文章编号:2096-1472(2024)04-0033-05

基于关键点检测的坐姿识别方法

徐寅哲,屠佳佳,李 洲,史伟民

(浙江理工大学机械工程学院,浙江 杭州 310018) ⊠ brgzz@qq.com; tjji23@163.com; 2521170001@qq.com; swm@zstu.edu.cn



摘 要:针对直接应用深度学习分类算法所得模型泛用性较差的问题,提出了一种分两个步骤完成人体坐姿识 别的方法。所提方法首先提取图片中人体上身关键点的坐标信息,在关键点检测环节采用高分辨率主干网络,并进 一步改进了模型结构,在下采样环节中引入 SE(Squeeze-and-Excitation)注意力机制,加强了空间位置特征的表达, 取得了更高的检测平均精准度;然后采用随机森林算法对关键点进行坐姿分类。实验结果表明:所提方法识别准确 率可以达到 94%以上,并且在陌生场景下有更好的泛用性,能适应实际应用中复杂的人物环境。

关键词:坐姿识别;高分辨率网络;人体关键点检测;随机森林算法 文献标志码:A

中图分类号:TP391.41



Sitting Posture Recognition Based on Keypoints Detection

XU Yinzhe, TU Jiajia, LI Zhou, SHI Weimin

(College of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China) ⊠ brgzz@qq.com; tjji23@163.com; 2521170001@qq.com; swm@zstu.edu.cn

Abstract: Aiming at the problem of poor generalizability of the model obtained from the direct application of deep learning classification algorithm, this paper proposes a two-step method to complete the human sitting posture recognition. In the proposed method, firstly the coordinate information of the keypoints of the upper human body in the picture is extracted, a high resolution backbone network in the keypoints detection link is adopted, and the model structure is furtherly improved. The Squeeze-and-Excitation (SE) attention mechanism is introduced in the downsampling link to strengthen the expression of spatial positional features, achieving a higher detection average accuracy. Then the Random Forest algorithm is used to classify the keypoints for sitting posture. Experimental results show that the recognition accuracy of the proposed model can reach over 94%, and it has better generalization in unfamiliar scenarios, which better adapts to environments with complex characters in practical applications.

Key words: sitting posture recognition; high resolution network; human keypoints detection; Random Forest algorithm

0 引言(Introduction)

机器视觉技术的成熟发展,使其快速应用于人们生活的方 方面面。将机器视觉技术与物联网技术相结合,是目前在个人 或家庭应用场景下的一个趋势。在人们的日常学习、工作与生 活中,坐姿是人体最主要的姿态之一,为了避免长期不正确的 坐姿给人体带来的健康问题,研究人员开始在智能家具中部署

坐姿提醒功能。

在采用视觉技术的识别方法中,叶启朗等[1]提出了一种基 于人体骨架连接关系的坐姿识别方法,通过正面的骨架图像, 使用残差网络模型进行坐姿识别分类。房志远等[2]采用特征 融合的方式将人体骨骼特征用于卷积神经网络训练,但识别准 确率不高。在图像识别分类研究中,通常直接采用深度学习网

络模型进行分类训练。但是,对人体坐姿的分类与对物体的分 类不同,更多关注的是姿态的变化。此外,生活场景复杂多样, 但是采集的图像数据是有限的,不可能覆盖所有环境,因此背 景和人物的变化对网络模型识别的准确率有较大的影响,导致 模型泛用性较差。针对坐姿图像识别方法中存在的问题,本文 提出一种人体坐姿识别方法,它分两步完成对人体不同坐姿行 为的细致区分,具有准确率高、泛用性强,以及不受环境和人物 变化影响等优点。

1 方法结构和理论分析(Algorithm structure and theoretical analysis)

1.1 方法总体结构

本文提出的坐姿识别方法的总体结构设计如图 1 所示,它 主要由两个部分组成。首先,图像经过深度学习训练的网络模 型获得人体关键点信息。相机的拍摄取景是人体正面坐姿,因 此关键点选取眼、耳、鼻及肩膀 4 处位置,即输出 7 通道的矩 阵,每一个通道包含一个关键点的坐标位置。其次,在得到 7 个关键点坐标后,以此作为输入参数通过机器学习训练的算法 模型预测坐姿的分类识别结果。这里设置正面、低头、倾斜、侧 身等多个坐姿分类结果。





Fig. 1 Overall design of sitting posture recognition methods

根据算法的整体结构可知,通过先识别关键点再分类的方法,已经将图像信息进行了一次预处理,相当于提取了与坐案高度相关的特征,再将这些特征做进一步分类,可以规避直接训练分类模型带来的低泛用性的问题。影响最终的坐姿识别结果的因素主要来自两个方面:一是人体关键点的提取,二是对关键点的分类。下文将着重对这两个方面进行研究及验证。

1.2 主干网络

对图像进行深度学习训练的本质在于.通过像素的特征提 取图片中所包含的特定信息,而关键点法测的目的在于准确得 到事物在图像中的位置信息。该过程包括两个任务:一是对不 同关键点的图像进行分类,二是对不同的关键点进行定位^[3]。 在常见的分类网络中,通常要经过下采样提取特征,随着卷积 神经网络层数的增加,特征层的维度逐渐增大,但特征图的宽 高尺寸随之缩小。这样的网络结构虽然能增强语义信息,有利 于分类任务的进行,但是丢失了大量的空间位置信息,不利于 关键点的准确定位。因此,针对关键点检测的应用场景,主干 网络应当能始终保持较高的分辨率,并通过与低分辨率特征层 进行融合,达到既有利于分类任务的进行,又不会丢失位置信 息的效果,由此可以提高关键点检测的准确性。

高分辨率网络^[4]采用了具有多个分支的网络结构,其得名 于始终有一条分支保持高分辨率特征层,而其他分支不断做下 采样,缩小了特征层的尺寸大小,增强了网络的特征表达能力。 多个分支并行,相互之间进行多尺度的特征融合,最终汇聚到 一起,主干网络结构示意图如图 2 所示。

这种通过并行连接的方式构建的网络对不同分辨率尺度 的特征进行了融合,使其能够在增强语义特征的同时保留空间





Fig. 2 Diagram of backbone network

信息。凭借这种结构,高分辨率主干网络可以实现准确的关键 点分类以及所在位置的检测。本文在关键点检测部分以此作 为主干网络,训练适用于坐姿场景的神经网络模型。

1.3 模型训练

将人体图片输入主干网络后,通过一个卷积核大小为1, 卷积核个数为7的卷积层,得到7通道的特征层,其中每一层 都是针对每一个关键点的热力图(Heat Map)^[5]。网络模型最 后输出的热力图的分辨率是原图的1/4。其中,热力图上最大 值所在的位置就是关键点的预测结果,将该位置映射回原图 中,就能得出该关键点检测的坐标值。

图 3 为关键点的热力图示意。对于每个关键点,以关键点 坐标为中心施加一个一维的高斯分布,其高斯分布热力图可以 表示为

$$= \exp\left[-\left(\frac{(x-a_i)^2}{2\lambda^2} + \frac{(y-b_i)^2}{2\lambda^2}\right)\right] \qquad (1)$$

其中: a, φ, 分别是关键点 p, 在图中所在的坐标位置, λ 为控制分布扩散大小的常数。用高斯分布对每个关键点坐标进行处理的目的是更好地进行网络收敛,如果只采用一个点作为正 减本,那么其他点都是负样本,导致正负样本比例不均,两者差 距悬殊,网络就会出现难以收敛的情况。将结合高斯分布后得 出的热力图与模型预测输出的热力图进行对比,可以计算均方 误差损失(MSE)。用到的损失函数为

$$loss = \frac{1}{KH_wH_h} \sum_{i}^{K} \left[G_i^* \left(x, y \right) - G_i^{gt} \left(x, y \right) \right]^2 \qquad (2)$$

其中:G_i^{*}(x,y)为网络模型预测得出的热力图,G^u_i(x,y)为数 据集中标注的关键点坐标应用了高斯分布得到的热力图。K、 H_w、H_h分别为关键点的个数和热力图的宽、高。计算出损失值 后,通过反向传播不断迭代权重参数,使得预测结果与标注信息 相接近。在使用训练好的网络模型进行预测时,可以选取模型 输出的热力图中最大值的所在位置求出相应关键点的坐标值。



图 3 关键点的热力图示意 Fig. 3 Schematic heat map of key points

1.4 机器学习分类方法

完成对人体关键点的检测后,就要通过几个关键点坐标得 出坐姿的分类结果。不同于图像输入的大量像素点信息,该分 类任务的参数变量较少,只有7个坐标点,即14个参数。出于 模型轻量化和分类计算速度的考虑,这里选择使用机器学习方 法对坐姿进行分类。用于机器学习分类的方法有许多种,根据 坐姿坐标的数据形式进一步筛选出K-近邻算法、随机森林算 法、BP神经网络等分类模型。这些模型都是有监督学习算法, 适用于学习已经有明确的类型标注的数据集。

随机森林算法是由大量的决策树构成,并在决策树的训练 过程中引入属性选择的随机性。在进行分类任务时,放入训练 样本后,森林中的每一个决策树都会独立完成学习,在对测试 样本进行预测时,每一个决策树得到了各自的分类结果,统计 这些输出结果中出现最多的类别,随机森林模型就以此作为最 终的输出。随机森林算法具有训练简单、快速,能判断特征重 要性,以及抗过拟合能力较强等优点,适用于维度较高、特征较 多的数据集。

K-近邻算法没有显式的学习过程,在训练阶段只是把训练 样本保存下来,当收到测试样本时,再进行处理和计算。特征 空间中的两个实例点的间距就是其相似程度的反映,当测试样 本的点位与某一类的训练样本的距离最接近时,就将其归为一 类,得到分类结果。K-近邻算法具有模型结构简单、无须训练 等优点,适用于样本容量较大的分类场景。

两个实例点的间距计算公式如下:

$$L_{p}(x_{i}, x_{j}) = \left(\sum_{i=1}^{n} |x_{i}^{(l)} - x_{j}^{(l)}|^{p}\right)^{\frac{1}{p}}$$
(3)

图 4 为 BP 神经网络的结构示意图。神经网络由输入层、隐 藏层、输出层 3 个部分组成。输入样本数据后,每个特征值乘以 相应的权重,不断向前传播,计算出结果后,与标注的样本结果 进行对比,根据误差进行反向传播,修正权重参数,然后重新计 算输出,以上步骤循环往复,最终得出与预期相符的结果,完成 映射模型的训练,并得到相应的权重文件。BP 神经网络有很强 的非线性表达能力,适用于复杂特征场景下的求解问题。



图 4 BP 神经网络的结构示意图 Fig. 4 Diagram of BP neural network structure

2 模型改进与对比实验(Model improvement and comparison)

2.1 引入注意力机制

注意力机制是从人类视觉及认知过程中得到启发的,被应 用于神经网络模型中^[6]。在之前的网络模型中,卷积、池化等 操作都着眼于提取图片中显著的特征而忽略了随意的线索,而 注意力机制的加入使得这个过程有了一个可以学习的参数,该 方法能使网络能够有偏向性地关注特征图中的重要信息,选择 性地将注意力集中在某一些输入数据上,从而达到提高模型预 测准确率和泛用性的效果。 HU等^[7]提出的 SE 注意力机制着重关注通道间的关系, 是一种经典的通道注意力方法,其通过自适应地学习不同通道 间的注意力权重,提高卷积神经网络对特征的表达能力。SE 注意力机制模块结构如图 5 所示,从图 5 中可以看出有 3 个阶 段,在 Squeeze 阶段是通过平均池化方法将特征图的高和宽降 为 1,每个通道仅用一个数表示,即压缩为一个特征向量,其计 算函数为

$$\boldsymbol{z}_{c} = \boldsymbol{F}_{sq}(\boldsymbol{u}_{c}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{u}_{c}(i,j)$$
(4)

其中:H和W分别表示高和宽,u。表示每个通道的特征矩阵。 在Excitation阶段通过叠加使用全连接层和非线性激活函数, 学习生成权重信息,其计算函数为

 $s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2\delta(W_1z))$ (5) 其中:W 权重就是注意力机制中的可学习参数,用于体现模型 所需的特征相关性;向量 s 为不同通道的权重信息。在 Scale 阶段,根据上一阶段生成的权重向量 s,对最初的特征图进行 权重计算,得到新的特征图作为输出。SE 注意力机制模块不 会改变特征图的尺寸大小,而是自适应地根据需要,加权调整 特征图每个通道的相应权重。该方法可以帮助网络有选择地 关注重要的特征通道,起到改进模型和提高其准确率的作用。



图 5 SE 注意力机制模块结构 Diagram of SE attention module structure

在高分辨率网络中存在多个分支并行计算,在多个分辨率 特征进行融合时,由于分辨率的尺度不同,因此需要进行上采 样或下采样等操作。在下采样环节,要根据目标尺度叠加多个 苍积核大小为3、步距为2的卷积计算,不断缩小特征图的分辨 率,在到达目标尺寸前的最后一个卷积中,除了继续缩小分辨 率外,还要将特征图通道扩大相应倍数,便于进行特征融合。 在此过程中引入SE通道注意力机制,改进后下采样模块示意 图如图6所示,通过引入SE通道注意力机制调整特征图中的 通道贡献权重,使得高分辨率分支在下采样后,在通道数扩增 的基础上加强重要特征的表达,以便在后续特征融合环节促进 信息的有效整合,将高分辨率分支所包含的空间位置特征信息 更显著地融入进来。



图 6 改进后下采样模块示意图

Fig. 6 Diagram of the improved downsampling module

2.2 关键点检测

在训练数据集方面,为了提高关键点检测模型的泛用性,本 实验采用 COCO 数据集^[8]进行模型训练,数据集涵盖了大量带 有人体关键点标注的样本实例,是目前主流的人体姿态估计数 据集之一。在正面坐姿检测场景中只需用到 7 个点位坐标,因 此本实验在 COCO 数据集的基础上进行了裁剪,仅选取了具有 上身关键点信息的样本。网络的训练基于 Ubuntu 操作系统以 及 Pytorch 深度学习框架。本实验采用高分辨率网络、本文改进 网络、残差网络^[9]以及改进残差网络(ResNeXt)^[10]4种主干网络 分别进行模型训练以及相应的验证,并进行检测效果对比。

在关键点任务中,一般采用 OKS(Object Keypoint Similarity)表示预测点坐标和真实点坐标之间的相似程度,其 值为 0~1,值越接近 1,表示模型预测的准确率越高。OKS 的 计算公式为

$$OKS = \frac{\sum_{i} \exp(-d_{i}^{2}/2s^{2}k_{i}^{2})\delta(v_{i} > 0)}{\sum_{i} \delta(v_{i} > 0)}$$
(6)

其中:*i* 表示第几个关键点, *δ* 表示关键点是否可见, *d_i* 表示预测 点和真实点之间的欧式距离, *s* 为目标尺度因子, *k_i* 为衰减常 数。对验证集样本的关键点相似度进行统计分析, 计算其平均 精确率和平均召回率, 并以此作为评判模型优劣的标准。

表1为不同主干网络在 COCO 验证集上的准确率表现,其 中 COCO 验证集仅选取了具有上半身关键点的部分样本,表中 计算了不同 IoU(Intersection over Union)下的平均精确率和平均 召回率。结果显示,相较于改进残差网络,高分辨率网络在平均 精确率和平均召回率上都有更好的表现,其提升幅度约2百分 点。而本文所提改进模型在高分辨率网络的基础上,进一步提 升了关键点的识别能力,在平均精确率和平均召回率上都有更 好的表现。

表1	不同主干网络在 COCO 验证集上的准确率表现	
Tab.1 Accu	uracy of different backbone networks on COCO validation set	t

主动的	准确率					
土丁四沿	AP	AP ⁵⁰	AP ⁷⁵	AP ^M	APL	AR
$\mathrm{HRNet}^{[4]}$	0.881 671	0.979 664	0.948 516	0.866 072	0.917 554	0.907 442
改进网络	0.884 138	0.970 019	0.958 936	0.867 571	0.920739	0.908 910
$\operatorname{ResNet}^{[9]}$	0.852 549	0.969232	0.947 408	0.830 866	0.893180	• 879 245
$\operatorname{ResNeXt}^{[10]}$	0.863 026	0.969231	0.945 927	0.842 217	0.903 855	0.890 566

此外,为了验证训练得到的模型在真实坐姿场最下的准确 率,除了使用 COCO 数据集中的验证集样本外,还采 7 500 张不同的人物和背景的正面坐姿图片,对其中的关键点进行标 注,作为测试集样本在完成训练后对模型做进 -步的准确率验 证。表2为不同主干网络在自建测试集上的 确率表现,模型 的平均精确率和平均召回率普遍可以及 到80%以上,可见训 练所得模型部署用于正面坐姿的场景中也有较好的检测效果。 此外,高分辨率网络相较于其他网络的精确率更高,与残差网 络相比,在精确度和召回率方面都有6%以上的提升,说明该 算法更加适用于本项目的应用场景。在使用改进高分辨率网 络训练的模型进行预测时,相比于原模型,准确率提升了1%。 图 7 为改进模型预测结果可视化,从图 7 中可以看出,模型准 确地标注出了人体上身7个正面坐姿关键点的坐标位置,并且 模型在多种复杂场景中的预测准确率表现均证明其具有良好 的检测效果。

表2	不同主干网络在自建测试集上的准确率表现	
àb.2 Accu	acy of different backbone networks on self-built test	set

十二团级	准确率				
土丁四组一	AP	AP^{50}	AP ⁷⁵	AR	
HRNet ^[4]	0.893 217	0.979 792	0.929 314	0.903 378	
改进网络	0.903 839	0.978 115	0.927 750	0.915 270	
$\operatorname{ResNet}^{[9]}$	0.788 879	0.958 926	0.857773	0.804 730	
ResNeXt ^[10]	0.836 024	0.967 813	0.887 298	0.847 568	



图 7 改进模型预测结果可视化

Fig. 7 Visualization of the improved model prediction results

2.3 机器学习分类

在分类实验中,将坐姿图片分成6组,分别为正面、低头、 左倾斜、右倾斜、左侧身、右侧身,每张图片以人体上身7个关 键点坐标作为输入数据,共14个特征。将6组图片全部汇总 后,按1:4的比例随机划分训练集和验证集,分别通过K-近 邻算法、随机森林算法、BP神经网络等分类算法训练模型,并 验证其分类准确率。此外,除了直接用像素坐标作为输入,还 增加了一组像素坐标值归一化后的样本,以相同的方法进行训 练,作为对比组。

如图 8 所示为不同分类算法的准确率。图 8 中的数值为 10 次训练获得的准确率的平均值。结果显示,3 种分类算法都 能完成坐姿分类任务且具有比较高的分类准确率,都能达到 95%以上。其中,随机态林算法的分类准确率最高,K-近邻算法 次之。此外,数据归一化操作对于 BP 神经网络训练有较大的提 升作用,但对于 K-近邻算法反而有不利的影响。





Fig. 8 Accuracy of different classification algorithms

由于随机森林算法的分类准确率最高,因此选用该算法模型进行坐姿分类识别,并在此做进一步的分析。图 9 为随机森林分类模型混淆矩阵。混淆矩阵作为一项评判分类模型效果的指标,可以直观地分析出各个分类项具体的识别情况。从图 9 中可以看出,在各项坐姿类别中,模型都有比较高的分类准确率。



图 9 随机森林分类模型混淆矩阵

Fig. 9 Confusion matrix of Random Forest classification model

2.4 实验结果的对比与分析

将关键点检测和机器学习分类算法结合,组成坐姿识别模型,将其与其他常见的图像分类网络模型进行对比。为了验证本文所提方法相较于采用深度学习做分类的方法在泛用性上的优势,设计了有针对性的实验。选取 VGG 网络^[11]、ResNet 残差网络、ShuffleNet 网络^[12]等经典的网络模型为例,对坐姿图像进行训练和测试。除了直接使用这些通用的分类模型外,还增添了一组文献[13]中提到的识别方案,在训练前对图像进行人物前景提取。为了验证模型的泛用性,准备了两份验证集,一份验证集是与训练集一起从上文的数据集中随机按比例划分出来的,由于人物和背景重复,因此这一份验证集对于模型来说是比较熟悉的。另一份验证集采集了与之前的背景和人物不同的图像,这些陌生场景都是之前训练数据集中从未出现过的。

表 3 为场景变化对不同模型准确率的影响,从表中可以看 出,对于在训练中已经学习过的场景和人物,各个模型都有比 较高的分类准确率,但当背景和人物发生变化时,模型的分类 准确率就出现了显著的下降,文献[13]中的方法虽然去除了背 景环境带来的干扰,但是在人物发生变化时,分类准确率仍受 到了较大影响。而本文提出的两步式分类模型基本不受场景 变化的影响,始终能保持较高的分类准确率,具有较好的泛用 性。可见,直接应用深度学习进行图像分类的方法在固定场景 下能得到较高的分类准确率,但是当模型应用于陌生场景时, 就难以保证分类的准确率,这对于机器视觉产品应用落地是不 利的,而本文提出的两步式分类模型能够更好地满足多变环境 中的坐姿识别需要。

表3 场景变化对不同模型准确率的影响

Tab.3 The impact of scene changes on the accuracy of different model

不同场景 准确率	本文 模型	VGG ^[11]	ResNet ^[10]	ShuffleNet ^[12]	文献方法 ^[13]
熟悉场景	0.972	0.959	0.979	0.966	0.946
陌生场景	0.942	0.683	0.733	0.767	♦ 0.832

3 结论(Conclusion)

本文提出了一种将深度学习和机器学习相结合的坐姿识 别方法,将识别任务划分成关键点检测和机器学习分类两个步 骤。第一步是基于多尺度分支的高分辨率主干网络,通过引入 SE 注意力机制改进模型,利用 COCO 数据集中大量的人体姿 态样本训练出高精确度的人体上身 7 个关键点的检测模型。 第二步是以检测到的关键点坐标作为特征值,采用随机森林算 法训练分类模型,并对比研究了多种分类算法在此应用场景中 的分类效果,最终得到了一种两步式的坐姿识别模型。在检测 部分的实验结果表明,改进主干网络后的模型相较于原始模 型,在平均精确率上提升了1百分点,人体上身关键点检测的 平均精确率可达90%。在分类部分的实验结果表明,两步式 分类模型的坐姿分类识别准确率可以达到94%以上,并且相 较于直接应用图像分类深度学习模型,具有更好的泛用性,更 适用于复杂生活场景。

参考文献(References)

[1] 叶启朗,李戴薪,南海. 一种基于人体骨架的任意角度坐姿 识别方法[J]. 计算机应用研究,2023,40(11):3509-3514.

- [2] 房志远,石守东,郑佳罄,等. 基于神经网络的骨骼特征融合 下坐姿快速识别[J]. 传感技术学报,2022,35(5):613-620.
- [3] 梁桥康,吴樾. 基于 HRNet 的轻量化人体姿态估计网络 [J]. 湖南大学学报(自然科学版),2023,50(2):112-121.
- [4] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation [C]//IEEE. Proceedings of the IEEE: 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 5686-5696.
- [5] BULAT A, TZIMIROPOULOS G. Human pose estimation via convolutional part heatmap regression[C]//LEIBE B, MATAS J, SEBE N, et al. Computer Vision -ECCV 2016:14th European Conference on Computer Vision EC-CV 2016. Cham:Springer,2016:717-732.
- [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[DB/OL]. (2023-08-02)[2023-10-24]. https://arxiv.org/abs/1706.03762.
- HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation networks C // IEEE. Proceedings of the IEEE: 2018
 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2018:7132-7141.
- [8] LIN TAMAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[DB/OL]. (2015-02-21)[2023-10-24]. https://arxiv.org/abs/1405.0312.
- P HE & M,ZHANG X Y,REN S Q, et al. Deep residual learning for image recognition [C] // IEEE. Proceedings of the IEEE:2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway:IEEE,2016:770-778.
- [10] XIE S N, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks [C]// IEEE. Proceedings of the IEEE: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 5987-5995.
- [11] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[DB/OL]. (2015-08-10)[2023-10-24]. https://arxiv.org/abs/1409.1556.
- [12] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet; an extremely efficient convolutional neural network for mobile devices[C]//IEEE. Proceedings of the IEEE; 2018 IEEE/ CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2018; 6848-6856.
- [13] 黄旭. 基于判别式深度学习的坐姿视觉识别方法研究[D]. 长沙:湖南大学,2018.

作者简介:

- 徐寅哲(1999-),男,硕士生。研究领域:机器学习,嵌入式 技术。
- 屠佳佳(1987-),男,博士生。研究领域:纺织装备智能化。
- 李 洲(1998-),男,硕士生。研究领域:嵌入式技术。
- 史伟民(1965-),男,博士,教授。研究领域:纺织装备机电控制 技术。