DOI:10.19651/j.cnki.emt.2208880

基于骨骼的双支融合模型的人体行为识别*

罗旭飞[1,2] 崔 敏[1] 张 鹏[1,2]
(1.中北大学仪器与电子学院 太原 030051; 2.中北大学南通智能光机电研究院 南通 226010)

摘 要: 针对循环神经网络存在提取特征单一,对特征的空间信息处理不充分的问题,提出一种基于骨骼的双支融合的人体行为识别模型。该模型由双向循环门网络和多尺度的残差网络融合的双支网络中进行特征提取,得到丰富的时间和空间上的特征信息,并且在双向循环门网络中增加注意力机制,进一步提升整个网络的性能,最后将特征信息经过分类器进行分类得到动作。分别使用 UCF101 和 HMDB51 数据集进行实验,准确率分别为 98.0%和 67.8%。通过实验测试,证明该模型能够获得更加完整的特征信息并且具有良好的性能指标。

关键词: 人体行为识别;双支网络融合;多尺度特征;深度学习
中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.99

Human behavior recognition based on double-branch fusion model based on skeleton

Luo Xufei[1,2] Cui Min[1] Zhang Peng[1,2]
(1. School of Instruments and Electronics, North University of China, Taiyuan 030051, China;
2. Nantong Institute of Intelligent Opto-Mechatronics, North University of China, Nantong 226010, China)

Abstract: Aiming at the problem that the recurrent neural network has a single feature extraction and insufficient processing of spatial information of the feature, a two-branch fusion human behavior recognition model based on bone is proposed. The model is extracted by the two-branched network of two-way cyclic gate network and multi-scale residual network, which obtains rich feature information in time and space, and increases the attention mechanism in the bidirectional cyclic gate network to further improve the performance of the whole network, and finally the feature information is classified through the classifier to obtain action. Experiments were conducted using the UCF101 and HMDB51 datasets, respectively, with an accuracy rate of 98.0% and 67.8%, respectively. Through experimental tests, it is proved that the model can obtain more complete feature information and has good performance indicators.

Keywords: human behavior recognition; dual branch network convergence; multiscale features; deep learning

0 引 言

随着图像处理的不断发展,视频监控的使用已经越来越广泛,通过视频监控的对人体行为的分析和判断是机器视觉一个重要的领域,由于机器视觉容易受到光照强度、空间背景以及视角的不同,人体行为识别的研究也越来越重要。人体行为的检测和识别主要分为基于传感器的行为识别以及基于计算机视觉的行为识别[1]。目前使用比较广泛的是基于计算机视觉的行为识别方式,该方法能通过与使用图形处理技术与深度学习和机器学习等算法相结合对视觉信息进行处理分类识别,并且能够满足数据量大、计算量大以及提高速度与精度的要求[2]。人体行为识别是从人工提取关键的特征信息进行处理的,逐渐转变为通过是深度学习的方式来进行分类处理,通过这种转变提高了算法的准确度和鲁棒性[3]。

基于机器视觉的人体行为识别的主要是对人体行为是被信息的提取与分类,在对人体行为提取的信息主要是包括光流图像信息[4]、RGB 图像信息[5]和人体骨架信息[6]等。通过使用 3D 光流法[7]提取到 3D 特征,并且经过数据预处理对所得到的特征信息进行编码设计。还可以使用时域扩张残差网络和双分支特征提取方法[8],获取 RGB 图像中的人体的空间和时间的语义信息。人体骨架信息主要是通过 Kinect 相机和 Openpose 算法等方式获取的。能够使用 Kinect 相机采集获取人体的骨架信息[9],然后对骨架信

收稿日期:2022-01-18
* 基金项目:军委装备发展部预研基金(41403010305)、装备预研兵器工业联合基金(6141B012907)项目资助

息进行标注处理。

目前研究基于机器视觉的人体行为识别主要是与深度学习相结合实现的。循环神经网络在对时间序列的信息处理时有这较为良好的效果[10]，所以可以使用 RNN 网络对人体行为的时间信息进行处理，并且 RNN 网络一般与 CNN 想结合能够有更好的识别效果[11]。由于 RNN 网络存在这梯度爆炸和梯度消失的问题[12]，所以改进 RNN 网络提出了长短时记忆网络(LSTM)[13]，LSTM 网络使用记忆细胞，通过对过去信息的筛选与当前时刻信息的结合，相比与 RNN 网络有着更加优异的表现[14]。由于 LSTM 网络的模型存在参数量大以及模型比较复杂的特点[15]，针对这些问题提出了门控循环单元(GRU)[16]网络，该网络主要简化了 LSTM 网络的参数模型，并且在一定条件下精度有所提高。

本文主要是针对循环神经网络中提取特征单一及对特征的空间信息处理不充分的问题，提出使用 Openpose 算法获得关键点信息，将信息输入到双向循环门网络和多尺度的残差网络双支融合网络实现特征提取。在双向循环们网络中加入注意力机制更精准的获得时间特征信息，使用多尺度的残差网络实现人体骨骼关键点的空间信息进行处理，将两个分支的进行融合得到完整的人体行为特征信息，最后经过的分类函数进行人体行为分类。实验结果表明，本文设计的模型在使用 F1 参数等图像处理的评价指标时获得了较好的结果，并且实验测试也表现了良好的性能。

1　模型总体框架

基于骨骼的双支融合模型的人体行为识别主要是使用 Openpose 进行人体行为的关键点信息进行检测，将检测的关键点信息通过数据预处理之后，输入出双向循环网络和残差网络所构建的双支网络，提取关键点的时间和空间的特征信息，最后将所得到的特征信息处理分类，得到人体行为识别的结果，整体的框架图如图 1 所示。

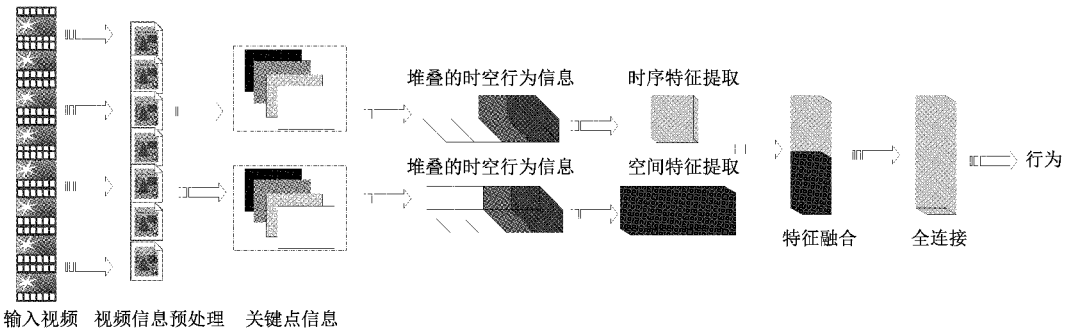

图 1　模型总体框架

1.1　骨骼关键点信息的获取

Openpose 是一种自上而下的姿态关键点检测算法，是由卡梅隆大学感知实验室提出的，该算法主要是通过使用 CNN 网络和 VGG19 网络进行处理来获取热力图来获得关键点的信息，然后通过使用向量图将所获得的关键点的信息进行连接，该算法的整体流程如图 2 所示。

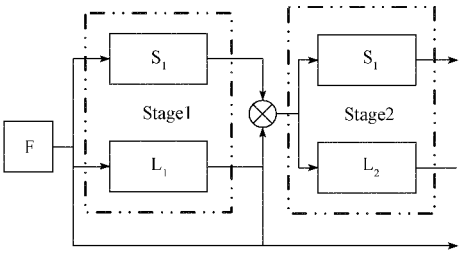

图 2　Openpose 算法原理

该图中的 S_t 就是热力图的信息，L_t 代表的是骨骼的向量，通过对 S 和 L 的处理将处理后的结果和原始特征进行融合，将融合后的特征在一次输入下级网络进行处理，数次迭代后进行连接获得骨骼的关键点，可用式(1)进行表示：

$$f=\sum_{t=1}^{T}(f_S^t+f_L^t) \tag{1}$$

1.2　数据预处理

人体行为是连续时间顺序上的动作，单帧的人体图像信息无法完全表达该图像上的人体行为动作，所以需要输入连续的时间帧的图像，表示该动作的信息，数据的预处理就是实现连续多帧图像的输入以及对数据的标注。通过对连续的单帧图像进行分组处理，使每个组都能获得连续的多帧的动作图像，保证每个组的图像帧数相同。将分组后的动作图像通过独立热编码进行标注分类，将分类好的数据投入深度学习的网络中进行训练以及测试。

独立热编码是一位有效编码，在所分的类别较少的分类中使用的比较广泛，它针对不同的类别提供一个独立的有效编码进行标注。在使用独立热编码时需要进行特征归一化的处理，特征归一化主要是将输出值映射到[0，1]的区间范围内，保证数据能够在相同量上进行处理，如式(2)所示。

$$X_{norm}=\frac{X-X_{\min}}{X_{\max}-X_{\min}} \tag{2}$$

通过多帧图像的输入能够保证一个动作的完整性以及获得该动作更多的特征信息，从而训练和测试的数据更加的准确。

2　双支模型网络设计

本文的双支融合模型主要是由双向循环门网络和注意力机制所组成的时间特征提取网络和多尺度的残差网络所组成的空间特征提取网络以及分类器等模块所组成的。该模型主要将时间特征提取和空间特征提取相结合，将所得到的信息通过特征融合后，放入分类器中进行分类，获得最后的人体行为的种类。

在双向循环门网络中加入通道注意力机制，使用 Relu 当做激活函数。使用注意力机制保证当前时刻信息的重要程度，从而使时间顺序上的特征信息能更好的表现出当前动作，将得到的特征信息输入打平均池化层中进行维度的调整，保证两个分支的特征信息的融合。

多尺度的残差网络的主要是分别使用 1 * 1、1 * 1 和一个 3 * 3 以及 1 * 1 和两个 3 * 3 的卷积核，对输入特征进行处理，得到多尺度的特征信息，将多尺度的特征信息通过使用 Concat 操作进行连接，将所获得特征信息在通过一个 1 * 1 的卷积核处理，将最后得到的特征输入到全局平均池化层。通过使用 Concat 将两个分支的特征信息进行连接，获得完整的时空特征信息，通过是同 softmax 分类器进行动作分类。整体的模型设计如图 3 所示。

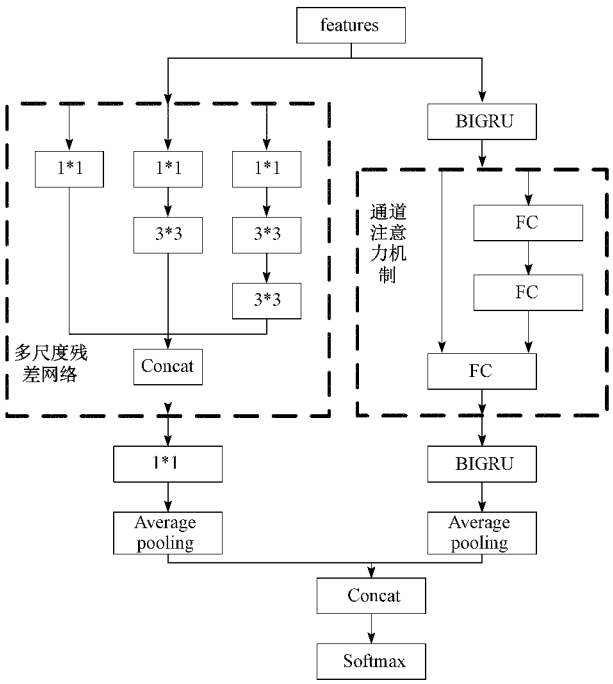

图 3　双支融合模型结构图

2.1　时间特征提取网络

时间特征提取网络主要是由双向循环门网络以及注意力机制网络所组成的，该部分网络主要是处理 Openpose 所获得关键点的时间序列的信息，因为循环神经网络对时间序列的处理较优秀，在此基础上增加注意力机制可以更好的获取到时间特征信息。

双向循环们网络的基本单元是 GRU 网络，该网络主要通过当前时刻的输入 x^t 和上一个时刻的状态 h^{t-1}，得到更新门 z 和重置门 r，h' 是对当前时刻状态的记忆，y^t 是该隐藏节点的输出，h^t 是传递该节点信息给下个节点，GRU 网络的基本结构如图 4 所示。

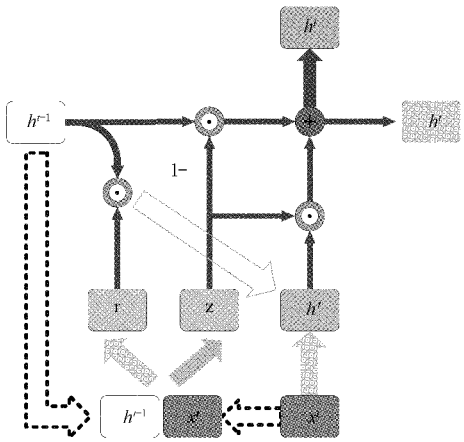

图 4　GRU 网络基本结构图

通过对更新门 z 和重置门 r 进行进行计算就能够同时实现遗忘和记忆，最后我们可以得到 h^t，如式(3)所示。

$$h^t = (1-z)\odot h^{t-1} + z\odot h' \tag{3}$$

在处理时间特征是加入通道注意力机制，增加对当前时刻信息的权重，更好的获得时间上的特征信息，通道注意力机制主要是将特征提取后通过 1 * 1 的卷积核操作，最后通过激活函数得到最后的特征，其结构如图 5 所示。

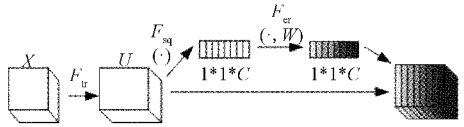

图 5　注意力机制的结构图

2.2　空间特征提取网络

空间特征提取网络主要是由多尺度的残差网络所实现的，该网络能够更好的提取到人体运动过程中的空间特征的变化，实现对人体行为特征的多元化的处理。

残差网络的提出主要是为了解决深度学习过程中，网络的深度过深存在梯度消失和梯度爆炸的问题所提出的，它主要是通过跳跃链接实现，特征信息从某一层直接传送至另一层，不需要通过顺序链接实现，其原理结构如图 6 所示。

本文主要是使用多尺度的残差网络进行特征提取，多尺度的残差网络主要是将分别通过 1 * 1、3 * 3 的卷积核操作分别对特征进行提取，可以得到不同尺度的特征信息，最后使用 Concat 连接三个分支的特征，输入到一个 1 * 1 的卷积核得到最后的特征信息。该信息能够更好的表现出人体行为的空间特征。

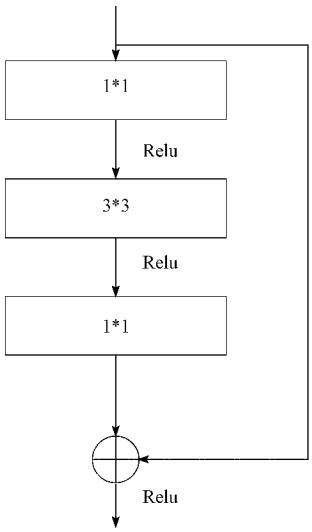

图 6 残差网络原理结构

2.3 分类器模块

通过分类器对所获得特征信息进行分类处理,其原理如下:构建 SoftMax 为激活函数,它为每一种分类都添加了一种概率,其中 Softmax 函数表达式如式(4)所示。

$$\sigma i(Z)=\frac{e^{Z_i}}{\sum_{j-1}^{m}e^{Z_i}} \tag{4}$$

其中,Z_i 的含义是指函数在第 i 个节点所具备的输出值,中的 j 作为的输出节点的类别总的个数。如神经网络原始输入 Z_1、Z_2、Z_3…此函数就可以把单个节点的输出值变成为概率分布。而 Softmax 激活函数所对应的损失函数为交叉双熵函数,其表达式如式(5)所示。

$$H(p,q)=-\sum_{i=1}^{n}p(x_i)\log(q(x_i)) \tag{5}$$

3 实验测试与结果分析

3.1 数据集及评价指标

本文使用是公开的人体行为数据集为 UCF101 和 HMDB51。UCF101 是行为类别和样本数量最多的数据库之一,其中包含 13 320 个视频和 101 个类别。HMDB51 包含 6 849 个视频,总共 51 个类别,每个类别至少包含 101 个视频。数据集按照 8∶2 的比例分为训练集以及测试集,进行模型的训练。

本文使用的评价指标主要是图像检测的相关指标包括准确率(Acc)、精准率(Pre)、召回率(Rec)以及 F1 分数(F1)进行评价,它们主要是通过正确的正样本个数(TP)、不正确的负样本个数(TN)、正确的负样本个数(FP)以及不正确的正样本个数(FN)进行计算得出的,具体计算如式(6)~(9)所示。

$$Acc=\frac{TP+TN}{TP+TN+FP+PN} \tag{6}$$

$$Pre=\frac{TP}{TP+FP} \tag{7}$$

$$Rec=\frac{TP}{TP+FN} \tag{8}$$

$$F_1=\frac{2*Pre*Rec}{Pre+Rec} \tag{9}$$

3.2 实验环境与参数设置

本文实验的硬件条件为 GPU GTX 3070Ti,CUDA 版本为 10.2,所使用的深度学习的框架为 Tensorflow,python 的版本使用的是 3.7,学习率初始值设置为 0.1,迭代次数设置为 50,Batch 值设置为 32,使用 Adam 进行参数更新,学习率的衰减设置为 0.000 1,采用交叉熵损失函数计算损失。

3.3 实验分析

1)消融实验

本实验主要是针对该模型的对各个分支以及与模型的对比,该实验主要是分别使用时间提取分支、空间提取分支以及两个分支融合提取的方式对特征信息进行处理,通过分类器模块对特征分类得到最后的分类的结果。

通过分别对各部分的训练得到实验数据计算评价指标进行对比,评价指标如表 1 所示。

表 1 各分支的评价指标 %

模型结构	Acc	Pre	Rec	F_1
Resnet	57.6	55.1	54.8	54.9
GRU	86.6	86.3	86.1	86.2
Re-GRU	98.0	97.7	97.8	97.7

从表 1 的各项评价指标可知,双支融合模型的各项指标均有较大幅度的提升,所以融合后的模型整体性能比其它两个模型要优越。

为了更加直观的显示实验的对比结果,该实验使用混淆矩阵来对各种动作的情况进行描述,如图 7~9 所示。

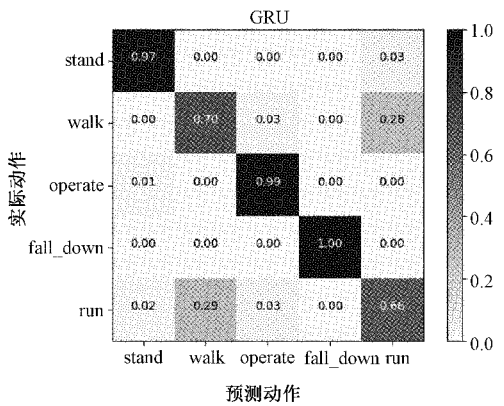

图 7 空间特征分支的混淆矩阵

通过对混淆矩阵对比可以看出,相似动作对于 3 个模型的影响比较大。如图 7 的空间特征分支(Resnet),除了

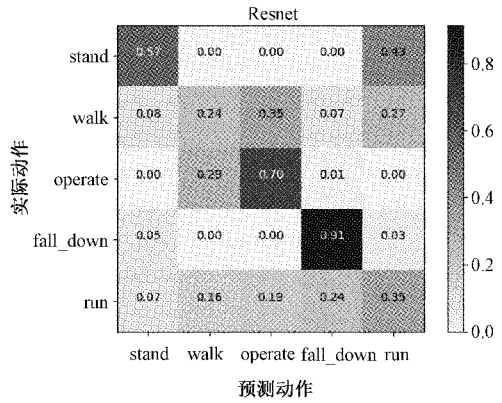

图 8　时间特征分支的混淆矩阵

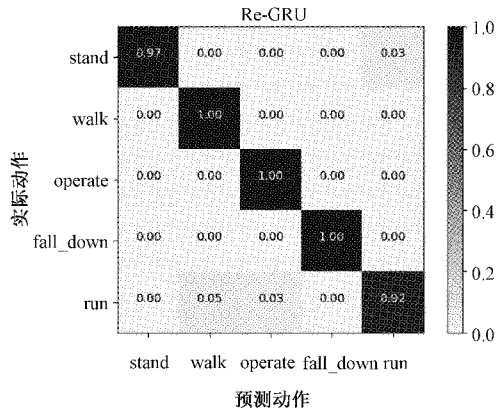

图 9　融合双支的混淆矩阵

相似动作较低外，非相似动作的准确率也比较低，如图 8 的时间特征分支(GRU)，除了相似动作的准确率有较小的提升外，非相似动作的准确率提升比较大。如图 9 的双支融合模型(Re-GRU)，除相似动作会有较小的误判，其它均能判断正确。

2)注意力机制实验

本实验主要是针对时间特征提取中有无注意力机制对整个网络准确度的影响，该实验分别在 UCF101 和 HMDB51 两个数据集上进行，实验结果如表 2 所示。

表 2　有无注意力机制实验　%

是否有注意力机制	UCF101	HMDB51
无	95.4	64.2
有	98.0	67.8

通过实验可以看出，是否添加注意力机制对整个模型的性能有一定的影响，在 UCF101 和 HMDB51 两个数据集上，在添加注意力机制后，Re-GRU 模型的准确率分别提升了 2.6%和 3.6%，增加注意力机制可以使整个网络的性能有所提升。

3)对比实验

本实验主要是通过使用相同的公开数据集将本文的算法模型与现阶段的一些算法模型进行对比，分别使用 UCF101 和 HMDB51 选取 5 类动作进行对比，实验数据如表 3 所示。

表 3　算法对比实验　%

算法	UCF101	HMDB51
CNN+LSTM	94.4	62.7
ResNeXt+Attention	95.2	65.6
Inceptionv3+Bi-LSTM+Attention	94.6	63.4
Clips+CNN+MTLN	93.7	60.8
TSN	94.2	69.4
Re-GRU	98.0	67.8

通过对上述算法在不同的数据集上对比可知，在 UCF101 数据集上的识别精度普遍比在 HMDB51 数据集上的识别精度要高。使用相同的数据集对比，本文的算法相比于其它的人体行为识别算法均有所提升，在 UCF101 数据集上本文的算法效果最好，准确度提升了 2.8%，同时在 HMDB51 数据集上本文的算法也能达到一个较高的准确率。

4)系统实验测试

本实验主要是通过使用 python 和 pyqt 搭建行为识别的系统，对人体的行为动作进行实时的检测和判断，实验测试的结果如图 10 所示。

由图 10 可以看出，使用该模型能够正确的识别多人的行为动作种类，然后对模型中的 5 类动作分别进行了多次实验的测试，5 类动作的准确率如表 4 所示。

表 4　动作准确率表

动作类别	实验总次数	正确次数	准确率/%
stand	200	185	92.5
walk	200	181	90.5
operate	200	184	92.0
fall_down	200	195	97.5
run	200	168	84.0

由表 4 可知，在实验测试时，相似动作的准确率相比于其它的较低，除了跑之外其它动作识别的准确率能够保证在 90%以上，实现了系统所预期的效果。

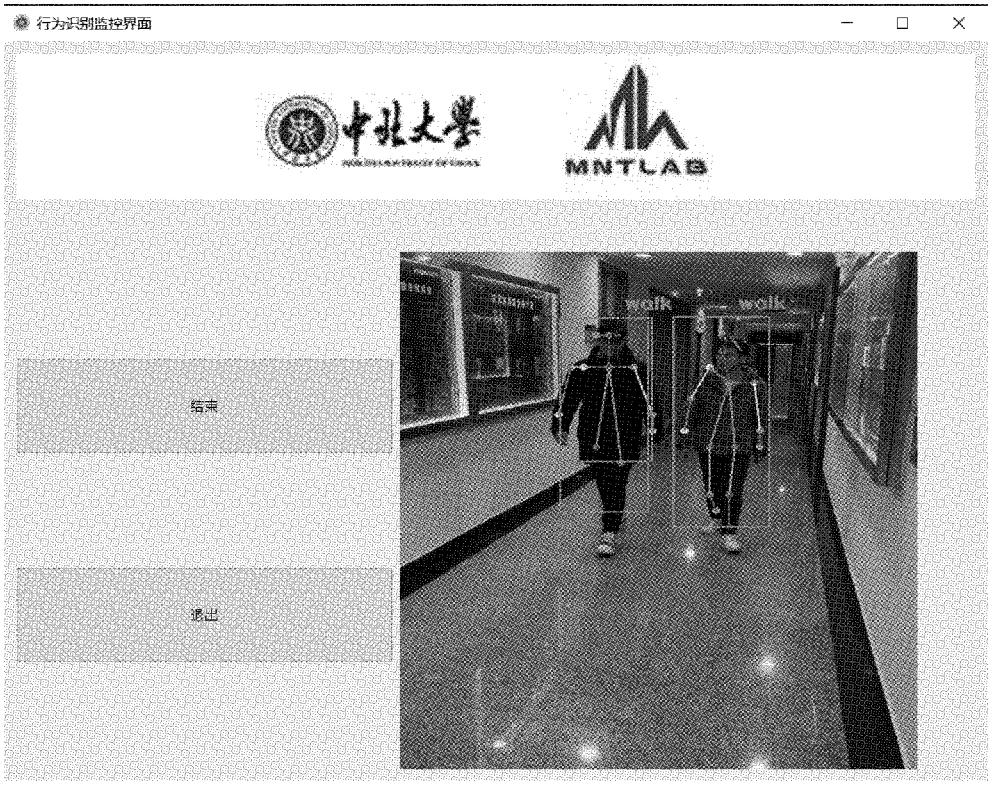

图 10 实验测试结果图

4 结 论

本文提出的基于骨骼的双支融合的人体行为识别模型，主要是通过使用多尺度的残差网络和带有注意力机制的双向循环门网络分别对人体行为过程中的空间和时间特征进行提取，保证获取到更加完整和有效的人体行为特征信息，提高识别的精度。实验结果表明，本文的算法在 UCF101 和 HMDB51 数据集上都有较高的准确率，并且通过系统实验测试，证明该模型有较好的性能。

参考文献

[1] 杨观赐，李杨，赵乐，等. 基于传感器数据的用户行为识别方法综述[J]. 包装工程，2021，42(18)：94-102，133，11.

[2] 游伟，王雪. 人行为骨架特征识别边缘计算方法研究[J]. 仪器仪表学报，2020，41(10)：156-164.

[3] 裴利沈，刘少博，赵雪专. 人体行为识别研究综述[J/OL]. 计算机科学与探索：1-23[2021-11-11].

[4] SIMONYAN K，ZISSERMAN A. Two-stream convolu，tiona-l networks for action recognition in videos [C]. Ad-Vances In Neural Information Processing Systems，M-onTreal，Mit Press，2014：568-576.

[5] FEICHTENHOFER C，FAN H，MALIK J，et al. Slowfastn-etworks for video recognition [C]. Proceedings Of The 2019 Ieee/Cvf International Conference On Computer Vision，Piscataway：Ieee，2019：6201-6210.

[6] 王攀. 基于骨骼的人体行为识别技术研究与应用[D]. 成都：电子科技大学，2021，DOI：10.27005/d.cnki.gdzku.2021.003598.

[7] ASHWAN A，LAI Y K，SUN X F. Saliency guided local and global descriptors for effective action recognition[J]. Computational Visual Media，2016，2(1)：97-106.

[8] 薛盼盼，刘云，李辉，等. 基于时域扩张残差网络和双分支结构的人体行为识别[J/OL]. 控制与决策：1-10[2021-11-25].

[9] 田皓宇，马昕，李贻斌. 基于骨架信息的异常步态识别方法[J/OL]. 吉林大学学报(工学版)：1-13[2021-11-25].

[10] 宁亚飞，赵英亮，吴美荣，等. 时空卷积自编码网络异常行为检测[J]. 国外电子测量技术，2020，39(10)：104-108.

[11] 周楠，陆卫忠，丁漪杰，等. 基于深度学习的人体行为识别方法研究综述[J]. 工业控制计算机，2021，34(8)：116-117，119.

[12] 吕亚兰，安建伟. 基于特征融合双流网络的人体行为识别[J]. 电子测量技术，2020，43(20)：121-126.

[13] 王丽君，刘彦戎，王丽静. 基于卷积长短时深度神经网络行为识别方法[J]. 电子测量与仪器学报，2020，34(9)：160-166.

[14] 曾明如,郑子胜,罗顺.结合 lstm 的双流卷积人体行为识别[J].现代电子技术,2019,42(19):37-40.

[15] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical e-valuation of gated recurrent neural networks on sequence modeling[J]. ArXiv Preprint, 2014, ArXiv:1412.3555.

[16] LI C, ZHONG Q, XIE D, et al. Skeleton-based action r-ecognition with convolutional neural networks [C]. Proceedings of 2017 IEEE International Conference on Multimedia & Expo Workshops. HongKong: IEEE, 2017: 597-600.

作者简介

罗旭飞,硕士研究生,主要研究方向为嵌入式硬件系统的开发,机器学习。

E-mail:luoxufeinuc @163.com

张鹏(通信作者),副教授,主要从事自动控制、嵌入式系统、存储测试及机器学习等方面的教学和研究工作。

E-mail:sxyczhangpeng@126.com