

基于改进 ResNet-18 的红外图像人体行为识别方法研究

周啸辉¹, 余磊¹, 何茜¹, 陈涵¹, 聂宏², 欧巧凤¹, 熊邦书¹

(1. 南昌航空大学 图像处理与模式识别江西省重点实验室, 江西 南昌 330063;

2. 北爱荷华大学技术系, 美国 爱荷华州 50614 - 0178)

摘要: 人体行为识别在安全监护、安防监控、智能家居等诸多领域具有重要的研究意义和广泛的应用价值。由于红外信息具有受光照影响小、保护隐私等特性, 因此基于红外信息的人体行为识别方法备受国内外学者关注。本文对包含7种行为类别的红外信息进行连续帧拼接处理, 构建红外图像数据集。传统的 ResNet-18 网络性能较为优异, 在可见光图像识别上一直表现良好, 但在红外图像识别中效果欠佳。本文根据红外图像特性, 对其进行相应改进: 首先, 构建多分支同构结构, 替换 7×7 卷积, 增强网络的表达能力; 其次, 结合最大池化与平均池化, 避免丢失有用信息; 最后, 引入非对称卷积块构成多重残差结构, 并与改进 CBAM 模块结合对残差块进行优化, 从而增加网络多样性, 提升网络的特征提取能力。实验结果表明, 改进 ResNet-18 网络识别率达到 99.96%, 不但高于传统的 ResNet-18 网络, 而且明显优于基于红外图像的其他网络。

关键词: 人体行为识别; 改进 ResNet-18 网络; 红外图像; 多重残差结构; 改进 CBAM 模块

中图分类号: TP391.4; TN219 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2021.09.011

Research on human behavior recognition method in infrared image based on improved ResNet-18

ZHOU Xiao-hui¹, YU Lei¹, HE Xi¹, CHEN Han¹, NIE Hong², OU Qiao-feng¹, XIONG Bang-shu¹

(1. Key Laboratory of Image Processing and Pattern Recognition of Jiangxi Province,

Nanchang Hangkong University, Nanchang 330063, China;

2. Department of Technology, University of Northern Iowa, Iowa 50614 - 0178, USA)

Abstract: Human behavior recognition is widely applied in safety monitoring, security monitoring, smart home, etc. and proves to be of great research significance in these areas. The recognition method of human behavior based on infrared information has attracted the attention of scholars at home and abroad because infrared information is barely influenced by illumination and able to protect users' privacy. In this paper an infrared image data set is constructed by sequential frame stitching of infrared information with seven behavior categories. The traditional ResNet-18 network which works perfectly on visible image recognition has poor performance on infrared image recognition. In this paper, corresponding

基金项目: 国家自然科学基金项目 (No. 61866027); 江西省自然科学基金项目 (No. 20202BAB202016); 南昌航空大学研究生创新专项基金项目 (No. YC2020043) 资助。

作者简介: 周啸辉 (1995 -), 男, 硕士研究生, 研究方向为图像识别与人工智能。E-mail: zhouxiaohuiiii@126.com

通讯作者: 余磊 (1984 -), 男, 副教授, 主要从事图像处理及应用。E-mail: yulei@nchu.edu.cn

收稿日期: 2020-11-17; **修订日期:** 2020-12-21

improvements are made according to the characteristics of infrared images. First, a multi-branch isomorphic structure is established to replace 7×7 convolution, which increases the expressive ability of the network. Second, maximum pooling is combined with average pooling to avoid useful information lost. Finally, the multiple residual structures are constructed by introducing asymmetric convolution block and combining with the improved CBAM module to optimize the residual block, then increase the network diversity and enhance the feature extraction ability of the network. The experimental results show that the recognition rate of the improved ResNet-18 network is 99.96%, which is higher than that of the original ResNet-18 network as well as other networks based on infrared images.

Keywords: human behavior recognition; improved ResNet-18 network; infrared images; multiple residual structure; improved CBAM module

1 引言

人体行为识别是计算机视觉和模式识别领域的核心课题之一,目标在于识别图像或视频中人的行为^[1]。对人体行为进行有效且精准地识别是许多智能服务的基础,在智能监控、智能家居、虚拟现实等诸多领域中具有重要的研究意义和广泛的应用价值,备受国内外学者的关注^[2]。

近年来,红外成像系统以其受光照影响小、保护隐私的独特优势受到越来越广泛的关注,众多优秀学者对红外行为识别领域展开深入研究。Zhu 等人^[3]提出了一种随机采样结构,并建立统计特征空间;Hilsenbeck 等人^[4]利用霍夫森林提取时空域的整体通道特征;Tan 等人^[5]提出一种基于运动偏差图像的相位特征描述符,在受试者个体差异较大的情况下仍具有较高的识别精度;Liang 等人^[6]提出一种分段架构,利用子动作间关系,并结合异构信息融合和类隐私保护协同表示,在多个数据集上取得了不错的表现。上述方法主要通过手动提取特征,并制定相应的判决标准进行人体行为识别。然而,这些方法提取的特征往往不能准确表征人体实际行为,从而导致识别准确率较低。深度学习能自动提取有效特征,为解决该问题的一种有效途径。

深度学习^[7]是机器学习的一个分支,在视觉识别和自然语言处理等领域取得了突破性进展,也给红外行为识别带来了前所未有的发展。目前,红外行为识别领域的深度学习主要方法分为两类:基于视频或基于图像。

基于红外视频,Gao 等人^[8]构建了一个红外行为识别数据集 InfAR,并提出一种双流卷积神经网络,平均识别率为 76.66%;Liu 等人^[9]提出基于全局时间表示的三流卷积神经网络,整合局部、全

局时间信息和时空信息,在数据集 InfAR 和 NTU RGB + D 上取得 79.25% 和 66.29% 的平均识别率;Yang 等人提出 3D 卷积神经网络结合 SVM 分类器的方法,识别精度达到 96.5%。

基于红外视频的方法虽然取得了较高的识别精度,但它们的模型复杂,处理时间长,效率较低。相较而言,基于红外图像的人体行为识别方法模型简单,处理速度较快,逐渐受到研究者关注。Akula 等人^[10]提出将 2D CNN 网络用于识别行走、站立、静坐、在桌子前静坐、跌倒和倒在桌子上 6 种行为,其对跌倒的识别精度较低;Gochoo 等人^[11]利用三个传感器,从 x 、 y 和 z 轴捕获人体瑜伽姿态,其总体识别率较高。然而,基于红外图像的方法仅对静止行为有较好的识别效果;对于跌倒等包含一系列连续动作的行为,由于难以依据单帧图像进行识别,这些方法往往不能达到令人满意的效果。

针对上述问题,本文提出红外图像拼接方法:通过将连续多帧图像拼为一张大图,实现从三维视频到二维图像的转变。这样既保留了数据的时空联系,有利于后续识别工作;又可简化识别模型,减少处理时间。残差网络^[12]便于训练,易于收敛,泛化能力强,性能优异,在图像识别方面得到了广泛应用。ResNet-18 作为一种典型代表,识别精确率较高且模型参数较少;其在可见光图像识别上一直表现良好,但在红外图像识别中效果欠佳。本文在 ResNet-18 网络的基础上,根据红外图像的特性对其进行改进:首先,构建多分支同构结构,替换 7×7 卷积;然后,结合最大池化和平均池化对池化层进行优化;最后,引入非对称卷积块构成多重残差结构,并将改进 ECA 模块与 CBAM 的空间注意力模块结合,加入残差块中。此外,考虑

到跌倒常对人,尤其对老年人造成较为严重的伤害^[13],因此本文选择跌倒和易与跌倒相混淆的站到坐、坐到躺等行为,以及站立、静坐、平躺、行走等常见行为作为研究对象,构建红外人体行为识别系统,并对这 7 种行为进行识别,取得了良好的效果。

2 人体行为识别系统

本文所设计的人体行为识别系统如图 1 所示。首先,选择德国海曼器件公司制造的 HT-PA80x64dR1L5.0/1.0 热成像阵列传感器采集红外信息,其分辨率为 80×64 像素。其次,为了尽可能减少视场盲区,将传感器安装于距离地面 1.8 m 的墙角处,与水平方向的夹角为 88° ,与垂直方向的夹角为 70° 。从图 1 可以看出,在 1.8 m 以下的区域中,除了阴影部分外,其余均在视场范围内。再次,将传感器采集的温度分布信息,通过 I²C 接口发送至树莓派,并进一步通过内置通信模块将其发送至 PC 端,从而获取红外图像。最后,对红外图像进行拼接,再经过改进 ResNet-18 卷积神经网络算法进行分类,完成人体行为识别。

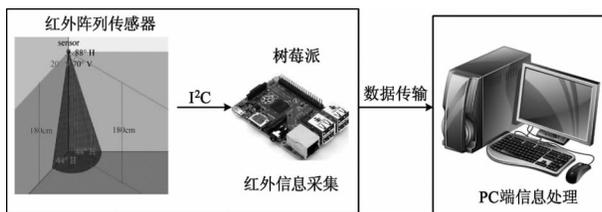


图 1 系统结构示意图

Fig. 1 System structure diagram

3 红外图像拼接

对于采集到的红外图像,考虑到帧与帧之间的关联性,基于视频流的深度学习方法是一个很好的选择。然而,传统基于视频的深度学习方法存在模型复杂,处理时间长等缺陷。基于图像的深度学习方法虽然具有模型相对简洁,运算处理快等优势;但人体行为往往包括一系列连续动作,难以仅依据单帧图像进行识别。为了充分利用数据的时空联系,方便后续识别,同时简化识别模型,提高处理效率,本文提出红外图像拼接方法:将连续多帧图像组合为一张大图,实现从三维视频到二维图像的转变。

由于跌倒,从站到坐,从坐到躺等人体姿态变

化过程的持续时间不超过 4 秒,且 HT-PA80x64dR1L5.0/1.0 热成像阵列传感器帧率约为 5 f/s,因此为了在一张大图中包含完整的行为过程,本文选取 25 帧连续图像进行拼接,图 2 为拼接后的跌倒状态示意图。

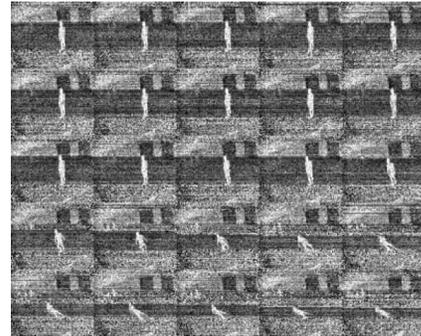


图 2 拼接图像示意图

Fig. 2 Schematic diagram of mosaic image

4 改进 ResNet-18 网络

近年来,由于在图像和语音识别方面的良好表现,卷积神经网络逐渐成为研究热点。残差网络是一种表现优异的卷积神经网络,在图像识别方面有着广泛的应用。ResNet-18 网络作为一种典型代表,其性能优异且模型参数较少,在保持较高准确率的同时仅占用较少的内存空间^[14]。

ResNet-18 网络在可见光图像识别上一一直表现良好,但在红外图像识别中表现欠佳。本文在 ResNet-18 网络结构的基础上,针对人体行为识别任务,并根据红外图像自身的特性,提出一种改进 ResNet-18 网络结构,该结构如图 3 所示。首先,构建多分支同构结构 Conv_x,替换 7×7 卷积;然后,结合最大池化和平均池化,对池化层 pool1、pool2 进行改进;最后,引入非对称卷积块构成多重残差结构,并将改进 ECA 模块与 CBAM 的空间注意力模块结合,加入残差块 Conv1_x、Conv2_x、Conv3_x 和 Conv4_x 中。

对于 Conv_x 中的每个卷积层,图 3 中依次给出了输入通道数、卷积核大小及输出通道数。网络的输出将会给出图像的类别,即图像是跌倒,站立,静坐,平躺,行走,站到坐以及坐到躺中的哪一类。表 1 给出了该网络的相关参数,包括各层的名称,结构参数和输出大小,及相应卷积层的核大小、输出通道数,池化层的核大小、步长。

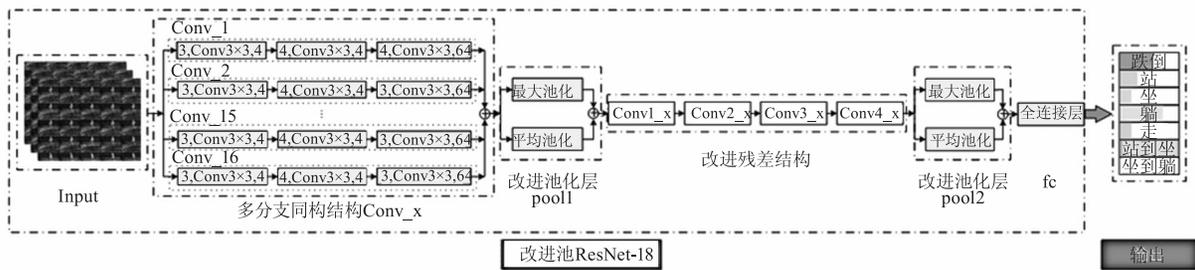


图3 改进 ResNet-18 网络结构

Fig. 3 Improve ResNet-18 network structure

表1 改进 ResNet-18 网络相关参数

Tab.1 Improve ResNet-18 network related parameters

各层名称	结构参数	输出大小
Input	—	400 × 320 × 3
Conv _x	$\begin{bmatrix} 3 \times 3, 4 \\ 3 \times 3, 4 \\ 3 \times 3, 64 \end{bmatrix} \times 16$, 并行连接	200 × 160 × 64
pool1	最大池化 平均池化 3 × 3, 步长为 2	100 × 80 × 64
Conv1 _x	$\left[\begin{matrix} 1 \times 3 \times 3, 64, 1, \begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \\ \text{改进 CBAM} \end{matrix} \right] \times 2$	100 × 80 × 64
Conv2 _x	$\left[\begin{matrix} 1 \times 3 \times 3, 128, 1, \begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \\ \text{改进 CBAM} \end{matrix} \right] \times 2$	50 × 40 × 128
Conv3 _x	$\left[\begin{matrix} 1 \times 3 \times 3, 256, 1, \begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \\ \text{改进 CBAM} \end{matrix} \right] \times 2$	25 × 20 × 256
Conv4 _x	$\left[\begin{matrix} 1 \times 3 \times 3, 512, 1, \begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \\ \text{改进 CBAM} \end{matrix} \right] \times 2$	12 × 10 × 512
pool2	最大池化 平均池化 3 × 3, 步长为 2	6 × 5 × 512
fc	输入有 15360 个神经元结点, 输出为 7 个	7

4.1 多分支同构结构

传统残差网络中第一层卷积为 7 × 7 的大卷积, 目的是在不增加通道数的情况下, 尽量保留原始图像的信息。然而红外图像对比度低、图像模糊^[15], 单一 7 × 7 卷积非线性表达能力不强, 特征提取效果不佳, 为了增强网络的表达能力, 提升网络模型对红外图像的分类性能, 本文设计了如图 3 所示的多分支同构结构。该结构由 16 组卷积构成, 即 16 条独立路径的同构结构。每组第一层卷积的步长为 2,

填充为 1, 其余层卷积步长为 1, 无填充。

4.2 池化层改进

最大池化能够提取图像的主要特征, 平均池化则能很好地保留图像的背景信息。传统残差网络使用单一的最大池化或平均池化, 这样不可避免地丢失了部分有用信息。为了尽量避免有用信息的丢失, 本文设计了如图 3 中 pool1 和 pool2 所示的池化结构, 将最大池化和平均池化提取的特征进行融合, 这样得到的特征将更丰富多元, 其表达能力也更强。

4.3 残差块改进

4.3.1 多重残差结构

传统残差块由 2 个堆叠的 3 × 3 卷积及恒等映射连接构成, 卷积结构单一, 提取特征方式固定。为增加网络的多样性, 使网络能够提取不同尺度的图像特征, 本文设计了多重残差结构。如图 4 所示。

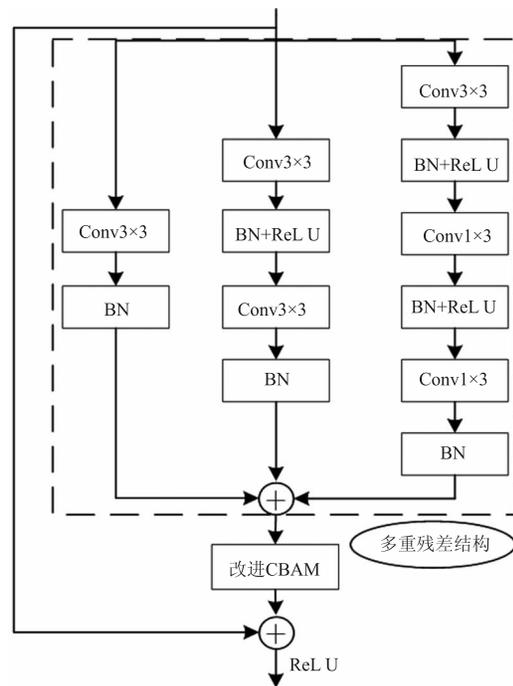


图4 改进残差块示意图

Fig. 4 Improved residual block diagram

多重残差结构,在原结构基础上增加两条并行的特征提取通路:一条通路由 3×3 卷积构成,另一条通路由 3×3 卷积和 1×3 卷积、 3×1 卷积构成的非对称卷积块堆叠而成。这种非对称卷积块抑制了模型的过拟合,增强了模型的非线性扩展性,不但能够提取更多空间特征,而且使提取的特征更加稳定与多元。

4.3.2 改进 CBAM

在数据采集的过程中,不可避免会受到噪声的干扰,而且不同情况下,人体在红外图像中所处位置以及呈现的形态是不一样的。这就要求识别方法能够更多地关注人体形态特征信息,而忽略其他无价值的信息;注意力机制很好地解决了这个问题。

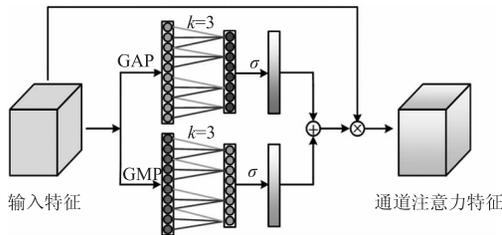


图 5 改进 ECA 模块示意图

Fig. 5 Improved ECA module diagram

Wang 等人^[16]提出 ECA 模块,给定通过全局平均池化(GAP)获得的聚合特征,经过大小为 k 的快速 1D 卷积实现跨通道信息交互,接着使用 Sigmoid 函数生成通道注意力权重,再将该权重与输入特征相乘得到通道注意力特征。但是该模块忽略了目标特征的另一重要表征形式——经过全局最大池化(GMP)获得的聚合特征。因此,本文同时使用 GAP 和 GMP 提取的特征对 ECA 模块进行改进,使网络能够更有效地增强有意义的特征通道。改进 ECA 模块如图 5 所示,其中 1D 卷积核大小 $k=3$, σ 表示 Sigmoid 函数。

ECA 模块使网络能更有效地选择有意义的特征通道,CBAM^[17]中的空间注意力模块则使网络能更多的关注有价值的区域信息。将改进 ECA 模块与空间注意力模块结合,形成改进 CBAM 模块,该模块如图 6 所示,并将其加入多重残差结构中,对残差块完成进一步的改进,示意图如图 4 所示。

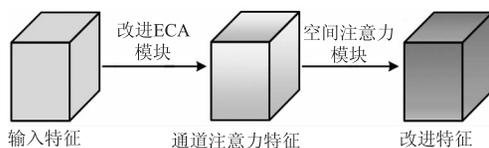


图 6 改进 CBAM 模块

Fig. 6 Improve CBAM module

5 实验结果与分析

5.1 实验数据

实验环境为室内普通办公房间,实验数据来自三位不同身高、体态的实验人员(两男一女);考虑季节对实验结果的影响,我们选择最具代表性的夏季与冬季进行数据采集;跌倒、站立、静坐、平躺、行走、站到坐以及坐到躺等 7 种行为的示意图如图 7 所示。

本文将拼接后得到的 6930 张红外图像按 2:1 的比例分成训练集与测试集,得到包含 4620 张图像的训练集和包含 2310 张图像的测试集。为了增加训练集的数量,减少模型的过拟合,采用水平翻转的方法将训练集扩充一倍,最终得到包含 9240 张图像的训练集和包含 2310 张图像的测试集。各个行为的实验次数、拼接后图像数、水平翻转前和水平翻转后训练集图像数、测试集图像数如表 2 所示。

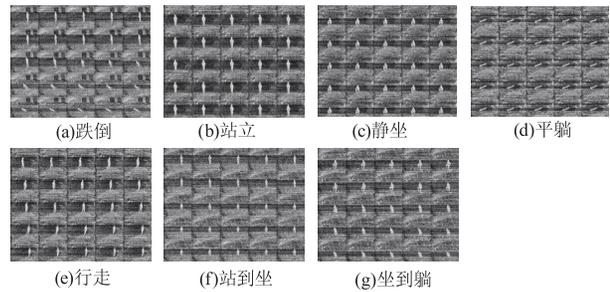


图 7 各种人体行为示意图

Fig. 7 Diagrams of various human behaviors

表 2 各行为相关图像数

Tab. 2 The number of images related to each behavior

类型	实验次数	拼接后图像数	水平翻转前训练集图像数	水平翻转后训练集图像数	测试集图像数
跌倒	87	990	660	1320	330
站立	80	990	660	1320	330
静坐	80	990	660	1320	330
平躺	80	990	660	1320	330
行走	80	990	660	1320	330
站到坐	91	990	660	1320	330
坐到躺	109	990	660	1320	330
7 种行为合计	607	6930	4620	9240	2310

5.2 环境与设置

实验运行环境:Debian 操作系统,Intel Xeon E5-2640 v4 处理器,16G 内存,NVIDIA GTX1080Ti GPU。使用 Pytorch 深度学习框架搭建网络,网络的

损失代价函数选择交叉熵损失函数;采用 Adam 优化方法,学习率设置为 0.001。

5.3 实验结果与对比分析

为了使实验结果更加可靠,将改进 ResNet-18 网络训练三次,每次训练 300 轮,取最优参数进行测试,结果如表 3 所示。

表 3 混淆矩阵

Tab. 3 Confusion matrix

		预测类别						
		跌倒	站	坐	躺	走	站到坐	坐到躺
实际类别	跌倒	330	0	0	0	0	0	0
	站	0	330	0	0	0	0	0
	坐	0	0	330	0	0	0	0
	躺	0	0	0	330	0	0	0
	走	0	1	0	0	329	0	0
	站到坐	0	0	0	0	0	330	0
	坐到躺	0	0	0	0	0	0	330

从表 3 可以看出,在 2310 个行为中,正确识别 2309 次,错误识别 1 次,准确率高达 99.96%,且异常行为——跌倒——均能准确识别,表明本文所提改进 ResNet-18 网络的识别效果良好。此外,误判的图像如图 8 所示。

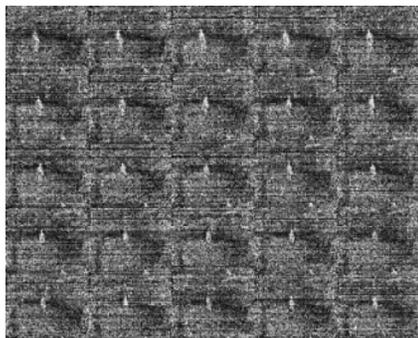


图 8 误判图像

Fig. 8 Image of misjudgment

从图 8 可以看出,人体特征不明显,使得行走图像易与站立图像相混淆。当人体距离传感器较远,且环境温度较高时,可能出现图像中噪声过大导致网络未能准确提取人体特征的情况。然而,行走与站立均为正常行为,误判不会造成严重后果。

在基于红外图像的人体行为识别领域,文献[10]识别的人体行为与本文相近,且其构建的 CNN 网络取得了比传统 LBP-KNN、HOG-KNN、LBP-SVM 等方法更优的识别精度;文献[11]构建的 CNN 网

络是目前人体行为识别效果最好的深度学习方法之一。为了进一步验证本文所提改进 ResNet-18 网络的性能,在本文构建的数据集下,与文献[10]、文献[11]构建的 CNN 网络以及 ResNet-18 网络进行对比,结果如表 4 所示。

表 4 各网络实验结果对比

Tab. 4 Comparison of network experiment results

网络	文献[10]	文献[11]	ResNet-18	改进 ResNet-18
识别率	93.64 %	98.14 %	99.65 %	99.96 %

从表 4 可以看出,与原始 ResNet-18 网络相比,改进 ResNet-18 网络的识别率有所提升,这证明了本文所提方法的有效性;此外,改进 ResNet-18 网络的识别率明显优于文献[10]和[11]中构建的 CNN 网络,这证明了本文所提方法具有明显的优势。

6 结论

本文提出了一种基于改进 ResNet-18 网络的红外图像人体行为识别方法。首先,通过热成像阵列传感器采集跌倒、站立、静坐、平躺、行走、站到坐及坐到躺 7 种行为的红外图像,并对图像进行拼接处理。接着,通过对 ResNet-18 网络进行相应改进,增强了网络的表达能力,降低了有用信息的丢失率,增加了网络的多样性,提升了网路的特征提取能力。经实验表明,改进 ResNet-18 网络识别率达到 99.96%,不但高于传统的 ResNet-18 网络,而且明显优于基于红外图像的其他网络,取得了理想的识别效果。

参考文献:

- [1] Li Junguo, Zhou Shuren, Cai Biye. Human action recognition based on scene-part [J]. Measurement & Control Technology, 2020, 39(2): 104 - 108. (in Chinese)
李俊国,周书仁,蔡碧野. 基于场景-部件的人体行为识别方法[J]. 测控技术, 2020, 39(2): 104 - 108.
- [2] Cai Qiang, Deng Yibiao, Li Haisheng, et al. Survey on human action recognition based on deep learning [J]. Computer Science, 2020, 47(4): 85 - 93. (in Chinese)
蔡强,邓毅彪,李海生,等. 基于深度学习的人体行为识别方法综述[J]. 计算机科学, 2020, 47(4): 85 - 93.
- [3] Zhu Y, Guo G. A study on visible to infrared action recognition [J]. IEEE Signal Processing Letters, 2013, 20(9): 897 - 900.
- [4] Hilsenbeck B, Münch D, Grosselfinger A K, et al. Action recognition in the longwave infrared and the visible spec-

- trum using hough forests [C] // 2016 IEEE International Symposium on Multimedia (ISM). IEEE, 2016: 329 - 332.
- [5] Tan Y, Yan W, Huang S, et al. A motion deviation image-based phase feature for recognition of thermal infrared human activities [J]. Engineering Letters, 2020, 28 (1): 48 - 55.
- [6] Liang C, Liu D, Lin Q, et al. Multi-modal human action recognition with sub-action exploiting and class-privacy preserved collaborative representation learning [J]. IEEE Access, 2020, 8: 39920 - 39933.
- [7] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18 (7): 1527 - 1554.
- [8] Gao C, Hauptmann G, Meng D, et al. InfAR dataset: Infrared action recognition at different times [J]. Neurocomputing, 2016, 212: 36 - 47.
- [9] Liu Y, Lu Z, Li J, et al. Global temporal representation based CNNs for infrared action recognition [J]. IEEE Signal Processing Letters, 2018, 25 (6): 848 - 852.
- [10] Akula A, Shah A, Ghosh R, et al. Deep learning approach for human action recognition in infrared images [J]. Cognitive Systems Research, 2018, 50: 146 - 154.
- [11] Gochoo M, Tan T H, Huang S C, et al. Novel IoT-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning [J]. IEEE Internet of Things Journal, 2019, 6 (4): 7192 - 7200.
- [12] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 - 778.
- [13] Taramasco C, Rodenas T, Martinez F, et al. A novel monitoring system for fall detection in older people [J]. IEEE Access, 2018, 6: 43563 - 43574.
- [14] Zhong Y, Qiu S, Luo X, et al. Facial expression recognition based on optimized resNet [C] // 2020 2nd World Symposium on Artificial Intelligence (WSAI). IEEE, 2020: 84 - 91.
- [15] Xi Lin, Sun Shanyuan, Li Linna, et al. Depth estimation from monocular infrared images based on SVM model [J]. Laser & Infrared, 2012, 42 (11): 1311 - 1315. (in Chinese)
席林, 孙韶媛, 李琳娜, 等. 基于 SVM 模型的单目红外图像深度估计 [J]. 激光与红外, 2012, 42 (11): 1311 - 1315.
- [16] Wang Q, Wu B, Zhu P, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11534 - 11542.
- [17] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module [C] // Proceedings of the European conference on computer vision (ECCV). 2018: 3 - 19.