

OpenPose 结合 BP 网络的行为检测方法

周德良*

(北京中电易达科技有限公司,北京 100190)

摘 要:采用 OpenPose 与 BP 网络相结合的方法对人体整体或局部行为进行分类检测,首先利用人体姿态估计算法获得人体骨架节点坐标数据,然后利用 BP 分类网络对节点坐标数据进行迭代训练与学习。检测不同整体行为的分类模型,训练准确率达 100%,网络损失仅为 0.091,实测各类准确率及总体准确率均达 100%;检测局部行为的分类模型,训练准确率亦达 100%,网络损失小于 10^{-6} ,实测各类准确率及总体准确率均达 100%;OpenPose 与 BP 网络相结合的方法不仅可以实现不同整体行为或局部行为快速、准确的分类检测任务,同时还克服了传统行为检测方法的不足,能够实现更高效、更准确、更快速的分类检测。

关键词:BP 神经网络;OpenPose;行为识别;隐藏层
中图分类号:TP391.4 **文献标识码:** A

随着人体行为识别技术和计算机视觉的快速发展,行为检测和识别^[1-3]在人机交互、视频监控和基于内容的视频检索等领域变得越来越重要,已引起了学术界和工业界的广泛关注,并成为计算机视觉领域的重要研究课题。行为检测主要是对视频或视频流中人体行为模式进行分析与识别,并分析判断是否发生了指定的目标行为,比如站立、坐着、躺卧等,当检测到指定目标行为后则执行后续相关操作。传统的行为检测方法主要包括模板匹配法、状态空间法以及基于模型^[4]的方法,这些方法普遍存在识别准确率低、识别可靠性差等问题,难以满足对海量图片进行准确、高效的行为检测的需求。本文采用“姿态估计”与“分类网络”相结合的方法,即利用 OpenPose^[5-8]与 BP 分类网络^[9-13]相结合的方法对视频或视频流中的目标行为进行识别:首先,基于 OpenPose 提取图像中的人体骨架节点坐标数据;再次,将骨架节点坐标数据作为 BP 分类网络的输入数据进行迭代训练与学习;最后,利用分类检测模型输出行为检测结果。

1 BP 神经网络

BP 神经网络^[11-13]是误差反向传播神经网络的

简称,是一种按照误差反向传播算法训练的多层前馈神经网络,其以网络实际输出值与期望输出值之间的误差均方差为目标函数,利用梯度下降法使目标函数最小以实现网络最优。BP 神经网络的结构如图 1 所示,包括输入层、隐藏层、输出层^[14],其中隐藏层可以包括一层或多层,BP 神经网络不存在层与层之间的局部连接,而是全部保持全连接状态。BP 神经网络具有很强的非线性映射能力,可以对任一种非线性输入输出关系进行逼近,在解决一些非线性问题方面非常突出,已广泛应用在分类识别、模式识别、函数逼近、回归预测等领域。

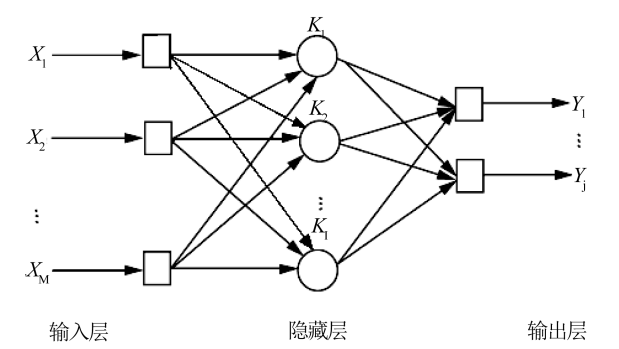


图 1 BP 神经网络拓扑结构图
Fig.1 Topology of BP neural network

BP 神经网络的训练流程:输入信号经输入层输入,通过隐含层的复杂计算由输出层得到网络输出,以网络输出与期望输出之间的误差均方差为目标函数,若目标函数不满足期望要求,则将误差信号通过误差反向传播算法,由输出层向输入层逐层反传并分摊给各层所有单元,并根据各单元误差信号计算梯度来更新各单元权值及相关参数,参数更新后输入信号再次由输入层输入网络,重复上述过程,直到目标函数满足期望要求,或迭代过程达到预先设定的训练次数为止。

2 OpenPose

OpenPose 算法^[5-8]是由美国卡耐基梅隆大学(CMU)基于卷积神经网络^[15-16]和监督学习并以caffe^[17-18]为框架开发的人体姿态估计算法,该算法是自下而上的估计算法,即先得到关节点位置再获得人体骨架。OpenPose 是第一个用于多人二维姿态检测的开源实时系统,它可以获得人体骨架节点并将它们很好地关联起来,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,同时,画面中人数增加不会影响 OpenPose 的检测实时性。OpenPose 可以实现二维多人关键点实时识别,包括 18 个身体关键点的识别、70 个面部关键点的识别、以及 42 个左右手关键点的识别,图 2 所示为 18 个身体关键点的识别效果图。因此,OpenPose 可以实现面部表情、躯干、四肢及手指的姿态估计,适用于单人和多人实时检测,且具有很好的鲁棒性。



图 2 身体关键点的识别效果图

Fig.2 Recognition effect of body key points

OpenPose 的主体网络结构采用 VGGNet 网络^[19]作为骨架,然后采用两个分支网络分别回归

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 VGGNet 提取的原始图像特征连结后继续下一阶段的训练。图 3 所示为 OpenPose 的网络结构图,其中 F 是通过 VGGNet 提取出的原始图像的特征,下标 1 和 2 分别表示多阶段迭代中的第一阶段和第二阶段。

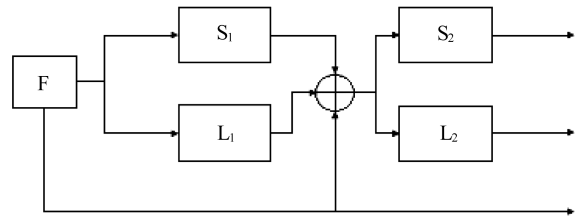


图 3 OpenPose 网络结构图

Fig.3 OpenPose network architecture

3 整体行为分析

3.1 数据采集与处理

本文共采集站立、坐着、躺着共 3 类整体行为图片,每类行为图片均利用 usb 摄像头采集 1 250 张,其中 1 000 张用于网络训练,余下 250 张用于网络实际测试。每类行为图片均利用摄像头从正面进行采集,且采集过程中四肢、躯干保持自然状态,无抬腿、摸头、倾斜等有意状态。利用 OpenPose 对上述 3 类行为图片中人体骨架节点的坐标位置信息进行提取,由此可得到躯干、四肢、头部等部位共 18 个骨架节点(包括鼻子、脖子、右肩、右肘、右腕、左肩、左肘、左腕、右髋、右膝、右踝、左髋、左膝、左踝、右眼、左眼、右耳、左耳)的位置信息,并保存在 json 文件内。

3.2 特征提取与学习

站立、坐着、躺着这 3 类整体行为主要与左右臂、左右腿和脖子共 13 个骨架节点有关,因此仅从 json 文件内提取这 13 个骨架节点的坐标位置信息作为 BP 网络的输入数据进行迭代训练与学习,然后利用训练所得模型对实际行为数据进行检测,从而得到最终的结果分类与输出,具体流程如图 4 所示。

13 个骨架节点坐标记为 $(P_1, P_2, \dots, P_{13})$, 将其展成一个骨架向量 $(P_{x1}, P_{y1}, P_{x2}, P_{y2}, \dots, P_{x13}, P_{y13})$ 作为 BP 神经网络的输入。输入数据既包含行为数据的点特征,同时包含点与点之间的线特征,以及线与线之间的面特征,这些特征通过 BP



图 4 算法流程图

Fig.4 Algorithm flow chart

分类网络进行提取与学习。本文设计的 BP 分类网络共有 3 层,即一个输入层、一个隐藏层和一个输出层,其中输入层输入训练数据,其节点数为 26,由输入数据包含的元素个数决定,代表输入数据的 26 个特征;隐藏层迭代学习输入数据的二维结构特征,其节点数为 130,由于其无相应理论指导,故由反复试验确定;输出层输出学习分类结果,其节点数为 3,由结果类别个数决定,代表输出数据的 3 个特征,具体网络参数见表 1。

表 1 参数选取

Tab.1 Selection of parameters

参数	取值
输入层节点数/个	26
隐层节点数/个	130
输出层节点数/个	3
学习率/%	0.01
最大迭代次数/次	50 000

3.3 结果与分析

利用 BP 神经网络对 3 类行为数据进行50 000 次迭代训练,得到的检测模型性能参数如表 2 所示,模型训练准确率达到 100%,模型网络损失仅为 0.091,其中训练模型的准确率在迭代 2 000 次后就迅速趋于 100%,此后随着迭代次数增加保持 100%的准确率不变;模型训练损失的曲线变化如图 5 所示,训练初始阶段网络损失为 400 左右,在 3 000次迭代训练后网络损失迅速下降,此后随着迭代训练次数的增加,网络损失最终趋于 0。

利用检测模型对每类 250 个共计 750 个行为数据进行实际测试,得到各个类别的检测准确率均达到 100%,模型总准确率亦为 100%。上述测试结果表明:

表 2 检测模型性能参数

Tab.2 Testing model performance parameters

参数	取值	参数	取值
准确率	1.0	网络损失	0.091

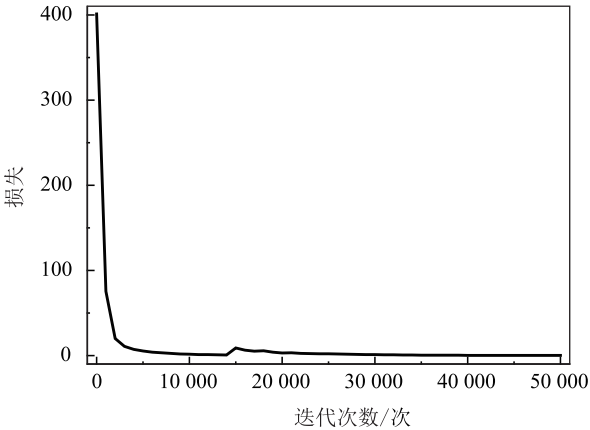


图 5 网络损失曲线

Fig.5 Network loss curve

(1)利用 OpenPose 与 BP 分类网络相结合的方法,可以实现对不同整体行为准确、高效的分类检测。

(2)将人体行为转换为二维坐标位置数据,即将人体行为二维数据化,可以将不同整体行为的表现特征以二维数据化的形式重新表现出来,这种表现形式不仅使行为特征更加量化具体,还使行为特征的训练学习变得更加简单有效,从而充分提高了训练模型的准确性和收敛速率。

(3)待检测行为的二维数据化结合 BP 分类网络具有很强的非线性映射能力,使行为数据的迭代训练与学习变得更加快速、准确、有效,进而保证了良好的行为检测效果。

OpenPose 与 BP 分类网络相结合的方法,可以对不同整体行为进行行之有效的检测。但是,如果将躯干整体拆分为一个个局部躯干,对于这些局部躯干的局部行为,上述方法又是否能达到同样良好的检测效果? 针对这一问题,接下来将检测人体的局部行为,分析讨论 OpenPose 与 BP 分类网络相结合的方法对人体局部行为的分类检测性能。

4 局部行为分析

4.1 数据处理与训练

共采集左臂平举、右臂平举、双臂平举共 3 类局部行为图片,每类行为图片均利用 usb 摄像头采

集 1 250 张,其中 1 000 张用于网络训练,余下 250 张用于网络实际测试。每类行为图片均利用摄像头从正面进行采集,且采集过程中使单臂平举的另一臂保持自然状态。利用 OpenPose 对上述 3 类行为图片中人体骨架节点的坐标位置信息进行提取,并将得到的躯干、四肢、头部等部位 18 个骨架节点的位置坐标信息保存在 json 文件内。

4.2 特征提取与学习

左臂平举、右臂平举、双臂平举这 3 类局部行为主要与左右臂共 6 个骨架节点有关,因此仅从 json 文件内提取这 6 个骨架节点的位置坐标数据作为网络的输入数据进行迭代训练与学习,然后利用训练所得模型对实际行为图片进行检测,从而得到最终的结果分类与输出,其具体流程与图 4 所示相同。

6 个骨架节点坐标记为 (P_1, P_2, \dots, P_6) , 将其展成一个骨架向量 $(P_{x1}, P_{y1}, P_{x2}, P_{y2}, \dots, P_{x6}, P_{y6})$ 作为 BP 神经网络的输入。输入数据不仅包含行为数据的点特征,同时包含点与点之间的线特征,以及线与线之间的面特征,这些特征通过 BP 分类网络进行提取与学习。特征学习所使用的 BP 分类网络共有 3 层,即一个输入层、一个隐藏层和一个输出层,其中输入层的节点数为 12 个,隐藏层节点数为 130 个,输出层的节点数为 3 个,其余与表 1 相同。

4.3 结果与分析

利用 BP 神经网络对 3 类行为数据进行 50 000 次迭代训练,得到的检测模型性能参数如表 3 所示,模型训练准确率达到 100%,模型训练损失小于 10^{-6} ,其中训练模型的准确率在迭代 1 000 次后就迅速趋于 100%,此后随着迭代次数增加保持 100%的准确率不变;模型训练损失的曲线变化如图 6 所示,训练初始阶段网络损失为 240 左右,在 2 000 次迭代训练后网络损失迅速下降,此后随着迭代训练次数的增加,网络损失最终趋于 0。

利用检测模型对每类 250 个共计 750 个行为数据进行实际测试,得到各个类别的检测准确率均为 100%,模型总准确率亦为 100%。上述测试结果表明:

表 3 检测模型性能参数

Tab.3 Testing model performance parameters			
参数	取值	参数	取值
准确率	1.0	网络损失	小于 10^{-6}

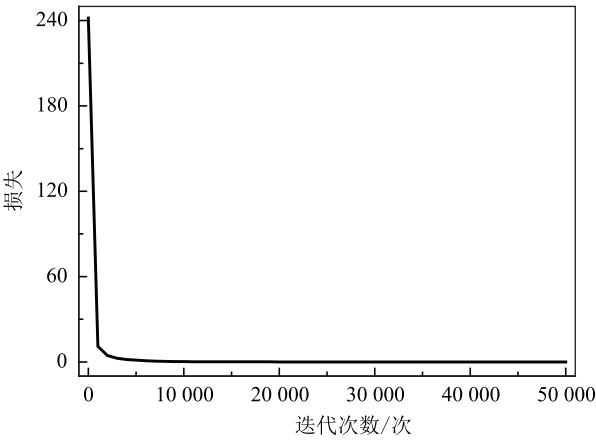


图 6 网络损失曲线
Fig.6 Network loss curve

(1) OpenPose 与 BP 分类网络相结合的方法,针对局部行为和整体行为均可达到准确、快速、高效的分类检测效果,该方法具有较好的行为检测普适性。

(2) 不论是局部行为,还是整体行为,行为的二维数据化都是行之有效的,它不会因为行为的整体性或局部性之分,导致行为特征的二维数据化表现程度出现偏好性或差异性。

4.4 讨论

利用 OpenPose 将人体行为转化为二维坐标位置数据,再利用 BP 分类网络对该数据进行迭代训练与学习,即 OpenPose 与 BP 分类网络相结合的行为检测方法,具有较好的行为检测普适性,其不仅可以对躯干整体产生的整体行为进行准确、快速、有效的分类检测,对局部躯干产生的局部行为同样可以达到很好的分类检测效果。OpenPose 与 BP 分类相结合的检测方法,收敛速度快、训练耗时短,相对于传统的行为检测方法大幅提高了特征学习效率。不仅如此,该方法还解决了传统行为检测方法准确率低、效率低、可靠性差等问题,可以在海量行为图片的分类任务中实现准确、高效、可靠的分类检测效果。

除 18 个身体关键点的识别检测外,对于面部 70 个关键点、以及左右手 42 个关键点的识别检测,使用 OpenPose 与 BP 分类网络相结合的检测方法,只要合理地设计网络及其参数,同样可以实现宏观与细节的普适检测,并具有良好的检测效果。这样,无论是整体面部表情、多手指姿态的识别,还是面部表情一个细节、单手指姿态的识别,都可以

实现准确、高效、可靠的识别检测效果。

5 结论

(1)对每类 250 个共计 750 个行为数据进行实际测试发现,无论是整体行为分类检测,还是局部行为分类检测,OpenPose 与 BP 分类网络相结合的方法均可以实现各类行为检测准确率达 100%,总体检测准确率亦达到 100%。

(2)OpenPose 与 BP 分类网络相结合的行为检测方法,具有较好的行为检测普适性,其不仅可以对躯干整体产生的整体行为进行准确、快速、有效的分类检测,对局部躯干产生的局部行为同样可以达到很好的分类检测效果。

(3)OpenPose 与 BP 分类网络相结合的方法收敛速度快、训练耗时短,且克服了传统行为检测方法效率低、准确率低、可靠性差等不足,能够对大批量行为图片进行更高效、更准确、更快速的行为分类检测。

参考文献:

[1] XIA L, CHEN C C, AGGARWAL J K. View invariant human action recognition using histograms of 3D joints[C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2012: 20-27.

[2] 陆中秋, 侯振杰, 陈宸, 等. 基于深度图像与骨骼数据的行为识别[J]. 计算机应用, 2016, 36(11): 2979-2984.

[3] ADAM N R, ATLURI V, HUANG W K. Modeling and analysis of workflows using Petri nets[J]. Journal of Intelligent Information Systems, 1998, 10(2): 131-158.

[4] 邹建. 基于 BP 神经网络的人体行为识别[J]. 网络与信息安全, 2017, 3(9): 1-6.

[5] SATO K, NAGASHIMA Y, MANO T, et al. Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning-based 2D pose estimator[J]. PloS one, 2019, 14(11): 1-15.

[6] LI X, FAN Z, LIU Y B, et al. 3D pose detection of closely interactive humans using multi-view cameras[J]. SENSORS, 2019, 19(12): 1-16.

[7] SONG C H, KIM S H. Robust vehicle occupant detection based on RGB-depth-thermal camera[J]. The Journal of Korea Robotics Society, 2018, 13(1): 31-37.

[8] PARK J H, SONG K H, KIM Y S. A kidnapping detection using human pose estimation in intelligent video surveillance systems[J]. Journal of the Korea Society of Computer and Information, 2018, 23(8): 9-16.

[9] 吕学志, 范保新, 尹建, 等. 基于 BP 神经网络的维修任务优先级分类方法[J]. 计算机工程与应用, 2014, 50(24): 250-254.

[10] 冷听, 张树群, 雷兆宜. 改进的人工蜂群算法在神经网络中的应用[J]. 计算机工程与应用, 2016, 52(11): 7-10.

[11] ZOU L, ZHAO Y X, JIAO B. Application of BP neural network in digital image recognition[J]. IOP Conference Series: Materials Science and Engineering, 2019, 490(7): 1-5.

[12] ZHANG L, WANG F L, SUN T, et al. A constrained optimization method based on BP neural network[J]. Neural Computing & Applications, 2018, 29(2): 413-421.

[13] ZHU Q L, WANG S G, SUN Q B, et al. Service classification based on improved BP neural network[J]. Journal of Internet Technology, 2018, 19(2): 369-379.

[14] GULIYEV N J, ISMAILOV V E. On the approximation by single hidden layer feed forward neural networks with fixed weights[J]. Neural Networks, 2018, 98(12): 296-304.

[15] SHI W W, GONG Y H, TAO X Y, et al. Improving CNN performance accuracies with min-max objective[J]. IEEE Transactions on Neural Networks & Learning Systems, 2018, 29(7): 2872-2885.

[16] ZHANG X B, CHEN F C, HUANG R Y. A combination of RNN and CNN for attention-based relation classification[J]. Procedia Computer Science, 2018, 131(4): 911-917.

[17] JIA Y Q, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[J]. Computer Science, 2014, 2: 675-678.

[18] MAKOTO I, TETSUYA O, LEONARD B. A vegetable category recognition system: a comparison study for caffe and Chainer DNN frameworks[J]. Soft Computing, 2019, 23(9): 3129-3136.

[19] LIU K H, ZHONG P S, ZHENG Y, et al. P_VggNet: A convolutional neural network (CNN) with pixel-based attention map[J]. PLOS ONE, 2018, 13(12): 1-11.

(责任编辑:于慧梅)

Behavior Detection Method Based on OpenPose and BP Network

ZHOU Deliang*

(Bei Jing Zhongdianyida Technology Co., Ltd., Beijing 100190, China)

Abstract: OpenPose and BP network is used to classify and detect the whole or local human behavior. Firstly, the coordinate data of human skeleton nodes are obtained by human pose estimation algorithm, and then the coordinate data of nodes are iteratively trained and learned by BP classification network. The accuracy of training models for detecting different overall behaviors is 100%, the network loss is only 0.091, and the accuracy of each category and the overall accuracy are 100%. The accuracy of training models for detecting local behaviors is 100%, and the network loss is less than 10^{-6} . The accuracy of each category and the overall accuracy are 100%. The combination of OpenPose and BP network not only achieves fast and accurate classification and detection tasks for different global or local behaviors, but also overcomes the shortcomings of traditional behavior detection methods, and achieves more efficient, more accurate and faster classification and detection.

Key words: BP neural network; OpenPose; behavior recognition; hidden layer

(上接第 86 页)

Research and Implementation of Topology Query for Remote Sensing Images

WEI Dandan^{1,2}, YU Jinsongdi^{1,2*}

(1.Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Spatial Information Research Center of Fujian Province, Fuzhou University, Fuzhou 350003, China; 2.The Academy of Digital China (Fujian), Fuzhou 350003, China)

Abstract: At present, the remote sensing image query method is mostly based on the form of key words. The topology-based query method has the spatial geometric invariance compared with the keyword-based query method, and is important in guaranteeing the recall and precision of query results. This paper studies the DE-9IM and its topology interface JTS Topology Suite, and bases on Solr to implement the Earth Observation 3 Intersection topology query function by JTS which provided topology judgement function. Experiments prove that JTS can be well applied to EO-3IM, and it also proves the enforceability of remote sensing image topology query.

Key words: topology query; JTS Topology Suite; EO-3IM; earth observation