第二章 人体行为识别和3D姿态估计研究概述

2.1 人体行为识别相关研究

2.1.1 传感器数据采集

2.1.2 特征提取与特征表示

2.1.3 人体行为识别方法的研究与比较

　 2.1.4 人体行为识别面临的挑战

2.2 3D人体姿态估计相关研究

2.2.1 视觉数据采集

2.2.2 特征提取与特征表示

2.2.3 3D姿态估计方法的研究与比较

2.2.4 3D姿态估计面临的挑战

2.3 本章小结

第三章 基于长短期记忆网络和卷积神经网络的人体行为识别

3.1 循环神经网络

3.2 长短期记忆网络

3.3 卷积神经网络

3.4 基于长短期记忆网络的人体行为识别模型

　 　3.4.1 模型整体框架

　　 3.4.2 模型结构设计

3.4.3 模型训练过程

3.5 基于卷积神经网络的人体行为识别模型

　 　3.5.1 模型整体框架

　　 3.5.2 模型结构设计

3.5.3 模型训练过程

3.6 实验过程与分析

　　 3.6.1 实验数据集

　　 3.6.2 实验环境与设置

3.6.3 评估方法

　　 3.6.4 试验结果与分析

3.7 本章小结

第四章 基于深度卷积和长短期记忆神经网络的行为识别

4.1 基于深度卷积和长短期记忆神经网络的人体行为识别模型

　 　4.1.1 模型整体框架

　 　4.1.2 模型结构设计

4.1.3 模型训练过程

4.2 实验结果与分析

　 　4.2.1 实验数据集

　　 4.2.2 实验环境与设置

4.2.3 评估方法

　　 4.2.4 试验结果与分析

4.3 本章小结

第五章 基于前馈神经网络的3D人体姿态估计

5.1 前馈神经网络

5.2 基于前馈神经网络的3D姿态估计模型

　　 5.2.1 模型整体框架

　 5.2.2 模型结构设计

5.2..3 模型训练过程

5.3 实验过程与分析

　　 5.3.1 实验数据集

　 　5.3.2 实验环境与设置

5.3.3 评估方法

　 5.3.4 试验结果与分析

5.4 本章小结

第六章 总结与展望

6.1 总结

6.2 展望

参考文献

附录

致谢

**2.1 人体行为识别相关研究**

人类行为识别旨在，给出一组关于他或她自己和周围环境的前提下，识别一个人的行为。通过利用从各种资源检索而来的信息，例如环境[90]或穿戴的传感器[91,92]，来完成识别的任务。有些方法已经采用，在不同的身体部位佩戴专用的运动传感器，如腰部、胸部、大腿和手腕[102-104]等，来实现良好的分类效果[93]。传感器的位置在人体行为识别有很重要的作用，例如一些行为，吸烟、饮食、写作、打字、喝咖啡和讲话，位于腰部或裤子口袋里的传感器不能有效地识别这些行为活动，因为这些活动主要涉及手部的运动，如果传感器绑在手臂上，则容易识别这些行为动作。另外一些行为，例如行走、上下楼梯等，则可以很好地被腰部和裤子口袋位置的传感器所识别。以上这些传感器通常都会令普通用户感到不舒适，因此，不能为行为监测提供长期的解决方案。

智能手机正在为以人为中心的应用带来新的研究机会，用户是场景信息丰富的数据来源，而手机是第一手的感应工具。最新的手机设备都带有嵌入式内置传感器，如麦克风、双摄像头、加速计、陀螺仪等。使用带有惯性传感器的智能手机，是人体行为识别的替代解决方案。这些大众化的设备为自动且全面地监控日常生活的行为（ADL）提供了一个灵活可负担得起同时提供电话服务的一体化解决方案。因此，在过去的几年中，一些研究工作提出，使用智能手机来实现理解人体行为。例如，采用Android智能手机上的嵌入式三轴加速度计来进行人体行为识别的始作俑者之一[94]；以及其他相关的报道[95,96]。

本节人体行为识别相关研究的内容包括：数据采集、特征提取、特征表示、识别方案与方法、以及当前人体行为识别面临的一些挑战和难题等，以下从各个方面详细描述。

**2.1.1 传感器数据采集**

人体行为识别实验中用到的数据集主要是利用智能手机上相关的加速度计和陀螺仪传感器收集而来。在数据采集的时候，一般手机会绑在收集人员或者实验志愿者的腰部或腿部。加速度计和陀螺仪信号已经被很多的研究和应用报道，对人体识别相关任务非常高效。因为，加速度计信号是一个三轴加速度计测定手机相对地球表面运动方向的改变值，由重力和身体运动部分组成；而陀螺仪信号是手机围绕某个轴向的旋转角速率值。

**2.1.2 特征提取与特征表示**

从传感器中收集而来的信号数据常常伴有噪音，会影响行为识别的效果，所以一般要进行数据预处理，清除和过滤冗余和无用的特征或数据。此外，数据分割的窗口大小也会影响行为识别的效果。对于简单的行为活动，如散步、慢跑、骑车[105-106]，2至5秒的小窗口就能够有效地识别。本实验中的数据集通过应用噪声过滤器清理掉不需要的特征之后，这些数据被划分成2.5秒时间左右大小的滑动窗口，且滑动窗口之间的数据实现50%重叠(128读数/每窗口)。特征提取是在时域和频域上进行的，特征向量从采样窗口获得。数据集中的每一行，即每个滑动采样窗口，共有561个具有时域和频域的特征向量、用户的行为标签和主题等。人体行为识别相关研究[100]中使用的标准度量，如均值、相关系数、信号值域和自回归系数[101]等将被用来作为特征映射和表示。详细情况见附录中的表1，运用在时域和频域信号特征的所有度量。为了简化性能评估，数据集被随机划分为两组，其中70％的数据被选择用于训练，其余30％用于测试。

**2.1.3 人体行为识别方法的研究与比较**

限制性Bozltman机器（RBM），即一种特定形式的对数线性马尔科夫随机场（MRF）[111]已经被提出，作为一种DNN技术来为行为识别[107]提取特征，它在第一级网络使用高斯可见单元，并以监督的方式训练网络。

**2.1.4 人体行为识别面临的挑战**

由于

1. **基于长短期记忆网络和卷积神经网络的人体行为识别**

神经网络有强大的模式分类能力，并且它是深度学习技术的基础。本章将在3.1节介绍浅层RNN的基本结构和相关内容，3.2节将简单介绍LSTM架构的基本内容以及工作原理，CNN的相关内容和网络结构在3.3节揭晓，3.4节和3.5节分别介绍实验中用到的LSTM人体行为识别模型和CNN识别模型，3.6节为实验详细的过程和相关分析结果，3.7节是最后的总结。

**3.1 循环神经网络**

RNN只要是针对时序数据特别设计的结构，在每个神经单元中都使用循环连接，每一个神经元的激活函数会将激活值返回自身并带有一个权重和时间延迟，即提供一个关于之前激活值的记忆，而这种机制能够学习到数据序列的时间动态。如图1所示，单个循环单元的一个表示。

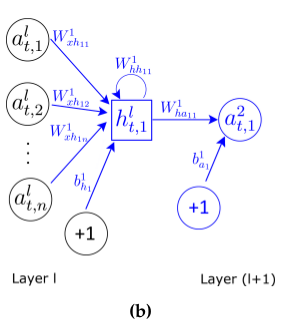


图1带有两个密集层的RNN

L+1层单元的激活和隐含值在相同的时间t计算而来

给定一个长度为T的时间序列输入，即t时刻l隐含层上的i单元的激活值alt,i，之后被RNN映射成隐含值序列，并通过递归公式迭代输出激活序列，递归公式如下所示：

(1)

其中，σ是非线性的激活函数，是隐含偏差向量，w是权重矩阵，是输入层至隐含层的权重矩阵，而是隐含层至隐含层的权重矩阵。以上这些循环单元的激活由以下公式定义：

(2)

其中，表示隐含层到激活层的权重矩阵，表示激活偏差向量。

虽然RNN等网络有图灵可计算性[114]，因而原则上适合学习序列。不过，它们所含有的特殊记忆机制会让它们在处理真实世界的序列问题时，学习会变得相当有挑战性[115]。

**3.2 长短期记忆网络**

LSTM循环网络是由RNN与记忆单元拓展而来，与循环单元不同，LSTM神经网络包含一个记忆单元，存储和输出信息，可以在时间序列任务中建模时序依赖，从而使得长时间的时序关系学习变得相对容易。LSTM网络使用“门”的概念，门是一种基于输入组件乘积的机制，而输入定义了每个记忆单元的行为。LSTM根据门的激活，不断更新神经单元的状态。传输到LSTM的输入被传播至不同的门，然后由门来控制在单元记忆上执行哪类操作，这类操作包括：写(输入门)、读(输出门)和重置(遗忘门)。LSTM神经网络单元的激活与RNN计算过程类似，每个LSTM单元隐含值ht的运算在每t时间步进行更新。LSTM层更新的向量表示如下：

(3)

(4)

(5)

(6)

(7)

其中，i、f、o和c分别是输入门、遗忘门、输出门、和神经单元激活向量，所有这些与向量h的尺寸大小一致，共同定义隐含值。σ代表非线性函数，at是在t时间传播到记忆单元层的输入，和都是权重矩阵，这些权重通过下表可以知道它们之间代表的关系。例如，Wai 是输入层到输入层的门矩阵，Whi是隐含层到输入层的门矩阵，和bo都是偏差向量。

**3.3 卷积神经网络**

虽然，无论是RNN还是LSTM神经网络，都可以接收原始的传感器信号作为输入。然而，

作为深度神经网络的一种，CNN[112]通过叠加卷积操作，可以作为特征提取器，抽取层次分明的抽象特征。这种模型能够自动地学习多重特征层，即特征表示学习。

**3.4 基于长短期记忆网络的人体行为识别模型**

**3.4.1 模型整体框架**

**3.4.2 模型结构设计**

**3.4.3 模型训练过程**

**3.5 基于卷积神经网络的人体行为识别模型**

在特定的应用程序中，手工设计特征需要特定的领域知识[107]。这个问题不是人体行为识别所独有，在其他领域已经被研究透彻。比如在图像识别[108]中，手写字识别要提取不同类型的特征，这与人脸识别不同。近些年来，由于硬件计算处理能力的进步，深度学习技术等到大力发展，并在诸多识别任务[109,110]中实现成功应用。本研究中，我们提出了基于CNN方法来识别不同场景下的行为。在人体行为识别任务中运用CNN，有如下的优势：

1. 局部依赖

CNN可以捕捉行为信号的局部依赖。在图像识别任务中，附近的像素之间通常有一定的强关系。同样，在人体行为识别中，加速度附近的数据很可能有相关性。

1. 尺度不变性

CNN保留特征尺度不变。在图像识别中，训练图像可能有不同的尺度。在人体行为识别中，一个人可以走不同长度的路（例如，不同的运动强度）。

在此，基于CNN方法的人体行为识别模型对于本研究中的贡献如下：

基于CNN方案的模型，再不需要特定领域知识的前提下可以提取人体行为的相关特征。该方法能够捕获活动行为信号的局部依赖性和尺度不变特征，因此，同一行为的变化可以通过已提取的特征有效地捕获。

**3.5.1 模型整体框架**

人体行为识别作为一个分类问题，输入是时间序列的信号，输出是行为类的标签。图1是人体行为识别模型的整体架构，主要分为训练阶段和分类阶段。在训练阶段，特征从原始时间序列数据中提取，之后，这些特征被用来训练一个分类模型。在分类阶段，首先从新来的原始数据中提取特征，然后用训练好的模型预测行为属于哪一类型。

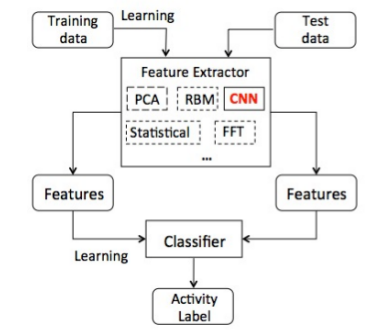


图1 人体行为识别CNN模型整体架构

**3.5.2 模型结构设计**

本节主要介绍基于CNN的人体识别模型结构设计。图2是行为识别CNN模型的主要结构，时间序列的数据被转换成矩阵形式后作为CNN模型的输入。本实验中的CNN模型主要有三种类型的网络层结构：(1) 输入层(即h0i层)，数值由输入数据决定；(2) 隐含层(即hli层)，数值大小由之前l-1层的数据推导而来；(3) 输出层(即hLi层)，最后一层隐含层决定其输出结果。CNN网络模型架构通过调节多组wli,j权重来学习，权重wli,j是输入层hli到其他单元hl+1j的权重，本实验中使用xli表示到uli单元(第l层的第i个单元)的所有输入，yli表示hli单元层的输出。

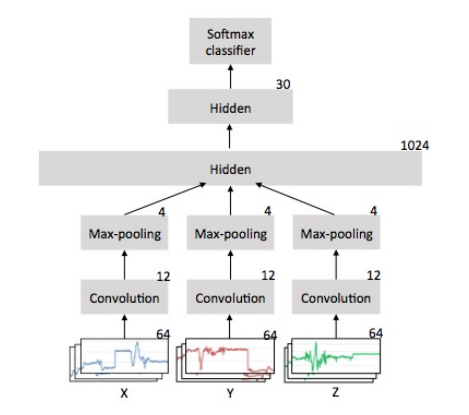


图2人体行为识别CNN模型结构

CNN模型的输入数据维度为140，输出维度为18，两层隐含层即全连接层的维度分别为1024和18，最顶层是Softmax分类层。

1) 卷积层

下文开始描述CNN网络如何捕捉局部依赖关系和固定标量特征。为了捕获数据中的局部依赖关系，可以强制约束相邻单元之间的局部连接。例如，图3中的中间层单元(神经元)只与输入层的本地子集单元连接。从生物学角度来看，视觉皮层中存在有复杂的细胞排列，它们对被称为感受野的输入小区域很敏感，并且这些细胞排列镶嵌在一起生成整个视野。这些滤波器在输入空间是局部的，因此适合于开拓隐藏在数据中的局部相关，所以它通常也被称为局部滤波器，或本地过滤器。对于局部滤波器来说，边的权重连接第i层与第j层单元，wi,j的大小可以通过wa卷积计算来减少，而且wi,j+m=wa，m是局部滤波器的宽度。在图3中，一维向量[w1,w2,w3]代表3个不同线性风格的局部滤波器，wi是连接两层单元的边的权重。卷积操作通过局部子集展开，这种拓扑约束对应于通过稀疏约束来学习权重矩阵，这不仅有利于模型提取局部依赖关系，而且减少了计算的复杂度。如图4所示，多个局部滤波器的输出组成了一个特征映射，在每个时序位置，不同映射中不同类型的单元计算不同类型的特征。

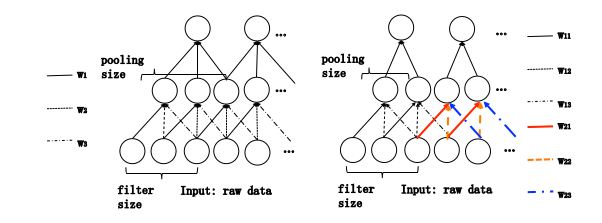
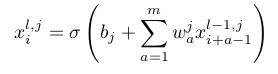


图3 CNN权重共享结构

相同风格的线表示共享权重

为了进一步形成更加丰富的数据表示形式，卷积层有一组多重特征映射构成。如图4所示，两层CNN包含左边3个特征映射(x(0), x(1))，右边2个特征映射。x(0)和 x(1)中的单元输出结果由左侧单元的卷积操作计算得到，即图4中矩形所描绘。假设将一些N单元大小的层作为输入，使用m大小宽度的滤波器w，那么卷积输出则为(N-m+1)个单元，卷积层l的输出为：

 (1)

其中，xl,ji是第i个单元上第j个特征映射的l卷积层输出，σ是一个非线性映射，通常是双曲正切函数tanh(·)。例如在图3中，第一个局部滤波器的首个隐含单元是



而第二个局部滤波器的第二个隐含单元是



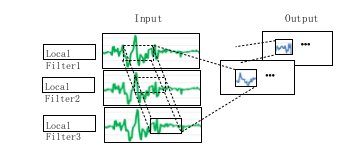
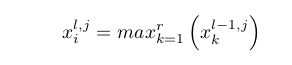


图4 特征映射

输入层的三个特征映射和输出层的两个输出结果

2) 最大池化层

在卷积层中某个特征一旦被检测出来，那么特征提取的位置就显得没有那么重要，只要确保近似的位置与其他特征相关即可。这样一来，通过在卷积层上额外叠加一个最大池化执行层，就可以减少输出的敏感度。保持特征尺度不变是CNN另一个重要的属性，主要是通过最大池化层来实现。在最大池化层，来自卷积层的特征会被分成几个部分，在每个部分都会使用最大值运算操作来计算和输出结果。CNN网络中最大池化层的激活函数为：

(2)

在局部时间窗口工作的滤波器，可以提供有效的方式来代表这些局部特征结构，而它们与整体时间轴的结合，最终可以识别不同的行为活动。

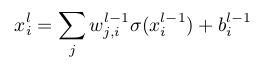
CNN网络结构可以包含一组或多组卷积层与最大池化层组合对，高层可以使用更抽象的滤波器来处理输入中更加复杂的部分。通常情况下，CNN高层由一个或多个全连接层神经网络堆叠而成。使用这些全连接神经网络的目的是，组合不同低层的局部结构来实现最终的分类。在本实验中，我们仅仅使用了一对卷积和最大池化层组合，以及两个常规的全连接神经网络。

**3.5.3 模型训练过程**

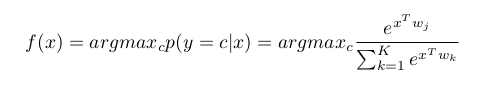
在 CNN模型的训练阶段，通过标准的前馈和反向传播算法来评估参数，可以实现目标函数结果的最小化。

1) 前向传播

首先，前向传播是在含有N个节点的卷积层中展开，如公式(1)所示。如果使用m\*m大小的滤波器w，卷积层的输出尺寸为N-m+1。 卷积层的输出将会作为输入被送入最大池化层，最大池化层计算过程见公式(4)。假如池化窗口的大小为k，而输入层有N个节点，那么输出则为最大池化层的N/k个节点，且每个k大小的窗口通过最大值函数，仅保留单个最大值。最大池化层之后是全连接网络，其中前向传播的公式为：

(3)

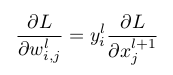
而顶层的Softmax分类函数为：

(4)

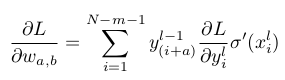
其中c是分类的标签，x是样本特征，y是标签变量，而w是权重向量，K是类的数量。

2) 后向传播

一旦前向传播的迭代完成，将会产生错误的数值，利用损失函数L(这里使用log概率的负值)能够通过梯度下降来更新每个边w。对于全连接层来说，梯度下降的计算过程如下所示：

(5)

其中，非线性映射函数是l+1层的节点j，。而卷积层梯度下降的计算方式为：

(6)

对于最大池化层的梯度，就如之前所描述，最大池化层仅仅通过引入稀疏性来减小卷积层输出的尺寸大小。在前向传播过程中，k大小的窗口被减小为单个最大值，之后仅仅依据该值产生的误差从卷积层开始来进行后向传播运算。由于最大值来自k大小窗口的一个位置，因此来自最大池化层的后向传播误差相对比较稀疏。

**3.6 实验过程与分析**

**3.6.1 实验数据集**

对于人体行为识别研究来说，有一些公开发布的数据集可以使用。比如，Opportunity项目[97]通过72种体态传感器，记录了一整套的日常行为。类似地，其他很多研究工作[98,99]也提供了公共数据集。公共数据集为不同学科和研究领域的学者们提供了可免费使用的资源，极大便利了研究工作的开展。本实验考虑选择两种公开数据集来训练和测试实验中的人体行为识别模型，最终来评估模型的表现效果。这两种数据集分别是Opportunity 行为识别数据集[84]和UCI人体行为识别数据集[85]，它们是不同场景下与人体行为相关的信号，在数据收集过程中都用到了三轴加速度计。这些信号数据通过一系列64个样本大小的滑动窗口分割而成，每个连续分割的窗口间的数据都有50%的数据分别与前后窗口重叠。之后，对这些数据实现0均值和标准方差归一化，归一化后的数据才能作为CNN等深度学习模型的输入。

1. Opportunity行为识别数据集

Opportunity行为识别数据集是关于家庭环境或厨房中早餐场景下的人体活动。该数据集使用了多个穿戴式传感器来收集64Hz频率下每天不同主题活动行为的信号，这些主题活动行为包括：开关冰箱、开关洗碗机、站着饮水、清洁、准备喝咖啡等，一共17种不同的行为，总共6个小时的记录。目前，该数据集可以在UCI机器学习库中下载和使用，而且已经有诸多第三方出版物[86-87]在使用。更重要的是，它还被做为大型公开行为识别竞赛的数据集，有大量的参与者使用不同的算法模型(如附录中表2所示)在运动模型识别中取得了不错的成绩。当然，现在可以从网站[88]下载这个公开的数据集。

1. UCI人体行为识别数据集

本实验一部分数据集来自UCI存储库的公开智能手机数据集。该数据集包含加速度计和陀螺仪的数据，由30名受试者进行6种不同的活动。这些活动包括：走路、上楼、走下楼、坐、站立和躺。21名随机受试者的数据为训练集，剩余9个数据用于测试。原始加速计和陀螺仪XYZ信号要进行均值为零的标准化(即减去均值后除以标准偏差)，对每个行为案例产生一组128个归一化值组成的向量。人体行为识别数据集已经被作为公共使用，并且呈现形式为原始的惯性传感器信号。

**3.6.2 实验环境与设置**

基于CNN架构的算法模型都是在基于Linux CentOS系统的服务器上运行，内存大小为64GB，GPU型号为Nvidia Tesla M40。

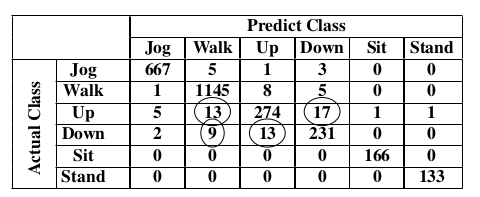
**3.6.3 评估方法**

**3.6.4 试验结果与分析**

实验中的CNN模型由卷积层和池化层组成，卷积核的大小为5，池化尺寸为3，顶层的两个全连接隐含层分别有1024和18个节点，最后的Softmax分类层产生状态概率。具体的人体行为识别分类结果如图5所示，其中坐的行为分类精度为92%，而召回率比较低为88%，尤其是它与站立的行为之间有明显被混淆和错分的情况。

为了更细致化地分析CNN模型的识别结果，所有行为分类的混淆矩阵记录在表3中。混淆矩阵表明，预测误差主要由行走和站立行为之间的混淆导致，因为这两种行为之间相似度比较大[94]。可能由于设备物理定位的问题，导致该行为很难识别和分类。后期的工作将进行必要的优化，从而改善更多类似行为被错分的情况。

表3 行为识别CNN模型在Opportunity数据集上的混淆矩阵



**3.7 本章小结**

在本节中，我们提出了LSMT神经网络和CNN网络结构的人体行为识别模型，并实现了自动提取特征，这些特征基于时间序列，拥有局部依赖性和特征尺度不变性等特点。实验结果表明，CNN模型对于人体识别是一种有效的解决方案。

对于以后研究实验中识别模型的鲁棒性来说，收集更多的数据是很有必要的。后期的优化和识别效果的提升，可能通过使用非监督式的预训练方法或者在多层CNN模型中重复池化操作来实现。

**第四章 基于深度卷积和长短期记忆神经网络的行为识别**

组合CNN与LSTM神经网络的整体框架，已经在基于时序信息建模的语音识别[113]领域取得了相当好的结果。这种类型的架构通过卷积操作，能够在特征提取中捕获时间依赖。

传统的神经网络与DNN一个主要的不同是，DNN可以在网络中有很多层，传统神经网络最多包含三层。DNN的一个关键优势是其输入特征的表示。DNN可以为不同的活动建模，只用相当少的训练数据，它可以与一些隐藏单元共享输入空间的类似部分，同时让其他单元对识别重要的输入特征子集保持敏感。另外，深度学习方法可能有潜力挖掘与人体行为动态紧密相关的特征，即从低层简单的动作映射到高层复杂的行为动态。

本节研究的主要贡献如下：

文中提出了基于深度CNN+LSTM的模型架构，即包含CNN结构和LSTM循环层结构的组合，能够实现自动地学习特征表示，以及在激活函数之间建模时序依赖。

本研究表明，该框架适合基于信号数据类型的人体行为识别任务，包括静态行为、周期行为、间断行为等。

本模型可以在不同的信号模式中实现无缝切换，而且可以结合不同形式的信号提升来提升识别的性能。

实验过程表明，本模型架构只需要借助很少的预处理流程，就可以直接从原始信号数据中展开训练和测试工作。这样从侧面体现了模型的泛化能力强，而且有助于减小工程偏差。

通过比较和讨论文中不同识别模型的结果，可以发现融合深度CNN 和LSTM框架的模型识别的效果比只用CNN结构或者单独的LSTM模型的识别效果好，这体现了深度架构的优势，也为未来开展这方面的研究指明了方向。

**4.1 基于深度卷积和长短期记忆神经网络的人体行为识别模型**

**4.1.1 模型整体框架**

**4.1.2 模型结构设计**

**4.1.3 模型训练过程**

**参考文献**

[84] Chavarriaga R, Sagha H, Calatroni A, et al. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition, Pattern Recognition Letters. 34(15):2033–2042, 2013.

[85] Anguita, Davide, et al. A Public Domain Dataset for Human Activity Recognition using Smartphones, in: ESANN, 2013.

[86] Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional Neural Networks for human activity recognition using mobile sensors. In Proceedings of the 6th IEEE International Conference on Mobile Computing, Applications and Services (MobiCASE), Austin, TX, USA, pp. 197–205, 6-7 November 2014.

[87] Gordon, D.; Czerny, J.; Beigl, M. Activity recognition for creatures of habit. Pers. Ubiquitous Comput. 18, 205-221, 2014.

[88] Opportunity Dataset. 2012. Available online: <https://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition>, 19 November 2015.

[89] Yang, J.B.; Nguyen, M.N.; San, P.P.; Li, X.L.; Krishnaswamy, S. Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition. In Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), Buenos Aires, Argentina, pp. 3995–4001, 25-31 July 2015.

[90] R. Poppe. Vision-based human motion analysis: An overview. Computer Vision and Image Understanding, 108(1-2):4-18, 2007.

[91] P. Lukowicz, J.A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner. Recognizing workshop activity using body worn microphones and accelerometers. Proceedings of the 2nd Int Conference Pervasive Computing, pages 18-22, 2004.

[92] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell, and B.G. Celler. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Transactions on Information Technology in Biomedicine, 10(1):156-167, 2006.

[93] R. Nishkam, D. Nikhil, M. Preetham, and M.L. Littman. Activity recognition from accelerometer data. In Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence, pages 1541-1546, 2005.

[94] J.R. Kwapisz, G.M. Weiss, and S.A. Moore. Activity recognition using cell phone accelerometers. SIGKDD Explorations Newsletter, 12(2):74-82, 2011.

[95] T. Brezmes, J.L. Gorricho, and J. Cotrina. Activity recognition from accelerometer data on a mobile phone. Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living, pages 796-799, 2009.

[96] W. Wu, S. Dasgupta, E.E. Ramirez, C. Peterson, and G.J. Norman. Classification accuracies of physical activities using smartphone motion sensors. Journal of Medical Internet Research, 14(5), 2012.

[97] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, and A. Ferscha. Collecting complex activity data sets in highly rich networked sensor environments. In Proceedings of the 7th International Conference on Networked Sensing Systems, 2010.

[98] E.M. Tapia, S.S. Intille, L. Lopez, and K. Larson. The design of a portable kit of wireless sensors for naturalistic data collection. In Proceedings of PERVASIVE 2006, pages 117-134, 2006.

[99] S. Dernbach, B. Das, N.C. Krishnan, B.L. Thomas, and D.J. Cook. Simple and complex activity recognition through smart phones. In 2012 8th International Conference on Intelligent Environments, pages 214-221, 2012.

[100] J.Y. Yang, J.S. Wang, and Y.P. Chen. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. Pattern recognition letters, 29(16):2213-2220, 2008.

[101] A.M. Khan, Y.-K. Lee, S.Y. Lee, and T.-S. Kim. Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis. In Proceedings of the 5th International Conference on Future Information Technology, pages 1-6, 2010.

[102] Bieber, G.; Fernholz, N.; Gaerber, M. Smart Watches for Home Interaction Services. In HCI International 2013-Posters’ Extended Abstracts; Springer: Berlin/Heidelberg, Germany, 2013; pp. 293-297.

[103] Mortazavi, B.J.; Pourhomayoun, M.; Alsheikh, G.; Alshurafa, N.; Lee, S.I.; Sarrafzadeh, M. Determining the single sest sxis for sxercise sepetition secognition and sounting on smartsatches. In Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Zurich, Switzerland, pp. 33-38; 16–19 June 2014.

[104] Trost, S.G.; Zheng, Y.; Wong, W.K. Machine learning for activity recognition: Hip versus wrist data. Physiol. Meas. 35, 2183-2189, 2014.

[105] Shoaib, M.; Scholten, H.; Havinga, P.J. Towards physical activity recognition using smartphone sensors. In Proceedings of the 2013 10th IEEE International Conference on (UIC) Ubiquitous Intelligence and Computing, Vietri sul Mare, Italy, pp. 80-87, 18–21 December 2013.

[106] Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J. Fusion of smartphone motion sensors for physical activity recognition. Sensors 2014, 14, 10146–10176, 2014.

[107] T. Plötz, N. Y. Hammerla, and P. Olivier. Feature learning for activity recognition in ubiquitous computing. In Proceedings of the Twenty-Second IJCAI Volume Two, pages 1729-1734. AAAI Press, 2011.

[108] D. G. Lowe. Object recognition from local scale-invariant features. In Computer vision, 1999. The proceedings of the seventh IEEE international conference on, volume 2, pages 1150-1157. Ieee, 1999.

[109] U. Bagci and L. Bai. A comparison of daubechies and gabor wavelets for classification of mr images. In Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference on, pages 676-679. IEEE, 2007.

[110] Y. Tang, R. Salakhutdinov, and G. Hinton. Robust boltzmann machines for recognition and denoising. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2264-2271. IEEE, 2012.

[111] G. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. Neural computation, 18(7):1527-1554, 2006.

[112] LeCun, Y.; Bengio, Y. Chapter Convolutional Networks for Images, Speech, and Time Series. In The Handbook of Brain Theory and Neural Networks; MIT Press: Cambridge, MA, USA, pp. 255-258, 1998.

[113] Sainath, T.; Vinyals, O.; Senior, A.; Sak, H. Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. In Proceedings of the 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, pp. 4580-4584, 19–24 April 2015.

[114] Siegelmann, H.T.; Sontag, E.D. Turing computability with neural nets. Appl. Math. Lett. 4, 77-80, 1991.

[115] Gers, F.A.; Schraudolph, N.N.; Schmidhuber, J. Learning precise timing with LSTM recurrent networks. J. Mach. Learn. Res, 3, 115-143, 2003.

**附录**

表 1. 特征向量计算的度量

|  |  |
| --- | --- |
| 函数 | 描述 |
| mean | 平均值 |
| std | 标准差 |
| mad | 绝对值中位数 |
| max | 数组最大值 |
| min | 数组最小值 |
| sma | 信号大小范围 |
| energy | 均方和 |
| iqr | 四分差 |
| entropy | 信号熵值 |
| arCoeff | 自回归系数 |
| correlation | 相关系数 |
| maxFreqInd | 最大频率分量 |
| meanFreq | 频率信号加权平均 |
| skewness | 偏频信号 |
| kurtosis | 频率信号峭度 |
| energyBand | 频率间隔的能量 |
| angle | 向量夹角 |

表 2 . 基准分类方法在Opportunity数据集上的运用

|  |  |
| --- | --- |
| Method | Description |
| LDA | 线性判别分析。基于正态分布特征的高斯分类器，且所有类具有相同的协方差矩阵。 |
| QDA | 二次判别分析。与LDA类似，这种技术也基于正太分布的特征，但类协方差可能不同。 |
| NCC | 最近邻分类器。测试样本与质心的欧氏距离，每一类样本都用于分类。 |
| 1NN | K近邻算法。计算测试样本和训练样本之间欧氏距离的懒惰算法，最频繁发生的k-最近样本是输出。 |
| 3NN | 与1NN类似，使用3个近邻。 |
| UP | Parma大学提交至OPPORTUNITY竞赛。使用均值、方差、最大值和最小值的模式比较。 |
| NStar | 新加坡大学提交至OPPORTUNITY竞赛。使用单一近邻和归一化数据的kNN算法。 |
| SStar | 新加坡大学提交至OPPORTUNITY竞赛。使用标量数据的SVM算法。 |
| CStar | 新加坡大学提交至OPPORTUNITY竞赛。使用最近邻和支持向量机的融合kNN算法。 |
| NU | Nagoya大学提交至OPPORTUNITY竞赛。使用均值、方差和熵值的C4.5决策树算法。 |
| MU | Monash大学提交至OPPORTUNITY竞赛。决策树嫁接算法。 |
| CNN | Yang[89]等提供的结果。对主题1、2、3使用平均性能计算的值。 |