b_j^{h+1} \quad (6)$$

式中: y_j^{h+1} 表示为第 $h+1$ 层第 j 个神经元的输出; x_i^h 表示为第 h 层的第 i 个神经元输入; ω_i^h 表示为第 h 层第 i 个神经元权重值; b_j^{h+1} 表示为第 $h+1$ 层第 j 个神经元的偏置。并且相邻两层之间有 ReLU 激活函数来保证网络的非线性。

线性模块能够学习特征的共性,是对具体特征进行学习。该模块的输入应与输出直接相关,因此需要对输入数据进行改进,用更具有“记忆化”的数据进行替代,即根据观察基本动作状态在一段连续时间内各个关键点的轨迹移动特点,根据经验值手动设计一些更匹配与所对应动作的输入融合参数。颈部、右臀和左臀 3 个关键点融合信息可以表示人体上半部分的姿态变化情况;右膝、左膝、右臀和左臀的融合信息可以表示人体下半部分的姿态变化情况。以此建立了上半部中心横坐标 U_x 、上半部中心纵坐标 U_y 、上半部分中心高度 U_h 、下半部分中心高度 L_h 和综合因子 CF 这 5 类融合输入参数,这些参数包含了重要关键点之间的进阶信息,更适合作为线性模块的输入参数,具体定义方式如式(7)~(11)所示。

$$U_x = \frac{\Delta_x[\text{neck}] + \Delta_x[\text{Rhip}] + \Delta_x[\text{Lhip}]}{3} \quad (7)$$

$$U_y = \frac{\Delta_y[\text{neck}] + \Delta_y[\text{Rhip}] + \Delta_y[\text{Lhip}]}{3} \quad (8)$$

$$U_h = \frac{[\overline{Rhip}]_y^{t_2} + [\overline{Lhip}]_y^{t_2} - [\overline{neck}]_y^{t_2}}{[\overline{Rhip}]_y^{t_1} + [\overline{Lhip}]_y^{t_1} - [\overline{neck}]_y^{t_1}} \quad (9)$$

$$L_h = \frac{[\overline{Rknee}]_y^{t_2} + [\overline{Lknee}]_y^{t_2} - [\overline{Rhip}]_y^{t_2} + [\overline{Lhip}]_y^{t_2}}{[\overline{Rknee}]_y^{t_1} + [\overline{Lknee}]_y^{t_1} - [\overline{Rhip}]_y^{t_1} + [\overline{Lhip}]_y^{t_1}} \quad (10)$$

$$CF = \frac{[\overline{Rhip}]_y^{t_2} + [\overline{Lhip}]_y^{t_2} - [\overline{neck}]_y^{t_2}}{[\overline{Rknee}]_y^{t_2} + [\overline{Lknee}]_y^{t_2} - [\overline{Rhip}]_y^{t_2} + [\overline{Lhip}]_y^{t_2}} \quad (11)$$

其中, $[key]$ 表示为关键点名称; Δ_x 表示为单位时间内的该关键点的 x 轴位置变化; Δ_y 表示为单位时间内的该关键点的 y 轴位置变化; $[\overline{key_1}]_y^{t_1} + [\overline{key_2}]_y^{t_1}$ 表示为关键点 key_1 和关键点 key_2 在 t_1 时刻 y 轴坐标的平均值; $[\overline{key}]_y^{t_2}$ 表示的是关键点 key 在 t_2 时刻的 y 轴坐标。

线性模块神经元间的传递关系如式(12)所示。

$$y_j = \omega_i^T x_i + b_j \quad (12)$$

其中, ω 为权重值第 i 个神经元权重值, b 为第 j 个神经元的偏置。

最后,线性模块和非线性模块的输出值会进行相加再最终通过 Softmax 分类函数得到各个动作状态的概率。在训练的过程中,两部分进行联合训练,共同优化参数。

2.5 损失函数

多分类问题使用交叉熵损失函数如式(13)所示。

$$Loss = - \sum_{i=1}^K y_i \log(p_i) \quad (13)$$

其中, K 为动作分类的数量; y_i 为符号函数,如果类别是 i , 则 $y_i = 1$, 否则取 0; p_i 为第 i 个神经网络的输出。

在反向传播的过程中,利用随机梯度下降(stochastic gradient descent, SGD)分别正向优化两个子模块的权重 ω 和偏置 b 。同时,非线性模块还利用了批标准化(batch normalization, BN)防止参数优化的过程中出现梯度消失和梯度爆炸的问题。

3 实验结果与分析

3.1 实验设计

本文模型训练在 Ubuntu18.04 和 Python3.6 环境下进行,采用 PyTorch1.0.0 作为深度学习框架,使用 4 块 NVIDIA 1080Ti GPU 进行数据处理;基于真实场景的实验以 Nvidia Jetson TX2 开发板为硬件基础,摄像头使用海康 HD3000 高清摄像头。

3.2 人体骨骺点检测实验结果

本文进行了人体骨骺点检测网络的实验以验证在轻量化改进下的 OpenPose 模型仍具有相对较高的准确性。网络采用微软公司的 COCO2017 数据集进行实验,该数据集分为训练、验证和测试集共 118 288 张图片,每张图片有 18 个人体骨骺关键点标注。本文分别使用随机旋转、尺度

缩放和左右翻转 3 种图像增强的方式,其中旋转角度为 $[-45^\circ, 45^\circ]$ 、尺度缩放因子为 $[0.65, 1.35]$ 以及 50% 的概率左右翻转。实验在 Ubuntu18.04 和 Python3.6 环境下进行,采用 PyTorch1.0.0 作为深度学习框架,使用 4 块 NVIDIA 1080Ti GPU 进行训练。初始学习率设定为 4×10^{-5} ,反向传播器为 Adam 优化器,最大迭代次数为 2.8×10^5 次,训练批次设置为 8。骨骼点检测网络选用平均精度均值(mean average precision, mAP)作为评价指标。表 1 所示为在 COCO2017 数据集下原 OpenPose 模型和轻量化改进后的 OpenPose 模型的实验结果,其中 AP^{50} 和 AP^{75} 分别表示阈值为 0.5 和 0.75 的关键点预测指标, AP^M 和 AP^L 分别表示人体尺寸较小和较大情况下的预测指标。

表 1 改进 OpenPose 网络测评结果 %

模型	AP	AP ⁵⁰	AP ⁷⁵	AP ^M	AP ^L
OpenPose	60.5	83.4	66.4	55.1	68.7
本文	57.9	80.4	63.1	51.7	67.9

由表 1 可知,改进后的轻量化 OpenPose 模型的预测指标与原模型相比略有下降,幅度约为 3%。但对于检测较大尺寸人体时,轻量化模型的 AP^L 值与原模型相差不大,这是由于本文采用的是 3×3 小卷积核代替 7×7 大卷积核,小卷积核在特征提取的过程中可以对大尺寸物体进行更多次的非线性映射,因此在这一方面表现较好。

图 10 为改进 OpenPose 模型在真实环境中的应用效果图。图 11 所示为终端运行最初 30 s 内的视频帧数变化折线图,由图可知视频帧数较为稳定,且从原来的 1.2 每秒传输帧数(frames per second, FPS)提升至 1.3~1.4 FPS,提升幅度达到 9%~16%。



图 10 人体关键点检测应用效果图

3.3 人体动作分类实验结果

人体动作分类网络采用自建数据集进行训练,该数据集包括 5 种基本动作标签在 0.5 s 内关键骨骺点之间的位置变化,初始包含了 1 500 组数据,通过检测图片组旋转 $\pm 30^\circ$ 将数据集扩充至 4 500 组数据,选择其中的 900 组数据作为测试集,剩余的 3 600 组数据作为训练集。神经网络的初始学习率设定为 0.01,迭代次数为 3 000,学习

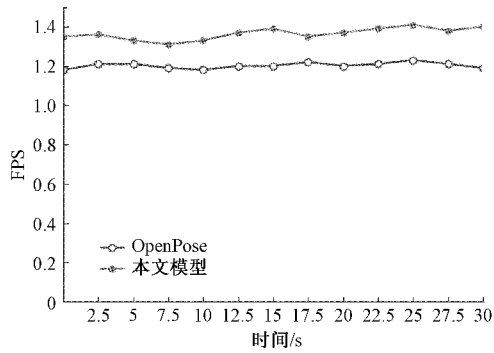


图 11 视频帧数对比图

率以 0.9 为基数每 100 轮衰减。表 2 所示为人体动作识别网络在迭代 100、500、1 000、2 000 和 3 000 次时 5 种动作分类的测试精度。

表 2 人体动作识别网络测评结果

动作分类	100	500	1 000	2 000	3 000
站立	0.143	0.428	0.617	0.808	0.877
坐着	0.137	0.411	0.583	0.753	0.835
走	0.128	0.377	0.554	0.732	0.793
坐下	0.122	0.397	0.564	0.746	0.815
起立	0.131	0.408	0.568	0.743	0.808

由表 2 可知,训练过程中网络的准确度都有一定的提升,且随着迭代次数的增加,准确度的增量也随之减少。其中,站立和坐着动作识别精度是最高的,而走和起立的识别精度最低,根据查阅数据,可能是由于 OpenPose 网络对于膝盖、脚踝部位的检测精度较低所引起。图 12 为动作识别效果,图像均来自于摄像头的实时捕捉,每组图像的时间差为 1 s 以内。结果显示,本文的方法能够及时检测动作的改变,达到实时判断基本动作状态的目的。



(a) 站姿转变为走姿



(b) 坐姿转变为起立

图 12 人体动作识别网络应用效果图

本文将轻量化深度模型和神经网络相结合,既减少了整体模型的计算量,同时也对骨骼点检测和动作分类拥有较高的识别精度。与当前其他方法相比,本文的识别算法弥补了当前深度卷积动作识别网络所需计算量大、硬件成本较高等缺点,所设计的网络结构更适用于当前的移动式嵌入终端。

4 结 论

本文在原有 OpenPose 模型的基础上进行了轻量化改进,将主干特征提取网络 VGG19 改为参数量更小、特征提取效果更好的 ResNet18 网络,并且使用深度可分离卷积对 OpenPose 模型中的预测网络进行了改进,在不损失过多识别精度的前提下缩减了模型计算量。接着对神经网络进行改进,在非线性网络中加入了线性模块,使得整体网络模型同时拥有记忆能力和泛化能力能够更好对人体动作进行学习,得到一个能够识别 5 种基本人体动作的分类网络。最后进行了基于真实场景的实验,改进后的骨骼点检测网络和人体动作分类网络拥有较高的识别精度并且能够在嵌入式设备中正常运行。

本文研究不足之处在于目前仅能识别 5 种基本人体动作,接下来进一步的研究应探寻更优秀的轻量化骨骼点检测网络以及更多关键点之间与特定动作的联系,并扩展数据集向种类更多和数据量更丰富的方向发展。

参 考 文 献

- [1] 王斌. 基于人工智能技术在高校实验室建设中应用与研究[J]. 国外电子测量技术, 2020, 39(7): 33-37.
- [2] 苏树智, 谢军, 平听瑞, 等. 图强化典型相关分析及在图像识别中的应用[J]. 电子与信息学报, 2021, 43(11): 3342-3349.
- [3] 王发明, 李建微, 陈思喜. 三维人体姿态估计研究综述[J]. 计算机工程与应用, 2021, 57(10): 26-38.
- [4] 韩孝天, 万旺根. 基于 MSP 的多人姿态估计算法[J]. 电子测量技术, 2019, 42(19): 79-84.
- [5] 邓益依, 罗健欣, 金凤林. 基于深度学习的人体姿态估计方法综述[J]. 计算机工程与应用, 2019, 55(19): 22-42.
- [6] 汪检兵, 李俊. 基于 OpenPose-slim 模型的人体骨骼关键点检测方法[J]. 计算机应用, 2019, 39(12): 3503-3509.
- [7] 张小娜, 吴庆涛. 基于深度学习的自顶向下人体姿态估计算法[J]. 电子测量技术, 2021, 44(9): 105-109.
- [8] IQBAL U, GALL J. Multi-person pose estimation with local joint-to-person associations[C]. European Conference on Computer Vision(ECCV) Workshops, Crowd Understanding, Springer International Publishing, 2016: 627-642.
- [9] PISHCHULIN L. DeepCut: Joint subset partition and

- labeling for multi person pose estimation[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 4929-4937.
- [10] NIE X, FENG J, XING J, et al. Pose partition networks for multi-person pose estimation [C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 684-699.
- [11] CAO Z, SIMON T, WEI S, et al. Realtime multi-person 2D pose estimation using part affinity fields[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 1302-1310.
- [12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. International Conference on Learning Representations, 2015: 1-14.
- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016:770-778.
- [14] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. ArXiv, 2017, ArXiv:1704. 04861.
- [15] 李舵,董超群,司品超,等. 神经网络验证和测试技术研究综述[J/OL]. 计算机工程与应用,2021: 1-16[2021-09-13].
- [16] SHI S, MIN Z, LU H, et al. Wide & deep learning in job recommendation: An empirical study [C]. Asia Information Retrieval Symposium, 2017:112-124.

作者简介

李一凡, 硕士研究生, 主要研究方向为智能信息处理。

E-mail: 19721377lyf@shu.edu.cn

袁龙健, 硕士研究生, 主要研究方向为嵌入式物联网、智能信息处理。

E-mail: yuanmochen@shu.edu.cn

王瑞, 教授, 主要研究方向为通信信号处理、传感器网络、几何代数等。

E-mail: rwang@shu.edu.cn