南京师范大学





多通道卷积神经网络人体运动识别研究

研究生:	黄文博
指导教师:	张雷 副教授
培养单位:	电气与自动化工程学院
一级学科:	控制科学与工程
二级学科:	控制理论与控制工程
完成时间:	2022年5月17日
答辩时间:	2022年5月22日

姓名: 黄文博 学号: 191802024

学位论文独创性声明

本人郑重声明:所提交的学位论文是本人在导师指导下进行的研 究工作和取得的研究成果。本论文中除引文外,所有实验、数据和有 关材料均是真实的。本论文中除引文和致谢的内容外,不包含其他人 或其它机构已经发表或撰写过的研究成果。其他同志对本研究所做的 贡献均已在论文中作了声明并表示了谢意。

学位论文作者签名: 黄文十季 日 期: 2022年6月7日

学位论文使用授权声明

研究生在校攻读学位期间论文工作的知识产权单位属南京师范 大学。学校有权保存本学位论文的电子和纸质文档,可以借阅或上网 公布本学位论文的部分或全部内容,可以采用影印、复印等手段保存、 汇编本学位论文。学校可以向国家有关机关或机构送交论文的电子和 纸质文档,允许论文被查阅和借阅。



摘要

随着普适计算技术和人工智能技术的不断发展,深度学习方法已在基于传感 器数据的人体运动识别任务上取得良好的效果。传感器技术为人们生活带来便利 的同时,也带来巨大的挑战。一方面,随着移动设备的小型化与轻量化,计算资 源变得十分有限,在有限的资源内实现准确度高且快速的动作识别成为许多研究 者的关注点;另一方面,卷积核中的冗余通道占用大量参数且对输出特征表示的 贡献度很低,传统的剪枝方法会直接将这些通道修剪掉,可大量节约参数,但这 种方法无法保证模型的泛化能力。保证泛化能力的同时实现准确且快速的动作识 别,使人体运动识别方法更加贴近人们的实际生活,是普适计算技术领域研究者 们当前的重要研究方向。

围绕人体运动识别的实际运用,本文主要研究冗余通道内蕴含的信息和冗余 通道重新激活的方法等一系列问题,研究内容及创新点如下。

首先,探索冗余通道在输出特征表示上的潜在作用。训练三层卷积神经网络 模型时,通过对冗余通道加入随时间呈指数衰减的高斯噪声,在多个国际上用于 测试模型性能的主流公开数据集进行测试,加入高斯噪声的做法已经略优于当前 的基准线模型,准确率已有小幅提升,这可证明冗余通道对输出特征表示具有潜 在贡献,可通过一系列激活方法进一步发掘。

其次,提出通道激活方法的核心思想——替换卷积核中的冗余通道为决定性 通道,使每个卷积核中大部分通道对输出特征表示均有贡献。将激活信息源设定 在网络内部。为提高通道多样性,进一步建立包括空间位移、跨通道交流和通道 平衡三种完整的内部通道激活方法。这些方法在保持泛化能力的前提下,测试时 取得更好的测试准确度,有效的降低冗余通道在模型中的占比。

最后,受到集成学习方法的启发,扩展通道激活的信息源为网络外部。集成 学习方法可进一步大幅提高通道间的差异,提高通道多样性。在测试时,外部通 道激活方法取得更进一步的提升,在三个数据集上的分类准确率、精度、召回率 和 F₁得分已趋于 100%。

在 UCI-HAR、OPPORTUNITY、UniMiB-SHAR、PAMAP2、WISDM 和 USC-HAD 六个数据集上,本文分别对基准线、三种内部通道激活方法和外部通道激活方法进行全面的评估。并将训练好的模型部署在树莓派和安卓智能手机上,进行实际测试以观测推理速度,使结果更加贴近实际。

T

关键词: 普适计算、人体运动识别、卷积神经网络、通道激活

Abstract

With the continuous development of ubiquitous computing technology and artificial intelligence technology, deep learning method has made breakthroughs in human motion recognition tasks based on sensor data. Sensor technology brings convenience to people's life, it also brings great challenges. On the one hand, with the miniaturization and lightweight of mobile devices, computing resources are becoming very limited. Many researchers focus on the realization of high accuracy and fast recognition effect within the limited resources. On the other hand, the redundant channels in the convolution kernel occupy a large number of parameters and contribute very little to the representation of output features. The traditional pruning method will directly trim these channels, which can save a lot of parameters, but this method cannot guarantee the generalization ability of the model. It is an important research direction for researchers in the field of pervasive computing to ensure generalization ability and realize accurate and fast motion recognition and make human motion recognition method closer to people's real life.

In this paper, focusing on the practical application of human posture recognition, the information contained in redundant channels and the method of reactivation of redundant channels are mainly studied. The research contents and innovations of this paper are as follows.

First, the potential role of redundant channels in output feature representation is explored. Training three layer convolution neural network model, based on the redundant channel add Gaussian Noise of the exponential decay over time, in many international public data sets to test the performance of existing test model, the practice of adding Gaussian noise has been slightly better than the current baseline model, accuracy improved slightly, it can prove the redundant channel said potential contribution to the output characteristics, Further discovery can be made through a series of activation methods.

Secondly, the core idea of channel activation method is proposed to replace redundant channels in convolution kernel as deterministic channels, so that most channels in each convolution kernel contribute to the output feature representation. The activation information source is set inside the network to improve channel diversity, and three complete internal channel activation methods including spatial displacement, cross-channel communication and channel balance are further established. These methods can achieve better test accuracy and effectively reduce the proportion of redundant channels in the model under the premise of keeping the generalization ability.

Finally, inspired by the ensemble learning method, the information source activated by the extended channel is external to the network. Ensemble learning method can further greatly improve the differences between channels and channel diversity. In the test, the external channel activation method has achieved further improvement, and the classification accuracy, accuracy, recall rate and F_1 score on the three data sets have approached 100%.

We comprehensively evaluated the baseline, three internal channel activation methods and external channel activation methods on UCI-HAR, OPPORTUNITY, UniMib-SHAR, PAMAP2, WISDM and USC-HAD data sets, respectively. The trained models were deployed on Raspberry Pi and Android smartphones for practical tests to observe reasoning speeds, which makes the results are closer to reality.

Keywords: ubiquitous computing, human activity recognition, convolutional neural network, channel activation

摘要	I
Abstract	II
目录	IV
第1章 绪	论1
1.1 引言	
1.2 研究	现状1
1.2.1	视频图像方法实现人体运动识别2
1.2.2	可穿戴设备上的人体运动识别3
1.2.3	多通道卷积神经网络的人体运动识别4
1.3 本文	主要研究内容5
1.4 各章	节安排5
第2章 数	据采集与预处理7
2.1 公开	人体数据集7
2.1.1	UCI-HAR 数据集
2.1.2	OPPORTUNITY 数据集8
2.1.3	UniMiB-SHAR 数据集9
2.1.4	WISDM 数据集10
2.1.5	PAMAP2 数据集11
2.1.6	USC-HAD 数据集12
2.2 本章	小结13
第3章 标	准卷积神经网络的人体运动识别14
3.1 卷积	神经网络14
3.1.1	卷积神经网络的构成15
3.1.2	卷积神经网络的训练17
3.1.3	应用在人体运动识别领域的卷积神经网络 19
3.2 实验	结果分析
3.3 本章	小结
第4章 内	部通道激活卷积神经网络的人体运动识别23
4.1 内部	通道激活的卷积神经网络23
4.1.1	空间位移
4.1.2	跨通道交流

4.1.3	3 通道平衡	
4.2 实	实验结果及分析	
4.2.1	1 定量对比	
4.2.2	2 通道冗余率	
4.2.3	3 内部通道激活效果	
4.2.4	4 移动端部署	
4.3 本	本章小结	
第5章	外部通道激活卷积神经网络的人体运动证	只别36
5.1 夕	外部通道激活的卷积神经网络	
5.1.1	1 双网络的外部通道激活	
5.1.2	2 多网络的外部通道激活	
5.2 实	实验结果及分析	
5.2.1	1 定量对比	
5.2.2	2 通道冗余率	
5.2.3	3 训练多样性的影响	
5.2.4	4 网络数量的影响	
5.2.5	5 移动端部署	
5.3 本	本章小结	
第6章	总结与展望	
6.1 论	论文主要研究工作总结	
6.2 后	后续工作及展望	
参考文文	〔献	
在读期间	间发表的学术论文及研究成果	53
致谢		

第1章 绪论

1.1 引言

随着近年来物联网技术的更新迭代和传感器技术的飞速发展,在小型可穿戴 设备上便可实现精准的人体运动识别已成为可能,这为人类的生活带来极大的便 利,基于传感器数据的人体运动识别^{[11}已成为普适计算技术^[2]领域内一个新研究 热点。内置加速度计、陀螺仪、磁力计和光感器等传感器的可穿戴设备已具备实 时采集时序信号数据的能力,计算机可以处理这些数据并送入相应的模型进行训 练,将模型部署到可穿戴设备上便可实现人体运动识别。无论是动作追踪^[3]和健 康检测^[4]等传统方向,还是智能家居^[5]和新型游戏设计^[6]等新兴方向,人体动作识 别都已被广泛应用。

早期,隐马尔科夫模型^[7]和支持向量机^[8]等传统监督学习模型多被用于人体运动识别,这些模型的局限性在于需要研究者人工设计数据中如峰值和均值等手工特征,才可以进一步进行分类,会耗费大量人力资源。近期,人工智能特别是深度学习^[9]技术广泛运用于各领域,例如卷积神经网络^[10]和循环神经网络^[11],使深度神经网络在人体运动识别中取得前所未有的突破。得益于深度神经网络可自动从输入数据中提取并学习特征的特点,人力资源耗费巨大的问题得以解决。尽管深度学习方法已在人体运动识别研究中取得重大进展,但仍有一些巨大的挑战,实现准确度高且泛化能力强的轻量化模型便是其中最为重要的之一。

普适计算技术作为一种技术创新,其核心是一种无处不在的计算,计算可以 在任何设备和任何位置以任何格式完成。如图 1.1 所示,在普适计算技术的加 持下,无论何时何地以何种方式,人们都可以通过自由的方式获取信息,生活不 会被计算设备妨碍。这便对部署到移动端的模型提出新的要求,需要在有限的资 源下完成准确度且快速的识别。而当前为了追求识别精度,叠加深度神经网络中 的卷积层已成为多数研究者的选择,与良好的识别精度一同出现的是复杂的模型。 可穿戴设备由于体积限制,各种如内存、计算力和续航等资源都相对有限,所以 复杂的模型并不适用于可穿戴设备。于是,本文的研究重点在于利用简单的模型, 在有限的资源内取得准确且快速的识别效果。

1.2 研究现状

当前,人体运动识别研究有两个主要方向:其一是利用视频与图像实现,通 过对视频与图像中的数据进行研究分析,可识别人类肢体上的关键节点,进而完



图 1.1 普适计算技术^[2]模式下的人体运动识别,人们可以更加自由的获取信息

成对人体运动的识别。其二是利用可穿戴传感器实现,通过处理传感器信号,从 而完成人体运动识别任务。两个方向各有优劣,利用视频与图像是一种非接触的 数据采集方式,人体活动的程度更加自由。但识别精度受光线、衣着等诸多外界 因素影响,此外被采集人员的隐私保护存在问题。而对于可穿戴传感器而言,几 乎不存在隐私保护问题,受外界因素影响小,同时,随着日新月异的设备升级, 小体积的可穿戴设备的功能日益完善,对人体活动的限制已越来越小,对比视频 与图像方法,其计算资源得到节省。

1.2.1 视频图像方法实现人体运动识别

当前,深度学习已在图像分类与识别领域取得突破,也被广泛运用在视频图像的人体运动识别中。视频图像方法实现的人体运动识别可分为单人模式与多人模式,而人体运动识别的思路也可分为自下而上和自下而上。自上而下的方法着重追求良好的精度,所以会优先检测人体再对单人进行运动识别,但速度较慢。例如 2018 年 Kocabas^[12]等人提出的多运动网络,第一部分为提取图片多尺度特征的卷积神经网络;第二部分的两个独立网络分别用于先后检测人体和人体关节点;第三部分的残差网络^[13]依据检测结果对检测到的人体关节点进行聚类,可得到被检测者的人体关节点聚集,这样便可实现人体运动识别。自下而上的方式则为先检测每个人的关节点,随后再与人体关联。由美国卡耐基梅隆大学的OpenPose^[14]项目可知名为部分区域亲和概念。如图 1.2 所示,此模型可实现自下而上的人体运动识别,检测人的关节节点。

而在国内,有许多研究者采用视频图像的方法来实现人体运动识别,例如余保玲^[15]等人和肖志涛^[16]等人。前者的多目标人体骨骼节点检测算法着重与 DeepPose 算法的输入,其输入改进为 Faster RCNN 检测出的含有人体关节点的



图 1.2 视频图像实现人体运动识别示例^[14](OpenPose 实现)

感兴趣区域,从而实现多个目标关节点的检测,完成人体运动识别。后者则提出 一种基于运动学动态图的人体运动识别方法,场景流特征图由 RGB 视频序列和 其对应的深度图序列生成。在此基础上,可计算运动学特征图序列;与此同时, 运动学动态图由经过分层排序池化编码的运动学特征图而成;最终,外观动态图 和运动学动态图会被送入双流卷积神经网络完成人体运动识别。

1.2.2 可穿戴设备上的人体运动识别

如图 1.3 所示,可穿戴设备上的人体运动识别为常见的模式识别与定位问题,浅层的神经网络已取得相当不错的识别效果。Straczkiewicz^[17]等人提出传感器数据与人体运动识别工作契合度高,可以轻松完成人体运动识别任务。Bao^[18]等人手动提取包括域熵、频率、均值和能量四种传感器特征数据,同时送入贝叶斯分类器、决策表、最近邻居和决策树中。Kwapisz^[19]等人同样采用手动的方式对从智能手机采集到的传感器数据进行特征提取,其中包括方差、均值和标准差等,随后采用六种不同的分类器进行人体运动的分类。但是,上述方法均为在特征工程的基础上完成的方法,特定方面的识别性能不佳,例如上楼和下楼动作,无法取得精准的结果。此外,手工提取特征耗费大量人力,需要进行合适的特征选择,这是一个前期相对困难的工作。

随着机器学习技术在多个领域内取得突破,深度神经网络也被广泛应用在人运动识别领域。由于深度神经网络可以自动提取传感器数据中的特征,大大减少人力资源的消耗,直接避免手工提取特征这一复杂前期工作。Ma^[20]等人利用一个三层卷积神经网络对传感器数据进行识别;Jiang^[21]等人则将原始传感器数据进行傅里叶变换,并用相邻排列算法转为二维传感器数据图,随后利用卷积神经网络进行人体运动识别;Ordóñez^[22]等人结合循环神经网络和卷积神经网络并提



图 1.3 可穿戴设备上的人体运动识别示例

出 DeepConvLSTM 模型,卷积神经网络会先提取传感器数据的局部特征,特征 会被送入循环神经网络,处理时序信号,完成人体运动识别;Teng^[23]等人则利用 逐层误差对卷积神经网络进行训练,在人体运动识别上获得更好的效果。Kumar^[24] 等人利用多模态模型,将传感器数据与视频数据进行融合,实现多模态数据的识 别;Kim^[25]等人为完成基于传感器的肢体动作识别提出名为 DeepGesture 的模型, 卷积神经网络用于前期传感器特征的提取工作,随后门控循环神经网络会捕获这 些特征中的长短期依赖信息。

国内亦有诸多研究者在此领域内进行研究,王震字^[26]等人提出基于深度卷积 和门控循环神经网络的人体运动识别方法,结合两者的优点,取得人体运动识别 的良好方法; 佟丽娜^[27]等人利用 LSTM 和惯性传感器,提出基于惯性传感器和 LSTM 网络的人体运动识别方法,由两层的 LSTM 对三维传感器数据完成人体 运动识别。

1.2.3 多通道卷积神经网络的人体运动识别

卷积神经网络的输入数据多为四维张量,图像数据和传感器数据均适用。经 过多次卷积处理后,模型会输出多个特征图,而卷积核中会有多个通道,用于处 理不同模态的数据。实际上,通道对输出的贡献并不是均衡的,有的通道对输出 起决定性作用,而有的通道对输出贡献很小甚至没有。体现在人体运动识别中则 为针对行走动作,加速度计的通道起决定性作用,而陀螺仪的通道作用很小。因 此,当前常规的卷积神经网络中含有相当多的冗余通道。Han^[28]等人提出通道剪 枝,在量化通道的不同贡献程度后,直接将贡献度低的通道修剪掉;Liu^[29]等人 提出模型瘦身方法,可应用于大部分常见的模型;在这些方法的帮助下,整个模 型的参数得到减少,通道的利用率和识别精度也得以提升。但是,剪枝与瘦身的 方法在模型泛化能力方面来说是非常激进的。参数得以减少的同时,泛化能力的 损伤也是极大的,具体表现为针对前进动作识别效果良好,而针对跳跃动作的识 别效果是很差的,这在实际运用中会导致不理想的识别效果。 近年来,为追求良好的识别效果,研究者们更加注重保护泛化能力的研究。 Wu^[30]等人提出的通道注意力模拟人类注意力,训练中会给予决定性通道更多的 注意力;后来出现各种注意力机制的变体,主体思路大同小异。但是,如何处理 冗余通道仍然是个问题,Sahay^[31]等人提出,模式识别与定位问题可模仿人脑海 马体进行工作,即强壮的神经元逐渐取代衰弱的神经元。以此思想为基础,研究 者们提出各种处理冗余通道的方法,例如郑^[32]等人提出的选择性通道训练神经网 络,在训练神经网络时,将后n个冗余通道替换为前n个决定性作用的通道,实 验结果表明,在保留泛化能力的情况下,模型的精度得以提升。带有选择性通道 的卷积神经网络是一种更加接近与人脑的神经网络模型,可以说是一种重新激活 冗余通道的思路。

1.3 本文主要研究内容

由上文重新激活冗余通道的思路作为启发,将激活方法分为内部激活和外部 激活。内部激活的激活信息源来自模型本身,在内部便实现冗余通道的激活工作; 外部激活的激活信息源来自模型外部,模型自身引入外部信息实现冗余通道的再 激活。在此思想的指导下,本文研究思路及内容如下。

首先,本文提出"冗余通道再激活"这一概念,通过为冗余通道加入高斯噪 声来观测模型的性能,证明冗余通道内含有潜在的可利用信息,引出激活信息源 为模型内部的内部通道激活方法;其次,搭建据有内部激活方法的卷积神经网络 模型,加入通道空间位移、跨通道交流和通道平衡模块以提高通道多样性,保护 泛化能力,在公开的人体运动数据集进行测试,并记录模型分类性能;最后,受 到集成学习思想的启发,扩展信息源为模型外部,使用外部通道激活方法,训练 多个不同卷积神经网络并将决定性通道整合入一个网络进行测试。

为证明方法的有效性,测试数据为六个用于测试模型性能、且为权威机构所 采集的公开人体运动数据集,模拟一个相对鲁棒的人体运动识别场景,否则有效 性的证明将失去意义。为使本文结果更加接近实际应用场景,训练完成的模型会 被部署至安卓(Android)智能手机和树莓派(Raspberry Pi)等可携带平台进行 实际测试,记录实际推理时间。

1.4 各章节安排

第一章:绪论。用于介绍研究内容的背景及意义,总结当前可穿戴设备的人体运动识别的研究现状,提出冗余通道再激活的概念,最终是对全文结构与章节的梳理。

第二章,数据采集与预处理。给出本文涉及数据的来源,主要为由权威机构

采集、仅用于测试模型性能的公开人体运动数据集。对每个数据集进行必要的介绍,并处理数据为适合送入卷积神经网络的形式。

第三章,标准卷积神经网络的人体运动识别。本章首先介绍常规的卷积神经 网络,研究其原理及训练过程并分析其结构。最终会利用公开人体运动数据集对 模型进行训练,实现标准卷积神经网络的人体运动识别。

第四章,内部通道激活卷积神经网络的人体运动识别。本章详细说明信息源 为网络内部时的冗余通道再激活,对此方法进行详细的研究,并提出三种提高通 道多样性保护泛化能力的方法。对比内部通道激活卷积神经网络与基准线模型在 公开人体运动数据集上的性能,同时提供在移动端的实际测试结果。

第五章,外部通道激活卷积神经网络的人体运动识别。本章指出内部通道激活仍可进一步改进以提高通道多样性。借鉴集成学习的思想,提出以外部模型作为信息源来进行冗余通道再激活的方法,在此基础上,进行与基准线模型的性能对比以及移动端的实际测试结果。

第六章,总结与展望。总结本文的工作,进一步展望此研究可改进的部分。

第2章 数据采集与预处理

明确数据来源可以更好的构建深度学习网络模型,主要使用包括 UCI-HAR^[33]、OPPORTUNITY^[34]、UniMiB-SHAR^[35]、WISDM^[36]、PAMAP2^[37]和 USC-HAD^[38]数据集,本章介绍这些数据集的特征及处理方法。如图 2.1 所示,首先 是一些内置传感器的可穿戴设备进行数据采集。其次对数据进行包括标准化的预 处理,目的是消除不同类型传感器之间量程间存在的差异。进而对数据进行采样, 传感器数据会延时间序列被切割为数据样本。最终,按照 70%训练集、10%验证集 和 20%测试集对样本进行划分。



图 2.1 数据的采集与预处理流程

2.1 公开人体数据集

2.1.1 UCI-HAR 数据集

为评估各种机器学习算法,来自加州大学欧文分校的研究团队挑选 19 岁至 48 岁的 30 名志愿者来采集数据建立此数据集。每位志愿者都将一部三星盖乐世 S2 (Samsung Galaxy S2)手机绑在腰间并要求在监督下进行 6 种不同的日常生 活活动,包括上楼、下楼、走路、平躺、站立和静坐。传感器信号由三轴加速度 计和陀螺仪以 50Hz 频率采样。

为在完成采集后进一步降低噪声,实验人员用截止频率为 20Hz 的三阶低通 巴斯沃德滤波器和中值滤波器对传感器信号进行预处理。由于人体动作蕴含的能 量近乎 99%都包含在 15Hz 以下,20Hz 足以保证捕获人体动作,而低通巴斯沃 德滤波器可以很好的分离加速度传感器信号中的人体动作分量和重力分量。同时, 实验人员所采用的滑动窗口重叠率为 50%,宽度为 128,原因有三点。首先,一 般人步行的节奏范围是 90 步每分钟到 130 步每分钟,也就是说最慢时完整的步 行周期为 1.5 步每分钟;其次,数据集照顾到包括残疾人和老年人在内的行为迟

动作类型	占比(%)
走路	18.75
上楼	18.75
下楼	18.75
静坐	15.63
平躺	15.63
站立	15.63

表 2.1 UCI-HAR 数据集的样本占比

缓人群。通过大量比对后,研究人员设定这类人群的最低速度相较于一般人慢 50%;最后,在频域中映射时,信号还需要通过快速傅里叶变换(Fast Fourier Transform, FFT),针对两个矢量的幂优化是必要的。UCI-HAR 数据集中每个样 本尺寸为(128,9),第一维为滑动窗口宽度,而第二维则代表传感器的9个轴,包括加速度计(x轴、y轴、z轴)、角速度计(x轴、y轴、z轴)和去除重力分量的加速度计(x轴、y轴、z轴)。样本类型如表 2.1 所示。

2.1.2 OPPORTUNITY 数据集

此数据集由苏塞克斯大学的 Roggen 等人在一个装有大量传感器的场景下采 集而获得,该场景包含 15 个无线和有线网络传感节点,具体来说为有 10 种模式 的 72 个传感器, OPPORTUNITY 数据集的子集包括未分段的传感器记录,通过

动作类型	占比(%)	动作类型	占比 (%)
拨开关	0.78	打开冰箱	1.6
用杯子喝水	1.07	关闭冰箱	0.79
清洁桌子	1.23	打开洗碗机	1.85
打开抽屉1	1.09	关闭洗碗机	1.32
关闭抽屉 1	0.87	开门1	1.87
打开抽屉 2	1.64	关门1	1.26
关闭抽屉 2	1.69	开门 2	6.15
打开抽屉 3	0.94	关门 2	1.54
关闭抽屉 3	2.04	NULL	15.63

表 2.2 OPPORTUNITY 数据集的样本占比

将传感器节点放置在 12 个不同的身体部位,从4 名受试者中收集,每个受试者 被要求在早餐场景中进行 17 种类型的活动,如开灯、关灯、开冰箱和关冰箱等, 重复运行 20 次,采样频率为 30Hz。

在采集过程中,受试者做相应动作的时间没有限制,动作的描述相对宽松, 例如吃早餐和泡咖啡,受试者共有5个回合的动作,在其中之一的演示回合重复 20次已被定义好的动作。在采集过程中,采集者们结合时间轴,对传感器数据标 注,每个时间点的动作数据都有对应的标注。如表2.2所示,该数据集的类型占 比不平衡,其中不属于17个分类中任何一个的空类型(NULL)占比高达七成。

在完成数据采集后, 蓝牙加速度计在传输过程中存在一部分数据丢失, 存在 不少残缺数据。为排除这一影响, 受此影响较大的轴上的数据会被移除, 共计 38 轴的数据被移除, 而剩余的 107 个轴上的传感器数据, 若存在缺失点, 其会被替 换为相邻最近的一个完整点数值。数据的切割方式依旧为滑窗, 每个窗口对应一 个样本。由于传感器的频率为 30Hz, 所有本文采用时间长度为 2 秒的滑动窗口, 重叠率为 50%, 即窗口长度为 64, 滑窗步长为 32, 最终处理完成的单个样本尺 寸为 (64, 107), 前 3 个志愿者的数据用作训练, 第 4 人用作测试。

2.1.3 UniMiB-SHAR 数据集

这是由米兰大学的 Micucci 等人收集的一种新型加速度传感器数据集,该数据集是为检测跌倒和监测人类活动而设计的。在 50Hz 的采样率下,研究人员采用安卓系统的智能手机对人体运动数据进行采集,从 30 名年龄在 18 岁到 60 岁之间的志愿者身上收集到全部 11771 个样本。所有样本被分为两个粗粒度类,包括 9 种日常生活活动和 8 种跌倒。

9种日常动作种类繁多,为最大程度减少志愿者在采集中的失误,每个志愿 者会按照制定好的顺序进行两次采集,在第一次采集中,智能手机会放置在志愿 者的右侧口袋中,而第二遍则放置在志愿者的左侧口袋,。这样的选择是因为日 常生活中人们也常把手机放置在这些位置。

摔倒动作是按照鼓掌、摔倒 6 次、鼓掌的顺序进行单独被采集的,其中,为 保护摔倒数据的采集不被影响,开始和结束的鼓掌都为他人完成。从安全考虑, 志愿者身体下方有厚达 15 厘米的床垫。在 6 次摔倒动作中,前 3 次智能手机位 于右侧口袋,而后 3 次智能手机位于左侧口袋。视频图像在此数据集中也被记录 用于辅助研究者确认开始和结束时刻。在进行标注时,研究者选取临近峰值 3 秒 的窗口为数据样本。由于窗口都在峰值附近,随后的窗口发生多个重叠的可能性 大幅提高,故研究者从用分割技术以取代重叠的滑动窗口,这也适合日常行为和 跌倒动作分割开的特性。与 UCI-HAR 数据集类似,采用 3 秒作为窗口大小的原 因时一方面,保证一个窗口包含两步的一个完整不行周期;另一方面,普通人的

动作类型(日常)	占比(%)	动作类型(跌倒)	占比(%)
上楼梯	7.87	向左跌倒	4.49
下楼梯	11.25	向右跌倒	4.47
静坐	1.7	向前跌倒	4.53
跳跃	6.34	向后跌倒	4.34
跑步	16.86	坐空时跌倒	3.69
走路	14.77	采取防护的跌倒	4.11
静坐后起立	1.3	与障碍物接触的跌倒	5.62
站立后躺下	1.83	失去知觉或眩晕	4.36
躺下后起立	2.51	-	-

表 2.3 UniMiB-SHAR 数据集的样本占比

平均步频在 90 步每分钟到 130 步每分钟。数据采集频率为 50Hz,每个样本的尺 寸为 (151,3),151 的窗口长度保证样本串口时间长度为 3 秒,3 为加速度计的 x、y 和 z 轴。最终数据集包含 7795 个日常动作样本,4192 个跌倒样本,共 11771。 样本占比如表 2.3 所示。

2.1.4 WISDM 数据集

无线传感器数据挖掘(WISDM)实验室使用现代移动设备,例如智能手机、 笔记本电脑和音乐播放器来收集这个传感器数据集。29 名志愿者将智能手机放 在自己的前腿上,收集上楼下楼、站立、慢跑、坐着、走路等6种活动。采样频 率设置为20Hz。该数据集含有10981个样本。

数据的收集由 WISDM 团队开发的图形界面应用进行收集,这个应用帮助实 验者记录用户的名字、开始和停止数据收集、并标记正在执行的活动,也可以选 择采集何种类型的传感器数据,采样频率亦可选择。每隔 50 毫秒会收集一次数 据,这样保证每秒可采集 20 个样本,为保证数据集的质量,一名研究者会对数 据的采集进行监督。采集数据时,志愿者被要求在规定时间内完成指定的动作。 数据被分成多个 10 秒的片段,每个部分的持续时间称为示例持续时间,一个 10 秒的间隔时间足够捕捉 6 个动作,6 个动作的占比如表 2.4 所示。步行、慢跑、 上下楼梯的周期可以用峰值之间的时间和加速度值的相对大小来表示。滑动窗口 依旧适用于此数据集的样本采样,沿着时间轴进行采样,样本尺寸为(200,3), 10 秒片段包含的 200 个数值特征,3 代表加速度计的 x、y和 z 轴。

动作类型	占比(%)
行走	37.2
慢跑	29.2
上楼梯	12
下楼梯	10.2
静坐	6.4
站立	5

表 2.4 WISDM 数据集的样本占比

2.1.5 PAMAP2 数据集

该数据集是欧洲环境辅助生活领域的一个项目,旨在帮助老年人,使他们的 生活尽可能健康和舒适。9名受试者参与数据采集过程,进行骑行、跳绳、步行 等18种活动。在100Hz的采样率下,每个受试者分别在手臂、脚踝和胸部安装 3个惯性测量单元(IMU)和一个心率监测器。为了便于比较,采样率被降为33Hz。 整个数据集都公开在互联网上。

三个惯性测量单元含有多个如加速度传感器、陀螺仪和磁力计等微机电系统 传感器,研究者们将传感器置于人体的三个不同位置,受试者胸口装有一个惯性 测量单元和心率计,第二个惯性测量单元装在受试者惯用手的手腕上方,第三个 惯性测量单元置于受试者惯用脚的脚踝上方。在9名受试者(8名男性1名女性) 中,有1名为左利手,剩余为右利手。遵循指定的动作方案(平躺、静坐、站立、 跑步、骑行、行走、熨烫衣服、吸尘器清洁、跳绳、上楼和下楼)可供选择动作

动作类型(日常)	占比 (%)	动作类型(跌倒)	占比 (%)
平躺	5.4	骑行	5.4
静坐	5.4	行走	19.6
站立	5.4	跳绳	3.6
熨烫衣服	5.4	上楼	1.8
吸尘器清洁	5.4	下楼	1.8
跑步	3.6	其他	23.2

表 2.5 PAMAP2 数据集的样本占比

方案(看电视、电脑前工作、开车、叠衣服、清理房间和踢足球),研究者们对 受试者进行数据采集,整个采集时间持续10小时,最终版本的数据集包含12个 不同的人体动作,其占比如表2.5所示。研究者们利用滑动窗口对数据集进行切 割,设定滑动窗口为3.6秒,重叠率为78%,最终样本尺寸为(120,86),第一 维代表滑动窗口的3.6秒,第二维则表示传感器的86个轴。

2.1.6 USC-HAD 数据集

南加州大学人类活动数据集(USC-HAD)由低水平的人类活动组成,被设计为比较机器学习算法的基准,尤其是在医疗保健场景下。共有12项体育活动(躺着、走路、坐在椅子上等),所有数据收集自14名志愿者。采样频率为100Hz。USC-HAD数据集被广泛应用于各种医疗保健应用,如老年人护理和健康监测。研究者们采用一个含有六个自由度的惯性测量单元(MotionNode, MN)放在志愿者前右侧大腿上,专门用于测量人体活动。MN内置一个三轴加速度计、一个三轴陀螺仪和一个三轴磁力计,MN是一个有线设备,通过一根USB线与笔记本电脑相连。在这种有线的采集情况下,采集到的传感器数据不会丢失,保真度也得到很好的保证。采集所用的数据线是软而长的,并不会对数据造成扭曲。

针对用户间的差异,14 名志愿者中包含 7 名男性与 7 名女性,他们的年龄、 身高和体重各异。在采集中,MN 的 x 轴指向地面并于 y 和 z 轴组成的平面垂 直,日常生活中最简单的动作被记录,覆盖绝大部分人的日常活动,仅记录加速 度计和陀螺仪数据,于 5 个不同的地点,每个地点进行时长 6 小时的采集,表 2.6 用于展示不同动作样本占比。在采样阶段,单个样本的尺寸为(512,6),第 一维度的 512 表示滑动窗口长度,第二维度的 6 则代表加速度计和陀螺仪的 x 轴、y 轴和 z 轴。

动作类型(日常)	占比 (%)	动作类型(跌倒)	占比 (%)
向前走	7.6	跳跃	4.9
向左走	10	静坐	2.7
向右走	11.6	站立	2.8
上楼	17.6	睡觉	12
下楼	16.7	电梯上行	5.2
向前跑	6.4	电梯下行	4.9

表 2.6 USC-HAD 数据集的样本占比

2.2 本章小结

本章用于介绍六个公开数据集,同时介绍数据的预处理方式。表 2.7 中展示 六个数据集的基本情况,公开数据集的采集通常有着较为严格的要求。典型的例 子便是 UniMiB-SHAR 数据集,其要求制定精确的动作采集顺序。在每个日常动 作间,受采集者会被要求以拍手鼓掌的方式对日常动作加以分割。标注过程中, 亦有视频作为对照,获得的标注样本精确可靠。

数据集	动作种类数	样本尺寸
UCI-HAR	6	128×9
OPPORTUNITY	17	64×107
UniMiB-SHAR	17	151×3
WISDM	6	200×3
PAMAP2	12	120×86
USC-HAD	12	512×6

表 2.7 数据预处理情况

第3章 标准卷积神经网络的人体运动识别

深度学习算法被用于实现本文的人体运动识别任务,本章详细介绍深度学习 算法中的典型方法——卷积神经网络,并做出具体的原理分析。深度学习方法学 习可以构建多层深度非线性卷积神经网络,卷积神经网络可以自动提取特征,复 杂的模型可以提取到更多有用的信息。最初,神经网络只有很少的层,而当今的 深度神经网络通常多于 5 层,深层的网络甚至可以达到成百上千层。在 2015 年 ILSVRC 图像分类领域的大赛中,深度残差网络^[39]的深度可达 152 层,这已经是 首届冠军亚历克斯网络(AlexNet)^[40]的 19 倍之多。

作为机器学习算法中的一部分,神经网络的本质仍然是学习,学习的目的是确定网络的权重值。在训练过程中,通过大量标注样本的持续训练,深度神经网络得以学习,网络内部的参数得到修改,同时构建损失函数,当损失函数达到局部最优解甚至是全局最优解时,训练完成。随后,研究者们利用训练完成的网络对未经过标注的数据进行识别。

3.1 卷积神经网络

卷积神经网络作为神经网络的一种,其特点是带有卷积结构。在 20 世纪 60 年代,生物学家 Hubel 和 Wiesel^[41]在对猫视觉皮层进行研究时,首次提出感受野 这个概念,这是神经网络的早期基础。

早在 1995 年, Yann Lecun^[42]等人于纽约大学提出完整的神经网络结构,其 主要作用是用来识别支票上的字符,如图 3.1 所示。实现特征提取的是卷积神 经网络内部的卷积层和池化层,最终全连接层和输出层会完成输出结果。值得一 提的是,由于权值共享和局部连接的特性,卷积神经网络的数量可以被有效的减 少,这些



图 3.1 卷积神经网络



图 3.2 卷积操作

特性也可一定程度上缓解模型的过拟合问题。时至今日,深度学习技术蓬勃发展,卷积神经网络也已广泛用于应对图像识别、语言处理和传感器信号定位等挑战。

3.1.1 卷积神经网络的构成

从以上描述可知,一个可以完成分类任务的卷积神经网络包含卷积层、池化 层和全连接层。卷积层。作为卷积神经网络的核心部件,卷积层是整个卷积神经 网络提取特征的关键,卷积层的卷积操作如图 3.2 所示,卷积运算公式为:

$$s_{i,j} = \left(\vec{W} * \vec{X}\right)_{i,j} = \sum_{m} \sum_{n} x(m+i,n+j) \cdot w(m,n)$$
(3-1)

在(3-1)中,卷积核为 \overline{W} ,*为表示卷积运算的符号。图 3.2中可见,输入数据的 尺寸为 4×4,卷积核的尺寸为 2×2,经过卷积核的卷积运算后,输出数据的尺寸 为 2×2。参考卷积运算的原理,输入尺寸为 $w^i \times h^i$,卷积核尺寸为 $w^k \times h^k$,在 经过卷积操作后,输出数据尺寸 $w^o \times h^o$ 为:

$$\begin{cases} w^{o} = \left[\frac{w^{i} - w^{k} + 2p}{S}\right] + 1\\ h^{o} = \left[\frac{h^{i} - h^{k} + 2p}{S}\right] + 1 \end{cases}$$
(3 - 2)

其中,*S*代表卷积步长 (Strides), *p*代表填充,图 3.2 中的步长为1,填充为0, []代表向下取整。当卷积核尺寸大于 1 时,卷积运算操作会改变输出数据的尺 寸,这种模式被称为 "Valid"。而另一种模式被称为 "Same",即卷积核的输入和 输出数据尺寸不变,通过改变填充*p*即可实现。

在实际应用中,输入数据和卷积核的维度是相同的。在每个卷积层中会有多 个卷积核,可以提取多种形式各异的特征,每一个卷积层都会在卷积核中产生一 个对应的输出:

$$a_j^l = \sigma(\vec{z}^l) = \sigma\left(\sum_{i \in M_i} w_j^l * a_j^{l-1} + b_j^l\right)$$
(3-3)

在上式中, *a*^{*j*}为第*l*层的第*j*个卷积层输出的特征图, *M*_{*i*}代表前一层输出特征图的 集合, *w*^{*j*}为第*l*层的第*j*个卷积核, *b*^{*j*}则表示第*l*层的第*j*个卷积核所对应的偏置。 在这里可以将卷积核的前向传播过程向量化为:

$$\vec{a}^{l} = \sigma(\vec{z}^{l}) = \sigma\left(\vec{W} * \vec{a}^{l-1} + \vec{b}^{l}\right)$$
(3-4)

 $\sigma(\cdot)$ 代表激活函数,研究者们常使用线性整流函数(Recitified Linear Unit, ReLU) 作为在卷积神经网络中的激活函数:

$$\sigma(\mathbf{x}) = \max(0, 1) \tag{3-5}$$

由于自身的稀疏性, ReLU 相比其他神经网络激活函数可以更有效的避免梯度消失及梯度爆炸的问题。同时,从(3-5)可见 ReLU 函数的形式简单,计算相对容易,这使得神经网络的训练时间得以有效减少。

卷积层后方通常会跟着一个池化层,池化层也可被称为下采样层,池化层可 以在对输入尺寸进行压缩的同时,尽最大可能保留有效信息。除此以外,池化层 还可以增加整个模型的鲁棒性,为特征映射加上输出缩放不变性、位移不变性和 旋转不变性。

最大池化和平均池化作为两种最常见的池化方式被研究者们广泛应用。在最 大池化使用时,池化层会取出池化区域的最大值作为池化层的输出值;而在平均 池化使用时,池化层会取出池化区域的平均值值作为池化层的输出值。两种池化 方式如图 3.3 所示。

对于输入数据而言,平均池化涉及池化区域的所有数值,可以更好的保留背景信息。最大池化则可以保留更多的纹理信息和边缘信息,与最大池化的计算相



图 3.3 最大池化与平均池化

比,平均池化更加简单,可以有效节约计算资源,池化层的前向传播表示如下: $\vec{a}^{l} = downsample(\vec{a}^{l-1})$ (3-6)

全连接层和输出层在卷积层和池化层后,更高维度的信息整合工作会由全连 接层完成,全连接层提取局部特征,换一种方式说,由卷积层得到的局部特征转 化为全局特征。在分类任务中,整合的信息会被送入输出层 Softmax。为方便全 连接层计算,经过卷积层和池化层的特征图都需要通过尺寸变化,转为特征向量。

3.1.2 卷积神经网络的训练

卷积层、池化层、全连接层和输出层共同组成卷积神经网络,重点在于解析 卷积层和池化层里误差的传播方向与方式,进而可以得到权值更新的方式。卷积 层中的反向传播中计损失函数为 $J(\theta)$, $J(\theta)$ 关于第l层的第j个未激活输出 z_j^l 的偏 导数计为 $\delta_l^l \equiv \partial J/\partial z_l^l$,则输出层L:

$$\delta_j^L = \frac{\partial J}{\partial z_j^L} = \frac{\partial J}{\partial a_j^L} \cdot \sigma'(z_j^L) \tag{3-7}$$

 $\sigma'(\cdot)$ 为激活函数的导数,将上式向量化并对L-1层未激活输出 z_j^{L-1} 求偏导数可得到:

$$\delta_{k}^{L-1} = \left(w_{1k}^{L} \cdots w_{jk}^{L} \cdots w_{nk}^{L} \right) \begin{pmatrix} \delta_{1}^{L} \\ \vdots \\ \delta_{j}^{L} \\ \vdots \\ \delta_{n}^{L} \end{pmatrix} \cdot \sigma' \left(\mathbf{z}_{j}^{L-1} \right)$$
(3-8)

继续向量化可得:

$$\vec{\delta}^{L-1} = \left[(\vec{w}^L)^T \vec{\delta}^L \cdot \sigma' (\mathbf{z}_j^{L-1}) \right] \tag{3-9}$$

其中的·代表两个向量间的元素相乘,通过 $L \boxtimes \delta_j^L$ 可计算 $L \boxtimes \overline{\delta}^{L-1}$,由L推理至l可得 到如下推论:

$$\vec{\delta}^{l-1} = \left[(\vec{w}^l)^T \vec{\delta}^l \cdot \sigma'(\vec{z}^{l-1}) \right] \tag{3-10}$$

计算损失函数关于每一层*l*的第*j*个未激活输出z^{*j*}的偏导数δ^{*j*},利用迭代优化,也就是梯度下降^[43]来更新权重参数。由于梯度的形式为多元函数对参数求偏导数的向量,表示的是函数变化程度最高的地方,所以沿着梯度向量的反方向,可以相对容易的找到局部甚至全局最小值。通过梯度下降法可以找到损失函数的最小值,可以使神经网络模型获得最优拟合。

小批量梯度下降法、批量梯度下降法和随机梯度下降法为三种研究者们最常用的梯度下降算法。在批量梯度下降法中,会使用全部输入的训练样本来进行权重的更新,这种方法最有可能获得全局最优解。但体量大的训练集进行训练必然

会导致过长的训练时间。在随机梯度下降法中,每次迭代训练仅仅更新一个输入 训练样本的权重,训练的速度因此大为加快,损失函数可以快速达到最小值,随 之带来的缺点是训练的稳定性只能说是差强人意,可能无法达到全局最优解。而 小批量梯度下降法是一种较为折衷的方法,这种方法会随即划分训练集为多个批 次,每个训练周期仅使用一个批次数据来更新权重,是一种满足寻求全局最优解 和加快训练速度的全面选择,本文遂使用此法。

本文为分类任务,输出层使用的激活函数为 Softmax,可使用交叉熵损失函数,形式为:

$$J(\vec{\theta}) = -\frac{1}{m} \left\{ \sum_{i=1}^{m} \sum_{j=1}^{k} \chi_A[y_i = j] \log \frac{e_t^T x_i}{\sum_{t=1}^{k} e_t^T x_i} \right\} + \frac{\lambda}{2} \sum_{i=1}^{k} \sum_{j=0}^{n} \theta_{ij}^2 \qquad (3-11)$$

逻辑函数 χ_A 中,真值时为1,假值时则为0。($\lambda/2$) $\sum_{i=1}^{n} \sum_{j=0}^{n} \theta_{ij}^2$ 是二次正则化项, λ 是正则化系数,加入正则化可以有效避免过拟合。m是小批量数据集中的样本 数量,在一个迭代周期的小批量数据,计算单个数据的交叉熵损失值并求均值作 为损失函数值,对该函数进行求导:

$$\nabla_{\theta_j}(\theta) = -\frac{1}{m} \sum_{i=1}^m \{ x_i [\chi_A(y_i = j) - p(y_i = j) | x_i; \theta] \} + \lambda \theta_j$$
(3-12)

$$\theta_j = \theta_j - \gamma \nabla_{\theta_j} J(\theta) \tag{3-13}$$

 γ 为学习率,通过前向传播可以得到权重 w_{jk}^l 和 b_j^l 对损失函数的偏导数,将其向量化为:

$$\frac{\partial \mathbf{J}}{\partial \vec{w}^{l}} = \begin{pmatrix} \delta_{1}^{L} \\ \vdots \\ \delta_{j}^{L} \\ \vdots \\ \delta_{n}^{L} \end{pmatrix} \cdot \left(a_{1}^{l-1} \cdots a_{k}^{l-1} \cdots a_{n}^{l-1} \right) = \vec{\delta}^{l} \cdot (\vec{a}^{l-1})^{T}$$
(3 - 14)

$$\frac{\partial J}{\partial \vec{b}^l} = \vec{\delta}^l \tag{3-15}$$

权重更新方式可表示为:

~ 1

$$\vec{w}^l = \vec{w}^l - \gamma \cdot \vec{\delta}^l \cdot (\vec{a}^{l-1})^T \tag{3-16}$$

$$\vec{b}^l = \vec{b}^l - \gamma \cdot \vec{\delta}^l \tag{3-17}$$

池化层在进行反向传播时,池化层中并未使用激活函数,对输入进行压缩。 可将池化层中的激活函数视为激活后为其输入本身,而池化层损失函数的偏导数 可假设为δ^l,此时求其前一层的δ^{l-1},需要将δ^l的所有子矩阵还原为未经池化的 尺寸,这称为反池化或者上采样。若前向传播中的池化方式为最大池化,需要还 原δ¹的所有子矩阵的值至未经池化区域最大值的位置。如使用平均池化,需要还 原δ¹的所有子矩阵的值至未经池化区域取均值之前的位置,用公式可表示为:

$$\delta_k^{l-1} = \left(\frac{\partial a_k^{l-1}}{\partial z_k^{l-1}}\right) \cdot \frac{\partial J}{\partial a_k^{l-1}} = upsample(\delta_k^l) \cdot \sigma'(z_k^{l-1}) \tag{3-18}$$

由于激活函数输出为输入本身,可视为 $\sigma(x) = x$,则 $\sigma'(x) = 1$,此时:

$$\vec{\delta}^{l-1} = upsample(\vec{\delta}^l) \tag{3-19}$$

3.1.3 应用在人体运动识别领域的卷积神经网络

如图 3.4 和图 3.5 所示,卷积神经网络的输入均为n 维数组(批数×通道 数×宽×高),所以本质上传感器的人体运动识别任务和计算机视觉的图像分类 是相同的。通常来说,图形分类输入的形状多为三通道(RGB)的正矩形,用于 处理图像数据的卷积核尺寸对应为宽高相同,并采用最大池化用于保留对图像更 为重要的边缘信息。图像数据量大,用于提取图像特征的卷积神经网络一般具有 较大的体量,模型的层数都较多,方便提取更加深层次的特征。因此,计算机视 觉模型的训练时间一般较长。

而传感器数据的宽高通常并不相同,且由于多传感器多轴的存在,通道数是 一个非固定值,在传感器多的环境下,通道的数目会远远多于3,所以在人体运 动识别中则常采用非正矩形的卷积核,卷积网络可以提取不同传感器轴上的数据。 也曾有研究者将传感器数据转化为传感器数据图像,图像形状的输入数据可将人 体运动识别任务近似的用计算机视觉任务对待。这样便可以使用正矩形卷积核来 提取特征,由于传感器数据量相比图像数据小很多,被广泛应用于完成传感器人 体运动识别任务的卷积神经网络一般体量较小,层数不要太多也可出色的完成特 征提取任务。综上所述,计算机视觉任务和人体运动识别任务在本质上都为将4 维张量送入卷积神经网络进行分类的任务,其不同主要在于输入数据的组织形式 训练成本。



图 3.4 计算机视觉的图像分类任务



图 3.5 传感器的人体运动识别任务

3.2 实验结果分析

本文利用上述的卷积神经网络,于 UCI-HAR、OPPORTUNITY、UniMiB-SHAR、PAMAP2、WISDM 和 USC-HAD 在内的六个数据集上完成训练、验证和测试,完成人体运动识别实验。

本文中的全部神经网络由 Python 编写并在 PyTorch 和 Jittor 框架下实现,完成训练的服务器搭载 6 代因特尔酷睿中央处理器(Intel 6th Gen Core I7-6850K),显卡为显存 24GB 的英伟达 RTX 3090(NVIDIA GeForce RTX 3090),内存为 64GB,实验流程图如图 3.6 所示,模型的训练都是在服务器端,模型会被保存 而部署到智能手机、英伟达 Jetson 和树莓派等平台。

在实验中,本文采用简单的3层卷积神经网络,Adam 优化器用于进行权重的更新,所有数据传入模型进行训练称为一个训练迭代周期(Epoch),训练迭代周期数为 200,根据不同的数据集,训练批量大小与每一层通道数量设置不同,详见下表,C(L_s)代表卷积核中有L_s个通道并输出L_s个特征图。

实验结果如下表所示,作为基准线(Baseline)卷积神经网络,分类精度已 稳定可靠。为全面展示实验结果,使用精度(Precision)、召回率(Recall)、F值 (F-measure)和准确率(Accuracy)多个指标作为度量标准。



图 3.6 实验流程图

数据集	第1层	第2层	此3层	批量
UCI-HAR	C (64)	C (128)	C (256)	128
OPPORTUNITY	C (64)	C (256)	C (384)	64
UniMiB-SHAR	C (64)	C (256)	C (384)	64
WISDM	C (128)	C (256)	C (384)	128
PAMAP2	C (128)	C (256)	C (384)	128
USC-HAD	C (64)	C (128)	C (256)	64

表 3.1 网络参数设置

$$Precision = \frac{TP}{TP + FP}$$
(3 - 20)

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3-21}$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Reall}$$
(3 - 22)

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$
(3 - 23)

其中TP、FN、FP和TN分别代表真正值、假负值、假正值和真负值。参数量与 浮点率用于度量网络体量大小,值得一提的是,基准线网络并未对通道进行任何 处理,下表统计实验结果,所有结果均为5次实验取均值。

数据集	UCI	OPPO	UniMiB	WISDM	PAMAP	USC
准确率	96.23	90.01	76.48	97.21	90.23	97.21
精度	95.96	89.93	76.57	97.33	90.14	96.98
召回率	96.16	90.25	76.66	97.22	90.36	96.97
F ₁ 得分	96.05	90.1	76.61	97.27	90.24	96.97
参数量	0.4	3.37	0.44	0.4	1.55	0.52
浮点率	43.8	520.67	25.84	34.84	488.62	116.87

表 3.2 作为基准线的实验结果(%),结果为 5 次实验取均值实验



图 3.7 Micro-average ROC 曲线(基准线)

本文进一步提供微平均受试者特征曲线(Micro-average Receiver Operating Characteristic Curve, ROC)供进一步展示算法性能。如图 3.7 所示,曲线下面积(Area Under Curve, AUC)与 1.0 的差距越小代表算法的性能越高。可见在基准线方法下,各数据集的 AUC 均未超过 0.97,有提升空间。

3.3 本章小结

在这一章中深度学习的基本概念得到详细的介绍,深度学习中的经典模型— 一卷积神经网络被进一步引出。在介绍卷积神经网络结构原理和训练过程的同时, 前向传播和反向传播的过程得到详细的推导。利用卷积神经网络,传感器数据上 的人体运动识别任务得以完成,在UCI-HAR、OPPORTUNITY、UniMiB-SHAR、 WISDM、PAMAP2 和 USC-HAD 数据集上得以验证,证明卷积神经网络非常适 合用于特征提取,为下文的选择提供依据。

22

第4章 内部通道激活卷积神经网络的人体运动识别

4.1 内部通道激活的卷积神经网络

实际上,传统的卷积神经网络中,通道对输出的贡献^[41]并不是均衡的,有的 通道对输出起决定性作用,而有的通道对输出贡献很小甚至没有。体现在人体运 动识别中则为针对行走动作,加速度计的通道起决定性作用,而陀螺仪的通道作 用很小。这些不含有决定性作用的通道被称为冗余通道,常规的卷积神经网络中 含有相当多的冗余通道,*l*₁范数可用于量化通道的贡献程度,通道的*l*₁范数计为 |γ_c|设定阈值, |γ_c|大于阈值时,该通道为决定性通道,反之则被称为冗余通道。

在用L₁范数量化通道的不同贡献程度后,直接将贡献度低的通道修剪掉的方 法被称为通道剪枝,在这些方法的帮助下,整个模型的参数得到减少,通道的利 用率和识别精度也得以提升。但是,剪枝与瘦身的方法在模型泛化能力方面来说 是非常激进的。参数得以减少的同时,泛化能力的损伤也是极大的,具体表现为 针对前进动作识别效果良好,而针对跳跃动作的识别效果是很差的,这在实际运 用中会导致不理想的识别效果。

由于高斯噪声^[45]已经广泛用于在卷积神经网络的权重初始化阶段,可提升卷 积神经网络的性能,本小节选择高斯噪声N(0,σ_t)作为信息源来激活冗余通道。 但是过多的噪声会导致模型难以收敛,所以本文的高斯噪声会随时间衰减。

$$\sigma_t = \alpha^t (0 < \alpha < 1) \tag{4-1}$$

图 4.1 为选择高斯噪声作为信息源的示意图。在六个公开数据集上测试 10 次后,记录分类准确度,从表 4.1 中可知高斯噪声激活后,冗余的通道可以获得 更大的|γ_c|,输出产生更多的贡献,小幅度提升分类准确度,这很好的证明冗余通



图 4.1 高斯噪声激活

米女	数据集	UCI	OPPO	UniMiB	WISDM	PAMAP	USC
高	准确率	96.31	90.33	76.67	97.35	90.43	97.51
斯	精度	95.99	90.4	76.69	97.27	90.75	97.77
噪	召回率	96.32	90.81	76.54	97.42	90.81	97.65
声	F ₁ 得分	96.15	90.6	76.61	97.34	90.78	97.21

表 4.1 加入高斯噪声后的分类表现(%),结果为 5 次实验取均值实验

道中包含潜在贡献。但这种激活方法不确定性较大且对性能提升十分有限,本文 采用小提琴图对加入高斯噪声后的分类准确率进行可视化,小提琴图可详细展示 分布的趋势以及上下限。如图 4.2 所示,加入高斯噪声后的分类准确率分布呈 四角星型,这说明实验中分类准确率的波动较大,加入高斯噪声对于准确度的提 升并不稳定。

为解决这一问题,研究者提出可模仿人脑海马体^[46]进行工作,即强壮的神经 元逐渐取代衰弱的神经元。以此思想为基础,在训练神经网络时使用模型内部的 信息使冗余通道重新对输出起作用的方法被称为内部通道激活,下面将介绍内部 通道激活的方法。

内部通道激活是一种可以从根本上解决冗余通道的做法。这种方法在单一模型中进行通道的激活, |γ_c|大于阈值的决定性通道这这里被作为信息源。核心思想 是将后 n 个冗余通道的权重全部换为前 n 个决定性通道的权重, 图 4.3 为内部



图 4.2 分类准确度分布



图 4.3 内部通道激活

激活示意图。虽然这些无效的过滤器权值被激活,输出效果更好,但仅仅替换用 决定性通道权值替换冗余通道的权值的做法是不够的,这种做法会导致相似通道 过多,模型的泛化能力不可避免的会受到伤害,究其根本是激活后通道多样性欠 佳,所以提高通道多样性是保护泛化能力的重要途径,下面的小结将介绍三种不 同的改进内部通道激活方法。

4.1.1 空间位移

在冗余通道被替换为决定性通道后,所有通道的位置是固定的。倘若对这里加上通道的空间位移,使相似的通道错位开,这样一来在卷积计算中则会得到不同的结果。在这里定义*X*的空间位移操作为Shift(*X*,*b*),则 x 轴和 y 轴上的空间位移操作可表示为:

Shift
$$(X, b)_x = \sum_{n=1}^{H} X_n \times \max(0, 1 - |x - n + b^h|)$$
 (4 - 2)

Shift(X, b)_y =
$$\sum_{m=1}^{W} X_m \times \max(0, 1 - |x - m + b^w|)$$
 (4 - 3)

这样一个即插即用的空间位移模块,可以在卷积神经网络的训练过程中大幅提升 卷积核中通道的多样性。如图 4.4 所示,从剖面图中可见绿色和黄色为位移后 的通道,卷积核的感受野被扩大,这说明空间位移还可以选择性扩大卷积核的尺 寸,可挖掘更多数据里的潜在信息,在提升信息利用效率的同时,模型的泛化能 力也可以被很好的保护起来。



空间位移前

空间位移后

图 4.4 空间位移

4.1.2 跨通道交流

这一节阐述另一种用于提高通道多样性的方法——跨通道交流激活,本文将 介绍卷积神经网络中的跨通道交流模块,并给出通道间实现通讯的相关公式。跨 通道交流模块是一个即插即用的轻量化模块,也就是说这样一个模块可以自由放 置在任何卷积神经网络内部的任何地方,且对于训练出的模型而言,参数量和浮 点率将会保持在同一水平。特征编码、权值传递和特征解码是跨通道交流模块的 三个主要部分。

特征编码部分用于提取所有通道的特征响应。具体来说,对于输出的特征图 *X_n*,首先将其扁平化为一维特征,然后传递给两个全连接层:

$$y_l^i = f_{enc}^{in} \left(x_l^i \right) \tag{4-4}$$

$$z_l^i = f_{enc}^{out} \left[\sigma(y_l^i) \right] \tag{4-5}$$

在卷积神经网络网络中,有 f_{enc}^{in} 和 f_{enc}^{out} 两种线性函数,可用 σ 表示激活函数 ReLU。

权值传递部分的作用是对不同的特征响应编码进行表示,保证通道之间的权值进行交换,同时更新各通道的特征响应。图卷积网络(GCN)^[47]就是跨通道交流的典型方法。本文提出了一种将软注意机制嵌入到网卷积神经络中的新架构,跨通道交流模块类似于图注意力神经网络。具体来说,在无向图中 $Z = \{z_l^i\}$ 为一个节点。 $s_{ij} = -f_{att}(z_l^i, z_l^j)$ 表示在两个节点之间的边缘强度。近期研究中有各种各样的方法学习 f_{att} ,而本文采用一种简单有效的计算边缘强度的方法:

$$\bar{z}_{l}^{i} = \frac{\sum_{k=1}^{h_{l}w_{l}} z_{l}^{i}[k]}{h_{l}w_{l}}$$
(4 - 6)

$$s_{ij} = -\left(\bar{z}_l^i - \bar{z}_l^j\right)^2 \tag{4-7}$$

其中 h_l 和 w_l 分别表示第l层 z_i^i 的高度和宽度。展平向量 z_i^i ,其第k个元素可以表示为 $z_i^i[k]$ 。在(4-6)中,特征编码器的平均输出可用于增加消息传递周期的鲁棒



图 4.5 跨通道交流

性。通过计算负的平方距离以保证权重在相似的通道之间有更多的交流,相似通 道的多样性和互补性得到提升。经过输出层 Softmax 后,可以得到标准化的注意 分数 a_i^i , n_i 为第l层的通道数,输出Z = { z_i^i }表示为:

$$Z_{j}^{i} = \sum_{j=1}^{n_{l}} a_{ij} z_{l}^{j}$$
 (4 - 8)

特征解码部分负责获取与校正所有已经过权重交流的通道,并将这些通道重新塑造为与原始输入相同大小。特征解码部分采用常规的卷积运算,将信息传递到下一层,特征解码器在获取通道更新方向输出Z后生效。

特征编码、权重传递和特征解码使卷积神经网络同一层上的所有相似通道的 权值进行交换与互补。来自同一卷积核的相似通道作为节点会首先被特征编码部 分计算,这些通道间的权值会通过消息传递模块进行交换,消息传递模块使用图 形神经网络相似通道的权值传递给所有通道。最后由解码器收集并还原,传送到 后续的卷积层。总之,跨通道交流模块的核心是图卷积神经网络,以每个通道为 节点进行权值的交换与互补,这种方法使大部分通道的多样性提高,加强模型的 分类性能,即插即用的模式也使模型的训练更加自由,在不增加模型体量的前提 下,模型的泛化能力得到很好的保护,图 4.5 为跨通道交流模块示意图。

4.1.3 通道平衡

多项研究表明, 冗余通道和批标准化层 (BN) 有关, 批标准化通过对输入特 征进行标准化, 使卷积神经网络的每一层都能更独立地进行学习, 这在广泛的人 体运动识别任务中起着不可或缺的作用。批标准化通常与 ReLU 共同工作, 这是 卷积神经神经网络较主流的构建模块, 批标准化后不可避免的产生许多 |γ_c| (*l*₁范 数)小至接近 0 的通道。常规应对这种现象的方法例如剪枝和注意力, 都着重于 花费更多的经历在决定性通道上, 本章则换一种思维, 将决定性通道与冗余通道 的贡献做出平衡, 进而提出通道平衡模块。 通道平衡模块通过对所有通道进行解相关(Decorrelation)或者白化(Whitening)操作,从而使所有的通道可以对特征表示做出或多或少的贡献。具体一些,对标准化后的通道进行解相关操作可表示为:

$$P_{nij} = D_n^{-\frac{1}{2}} \left[Diag(\gamma) \bar{x}_{nij} + \beta \right]$$
(4-8)

$$P_{nij}, \bar{x}_{nij}, \gamma, \beta \in \mathbb{R}^{C \times 1} \tag{4-9}$$

其中 P_{nij} 是C维向量(如C个通道),表示通道均衡模块的输出。将所有通道上的元 素 \bar{x}_{ncij} 叠加成一个列向量 \bar{x}_{nij} ,以同样的做法将 β_c 和 γ_c 叠加,分别得到 β 和 γ 。 $Diag(\gamma)$ 表示对角元素为 γ 的对角矩阵。 $D_n^{-1/2}$ 表示去相关或白化操作,可以平衡 决定性通道和冗余通道对特征表示的贡献。因为模型通常被划分为多批次训练, 而推理阶段却是针对单个样本的,所以对每个样本进行解相关操作是至关重要的。 考虑到通道平衡模块需要统计小批量和每个样本中涉及的通道依赖关系,这里将 D_n 写为:

$$D_n = \lambda \Sigma + (1 - \lambda) Diag(v_n) \tag{4-10}$$

$$v_n = f(\tilde{\sigma}_n^2) \tag{4-11}$$

$$f: \mathbb{R}^{C \times 1} \to \mathbb{R}^{C \times 1} \tag{4-12}$$

$$v_n, \tilde{\sigma}_n^2 \in \mathbb{R}^{C \times 1} \tag{4-13}$$

$$\Sigma \in \mathbb{R}^{C \times C} \tag{4-14}$$

$$\lambda \in (0,1) \tag{4-15}$$

其中,对标准化{ \bar{x}_n } $_{n=1}^N$ 后的整批样本计算协方差矩阵 Σ 。 $\tilde{\sigma}_n^2$ 是对所有通道估计的 方差向量,通过变换函数f来调整每个样本通道相关性。在经过詹森不等式 (Jensen inequality)之后, $D_n^{-1/2}$ 可以进一步被拉伸为:

$$D_n^{-\frac{1}{2}} = [\lambda \Sigma + (1 - \lambda)Diag(v_n)]^{-\frac{1}{2}} \leq \Sigma^{-\frac{1}{2}} + (1 - \lambda)[Diag(v_n)]^{-\frac{1}{2}} \qquad (4 - 16)$$

综上所述, D_n 主要由批量解相关(Batch Decorrelation, BD) $\Sigma^{-1/2}$ 和实例重加权(Instance Reweighting, IR) [$Diag(v_n)$]^{-1/2}两个分支组成,前者利用整个小批量估计的协方差矩阵进行批量解相关,后者则通过调整每个样本特征通道之间的相关性进行实例加权,本文使用一个可学习的比率 λ 来平衡批量解相关和实例重加权在整个通道平衡模块中的比例。

通道平衡模块为可穿戴设备的人体运动识别任务带来两个额外的好处: 一方面,更快的推理速度。 $D_n^{-1/2}$ 可以作为一个移动平均统计量来进行计算,这可以加快推理速度;另一方面,微不足道的计算量。只需计算一次根号倒数 $D_n^{-1/2}$, $[Diag(v_n)]^{-1/2}$ 也很容易计算。

接下来将继续介绍批量解相关 $\Sigma^{-1/2}$ 和实例重加权[$Diag(v_n)$]^{-1/2}两个分支。



图 4.6 跨通道平衡

直观地说,协方差矩阵 D_n 应该在小批量样本上计算。假设x是一个标准化的特征 张量,第i个通道与第j个通道之间的依赖关系用 Σ_{ij} 表示,使用 $\gamma_i\gamma_j$ 对 Σ_{ij} 标准化后 缩放。协方差矩阵可以表示为:

$$\Sigma = \gamma \gamma^T \odot \frac{1}{N \cdot H \cdot W} \bar{x} \bar{x}^T \qquad (4 - 17)$$

$$\bar{x} \in \mathbb{R}^{C \times M} \tag{4-18}$$

 ○意为元素间相乘,通过执行奇异值分解(Singular Value Decomposition, SVD)
 ^[48]或特征分解来计算Σ^{-1/2}。为了避免繁重的计算,本文在训练阶段采用牛顿迭 代法^[49]来计算反平方根,可以很容易地集成到卷积层中,计算量几乎可以忽略不 计。除批量解相关(Batch Decorrealtion)^[50]外,每个样本也需要去相关,实例重加 权可被定义为:

$$\tilde{\sigma}_n^2 = Diag\left[\gamma\gamma^T \odot \frac{(\sigma_{IN}^2)_n}{\sigma_{BN}^2}\right]$$
(4 - 19)

$$\tilde{\sigma}_n^2$$
, $Diag(\gamma\gamma^T)$, $(\sigma_{IN}^2)_n$, $\sigma_{BN}^2 \in \mathbb{R}^{C \times 1}$ (4 - 20)

其中 $Diag(\gamma\gamma^T)$ 用于提取给定矩阵的对角线, $\sigma_{BN}^2 和(\sigma_{IN}^2)_n$ 分别代表批标准化和实例标准化估计的方差。在(3-11)中, 变换函数f是一个用于重新参数化的 s 型激活函数 (Sigmoid), 可以表示为:

$$s = \frac{1}{NC} \sum_{n,c}^{N,C} (\tilde{\sigma}_n^2)_c \qquad (4-21)$$

$$[Diag(v_n)]^{-\frac{1}{2}} = Diag[\tilde{f}(\tilde{\sigma}_n^2;\theta)] \cdot s^{-\frac{1}{2}}$$
(4 - 22)

其中s⁻¹2表示对所有小批量和所有样本方差的平方根倒数。在原理类似的挤压与激励模块^[51]和全局上下文模块^[52]的指导下,本文利用一个带有θ参数子网络来调整信道依赖关系。S型激活函数f可以学习一组权值来控制所有通道方差的平方根,保证其输出幅值接近批量解相关分支中方差或协方差的平方根。这说明在信道均衡中,批处理去相关和实例加权都不占主导地位,整个通道平衡模块的示意图如图 4.6 所示。

4.2 实验结果及分析

本章使用内部通道激活的三种改良,在6个不同的公开人体运动数据集上进行实验,记录包括定量对比、通道冗余率等实验数据。

4.2.1 定量对比

本文将基准线和加入改良的内部通道激活的卷积神经网络进行定量对比。所 有模型均测试次,并记录平均值。在表 4.2 中,准确率、精度、召回率和 F₁ 得 分这四项指标被记录,所有实验均进行五次并取均值记录。

取准确率作为主要指标,空间位移方法分别在 UCI-HAR、OPPORTUNITY、 UniMiB-SHAR、WISDM、PAMAP2 和 USC-HAD 数据集上取得 96.77%、91.23%、 77.26%、98.52%、91.46%和 98.25%的准确率;跨通道交流方法取得 96.98%、 91.03%、77.42%、98.28%、91.93%和 98.33%的准确率;跨通道平衡方法则取得 97.35%、93.82%、78.65%、99.04%、92.14%和 99.17%的准确率。

与基准线 96.23%、90.01%、76.48%、97.21%、90.23%和 97.21%的准确率相比,均有不同程度的提升,而在精度、召回率和 F1得分这三项指标中,提升也是

	数据集	UCI	OPPO	UniMiB	WISDM	PAMAP	USC
	准确率	96.77	91.23	77.26	98.52	91.46	98.25
空 间	精度	96.79	91.35	77.62	98.67	91.88	98.57
位移	召回率	96.89	91.21	77.15	98.33	91.34	98.15
	F_1 得分	96.93	91.28	77.38	98.5	91.61	98.36
	准确率	96.98	91.03	77.42	98.28	91.93	98.33
跨通道	精度	96.77	91.55	77.66	98.46	92.07	98.45
坦交流	召回率	96.94	91.01	77.37	98.15	91.88	98.12
νu	F1得分	96.85	91.28	77.51	98.3	91.97	98.28
	准确率	97.35	93.82	78.65	99.04	92.14	99.17
通道平衡	精度	96.93	93.33	78.69	99.12	91.94	98.85
	召回率	97.33	94.04	78.88	99.26	92.43	99.02
	F1得分	97.12	93.68	78.78	99.18	92.18	98.93

表 4.2 定量对比 (%),结果为 5 次实验取均值实验



图 4.7 Micro-average ROC 曲线(内部通道激活)

类似的,这说明使用卷积神经网络内部的通道作为信息源确实可以提升模型的分 类性能,为进一步证明内部通道激活方法的有效性,如图 4.7 所示,ROC 曲线 与 AUC 可充分证明内部通道激活方法对模型性能的提升。鉴于这三种方法均为 内部通道核心思想加即插即用的模块改良而来,训练出的模型参数量与浮点率也 维持在同一水平,也就是说在在实际测试中,响应时间不会有巨大变化,实际测 试则在接下来的章节体现。

4.2.2 通道冗余率

为验证内部通道激活对降低冗余通道率的效果,本文设定|γ_c|的阈值为 1e⁻² 和 1e⁻³。由于对内部通道激活的改进方法均为加入即插即用的模块,本文分别于 UCI-HAR、OPPORTUNITY 和 USC-HAD 数据集上以同等参数进行训练,取 5 次 实验的均值,对冗余通道比率进行评估,实验结果如图 4.8 所示。

在柱状图中,两种蓝色分别表示的阈值为 1e⁻²和 1e⁻³的情况下,基准线卷积 神经网络的冗余通道率;两种橙色分别表示的阈值为 1e⁻²和 1e⁻³的情况下,基准 线卷积神经网络在空间位移激活改进下的冗余通道率;两种绿色分别表示的阈值 为 1e⁻²和 1e⁻³的情况下,基准线卷积神经网络在跨通道交流激活改进后的冗余通 道率;两种红色分别表示的阈值为 1e⁻²和 1e⁻³的情况下,基准线卷积神经网络在 通道平衡激活改进下的冗余通道率。于表 4.3 可以观察到阈值为 1e⁻²时,与基准 线卷积神经网络相比,三种内部通道激活方法都将冗余通道的比率降低近一半。 为证明内部通道激活方法的有效性,本文将|γ_c|阈值减低至 1e⁻³并继续观察,冗



图 4.8 不同内部通道激活方式下的冗余通道率

阈值	方法	UCI	OPPO	USC
	基准线	51.41	49.32	53.52
12	空间位移	26.09	24.58	25.47
1e ⁻	跨通道交流	26.17	24.38	36.02
	通道平衡	26.09	24.06	25.78
	基准线	46.95	45.42	49.53
1e ⁻³	空间位移	5.08	5.93	5.63
	跨通道交流	4.14	6.09	5.23
	通道平衡	4.45	6.09	5.87

表 4.3 不同内部通道激活方式下的冗余通道率(%),结果为 5 次实验取均值实验

余通道率已低至 5%左右,这说明大部分冗余通道的|γ_c|是介于 1e⁻² 和 1e⁻³ 之间 的,在三个都出现类似的现象也进一步证明内部通道激活方法的有效性。

4.2.3 内部通道激活效果

本节的实验是在 PAMAP2 数据集上完成的,为了研究内部通道激活如何增强冗余通道的特征响应,本文尝试建立一个可视化的类激活图(CAM)^[53],可视化



图 4.9 可视化类激活图

类激活图是一种检测通道间变化的方法,其特点是有效而清晰直观,在 PAMAP2 数据集上对带有内部通道激活的卷积神经网络进行训练。在图 4.9 中,可以看 到从模型中提取到的可视化类激活图,内部通道激活方法能够在传感器时间序列 中定位突出通道对特征响应的贡献。例如图 4.9 展示一个相对简单的活动—— 坐。读者们可以直接观察到对应轴的通道(手腕处:z轴;脚踝处:x轴;胸口处:v 轴)比其他通道做出更多的贡献。而一并展示的平躺动作比静坐更复杂一些,因 此该动作可视化类激活图中对应轴的通道(手腕处:y 轴和 z 轴;脚踝处:x 轴;胸 口处:x轴和 y轴)比静坐着有更多的带有贡献的通道。继续研究一种更复杂的活 动一一步行,与静坐着和平躺相比,步行中受试者的脚踝处有更多的运动,所以 在可视化类激活图中有更多轴对应的通道(手腕处:y轴;脚踝处:x轴、y轴和 z 轴;胸口处:x轴)成为对特征响应带有贡献的通道。在图 4.9 中,其形式最复杂 的动作为骑行,受试者的脚踝处、手腕处和胸口处有各种各样的动作,甚至比步 行还要复杂。因此从图中可以看到更多不同的的轴对应的通道(手腕处: y 轴和 z 轴;脚踝处:x轴、y轴和z轴;胸部:x轴和y轴)成为对输出特征响应有实际贡献 的通道。这些可视化类激活图直观展示内部通道激活方法在基于卷积神经网络中 的特征提取中效果,在内部通道激活方法下,大部分通道的能够以更全面的方式 进行学习,可对输出特征表示的贡献变得更多。

4.2.4 移动端部署

考虑到内部通道激活方法的实际用处,本文在嵌入式树莓派(Raspberry Pi) ^[54]平台上对内部通道激活方法的实际运行情况进行评估。主要通过三个主要步骤 来实现基于传感器数据的人体运动系统。首先,利用 USC-HAD 数据集中的训练



图 4.10 用户界面



图 4.11 树莓派 Model 3 B+上的推理时间曲线

集对模型进行训练并保存最佳模型为.pth 格式;其次,将保存的.pth 模型加载并 部署到嵌入式树莓派平台。最后,加载.pth 模型并实时运行推理,统计计算推理 时间。用于测试的平台是带有 ARM Cortex-A53 64 位 1.4GHz 和 1GB LPDDR2 SDRAM 的树莓派 3 Model 3B+。微型 SD 卡用于加载 Raspbian 操作系统和存储 数据。值得一提的是,PyTorch 库与树莓派的操作有很好的兼容性,图 4.10 是 用 Python 开发的图形用户界面。

本文进行 500 次实际, 推理时间曲线图如图 4.11 所示, 可以看出鉴于基准 线模型本身就为 3 层卷积神经网络, 推理时间已可以维持在一个较低的水平, 无

测试次数	基准线	空间位移	空间位移 跨通道交流	
1	68.73	67.89	65.77	69.81
2	61.83	65.85	68.55	64.81
3	64.92	64.22	64.38	65.83
4	64.85	65.92	64.99	65.82
5	64.76	65.71	64.25	65.83
平均	64.42	65.92	65.59	66.42

表 4.4 实际推理时间 (窗口/毫秒), 记录 500 次实验结果

论是空间位移、跨通道交流还是通道平衡都能够保持与基线几乎相同的推理时间, 这对于可穿戴设备上的实际人体运动识别应用具有重要意义。

在表 4.4 中,有关于前五次推理时间和平均推理时间的详细记录,本身基准 线便可做到平均 65 毫秒一个动作的快速推理时间,而三种改进的内部通道激活 方法均为使用即插即用模块,这使得经过内部通道激活的卷积神经网络在实际测 试中,推理单个窗口的时间可以和基准线保持在同一较低水平上。

4.3 本章小结

本章提出信息源为内部通道的激活方法,核心思想是用贡献度大的决定性通 道替换冗余通道,在此基础上对内部通道激活方法进行改进,使用即插即用的空 间位移、跨通道交流和通道平衡模块使通道的多样性得以加强,保护模型的泛化 能力。实验结果可证实这些方法的有效性,与基准线模型相比在分类性能上有较 大提升,且实际测试中的推理时间也同样维持在较低水平。而信息源终归为内部 信息,下一章中本文将考虑把信息源扩展到外部网络,再进一步提升通道的多样 性。实现外部通道激活卷积神经网络的人体运动识别。

35

第5章 外部通道激活卷积神经网络的人体运动识别

上一章中,主要介绍的是通道激活信息源为统计卷积神经网络内部通道的内 部通道激活方法,并针对相似通道过多导致泛化能力下降的问题。提出包括空间 位移、跨通道交流和通道平衡三种即插即用的模块来提高通道的多样性。实验表 明,内部通道激活方法确实可以提高模型的分类性能,在实际应用中也是可以保 证较低水平的推理时间。然而,此时出现一个新的挑战:无论用什么样的模块以 提高通道的多样性,信息源却都是卷积神经网络内部的通道,它们在一开始便很 相似,采用即插即用模块提升多样性的做法终究对通道多样性的提升是有限的。

近期,集成学习^[55]在深度学习中获得广泛的关注。集成学习通过对多个个体 学习器进行整合,可得到一个性能良好且多样性丰富的模型。在这一思想的指导 下,本文跳脱出单一网络内部,将信息源扩展至网络外部,意在使用外部网络的 通道对目标网络进行激活。

5.1 外部通道激活的卷积神经网络

外部通道激活需要同时训练多个卷积神经网络,它与集成学习和知识蒸馏^[68] 有许多相似之处。与知识蒸馏不同,外部通道激活只需要用一个阶段就可以训练 多个模型,而知识蒸馏需要用两个阶段来训练教师模型和学生模型,第一阶段需 要优先训练出教师模型,第二阶段教师模型会指导学生模型的训练过程。训练所 需阶段数的不同可充分体现外部通道激活和知识蒸馏的明显区别。外部通道激活 方法会在同一个阶段内要求多个卷积神经网络模型相互学习。外部通道激活通常 比知识蒸馏训练出的模型表现得更好,前者训练出的模型在相互学习的过程中可 以获得更多的差异,多样性可以得到更好的保证。

值得一提的是,外部通道激活并不会增加通道提升多样性的交流成本,因为 所有的通道在每个训练周期被激活,而不需要额外增加训练周期,冗余通道的激 活是与训练共同完成的。陈^[57]等人提出了一种新的基于智能手机的人体运动识别 集成学习技术。与他们的工作相比,外部通道激活具有明显的优势,在外部通道 激活方法中,所有参预训练网络内的决定性通道会相互学习,最终来自不同卷积 神经网络的决定性会被集中到同一个卷积神经网络模型中,只需要在嵌入式系统 上部署单个网络,所以外部通道激活的卷积神经网络可以做到更加轻量级。而外 部通道激活方法与剪枝方法相比,不需要改变网络结构,保护模型的泛化能力。 与内部通道激活方法相比,来自外部的通道信息从一开始便保证有足够大的差异, 这使得通道的多样性也得到很好的保证。

5.1.1 双网络的外部通道激活

简单起见,本文首先讨论双网络的外部通道激活,即提供外部信息的网络仅 有一个,对网络 M₁和 M₂进行并行训练。在训练阶段,外部通道激活方法利用 M₂的有效过滤器作为信息源替代 M₁的无效过滤器。与内部激活相比,外部激活 有两个调整,一方面,由于初始权重的值和冗余通道的位置在两个卷积神经网络 中是不同的,如果简单的进行任意从 M₂迁移通道至 M₁的操作很有可能会损害 层的一致性,因此,本文选择卷积神经网络内部的每一个层进行外部通道激活。 卷积神经网络 M₂某一层中对输出具有决定性作用的通道被用来激活卷积神经网 络 M₁中同一层的冗余通道,在这一个学习过程中,M₁和 M₂可以相互学习信息, 实现通道间的互补与共享;另一方面,本文通过以下公式表示外部通道激活:

$$W_i^{M_2} = \alpha W_i^{M_2} + (1 - \alpha) W_i^{M_1} \quad (0 < \alpha < 1) \tag{5-1}$$

其中 $W_i^{M_2}$ 是卷积神经网络 M_2 在第i层的权重, $W_i^{M_1}$ 则是卷积神经网络 M_1 在第i层的权重。假设 $W_i^{M_1}$ 的信息量是小于 $W_i^{M_2}$ 的, α 应该大于 0.5,反之亦然。图 5.1 用于形象的解释信息源为双网络情况下的外部通道激活方法。

为提高两个网络的多样性,本文从一开始便设定两个卷积神经网络网络的初 始化方式为不同,并将包括优化器与学习率等超参数设置为不同,这样的设计方 法可以加大两个网络之间的差异,可以使两个卷积神经网络具有不同的权值,又 不会破坏每个层之间的一致性。在双网络的情况下值得注意的是,在外部通道激 活操作之后,如果只在每个训练周期实现激活,会有可能出现两个网络权值相同 的情况,但由于整个训练过程的多样性,训练出的权值会较大的差异性。当进一 步提升多样性并将外部通道信息源扩展到多个卷积神经网络的情况下,这种出现 同一训练周期中权值相同的现象就会大大减小甚至完全消失。



图 5.1 双网络的外部通道激活

5.1.2 多网络的外部通道激活

在多网络外部激活方法中,权值的变化需要被额外关注。假设对于所有卷积 神经网络,通道中的所有权值都是相同的,那么无论上文中α的绝对值是多少, 在常规的卷积运算下,卷积核中的每个通道对输出特征表示的贡献是相等的。而 使用多于两个网络作为信息源的目的为提高通道间的多样性,要在不破坏对应每 个卷积层一致性的前提下,这是一个需要被解决的问题,因此在多网络的外部通 道激活方法中会把每个层的权重*W_i*作为一个整体来计算,以下是权重信息值的 计算公式:

$$H(W_{i,j}) = -\sum_{k=1}^{B} p_k log p_k$$
(5-2)

在这一小节中,对两个模型的权重采用自适应加权策略进行加权。 $W_i^{M_1}$ 是网络 M_1 中第i层的权值,而 $H(W_i^{M_1})$ 表示网络 M_1 中该层的信息,根据 (5-2)可以计算出 $H(W_i^{M_1})$,而计算系数 α 需要满足两个条件。一方面,如果 $H(W_i^{M_1}) = H(W_i^{M_2})$,则 (5-2)中 α 的值为 0.5;如果 $H(W_i^{M_1}) < H(W_i^{M_2})$, α 的值则需要大于 0.5;另一方面,无论 $(W_i^{M_1}) < H(W_i^{M_2})$ 或者 $(W_i^{M_1}) > H(W_i^{M_2})$,包括卷积神经网络 M_1 在内所有作为外部激活信息源的卷积神经网络 M_2 到 M_k 都被要求贡献出内在的通道,并于其他网络共享。

自适应加权系数可以被用来满足上述两个要求:

$$\alpha = A \times \left(\arctan\left(c \times \left(H\left(W_i^{M_2}\right) - H\left(W_i^{M_1}\right) \right) \right) \right) + 0.5$$
 (5-3)

(5-3)中的A和c是一对固定的超参数,用于计算加权系数。在图 5.2 中,我们将满足上述两个需求的功能可视化。



图 5.2 功能可视化



图 5.3 多网络的外部通道激活

进一步将外部通道激活扩展到多网络的情况,如图 5.3 所示,并行训练多 个网络 M₁ 到 M_k,在训练过程中 M_{k-1}在每个训练周期中训练的同时将其自身有 决定性作用的通道传递给 M_k,在外部通道激活方法的加持下所有的网络可以在 特定的训练时间中交换信息,计算加权系数是根据自适应加权策略得出的。

5.2 实验结果及分析

本章作为第四章内部通道激活方法的扩展,主要目的是提高通道的多样性以 进一步提升分类性能,以相同的实验硬件完成基于传感器数据的人体运动识别任 务并详细记录测试数据。

在记录多项测试数据后,本文按照上一章对冗余通道进行量化,取多个阈值 冗余通道率进行观测。在改变参与训练的网络数量和训练多样性后进行实验,通 道的多样性对实验结果的影响也在本节中进行探究。

5.2.1 定量对比

本文将基准线和加入外部通道激活的卷积神经网络进行定量对比。所有模型 均训练并测试5次,记录平均值。

在表 5.1 中,准确率、精度、召回率和 F₁得分 4 项指标被记录。取准确率 作为主要指标,外部通道激活方法分别在 UCI-HAR、OPPORTUNITY、UniMiB-SHAR、WISDM、PAMAP2 和 USC-HAD 数据集上取得 99.65%、94.31%、79.56%、 99.46%、92.23%和 99.47%的准确率。与基准线 96.23%、90.01%、76.48%、97.21%、 90.23%和 97.21%的准确率相比,均有不同程度的提升,在 UCI-HAR、WISDM 和 USC-HAD 上更是取得非常接近 100%的准确率,这些提升相比于上一章的提

	数据集	UCI	OPPO	UniMiB	WISDM	PAMAP	USC
外	准确率	99.65	94.31	79.56	99.46	93.23	99.47
部通	精度	98.93	94.33	78.96	98.96	92.95	99.51
道激	召回率	99.33	94.08	79.24	99.38	93.18	99.21
活	F ₁ 得分	99.13	94.2	79.1	99.17	93.06	99.36

表 5.1 定量对比(%),结果为 5 次实验取均值实验

升而言更加明显,而在精度、召回率和 F₁得分这三项指标中,提升也是类似的, 这说明使用卷积神经网络外部的通道作为信息源确实可以提升模型的分类性能, 且在更强的通道多样性加持下,提升对比使用卷积神经网络外部的通道作为信息 源是更加可观的。由于训练出的模型只有一个会保存下来,每个层的一致性也保 持不变,所以训练出的模型参数量与浮点率也维持在同一水平,也就是说在在实 际测试中,响应时间不会有巨大变化,实际测试在下节用相同的方式体现。

5.2.2 通道冗余率

为验证内部通道激活对降低冗余通道率的效果, |γ_c|的阈值范围设定为 1e⁻⁴ 到 1, 本文分别于 UCI-HAR 和 OPPORTUNITY 数据集上进行训练, 取 5 次实验 的均值, 对冗余通道比率进行评估, 实验结果如图 5.4 所示。



图 5.4 通道冗余

当设定|γ_c|的阈值为1时,冗余通道率在UCI-HAR和OPPORTUNITY数据 集上分别为51.6%和61.1%,而通过外通道激活可以将冗余通道率降低至38.8% 和50%;降低阈值至1e⁻¹时,UCI-HAR和OPPORTUNITY上的冗余通道率为 44.3%和50.3%,有外部通道激活则将冗余通道率都降低到27.1%;继续降低阈 值到1e⁻²,此时的冗余通道率为35.2%和42.5%,而外部通道激活将冗余通道率 降低至更低的19.3%和13.4%;进一步降低阈值为1e⁻³,可以观测到冗余通道率 在两个数据集上分别为16.9%和37.2%,外部激活方法则可以进一步减低冗余通 道率,冗余通道率表现为3.5%和5.4%,降低比率明显;最终阈值低至1e⁻⁴时, 可以看到冗余通道率分别为15.6%和36.6%,而外部通道激活方法一样可以减低 冗余通道率,只是相比阈值为1e⁻³时,效果并不明显。这样的实验结果和内部通 道激活方法相比,降低的比率更大,这也一样说明冗余通道的|γ_c|大部分集中在 1e⁻²和1e⁻³之间。

5.2.3 训练多样性的影响

集成学习的多样性即参与模型之间的差异,在外部通道激活中起着重要的作用,对最终的性能至关重要。外部通道激活的变化越多,学习效果就越好,但硬件条件有限,无法设置无限大数量的网络参与训练。本文通过设置不同的学习率和训练样本顺序提高作为信息源的多个网络的多样性。为简便起见,6个不同的网络 M1、M2、M3、M4、M5 和 M6 的初始学习率分别为 5e⁻³、5e⁻⁴、2e⁻³、2e⁻⁴、1e⁻³、1e⁻⁴。他们的学习率分别每 10、20、25、30、40 和 50 个训练周期衰减到 95%。如表 5.2 所示,不同的学习策略和不同的样本顺序可以显著增强训练的多样性,从而提高分类性能。本文也鼓励进一步研究如何通过提高集合多样性来提高滤波器的激活性能,而其中一种做法则是增加作为信息源的网络数量。

不同的学习率	不同的样本顺序	UCI-HAR	OPPORTUNITY
X	Х	96.88	91.32
X	\checkmark	97.35	91.98
\checkmark	Х	98.53	93.38
\checkmark	\checkmark	99.65	94.31

表 5.2 训练多样性的影响,结果为 5 次实验取均值实验

5.2.4 网络数量的影响

对于外部通道激活,作为信息源的卷积神经网的数量是一个重要的超参数,本节分析该参数对 UCI-HAR 和 OPPORTUNITY 识别精度的影响。于中可知,作为信息源的网络数量为1到8,以同等条件依次实验5次并取均值,可以清楚地观察到对于人体运动识别任务而言,分类准确率随着信息源的复杂度(网络数目)的增加而提高,而当信息源的复杂度增加到一定量时,分类准确率趋于饱和,在此基础上继续增加则会导致分类准确率的下降,其原因在于基于传感器数据的人体运动识别任务,并不需要非常复杂的模型,复杂度达到饱和时继续增加复杂度则会导致信息的饱和,对分类准确率有负面影响。

信息源复杂度	UCI	OPPO
1(基准线)	96.77	91.23
2	96.87	91.4
3	96.93	92.53
4	97.45	93.6
6	99.13	94.2
8	98.98	94.94

表 5.3	网络数量的影响
-------	---------

在所有模型中,有6种网络参与的激活效果最好。随着我们不断增加网络的 数量,识别的准确性略有下降。外部通道激活方法可以帮助整个卷积神经网络从 外部信息源中学习到更多有用的信息,从而大大提高模型的分类能力。与传统的 集成学习不同,外部激活集成更有用的过滤器只进入一个单一的网络进行最终测 试,这不会增加任何内存和计算开销。

5.2.5 移动端部署

与上一章类似,一样需要考虑到外部通道激活方法的实际作用,一样需要对 模型进行移动端的部署,而这次则除树莓派之外加入安卓智能手机^[58]作为平台完 成实验。本文对 WISDM 数据集进行训练。

同样的,本文选择电脑端安卓系统虚拟机对外部通道激活方法进行实际测试,为便于部署,在 WISDM 数据集上训练好的模型保存为.pb 文件,安卓系统虚拟机的型号为谷歌 Nexus 6,该机型搭载的操作系统为安卓 11.0,是一个相对稳定的、兼容性良好且方便部署的旧版本系统,交互页面如图 5.5 所示。

测试平台	方法	1	2	3	平均
树莓派	基准线	69.25	68.56	67.48	68.42
	外部通道激活	69.65	69.12	67.51	68.75
安卓机	基准线	34.11	32.12	31.06	32.43
	外部通道激活	34.25	33.01	31.22	32.83

表 5.4 实际推理时间

本文进行 500 次实际,实际推理时间由表 5.4 记录,由于基准线模型本身就 为 3 层卷积神经网络,可以保证一个较低的水平的推理时间。推理时间基本没有 受到外部通道激活方法影响,而安卓智能手机性能强于树莓派,所以推理时间在 智能手机上缩短近一半,处理器过热问题也有较大的改善。



图 5.5 安卓智能机上的交互界面

5.3 本章小结

本章进一步增加信息源中通道信息的多样性并研究多样性对识别结果的决 定性作用,是研究结果表明,外部通道激活方法可以切实提高模型分类性能,通 道多样性对外部通道激活方法起决定性作用。

第6章 总结与展望

6.1 论文主要研究工作总结

传感器数据内丰富的信息在改变人们生活方式的同时,也带来巨大的挑战。 准确又节能的人体运动识别方法是研究者们的不断追求,然而卷积核中的冗余通 道占用大量参数又对输出特征表示无太大作用,传统的剪枝方法会直接将这些通 道修剪掉,这样可以大量节约参数。这些方法也有一个缺点——无法保证模型的 泛化能力,即模型在测试中表现不佳。本文则认定直接修剪模型中通道的方法是 激进的,冗余通道需要被重新考虑是由具有作用。因此,重新激活冗余通道并使 其对输出特征表示重新具有贡献是值得做出深入研究的。

本文认为直接修剪模型中通道的方法是激进的,围绕人体运动识别的实际运用,主要研究冗余通道内蕴含的信息和冗余通道重新激活的方法等一系列问题。 本文的研究内容及创新点如下。

首先,探索冗余通道在输出特征表示上的潜在作用。训练3层卷积神经网络 模型时,通过对冗余通道加入随时间呈指数衰减的高斯噪声,在多个国际上现有 测试模型性能的公开数据集进行测试,加入高斯噪声的做法已经略优于当前的基 准线模型,准确率已有小幅提升,这可证明冗余通道中对输出特征表示具有潜在 贡献,可通过一系列激活方法进一步发掘。

其次,提出通道激活方法的核心思想——替换卷积核中的冗余通道为决定性 通道,使每个卷积核中大部分通道对输出特征表示均有贡献。将激活信息源设定 在网络内部以提高通道多样性,进一步建立包括空间位移、跨通道交流和通道平 衡三种完整的内部通道激活方法。这些方法在保持泛化能力的前提下,测试时取 得更好的测试准确度,且有效的降低冗余通道在模型中的占比。

最后,受到集成学习方法的启发,扩展通道激活的信息源为网络外部。集成 学习方法可进一步大幅提高通道间的差异,提高通道多样性。于测试时,外部通 道激活方法取得更进一步的提升,在三个数据集上的分类准确率、精度、召回率 和 F₁得分已趋于 100%。

在 UCI-HAR、OPPORTUNITY、UniMiB-SHAR、PAMAP2、WISDM 和 USC-HAD 六个数据集上,本文分别对基准线、内部通道激活方法和外部通道激活方法进行全面的评估;同时,本文将训练好的模型部署至移动端平台进行实际测试,记录实际每个窗口的推理时间,使实验结果更加贴近实际生活中的运用。本文源代码均已上传互联网并开放下载。

6.2 后续工作及展望

本文提出的方法为激活冗余通道,并针对通道多样性这一项对通道激活方法 进行包括插件和信息源的改进,但也有问题值得继续谈论与研究。

文中所完成的分类动作相对简单,并无太多复杂的复合动作,且当前分类器 有完成训练需要对大量的样本进行标注的限制,只能识别已标注的动作,无法推 理识别其他的动作,这和理想的人工智能还有一定差距。当前有小样本学习技术 (Few-shot Learning, FSL)^[59]和零样本学习技术(Zero-shot Learning, ZSL)^[60], 即为模拟人类推理辨别新事物的能力。如图 6.1 所示,小样本学习为利用含多 类别而样本量少的支持集对训练好的模型进行比对,以少量的样本对模型再次进 行训练;而零样本学习则为由训练好可识别多个特征的模型,预测目标含有复合 特征。在传感器数据的人体运动识别中,小样本学习和零样本学习体现为少量动 作样本推理新动作或者通过简单的人体动作推理出复杂的复合动作,通过更加精 密的传感器,还可以实现更多高细致度的人体运动识别。



图 6.1 小样本学习与零样本学习技术

另外,文中完成模型的部署实现的功能相对单一,但目前已找到较好的框架, 预留有很多接口,可以实现更多的检测,如当前的动作、消耗卡路里数、心率、 距离、海拔以及体力等,人们还可以带上更多的传感器,监测更多健康有关的指标,若与上述小样本学习与零样本进行结合则可以实现更多有意义的功能,这对 人们健康监测的意义是重大的,而后续许多监测健康指标的算法需要开发,应用 的兼容性也需要专人维护。如图 6.2 所示,本文也在此初步设计一个功能全面 的框架,后续期待有更多的数据接入,让应用的功能更加丰富。



图 6.2 APP 框架

参考文献

- Golestani N, Moghaddam M. Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks[J]. Nature communications, 2020, 11(1): 1-11.
- [2] Anderson C, Hübener I, Seipp A K, et al. A survey of attention management systems in ubiquitous computing environments[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(2): 1-27.
- [3] Chen Z, Cai C, Zheng T, et al. RF-based human activity recognition using signal adapted convolutional neural network[J]. IEEE Transactions on Mobile Computing, 2021.
- [4] Gupta A, Gupta H P, Biswas B, et al. A fault-tolerant early classification approach for human activities using multivariate time series[J]. IEEE Transactions on Mobile Computing, 2020, 20(5): 1747-1760.
- [5] Yang P, Xie L, Wang C, et al. IMU-Kinect: A motion sensor-based gait monitoring system for intelligent healthcare[C]. Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers. 2019: 350-353.
- [6] Bu Y, Xie L, Gong Y, et al. Rf-dial: Rigid motion tracking and touch gesture detection for interaction via rfid tags[J]. IEEE Transactions on Mobile Computing, 2020.
- [7] Eddy S R. What is a hidden Markov model?[J]. Nature biotechnology, 2004, 22(10): 1315-1316.
- [8] Pisner D A, Schnyer D M. Support vector machine[J]. Elseiver Machine learning, 2020: 101-121.
- [9] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [10]Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems (NeurIPS), 2012, 25.
- [11] Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. IEEE

Transactions on Signal Processing, 1997, 45(11): 2673-2681.

- [12]Kocabas M, Karagoz S, Akbas E. Multiposenet: Fast multi-person pose estimation using pose residual network[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 417-433.
- [13] Wu Z, Shen C, Van Den Hengel A. Wider or deeper: Revisiting the resnet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119-133.
- [14]Cao Z, Simon T, Wei S E, et al. OpenPose: Realtime multi-person 2d pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine intelligence. 2019, 43(1): 172-186.
- [15]余保玲, 虞松坤, 孙耀然, 等. 基于 DeepPose 和 Faster RCNN 的多目标人体骨骼节点检测算法[J]. 中国科学院大学学报, 2020, 37 (6): 828-834.
- [16]肖志涛, 张曌, 王雯. 基于运动学动态图的人体动作识别方法[J]. 天津 工业大学学报, 2021, 40(1): 53-59.
- [17] Straczkiewicz M, James P, Onnela J P. A systematic review of smartphonebased human activity recognition methods for health research[J]. Nature Partner Journal: Digital Medicine, 2021, 4(1): 1-15.
- [18]Bao L, Intille S S. Activity recognition from user-annotated acceleration data[C]. International conference on pervasive computing. Springer, Berlin, Heidelberg, 2004: 1-17.
- [19]Kwapisz J R, Weiss G M, Moore S A. Activity recognition using cell phone accelerometers[J]. ACM SigKDD Explorations Newsletter, 2011, 12(2): 74-82.
- [20] Ma H, Li W, Zhang X, et al. AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition[C]. International Joint Conferences on Artificial Intelligence (IJCAI). 2019: 3109-3115.
- [21] Jiang W, Yin Z. Human activity recognition using wearable sensors by deep convolutional neural networks[C]. Proceedings of the 23rd ACM international conference on Multimedia (MM). 2015: 1307-1310.
- [22]Ordóñez F J, Roggen D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition[J]. Sensors, 2016, 16(1): 115.
- [23] Teng Q, Wang K, Zhang L, et al. The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition[J].

IEEE Sensors Journal, 2020, 20(13): 7265-7274.

- [24]Kumar P, Mukherjee S, Saini R, et al. Multimodal gait recognition with inertial sensor data and video using evolutionary algorithm[J]. IEEE Transactions on Fuzzy Systems, 2018, 27(5): 956-965.
- [25]Kim J H, Hong G S, Kim B G, et al. deepGesture: Deep learning-based gesture recognition scheme using motion sensors[J]. Elseiver Displays, 2018, 55: 38-45.
- [26] 王震宇, 张雷. 基于深度卷积和门控循环神经网络的传感器运动识别[J]. 电子测量与仪器学报, 2020, 32(1): 1-9.
- [27]佟丽娜, 马航航, 彭亮. 基于惯性传感器和 LSTM 神经网络的人体动作 识别方法[J]. 传感技术学报, 2020, 33(11): 1536-1543.
- [28]Han S, Mao H, Dally W J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding[J]. arXiv preprint arXiv:1510.00149, 2015.
- [29]Liu Z, Li J, Shen Z, et al. Learning efficient convolutional networks through network slimming[C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017: 2736-2744.
- [30] 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.
- [31]Sahay A, Scobie K N, Hill A S, et al. Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation[J]. Nature, 2011, 472(7344): 466-470.
- [32] Jeong J, Shin J. Training CNNs with selective allocation of channels[C]. International Conference on Machine Learning (ICML). PMLR, 2019: 3080-3090.
- [33] Anguita D, Ghio A, Oneto L, et al. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine[C]. International workshop on ambient assisted living. Springer, Berlin, Heidelberg, 2012: 216-223.
- [34]Roggen D, Calatroni A, Rossi M, et al. Collecting complex activity datasets in highly rich networked sensor environments[C]. 2010 Seventh international conference on networked sensing systems (INSS). IEEE, 2010: 233-240.
- [35] Micucci D, Mobilio M, Napoletano P. Unimib shar: A dataset for human

activity recognition using acceleration data from smartphones[J]. Applied Sciences, 2017, 7(10): 1101.

- [36]Kwapisz J R, Weiss G M, Moore S A. Activity recognition using cell phone accelerometers[J]. ACM SigKDD Explorations Newsletter, 2011, 12(2): 74-82.
- [37]Reiss A, Stricker D. Introducing a new benchmarked dataset for activity monitoring[C]. 2012 16th International Symposium on Wearable Computers. IEEE, 2012: 108-109.
- [38]Zhang M, Sawchuk A A. USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors[C]. Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp). 2012: 1036-1043.
- [39]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 (CVPR). 2016: 770-778.
- [40]Ballester P, Araujo R M. On the performance of GoogLeNet and AlexNet applied to sketches[C]. Thirtieth AAAI Conference on Artificial Intelligence (AAAI). 2016.
- [41]Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of physiology, 1962, 160(1): 106.
- [42]LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series[J]. The handbook of brain theory and neural networks, 1995, 3361(10): 1995.
- [43]Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828.
- [44]Gao X, Zhao Y, Dudziak Ł, et al. Dynamic Channel Pruning: Feature Boosting and Suppression[C]. International Conference on Learning Representations (ICLR). 2018.
- [45] Luisier F, Blu T, Unser M. Image denoising in mixed Poisson–Gaussian noise[J]. IEEE Transactions on Image Processing, 2010, 20(3): 696-708.
- [46] Sahay A, Hen R. Adult hippocampal neurogenesis in depression[J]. Nature neuroscience, 2007, 10(9): 1110-1115.
- [47] Wu F, Souza A, Zhang T, et al. Simplifying graph convolutional networks[C].

International Conference on Machine Learning (ICML). PMLR, 2019: 6861-6871.

- [48] Herviou L, Bardarson J H, Regnault N. Defining a bulk-edge correspondence for non-Hermitian Hamiltonians via singular-value decomposition[J]. Physical Review A, 2019, 99(5): 52-118.
- [49]Garcés A. On the convergence of Newton's method in power flow studies for DC microgrids[J]. IEEE Transactions on Power Systems, 2018, 33(5): 5770-5777.
- [50]Hua T, Wang W, Xue Z, et al. On feature decorrelation in self-supervised learning[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR). 2021: 9598-9608.
- [51]Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 7132-7141.
- [52]Cao Y, Xu J, Lin S, et al. Global context networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
- [53] Vinogradova K, Dibrov A, Myers G. Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract) [C]. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). 2020, 34(10): 13943-13944.
- [54] Upton E, Halfacree G. Raspberry Pi user guide[M]. John Wiley & Sons, 2014.
- [55] Zhou Z H. Ensemble learning [M]. Machine learning. Springer, 2021: 181-210.
- [56]Cho J H, Hariharan B. On the efficacy of knowledge distillation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 4794-4802.
- [57]Chen Z, Zhu Q, Soh Y C, et al. Robust human activity recognition using smartphone sensors via CT-PCA and online SVM[J]. IEEE Transactions on Industrial Informatics, 2017, 13(6): 3070-3080.
- [58]Chin E, Felt A P, Greenwood K, et al. Analyzing inter-application communication in Android[C]. Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys). 2011: 239-252.
- [59]Sung F, Yang Y, Zhang L, et al. Learning to compare: Relation network for few-shot learning[C]. Proceedings of the IEEE Conference on Computer

Vision and Pattern Recognition (CVPR). 2018: 1199-1208.

[60]Xian Y, Lampert C H, Schiele B, et al. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(9): 2251-2265.

在读期间发表的学术论文及研究成果

学术论文:

- Wenbo.Huang, Lei.Zhang*, Hao.Wu, et al. Channel-Equalization-HAR: A Light-weight Convolutional Neural Network for Wearable Sensor Based Human Activity Recognition [J], DOI: 10.1109/TMC.2022.3174816, IEEE Transactions on Mobile Computing, 2021. (SCI, 第一作者, IF=5.577)
- Wenbo.Huang, Lei.Zhang*, Qi.Teng, et al. The Convolutional Neural Networks Training with Channel-Selectivity for Human Activity Recognition Based on Sensors [J], DOI: 10.1109/JBHI.2021.3092396, IEEE Journal of Biomedical and Health Informatics (Old Name: IEEE Transactions on Information Technology in Biomedicine), 2021. (SCI, 第一作者, IF=5.772)
- Wenbo.Huang, Lei.Zhang*, Wenbin.Gao, et al. Shallow Convolutional Neural Networks for Human Activity Recognition using Wearable Sensors [J], DOI: 10.1109/TIM.2021.3091990, IEEE Transactions on Instrumentation and Measurement, 2021. (SCI,第一作者,IF=4.016)
- Wenbo.Huang, Lei.Zhang*, Shuoyuan.Wang, et al. Deep Ensemble Learning for Human Activity Recognition Using Wearable Sensors via Filter Activation [J], Major Revision, ACM Transactions on Embedded Computing Systems, 2021. (SCI, 第一作者, IF=1.193)
- Xing.Wang, Lei.Zhang*, Wenbo.Huang, et al. Deep convolutional networks with tunable speed-accuracy trade-off for human activity recognition using wearables [J], DOI: 10.1109/TIM.2021.3132088, IEEE Transactions on Instrumentation and Measurement, 2021. (SCI, 第三作者, IF=4.016)
- Wenbin.Gao, Lei.Zhang*, Wenbo.Huang, et al. Deep Neural Networks for Sensor Based Human Activity Recognition Using Selective Kernel Convolution [J], DOI: 10.1109/TIM.2021.3102735, IEEE Transactions on Instrumentation and Measurement, 2021. (SCI, 第三作者, IF=4.016)
- Shige.Xu, Lei.Zhang*, Wenbo.Huang, et al. Deformable Convolutional Networks for Human Activity Recognition Using Wearable Sensors [J], DOI: 10.1109/TIM.2022.3158427, IEEE Transactions on Instrumentation and Measurement, 2022. (SCI, 第三作者, IF=4.016)
- 8. Chaolei.Han, Lei.Zhang*, Yin.Tang, Wenbo.Huang, et al. Human Activity

Recognition Using Wearable Sensors by Heterogeneous Convolutional Neural Networks [J], DOI: 10.1016/J.ESWA.2022.116764, Elsevier Expert Systems with Applications, 2022. (SCI, 第四作者, IF=6.954)

- Shuoyuan.Wang, Lei.Zhang*, Xing.Wang, Wenbo.Huang, et al. A Novel All-MLP Architecture for Real-time Human Activity Recognition in Wearable Devices [J], Major Revision, Elsevier Applied Soft Computing, 2021. (SCI, 第四作者, IF=6.725)
- 10. Zhi.Wang, ShangXian.Cao*, Xianyong.Peng, HuaChun.Zhou, Wenbo.Huang, et al. A Dynamic Modeling Method Using Channel-selection Convolutional Neural Network: A Case Study of NOx Emission [J], Under Review, Elsevier Energy, 2022. (SCI, 第五作者, IF=7.147)

发明专利与软件著作权:

- 1. 张雷; 黄文博, 一种基于信道选择卷积神经网络用于可穿戴设备的人体 运动识别方法[P]. 中国专利: CN111860191A
- 2. 黄文博,基于信道选择卷积神经网络的人体运动识别系统,2019-1-8.
- 3. 黄文博, 基于 CBAM 注意力信道选择卷积神经网络的人体运动识别系 统, 2020-6-1.
- 4. 黄文博,基于残差信道选择卷积神经网络的人体运动识别系统,2020-6 1.

科技竞赛:

 2021年第十七届"挑战杯"全国大学生课外学术科技作品竞赛"黑科技" 专项赛 省三等奖

荣誉奖励:

- 1. 2021年研究生国家奖学金
- 2. 2021年南京师范大学一等学业奖学金
- 3. 2021年南京师范大学优秀研究生
- 4. 2022年南京师范大学优秀毕业生

学术服务:

- 1. MDPI Sensors 审稿人
- 2. Current Computer-Aided Drug Design 审稿人
- 3. Book entitled: "Digital Innovation Adoption: Architectural Recommendations and Security Solutions." 评审人

致谢

写到这里时,我意识到该说再见了。在这不长不短的三年里,有因为失败而 灰心过,更多的却是对未来的期待,一直学着如何写好论文,却忘了学习如何表 达自己的情感,谨以此感谢这三年鼓励我帮助我支持我的每个人。

求学十多载,家人一直给予我最大的支持。是爸爸怕我骄傲自满而不愿对我 表达的认可却对外人从不吝啬的夸赞我;是妈妈与我无话不谈替我排忧解难还一 直强调吃点好的别饿着自己;是大伯和伯母教会我广交益友做一个富有人情味的 人;是爷爷和奶奶一直对我学业的关心;是从外公和外婆家出来吃到撑的肚子; 也是家里起名橙子的猫对我呼噜呼噜……这点点滴滴的关心汇成我求学路上最 大的动力。在外,大家看到我飞的多高,而在家,只有家人见到我摔的多惨。在 此,除了感恩以外,我也必将再接再厉、力争上游。

"年轻的时候,总有那么关键的几年,你需要让自己的心跳和节奏快于别人, 才能从卷在一起的人里爬出来!"张雷老师作为我的导师,他的这段忠告也已在 我心中枝繁叶茂,我告诉自己:"如果与他人保持步伐一致,那就无法成为顶尖。" 在学术中,张老师博闻强识:在相处中,张老师亦师亦友。从前,我只听说科技 改变着人们的生活,一切都是模棱两可的概念。而张老师带我体会人工智能,深 层次的理解其中的原理,在一次次实验中,我看到电子产品认识我的脸、知道我 在干什么、听懂我说什么甚至与我对话,而这一切都被一段段代码实现,被一篇 篇论文讲述,人类的智慧就这样在我眼前尽数展现。时至今日,我也能说出我自 己对科技的理解:"展望时大张旗鼓,实现时悄无声息。"这短短一句理解的背后 是张老师一直强调的自我驱动、代码能力、英文水平和理论学习。张老师曾比喻 学术高手为洪七公,叮嘱我要跟着洪七公这样的高手学习,但在我心中张老师就 是洪七公。在此,我对张雷老师献上最真挚的感谢,祝您未来生活幸福,事业顺 利。

也感谢科研路上的每一位科研工作者。是 401 实验室的钱伟行和谢非老师, 信息与控制技术研究所这样一个充满进取的环境离不开他们的努力;是与我们合 作紧密的闵富红、武浩、何军、宋爱国老师,你们严谨的治学态度也时刻影响着 我;也是我申请博士路上的谢磊、杨健、潘金山、舒祥波、韩亚洪、程明明、何震 宇、苏敬勇、廖晓峰、张文强、张海军、韩恺老师还有每次投稿论文的同行审稿 人,在与诸位高手的交流中,我学到了很多也得到了认可。

同组的滕起、王焜、王震宇、王幸、韩朝磊、徐士戈、程鑫、权威铭、刘天 一、王朔远、程东洲、卜灿、梁俊杰、湛玉峰、杨光宇、秦卢通、熊婷、王帅帅,

55

同住的唐寅、高文彬和孟波,同窗的李冠宇、袁珊、杨淑琴、张瑞、瞿安朝、王 慧敏、刘丛昊、梁秦嘉、吕昊和张卉琳等。与你们结识是我这辈子的幸运,本来 平淡的生活在你们的陪伴下变得多姿多彩了起来,借此机会,表达下我对你们的 感谢。还有很多话都放在我的主页里了,本想放个二维码,但由于论文审阅机制 就不放在这里了,希望这份情谊通过数据延续下去。陪伴我们一起做实验到深夜 的几位实验室物业阿姨,你们也辛苦了。

与此同时,谢谢各位评审专家和答辩老师在百忙中抽空审阅本文,你们的宝 贵意见也是我前进的动力,我会将校训"正德厚生,笃学敏行"牢记在心,力求 做到最好。

"流水不腐,户枢不蠹。"我将这句话赠与大家,希望大家保持思考,坚持 运动;流年笑掷,未来可期。