

基于低分辨率红外传感器的深度学习动作识别方法

张昱彤¹, 翟旭平¹, 聂宏²

(1. 上海大学 特种光纤与光接入网重点实验室, 上海 200444; 2. 美国北爱荷华大学 技术系, 爱荷华州 锡达福尔市)

摘要: 近年来动作识别成为计算机视觉领域的研究热点, 不同于针对视频图像进行的研究, 本文针对低分辨率红外传感器采集到的温度数据, 提出了一种基于此类红外传感器的双流卷积神经网络动作识别方法。空间和时间数据分别以原始温度值的形式同时输入改进的双流卷积神经网络中, 最终将空间流网络和时间流网络的概率矢量进行加权融合, 得到最终的动作类别。实验结果表明, 在手动采集的数据集上, 平均识别准确率可达到 98.2%, 其中弯腰、摔倒和行走动作的识别准确率均达 99%, 可以有效地对其进行识别。

关键词: 动作识别; 双流卷积神经网络; 低分辨率红外传感器; 深度学习

中图分类号: TP319.4 **文献标识码:** A **文章编号:** 1001-8891(2022)03-0286-08

Deep Learning Method for Action Recognition Based on Low Resolution Infrared Sensors

ZHANG Yutong¹, ZHAI Xuping¹, NIE Hong²

(1. Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai University, Shanghai 200444, China;

2. Department of Technology, University of Northern Iowa, Cedar Falls 50614-0507, USA)

Abstract: In recent years, action recognition has become a popular research topic in the field of computer vision. In contrast to research on video or images, this study proposes a two-stream convolution neural network method based on temperature data collected by a low-resolution infrared sensor. The spatial and temporal data were input into the two-stream convolution neural network in the form of collected temperature data, and the class scores of the spatial and temporal stream networks were late weighted and merged to obtain the final action category. The results indicate that the average accuracy of recognition can reach 98.2% on the manually collected dataset and 99% for bending, falling, and walking actions, indicating that the proposed net can recognize actions effectively.

Keywords: action recognition, two-stream CNN, low resolution infrared sensor, deep learning

0 引言

随着世界各国人口老龄化问题越来越严重, 在日常生活中, 高龄独居老人极易因为发生跌倒等意外情况却得不到及时的救治而残疾甚至死亡^[1], 因此对室内跌倒检测算法的研究成为了热点。目前, 已有许多相关研究, 根据采集数据的装置分为摄像装置和传感装置, 传感装置可以细分为需穿戴传感器和无需穿戴传感器。在日常生活中, 摄像装置^[2]不仅会暴露用户的日常隐私, 且极易受到光照影响。加速度传感器^[3]等需穿戴传感装置需要老人时刻穿戴在身, 一定程度上也造成了不便, 而压力传感器^[4]、二值传感器^[5]等无需穿戴传感装置也往往受到使用环境的约束和影响,

系统鲁棒性较差。为了减少以上因素的影响, 有人提出使用被动红外传感器采集数据。此类传感器通过接收外界的红外辐射工作, 采集的数据为探测区的温度, 这样既保护了用户的隐私, 不易受光线影响, 也易于安装且不约束老人的行为。

基于此类传感器的动作识别方法主要有传统计算机视觉方法和深度学习方法两种, 传统计算机视觉方法^[6-8]需要人为地提取特征进行可靠的识别, 这就要求对采集的数据有很深的了解, 此类方法同时也需要设置很多的阈值, 阈值的选取一定程度上影响了算法的性能。随着深度学习的提出和实验平台的更新加强, 卷积神经网络(convolutional neural network, CNN)被越来越多的应用在动作识别领域, CNN可以自行从

收稿日期: 2021-04-21; 修订日期: 2021-06-02.

作者简介: 张昱彤(1996-), 男, 江苏盐城人, 硕士研究生, 主要从事基于红外图像的人体动作识别算法研究工作, E-mail: zyt164819285@163.com.

输入集数据中学习特征,免去了人为提取时带来的不确定性。Aparna 团队^[9]基于 LeNet 构建网络,对通过红外摄像头的随机采样采集数据送入网络学习,并与传统方法进行对比,识别率较高。王召军^[10]等人基于 VGGNet 搭建网络,不仅进行了动作识别还进行了身份识别,都取得了较高识别率。在许多方法中,还将 CNN 与长短期记忆模型(long short term memory, LSTM)进行结合搭建网络,也取得了不错的效果。Takayuki 团队^[11]用 16×16 的红外阵列传感器采集数据,将红外帧差图像序列和原始红外图像序列一起放入网络进行学习,并且全连接层后加上了 LSTM 提高准确率。Xiuyi Fan 团队^[12]用多个红外传感器采集数据放入到网络进行训练学习,并对比了门控循环单元与 LSTM、MLP (Multilayer Perceptron) 的结果。Felix POLLA 团队^[13]采用了 C3D 网络对连续帧序列进行处理,同时也加上了 LSTM 网络充分利用了时间关联信息。但是以上方法考虑了动作的时间信息,却没有过多捕捉运动特征,在可见光领域中,双流 CNN 得到了广泛的研究。2014 年 Simonyan 团队^[14]提出了双流 CNN 网络结构,将单帧图像和光流图分别作为空间和时间流一起放入网络训练,同时对时间与空间特征进行提取。Christoph 团队^[15]则提前进行了空间和时间特征的融合,可以保证更好地从空间到时间的映射关系,极大地发展了双流 CNN 网络。

为了结合时空信息和动作特征,本文将双流 CNN 网络结构应用在基于低分辨率红外传感器的人体动作识别中。采用 HEIMANN 型号为 HTPA80x64 dR1L5.0/1.0 的红外阵列传感器采集弯腰、坐下、站起、行走和摔倒 5 种日常动作的温度数据,分别构建空间和时间数据。常见深度学习动作识别方法以单帧 RGB 图像或者连续帧视频图像作为输入,本文方法中以原始温度数据作为输入,为了增加样本多样性,提高模型泛化能力,本文通过随机裁剪、翻转操作进行数据增强。在网络模型方面,首先根据双流的输入尺寸在原始双流 CNN 的基础上调整网络的层次,搭建时间流和空间流网络,接着对双流的输出进行加权融合,得到最终的分类结果。将双流网络分别与单流网络的结果进行比较,并与原始双流 CNN 网络、基于预训练网络迁移学习以及两种手工提取特征方法的识别结果进行对比,以验证所提出方法的性能。

1 动作识别方法设计

本文提出的基于低分辨率红外传感器深度学习动作识别方法分为训练流程和识别流程。训练时,分

别构建单个训练数据样本的空间数据和时间数据。在构建样本空间数据时,首先进行前景提取,本文针对多数前景提取方法采用经验值设置阈值的问题,提出一种基于奈曼-皮尔逊准则设置最佳阈值的前景提取方法,较为完整干净地提取出前景,对各个像素点时域上累加运动帧内的数据,得到累计温度矩阵作为该样本的空间数据。在构建时间数据时,将代表运动过程的 9 个单帧数据按照时序顺序以 3×3 的形式拼接成总温度矩阵,并分别一一对应保存样本的空间数据和时间数据。数据类型转换之后,通过随机位置裁剪固定尺寸、随机水平翻转进行数据增强,提高样本多样性,改进原始的双流 CNN 网络的空间流网络结构,将构建的时空训练数据送入改进的双流 CNN 网络中进行训练,得到训练数据的分类结果,将其与训练数据的标签,即训练数据的真实动作类别进行比较,分类结果的误差由损失函数量化。上述训练过程经不断迭代后,分类结果和数据标签之间的误差将不断缩小,训练误差值每次减少时,进行网络的权值更新,保存误差值最小时的网络模型作为最优模型。在进行识别时,将待识别数据通过以上方法构建空间数据和时间数据之后首先转换数据类型,按照与训练数据相同的输入尺寸进行裁剪预处理操作后送入最优模型中,得到待识别数据的动作识别结果,方法流程如图 1 所示。

2 时空数据的构建

本文基于红外传感器采集到的温度数据,构建时空数据,将原始温度值输入到提出的双流 CNN 网络中完成训练和识别。

2.1 空间数据的构建

首先采取背景减法提取前景,为了解决根据经验值选取阈值^[16]的问题,本文方法基于奈曼-皮尔逊准则选取每帧的最佳分割阈值。设像素点为背景点的状态设为 H_0 ,为前景点的状态设为 H_1 ,该准则是在 $P(H_1/H_0) = \alpha$ 的约束条件下,使得检测概率达到最大。

可以利用拉格朗日乘子 μ 构建目标函数:

$$J = P(H_0/H_1) + \mu [P(H_1/H_0) - \alpha] \quad (1)$$

根据要求,即求目标函数 J 的最小值,将公式(1)转化为积分运算,且由于:

$$\int_{Z_1} P(x/H_0) dx + \int_{Z_0} P(x/H_0) dx = 1 \quad (2)$$

式中: Z_0 和 Z_1 分别为判决区域,公式(1)可写为以下形式:

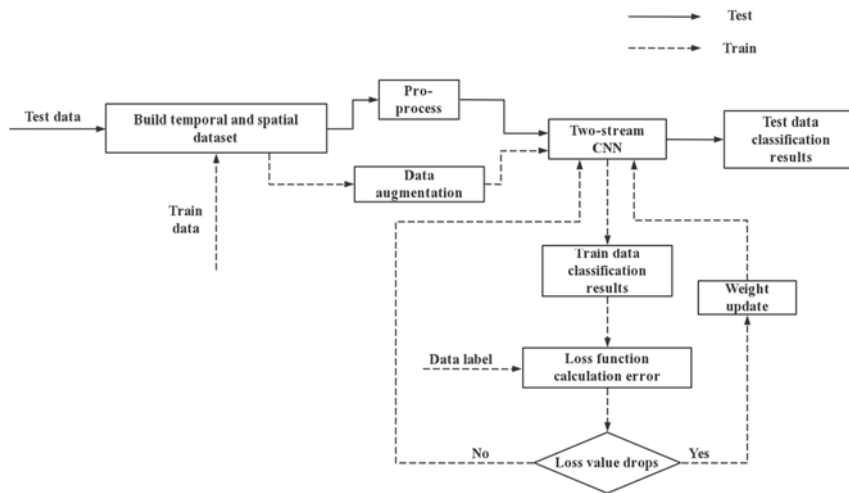


图1 本文动作识别方法流程图

Fig.1 Flowchart of the proposed activity recognition approach

$$J = \mu(1 - \alpha) + \int_{z_0} [P(x/H_1) - \mu P(x/H_0)] dx \quad (3)$$

若要求得公式(3)的最小值,被积函数部分应该取负,可得以下关系:

$$\frac{P(x/H_1)}{P(x/H_0)} \underset{H_0}{\overset{H_1}{\geq}} \mu \quad (4)$$

式中:判决门限 μ 可由约束条件得到:

$$P(H_1/H_0) = \int_{z_1} P(x/H_0) dx = \int_{\mu}^{+\infty} P(x/H_0) dx = \alpha \quad (5)$$

通过每帧数据得到的判决门限 μ 即为每帧的最佳

判决阈值 x_{kth} 。本文采用虚警率 α 为 0.01 约束下获得每帧的最佳判决阈值 x_{kth} , 前景提取公式如下:

$$f(x_k) = \begin{cases} x_k(i, j), & \text{if } x_k(i, j) \geq x_{kth} \\ 0, & \text{else} \end{cases} \quad (6)$$

得到每一帧的前景之后,对各个像素点时域上累加运动帧内的数据,得到累计温度矩阵 $g(x)$, 其二维热图如图2所示。

$$g(x) = \sum_{i=k_s}^{k_e} f(x_i) \quad (7)$$

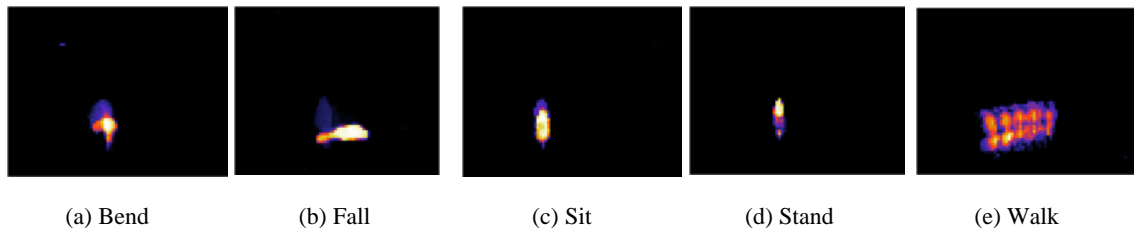


图2 空间数据二维热图

Fig.2 Two-dimension heatmap for spatial data

由图2可以发现,空间数据仅仅展现了运动目标的位置,而不包含位置变化的时序信息。公式(7)中的 k_s 、 k_e 分别代表运动开始和结束帧,由如下算法确定。探测区无人活动时,温度值在时域上波动较小,而探测区中存在活动时,时域温度波动较大。根据这一特点,可以通过提取每帧的最大温度方差得到 k_s 和 k_e 。

a_{ijk} 为第 k 帧的第 (i,j) 个像素点的温度值。第 k 帧每个像素点的方差计算公式为:

$$v_{ijk} = \frac{1}{\tau} \sum_{l=k-(\tau-1)}^k (a_{ijl} - \overline{a_{ijk}})^2, k \geq \tau \quad (8)$$

式中: τ 为滑动窗口大小,本方法通过计算窗口 τ 为 15

帧内的每个像素点的方差来进行人体活动检测。 $\overline{a_{ijk}}$ 为第 k 帧第 (i,j) 个像素点前连续 τ 帧的平均温度。

经过公式(8),可以得到每一帧的温度方差分布矩阵 v_k , 根据之前的分析,通过公式(9)提取每帧的最大温度方差:

$$v_{kmax} = \max(v_k) \quad (9)$$

每帧最大温度分布方差 v_{kmax} 大于等于阈值 v_{th} 时,则表明运动开始,该帧计为 k_s , 否则视为运动结束,计为 k_e 。本文确定阈值 v_{th} 的方法为在无人的环境下,在实验前、实验中和实验后运行装置至最大帧,重复3次,分别计算9组数据的最大温度方差,选取最大

值作为阈值 v_{th} 。为了研究环境温度和传感器安装场景对阈值的影响, 由于实验的限制, 在如表 1 所示的环境温度和场景下进行了对比试验。

表 1 不同环境温度、安装场景下的阈值对比

Table 1 Comparison of thresholds under different ambient temperatures and installation spaces

Ambient temperature	Laboratory	Living room
22°C	1.66	1.65
26°C	1.71	1.73
30°C	1.75	1.76

经对比实验后发现, 相对而言, 传感器安装空间对阈值变化影响很小, 环境温度变化对其影响更大, 根据实验时的环境温度, 最终阈值设为 1.7。

2.2 时间数据的构建

不同于基于视频图像的双流 CNN 识别方法, 本文采用的数据是从红外阵列传感器中获得如公式(10)所示的温度分布数据矩阵:

$$x_{T_k} = \begin{bmatrix} x_{T_{a11k}} & x_{T_{a12k}} & \cdots & x_{T_{a180k}} \\ x_{T_{a21k}} & x_{T_{a22k}} & \cdots & x_{T_{a280k}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T_{a641k}} & x_{T_{a642k}} & \cdots & x_{T_{a6480k}} \end{bmatrix} \quad (10)$$

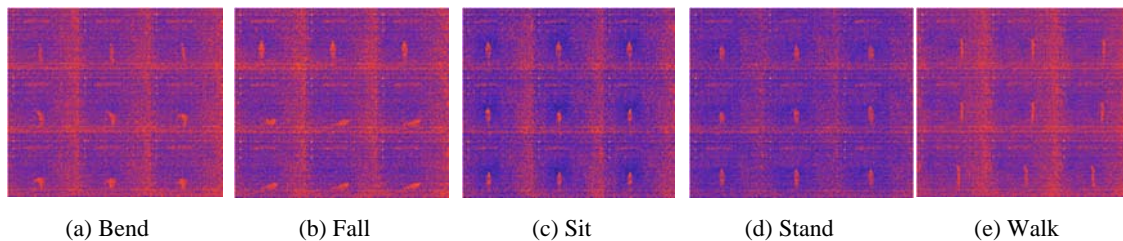


图 3 时间数据二维热图

Fig.3 Two-dimension heatmap for temporal data

3 网络设计

由于样本数量有限, 为了提高模型泛化能力, 在训练前需要进行数据增强, 训练数据在数据类型转换之后, 进行随机位置裁剪固定尺寸、随机水平翻转操作。裁剪操作既可以增加原始样本的多样性, 也可以减少网络的训练计算量, 确定固定尺寸大小时, 需尽可能保留数据中的重要信息, 原始空间数据大小为 64×80 , 时间数据大小为 192×240 , 经过实验对比, 空间数据裁剪为 60×60 的大小输入空间流网络, 时间流网络输入尺寸则为 180×180 。原始双流 CNN 的网络结构^[14]使用的是基本上与 AlexNet 同一种思路设计的 CNN_M 网络结构, 时间流与空间流网络结构均包含 5 层卷积层和 2 层全连接层。由于本文空间数据输

入尺寸较小, 所以空间流网络采用 VGG 网络的设计思路基于原始 CNN_M 空间网络结构进行改进, 最终将时间流与空间流网络的 Softmax 结果进行加权融合, 得到最终的识别结果。

上式表示第 x 个样本第 k 帧的温度分布数据。基于视频图像的双流 CNN 方法往往使用光流图作为时间流网络的输入, 在本文的方法中, 为了减少网络计算量, 提升网络的有效性, 将代表运动过程的 9 个原始单帧数据按照时序顺序以 3×3 的形式拼接成总温度矩阵作为时间流网络的输入。

在构建空间数据时可以求得动作过程的开始帧 k_s 和结束帧 k_e , 由于要选取 9 帧数据, 所以帧间隔 β 为:

$$\beta = \left\lfloor \frac{k_e - k_s}{9} \right\rfloor \quad (11)$$

式中: $\lfloor \cdot \rfloor$ 表示向下取整, 所以得到的 9 帧分别记为 T_{k_s} 、 $T_{k_s+\beta}$ 、 $T_{k_s+2\beta}$ 、 $T_{k_s+3\beta}$ 、 $T_{k_s+4\beta}$ 、 $T_{k_s+5\beta}$ 、 $T_{k_s+6\beta}$ 、 $T_{k_s+7\beta}$ 、 T_{k_e} , 将 9 帧数据以公式(12)的形式进行拼接成像素数为 192×240 的 $h(x)$ 作为时间数据, 其二维热图如图 3 所示。

$$h(x) = \begin{bmatrix} x_{T_{k_s}} & x_{T_{k_s+\beta}} & x_{T_{k_s+2\beta}} \\ x_{T_{k_s+3\beta}} & x_{T_{k_s+4\beta}} & x_{T_{k_s+5\beta}} \\ x_{T_{k_s+6\beta}} & x_{T_{k_s+7\beta}} & x_{T_{k_e}} \end{bmatrix} \quad (12)$$

入尺寸较小, 所以空间流网络采用 VGG 网络的设计思路基于原始 CNN_M 空间网络结构进行改进, 最终将时间流与空间流网络的 Softmax 结果进行加权融合, 得到最终的识别结果。

3.1 空间流网络

空间流输入数据的原始大小为 64×80 , 尺寸较小, 若采用原始网络中尺寸较大的卷积核, 网络的深度会受到限制, VGG 网络^[17]用多个较小的卷积核代替单个较大卷积核, 该方法不仅可以加深网络深度以学习更高级的特征, 也降低了计算复杂度, 因此本文采用此设计思路用 2 个 3×3 的卷积核代替原始较大尺寸的卷积核, 增加网络的深度提取深层空间特征, 本文采用的空间流网络结构如图 4 所示。

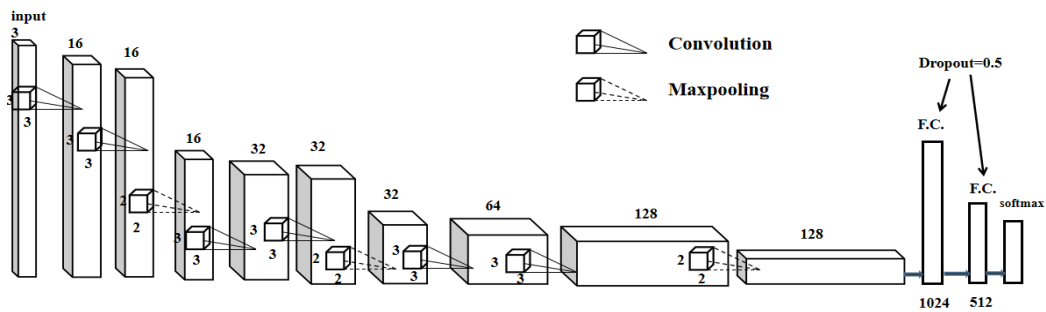


图4 空间流网络结构图

Fig.4 Spatial network architecture

改进后的空间流网络结构主要包括6层卷积层和3层最大池化层以及2层全连接层。空间流的输入通道数为1，由于输入尺寸较小，尺寸较大的卷积核会导致输出的特征图较小，不利于特征的学习，因此卷积层均采用大小 3×3 ，步长为1的卷积核，第1、2卷积层通道数均为16，第3、4卷积层通道数均为32，第5卷积层通道数为64，第6卷积层通道数为128，一定程度上加深网络的深度和宽度，在学习更加丰富的深层特征的同时，降低了参数的个数以及计算复杂度。池化层采用尺寸为 2×2 的窗口，步长为2，以筛

选特征，降低特征图的维度，保存主要特征。最后经过2层全连接层，全连接层中采用 dropout 为0.5的操作防止网络过拟合。卷积层和全连接层均采用整流线性单元 (Rectified Linear Units, ReLU) 作为激活函数。采用 softmax 函数计算空间流网络的最终结果。

3.2 时间流网络

时间流网络基于原始双流 CNN 时间流网络^[14]，为了减少参数量，防止过拟合，在原始网络基础上减少了一层卷积层，时间流网络结构如图5所示。

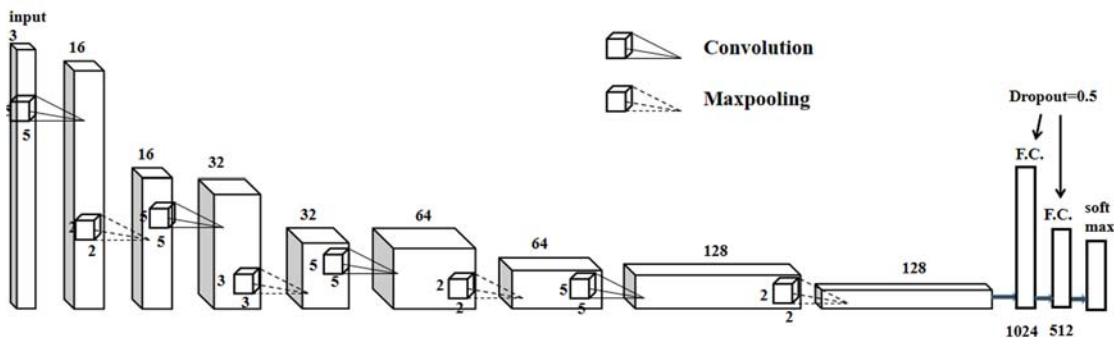


图5 时间流网络结构图

Fig.5 Temporal network architecture

时间流网络结构主要包括4层卷积层、4层最大池化层以及2层全连接层。卷积层均采用尺寸为 5×5 的卷积核，步长为1。第1层卷积层通道数为16，第2层通道数为32，第3层通道数为64，第4层通道数为128，每个卷积层之后均进行最大池化操作，在第1、3、4池化层选用大小为 2×2 的窗口，步长为2，第2池化层采用大小为 3×3 的窗口，步长为3。经过卷积和最大池化操作后经过2层 dropout 均为0.5的全连接层。卷积层和全连接层均采用 ReLU 函数为激活函数，softmax 函数计算时间流结果。

3.3 双流 CNN 网络

最终需要将时间流网络和空间流网络的概率向量经过下式加权融合后得到最终结果：

$$P(w) = b \cdot P_s(w) + (1 - b) \cdot P_t(w), \quad 0 \leq b \leq 1 \quad (13)$$

式中： b 为空间流网络权重； $P(w)$ 为最终的双流概率输出； $P_s(w)$ 为空间流概率输出； $P_t(w)$ 为时间流概率输出，双流 CNN 网络结构如图6所示。

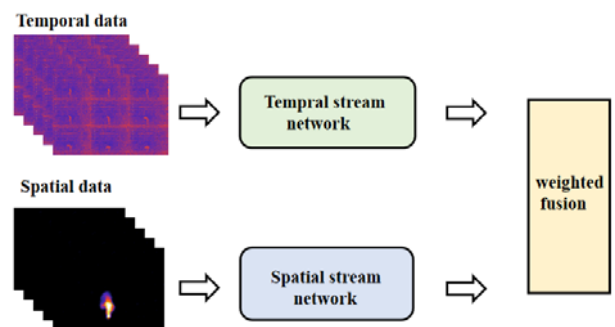


图6 双流 CNN 网络

Fig.6 Two-stream network

训练过程中模型的损失函数采用交叉熵函数进行计算，优化算法采用随机梯度下降（stochastic gradient descent, SGD）算法^[18]，SGD算法进行一次更新操作时，会对每个样本进行梯度更新，样本较多时，不会产生冗余，且速度较快。经过对比试验后，模型的学习率初始化为 0.01，动量为 0.5，不进行学习率的更新。

网络的训练轮数为 100 轮，训练时，每批次从头到尾读取 20 组数据载入内存训练，空间和时间流的数据必须保持一一对应，即每次读取的空间、时间数据均来自于同一样本。一次迭代结束后将训练数据随机打乱再进行下一次的读取数据和训练。所有轮数训练完成后，保存损失值最小的模型作为最终的模型，将测试数据放入该模型得到最终的分类结果。

4 实验与性能分析

4.1 实验平台及数据采集

本文实验基于 Pytorch 深度学习框架和型号为 NVIDIA GeForce GTX 1050Ti 的 GPU，计算框架为 CUDA。采用德国 HEIMANN 型号为 HTPA80x64 dR1L5.0 的热电堆阵列传感器，在校内普通实验室采集数据。室温为 26℃，将传感器置于墙壁，距离地面 2.6m。测试人员共 10 名（7 男 3 女），被测人员在探测区依次完成弯腰、站起、坐下、摔倒和行走 5 种动作，每种动作重复 30 次，最终每个动作获得 300 组数据，一共 1500 组，按照 2:1 的比例划分为训练集和测试集，进行实验，训练集为 1000 组，测试集为 500 组。

4.2 实验结果与对比

本文的空间流网络不同于原始双流 CNN 的空间流网络，基于 VGG 网络的设计思想，采用多个小卷积核代替中的大卷积核，在加深网络的同时，降低计算复杂度。训练集在本文改进的空间流网络与原始双流 CNN 空间流在训练过程中训练误差值对比如图 7 所示。

如图 7 所示，实验证明改进后的空间流网络比原网络收敛速度更快，损失值下降得更低。空间流和时间流两路的输出最终进行加权融合得到最终的分类结果，不同权重的识别准确率如图 8 所示，最终选择权重为 0.5，即平均融合。

为了验证所提出的双流 CNN 的网络性能，将双流 CNN 与空间单流、时间单流进行对比，不同动作的识别率如表 2 所示。

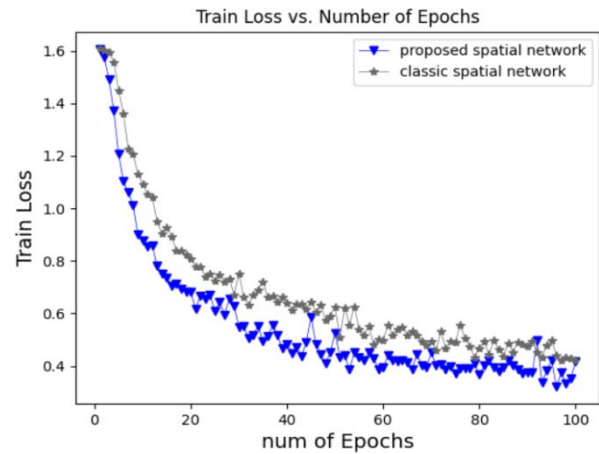


图 7 两种空间流网络训练误差值对比

Fig.7 Comparison of trainloss of two spatial stream networks

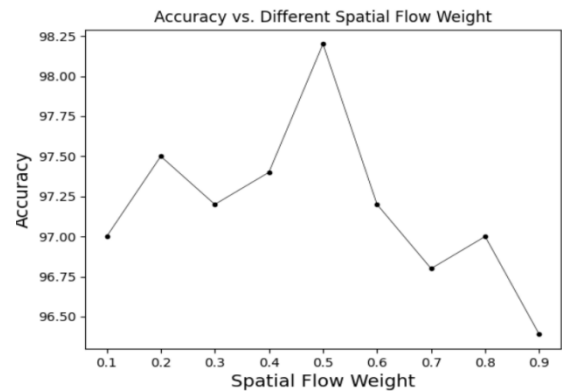


图 8 不同权重的识别准确率

Fig.8 Recognition accuracy of different weights

表 2 单流、双流 CNN 识别准确率

Table 2 Accuracy of one-stream and two-stream

	Temporal one-stream	Spatial one-stream	Two-stream
Bend	97%	88%	99%
Fall	96%	91%	99%
Sit	84%	85%	96%
Stand	91%	81%	98%
Walk	98%	97%	99%
Accuracy	93.2%	88.4%	98.2%

由表 2 可以发现，对于采集的数据，双流 CNN 的准确率较单流 CNN 网络有了较高的提升，弯腰、行走和摔倒动作准确率均达到 99%，由图 9 的混淆矩阵图可以发现，弯腰动作有 1 个测试样本错判成了站起，摔倒动作有 1 个测试样本错判成了弯腰，行走动作有 1 个测试样本错判成了摔倒，且由于传感器放置位置的原因，在坐下和站起的时间数据二维成像图中可以发现，两种动作较大的差别仅在 9 帧数据的中间 2 帧中，其他帧数据的差别很小，而在空间数据的构建过

程中,温度值得到了叠加,呈现出的两个动作变化的趋势差别较为细微,所以相比于其他的动作来说,坐下有4个测试样本错判为站起,而站起有2个测试样本错判为坐下,误判率较高。

为了进一步验证模型性能,将数据在本文模型的结果与文献[14]提出的原始双流 CNN 模型、在 ImageNet 上预训练的 VGG19^[17]和 ResNet^[19]这两种模型上进行迁移学习的表现结果进行对比。预训练的 VGG 和 ResNet 网络输入为 3 通道,输出类型也与本文类别不同,所以仅仅改变了第一层卷积层输入通道为 1,以及最后的输出层为 5。除此之外,还将本文结果与文献[6]以及文献[7]提出的两种人工设计特征方法的识别结果进行了对比,结果如表 3 所示。

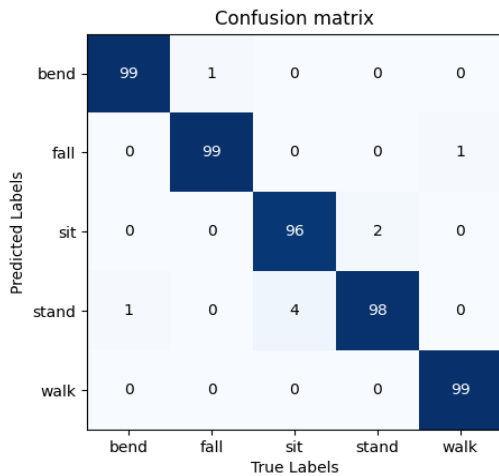


图9 双流 CNN 混淆矩阵图

Fig.9 Confusion matrix of two-stream CNN

表3 不同网络与方法结果对比

Table 3 Comparison of different network architectures and methods

Model	Accuracy
Simonyan ^[13]	95.3%
Pre-trained VGG16	93.4%
Pre-trained ResNet18	92.8%
Pre-trained ResNet50	60.6%
Mashiyama ^[6]	88.9%
Renbin Yang ^[7]	89.6%
Proposed method	98.2%

由表 3 可以看出,采集的数据在本文模型上有着较高的识别率,由于空间数据的尺寸较小,层数较深的网络模型最终的识别准确率反而降低。相比于文献[6]和文献[7]两种人工设计特征的传统方法,本文提出的深度学习方法的识别率有了较高的提升。经过以上的实验对比充分说明本文所提出的双流 CNN 模型能

够较好地地区分所采集的 5 种动作,达到了实验设计初期的要求。

5 结论

本文针对通过低分辨率红外传感器采集的温度数据提出一种基于双流 CNN 的人体动作识别方法。不同于常见深度学习方法中以图像的形式作为输入,本文方法以原始温度数据作为输入。采用背景减法提取前景,基于奈曼-皮尔逊准则确定最佳判决阈值,减少了虚警率,叠加运动帧内的前景数据,将得到的累计温度矩阵作为空间数据。将代表动作整体过程的 9 帧数据,按照时序顺序拼接得到 192×240 的温度数据矩阵作为时间数据。采用数据增强的方式提高样本的多样性,改进了原始的双流 CNN 网络结构,调整超参数,最终通过加权融合的方式融合时空特征得到最终输出的结果。结果表明在本文采用的数据上,5 种动作平均识别准确率为 98.2%,相比于其他网络及人工设计特征方法准确率有了较高的提升,其中弯腰、摔倒和行走这 3 个动作准确率均达 99%。由于传感器放置位置的原因,相较于其他动作,站起和坐下这两种易错分的动作的误判率还是很高。未来,可采用多个红外传感器从多角度采集数据,进一步提升这两种动作的识别准确率。

参考文献:

- [1] 张艳梅, 马晓霞, 赫继梅, 等. 失能老人跌倒的影响因素及长期照护服务需求[J]. 中国老年学杂志, 2019, 39(17): 4355-4357.
ZHANG Yanmei, MA Xiaoxia, HE Jimei, et al. Influencing factors of falls and demand for long-term care services in the disabled elderly[J]. *Chinese Journal of Gerontology*, 2019, 39(17): 4355-4357.
- [2] Hsieh Chungyang, LIN Weiyang. Video-based human action and hand gesture recognition by fusing factored matrices of dual tensors[J]. *Multimedia Tools and Applications*, 2017, 76(6): 7575-7594.
- [3] 王玉坤, 高炜欣, 王征, 等. 基于加速度传感器的人体姿态实时识别[J]. 计算机工程与设计, 2016, 37(11): 3092-3096.
WANG Yukun, GAO Weixin, WANG Zheng, et al. Real-time human activity pattern recognition based on acceleration[J]. *Computer Engineering and Design*, 2016, 37(11): 3092-3096.
- [4] 杜英魁, 姚俊豪, 刘鑫, 等. 基于电阻式薄膜压力传感器组的人体坐姿感知终端[J]. 传感器与微系统, 2020, 39(1): 78-81.
DU Yingkui, YAO Junhao, LIU Xin, et al. Human body sitting posture sensing terminal based on resistive thin film pressure sensor groups[J]. *Transducer and Microsystem Technologies*, 2020, 39(1): 78-81.
- [5] Kobiyama Yuta, ZHAO Qiangfu, Omomo Kazuk, et al. Analyzing correlation of resident activities based on infrared sensors[C]//IEEE

- International Conference on Awareness Science and Technology*, 2015: 1-6.
- [6] Mashiyama S, HONG J, Ohtsuki T. Activity recognition using low resolution infrared array sensor[C]//*IEEE International Conference on Communications*, 2015: 495-500.
- [7] 杨任兵, 程文播, 钱庆, 等. 红外图像中基于多特征提取的跌倒检测算法研究[J]. *红外技术*, 2017, **39**(12): 1131-1138.
YANG Renbin, CHEN Wenbo, QIAN Qing, et al. Fall detection algorithm based on multi feature extraction in infrared image[J]. *Infrared Technology*, 2017, **39**(12): 1131-1138.
- [8] 张昱彤, 翟旭平, 汪静. 一种基于低分辨率红外传感器的动作识别方法[J]. *红外技术*, 2022, **44**(1): 47-53.
ZHANG Yutong, ZHAI Xuping, WANG Jing. Activity recognition approach using a low-resolution infrared sensor[J]. *Infrared Technology*, 2022, **44**(1): 47-53.
- [9] Akula A, Shah A K, Ghosh R. Deep learning approach for human action recognition in infrared images[J]. *Cognitive Systems Research*, 2018, **50**(8): 146-154.
- [10] 王召军, 许志猛. 基于低分辨率红外阵列传感器的人体身份和动作识别[J]. *电气技术*, 2019, **20**(11): 6-10, 26.
WANG Zhaojun, XU Zhimeng. Human identity and motion recognition based on low resolution infrared array sensor[J]. *Electrical Engineering*, 2019, **20**(11): 6-10, 26.
- [11] Takayuki Kawashima, Yasutomo Kawanishi, Ichiro Ide Hiroshi Murase. Action recognition from extremely low-resolution thermal image sequence[C]//2017 14th *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017: 1-6, Doi: 10.1109/AVSS.2017.8078497.
- [12] FAN Xiuyi, ZHANG Huiguo, LEUNG Cyril, et al. Robust unobtrusive fall detection using infrared array sensor[J]. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2017, **5**(4): 194-199.
- [13] Polla F, Laurent H, Emile B. Action recognition from low-resolution infrared sensor for indoor use: a comparative study between deep learning and classical approaches[C]// 2019 20th *IEEE International Conference on Mobile Data Management (MDM)*, 2019: 409-414.
- [14] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos[J]. *Advances in Neural Information Processing Systems*, 2014, **1**(4): 568-576.
- [15] Feichtenhofer C, Pinz A, Zisserman A. Convolutional two-stream network fusion for video action recognition[C]// *IEEE International Conference on Computer Vision and Pattern Recognition CVPR*, 2016: 1933-1941.
- [16] Trofimova A A, Masciadri A, Veronese F, et al. Salice, indoor human detection based on thermal array sensor data and adaptive background estimation[J]. *Journal of Computer and Communications*, 2017, **5**(4): 16-28.
- [17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//*ICLR*, 2015: 1-14.
- [18] WANG L, ZANG J, ZHANG Q, et al. Action recognition by an attention-aware temporal weighted convolutional neural network[J]. *Sensors*, 2018, **18**(7): 1979.
- [19] HE K, ZHANG X, Ren S, et al. Deep residual learning for image recognition[C]//*CVPR*, 2016: 770-778. Doi: 10.1109/CVPR.2016.90.