

文本复制检测报告单(全文标明引文)

№:ADBD2018R_2018053015312720180530154834440174128006

检测时间:2018-05-30 15:48:34

检测文献: 53140528_吴苛勃_计算机科学与技术_基于视频流的人体行为分类建模研究(2)

作者: 吴苛勃

检测范围: 中国学术期刊网络出版总库

中国博士学位论文全文数据库/中国优秀硕士学位论文全文数据库

中国重要会议论文全文数据库

中国重要报纸全文数据库

中国专利全文数据库

图书资源

优先出版文献库

大学生论文联合比对库

互联网资源(包含贴吧等论坛资源)

英文数据库(涵盖期刊、博硕、会议的英文数据以及德国Springer、英国Taylor&Francis 期刊数据库等)

港澳台学术文献库

互联网文档资源

CNKI大成编客-原创作品库

个人比对库

时间范围: 1900-01-01至2018-05-30

⚠可能已提前检测, 检测时间: 2018/5/26 12:18:31, 检测结果: 1.3%

检测结果

总文字复制比: 0.4%

跨语言检测结果: 0%

去除引用文献复制比: 0.4%

去除本人已发表文献复制比: 0.4%

单篇最大文字复制比: 0.4% (面向智能电视应用的交互情景事件感知算法研究)

重复字数: [81]

总段落数: [5]

总字数: [20913]

疑似段落数: [1]

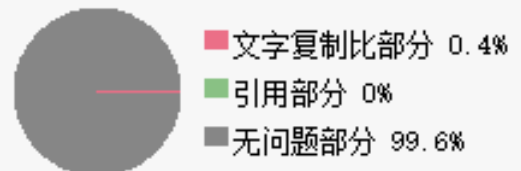
单篇最大重复字数: [81]

前部重合字数: [0]

疑似段落最大重合字数: [81]

后部重合字数: [81]

疑似段落最小重合字数: [81]



指标: ☐ 疑似剽窃观点 ☐ 疑似剽窃文字表述 ☐ 疑似自我剽窃 ☐ 疑似整体剽窃 ☐ 过度引用

表格: 0

公式: 0

疑似文字的图片: 0

脚注与尾注: 0

0% (0) 中英文摘要等 (总2693字)

2% (81) 第1章绪论 (总4045字)

0% (0) 第2章浅层学习 (总2194字)

0% (0) 第3章深度学习 (总10491字)

0% (0) 第5章总结与展望 (总1490字)

(注释: 无问题部分 文字复制比部分 引用部分)

1. 中英文摘要等

总字数: 2693

相似文献列表 文字复制比: 0%(0) 疑似剽窃观点: (0)

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

摘要

基于视频流的人体行为分类建模研究

当今正处于信息科技飞速进步的时代, 计算机、网络和数字技术正在阔然而又深切地改变着人们的平素生活和工作的习

惯。虽然比以往的生活较之方便快捷了许多，但是由于计算机还是无法辨别和理解人类的感受、想法及情绪等复杂的内心活动，从而不能主动地以类似人与人之间的方式来为人类的服务和进行与用户之间的信息交流。致使人们越来越重视和大力发展人工智能，希望通过隐式的人机交互方式，使得用户不必再拘泥于古典的人机交互模式，即必须对键盘、鼠标等输入设备进行显式而又繁琐地操作才能实现自己的目的，而是智能设备在人们所生活的三维物理空间中无需人类操作就能自主通过人们的体态语言等信息主动为人类进行服务。

故而基于视频流的人体行为分类研究是现下计算机视觉方面中的一个十分紧俏的探索钻研方向。本文主要是通过对于从包含多种人体复杂行为的视频流中进行对人类躯干的检测、人类躯体运动的追踪以及对躯体行为的辨别和理解这三个阶段进行基于视频流的人体行为分类建模研究。

本文在探讨了古典的分类模型及深度学习方法中的卷积神经网络二者之间的区别后，我们将视频流切分成帧进行检测，使用深度残差网络对视频流进行人体行为特征的提取，而后运用SVM分类器中两种降维方式对得到的图像特征值进行分类，而后对得到的结果进行评价，达到了我们预期中的效果。

关键词：行为分类深度残差网络 SVM分类器

Abstract

Research on Human Behavior Classification Modeling Based on Video Streaming

At the time of the rapid progress of information technology, computer, network and digital technology are changing the habits of people's ordinary life and work deeply and deeply. Although it is more convenient and quicker than the previous life, the computer can not distinguish and understand the complex inner activities of human feeling, idea and emotion, so it can not actively serve human beings and communicate with users in a similar way between people and people. People pay more and more attention to and vigorously develop artificial intelligence. Through implicit human-computer interaction, we hope that users will not have to adhere to the traditional human-computer interaction mode, that is, it is necessary to carry out explicit and tedious operation of keyboard, mouse and other input devices to achieve their own purpose, but the intelligent equipment is born in people. Living 3D physical space can automatically serve humans through human body language and other information without human operation.

Therefore, the research of human behavior classification based on video streaming is a very urgent research direction in computer vision. This paper focuses on the research of human behavior classification based on video flow in the three stages of human body detection, tracking of human body movement and the identification and understanding of human behavior from the video stream containing a variety of complex human behavior.

After discussing the difference between the two kinds of convolution neural networks in the traditional classification model and the deep learning method, we divide the video stream into frames, and use the depth residual network to extract the human behavior features of the video stream, and then use the two dimensionality reduction methods of the SVM classifier to obtain the image features. The value is classified, and then the results are evaluated to achieve our expected results.

Key words : human behavior classification , ResNet , SVM classifier

目录

第1章绪论	4
1.1选题背景	4
1.2国内外研究现状	5
1.3本文的研究内容和组织结构	7
1.3.1研究内容	7
1.3.2组织结构	7
第2章浅层学习	8
2.1浅层学习	8
2.1.1古典的人体检测方法	8
2.1.2传统的人体特征提取方法	9
2.2古典的HMAX模型	10
第3章深度学习	11
3.1深度学习	11
3.2人工神经网络	12
3.2.1神经元	13
3.2.2激活函数	14
3.2.3梯度下降	17
3.2.4反向传播算法	17
3.3卷积神经网络	18

3.3.1卷积层19
3.3.2池化层20
3.3.3全连接层21
3.4深度残差网络22
3.5分类器23
4.1实验需求26
4.2结构设计26
4.3系统实现26
4.4数据集预处理27
4.5训练与测试29
4.6小结34
第5章总结与展望35
参考文献36
致谢39

Table with 2 columns: Index, Content, and Statistics. Row 1: 1, 面向智能电视应用的交互情景事件感知算法研究, 2.0% (81). Row 2: 2, 指环可穿戴设备自然交互技术研究, 1.7% (68).

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

第1章绪论
1.1选题背景
近十年来, 本国六十岁及以上的人口逐年递增, 人口红利出现大幅度下降, 据统计, 老年化比率却由2007年的11.6%到2017年的17.3%。截止目前, 本国是现下世界上绝无仅有的老年人人口过两亿的国度, 并且老年化进程不断攀升, 预计2040年本国老年化进程达到顶峰, 本国年长人口数目日益增多的难题格外严峻。
老年人人口逐渐增多, 由于他们自身的身体机能衰退, 抵抗力减弱和城市环境污染的日益严重, 很容易就滋生疾病, 尤其是平日里无人照顾的独居老人, 有的时候并不在意的一些细微的变化和小疾病, 在突发状况时又因为单身一人不能及时的被救助而极有可能发生意外。对他们的健康和安全构成了严重的威胁, 因此, 这是对健康提出了新的问题和挑战。
目前, 本国已经将治理人口老年化问题作为了一项长期战略使命, 把妥善解决人口老年化带来的社会问题提升到了“事关社稷发展大局、事关黎民福利”的战略程度。由此吉林省发布了针对“独居老人安全保障与危险预警”的科技平台。专门针对老年人独居的人工录像监控步入人们视线, 越来越受到人们的青睐, 其中基于视频流的人体行为分类建模研究是人工录像监控的关键技术。
人工录像监控实际上就是在以往古典的录像监控技术根基上加入了计算机视觉技术和人机交互技术, 实现了“主动式”监控, 即就是在没有人为过问的意况下, 通过计算机视觉技术对连续的动态次序排列实行分析, 把监控场景中的人类躯体进行检查、追踪、辨别和理解。相比较于古典的录像监控技术来说, 录像监控系统的功能不再仅仅将监控器拍摄下来的视频传送到监控室中去, 然后需要由专门的人员进行不间断的审查。当突发意况发生时, 人工录像监控能够实时地进行对录像监控画面的检测、提取、分析和理解, 并且能够在很短的时间段内作出反应并发出警报, 不仅因为没有了工作人员的检测时间和反应时间而大大缩短了损耗时间, 而且还可以很好地避免因人眼过度疲劳而导致的疏忽和错误的发生而引起的不必要的不可挽回的巨大损耗, 下降了错误的发生率, 还减少了人口劳动力的使用以及物力资本的浪费。这样监控的有效性提高, 监控视频不再仅仅拘泥于作为事件发生后的有效证据, 还能在事件发生前进行目标的追踪和处理, 使得监控人员就可以对事件的发生和发展作一个预估的判断和处理, 不再被动地只能在事情已经发生后进行弥补和修护。同时工作人员也不需要再和以前一样不间断地进行对多个屏幕的监控, 只需在监控系统发出提醒时进行判断, 不仅下降了人的劳动强度同时也提升了监控系统的精确性。同时, 在周围环境变化差别灵敏度低的时候, 或者如清晨和黄昏这种光线不足会使得人脑下意识地不可避免会因疲乏而疏忽松懈麻痹大意的时候或周遭环境被进行伪装的意况下, 往往可以使用人工录像监控系统通过人工智能的技术和方法处理监控视频能够得到肉眼无法观测得到的变化意况, 提升了监控效果的精度, 减缓了事件的发生率。
而专门为老年人服务的人工录像监控系统, 为了将使得独居的老年人生活的时间尽可能更长作为目的, 除了应该满足上述条件外, 还应该能够辨别出老人正常的日常生活动作 (activity in dairy living, ADL), 如: 跑、跳、行走、弯腰等, 以作为能和异常行为和突发意况作出对比的数据集, 才能更好地为老年人服务。本文研究的基于视频流的人体行为分类建模通过对视

频流中单一人物进行简单的日常行为动作的辨别和理解，对人工录像监控技术的发扬前进有必然的参考价值。

1.2国内外研究现状

基于视频流的人体行为分类建模研究是目前当今计算机视觉研究范畴内的热门问题。但是早在二十世纪中叶七十年代初，瑞士著名心理学家Johansson【1】在对于人体动作分析的研究中首次运用了通过利用图像次序排列来对人类躯体动作行为展开分类探索。他在实验者的身体表面上的十二个活动骨节处（例如肩部骨节、腕掌骨节、膝骨节、踝骨节等等）贴上LED小灯后，并让实验者处于无光黑暗的环境下，使得此时屋内只有十二处LED小灯可被人看见，然后让实验者进行人体活动，同时让观察者记录下来LED小灯光点的运动组合来作为实验者动作行为的特征。他的实验证实了通过这些LED小灯光点的运动组合可以在黑暗中辨别人体动作行为。自此以后，各个国家的许多科研工作者都纷纷对此产生了浓厚的兴趣，开始投入到基于视频流的人体行为分类研究中来。而后经历了二十多年的研究和努力，在一九九七年，美国国防部高级研究项目署Defense Advanced Research Projects Agency联合卡内基梅隆大学Carnegie Mellon University、麻省理工学院Massachusetts Institute of Technology等高校共同研制了视觉监控项目Visual Surveillance And Monitoring (VSAM)【2】【3】。该项目的研究目的在于开发一种人工录像监控系统来全方位地对战场进行监测，从而解决在一些危险的环境下人无法实地监测的问题。而对于人口老年化非常严重的英国、法国、德国等发达国家和地区，自从上个世纪九十年代起，就已经开始尝试运用了普适计算 (ubiquitous / pervasive computing, UPC)和背景智能 (ambient intelligence, AMI)等计算机科学技术来开发一种能够帮助独居老人在家中生活而不是被迫去敬老院受人照顾看护的“智能辅助生活”(ambient assisted living, AAL)、“普适医疗保健”(pervasive computing technologies for healthcare, pervasiveHealth)或“智能家居”(smart home)等【4】【5】【6】【7】。在普适计算研究领域中，运用大量接触式感应器和非接触式感应器采集人们日常生活中的动作行为数据以期用来创建时间和运动粒度下的人类日常行为活动模型，并且运用了数据挖掘的技术方式来分析分类研究行为活动模型。因为是从大量人口中得来的数据，所以有可能建立成功真实的数据挖掘应用和发现行为模式【8】。但是现在为止，建立的较为准确的人体日常行为活动模型仅限于分辨一些简单的动作，但是对于用户的目的和意向还不能够进行正确的推理预估和判断（例如，对于像“吃饭”、“阅读”这种类似的人体行为动作尺度较小的动作在没有足够的可以使用的上下文信息时就很难进行区分和辨别）【9】。

但在计算机视觉领域内，科研工作者们着重于从动态图像次序排列中来获取人体动作行为的理论、模型和智能系统。主要目标从过去的估算判断动作行为是在什么时间、在什么地点、是什么人和做什么运动的定量参数（即W4, when, where, who and what，一般可以根据实际的需求来从四个定量参数中选择一个至多个参数来达到想要的效果和目的）【10】【11】，例如已知的对象在一段事件中的空间位置【12】【13】，到近些年来研究的热门关注点开始转变为分析动态图像次序排列中所表达出来的认识体现过程，然后从这个过程中能够理解识别出所记录下来的动作行为【14】【15】【16】。当前许多研究者们普遍承认“行为理解就是分析和辨别人体运动模式并产生关于动作与交互的高层次描述。”并且人们按照这种想法观点将人类躯体行为动作的辨别和理解分为两个阶段，即第一阶段：通过建立行为动力学模型进行人体动作行为的分类以完成对其的动作辨别，这个阶段也是最为关键的；第二阶段：就是用非专业人员就可以理解的自然语言对人体的行为动作和人机之间的交互作一个高层次的描述，即解释【17】【18】。但是在第一阶段中，在通过已有数据建立好的模型对实验数据来进行检测、追踪、辨别和理解，但是这样的辨别理解在固定的上下文信息时准确率较高，因为对于在不同的上下境中，人们的行为动作的典型特征也会随之而变化从而导致分类和辨别时的错误率上升。但是由于对上下文信息的处理过于纷繁复杂，存在了多义性的表现形式和内涵意义，无法将所有的上下文信息全部涵盖，所以至今这个问题亟待解决。

1.3本文的研究内容和组织结构

1.3.1研究内容

本文的研究方向是基于视频流的人体行为分类建模研究，本文重点在于对比分析了浅层学习方法和深度学习方法后，在利用深度学习的技术通过视频流做出来的动态图像次序排列进行了对人类躯干的检测、追踪、特征提取，而后用得出来的数据构建出人体行为分类模型，并通过对模型进行的测试，得到模型的辨别分类视频流数据集中人体行为动作的正确率和精确性的结果，并同时估计评价结论的可靠性。

1.3.2组织结构

本篇论文共分为六章，其内容组织结构如下：

第一章是绪论，其中分为三个部分，其中第一部分阐述了对当前世界人口老年化严重及本国老年人占总人口的比率逐年递增，对于独居老人的日常生活所希望得到的辅助的人工人工录像监控系统的需求日益加剧，而其关键技术即为基于视频流的人体行为动作分类建模方法至今仍不成熟的背景，充分说明了本文选题的实际意义。第二部分为人体辨别技术的起源和历史，以及目前国内外的大量研究者对于该问题的解决现状及实际困难。第三部分为本文的主要内容和组织架构。

第二章是浅层学习，主体分析了在深度学习之前的“浅时代”，紧要分析了古典的基于人工神经网络的HMAX模型的工作方式和优缺点。

第三章是深度学习，主要分析了深度学习方法的工作方式及优缺点和HMAX模型比较。并且详细阐述了人工神经网络的结构化模型。同时对于本文运用的深度残差网络ResNet进行公式化详解。

第四章是系统的实现与测试结果，通过对于有人体的挥手、化妆、太极拳、弹钢琴等共101种动作行为的视频流的测试和分析，得出系统的有效性及精确性，并得出结论。

第五章是总结与展望，对于本文所设计出来的系统进行总结，并对于相关不足和缺陷进行下一步的设想和提出问题。

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

第2章浅层学习

2.1浅层学习

基于视频流的人体行为分类建模探索总体是将包罗了人体动作行为的视频流转化为动态图像次序排列后逐帧进行人体检测、人体特征的提取以及对人体动作的辨别和理解等三个步骤进行处理的。而古典的基于视频流的人体行为分类建模研究与深度学习方法的区别主要是在人体检测和人体特征的提取这两个步骤上。古典的方法最开始是通过有限个非常简单的几何特征来进行对人体的辨别，这样的通过提取几何特征的方法的优点是特征维数少，所以不会遭受到维数灾难的问题。然而由于这种方法的稳定性差、而且区分能力很弱，并且需要人工进行，难以自动化的原因，这种方法逐渐被抛弃。取而代之的是模板匹配技术，也就是对像素区域的相似度进行直接比较。

图2-1 人体行为分类研究的流程

2.1.1古典的人体检测方法

在古典的视频流人体检测算法中广泛使用的有光流法、帧间差分法和背景减除法三种。

光流法的主要工作定理是使用动态图像次序排列中像素随着时间的变化而发生了变化的光流特性以及相邻帧与帧之间的关联联系来发现上一帧与正输入进来的帧之间所暗含的应和关系从而数算出物体的动作信息【19】。该方法主要分为（1）基于特征的匹配方法；（2）基于频域的匹配方法；（3）基于梯度的匹配方法。光流法的优点在于不需要限制摄像机位置必须固定一成不变，并且可以在相对来说比较复杂的上下境中进行对人类躯干的检测。但是光流法的计算方法相当复杂且实现起来鲁棒性比较差，除非提供大型的计算机，否则不建议采用。

帧间差分法的人体检测原理是将视频流中相邻或相隔 n 帧图像中像素值作差，而后进行运用阈值化来提取运动区域【20】【21】。帧间差分法算法简洁并且难于受到处境光线的变化作用，对动态的处境变化具有较强的鲁棒性；但同时，帧间差分法不能够用于移动的摄像机上，并且对于比较缓慢或静止的人体运动动作（例如：“阅读”“睡觉”等）和意况下几乎不能检测，效果很差强人意，并且如果在待检测的人体目标表面具有大面积的相似灰度值区域的意况下，提取出的人体特征并不完全，常常可能会显现出孔洞等意况。

背景减除法是眼前为止用处最广泛的图像检查方式，是一种对静止的场景进行人体动作行为的切分和提取方式。其原理是将不包含待检测人体的背景帧图像先提前保存好后，再将当前得到的图像帧与背景帧的图像相减作差后，得到人体动作行为的运动区域的灰度图后将结果二差化提取待检测人体特征。背景减除法的缺点很明显，一是同帧间差分法一样，不能够用于移动的摄像机上，而是对于背景帧图像的更新过于复杂，在数据集大的时候难以实现，必须要利用专门的背景建模模型（例如：平均值法等）；但是背景减除法的算法简便易于落实，对特征的提取最为完整，不易出现孔洞的意况。

2.1.2传统的人体特征提取方法

而对于人体的特征提取方面，SIFT算法逐渐出现在人们的视野范围中，并受到绝大多数的研究者们青睐，这是一种通过从图像中提取特征的方法，可以用在可移动摄像即不同视角之间的待测人体的可靠的特征提取方法【22】。SIFT方法关键点在它对于特征的提取具有尤为突出的鲜明性，是用关键点与已知的大型数据库中的关键点来进行匹配。这种鲜明性是由用代表图像梯度高维度的矢量来实现的，并将它装置配置在图像中的局部区域内部。并且关键点对图像的尺度大小及旋扭变幻具有不变性，对于跨越很大尺度的模拟、映射等变动具有非常好的适应性。因为这种方法可以从典型的图像中可以提取大量的特征点，所以能够在相对较复杂的背景下提取小目标。同时还可以将从整个尺度范围提取出来的十分小的局部特征与不仅小还高度遮挡的目标进行匹配，而提取出来的大的特征点可以在对图像噪音和图像模糊时发挥作用。SIFT方法的计算是十分高效的，在一般配置的环境下，就可以从典型图像中提取出来几千个特征点。

2.2古典的HMAX模型

HMAX模型第一次出现在人们的视线中是在2007年由Poggio以及Serre等人提出的。这是首个模仿人脑的、基于特征组的对象特征辨别提取框架【23】。该模型的理论根基是特征组合以及分层次的物体辨别的思想，模拟了生物学中人眼视皮层组织中的神经元细胞对于对象辨别处理的过程，相当于是一个四层的神经网络，实际上可以说已经初步具备了现代深度学习模型的雏形。HMAX模型运用通过事先设定好的Gabor滤波器，而后求取最大值操作和交替进行模板匹配的方法来进行对待辨别对象的特征提取，以期得到复杂性及不变性越来越好的特征表示。

HMAX模型分为S1层、C1层、S2层、C2层四层。在S1层输入的图像经过一组包含了4不同方向不同尺度的共64个的滤波器（所使用的Gabor滤波器的公式如下）处理后，得到共16组的64张图片。

到C1层以频带为单位将64张图片进行相邻两组之间的两两组合，而后在每个组之中再划分 $N \times N$ 个格子后，再分成 $N \times N$ 个更小的格子，从每个格子中取得具有最大值的子格子的值后，每张响应图像都能够得到一张最大值图，而后得到输入图像的各种特征值后进行输出。

在S2层中，将输入进来的数据根据特征块的模板 P_j 进行比较，然后通过下面的公式进行匹配运算后，重新生成一个新的

组合Y'作为S2层的输出进入C2层。

最后在C2层将输入进来的Y'进行排序并取最大值，最终得到一个特征向量作为输出，即HMAX特征，且维数为N。

第3章深度学习

3.1深度学习

在2006年的某一期《Science》期刊上，“深度学习”这个概念被Geoffrey Hinton等人提出来，这是一种拥有预先训练的可以通过多层次人工神经网络来模拟出数据之间真实的非线性关系函数的学习方法【24】【25】。深度学习的工作原理就是对人的脑视觉皮层进行效法。在二十世纪八十年代初，Hubel等人发现人脑对于视觉图像的理解过程是分层次的：第一步先是由眼睛来“摄入”图像信息（类比于摄像机的工作原理），并将其传入大脑皮层中；第二步是由一些神经元细胞进行对图像的“预处理”（例如对于图像的边缘信息的处理），而后再反馈给大脑皮层；最后由大脑皮层来对图像进行更高级的抽象化处理并做出判断【26】。坦白来说，深度学习的本质就是使用有很多个节点（也就是神经元）构成的多层感知机对数据进行非线性化的处理，得到抽象的模型。

在机器学习这个探索方向之中，深度学习被广泛地发展以及应用。而在这其中，对于图像的检测和图像的特征提取，尤其是在人体检测和人体特征的辨别方面，深度学习与之前的古典的方法最大的区别就是：深度学习能够通过利用半监督模式或非监督模式的特征获取方法和划分层次式的特征提取的算法来取代纯人工自己来设置参数从而获取特征和检测。

深度学习中的“深度”，需要被特别地指出；从一个输入处到一个输出处所经历的计算可以被称作一个流向图，而在这个流向图中的最长路径，就被称为这个流向图的“深度”。

图3-1含有多个隐藏层的深度学习模型

3.2人工神经网络

在很早的二十世纪中叶四十年代初，著名的心理学家McCulloch联合数学逻辑学家Pitts等人通过对人脑神经网络结构的分析搭建了关于神经网络的数学模型（即为被称作MP模型）。MP模型，是对于运用以神经元模式的抽象化的数学描摹和网络搭建组织手段，从而证明了单个神经元结构模型可以执行和运算线性的数学函数。由此，“人工神经网络（ANN）”【27】的概念开始逐渐被人们所熟知和理解，并有无数的研究者们对此进行了大量的研究，人工神经网络的时代正式到来。

ANN神经网络是机器学习的一个重要派系，同时也是深度学习，这个目前最热门的研究方向的根基。ANN神经网络的建构原理就是运用多量的彼此之间相互链接的神经元（neuron）而集合。如图3—2所示，这是一个经典的包含了输入层、隐藏层和输出层三个层次神经网络结构图。在这里，除了输入层以外的隐藏层和输出层均为计算层，其中隐藏层作为中间层可以不止一层。输入、输出两层系之中的neuron的数目往往是根据条件和所求已经设定好的，而隐藏层作为运算和抽象模拟的过程，其中的neuron个数可以进行自由的制定。并且ANN网络结构图之中的关键点，或者是说重中之重地方不是对于neuron数量多少的设定，而是在于神经元之间的连接线。每一条首尾两端链接两个神经元的连接线都对应着通过该连接线的连接信号的加权值。在这里我们称之为“权重”。权重，是需要对该ANN神经网络结构模型进行训练而得到的。

下面我们对人工神经网络中的每个组成部分进行逐一地描述，来介绍复杂的深度学习结构模型的构成。

图3-2 一个经典的ANN网络结构图

3.2.1神经元

ANN神经网络中的neuron模型是根据人体大脑皮层中的生物神经元结构制作而成。如下图3-3、3-4所示，神经元模型中包含了三个部分，分别是输入、运算函数以及输出；其中neuron的输入a1、a2、a3可以看作是生物神经元中的“树突”，而运算函数F1、F2则可以看作是生物神经元细胞中的“细胞核”，而neuron中的输出s则是生物神经元细胞中的“轴突”，而神经元中的w1、w2、w3则是链接前后两个神经元的连接线上的权重。我们正是通过利用输入进来的已知属性而进行的抽象化运算从而得到未知属性而进行预测的。所以，只有将权重的值不断调整到最优，该人工神经网络结构的预测才能最为精准。

图3-3生物神经元

图3-4 ANN网络结构中的神经元

3.2.2激活函数

在人工神经网络结构的研究中，由于很多的数据分析分类问题不是线性问题，所以很难仅仅只用一个单层次的人工神经网络结构（即感知器）就可以来解决我们的问题。而若要使用多个感知器来进行叠加运算以后就会发现我们所做出来分类仍然是一个线性分类（如下图3-5所示），虽然我们可以使用无限多个直线来进行对曲线的取替，但是计算量过于庞大，所以并不能实现。因而我们通常需要使用一个非线性函数，这样就可以很轻易地来帮助我们建立数据之间的非线性关系模型（如下图3-6所示），而这个函数就叫做激活函数（activation function）【28】。

图3-5使用多个感知器所获得的线性分类器

图3-6使用激活函数所获得的非线性分类器

常用的几种激活函数由sigmoid、tanh以及Relu等。sigmoid函数的公式为.其函数值分布图像如下图3-7所示，它的输出值在开区间 $(0, 1)$ 上，中间斜率大的输入区域可以看作是生物神经元细胞中的活跃区，而两边函数曲线较为平缓的输入区域则可以看成是抑制区。

图3-7 sigmoid函数

tanh函数为.其函数值分布图像如下图3-8所示，tanh函数是一个双曲线正切函数，其形状与sigmoid函数相似，但是需要指出的一点不同的是它们的输出区间不同，tanh函数的输出区间是在开区间 $(-1, 1)$ 之间，而且整个函数是以0为中心的。所以根据对反向传播算法的运用过程中的考虑，一般的意况之下，隐藏层会使用tanh函数作为激活函数，而输出层则可以用sigmoid函数来作为激活函数。

图3-8 tanh函数

而本文所用的Relu函数是目前当下运用最为广泛的函数。Relu函数的公式为.其图像如下图3-9所示，Relu函数的输入在不为负数的意况下输出本身，这样做的好处是不存在梯度饱和问题，而就sigmoid函数和tanh函数而言，当输入为比较大的正数时，函数的梯度很小。而在反向传播对运用该函数的隐藏层进行微分链式法则来计算权重的微分时，这个微分的值可能会几乎接近于零，从而并不利于对于权重的优化问题，使得梯度饱和。并且，Relu函数是线性的，计算简便，同时也大大地提高了计算速度。

图3-9 Relu函数

3.2.3 梯度下降

在运用人工神经网络结构实行判断和预算估略时，我们设真实的输出值为 y ，将估算出来的预测输出值设为 y' ，定义函数值.我们将这个值称为这个人工神经网络结构的损失，这个函数称为这个ANN神经网络结构的损失函数。而损失值愈小，说明我们的人工神经网络结构愈好，所以我们希望通过对结构的优化来减少损失，使得损失值愈来愈小。而常用的优化方法就是梯度下降法。

梯度下降法，就是对于损失函数求出它关于当前输入值的偏导数，而偏导数又是损失函数上升最快的方向，所以我们取偏导数的反方向，让参数沿着这个偏导数的相反方向进行前进一个单位，这个偏导数就是梯度，从而这样就可以将损失函数的损失值减小。而后不断地重复这个过程，直至梯度接近于零值点的时候为止，此时的输入值就达到了令损失函数为最小值时候的状态。而比较理想的梯度下降法既要满足这一条件，又要能够进行能全局收敛这一条件。这样，我们就进行了对最小优化问题的探索 and 解决。

3.2.4 反向传播算法

由于在多层次的人工神经网络结构之中，单纯地利用梯度下降法进行对损失函数的优化问题计算量过于复杂和庞大，所以，在梯度下降法的根基之上，我们运用了反向传播算法（即BP算法）。反向传播算法的本质就是输入与输出的对应相联关系。反向传播法的信息处理理论由来于对多个简易的非线性函数进行的多次重合叠加，从而拥有了高度抽象拟合复现函数的能力，然后从一个N维的集合空间（即输入）进行连续的这种高度非线性映射到一个M维的集合空间（即输出）。反向传播法大大地减少了梯度的计算，如下图3-10所示（其中E是求梯度的运算），这种方法不再简单粗暴地将所有参数地梯度统一计算，而是将顺向传播过程之中的由输入层进入途径n层隐藏层处理后传输到输出层的输入信息在没有捕捉到预期中的输出值时，进行对其的损失函数的损失值进行反向传播运算，由输出层开始，从后往前地进行逐层梯度运算，组成了权重向量的梯度借此来修改权值，然后进行循环往复地修改，直至误差减小至忽略不计时，停止迭代。

图3-10 反向传播算法示意图

3.3 卷积神经网络

我们前面说过，深度学习模型是一种多层系的ANN神经网络构造模型，由于结构过于复杂，在反向传播过程中往往由于误差变得分散而无法收敛起来所以优化效果并没有达到令人满意的效果，所以往往并不能直接使用反向传播的算法。所以我们运用了权值共享的思想，让神经元使用相同的权值进行运算，而这种结构被称为“卷积神经网络”（CNN）【28】。卷积神经网络是在二十世纪六十年代初被提出，目前已经成为了一种引起广大研究者们高度重视的高效而又便捷的辨别方法。尤其是在人体辨别分类这一课题中，由于卷积神经网络不需要对图像进行纷繁复杂地预处理工作，而是可以直接地将原始图像输入进来，故而得到了非常广泛地传播和应用。

卷积神经网络是一种向前反馈型神经网络，即各个神经元划分层系排列好，而每一层神经元只能接收前一层传递来的信息作为输入信息，再将输出信息传递给下一层神经元，而各层系之间没有反向的传递信息方式。由于卷积神经网络运用隐式地进行特征的提取训练方式来取代显式提取，并且再加上让神经元使用相同的权值进行运算，故而该神经网络可以进行并行地学习方式，从而降低了对人体特征的提取与分类过程中重新建立数据模型的复杂度。

卷积神经网络按照功能可以分为卷积层、池化层、激活层以及全连接层四个阶段，其中激活层就是使用激活函数进行运算，从而在这里我不再作赘述，下面就是从其它三个阶段对卷积神经网络加以介绍。

3.3.1 卷积层

卷积层就是进行卷积特征提取的阶段，卷积层的工作原理是由于图像具有的统计平稳性，从而在图像上的某一部分所训练学习得到的特征也可以将其用在图像上的另一部分之中，故而对于这个图像上的所有位置，我们都可以使用同样的学习特征进行提取。

卷积层的数学原理是

其中，我们把输入值（Input）称为X，把W称为卷积核。其中值得注意的一点是，W的张量同X张量大小相同。卷积核（Kernel）可以看成是一种过滤器，我们首先将卷积核作为一个小的样本进行特征的提取，而后通过这个卷积核在原图像上进行恣意位置的探测，从而得到了对整张图像进行这个特征的激活值。此外，实际的卷积核里的每一个参数都是通过多次运用反向传播算法进行推演计算而后得到的。由于一张图像会有多个特征提取的需求，所以卷积层中可以包含着多个卷积核来进行对图像不同特征提取的处理。

例如在图3-11之中，展示了一个5*5的图像，由于图像有RGB三个颜色通道，所以输入值X是三维的，从而可以得出卷积核W1、W2也是三维的，所以每个卷积核由是三个3*3的矩阵。为了使输出值（Output）得到一个我们设想好的图像大小的尺寸，模型除了会预先设置好每次卷积核移动的步长（Stride）和默认偏移值（通常为1）外，还会在输入图像的外围加上一个有n层大小边框，而这个边框被称为padding。本例中，Stride设置为2，padding设置为1，从而通过公式

我们可以得到输出的两张图像的大小为3*3。其中，图片上所示的过程为分别将 $x[:, :, 0]$ 与 $w[:, :, 0]$ 、 $x[:, :, 1]$ 与 $w[:, :, 1]$ 、 $x[:, :, 2]$ 与 $w[:, :, 2]$ 进行卷积训练：

由此可得到最终输出的结果为。

图3-11 卷积层运算示例

3.3.2池化层

当输入图像经过卷积层后输出的特征矩阵依然非常复杂，所以我们需要通过运用池化层对其输入进行再次的降维度和化简压缩，这个过程被称为池化作用（Pooling）。一般来说，池化作用就是将特征矩阵划分成N个a*b型的池化区域，然后对于每个池化区域里的数值进行取平均值或最大值的操作，从而用所得的结果进行参加后续的训练过程。

如图3-12所示，这是一个4*4的输入特征矩阵，将其划分成4个2*2的池化区域后取每个池化区域中参数的最大值，最后得到一个2*2的输出矩阵传入下一层。

图3-12 池化层作用示例

3.3.3全连接层

将进行多次卷积和池化作用的特征矩阵输入到全连接层进行高度得抽象化。卷积神经网络结构中的全连接层通常是上一层的神经元与当前层的所有神经元相链接，将所得到的特征矩阵转变为1*n的链式特征向量，而后与输出层的每个神经元也都链接（如图3-13所示范例），而后通过softmax函数进行辨别。虽然此时的特征十分简练，甚至已经丢失了不少信息，但是却将图像中最明显的特征保存下来以此为凭借进行分类。

我们给定训练集，其中 $x(i)$ 是第i个输入值，而 $y(i)$ 则是其分类的标签，设第j类的第i个预测值为 \hat{y}_{ij} ，从而，我们利用于分类的softmax函数公式：来进行对其正则化处理，而后输出。

图3-13全连接层结构

3.4深度残差网络

研究者们渐渐发现，随着对人工神经网络的深度不断地加深时，不仅对实验数据集的预测准确率会有一个不断上升达到峰值后会很快地下降，而且对于训练数据集的准确率也会有先增高再下降的曲线趋势。于是针对于这个问题，随着研究者的不断探索，“深度残差网络”（ResNet）的概念在2015年被提出来，深度残差网络允许对于ANN神经网络层系数的增加。

深度残差网络较之以前的神经网络添加了残差块，结构如图3-14所示，这种结构新添了一种新的连接方式“shortcut connection”，即将某一层的输出不经过下一层而直接“走捷径”输出到下n层的位置输入。

图3-14 经典残差块结构

在本文中，我们使用的是ResNet34结构，ResNet34结构通常被分成分别是：conv1、conv2_x、conv3_x、conv4_x、conv5_x这五个部分再加上一个全连接层用于分类。在conv1部分包含了卷积层和池化层，而后分别在conv2_x部分包含了三个残差块、conv3_x部分包含了四个残差块、conv4_x部分包含了六个残差块、conv5_x部分包含了三个残差块。而如图3-15所示，是ResNet34模型中某个残差块的结构，可以看到每个残差块均有两层结构，所以ResNet34确实是有层（其中不包含池化层）。而ResNet34区别于以往的34层神经网络结构的优点是保证了信息的完整性，在利用梯度下降法作优化问题时，以往的神经网络结构传递梯度时最后接收梯度的层往往容易使得梯度趋近于零，无法进行矫正修改，而ResNet34则不再有这个忧患，从而使训练的模型更加精准，从而简化了学习目标和难度。

图3-15 ResNet34残差块的结构

3.5分类器

分类器的作用就是利用已经明确给定好类别的训练集数据进行分类训练后再对未知类别的实验数据集根据提取到的特征进行预估和分类。根据监督方式的不同，通常将分拣归类器分为有监督式分类、半监督式分类及无监督式分类三种。其中：有监督式分类是运用人工标注过的训练集进行分析训练，目前比较火的支持向量机算法、人工神经网络以及感知器算法均属于有监督式分类；而半监督式分类则是运用仅有一部分进行人工标注的训练集，不仅运用其人工标注同时还要利用该训练集本身信息进行分类学习的分类方法；无监督式分类是指运用未经过人工进行任何标注的训练集仅仅通过其训练集本身所拥有的信息进行学习分类的分类方法。本文研究的是利用支持向量机分类器来代替卷积神经网络中的全连接层来进行分拣归类。

1995年，Cortes等人提出了支持向量机算法（即为SVM）。支持向量机分类器除了对于可线性分类意况能够处理分析之

外，还能够将低维度的线性不可分意况转成高纬度问题，而后在高纬度空间下应用核函数展开定理来进行线性分析分类。如图3-16所示的例子中，在一个二维平面之中有一个由有限个二维坐标值构成的线性可分点集合，而其中的就是支持向量，如何在许多条划分直线中找到最好的那条划分直线？

图3-16 某个二维线性可分集合

如图3-17所示，利用支持向量机算法，首先我们可以通过公式，令 $g(x)=0$ 找到一个超平面将两个分类集合完全分开， $g(x)$ 可以看成是到超平面的距离。若要求得最优超平面，则需要使得这个超平面距离支持向量愈远愈好。而后设与超平面相垂直的向量为 w ， r 为到超平面的垂直距离，所以我们可以得出。为了追求最有分界面，显然 r 的值愈大愈好，由此可得出结论，当最小时，我们获取了最优分界面。所以最后整理可得。

图3-17 支持向量机算法

第4章系统实现与测试

4.1实验需求

基于视频流的人体行为分类建模研究实验所要研究的内容是对含有待检测人类躯体动作行为的实验视频流，进行人体信息的监测，人体特征的提取，建立行为分类模型，并评估结论的准确性和可靠性。

4.2结构设计

图4-1 结构流程图

本次实验的流程如图4-1所示，我们首先通过运用FFmpeg将视频流进行预处理，将视频流切分图像，然后将处理好的图像次序排列输入到已经训练好的3D-ResNet34模型进行特征提取，然后使用SVM分类器代替模型中的全连接层对特征数据实行分类运算，最后对分类结果进行评测估略。

4.3系统实现

虽然2D-ResNet34模型对于图像中方向关系有一定的辨别能力，但是没有涉及到视频流的“时间”维度，所以我采用了3D-ResNet34模型，对图像中在三维时空内部对人体的定位更加精确。3D-ResNet34模型较之2D-ResNet34模型来说多了一个“depth”参数，从而可以将连续的几帧图像按照次序堆叠起来（如图4-2所示），这样就解决了动作的连续性问题。但是由于从训练一个新的3D-ResNet34模型作为开始既要耗费大量的时间和精力，同时也对于设备的要求非常高，所以如果使用一个已经预先训练好的ResNet34模型作为我们的模型根基将会大大节省我们的时间提高了我们的效率，所以我们采用了一个在Kinetics数据库上预先训练好的ResNet34模型，并在此根基之上加以改良。我们在ubuntu16.04操作系统下将以Pytorch作为深度学习框架。PyTorch是由Facebook公司开发的一款开源深度学习框架，因为它能够自动求导，让我们无需再人工编写代码进行梯度求解的过程，同时，PyTorch不仅基于Python语言，使我们操作简单，并且在底层接口使用了C/C++，运行快捷。况且，PyTorch还支持GPU加速，所以我们还使用了英伟达TitanXp作为硬件加速卡进行了对我们残差神经网络的训练数据集及实验数据集的图像处理。

图4-2 3D-ResNet34使用3D卷积核进行卷积操作

4.4数据集预处理

我们使用了UCF101数据库和HMDB51数据库中的视频流，如图4-3所示，UCF101数据库包含来自BBC以及ESPN电视频道采集到的人体的101类行为动作视频样本，其中每一类由25个人做动作，每个人做4-7组（如图4-4所示为做太极拳的动作的视频流集合），共有13320个视频，视频流的分辨率为320*240。而HMDB51数据库是由美国Brown大学在2011年发布的包含来源于电影中的人体动作行为的51类动作视频样本，而每类之中至少包含101个视频流样本，一共6849段视频。然后对数据集中的每一个视频流都用FFmpeg进行切分，将视频流按每秒10帧进行切分，将得到的图像按照顺序进行排列放入一个文件夹之中。图4-5所示为其中一个太极拳视频流经过切分后得到的部分图像数据集。

而后将切分后的图片进行预处理操作，首先将图片压缩成112*112大小，而后将图片进行归一化处理，即通过数学计算，我们把图像中的RGB通道的值以（114.7748，107.7354，99.4750）为均值，方差为（1，1，1）进行正则化处理。

图4-3 原有UCF101数据集分类

图4-4 含有太极拳动作的未切分数据集

图4-5 一个含有太极拳动作的视频流切分后的图像集合

4.5训练与测试

我们用预训练好的3D-ResNet34模型将处理好的图像次序排列逐一读取，进行对图像中人类躯干的检测及特征提取，此时模型输出的图像的特征数据生成.json格式的文件进行保存（如图4-6，是对图4-5中的图像集合进行特征检测得到的特征数据）。

图4-6视频流经3D-R34模型提取出的特征数据

此时，我们再分别使用取平均值和最大值两种特征降维处理方法的SVM分类器将分别按照不同的比例随机划分训练集和实验数据集进行分拣归类的训练和实验，来求解不同降维方法以及训练集所占百分比对实验分类正确率的影响。得到如图4-7的结果，可以很明显的看到当训练集的比例占总比例愈来愈高时，花费的时间代价愈来愈大，对实验数据集分类的精确率也逐渐提高。

图4-7 实验结果图

如图4-8所示，是SVM分类器对于UCF101数据集上的视频流进行分类经过十次随机抽取得到准确率均值的结果。通过进

一步分析可以看到，无论是运用取平均值的降维方法还是取最大值的降维方法对于分类效果的影响并不大，而训练集的大小比例对于SVM分类器的分类精准度影响较大。当训练集为10%的时候，准确率仅为87.0%左右，但是随着比率不断升高，分类的准确率也不断升高，最终在训练集为90%时，准确率达到最高约为97.5%左右。

图4-8 UCF101数据集实验结果

通过图4-9的折线对比图可看出，当训练集在从10%增加50%的过程中，准确率极速升高，但是升高至50%以后，折线趋于平缓，但是由图4-7可知，在这个过程中需要花费的时间却18.60s到30.41s提高至原有的1.6倍左右。所以由此我们可得出结论，在此次实验中，倘若我们对分类结果的准确率只是要求在95%以上时，我们可以选择与实验数据集数量大小相同或相近的数据集作为训练集即可。

图4-9 UCF101数据集实验结果均值对比分析

如图4-10所示，是SVM分类器对于HMDB51数据集上的视频流进行分类经过十次随机抽取得到准确率均值的结果。同样的，训练集的大小比例对于SVM分类器的分类精准度影响仍然较大。当训练集为10%的时候，准确率仅为52.5%左右，但是随着比率不断升高，分类的准确率也不断升高，最终在训练集为90%时，准确率达到最高约为65.0%左右。但是经过十次反复随机选取训练集和实验集数据进行测评时，我们发现无论怎样选取训练集的数据进行对SVM分类器的训练后再进行对实验数据集的分类，其分类的准确率较之在UCF101数据库上进行训练后分类的高准确率来说会远远低于这个值。笔者通过对于UCF101数据库及HMDB51数据库中的视频流进行分析和对比之后，发现UCF101数据库中的每一个动作行为分类一百多个含有人体动作行为的视频流样本中，二十五个实验者的体型、体态、甚至动作幅度都惊人地相似，而且对于图像序列中的背景环境比较简单，不易造成干扰。但是在HMDB51数据库中的动作行为分类中，虽然每一类中也包含有至少一百个视频流，但是由于这些视频流是人为地从不同的电影视频中提取出来，往往视频流中包含的人物的形象、体型、体态和动作幅度都不同，虽然可能有相同的人物，但是演员对于不同的电影中的不同角色的演绎，往往相同的动作也会有通过不同的动作幅度和细节处理来表达对于不同环境不同性格的演出。况且，再加上导演对于电影的艺术化处理，往往这些提取出来的图像序列中的背景很复杂，会给我们的3D-ResNet34模型在特征提取时造成很大的阻碍，从而使得分类效果并不理想。

图4-10 HMDB51数据集实验结果列表

如图4-11所示的折线对比图可看出，SVM分类器在HMDB51数据集上进行分类时，通过折线变化可以分析得到，与在UCF101数据集上分类无论是运用取平均值的降维方法还是取最大值的降维方法对于分类效果的影响并不大的这一特点不同的是，当训练集占总视频流的数量比超过40%时，取平均值操作来降维的方式要远远比取最大值的降维操作来进行分类精准。而与在UCF101数据集上分类相似的一点是当训练集在从10%增加60%的过程中，准确率极速升高，但是升高至60%以后，折线趋于平缓，并且由图4-11可知，在这个过程中需要花费的时间却9.32s到17.49s提高至原有的1.8倍左右，时间代价却高出了接近两倍。所以由此我们可得出结论，在此次实验中，倘若我们对分类结果的准确率只是要求在60%以上时，我们可以选择与实验数据集数量大小相同或相近的数据集作为训练集即可。

图4-10 HMDB51数据集实验结果均值对比分析

图4-11 HMDB51数据集实验运行结果

4.6小结

本章首先对实验需求的分析和理解，然后给出了实验系统的框架设计，并给出了流程图。并且，通过对目前现有的基于视频流的人体行为分类建模技术的了解再结合自己现有的硬件设备条件，决定使用了3D-ResNet模型进行实验，并结合SVM分类器以提升其分类准确率。本实验通过选取不同视频流数据库上的样本进行实验来测试该架构模型，并配上对不同条件下得到的实验结果数据图以及折线图的对比分析来总体评价所设计的结构模型的优点和不足。最后得出结论，本文所设计的分类模型满足了实验需求，尤其是在相对简单背景下，对与和训练集包含有相同或相似的人体动作行为的实验集进行分类的效果最佳。

5. 第5章总结与展望 总字数：1490

相似文献列表 文字复制比：0%(0) 疑似剽窃观点：(0)

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

第5章总结与展望

本文将UCF101数据集和HMDB51数据集的视频流首先通过运用FFmpeg软件将视频流进行预处理，将视频流切分图像，然后将处理好的图像次序排列输入到已经训练好的残差神经网络3D-ResNet34模型进行特征提取，然后使用SVM分类器代替模型中的全连接层对特征数据实行分类运算的方法进行了对人体行为分类建模的研究，建立了一个基于视频流的人体行为分类的模型，并对其进行了训练与测试。对不同的训练集、不同的特征降维操作以及训练集所占比例的不同，经过多次实验后，我们对结果的对比分析，最终得出了在UCF101数据集上运用取最大值的降维操作随机取90%的数据当作训练集时，实验得到的分类准确率最高，约为97.6%；在HMDB51数据集上运用取平均值的降维操作随机抽取90%的数据当作训练集时，实验得到的分类准确率最高，约为68%。基于视频流的人体行为分类建模研究在许多领域都有应用，例如在自动录像监控系统中对于人像识别、对于辅助独居老人日常生活方面的智能监控系统的实现等方面都有着建设性的作用。但是仍存在有许多待解决的问题。

题，例如：切分视频流的方法会丢失一些原始视频中的信息从而导致了准确率的下降；在3DResNet34模型中对于连续图像的跳帧选取而导致特征的提取不够全面；对于在较之稍复杂的场景下的人体动作行为的特征提取技术还不够成熟；对于视频流中对同一动作的不同人会有人体的性别、体型、体态、动作幅度等方面的差异从而导致分类的准确率不高等问题存在。对此，为了改进这些不足，我们还将进行更深一步的探索和研究。

参考文献

- [1] G Johansson . Visual Perception of biological motion and a model for its analysis [J] . Perception & Psychophysics . 1973 , 14 (2) . 201—211 .
- [2] Collins R , Lipton A , Kanadeand T . A System for Video Surveillance and Monitoring . VSAM Final Report [R] . Robotic Institute Carnegie Mellon University , 2000 .
- [3] Collins R , Lipton A , Kanadeand T . Introduction to the Special Section on Video Surveillance [J] . IEEE Transaction on Pattern Analysis and Machine Intelligence , 2000 , 22 (8) : 745—746 .
- [4] Gaßner K . , Conrad M . 2010 , ICT enabled independent living for elderly , Astatus—quo analysis on products and the research landscape in the field of Ambient Assisted Living (ALL) in EU-27 , [http : //www .aal—deutschland .de / deutschland/ dokumente / ict_for_elderly_webversion . pdf](http://www.aal—deutschland.de/deutschland/dokumente/ict_for_elderly_webversion.pdf) .
- [5] Gustafsson L . 2011 . The Ambient Assisted Living Joint Programme(AALJP) . [http : //www . helmholtz-muenchen . de/ fileadmin / JOIN/PDF / AALProjectsPorfolio . pdf](http://www.helmholtz-muenchen.de/fileadmin/JOIN/PDF/AALProjectsPorfolio.pdf) .
- [6] Taleb T , Bottazzi D , Guizani M , Nait—Charif H . 2009 . ANGELAH : A Framework for Assisting Elders at Home . IEEE Journal on Selected Areas in Communications , 27 (4) .
- [7] Jara A J , Zamora—Izquierdo M A , Gomez—Skarmeta A F . 2009 . An Ambient Assisted Living System for Telemedicine with Detection of Symptoms . Bioinspired Applications in Artificial and Natual Computeation , Lecture Notes in Computer Science , 5602 : 75—84 .
- [8] Eagle N , Pentland A . 2006 . Reality mining : sensing complex social systems [J] . Personal and Ubiquitous Computing , 10 (4) : 255—268 .
- [9] Logan B , Healey J , Philipose M , et al . 2007 . A long-term evaluation of sensing modalities for activity recognition [C] . Ubiquitous Computing .
- [10] Har Itaoglu I , Harwood D , Davis L . W4 : real—time surveillance of people and their activities [J] . IEEE Trans on Pattern Analysis and Machine Intelligence . 2000 , 22 (8) . 809—830 .
- [11] Gonzalez J , Rowe D , Varona J , et al . 2009b . Understanding dynamic scenes based on human based on hunman sequence evaluation [J] . Image and Vision Computing , 27 (10) : 1433-1444 .
- [12] Aggarwal J K , Cai Q . 1999 . Human motion analysis : a review [J] . Computer Vision and Image Understanding , 73 (3) : 428-440 .
- [13] Gavrila D M . 1999 . The visual analysis of human movement : a survey [J] . Computer Vision and Image Understanding , 73 (1) : 82—98 .
- [14] Buxton H . 2002 . Learning and understanding dynamic scene activity : a review [J] . Image and Vision Computing , 21 (1) : 125—136 .
- [15] Wang L , Hu W , Tan T . 2003 . Recent developments in human motion analysis [J] . Pattern Recognition , 36 : 585—601 .
- [16] Moeslund T B , Hilton A , Kruger V . 2006 . A survey of advances in vision—based human motion capture and analysis [J] . Computer Vision and Image Understanding , 104 : 90—126 .
- [17] Wang L , Hu W , Tan T . 2003 . Recent developments in human motion analysis [J] . Pattern Recognition , 36 : 585—601 .
- [18] Du Y , Chen F , Xu W , et al . 2008 . Activity recognition through multi—scale motion detail analysis [J] . Neurocomputing , 71 : 3561—3574 .
- [19] A. Verri , S. Uras , E. DeMicheli . Motion Segmentation from optical flow [A] . The 5th Alvey Vision Conference [C] . 1989 . 209—214 .
- [20] Lipton A , Fujiyoshi H , Patil R . Moving target classification and tracking from real—time video . In : Proc IEEE Workshop on Applications of Computer Vision , Princeton , NJ , 1998 . 8—14 .
- [21] Anderson C , Bert P , Vander Wal G . Change detection and tracking using pyramids transformation techniques . In : Proc SPIE Conference on Intelligence Robots and Computer Vision , Cambridge . MA , 1985 . 579. 72—78 .
- [22] Lowe D G . Distinctive image features from scale-invariant keypoints [J] . International journal of computer vision , 2004 , 60 (2) : 91—110 .
- [23] T Serre , L Wolf , S Bile chi , M Riesenhuber , T Poggio . Robust object recognition with cortex-like

mechanisms [J] . IEEE Transactions on Pattern Analysis and Machine Intelligence , , 2007 . 29 (3) : 411—419 .

[24] Schmidhuber J . Deep Learning in neural networks : An overview . [J] . Neural Netw , 2014 , 61 : 85—117 .

Bengio Y . , Courville A . , Vincent P . Representation Learning : A Review and New Perspectives . IEEE Transactions on Pattern Analysis and Machine Intelligence . 2013 , 35 (8) : 1798—1828 . arXiv : 1206 . 5538 .

[25] Hubel David , Wiesel Torsten . David Hubel and Torsten Wiesel . [J] . Neuron , 2012 , 75 (2) .

[26] Chen W , Pourghasemi H R , Kornejady A , et al . Landslide spatial modeling : Introducing new ensembles of ANN , MaxEnt, and SVM machine learning techniques [J] . Geofisica Internacional , 2017 , 305 : 314-327 .

[27] Surhone L M , Tennoe M T , Henssonow S F . Activation Function [M] . 2010 .

[28] Chua L O , Roska T . CNN paradigm [J] . IEEE Transactions on Circuits & Systems I Fundamental Theory & Applications , 1993 , 40 (3) : 147-156 .

致谢

时间如白驹过隙，经历了半年的时光，从论文的选题时候的无措彷徨到坚定方向确立题目；至为了搜集和整理大量资料而心慌意乱到沉着从容；从开题报告的慌乱不安到写论文时反复修改的焦躁难安；再到现在的尘埃落定，期间经历了痛苦彷徨和喜悦自信，在写作论文的过程中不仅仅是心情喜怒哀乐的变化，更多的是对自己心态的历练以及自我能力的修炼。如今，伴随着这篇毕业论文的最终成稿，自己甚至还些许有了一点成就感。

首先在这里，我要向我的导师于哲舟教授致以我最诚挚的感谢。于教授在我论文定题、搜集资料、构思到定稿期间给予了我很多的帮助。于教授作为老师，春风风人，夏雨雨人，在百忙之中仍然抽出时间悉心指导，诲人不倦；于教授作为长辈，在我茫然失措时谆谆教导，指点迷津，以自身为榜样，在我人生的道路上为我点亮前行的道路。无论是为人还是治学，于教授都是我为人做事的榜样，能够师从于教授是我的幸事。

其次，我要感谢我的父母。诗经《蓼莪》曾言“哀哀父母，生我劬劳；哀哀父母，生我劳瘁”，在我彷徨于时，永远在我的身后，作为我最坚实的后盾，让我依靠给予我力量让我重新扬帆起航，用自己不再年轻的身体为我保驾护航，他们给予我的关怀和爱护是我不断前行的动力源泉。

与此同时，我要感谢我的同学们，同窗四年，我们在一起经历了无数难忘的时光。感谢有你们，一如既往的关心与支持，鼓励和帮忙。在我遇到困难和难题时，无私地帮助和热心的关怀，使我如沐春风，增添信心。四年间同学之间的情谊，必将点点滴滴留存与我心间。

最后，我必须还要感谢我的母校——吉林大学，是她给了我一个广阔无垠的知识海洋和展现自我的舞台，让我不断更新自我，不断地改善自我，洗涤我的灵魂，增强我的意志。

宁静、温馨的四年校园生活就此远去，但是我们的生命历程中挑战的还没有结束。在未来的生活航海之中，我们肯定还会遇到更加多元化的挫折与艰难，海浪与暴风，但是不论怎样，我仍然会披荆斩棘，不断奋进，不断地完善自我，前进的号角也永不停歇！

说明：1.总文字复制比：被检测论文总重合字数在总字数中所占的比例

2.去除引用文献复制比：去除系统识别为引用的文献后，计算出来的重合字数在总字数中所占的比例

3.去除本人已发表文献复制比：去除作者本人已发表文献后，计算出来的重合字数在总字数中所占的比例

4.单篇最大文字复制比：被检测文献与所有相似文献比对后，重合字数占总字数的比例最大的那一篇文献的文字复制比

5.指标是由系统根据《学术论文不端行为的界定标准》自动生成的

6.红色文字表示文字复制部分;绿色文字表示引用部分

7.本报告单仅对您所选择比对资源范围内检测结果负责



 amlc@cnki.net

 <http://check.cnki.net/>

 <http://e.weibo.com/u/3194559873/>