

基于 OpenPose 模型与 ELM 的人体行为检测

张 青 章成旭 刘湘政 汪 洋

(吉首大学 信息科学与工程学院, 湖南 吉首 416000)

摘 要: 本文使用基于深度学习的 OpenPose 模型结合极限学习机 (ELM) 算法, 对视频或图像中人体行为进行分析与识别, 首先利用二维人体姿态估计模型获得人体的关节点坐标数据, 进而对获取的骨架关节点数据进行处理, 再使用 ELM 对数据进行分类, 从而识别出当前的行为动作。在实验中使用自制数据集进行测试, 算法的准确率达到 90.41%。

关键词: 行为识别; 人体姿态估计; ELM; OpenPose

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1003-9767 (2020) 24-040-04

Human Behavior Detection Based on Openpose Model and ELM

ZHANG Qing, ZHANG Chengxu, LIU Xiangzheng, WANG Yang

(School of Information Science and Engineering, Jishou University, Jishou Hunan 416000, China)

Abstract: In this paper, we use OpenPose model based on deep learning combined with extreme learning machine (ELM) algorithm to analyze and recognize human behavior in video or image. Firstly, we use two-dimensional human posture estimation model to obtain human joint point coordinate data, and then process the obtained skeleton joint point data, and then use elm to classify the data, so as to identify the current behavior Make. In the experiment, the self-made data set is used to test, and the accuracy of the algorithm reaches 90.41%.

Keywords: behavior recognition; human pose estimation; ELM; OpenPose

0 引言

随着人工智能技术的快速发展, 计算机视觉技术在智能检测方面应用也越来越多。针对人体的姿态、行为识别等研究方向逐渐成为计算机视觉领域内的研究重点, 吸引大量学者与工程人员的广泛关注。

行为识别主要是对视频或图像中人的行为进行分析与识别, 判断检测后的行为是否出现设定的行为^[1-2]。比如击打、使用工具伤人等危险行为, 当算法在视频与数据的分析中检测到相对应的动作时, 则会发出相应的信息。传统的行为识别方法主要包括模板匹配法^[3]、状态空间法^[4]等方法。但是这些方法普遍存在的问题是准确率低, 识别可靠性差等问题, 难以在巨大数据量的情况下进行快速检测。深度学习方法依赖数据中的特征信息, 会影响模型的检测精度^[5]。

人体姿势估计是计算机视觉研究领域备受关注的重大问题, 是了解图像和视频人物行为的关键步骤。人体姿势估计分为单人姿态估计与多人姿态估计, 它被定义为图像或

视频中人体关节的定位问题。在多人姿态估计中, 由于无法预先得知图像中每个人的位置和总人数, 因此多人姿态估计比单人姿态估计更困难。在多人姿态估计中, 分为自顶向下 (top-down) 与 (bottom-up) 自底向上的检测方法。自顶向下的方法是首先检测出图像中的多个人, 再对每个人分别进行姿态估计, 这类方法的计算准确率较高, 但整体模型的运行速度慢, 其中代表性的算法有 CPN^[6]、RMPE^[7]、G-RMI^[8] 等。**自底向上**的方法是先检测出图像中的所有关节点, 再分析关节点的归属, 最后将所有关节点按照顺序拼接成人体的骨架。这类方法的优势在于运行速度较快, 模型几乎可以做到实时处理, 其中代表性的算法有 DeeperCut^[9]、OpenPose^[10]、Multi PoseNet^[11]。

本文采取人体姿态估计与特征分类相结合的方法, 利用 OpenPose 模型与 ELM 算法相结合的方法进行检测, 首先运行 OpenPose 模型获取人体关节点的坐标数据, 并作为极限学习机的输入进行训练学习, 在完成后输出检测到的行为识别结果。

基金项目: 吉首大学 2019 年度研究生校级科研项目 (项目编号: Jdy19045)。

作者简介: 张青 (1995—), 男, 甘肃兰州人, 硕士研究生在读。研究方向: 计算机视觉。

1 人体姿态信息的提取

OpenPose 模型由美国卡耐基梅隆大学的研究人员提出,是基于卷积神经网络和监督学习开发的二维人体姿态估计的一种算法。该算法采用了自底向上的获取方法,关节点结构图如图 1 所示。OpenPose 模型在实时检测中不会受到画面中人数、人的位置与比例以及人与人肢体互动的影响,具有很强的鲁棒性。

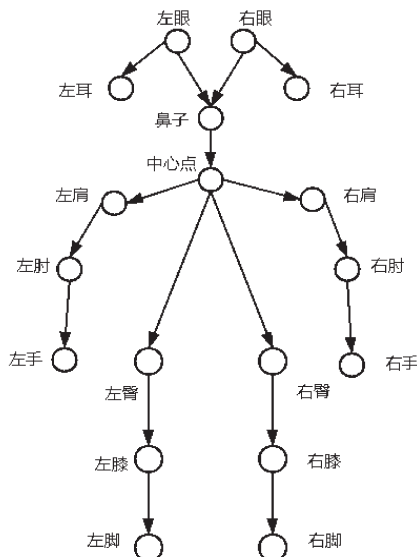


图 1 人体姿态关节点分布情况

OpenPose 的主体网络结构采用 VGG 卷积神经网络作为骨架,采用两个分支子网络分别回归关节点位置 S 和像素点在骨架中的走向 L ;两个分支网络的结构是多阶段迭代的,每个阶段计算一次损失函数,然后将 L 、 S 以及通过 VGG 卷积神经网络提取的原始图像特征连结后继续下一阶段的训练。

局部亲和场算法的作用是计算候选关键点部位相互连接组成候选肢体的置信度,如图 2 所示。设 j_1 、 j_2 是两个不同的关键部位, v 为从 j_1 指向 j_2 的单位向量, v_T 为向量 v 的垂直向量,肢体 c 的长度和宽度分别为 l_c 、 σ , X_{j_1} 、 X_{j_2} 分别表示候选关键部位 j_1 、 j_2 的位置坐标。

首先,如式 (1) 所示,计算图像上某一点 p 在肢体 c 上的置信向量 $A_c(p)$ 。如果点 p 在肢体 c 的单位方向向量 v ,否则为零向量。式 (2) 和式 (3) 为判断点 p 是否在肢体 c 上的公式,若均成立则点 p 在肢体 c 上,否则不存在。

$$A_c(p) = \begin{cases} v, & p \text{ 在肢体上} \\ 0, & p \text{ 不在肢体上} \end{cases} \quad (1)$$

$$0 \leq v \cdot (p - X_{j_1}) \leq l_c \quad (2)$$

$$|v_T \cdot (p - X_{j_1})| \leq \sigma_c \quad (3)$$

之后,可用式 (4) 得出肢体 c 的置信度 E_c ,它是通过计算关键部位 j_1 、 j_2 所在连线上的点置信向量定积分得来的。

$$E_c = \int_0^1 A_c(p(u)) \frac{X_{j_2} - X_{j_1}}{\|X_{j_2} - X_{j_1}\|} du \quad (4)$$

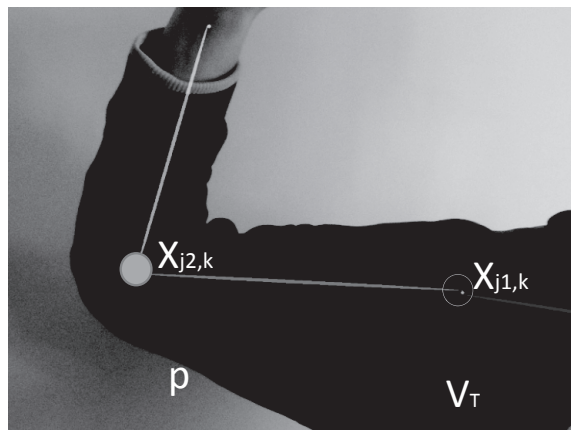


图 2 关节点间的连接

2 姿态分类器—极限学习机

极限学习机 (Extreme Learning Machine, ELM) 是由新加坡南洋理工大学的 Huang 等人提出的一种针对单隐层前向神经网络 (Single-Hidden Layer Feedforward Network, SLFN) 的新算法^[12]。不同于传统的神经网络训练算法,ELM 算法对输入层的权值和偏置进行随机赋值,再利用 Moore-Penrose 广义逆矩阵的方法直接解出隐层到输出层的权值。ELM 算法的优势是:需要手动设置的参数只有隐含层节点个数,算法执行过程中不需要人工调整参数;避免了传统训练算法反复迭代的过程,快速收敛,极大地减少了训练时间,并且所得解为唯一的最优解。一个典型的 SLFN 由输入层、隐含层和输出层构成,如图 3 所示。

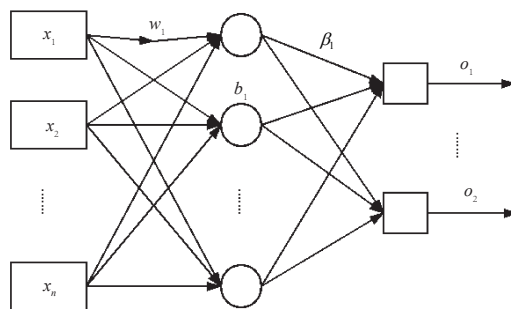


图 3 极限学习机网络结构

其模型可以描述为:给定一个训练集 $S = \{(x_i, t_i) | x_i \in R_n, t_i \in R_m, i=1, \dots, N\}$, 激活函数 $g(x)$ 和隐含层节点个数 N 。算法的实现步骤如下:①确定隐含层个数,随机设置输入层到隐含层的权值 w_i 和隐含层的偏置 b_i 的数值;②选择一个隐含层神经元激活函数,并计算隐含层的输出矩阵 H ;③计算隐含层到输出层的权值 β , $\beta = H^+ T$, 其中 H^+ 是矩阵 H 的 Moore-Penrose 广义逆矩阵。

3 实验结果分析

3.1 数据分析与处理

本文共收集侧身看屏幕、背对摄像机、弯腰、板书、低

头 5 个类别的动作图片,每类图片利用照相机采集 100 张图片,其中训练集与测试的比例为 7 : 3。每类图片均为使用照相机从正面采集,共 340 张图片。利用 OpenPose 模型对上述 5 类行为图片中的人体关节坐标的位置信息进行提取,其中模型会提取出 18 个特征点。所有关节点的分别为鼻子、中心点、左肩、左肘、左手腕、右肩、右肘、右手腕、左臀、左膝、左脚腕、右臀、右膝、右脚腕、左眼、左耳、右眼和右耳,如图 4 所示。这些关节点的坐标信息在提取完毕后保存在 csv 格式的文件中。

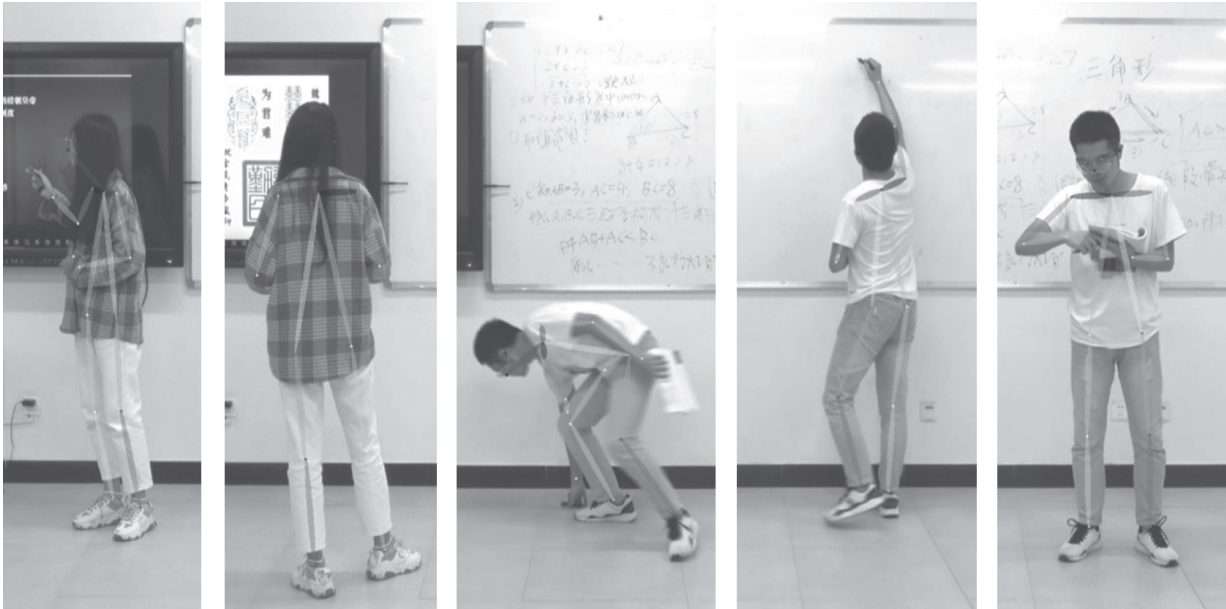


图 4a 侧身看屏幕 图 4b 背对摄像机 图 4c 弯腰图 4d 板书图 4e 低头

图 4 5 个类别的行为图片

3.3 结果分析

在现有的机器学习分类器算法中,基于传统的方法较为成熟,并大量应用于分类与识别领域。本文选择较为典型的支持向量机(Support Vector Machine, SVM)、决策树(Decision Tree, DT)、反向传播神经网络(Back Propagation Neural Network, BPNN) 和 K 近 邻(K-NEarest Neighbors, KNN) 4 种分类算法进行对比。

利用算法对每个行为类别 30 个共计 150 个图像进行实际测试,且迭代次数为 100,并记录各个模型在分类时的运行时间。各个算法的测试精度结果如表 2 所示,运行时间结果如表 3 所示。

表 2 迭代 100 次后各个算法的平均精度对比	
分类器	测试精度 /%
支持向量机 (SVM)	75.01
决策树 (DT)	86.46
反向传播神经网络 (BPNN)	84.88
K 近邻 (KNN)	77.65
极限学习机 (ELM)	90.41

3.2 模型测试

训练与测试的环境配置如表 1 所示。

表 1 训练与测试的环境配置	
软件硬件实验平台	具体参数型号
操作系统	Windows 10 专业版
机器学习框架	Sklearn
实验软件	Pycharm
CPU	Intel Core i7-6700k 3.40GHZ
GPU	NVIDIA GEFORCE GTX1080TI

表 3 迭代 100 次后各个算法的运行时间对比	
分类器	运行时间 / 秒
支持向量机 (SVM)	0.199 0
决策树 (DT)	0.104 9
反向传播神经网络 (BPNN)	102.78
K 近邻 (KNN)	0.150 1
极限学习机 (ELM)	1.812 5

其中,反向传播神经网络的参数:输入层节点个数为 20,隐藏层节点个数为 20,输出层节点个数为 20,学习率为 0.01,最大迭代次数为 5 000 次。支持向量机的参数:核函数为高斯模型分类器 (rbf), 惩罚参数值为 10, gamma 值为 30。决策树的参数: criterion 为 gini 系数, splitter 为 best, max_depth 为 2, min_samples_split 为 2, min_samples_leaf 为 1, min_weight_fraction_leaf 为 0。KNN 的 参 数: n_neighbors 为 5, weights 为 uniform, algorithm 为 auto, leaf_size 值为 30。

经过实验对比可以发现,极限学习机的测试精度优于 BP 神经网络,支持向量机,表明极限学习机对高维数据时表现的处理性能较好. 同时在相同的迭代次数下,极限学习机的运行时间也低于 BP 神经网络的运行时间。

4 结 语

本文提出了一种基于 OpenPose 模型与 ELM 算法的人体异常行为识别方法,通过 OpenPose 模型获取人体关节点数据,将提取后的特征放入 ELM 中进行分类识别。通过训练与测试,取得了较好的精度,克服了传统行为识别算法的不稳定、运行速度低等不足。同时,运行速度也优于传统模型度,能够对数量较大的图片集进行准确、快速、稳定的行为识别检测。

参考文献

- [1]Chao M W, Lin C H, Assa J, et al. Human Motion Retrieval from Hand-Drawn Sketch[J]. IEEE Transactions on Visualization & Computer Graphics, 2012, 18(5):729-740.
- [2]Wang X.Intelligent multi-camera video surveillance: A review[J].Pattern Recognition Letters,2013,34(1):3-19.
- [3] 贾贺林. 基于 DTW 的人体行为模式识别算法 [J]. 电子技术与软件工程,2019(5):61.
- [4] 吴琦. 人体行为识别算法研究与实现 [D]. 南京:南京财经大学,2018.
- [5]IJJINA E P,MOHAN C K.Hybrid deep neural network model for human action recognition[J].Applied Soft Computing,2016(46):936-952.
- [6]Chen Y,Wang Z,Peng Y,et al.Cascaded pyramid network for multi-person pose estimation[C]//Proceedings of the IEEEConference on Computer Vision and Pattern Recognition,2018.
- [7]Fang H S,Xie S,Tai Y W,et al.Rmpe:Regional multi-person pose estimation[C]//Proceedings of the IEEE International Conference on Computer Vision,2017.
- [8]Papandreou G,Zhu T,Kanazawa N,et al.Towards accurate multi-person pose estimation in the wild[C]// Proceedings of the IEEEConference on Computer Vision and Pattern Recognition,2017.
- [9]Insafutdinov E,Pishchulin L,Andres B,et al.Deepercut:A deeper,stronger,and faster multiperson pose estimation model[C]//European Conference on Computer Vision,2016.
- [10]Cao Z,Simon T,Wei S E,et al.Realtime multiperson 2d pose estimation using part affinity fields[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017.
- [11]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,2018.
- [12]Huang C B,Zhu Q Y,Siew C K.Extreme learning machine:a new learning scheme of feedforward neural network[C]//Proceedings of 2004 IEEE International Joint Conference on Neural Networks,2004.