

**专 业 学 位 硕 士 学 位 论 文**

**基于图卷积神经网络和长短时记忆模型的行人通行意图预测方法**

**Pedestrian Intent Prediction Method based on Graph Convolutional Networks and** **Long Short-Term Memory Model**

作 者 姓 名： 许晓君

工 程 领 域： 车辆工程

学 号： 31803163

指 导 教 师： 祝雪峰

完 成 日 期：

大连理工大学

Dalian University of Technology

大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目：

作 者 签 名 ： 日期： 年 月 日

摘 要

在自动驾驶场景中，准确预测行人行为意图是至关重要的技术之一。现有方法中，行人意图的预测多是基于运动轨迹和姿态动作，对行人身体固有的信息利用较少。因此，本文在使用行人过马路时的骨骼关键点信息的同时，增加面部关键点信息,两者共同作为预测网络的输入。构建基于图卷积神经网络(GCN)和长短时记忆模型(LSTM)的行人通行意图预测网络，提取空间序列特征和时间序列特征，以获得良好的预测结果。本文的主要研究内容如下：

(1)制作行人马路通行行为数据集。通过摄像机进行学校内部以及周围区域行人通行视频的采集，视频分为两类采集，分别是行人过马路和不过马路的视频。将采集到的行人视频进行合理的裁剪并进行视频帧分离，针对分辨率不高的视频帧进行分辨率的重建以提高视频帧质量。

(2)骨骼关键点和面部关键点的提取。本文中需要对处理后的视频帧进行行人骨骼关键点和面部关键点的检测，从而根据提取到的关键点的位置信息来预测行人是否会过马路。将提取到的行人关键点位置制作成CSV格式的数据集，便于输入预测网络。本文选用开源库Openpose进行关键点信息的提取，保证检测的准确度和鲁棒性。关键点包括人体骨骼的25个关键点和面部的70个关键点，共95个关键点。

(3)基于图卷积神经网络(GCN)和长短时记忆模型(LSTM)来预测行人通行意图。将CSV格式的数据经过处理后，输入到图卷积神经网络和长短时记忆模型结合的预测网络中，对视频帧的关键点信息进行时间和空间上的特征提取。通过训练网络，找到行人通行意图与行人的骨骼关键点和面部关键点的位置之间的内在联系，使得该预测网络能够根据行人的关键点的位置变化对行人是否通过马路进行判断。

骨骼关键点的变化直接反应了行人的行走动作规律。面部关键点的变化可以从侧面反映出人的意图从而预测人的行为。因此，我们使用骨骼关键点与面部关键点双重信息，将其输入到神经网络进行训练，结合图卷积神经网络处理空间序列信息的能力和长短时记忆模型处理时间序列信息的能力，从而使得网络能够学会根据关键点信息预测行人是否过马路，更有效的预测行人的通行意图。

关键词：行人通行意图预测；图卷积神经网络；长短时记忆神经网络；骨骼关键点；面部关键点

Prediction Method of Pedestrian Crossing Intention Based on Graph Convolutional Network and Long Short-Term Memory Model

Abstract

Accurately predicting pedestrian intentions is one of the most important technologies in autopilot scenarios. Of the existing methods, the prediction of pedestrian intention is mostly based on movement trajectory and posture movement, and it makes little use of the inherent information of pedestrian body. Therefore, in this paper, not only key points information of bone but also key points information of face are used as the input of the prediction network. A pedestrian intention prediction network based on graph convolutional network (GCN) and long short-term memory model (LSTM) is established in this paper, and spatial sequence features and time sequence features are extracted to obtain good prediction results. The main contents of this paper are as follows:

(1) Dataset of pedestrian crossing behavior production. The videos of pedestrians crossing were collected by camera in and around the school. The videos can be divided into two categories: the videos of pedestrians crossing the road and the videos of pedestrians failing to pass the road. The collected pedestrian videos were clipped reasonably and the video frames were separated. The video frames with low resolution were reconstructed to improve quality.

(2) Extraction of key points of bone and face. In this paper, the key points of bone and face were detected from the processed video frames, so as to predict whether the pedestrians will cross the road according to the location information of the extracted key points. The location information of the extracted key points was made into CSV format to input into the predictive network easily. In this paper, the open source library Openpose is used to extract key points to ensure the accuracy and robustness of detection. The key points include 25 key points of human bone and 70 key points of face, a total of 95 key points.

(3) The pedestrian intention is predicted based on graph convolutional network (GCN) and long short-term memory model (LSTM). After processing the data in CSV format, the processed data was used as input for predicting network combining LSTM and GCN, and the feature extraction in time and space was performed on the key points of the video frames . By training the network, the internal connection between pedestrian's intention of pedestrians and the location of the pedestrian's key points of bone and face will be found, so that the network could judge whether the pedestrian will cross the road according to the location change of the pedestrian's key points.

The change of the key points of bone directly reflects the walking movement rule of the pedestrian. The change of the key points of face can reflect the person's intention from the side and predict the person's behavior. As a result, we use key points of bone and key points of face as input to the neural network for training, combine the ability of graph convolutional network to process the spatial sequence information of graph structure and the ability of long short-term memory model to process the time sequence information, so as to make the network learn to predict whether the pedestrian will cross the road according to the key points information and predict the intention of the pedestrians more effectively.

**Key words:** Pedestrian Intention Prediction; Graph Convolutional Network; Long Short-Term Memory Model; Key Points of Bone; Key Points of Face

目 录

[摘 要 I](#_Toc40470432)

[Abstract II](#_Toc40470433)

[1 绪论 1](#_Toc40470434)

[1.1 课题研究背景和意义 1](#_Toc40470435)

[1.2 行人通行意图预测研究现状 2](#_Toc40470436)

[1.2.1 行人通行意图预测国内外相关研究现状 2](#_Toc40470437)

[1.2.2 行人通行意图预测算法的不足 6](#_Toc40470438)

[1.3 深度学习研究现状 6](#_Toc40470439)

[1.3.1 深度学习的概念理解 6](#_Toc40470440)

[1.3.2 深度学习的发展历史 7](#_Toc40470441)

[1.3.3 深度学习在自动驾驶上的应用 10](#_Toc40470442)

[1.4 本文主要研究内容 13](#_Toc40470443)

[1.5 本文技术路线 14](#_Toc40470444)

[1.6 本章小结 15](#_Toc40470445)

[2 卷积神经网络和长短时记忆神经网络概述 16](#_Toc40470446)

[2.1 卷积神经网络概述 16](#_Toc40470447)

[2.2 长短时记忆神经网络模型 27](#_Toc40470448)

[2.3 本章小结 31](#_Toc40470449)

[3 数据集的制作 32](#_Toc40470450)

[3.1 视频的采集与处理 32](#_Toc40470451)

[3.2 视频帧的分辨率重建 32](#_Toc40470452)

[3.3 行人关键点的提取 34](#_Toc40470453)

[3.3.1 Openpose介绍 34](#_Toc40470454)

[3.3.2 Openpose原理 35](#_Toc40470455)

[3.3.3 关键点位置信息表示 39](#_Toc40470456)

[3.4 CSV格式的数据集制作 40](#_Toc40470457)

[3.5 本章小结 41](#_Toc40470458)

[4 基于图卷积神经网络和LSTM的行人通行意图预测 42](#_Toc40470459)

[4.1 图卷积神经网络原理概述 42](#_Toc40470460)

[4.2 基于图卷积和LSTM的预测网络 44](#_Toc40470461)

[4.2.1 网络结构 44](#_Toc40470462)

[4.2.2 损失函数 46](#_Toc40470463)

[4.2.3 优化方法 47](#_Toc40470464)

[4.2.4 Sigmoid分类函数 47](#_Toc40470465)

[4.3 预测结果对比及验证 47](#_Toc40470466)

[4.4 本章小结 49](#_Toc40470467)

[5 总结与展望 51](#_Toc40470468)

[5.1 论文总结 51](#_Toc40470469)

[5.2 不足与展望 51](#_Toc40470470)

[5.2.1 本方法的优势与不足 51](#_Toc40470471)

[5.2.2 未来展望 52](#_Toc40470472)

[参 考 文 献 53](#_Toc40470473)

[致 谢 58](#_Toc40470474)

[大连理工大学学位论文版权使用授权书 59](#_Toc40470475)

1 绪论

1.1 课题研究背景和意义

近年来，自动驾驶已经成为汽车领域的研究热点之一。自动驾驶汽车能够避免由于驾驶员的错误操作引起的交通事故，而且能够避免酒后驾驶、恶意驾驶等不当驾驶行为，从而提高道路交通的行驶安全性。尽管自动驾驶汽车发展的越来越成熟，但是仍然不能打消行人对其安全性的疑虑，比如当自动驾驶汽车在斑马线前面停止前进，车前行人会怀疑是否可以在这个时候通过马路。当车人交流出错时，就会容易发生碰撞事故或者行人并无过马路的意图时车辆却在原地长时间等待造成的时间浪费。优步科技公司的一辆正在公路上测试的自动驾驶汽车撞上了一名推着自行车过马路的女子，最终女子被送往医院后不治身亡。美国汽车协会对装备了自动驾驶传感器和软件的雪佛兰迈锐宝、本田雅阁、特斯拉Model3和丰田凯美瑞做了一次汽车与行人碰撞测试，结果不尽人意，选取的车辆都没有很好的对行人做出规避，甚至发生了直接撞上假人的情况。如果车辆在真实的交通车道上行驶，这样的结果带来的后果是十分严重的。

行人作为交通环境的主要参与者，主动性强，动作灵活多变，运动轨迹变化迅速，行人们可以在很短的时间内就改变自身的行走速度或者方向。如果对行人的通行意图产生了误判可能会发生严重的后果，因此准确的预测行人的通行意图至关重要(如图1.1)。研究表明，运动预测的错误率与意图估计之间呈次线性关系。因此，尽管在大于1秒的时间范围内，运动预测的性能较低，但意图估计仍然相对稳定[1]。所以预测行人意图具有一定的稳定性和可预测性。自动驾驶汽车如果能够更好的预测行人的通行意图，就可以减少因车和行人交叉前进造成的碰撞伤害以及驾驶员误以为行人过马路而造成的长时间等待，节约驾驶员和行人在马路上的时间消耗，提高行人通行马路的安全性。



图1.1 行人过马路意图预测

Fig. 1.1 Pedestrian crossing intention prediction

目前的自动驾驶汽车需要一个方法能够采取行人过马路时的有效信息提供给汽车，使得驾驶员可以通过传递回的信息来辅助驾驶员做出等待行人通过或者直接前行的决定。常见方法有：基于行人运动路径、基于行人的动作、基于场景的上下文信息以及时空轨迹建模等，但是这些方法均未考虑行人的面部关键点信息，没有将行人的骨骼关键点信息和面部关键点信息进行结合，信息利用的不够充分。这是亟待解决的一个问题。

1.2 行人通行意图预测研究现状

1.2.1 行人通行意图预测国内外相关研究现状

本研究主要目的是利用神经网络学习行人过马路时的骨骼关键点和面部关键点信息的内在变化规律，从而对行人的通行意图进行预测，辅助自动驾驶汽车做出决定，避免或减少行人与汽车发生碰撞的可能性或者由于误判行人通行意图而导致的长时间等待。因此，下文就行人通行预测的相关国外研究现状、国内研究现状进行讨论。目前国内外研究可以大致分为三类：基于模型、基于算法以及基于数据，前两类是主要方法。接下来的国内外研究现状分析会依次从这三方面展开。

（一）国外研究现状

Himanen等人提出一个多项的logit模型被用来检查行人和驾驶员对在人行横道上发生的“偶遇”的反应。他们根据行人、环境和交通状况的不同，可以确定驾驶员刹车或迂回的可能性，以及行人在遭遇事故后继续通过马路的可能性[2]。印度统计研究所Manuszak等人研究了行人希望在有信号的十字路口穿过车流的行为。提出了参数化和非参数化方法来估计行人数量中临界缺口的分布[3]。Gandhi等人综述了基于形状等线索的各种方法，用可见和非可见光传感器探测行人的运动。对行人行为概率建模研究，用于预测行人与车辆之间的碰撞[4]。德国阿沙芬堡大学Goldhammer等人主要研究行人在人行横道上起始过程中短时轨迹的早期预测，提出两种模型，分别为分段线性模型和乙状结肠模型，用于预测从脚跟开始的轨迹。研究结果为基于车辆的碰撞风险估计提供了依据，从而为驾驶员可能的预警、自主紧急制动或规避转向提供了依据[5]。Bonnin提出了一个通用的基于上下文模型来预测城市中心的行人过马路行为，并证明模型在早期提供了准确的预测。开发了一个附加的特定模型以适应斑马线的环境。实验表明，该模型能在特定的环境下产生更好、更早的预测[6]。Volz等着重于预测行人在接近人行横道时的过马路时间。由于底层模型的复杂性，提出了一种通过回归模型学习目标变量的数据驱动方法，使用分位数回归代替标准均值回归。实验证明这种特殊类型的回归更适合于描述真实世界中行人轨迹的可变性[7]。Zhao等人总结了使用路边激光雷达传感器监测和预测行人过马路意图的实践，提出了一种基于深度自编码的人工神经网络(DA-ANN)方法来处理行人过马路的数据并预测行人过马路的意图。实例分析表明，该模型对实时系统的预测精度和计算效率均在95%左右[8]。

Abramson等人提出了一种高效的行人图像分类算法，该算法利用粒子滤波对各种算法的结果进行融合，并对行人进行跟踪，为预测未来行人的运动提供了一个很好的工具[9]。Ngai等人介绍了一种针对特定场景的最优应急机动识别算法，与现有技术相比，其优势在于考虑了行人的受伤风险以及不确定的未来动作[10]。Coelingh等人介绍了一种最新的汽车全制动行人检测系统(CWAB-PD)。它通过发出警告，并在必要时使用全制动功率自动制动，帮助驾驶员避免追尾事故和行人事故。实验结果表明，CWAB-PD系统可以避免高达35公里/小时的事故，并可以减轻事故，实现35公里/小时的冲击减速[11]。Goldhammer等人提出了一种基于交互多模型扩展卡尔曼滤波(IMM-EKF)的行人跟踪方法。此外，提出了一种基于运动轮廓图像的描述符(MCHOG)与支持向量机(SVM)分类相结合的新方法，该方法在智能基础设施的约束下，在初始步骤中就能达到99%的准确率[12]。Kohler等人提出了一种主动的行人保护系统，在城市交通中，当无法通过刹车来避免碰撞时，自动保持车道的规避策略。通过基于视频的运动梯度运动轮廓图和静止检测，证明了车辆在行人足部进入车道前就可以进行规避机动[13]。Braeuchle等人介绍了一种针对特定场景的最优应急机动识别方法。与现有技术相比，其优势在于考虑了行人的受伤风险以及不确定的未来动作。因此，可以量化某些应急操作的差异，并获得最佳效益[14]。Gepperth等人提出了一个利用图形处理单元(GPU)计算能力的实时行人检测和姿态分类系统，提出的姿势分类可以确定行人的方向和未来可能的运动[15]。Kropf等提出了一种基于不确定性行人模型的概率态势分析方法，考虑到行人受伤风险的最优降低，选择紧急机动[16]。Keller等人研究了行人路径预测和短时亚秒级行为分类问题。新提出的方法可以在停车0-0.77s的时间范围内，实现10-50cm的更准确的位置预测[17]。Kwak等人提出了一种新的算法可以预测行人的意图。为了在连续序列中预测行人的意图，使用动态模糊自动机(DFA)方法，利用路缘与行人之间的距离以及行人的速度和头部方向等时空待征进行预测[18]。Voelz等人提出了一种将运动跟踪算法与数据驱动相结合的行人运动预测方法。该方法建立在层次结构上，首先对每个行人的意图进行分类，然后计算几个定性指标来进行预测[19]。

Oxley等人做了两个实验，在一个模拟的十字路口环境中，检查行人年龄差异对人车差距选择决策的影响。结果表明，在所有年龄组中，间隔选择主要基于车辆距离，较少基于到达时间[20]。Schmidt等人开发了一个认知驾驶员协助系统，该系统可以捕捉交通状况，分析交通状况，并在行人构成潜在危险时提醒驾驶员。数据分析表明，行人的决策取决于车辆的距离，而不是碰撞的时间[21]。Brechtel等提出了一种车辆对行人避碰的预测方法，该方法通过观察行人的运动实例，学习并预测行人的行为[22]。Volz等人使用一个真实的行人轨迹数据库来学习一个能够预测行人是否会穿过街道的模型。首先引入大量可能适合描述行为的特性，然后执行相关性分析以确定那些必要的特性，以达到最佳的泛化性能[23]。Knorr等人进行了两个实验来研究个人或情境特征在多大程度上影响了人类运动中成功避免碰撞的角色归因和贡献，得出行人让路对成功避免碰撞具有更大的相关性[24]。

目前研究成果中，国外汽车制造商正在开发可以与行人进行“眼神交流”的自动驾驶汽车。捷豹路虎为自己的自动驾驶汽车的车厢安装了虚拟的“眼睛”，通过自动驾驶汽车的雷达传感器和摄像头与行人进行互动。当有行人经过时，它们会对行人进行礼让，同时也会观察从侧道出来的司机或者是接近的司机。SEMCON公司构思出一种新型的笑容指示灯，该指示灯使用追踪摄影机以及雷射科技，对行人头部的转向进行捕捉，由此观测行人的眼神是不是看向汽车，如果确定行人的眼神是看向自动驾驶汽车，那么车头安装的笑容指示灯，就会发出犹如一道微笑的灯光，以此来提示行人优先通过马路。

(二)国内研究现状

Rong-Ben等人叙述了世界范围的智能车辆行人检测技术，集中说明了主要发达国家对智能车辆行人检测技术的研究，同时阐述目前智能车辆行人检测技术的不足之处，对未来的技术趋势做出展望[25]。Yao等人提出了一种基于卡尔曼预测的行人检测和实时预测行人位置的方法。该方法可以对行人与车辆碰撞点进行预先判断[26]。刘光新等人分析了信号过渡时候不同信号灯显示下行人过街行为的影响因素,同时构建模型对行人过街行为进行预测。采用Pearson x2检验法分析影响行人选择行为的因素,并建立了二元Logit模型预测行人在信号过渡时间内的过街行为[27]。朱芳芳等人构造了行人在过街时间内选择行为的概率与安全间隙之间的关系模型,并对模型进行了测验,该模型的拟和度及其精度较好[28]。廖明军等人采集吉林大街的两个路段的过街行人数据，通过人为观测视频来获取行人过街行为数据，利用Binary Logit模型建立模型并且预测的准确率达89%[29]。楼佳妮对国内主要大型活动的行人交通行为进行了理论研究与实证分析，并构建了适用于行人的交通分配模型，重点对其中的路阻函数参数进行了行人化标定[30]。黄智等人通过一阶马尔科夫模型描述行人过街运动,基于无味变换,有效的预估人和车发生碰撞的概率以及碰撞速度和时间的概率密度,预测得到人车冲突中行人受到伤害的概率[31]。

Chen等人在前人研究的基础上，提出了一种改进的多层次行人行为预测方法。改进主要集中在轨迹与聚类MP的相似匹配准则上，结果表明，该方法适用于实际动态变化环境，预测结果与实际情况相一致[32]。车领在全面考虑影响行人过街的各种条件下,构建以贝叶斯网络为基础的行人过街意图预测模型。确定每个节点变量彼此的关系,使用统计软件和Tabular CPD函数分析数据之间的相关性和对先验概率进行初始化,通过联合树推理算法获知每个节点变量对行人过街意图的影响,使用模型基于所给数据进行预测获得行人过街的概率,预测结果符合实际要求[33]。Zhang等人根据行人检测系统的框架，依次总结了利用机器视觉以及基于多传感器融合下的行人的检测方法、行人跟踪原理和行人行为分析，对利弊进行了分析和总结[34]。Likun等人提出了一种交叉口行人轨迹预测算法。为了获得行人运动的特征，开发了一种数据标记和行人身体方向回归的方法，预测精度也超过了交互多模型KF和动态贝叶斯网络(DBN)等方法[35]。董海荣等人发明公开了一种基于马尔可夫状态跳变的行人行走行为预测方法，为正确疏导人群提供了参考，对提高人群疏散效率具有重大意义[36]。黄文恺等人提出基于HMM算法的行人过马路的预测方法和系统，所述方法包括：接收采集装置采集的当前数据；根据所述当前数据和预先存储的历史数据基于HMM算法建立深度学习模型。此外，在建立深度学习模型的过程中，应用了道路数据，通过加入道路信息对行人是否过马路的判断，使得预测结果更加的准确[37]。闫旭等人研究了利用有限的文献资料如何分析行人路径选择的理论依据，对模型的适用情境以及专业理论的择取对分析过程十分关键[38]。

2005年第11届中国交通运输专家国际会议针对我国存在的严重的行人安全问题和信号交叉口行人交通混乱的现状，重点研究了道路交通违章行为的类型、违法性以及影响行人违章行为的因素。研究可以帮助了解行人在信号交叉口人行横道上的行为[39]。杨燕等人为理解行人在交叉路口过街的违章行为,调查了长周期信号控制的交叉路口行人违章过街的心理因素,基于计划行为理论制成违章过街行人的心理问卷。问卷结果可知知觉行为控制和行为意向可以推知行人的现实违章行为,准确率为73.9％[40]。

1.2.2 行人通行意图预测算法的不足

从1.2.1中我们可知，目前对于行人的通行预测主要包括基于行人动作的预测方法和基于行人路径的预测两种方法，对于行人其他的信息利用较少尤其缺乏对面部信息的利用。此外，行人通行数据也较为奇缺，特别是缺乏同时标示动作信息和面部表情的数据库。

本文通过引入行人面部关键点信息和骨骼关键点信息，利用深度学习算法预测行人是否通过马路。骨骼关键点信息包含了一个人动作的表达规律。微表情涵盖了一个人将要做出某种动作的信息，面部关键点作为行人微表情的直接表达，可从中有效学习到行人过马路时候面部的变化规律，因此本文提取行人面部关键点和骨骼关键点，利用行人动作信息和面部信息，实现信息多元化。建立图卷积神经网络GCN和长短时记忆模型LSTM结合的预测网络，实现空间序列和时间序列上的特征提取，加上对关键点的信息利用，对行人通行意图进行有效预测。

1.3 深度学习研究现状

1.3.1 深度学习的概念理解

深度学习是包含于机器学习的一种复杂的算法，是机器学习中一个新的研究方向，为了能够更加接近于人工智能的目标而被提出。2006年，机器学习领域的领头人物、多伦多大学学者Geoffrey Hinton及其学生联合发表了一篇论文，深度学习的概念由此被提出[41]。深度学习方法将底层简单特征组合形成更为抽象的高层特征，从中获得数据的内在联系，还可以对声音、图像等进行建模。深度学习可以通过学习非线性的深层网络结构，从而对复杂的函数进行拟合，可以从很少的样本数据中学习到潜在的数据特征。通过构建模拟人的大脑的处理结构，模仿大脑处理声音、图片等数据的方式。深度学习通过学习大量的数据，从中学习到数据的本质规律，获取代表性的特征，从而实现更有效的分类预测。

基本的深度学习思想阐述如下：设想存在一个系统S，这个系统可以分为n层（即S1，…，Sn），系统的输入表示为I，输出表示为O，系统可表示为：I=>S1=>S2=>…=>Sn=>O。为了使系统输出O能够尽量接近系统输入I，可以通过不断地改变系统中的参数，从而获得系统输入I的一连串的层次特征S1，S2，…, Sn。深度学习基本思想就是通过这种方式来自动学习层次特征从而实现对系统输入信息的分级表达，使得输入输出在一定的规则限制内尽量缩小差距[42]。

1.3.2 深度学习的发展历史

深度学习的发展历程基本分成三个阶段(如图1.2)，从20世纪40年代到60年代的阶段中，深度学习以控制论的名字发展，从20世纪80年代到20世纪90年代时段，深度学习被称作为连接机制，这时候的深度学习可以使用反向传播进行训练，相对于第一阶段有了很大的进步。第三阶段是深度学习概念真正诞生的时间段。2006年，以机器学习先导人物Hinton提出深度信念网络（DBN）为开端开始了深度学习发展的第三阶段。

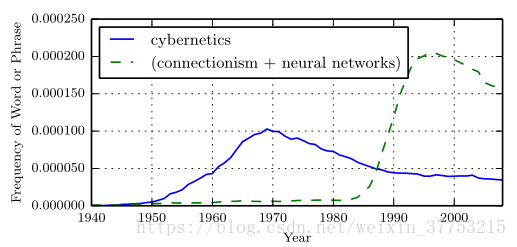


图1.2 深度学习的发展

Fig. 1.2 Development of in-depth learning

人工神经网络这个概念最早由美国数学家 Walter Pitts 和心理学家 Warren McCulloch在1943年合作提出，打开了人们对人工神经网络研究的的大门。同年，由神经科学家麦卡洛克和数学家皮兹建立了MCP模型，即模拟生物大脑的神经元结构及其作用原理构造出的模型[43]。1957年Frank Rosenblatt提出了感知机perceptron算法，神经网络和深度学习起源于此算法。感知器被人们称为一种最简单的前馈式人工神经网络，是一个简单的二分类线性分类器。

1969年，美国数学家及人工智能领域的联合创始人Marvin Minsky证明感知器实际上只是一种线性模型，因此只能解决线性分类的问题，仅仅是对于简单的亦或问题，感知器都不能做到正确处理。从此之后，神经网络就陷入了瓶颈期，长达20年处于停滞不前的状态。

1974年，美国哈佛大学博士毕业生Paul Werbos在他的博士论文里，首次提出使用误差的反向传播方法进行一般的人工神经网络训练[44]。1980 年，来自多伦多大学的Jeffrey Sinton 教授基于传统的传感器，由多隐藏层深度结构来替代感知器的单层结构。其中，多层感知器模型（Multi\_Layer Perceptron）最具代表性。1982年约翰·霍普菲尔德构建的名为Hopfield的神经网络是最典型的相互连结型网络，该网络也是起源最早的递归神经网络[45]。Hopfield模型分为离散型和连续型。

1984年,日本学者福岛邦彦提出了一个有卷积层、池化层的神经网络结构，这个神经网络结构是卷积神经网络的原始模型，被称为Neocognitron，该模型能够进行有变形或位移的模式识别[46]。

1986年，Rumelhart和McClelland等人提出了BP神经网络的概念，可以较为系统地解决多层神经网络隐藏层存在的权重学习问题[47]。Hinton提出适用于多层感知器（MLP）的BP算法，采用了sigmoid函数进行一个非线性映射，从而有效解决了非线性分类的问题。至此以后，深度学习的发展迎来了第二次发展浪潮，深度学习发展进入了新阶段。

1989年，Cybenko和Hornik两人提出了万能逼近定理，说明了任何函数都可以被三层的神经网络以任意的精度逼近，从理论角度上说深度学习可以做到近似任何的函数，是深度学习的一个极其重要的理论基础[48]。同年，Yann LeCun等人提出了具有五层神经网络层的卷积神经网络LeNet,这个网络作为最早的神经网络之一，推动了深度学习的发展。这个模型十分有效的解决了手写字体的识别问题，通过十分合理巧妙的网络结构设计，利用参数共享、池化以及卷积等操作，有效的规避了繁琐额计算过程，节约了计算成本，网络的最后是一个全连接层，利用全连接层进行分类。1991年，人们发现了反向传播的一个很大的漏洞即梯度消失问题，随着隐藏层层数的不断增加，准确率却随之不断下降。

1997年，来自瑞士的深度学习大神Juergen Schmidhuber和人工智能领域著名学者Sepp Hochreiter在论文《LONG SHORT-TERM MEMORY》中联合提出了一个带有记忆功能的循环网络的变体,就是著名的长短时记忆网络LSTM,可以用来解决之前出现的梯度消失的问题[49]。长短时记忆网络是为了针对性解决一般的循环神经网络所存在的长期依赖的问题而提出来的，它作为是递归神经网络的延伸，可以在时间延迟十分长的情况下获得有效的学习结果。该网络内部含有LSTM区块，区块中有三个门，分别是遗忘门、输入门和输出门，其中由谷歌的实验测试所得，遗忘门是最重要的一个门，它能够决定输入门输入的哪些部分能够被记住，哪些部分能够被输出门输出。

2006年以后深度学习进入了新一轮的快速发展时期。深度学习大师Hinton团队在2006年发表论文《A fast Learning Algorithm for Deep Belief Nets》将深度学习拉回了大众的视线，在该论文中，提出了一个新的深度学习模型就是深度置信网络，并给出了一种高效的半监督算法[50]。

2010年，Xavier Glorot和Yoshua Bengio在论文《Understanding the difficulty of training deep feedforward neural networks》中提出了Xavier初始化方法，基于该初始化方法，能够使得我们在进行神经网络的初始化的时候让权重保持合适的大小，从而能够避免在反向传播的过程中因为梯度递减而造成的梯度消失问题，参数在经过初始化后，依然能够保持输出接近于高斯分布[51]。但是，Xavier初始化被发现只是适用于线性激活函数，这达不到深度学习的要求。

2011年，深度学习先驱人物Yoshua Bengio等人在《Deep Sparse Rectifier Neural Networks》一文中首次提出了ReLU激活函数并证明了该激活函数可以有效的解决梯度消失问题，为深度学习的进一步发展扫清了一部分阻碍[52]。当时引入ReLU激活函数是因为Sigmoid激活函数基本存在以下问题：第一，Sigmoid在使用时的计算成本大，但是ReLU函数的计算量相对来说小很多，求导可以很容易的实现。第二，Sigmoid函数在训练的时候很容易造成梯度消失的问题，这个问题也是阻碍深度学习发展的绊脚石，但是ReLU函数可以有效解决，有利于深度神经网络的反馈传播。第三，ReLU函数相对于Sigmoid函数来说，更具稀疏性，它可以在学习传播过程中使得一些神经元输出变为零，从而有效避免过拟合问题的发生。

2012年，Dropout方法被提出，该方法通过随机丢弃部分神经元从而有效的避免神经网络出现的过拟合问题，提高神经网络的泛化能力。同年，深度学习领域三巨头之一的Hintona的学生Alex通过AlexNet以领先于第二名10.9%的准确率夺取ImageNet 2012 图像识别大赛冠军宝座[53]。AlexNet 有八层卷积神经⽹络，通过使用ReLU激活以及采用过多个GPU进行训练，使用最大池化以及采用概率为0.5的Dropout函数等措施使得训练结果达到了一个很低的错误率。

2014年，R-CNN的提出使得目标检测技术进一步发展。2015年，又提出了Fast R-CNN,有效提高了目标检测的效率[54]。但是，由于在Fast R-CNN中region proposer部分限制了算法的整体效率的提高，所以在2016年，微软研究团队又提出了Faster R-CNN的方法来进一步提高检测效率。

2017年AlphaGo打败了顶级围棋棋手李世石，使得强化学习风靡一时。不久之后，又出现了名为AlphaGo Zero的新版本，该版本靠卷积神经网络实现了自学成才。2017年年底，AlphaGo Zero进一步更新为Alpha Zero，技能变得更为强大。

1.3.3 深度学习在自动驾驶上的应用

深度学习技术的飞速发展，对很多方面起到了推动作用，在许多领域都取得重大进展。在大数据时代，面对各种复杂繁多的信息，传统的各种人工方法都不足以满足人们的需求。深度学习模拟人脑的工作机制，构建神经网络模型，从数量巨大的数据中提取有效特征，甚至处理的效果超过了人类。深度学习的应用领域繁多，本文主要就深度学习在自动驾驶方面的国内应用进行陈述。深度学习技术的广泛使用使得传统汽车变得智能化，互联网技术、高精度地图与深度学习技术的组合应用使得自动驾驶技术在各方面都有了很大的进步。

(一) 深度学习技术在物流配送方面的应用

苏宁物流完成了国内电商企业在大型无人车领域的第一次探索。经由苏宁物流自主研发的无人配送车“卧龙一号”(如图1.3)，能够按照使用者的提前设置好的路线进行自动行驶，从而实现送货上门。该无人配送车融合了多种技术，包括云计算、AU以及物联网等，能够自主躲避障碍物以及在路途中遇到的行人，对于红绿灯也能自动识别。除此之外，苏宁物流还推了无人重卡“行龙一号”(如图1.4)，可有效避免司机疲劳驾驶，提高运输效率。



图1.3 无人配送车“卧龙一号”

Fig. 1.3 Unmanned delivery vehicle “Wolong No.1”



图1.4 无人重卡“行龙一号”

Fig. 1.4 Unmanned heavy truck “Xinglong No.1”

美团推出了无人配送开放平台，其中配送车的无人驾驶界别可达到L4级。美团的无人配送车使用了激光雷达、超声波以及摄像头等多种传感器，对这多种传感器进行技术融合人，从而使得研制的无人配送车能够实现在城市路况下进行低速度的行驶的能力。在未来的很长一段时间内，美团配送都会采用人车混合配送的配送方式。相对于美团骑手来说，无人配送车能够长时间的连轴工作，弥补骑手的配送缺口，承担大部分的夜间配送工作，提高骑手的工作安全性。通过发挥骑手和配送车两方面的优势，扩大配送效率，最大程度完成用户的要求。

京东在JD CUBE大会上宣布其自主研发的自动驾驶重卡预计将在2020年完成商业化运营部署工作。西井科技公司发布全时多场景的无人驾驶电动重型卡车Q-Truck，采用无驾驶室的模块化设计。智行者研制出的主要用于小区环境下的无人物流配送车“蜗必达”进入量产阶段。

(二) 深度学习技术在共享出行方面的应用

无人驾驶技术在汽车共享出行方面的应用有效的解决了它之前存在的一些痛处，将主动身份变换为车，不再是以前的人找车、人找位，而变化为车找人、车找位，通过无人驾驶技术的支持，还可以实现用户的一键泊车等功能，为以后的智能化城市建设做好准备。国内众多的共享出行的无人驾驶企业队伍开始了共享出行汽车的研发以及测试，其中包括百度、宇通、金龙、滴滴、中智行科技、驭势科技以及美团等。

国内的驭势科技旨在运用AI技术发展未来的出行方式。在北京举行的共享汽车分时租赁品牌GoFun(如图1.5)出行的4.0新发行版本发布会上，驭势科技在会上指出以后的汽车只需要使用者在手机上下载相应的软件进行下单，共享汽车就能根据订单要求自主完成车位的寻找、路径规划以及自动泊车等，通过智能技术，节约使用者的出行时间。伴随着车联网技术的发展，能够实现停车地点的更为有效的调度，对汽车的迁移提出更有效的管理方案，整体的运营成本实现大幅度降低。该公司通过自动代客泊车可以实现共享汽车的精确停车，公司提出的方案已经投入使用，这样可以有效节约社会资源，解决传统车辆行驶的拥挤问题。



图1.5 驭势科技共享汽车GoFun

Fig. 1.5 Shared car “GoFun” of UISEE

(三) 深度学习技术在公共交通方面的应用

国内已经有很多企业开始对无人驾驶如何运用在公共交通方面开始了研究。通过使用深度学习技术可以实现无人驾驶汽车的多种功能。

宇通开发的无人驾驶客车(如图1.6)已经完成了开放道路的测试步骤，这款无人驾驶客车属于L4级别，现在已经具备了自主巡航、躲避障碍物、准确停靠以及紧急制动等功能。在有市民准备乘车时，该客车就会在乘车点主动靠停，在市民上车后，随即关门行驶。车辆在行驶过程中能够做到准确识别行人，如果遇到有横穿马路的行为，能做到及时减速躲避，避免造成行人的人身伤害。及时调节车内的温度，将温度保持在一个合适的范围，提高市民的舒适度。可以根据实际的运行情况对路线进行调节优化，也可根据乘客数量的增加自动追加客车。



图1.6 宇通无人驾驶客车

Fig. 1.6 Driverless passenger car of YUTONG

百度与金龙客车联合设计了L4级自动驾驶汽车“金龙阿波龙”，深兰科技研发 “熊猫智能公交车”已经成功试行。中国重汽集团的首辆L4级别的全智能无人驾驶客车在山东正式上路运营测试。

1.4 本文主要研究内容

本文的研究目的是通过采集行人数据提取行人的骨骼关键点和面部关键点，结合图卷积神经网络和长短时记忆模型构建预测网络来预测行人是否过马路。在人的行走过程中，人的姿态估计可以从骨骼关键点得到，骨骼关键点的变化反应了行人的行为规律。人的面部蕴含着许多的信息，脸上的微表情可以体现出一个人做事的动机，是人是否做一件的事情的细微体现。面部关键点的变化可以从侧面反映出人的情绪变化从而预测人的行为。因此，我们采集行人的骨骼关键点与面部关键点信息，提取关键点的时间和空间序列特征，将他们输入到图卷积神经网络和长短时记忆模型结合的预测网络中进行训练，从而使得网络能够学会利用行人关键点信息有效的预测行人过马路的意图。全文共分为五个章节：

第一章，主要介绍了本文课题研究背景以及研究意义。通过研究目前行人行为预测的现状，在本文中提出目前研究的不足之处，说明了本文所研究方法与之相比较所存在的优势。提出了基于图卷积神经网络和长短时记忆模型的预测网络，结合行人关键点信息进行行人过马路意图预测。作为铺垫的理论基础，介绍了涉及的深度学习的发展历程以及有关理论知识，详细举例说明了深度学习在自动驾驶领域的应用。

第二章，主要介绍了卷积神经网络和长短时记忆模型的知识。在本文中使用的关键点检测库是通过卷积神经网络来实现的，卷积神经网络在图像处理领域有很大的优势，可以有效提取图像信息。同时介绍了长短时记忆网络LSTM的三个“门”的操作过程，为预测网络搭建做理论铺垫。

第三章，主要介绍了行人关键点提取的原理和CSV数据集制作过程。在学校及其周围区域拍摄行人视频，视频经过视频帧分离以及分辨率重建等处理后作为关键点提取的数据来源输入。本文中骨骼关键点和面部关键点的提取是通过Openpose实现的，因此我们在第三章主要介绍了Openpose的来源背景和使用优势等，然后陈述了使用Openpose提取关键点的原理以及提取后的效果说明，对关键点坐标的表示进行了说明。将Openpose提取的关键点信息作进一步的处理得到本文需要的CSV格式的数据集，为后文中使用该数据集进行训练做好铺垫。

第四章，本章主要介绍行人通行意图预测网络的设计。本文中使用的预测网络是结合图卷积神经网络和长短时记忆网络长短时记忆模型进行设计的，对关键点在时间上和空间上的特征信息进行提取处理，得到最终的预测结果。因此第四章主要介绍了图卷积神经网络原理以及预测网络的结构设计，并对长短时记忆模型预测网络和添加图卷积处理的预测网络的预测结果进行了对比分析，对本文提出的预测网络的预测效果进行了验证。

第五章，在本章中主要对全文的总结与展望。说明本文中提出的行人通行意图预测方法相对于之前提出的方法的优势所在，指出了图卷积神经网络和LSTM结合的行人过马路意图预测方法所存在的不足之处以及在未来后期中希望做出的改进。

1.5 本文技术路线

本文提出了一种结合图卷积神经网络和长短时记忆模型的行人通行意图预测方法，该方法能够根据行人的骨骼关键点信息和面部关键点信息来预测行人是否过马路。通过充分利用行人信息，节约汽车在行驶过程中驾驶员的等待时间。同时，更加有效的保护了行人的人身安全。本文的整体思路如图1.7所示。

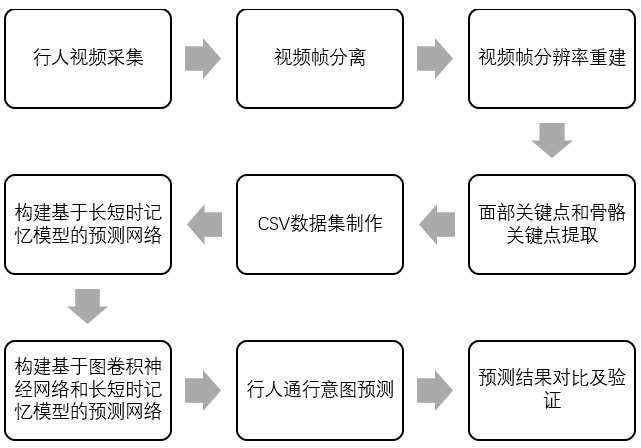


图1.7 全文技术路线  
Fig. 1.7 Full-text Technical Route

1.6 本章小结

本章主要介绍了本论文课题研究背景以及研究意义。提出了目前研究的不足之处，概述了国内外发展的现状，说明了本文所研究方法所存在的优势。阐述了深度学习的相关知识以及其在自动驾驶领域的应用，最后以流程图的形式阐述了本文的技术路线。

2 卷积神经网络和长短时记忆模型概述

2.1 卷积神经网络概述

1943年，Warren McCulloch和Walter Pitts在论文中首次提出了人工神经网络的概念，提出了经典的M-P模型(如图2.1)，是模仿生物神经元处理信息的简化模型。

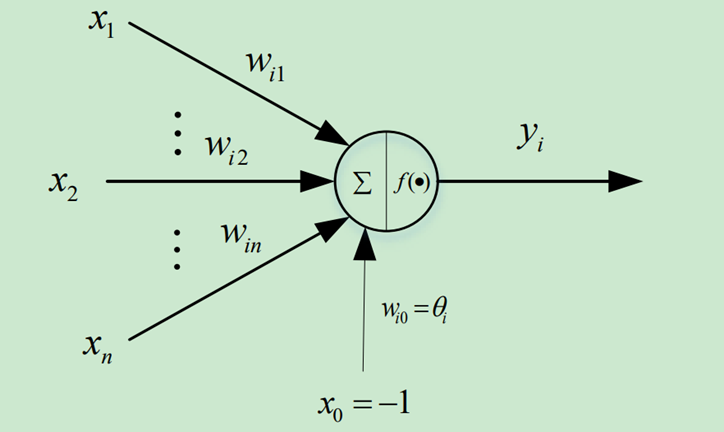


图2.1 M-P模型

Fig. 2.1 M-P model

其中， (i=1,2,3,…,n)代表当前神经元的输入，这些输入是由与当前神经元连接的周围神经元对其传入的信号。 (j=1,2,3,…,n)代表神经元i和神经元j之间连接的权值，是激活神经元需要达到的阈值，函数f称为神经元的激活函数,表示最终神经元i的输出。可以表示为以下形式:

神经元i接收与其连接的神经元传递来的信息，由神经元之间连接的权重来决定输入信息的表达，然后神经元i将这些信息按照权重进行叠加。将信息与阈值比较，并通过激活函数进行输出。

1958年，纽约心理学家Frank Rosenblatt在其著名文章《Electronic ‘Brain’ Teaches Itself.》中提出了感知器模型(如图2.2)，是最基础的用于线性可分模式下的神经网络。

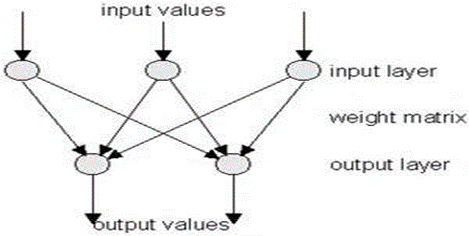


图2.2 感知器模型

Fig. 2.2 Perceptron model

感知器都是由M-P神经元组成，它是一种典型的二元分类器，把矩阵的数值输入转换成一个二元的输出值。感知器的权重决定每个输入的重要程度，偏置是激励函数的偏移量。感知器通过加权计算函数进行决策，通过不断的改变模型内部的权重大小和偏移量进行学习，是形式上最简单的一种前馈式人工神经网络。感知器的结构比较简单，对于能够解决的问题，存在着收敛算法，并且能够从数学上进行严格的证明，因此感知器对神经网络的研究有着十分重要的推动作用。

1982年，美国物理学家J.J.Hopfield在《Neural networks and physical systems with emergent collective computational abilities》中提出了单层反馈神经网络—Hopfield网络(如图2.3)。作为一种递归神经网络，Hopfield神经网络能够保证向局部极小收敛。该神经网络中存在的每一个神经元既可以作为输入神经元也可以作为输出神经元，所有的神经元都会将自己的信息输送给神经网络中其他的所由神经元，在输出的同时也会接收其他所有神经元通过连接权传递过来的信息。因为神经元之间信息的互相传递，因此权值矩阵是一个对称的矩阵。Hopfield神经网络可以分为两种类型：连续型和离散型。这两种类型的模型也分别有适用的场合，离散型的Hopfield神经网络适用于联想记忆，而连续型的适用于优化问题。

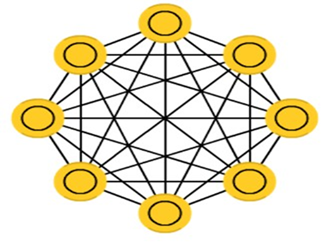


图2.3 Hopfield神经网络

Fig. 2.3 Hopfield neural network

支持向量机SVM(如图2.4)由Vapnik和Alexey Y. Chervonenkis于1964年被提出，后期在20世纪九十年代得到了进一步的改进，广泛被人们应用在人脸识别、文本分类等方面。SVM是一种基于定义在特征空间上的最大间隔提出的二元的线性分类器。目的就是最大化特征空间上的间隔，求解的过程可以具体化为求解一个二次规划的问题。SVM的基本思想是能够将训练数据准确划分并且能够分离出在空间唯一的间隔最大的超平面。在支持向量机中，采用的是合页损失函数，同时添加正则化项来优化，因为结构稳健，具有稀疏性。支持向量机还通过核方法完成进一步改善，使得结构能够进行非线性分类,得到了非线性SVM。

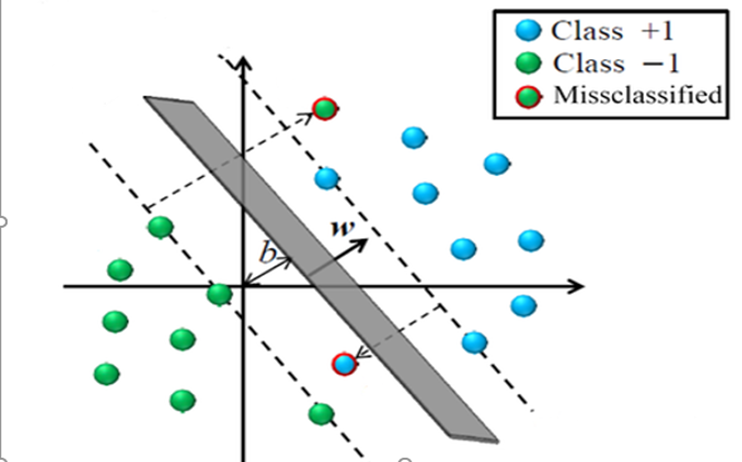


图2.4 支持向量机

Fig. 2.4 Support vector machine

常见的激活函数分为Sigmoid函数、Tanh函数以及ReLU函数。激活函数能够让神经网络表达非线性映射，对神经网络学习非线性函数具有非常重要的作用。如果没有激活函数，那么神经网络的层与层之间只是简单的矩阵相乘，无论中间添加了多少层，模型也只能表达线性映射。激活函数使得神经网络能够拟合非线性数据，当遇到线性不可分的数据，激活函数就发挥了不可或缺的作用，使得神经网络模型的表达能力大大提高。

Sigmoid函数(如图2.5)可以把输入的值压缩至0和1之间，负数很大时，输出接近于0，当数值大于0且很大时，输出接近于1[55]。输入的绝对值越大时，激活函数的导数梯度越小。Sigmoid函数属于饱和激活函数，在使用时很容易达到饱和，当输入的绝对值接近于无效大，梯度接近于0。神经网络在反向传播时，如果这时候的导数基本是0，那么权重的更新就会很困难，从而造成神经网络无法继续学习，导致所谓的梯度弥散的问题。Sigmoid函数的输出不是零均值，导致神经网络后面的网络层的输入就不是零均值的，这样神经网络就会收敛的十分缓慢。除此之外，使用Sigmoid函数的计算量大，对计算机造成负担。因此Sigmoid函数在使用一段时间后，逐渐降低了使用热度。

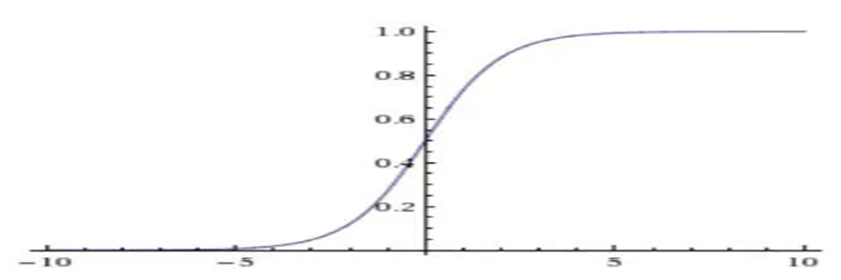
图2.5 Sigmoid函数

Fig. 2.5 Sigmoid function

Tanh函数(如图2.6)在Sigmoid函数上面做了改进，解决了Sigmoid函数的非零均值问题，将最终的值压缩至-1到1之间。但是它并没有解决饱和神经元的情况下的梯度消失等问题[56]。计算量大，仍然是指数运算，拖慢了计算机的运算速度。

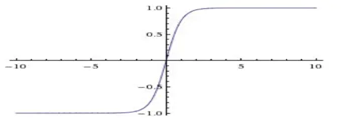


图2.6 Tanh函数

Fig. 2.6 Tanh function

为了进一步对激活函数进行改进，深度学习先导Hinton等人在2012年发表的论文《ImageNet Classification with Deep Conolutional Neural Networks》中提出了一种非饱和激活函数，这就是ReLU激活函数(如图2.7)。ReLU激活函数是目前应用很广泛的激活函数，输入小于0的时候，输出为0，当输入大于0的时候，输出等于输入。

ReLU函数解决了之前激活函数没有解决的梯度消失等问题。函数本身是线性的，因此计算量相对于之前的指数运算大大减少[57]。但是，如果输入神经网络的值是小于0的，那么输出就会是0，这就会导致后面的网络层的输入全部变为0，权值无法进行更新，从而造成神经网络大片神经元不会再学习。人们在后来提出了eLU函数，相对于ReLU激活函数，它的输入小于0时，是可以具有一定的输出值的，因此避免了神经元大面积失去学习能力的问题。

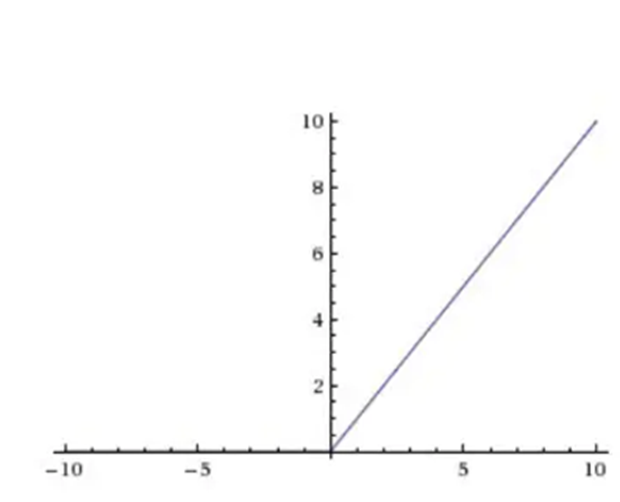


图2.7 ReLU函数

Fig. 2.7 ReLU function

神经网络(如图2.8)由许多的神经元组成，多层神经网络主要由输入层、输出层以及夹在中间的隐藏层三部分组成。

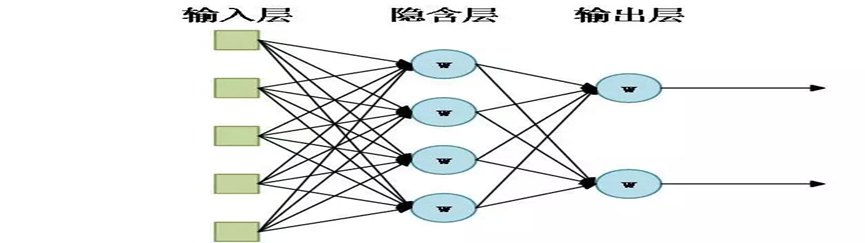


图2.8 神经网络

Fig. 2.8 Neural Network

输入层就是代表输入到当前神经元的信息输入层，通常在信息传入输入层之前会对信息进行一定的处理。隐藏层占据了神经网络的多部份神经元，是将数据按需输出的关键部分，信息的选择性输出由权重和偏差等进行调节。输出层将隐藏层处理后的信息输出。

神经网络基本可以分为四种类型：前馈神经网络、反馈神经网络、随机型神经网络以及自组织竞争型神经网络。前馈神经网络（Feedforward Neural Network，FNN）是最早发明的简单人工神经网络。

前馈神经网络(如图2.9)是神经网络中最简单的一类，它的每一层神经网络层只与前一层相连。每一层收到前一层的输出信息后将其输入给下一层，层与层之间没有反馈，信号是一个简单的单向传播的过程。在同一神经网络层中的神经元之间不存在相互连接，信息只在层与层之间传递并只向着单一方向传递。

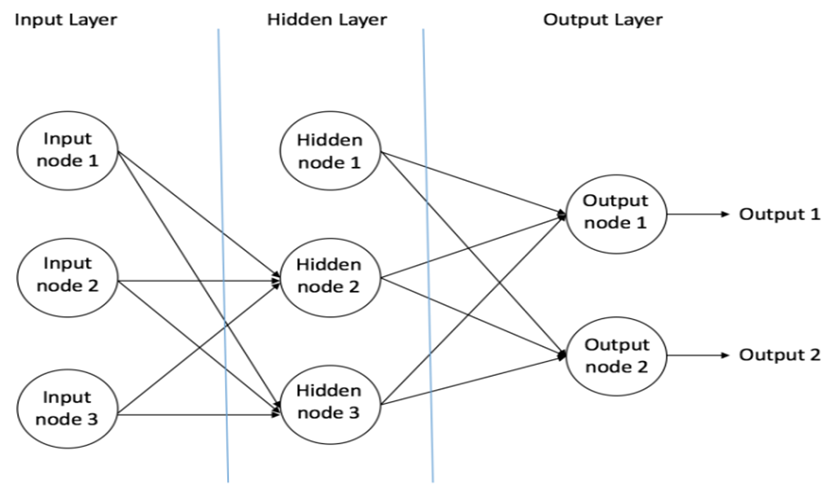


图2.9 前馈神经网络

Fig. 2.9 Feedforward neural network

前馈神经网络主要是由输入层、隐藏层和输出层组成，神经网络只能由一个输入层和一个输出层，但是隐藏层却可以有0个或者若干个。在这三种神经网络层中由分别包含了三种类型的节点。分别是输入层节点、隐藏层节点和输出层节点。

输入层节点将外部信息传入神经网络中，输入层的没一个输入层节点不负责计算，他们的任务只是负责信息的传递，将外界信息传给隐藏层节点。隐藏层节点是夹在输入层和输出层之间的，没有与外界信息的直接接触。他们负责接收输入层节点传递进来的信息，将这些信息进行相应的计算后传递给输出层节点。输出层节点负责将隐藏层节点传递出的信息传输到外部。

深度学习学习任务实际上可以称为一个函数的拟合过程，通过不断地学习，连续的接近真实值。在学习的过程中，损失函数就会起到一个指导作用。神经网络常见的任务一般就是分类和回归两种。分类输出的是离散值，一般判断图片中出现的物体的属性类别。回归输出的是连续值，一般是预测的回归值与真实值之间的偏移差距。损失函数通过比较模型的输出与真实值之间的差距来判断模型拟合性能的好坏。当预测的值与实际需要输出的值之间的差距很大的时候，损失函数的值将会很高。因此在训练时，可以通过损失函数有效的惩罚模型以进行优化。在损失值较大时，通过网络的传播过程，推动权值的更新。在不断接近真实值的过程中，权值的变化也会越来越小。损失函数的本质作用就是用来衡量模型的预测能力的好坏。损失函数分为分类损失函数和回归损失函数。

常见的分类损失函数有交叉熵损失函数、坡道损失函数以及合页损失函数等。

交叉熵原本是信息论中的一个定义，用来估计平均编码的长度。后来将其用于神经网络作为损失函数核心概念来使用。交叉熵损失函数公式如下：



其中，为真实值，为预测值，为损失值。

当真实值y=0时，，这时候L对于单调递增。越小，与真实值y越接近，对应的损失值L越小。同理，当y=1,L对于单调递减。所以越大，与y越接近，计算得到的L值越小。L越大，对模型的惩罚力度就会越大，从而使得预测值与真实值越来越接近。交叉熵损失函数可以分为Sigmoid交叉熵损失函数和Softmax交叉熵损失函数。

Sigmoid交叉熵损失函数将实际值映射到0到1之间，其输出的结果相互独立，Sigmoid函数将一个值转换为一个概率，设置一个阈值，当概率值大于阈值时，则认为该判断属于某一个类别。如果输出的概率值小于某一个阈值，则认为该分类不属于某一类别。因此，Sigmoid交叉熵损失函数一般被用于多标签分类，这样只要输出值满足阈值要求，那么该判断就可以被判别为属于多个类别。

Softmax交叉熵损失函数将n维的实际向量（a1,a2,…,an）映射为(b1,b2,…,bn),映射值均为0到1的常数。Softmax函数作用下的输出值之间是相互关联的，因此映射之后的向量值加起来的和为1。转换之后的值可以看作概率，取最大值对应类别作为判断后最终的分类结果。因此，Softmax交叉熵损失函数多用于多类别分类。

常见的回归损失函数有L1损失函数、L2损失函数以及Smooth-L1损失函数。

L1损失函数即最小绝对值误差（LAE），它表示估计值与真实值之间的差距绝对值之和。L1损失函数十分稳定，对于误差的导数为常数，训练时比较稳定。但是也是因为损失函数导数的绝对值较小，对于学习率比较小的情况下，损失函数的值就会一直在某一处波动，难以在短时间内收敛。当真实值与预测值之间的差距很小时，这时候的损失函数的导数相对于差距较大。除此之外，L1损失函数在0处到处不唯一，这也是阻碍函数收敛的一个因素。L1损失函数公式如下：



其中，S是差值绝对值总和，目标值是 Yi，f(xi)是估计值。

对于大部分的卷积神经网络，比较常用的损失函数是L1损失函数而不是L2损失函数，因为L2损失函数的导数比L1损失函数的大，收敛速度快得多。但是相对来说，L2损失函数增加的是误差的平方级，当我们选用L2损失函数时，鉴于误差较大的情况下，L2损失函数对于估计值的梯度比较大，这时的训练变得十分不稳定，需要对模型做很大的调整。L2损失函数公式如下：



其中S是差值平方总和，i是样本的顺序，目标值是Yi，f(xi)是估计值。

针对两者的缺陷之处，提出了Smooth-L1损失函数。它能够保证真实值与估计值的差距较小时，梯度足够小。在真实值与估计值的差距较大时，防止因为太大的误差主导损失函数，甚至造成梯度爆炸。Smooth-L1损失函数在零处也是函数平滑可导的，防止在零处函数导数不唯一产生不收敛问题。综上，Smooth-L1损失函数避免了L1损失函数和L2损失函数各自存在的一些问题，结合了L2损失函数收敛更快且在零点处有导数，便于收敛的优点。同时，在边界区域结合了L1损失函数的优点，让网络面对异常值时更加稳定，能够在偏移值较大时仍然保持梯度稳定。三者损失函数的对比如图2.10所示。

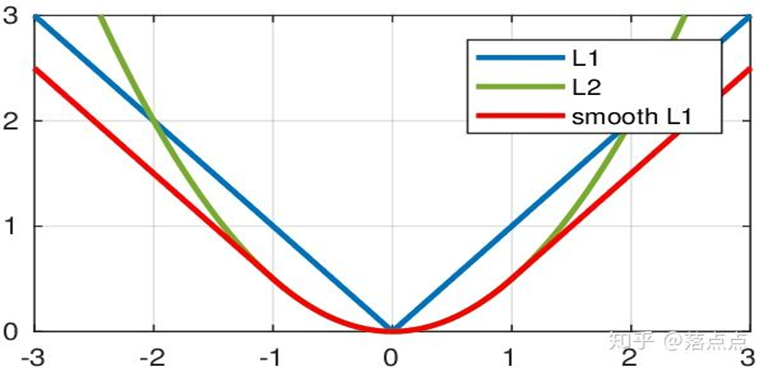


图2.10 三种损失函数对比图

Fig. 2.10 Contrast figure of three loss functions

卷积神经网络是一种包含卷积计算的深度前馈神经网络，对于图像处理有很出色的表现。其中的神经元只连接相邻部分神经元，它包含卷积层、池化层以及全连接层等。

卷积层由若干个卷积单元组成，卷积层的目的是提取特征。从第一层卷积层开始逐渐提取更高级的特征，最终对这些特征进行组合[58]。卷积单元中的参数不断的学习优化，通过反向传播过程进行更新。卷积通过不同的二维数字矩阵即卷积核对输入图像进行数学运算。卷积过程的示例如图2.11所示。



图2.11 卷积操作示意图。左边矩阵为输入，中间的二维矩阵为卷积核，右边为输出。卷积核和图像像素值进行点乘操作，从而得到特征值

Fig. 2.11 Schematic diagram of convolution operation.The matrix on the left is the input, the matrix in the middle is the convolution kernel, and the matrix on the right is the output.The convolution kernel is dotted with the pixel value of the image to obtain the eigenvalue

由于图像中临近区域的像素值一般变化不大，值的大小十分相近，因此卷积之后输出的像素值也具有相似的大小。也就是说，我们提取到的一些信息的作用是相似的，这就造成了信息冗余。池化层的作用就是为了解决这个问题，通过减少网络层输入的大小进行特征压缩。池化分为最大池化和平均池化，较为常用的是最大池化。最大池化作用如图2.12所示，用2\*2的滤波器，选用步长为2进行最大池化。

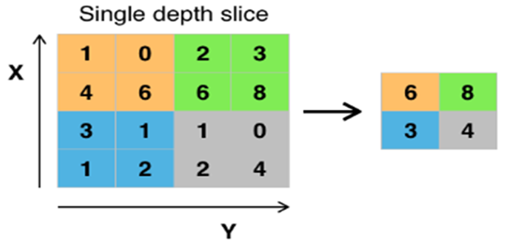


图2.12 最大池化

Fig. 2.12 Max pooling

全连接层可以看作是多层感知机，这一层的每一个结点都与上一神经网络层的所有结点进行相连，从而对之前提取的特征进行综合，这样也导致了全连接层的参数量较大。全连接层的激活函数一般是选用ReLU激活函数，在整个神经网络中起到了最终决策的作用。图2.13中的简单示意图展示了全连接层的推导过程。x1,x2,x3是全连接层的输入，a1,a2,a3是全连接层的输出。公式如下：



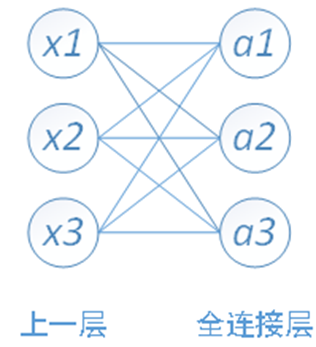


图2.13 全连接层信息传递过程

Fig. 2.13 Information transfer process of full connection layer

在卷积神经网络的一个卷积层中，一般含有许多的特征图（Feature Map），同一个特征图（Feature Map）共享权值。卷积结构使得深层网络占据的内存量大大减少，卷积神经网络中关键的三个处理分别是，局部感知、权值共享和池化，这三个步骤有效的减少了神经网络的参数量。

(一) 局部感知

首先讲解局部感知的作用。感受野就是卷积神经网络CNN中每一层的特征图（Feature Map）上的像素点映射回原始图像的区域大小，控制每一层的输出结果中单一元素所影响的输入层的区域。卷积神经网络的感受野可以用中心位置(center location)和大小(size)来表示。对于卷积神经网络来说，图像的像素值并不是具有相同的地位。如果一个像素点的位置更加靠近感受野的中心位置，那么它对于特征计算的贡献度更大。

在卷积神经网络中，如果每一个神经元都与上一层的的所有神经元进行连接，那么参数量就会大大增加，学习效率就会大幅度降低。因此，在卷积神经网络中，采取的是局部感知的方式。每一个神经元只与上一层的某一处区域的神经元进行连接，每一个神经元都不再需要感知全局图像。在更高层的卷积神经网络层中，就会将神经元对全局图像分部分进行感知的感知结果综合起来从而得到全局信息。最终这样，我们可以减少网络连接的数量，使得卷积神经网络参数量大大减少。

(二) 权值共享

卷积神经网络中，对于给定的一个图片作为输入，会用卷积核去滑行扫这个输入图片。卷积核的定义为在进行图像处理时，给定的一个输入图像，将此图像中一个局部小区域中的像素进行加权平均后作为输出图像中对应位置的单个像素，其中权值由卷积核进行定义。而卷积核设计的数字就是权重，这些数字不会因为在图像内的位置不同而发生改变，只要是被同一个卷积核扫过，权系数都是一样的，也就是所谓的共享。权值共享表示过滤器在遍历整个图像的时，它的参数值是不会发生变化的，假设共有三个过滤器，每个过滤器都会遍历整个图像，在遍历时，参数值还是和原来的值一样，即图像中的所有元素都“共享”了一样的权值。权值共享大大降低卷积神经网络的参数数量。

(三) 池化

在卷积神经网络中，我们通过池化来进一步降低参数数量，同时可以防止神经网络过拟合，减小网络的计算复杂度。除此之外，池化还能够保持了旋转、平移、伸缩等特征的某种不变性。池化一般分为两种，一种是平均池化，一种是最大池化[59]。

平均池化就是将图像某一区域中的平均值当成池化操作后的值，最大池化就是选取局部接受域中的最大值点作为池化后的值。平均池化相对于最大池化更能够降低因为邻域大小受到限制而产生的误差，更多的留住的图像的背景信息。而最大池化相对于平均池化来说更能够降低卷积层存在的参数误差导致的估计均值的偏移，图像的纹理信息得到更多保留。因此两者需要根据具体需要进行选择。

2.2 长短时记忆模型

长短时记忆神经网络即LSTM是由Jürgen Schmidhuber在1997年提出的神经网络。长短时记忆网络（LSTM）属于循环神经网络，是一种时间递归神经网络(如图2.14)。神经网络前面的神经元的信息会对后面的神经元信息造成影响，为了解决之前的神经网络不能建立起这种信息传递的问题，出现了递归神经网络。与传统的循环神经网络不同之处在于，LSTM擅长处理时间序列相关的未知大小的长时间延长。传统递归神经网络针对需要过去多个信息的预测问题，并不能有效解决这种长期以来问题，而LSTM能够解决传统的循环神经网络存在的长期依赖问题。LSTM可以对时间序列相关的数据进行分类、预测等，在手写识别、语音识别以及图片描述等领域取得了优于其他的神经网络的成果。

当信息之间的间隔不断变大的时候，传统的递归神经网络难以学习远距离信息。LSTM通过对结构的特殊设计，有效的避免了长时间依赖的问题，能够记住长期信息，使得信息能够长久化。区别于传统循环神经网络简单的重复性模块的链式结构，LSTM重复的单元内部拥有不同的结构，内部结构是四个神经网络层，四个神经网络层之间通过一种特殊的形式进行信息的交互。

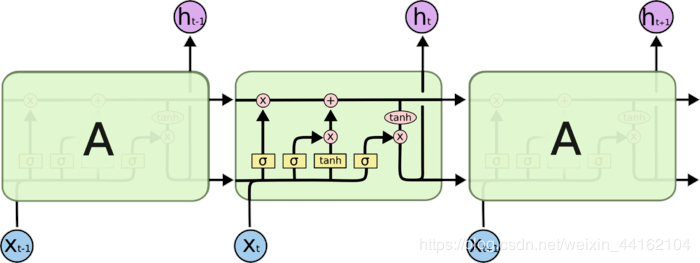


图2.14 LSTM模型

Fig. 2.14 LSTM model

传统的循环神经网络的隐藏层只有一个状态，它对于神经网络短期的输入表现得十分敏感。但是在LSTM中又再增加一个了状态，即状态c，用它来保存长期的状态，将其称之为细胞状态(cell state)。长短时记忆网络LSTM的核心就是细胞状态，信息在经过细胞的时候很大程度的保持不变的状态，只是很小的部分经由一些线性交互进行信息的更有选择性的传递。LSTM对于信息的增添与删除取决于一种称为门的结构，门结构是由Sigmoid神经层和逐点乘法操作组成(如图2.15)。其中，信息通过Sigmoid神经层输出0到1之间的数值，输出的数字表示信息通过的比重，0和1表示两种极限情况。0表示信息完全不被通过，1表示信息可以完全通过。LSTM中一共有三种门，通过这三种门来进行信息的控制，分别是输入门、遗忘门和输出门。

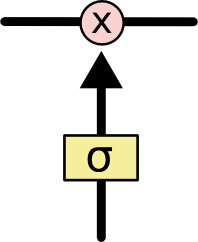


图2.15 单个信息节点由sigmoid函数和点乘操作组成，其中表示Sigmoid函数

Fig. 2.15 A single information node consists of a sigmoid function and a dot product operation, where the Sigmoid function is represented

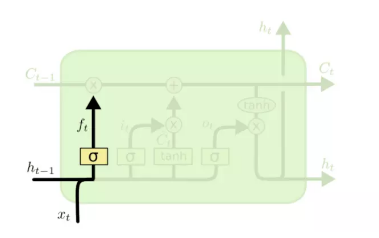
首先信息经由LSTM处理的第一步为决定哪些信息需要从细胞状态中被丢弃，这个步骤由遗忘门来实现。遗忘门读取上一时刻的输出ht-1以及当前时刻的输入信息xt，并且将上一时刻的细胞状态Ct-1中的所有数字经由Sigmoid函数输出为0到1之间的数值。0表示完全删除，1表示完全保留。当门的输出为0时，其他的任何向量与其相乘最终都会变成零向量，表示信息完全没有通过。当门的输出是1时，其他的任何向量与其相乘最终都会保持原来的状态，这就表示信息完全通过。遗忘门结构如下图2.16所示。

图2.16 遗忘门

Fig. 2.16 Forget gate

遗忘门作用公式如下所示：



其中，表示Sigmoid函数，表示权值，表示上一时刻的输出，表示当前时刻的输入，表示偏置值。

下一步是决定将哪一些新的信息加入到细胞状态中，这一步分为两个部分。首先，输入门层中的Sigmoid层选定哪些信息进行更新。然后，经由一个Tanh层生成一个候选的向量，这一向量将会被加入到细胞状态中。结构如下图2.17所示。

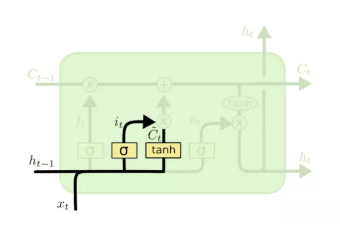


图2.17 细胞状态更新

Fig. 2.17 Update cell status

两个部分的公式如下：





经过这两个部分的处理完成了对细胞状态的更新。将当前的细胞状态和长期细胞状态组合在一起，就形成了新的细胞状态。经过遗忘门的处理来保存之前的长时间的信息，经过输入门的处理可以阻止不重要的信息进去到记忆中。将上一个状态值乘以ft，得到需要忘记的部分的表达式。然后将得到表达式的值加上得到的是新的候选值，按照决定需要更新每个状态值的多少来衡量。最终得到更新值，表达式如下：



其中，为遗忘门操作公式，为上一时刻的细胞状态，为输入门操作公式，为候选向量。

输出门最终决定要输出的信息。 此输出将基于细胞状态，得到一个过滤版本。 首先，信息通过Sigmoid层，它决定了要输出哪些部分的细胞状态。再将细胞状态由Tanh函数处理后乘以Sigmoid门的输出，至此只输出了最终决定输出的那些部分信息。输出门的结构图如图2.18所示。

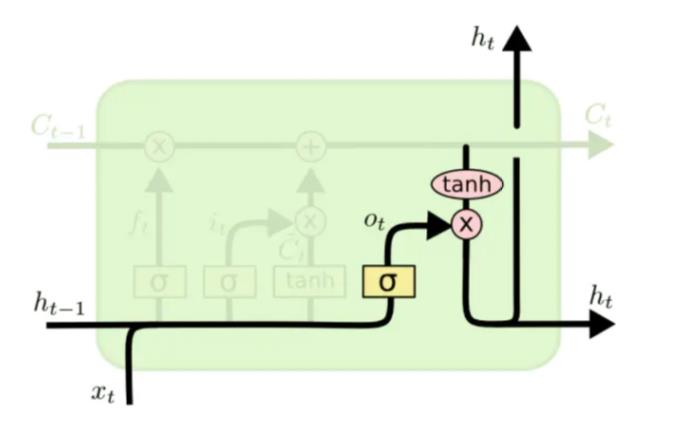


图2.18 输出门

Fig. 2.18 Output gate

操作公式如下：





其中，表示输出门操作，表示最终的输出值。

2.3 本章小结

本章节主要讲述了卷积神经网络的理论概述以及长短时记忆模型的相关理论知识。从神经网络的起源出发，讲述了神经网络的发展历程，为接下来讲解卷积神经网络打好基础。卷积神经网络一般包括卷积层、池化成以及全连接层，是一种直接的输入到输出的映射，有强大的处理图像和提取特征的能力。对卷积的核心内容进行了讲解，并对卷积的三大特点：局部感知、权值共享和池化做了详细的讲述。本节关于长短时记忆网络部分主要讲了三个“门”的操作过程。

3 数据集的制作

本文主要是通过在不同的地点采集行人过马路的视频，对视频进行一定的帧处理后，对视频帧中的行人进行骨骼关键点和面部关键点的提取。将提取后的关键点位置信息进行转换后输入到预测网络中，使得网络学会通过关键点信息预测行人是否过马路。这个章节将会分为以下四个部分讲解，分别为：视频的采集与处理、视频帧的分辨率重建、行人关键点的提取以及CSV格式数据集的制作。

3.1 视频的采集与处理

本文是在学校内以及周围区域对行人进行过马路和不过马路的视频采集，视频中行人面部基本正对镜头。将采集的视频进行合理的长度裁剪，因为我们要做的是预测行人是否过马路，所以将采集的视频中过马路之前的等待片段裁剪出来。将处理后的视频进行视频帧的分离，每个行人选择8帧的视频帧。视频示例帧如图3.1所示。



图3.1 视频帧示例

Fig. 3.1 A video frame example

3.2 视频帧的分辨率重建

在采集过程中，有些视频由于分辨率不够高，所以会影响关键点检测的准确度，因此我们使用ESPCN网络(如图3.2)进行分辨率的重建。

论文提出了图像分辨率重建网络ESPCN。ESPCN网络能够在低分辨率图像上直接进行卷积得到高分辨率图像。ESPCN提出一种新的卷积神经网络结构体系，在低分辨率空间中进行特征映射的提取[60]。ESPCN提出的核心理念是亚像素卷积，亚像素卷积层通过学习一组尺度扩展滤波器，将最终的低分辨率特征映射放大到高分辨率输出中。图像利用亚像素卷积从低分辨率放大到高分辨率，其中，插值函数在此过程中被隐含地包含在前面的卷积层中，能够被网络自动学习到。图像大小变换只在ESPCN网络最后一层进行，前面的卷积运算只是在低分辨率图像上操作，因此效率较高。使用公开的数据集进行对比测试发现该方法比之前基于卷积神经网络的方法速度快了一个数量级。

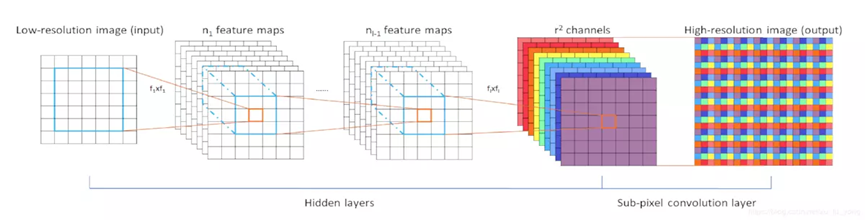


图3.2 ESPCN网络

Fig. 3.2 ESPCN network

为了衡量经过ESPCN处理后的图像质量，我们通常会参考PSNR值来确定分辨率是否提高。PSNR指的是峰值信噪比，是评价图像质量的一种客观标准。PSNR越大，代表失真越小，图像质量越好。经过ESPCN处理后的部分低分辨率视频帧的PSNR变化如图3.3所示。

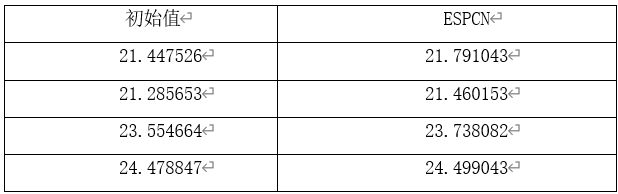


图3.3 ESPCN网络对部分低分辨率帧的分辨率重建结果

Fig. 3.3 Resolution reconstruction of some low-resolution frames by ESPCN network

3.3 行人关键点的提取

本文中需要将处理后的视频帧进行行人骨骼关键点和面部关键点的检测，从而根据提取到的关键点的位置信息来预测行人是否会过马路。本文中选用Openpose来进行关键点的提取。

3.3.1 Openpose介绍

Openpose作为第一个实时的多人姿态识别系统是由美国卡耐基梅隆大学Gines Hidalgo等人开发设计的，该开源库是基于卷积神经网络和监督学习进行开发的，以Caffe作为框架。作为全球第一个根据深度学习技术开发的实时多人的二维姿态估计项目，Openpose具有很好的实时性以及鲁棒性。Openpose可以联合检测人体、手脚以及面部的135个关键点。人体姿态估计分为top-down和bottom-up两种方法，bottom-up就是先检测人体的各部分构件，然后将多个人身上的不同构件分配到不同的人身上。top-down就是先检测人体，然后再检测每一个人的人体关键点。Openpose使用的是bottom-up方法，速度相对较快且对其进行了一定的改进。

该库和Alpha-Pose (fast Pytorch版本)以及 Mask R-CNN做了运行时间的比较(如图3.4)。对每个算法使用相同的图像进行分析，批处理大小为1。每个分析重复1000次，然后取平均值。实验都是在Nvidia 1080 Ti和CUDA8的系统上完成的。Openpose基本可以做到运行时间不随着图像中人数的增加而发生很大幅度的变化，运行时间基本始终保持水平线。Opnepose可以有多方面的应用，比如人体姿态三维重建，多人运动监测骨骼、面部以及手脚的关键点检测等。

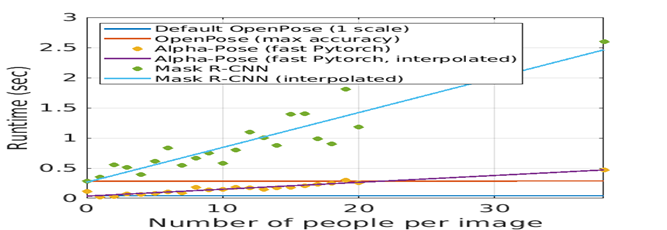


图3.4 Openpose与其他算法的实时性比较

Fig. 3.4 Real-time comparison of Openpose and other algorithms

3.3.2 Openpose原理

Openpose是基于卷积神经网络以及监督学习开发的人体姿态估计开源库。该开源库的作用过程基本可以分为四个部分：卷积神经网络获得heatmap、分支卷积网络预测关键点位置以及置信度、寻找肢体连接、肢体拼装成完整体。

首先第一步是通过VGG19卷积神经网络(如图3.5)的前十层神经网络层进行图像特征的提取。

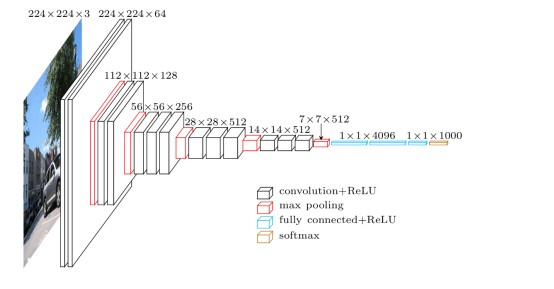


图3.5 VGG19卷积神经网络

Fig. 3.5 VGG19 convolutional neural network

VGG19神经网络是经典的卷积神经网络之一，[VGG](http://link.zhihu.com/?target=https%3A//arxiv.org/abs/1409.1556)是英国牛津大学课题组提出的，共有19个参数层[61]。VGG中，使用了3\*3卷积核来提升神经网络的作用效果。

(一) 使用3\*3的卷积核。VGG19对于AlexNet的一个比较大的不同之处在于，使用了一连串的小的卷积核去代替原先的AlexNet网络中的5\*5、7\*7以及11\*11的大卷积核。这样的处理之后，增加了神经网络的层数以便于学习更加复杂的特征，参数量有了明显的减少。当感受野相同时，提高了神经网络的学习效果，能够学习到更加细节化的特征。

(二) 使用2\*2的池化核

在AlexNet神经网络中，使用的是3\*3的池化核，而在VGG19神经网络中，使用的是小池化核，大小为2\*2。这样使得层数更深时，得到更宽的特征图，这样模型构架更深更宽，计算量的大小增加缓慢。

(三) 使用更多的通道数

在VGG网络中，很大的增加了通道数的数量。它的第一层的通道数为64，其后的每一层都将通道数进行翻倍，最多可达512个。通过增加通道数的数量，使得神经网络可以从输入数据中提取到更多的信息，提取到更加丰富的特征。

第二步是将提取到的特征输入到接下来的多级卷积神经网络(如图3.6)中，网络分为两个分支，每个分支都有t个阶段，每一个阶段会对得到的特征图进行融合。第一分支用来预测部位置信图Part Confidence Maps，第二分之用来预测部位亲和域Part Affinity Fields，Part Affinity Fields用来描述关节点的走向。将每个阶段的预测及其对应的图像特征连接到每个后续阶段。具体过程如下，图像首先由卷积网络生成一组特征映射F，输入到每个分支的第一阶段。在第一阶段,网络产生一组检测置信度图和一组part affinity fields ,和是的第一阶段的推理CNN。在随后的每个阶段，将前一阶段的两个分支的预测与原始图像特征F进行连接，用于产生精细化的预测。和 是t阶段的推理CNN。t阶段的和的表达公式如下：





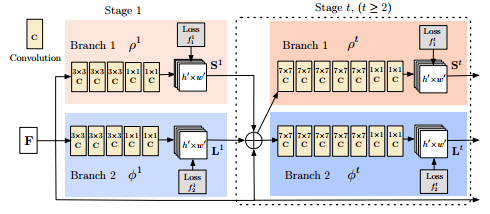


图3.6 二分支的多级CNN结构

Fig. 3.6 Two-branch multistage CNN structure

第一个分支的每个阶段预测置信度图，第二个分支的每个阶段预测PAFs 。每个阶段后，将两个分支的预测与图像特征连接起来，输送到下一个阶段。

为了能对不同的目的进行预测，使用了不同的损失函数，损失函数采用损失，并对损失函数进行了一个空间上的加权。当图像位置缺乏标记时，公式中的二进制掩码等于零,防止在训练过程中错误的惩罚真阳性预测。在每一个阶段，会添加中间监督，通过周期性地补充梯度来有效避免梯度消失问题的发生。两个分支在t阶段的损失函数分别为:



其中，是置信度图的真值，是PAFs的真值，是二进制掩码，为图像位置。

对于Confidence Maps的检测，若处于理想状态下，一个图片中如果出现单人，且它的关键点的相应部位是可见的，那么每一个关键点对应的置信度图就会有一个峰值出现。当图片中出现多个人时，这时候每一个关键点对应的部分就会有多个峰值。置信度图片中出现的峰值就是关键点。我们取置信度图的最大值而不是平均值，这样峰值附近的精度才会保持不同。在测试阶段，我们在进行置信度图的预测时，通过非极大值抑制来得到heatmap中的峰值。

第三步，我们使用获得的用关节信息和Part Affinity Fields来获得肢体连接关系。我们已经获得了一组检测到的人体关键点，接下来就是将它们进行组装。在之前的检测中，存在着两个重要的问题：(1)只对肢体的位置信息进行编码，而没有对肢体的方向进行编码;(2)将肢体的支撑区域缩小为一个点，仅仅用一个点去表示一个区域。因而检测错误率一直很高。Part Affinity Fields (PAFs)解决了以上存在的问题。它们在检测时不仅仅保存了肢体支撑区域的位置信息还保存了它的方向信息。对于特定的肢体区域内的每个像素，PAF作为一个二维向量，编码了肢体区域内从一个关键点指向另一个关键点的方向，弥补了之前方法存在的缺陷。例如，要寻找所有的左上臂，已经有左肩和左肘part信息，以及左上臂的Part Affinity Fields信息。将左肩中的所有part与左肘的所有part进行匹配，计算两个part间的Part Affinity Fields信息的积分，作为该part相连的score。最后将所有score排序，来确定part之间是否相连。这样我们就获得了所有的连接信息，每一个连接都可以看作是一个肢体。

最后一步，我们将肢体拼装成完整个人，在获得了所有肢体后，将拥有相同关节的肢体看作是同一个人的肢体。将同一个人的关节点进行连接，由于Part Affinity Fields自身所有的矢量性，使得生成的偶匹配很正确，将拥有同一个关节的肢体看作属于同一个人的肢体，最终合并为一个人的整体骨架。

当图像中出现多人时，我们首先对关键点进行检测，对检测的置信度图进行非极大值抑制操作从而获得关键点位置。将检测得到的关键点的候选集合中的每个候选关键点与其他的关键点候选进行一个关联性的匹配，通过寻找匹配，找到关键点对，从而通过最有匹配形成肢体。论文中使用贪婪松弛算法，保证匹配的高质量，使得肢体都能够对应到每个人上。文中增加了两个松弛：1.选择数量最小边来得到一个人的姿态的生成树骨架，不使用完整的图。2.将匹配问题进行分解，分为一系列的二部匹配子问题，分别确定相邻树节点的匹配。分别获得每个肢体类型的候选肢体连接。对于所有的肢体连接候选，将共享同一关键点检测候选体的肢体进行组合，组合成多人的全身姿势，这种方法大大加快了检测的速度。

一般的人体姿态检测都是运用top-down的方法，也就是说将人检测出来，在基于每个人进行关键点检测，因此会造成第一步如果无法进行，第二步就会失败。其次，一般的检测方法的运行时间都是和人数正相关。人越多，消耗的时间就越多。Openpose使用了bottom-up方法，为了克服无法利用全局上下文信息的特点，运用了Part Affinity Fields部分亲和域的理念，编码肢体位置和方向。利用贪婪算法对全局上下文进行重组的编码，从而获取高质量的检测结果，只是以增加小部分的计算成本为代价，运行基本达到实时状态，最终消耗时长与图像中出现的人的数量没有强关联，且具有很高的鲁棒性。

本文在视频帧上使用Openpose检测之后获得的关键点检测效果如图3.7所示。



图3.7 视频帧关键点检测示例

Fig. 3.7 Keypoints detection on a frame sample

3.3.3 关键点位置信息表示

Openpose检测得到的人体骨骼关键点和面部关键点信息包括了x、y关键点坐标以及检测置信度。关键点的坐标用科学计数法表示。关键点坐标是以图片坐标系的左上角作为原点，以人脸面对的朝向，x轴方向向右，y轴方向向下。Openpose检测获得的骨骼关键点有25个，面部关键点有70个，所以一共是95个关键点。骨骼关键点包括鼻子、肩膀、脖子、手腕、手臂以及脚后跟等部位。面部关键点涉及眼睛、嘴巴、眉毛以及鼻子等部位。骨骼关键点检测效果图如图3.8所示，面部关键点检测效果图如图3.9所示。

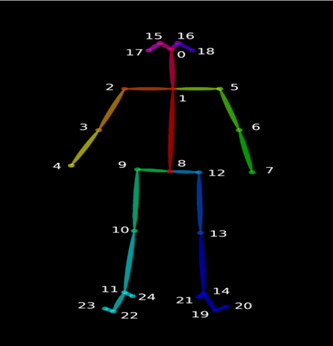
图3.8 视频帧关键点检测

Fig. 3.8 Keypoints detection of skeleton

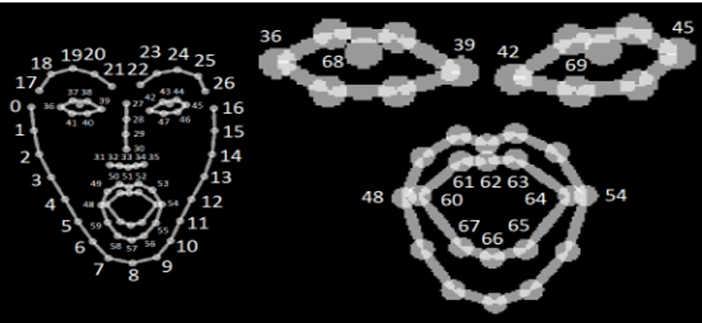


图3.9 面部关键点检测

Fig. 3.9 Keypoints detection of face

3.4 CSV格式的数据集制作

Openpose输出的数据集格式为json或者yml，本文中按照需要将其转换为CSV格式，只使用行人的关键点位置信息。将关键点坐标从科学计数法表示转换为一般表示(如图3.10)，并进行归一化处理(如图3.11)。数据原始大小为190\*8，即每个行人的95个关键点的x和y坐标共190个数值以及每个行人取8帧视频帧进行预测。将行人ID、帧序号以及标签加入数据后，数据大小变化为193\*8，其中未检测到的关键点的坐标值为0。数据的标签分为1和0，分别表示过马路和不过马路。因为原始的Openpose提取出来的关键点的坐标不仅仅和行人在图像中出现的位置有关系，还与行人与摄像机距离的大小也有关系。因此，我们需要减小这些因素对预测结果的影响，将坐标大小进行归一化。归一化的公式如下所示：





其中，x表示横坐标，y表示纵坐标，x min或y min表示当前帧的 x或y的最小值，x max或y max表示当前帧 x或y的最大值。

最后共获得了()条数据。其中，训练集()个，测试集()个。

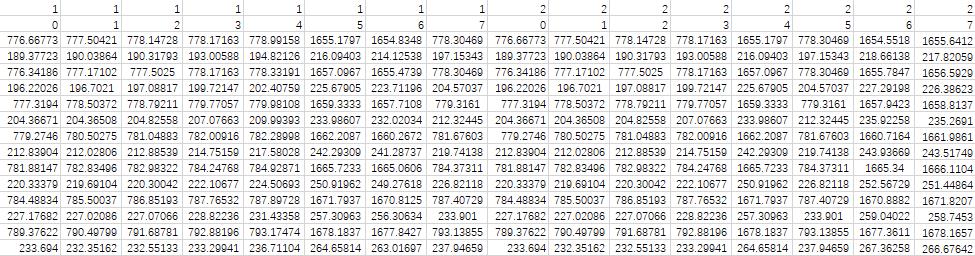
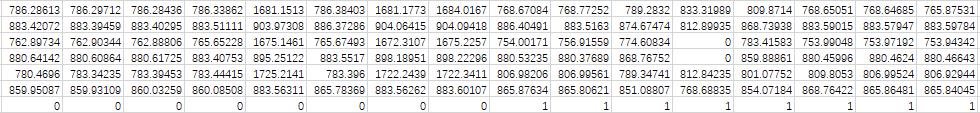
图3.10 部分归一化之前的行人数据

Fig. 3.10 Partial pedestrian keypoints Data before normalization

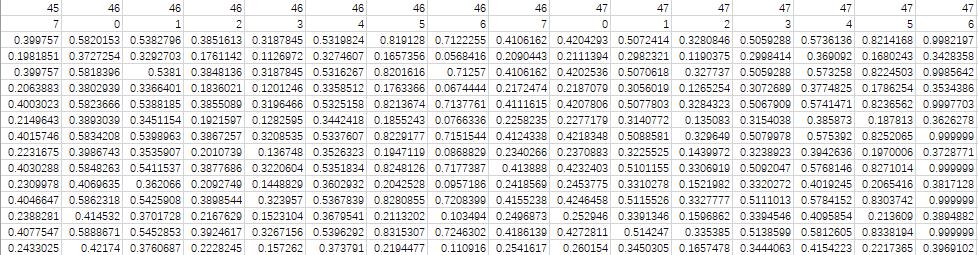
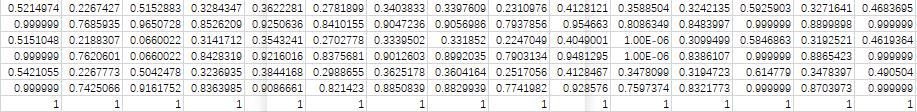
图3.11 部分归一化之后的行人数据

Fig. 3.11 Partial pedestrian keypoints Data after normalization

3.5 本章小结

本章节主要包含以下内容：视频的采集处理、行人骨骼关键点和面部关键点的提取以及CSV的制作。将采集到的视频数据进行裁剪和分辨率重建等处理，使得视频帧能够更加符合使用要求。对使用工具Openpose开源库的关键点提取原理进行了详细的说明。首先介绍Openpose相对于其他的关键点提取工具所具有的优势之处，随后对Openpose库提取关键点的过程进行了讲解。介绍了关键点坐标的标注特点，最后讲述了CSV数据集的制作过程。

4 基于图卷积神经网络和长短时记忆模型的行人通行意图预测

4.1 图卷积神经网络原理概述

图（Graph）在不同领域有着不同的表现形式，数学中，图是图论的主要研究对象，在计算机科学领域，图又是一种常见的数据结构。图可以表示为顶点（Vertex）和连接顶点边的集合，记作G=（V，E），其中V是顶点集合，E是边集合，如图4.1所示。

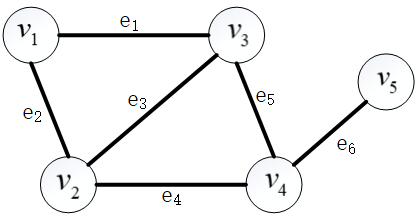


图4.1 图G结构图

Fig. 4.1 Structure of Graph G

作为一种基本的数据结构，图在计算机中一般采用邻接矩阵的形式存储，当使用邻接矩阵进行图的存储时，需要使用一个一维数组表示顶点的集合，一个二维数组表示邻接矩阵，描述图中顶点的关联，定义为：



其中，表示为图中顶点的连通性，特别的，由于图连接的特性，在邻接矩阵中会出现大量的0值，因此可以使用稀疏矩阵的形式来存储邻接矩阵，使邻接矩阵的空间复杂度控制在的范围内。图4.2给出了图4.1的邻接矩阵和顶点集合的存储表示，如图所示，图的邻接矩阵是关于对角线对称的，即。

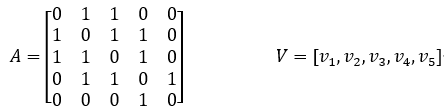


图4.2 图G的邻接矩阵以及顶点集合

Fig. 4.2 The adjacency matrix and vertex set of graph G

图卷积的定义：（需不需要另起标题啊 前面是图的定义 下面这些是图卷积的定义）

常规欧几里得数据（图像或文本）的标准卷积不适用于一般的图结构数据，借鉴卷积神经网络（CNN）在计算机视觉以及图像处理领域的成功实践，作为一种高速有效的数据处理方法，将图信号处理（Graph Signal Processing，GSP）中定义的卷积操作（式4.2）推广到图数据处理任务中，在自然语言处理、图生成、计算机视觉以及知识图谱方面具有重要的意义。



其中，，是定义在图G中的两组信号，表示哈达玛积，GFT表示图傅里叶变换（Graph Fourier Transform，GFT），IGFT表示图上的逆傅里叶变换（Inverse Graph Fourier Transform，IGFT）。

将图数据转化为结构化的数据形式，目前有两种基本方法。一个是扩展卷积的空间定义，另一个是用图傅立叶变换在谱域中进行操作。前一种方法是将顶点重新排列成某些网格形式，该形式可以通过正常的卷积运算来处理，其中卷积滤波器直接应用于图节点及其邻居，称为空域图卷积。后者是在频谱域中应用卷积的频谱框架，通常称为频域图卷积。本文所采用的是基于后者的方法。现有的几项频域研究主要通过将图卷积的计算复杂度降低到线性来使图卷积更有效率。

论文中，通过引入图卷积算子“”对图信号进行频域空间的图卷积操作，将信号与核H（频率响应函数）做卷积操作：



其中，为图的拉普拉斯矩阵，其定义为：，为对角矩阵，表示的是节点的度，为邻接矩阵，由于是是实对称矩阵，对其进行正交对角化，可得：，是拉普拉斯矩阵的特征向量，是拉普拉斯矩阵特征值组成的对角矩阵。滤波器也是对角矩阵。

式(4.3)即为频域视角下的滤波（空域下的卷积）操作，其变换过程由三步组成：

1.通过图傅里叶变换（GFT），即把图信号变换到频域空间；

2.通过滤波器对频域空间的信号进行处理；

3.通过逆图傅里叶变换（IGFT），将频域空间处理的信号反解为图信号。

文中，在进行网络训练的过程中，为加强训练学习时的数值稳定性，降低进行特征分解时需要计算拉普拉斯矩阵的特征值以及特征向量所造成的计算时间的时间复杂度，仿照拉普拉斯矩阵对L进行重归一化处理。令，，。其特征值范围在重归一化后范围为(-1,1)，该操作可以有效的防止网络训练优化过程中的梯度消失及梯度爆炸现象。定义网络训练的参数化矩阵对输入信号进行高维映射。式(4.4)定义了图卷积层，表示了图卷积网络的卷积过程：



其中，为输入信号矩阵经过图卷积层（GCN layer）的输出。

4.2 基于图卷积神经网络和长短时记忆模型的预测网络

基于图卷积神经网络和长短时记忆模型的预测网络提取了人体关键点的时间信息和空间信息，相对于单纯的长短时记忆模型对时间序列特征的处理，增加了对关键点空间序列特征的处理。提取的骨骼关键点和面部关键点在图像上的显示是拓扑结构，因此本文利用图卷积神经网络对拓扑图的处理能力，提取关键点的空间特征，与长短时记忆模型结合，获得更好的预测效果。

4.2.1 网络结构

基于人体骨骼关键点和面部关键点的图卷积神经网络设计：

(一) 输入模型的坐标向量生成

本文中的预测网络是基于图卷积神经网络和长短时记忆模型创建的。关键点信息利用Openpose开源库处理视频数据获得。视频数据是8帧视频帧，每帧都有一组关键点坐标，包含行人的70个面部关键点和25个骨骼关键点共95个关键点，坐标包括x坐标和y坐标。使用给定二维坐标形式的关键点序列，构建一个无向空间图，其中图节点为人体结构中的关键点，关键点的自然连接为图边。因此，图卷积神经网络的输入是图节点上的联合坐标向量，输入经过维度处理后为四维张量(N,8,95,2)，N是数据集中行人的数量。

(二) 关键点空间图模型的搭建

关键点序列由每帧中每个人的关键点2D坐标表示。连接所有关键点的坐标向量，以形成每帧的单个特征向量。在论文中，我们利用空间图形成骨架序列的层次表示。特别地，我们在具有N个关键点和T帧的骨架序列上构建T个无向空间图，本文中N=95，T=8。是一组有限的顶点，对应于行人的95个关键点;E是一组边，表示关键点之间的连通性，而表示图的邻接矩阵。

在图G中，第t帧的节点集，包括骨架序列中的所有关键点。因为本文中只涉及x和y,不涉及置信度, 所以作为网络模型的输入即某个点上的特征向量不包括置信度cor，即，由帧t上的第i个关键点的坐标向量组成。根据人体结构的连通性，将第t帧内的具有连通性的关键点通过邻接矩阵连接。本文中的邻接矩阵大小为95\*95，其中面部关键点序号为0到69，骨骼关键点序号为70到94，根据第三章中图3.8和图3.9的关键点的连接关系，填充邻接矩阵，0表示无连接，1表示连接。每帧的数据经过图建模后，将关键点的特征向量经过多层感知器MLP映射到高维空间作图嵌入,如式(4.5)所示，数据大小变为(N,8,95,128)，之后经过图卷积神经网络GCN层对关键点图模型进行卷积处理提取关键点的高维特征,如式(4.6)所示，维度变为(N,8,95,64)，其中N为行人数量。最后将数据经过全连接层输出。





其中，为图嵌入层和图卷积层的权重矩阵。式(4.6)中的操作步骤相当于基于图像的卷积神经网络的卷积操作，其中卷积神经网络的输入由图像的像素强度矢量构成。

之后将每个时间帧上的特征拼接，拼接后的向量表示如下所示：



本文的关键点图模型搭建示意图如图4.3所示。

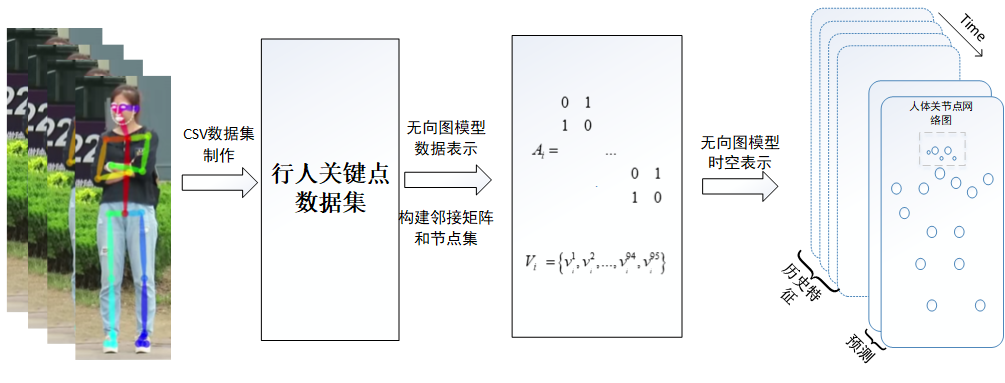


图4.3 模型搭建

Fig. 4.3 Model structure

(三) LSTM层输出预测结果

经过维度处理后作为LSTM网络层的输入，维度大小为(8,N,3040)，N为行人的数量。经过LSTM层提取8帧的时间序列特征，经过多层感知器（MLP），使用Sigmoid分类函数获得当前帧的动作概率，设置阈值为0.5,与输出的概率值进行比较，判断行人是否过马路。当概率值大于阈值时，判断行人意图为过马路。当概率值小于阈值时，判断为行人不过马路。

本文设计的预测网络整体结构如下图4.4所示。

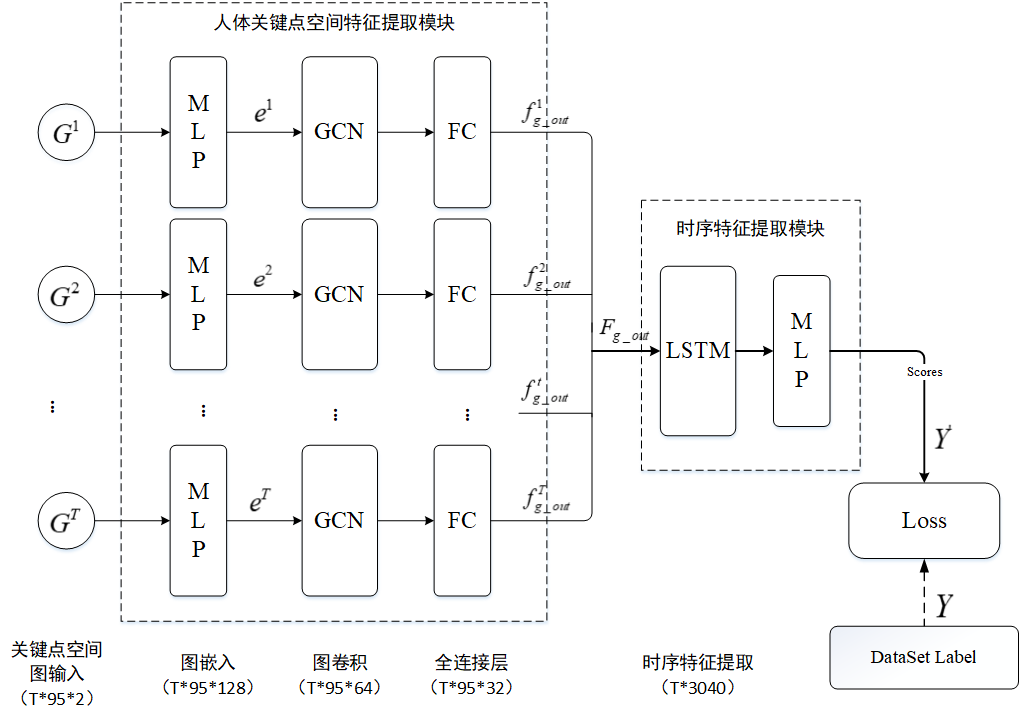


图4.4 预测网络结构

Fig. 4.4 Prediction network structure

4.2.2 损失函数

因为预测网络最终输出的是行人是否过马路的判断结果，而行人过马路的标签分为0和1，分别表示不过马路和过马路。因此我们选择使用二元交叉熵来衡量损失。二元交叉熵损失函数的表达式定义为：



其中L指的是网络预测值和真实值之间的误差。N指的是样本数量，i指的是样本的顺序。指的是真实值，指的是神经网络的预测值。

4.2.3 优化方法

本文采用Adam优化器来进行优化，Adam引进二阶动量，结合AdaGrad和RMSProp两种优化算法的优点。综合考虑梯度的均值和梯度的未中心化的方差，计算出更新的步长。实现起来简单方便，计算效率高，对于内存的要求低。其中，初始学习率设置为0.001。

本文采用Dropout正则化，Dropout正则化是解决模型过拟合的一种十分常见的处理方法。在训练的过程中，使用Dropout会随机的舍弃一部分的神经元，相当于训练多个不同的网络最后进行了取均值的效果。每次都会针对不同的一批数据生成不同的网络模型，减小了神经元之间的相互依赖，最后将这些不同的模型进行组合，使得模型的泛化能力大大提高。在测试时，Dropout的功能就会被舍弃。本文Dropout概率设置为0.2。

4.2.4 Sigmoid分类函数

本文中选用Sigmoid函数作为分类函数，数据经由Sigmoid函数处理后变为一个0到1之间的概率值，我们将概率阈值设置为0.5。当概率值大于0.5时，判断行人意图为过马路，当概率值小于0.5时，判断行人意图为不过马路。函数的公式如下：



其中，为Sigmoid层的输入值，为输出的概率值。

4.3 预测结果对比及验证

我们使用Tensorboard生成网络在测试集上的曲线图。Tensorboard对Tensorflow，Keras，PyTorch都有很好的支持，在Keras中通过回调函数callbacks模块调用Tensorboard实现网络的预测结果展示。本文中训练预测网络所用的电脑硬件和软件配置环境如表4.1所示。

表4.1 电脑配置

Tab. 4.1 Computer configuration

|  |  |
| --- | --- |
| 项目 | 配置 |
| 操作系统 | Ubuntu 18.04 LTS |
| GPU | NVIDIA Titan X |
| 深度学习框架 | Keras |
| 数据处理 | Python |

本文的对比环节中，首先只使用长短时记忆模型构建的预测网络对行人意图进行预测，后与本文所提出的图卷积神经网络结合长短时记忆模型构建的预测网络结果进行对比。其中为保证实验效果的准确性和对比结果的可靠性，长短时记忆模型构建的预测网络的训练环境和参数设置与图卷积神经网络结合长短时记忆模型构建的预测网络的训练环境和参数设置（如损失函数等）保持一致。对比实验中，长短时记忆模型构建的预测网络直接读取CSV数据作为输入。数据经过LSTM层和全连接层输出，Sigmoid函数将其处理为概率值后与0.5的阈值作比较。其中，大于阈值的判断为过马路，小于阈值的判断为不过马路。最终在测试集上预测准确率为73%，如图4.5所示。其中，横坐标为测试集数据读取顺序，过马路和不过马路的数据各100个，共200个。纵坐标为类别。上半部分折线图为标签，下半部分折线图为预测结果。准确率为测试集中预测正确的数据数量除以测试集总数量。（换图）

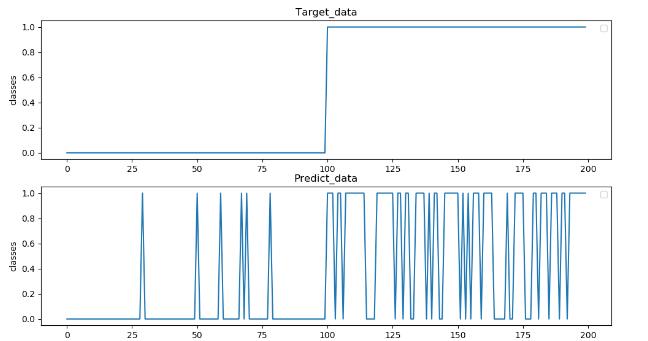


图4.5 测试集在基于LSTM构建的预测网络上的预测结果

Fig. 4.5 The prediction results of the prediction network based on LSTM on test set

图卷积神经网络结合长短时记忆模型构建的预测网络在测试集上的预测结果如图4.6所示。

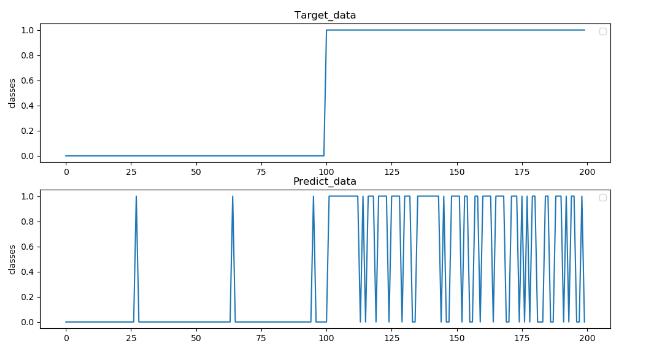


图4.6 测试集在基于LSTM和GCN构建的预测网络上的预测结果

Fig. 4.6 The prediction results of the prediction network based on LSTM and GCN on test set

该结果与图4.5表示的基于LSTM的预测网络的预测结果对比可知，GCN结合LSTM的预测网络在测试集上的准确率较LSTM预测网络提高了6%，准确率为79%。图中，横坐标为测试集数据的读取顺序，过马路和不过马路的数据各100个，共200个。纵坐标为类别。上半部分折线图为标签，下半部分折线图为预测结果。准确率为测试集中预测正确的数据数量除以测试集总数量。

本节在测试集上使用训练过的预测网络模型进行测试，通过测试模型在测试集上的准确率对网络的性能进行评估。结果表明本文提出的网络模型在测试集上有较为良好的表现，通过添加面部关键点信息，将骨骼关键点与面部关键点结合，充分利用了对行人意图预测有效的信息，使得预测网络获得了较为准确的预测结果。图卷积神经网络的前处理过程相对于单纯的长短时记忆模型预测增加了对空间序列信息的处理，有效的对关键点拓扑图进行空间特征的提取，发挥了图卷积处理拓扑图的优势。而长短时记忆（LSTM）模块通过对时间序列上的信息进行处理，实现了时间信息和空间信息的结合，准确率的提升说明图卷积神经网络的前处理对预测效果有积极的提升作用。

4.4 本章小结

本章节主要是针对预测网络的构建过程和最后的预测结果进行了描述。本文通过图卷积神经网络提取空间特征生成更高级别的特征映射。然后，将各个时间帧上的特征映射通过LSTM网络层处理，提取时序特征，经过多层感知器（MLP），由Sigmoid分类函数获得预测时刻的动作概率，与阈值比较后，判断行人是否过马路。通过构建该预测网络，学习提取的行人骨骼关键点和面部关键点数据的内在规律，使得预测行人通行意图的目的得以实现。综上所述，本章将图卷积神经网络与长短时记忆模型结合，构建了一种能够进行行人通行意图预测的预测网络。该预测网络能够学习数据的内在规律，基于行人的关键点位置信息，最终预测行人是否过马路的通行意图。

5 总结与展望

5.1 论文总结

针对行人通行意图预测的目的，本文提出了一种基于图卷积神经网络和长短时记忆模型的行人通行意图预测方法。本文的所做主要工作以及研究内容如下：

(一) 视频的采集与数据的制作

使用摄像机在学校内部以及周围采集行人视频，然后进行视频的帧分离，再将分离帧进行分辨率重建。经Openpose提取行人关键点信息并制作为CSV格式的数据集，包括帧信息、行人ID、归一化后的行人面部关键点和骨骼关键点位置坐标以及标签。标签分为两类过马路和不过马路，分别为1和0。

(二) 构建预测网络结构

本文中利用2D坐标形式的身体关键点序列，构建一个无向空间图，其中图节点为人体结构中的关键点，关键点的自然连接为图边。将每帧的数据经过图建模后，将关键点的特征向量经过多层感知器MLP映射到高维空间作图嵌入，之后经过GCN层对人体关键点图模型进行卷积处理提取关键点的高维特征。对输入数据进行多层空间上的图卷积运算，从而在图上生成更高级别的特征映射。最后经过全连接层输出。然后，将各个时间帧上的特征映射通过LSTM网络层处理，提取时序特征，经过多层感知器（MLP）处理，由Sigmoid分类函数获得动作概率，与设定阈值进行比较，判断行人是否过马路。

(三) 网络的结果对比及验证

将基于长短时记忆模型的预测网络的预测结果与图卷积神经网络和长短时记忆模型结合的预测网络的预测结果进行对比。通过在测试集上得到的准确率对网络进行效果评估。通过图卷积神经网络对数据的前处理使得预测准确率提高了6%，最终获得79%的准确率，说明本文提出的预测网络能够较好的对行人意图进行预测且图卷积神经网络的处理对预测结果起积极作用。

5.2 不足与展望

5.2.1 本方法的优势与不足

本文中提出的行人通行意图预测方法相对于传统方法不同，传统方法基本都是基于行人的轨迹和动作等来进行行人意图预测的，对于行人本身的信息的利用度不高。本文中提出的方法增加了面部关键点信息，面部关键点可以体现人的微表情，而微表情与人即将做出的动作息息相关，增加面部关键点信息使得行人意图预测的信息利用更充分。文中提取行人的骨骼关键点和面部关键点，利用图卷积神经网络对关键点的空间特征进行处理，利用长短时记忆模型对时序特征进行处理，获得了较好的预测效果。

本方法还有很多不足之处，本文提出的行人通行意图预测方法使用时有一定的局限性。首先，数据的采集都为正面行人视频采集，普适性不高，存在特殊性。其次，关键点存在遮挡问题，部分帧难以提取到完整的行人关键点信息，每个行人帧数的选取以及判断阈值未使用其他值进行效果比较。由于时间原因，这些问题还有待解决，数据类型有待丰富。

5.2.2 未来展望

本文提出的方法可以与未来的智能交通设备相结合，通过设备采集到行人的关键点信息后输入到我们的行人意图预测网络中，当行人表现出通行意图时，能够方便驾驶员及时收到反馈，停车礼让行人。当行人没有检测出通行意图时，驾驶员可以安心行驶，避免不必要的等待时间，节约时间成本。未来可采用准确度更高的骨架关键点和面部关键点提取算法，用时更少，更大程度的节约驾驶过程中的等待时间。多角度收集行人的关键点信息，使得信息的密度更高，通过提高数据的质量，使得该预测方法可以具有更好的实用性和普适性。

参 考 文 献

[1] VEHITS 2018 - Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems, v 2018-March, p 341-348, 2018

[2] Himanen V , Kulmala R . An application of logit models in analysing the behaviour of pedestrians and car drivers on pedestrian crossings[J]. accident analysis & prevention, 1988, 20(3):187-197.

[3] Manuszak M , Das S , Manski C . Walk or Wait? An Empirical Analysis of Street Crossing Decisions[J]. Gsia Working Papers, 2004.

[4] T. Gandhi, M.M. Trivedi. Pedestrian collision avoidance systems: a survey of computer vision based recent studies[C]// IEEE Intelligent Transportation Systems Conference. IEEE, 2006.

[5] Goldhammer M , Gerhard M , Zernetsch S , et al. Early prediction of a pedestrian's trajectory at intersections[C]// Intelligent Transportation Systems (ITSC), 2013 16th International IEEE Conference on. IEEE, 2013.

[6] Bonnin S , Weisswange T H , Kummert F , et al. [IEEE 2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC) - Qingdao, China (2014.10.8-2014.10.11)] 17th International IEEE Conference on Intelligent Transportation Systems (ITSC) - Pedestrian crossing prediction using multiple context-based models[J]. 2014:378-385.

[7] Volz B , Mielenz H , Siegwart R , et al. [IEEE 2016 IEEE Intelligent Vehicles Symposium (IV) - Gotenburg, Sweden (2016.6.19-2016.6.22)] 2016 IEEE Intelligent Vehicles Symposium (IV) - Predicting pedestrian crossing using Quantile Regression forests[J]. 2016:426-432.

[8] Zhao J , Xu H , Liu H , et al. Detection and tracking of pedestrians and vehicles using roadside LiDAR sensors[J]. Transportation Research, 2019, 100(MAR.):68-87.

[9] Abramson Y , Steux B . Hardware-friendly pedestrian detection and impact prediction[C]// IEEE Intelligent Vehicles Symposium. IEEE, 2004.

[10] Chen Z , Ngai D C K , Yung N H C . Pedestrian Behavior Prediction based on Motion Patterns for Vehicle-to-Pedestrian Collision Avoidance[C]// Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on. IEEE, 2008.

[11] Coelingh E, Eidehall A, Bengtsson M. Collision warning with full auto brake and pedestrian detection-a practical example of automatic emergency braking[C]. The 13th Int IEEE Conf on Intelligent Transportation Systems(ITSC). Funchal: IEEE, 2010: 155-160.

[12] Kohler, S , Goldhammer , M , Bauer, S , et al. Early Detection of the Pedestrian's Intention to Cross the Street[J]. 2012.

[13] Kohler S , Schreiner B , Ronalter S , et al. Autonomous evasive maneuvers triggered by infrastructure-based detection of pedestrian intentions[C]// IEEE Intelligent Vehicles Symposium. IEEE, 2013.

[14] Braeuchle, C , Flehmig, F , Rosenstiel, W , et al. Maneuver decision for active pedestrian protection under uncertainty[C]// International IEEE Conference on Intelligent Transportation Systems. IEEE, 2013.

[15] Gepperth A , Garcia Ortiz M , Heisele B . [IEEE 2013 16th International IEEE Conference on Intelligent Transportation Systems - (ITSC 2013) - The Hague, Netherlands (2013.10.6-2013.10.9)] 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013) - Real-time pedestrian detection and pose classification on a GPU[J]. 2013:348-353.

[16] Kropf, T , Rosenstiel, W . Situation analysis and decision making for active pedestrian protection using Bayesian networks[J]. 2013.

[17] Keller C. G., Gavrila D. M. “Will the pedestrian cross? a study on pedestrian path prediction,”IEEE Transactions on Intelligent Transportation Systems, vol. 15, pp. 494–506, 2014.

[18] Kwak J Y , Ko B C , Nam J Y . Pedestrian intention prediction based on dynamic fuzzy automata for vehicle driving at nighttime[J]. Infrared Physics & Technology, 2017, 81:41-51.

[19] Voelz B , Mielenz H , Gilitschenski I , et al. Inferring Pedestrian Motions at Urban Crosswalks[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(2):544-555.

[20] Oxley J A , Ihsen E , Fildes B N , et al. Crossing roads safely : An experimental study of age differences in gap selection by pedestrians[J]. Accident Analysis & Prevention, 2005, 37(5):962-971.

[21] Schmidt S , F?Rber B . Pedestrians at the kerb – Recognising the action intentions of humans[J]. transportation research part f psychology & behaviour, 2009, 12(4):0-310.

[22] Brechtel S , Gindele T , Rüdiger Dillmann . Learning Context Sensitive Behavior Models from Observations for Predicting Traffic Situations[C]// International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, 2013.

[23] Volz B , Mielenz H , Agamennoni G , et al. Feature Relevance Estimation for Learning Pedestrian Behavior at Crosswalks[C]// IEEE International Conference on Intelligent Transportation Systems. IEEE, 2015.

[24] Knorr A G , Willacker L , Hermsd?Rfer J , et al. Influence of person - and situation-specific characteristics on collision avoidance behavior in human locomotion.[J]. Journal of Experimental Psychology: Human Perception and Performance, 2016, 42(9):1332-1343.

[25] Lie G , Rong-Ben W , Bai-Yuan G U , et al. Overview of Worldwide Intelligent Vehicle Pedestrian Detection Technology[J]. Journal of Highway and Transportation Research and Development, 2005.

[26] Ma G S , Yao J , Yang X G . Prediction of the Position of Pedestrian Crossing Road Section Based on Kalman Predictor[M]. IEEE Computer Society, 2009.

[27] 刘光新, 李克平, 孙剑. 信号控制交叉口行人过街等待时间研究[J]. 中国安全科学学报, 2009(09):2+163-170.

[28] 朱芳芳. A Model of Pedestrian Gap Selecting in Crosswalks at Unsignalized Intersections%无信号控制交叉口行人过街间隙选择模型研究[J]. 交通信息与安全, 2011, 29(2):36-39.

[29] Mingjun L , Kaiying W , Yuying S , et al. Pedestrian Crossing Decision-Making Model at Unsignalized Mid-Block[J]. journal of beihua university(natural science), 2015.

[30] 楼佳妮. 大型活动行人交通分析、预测及仿真[D]. 北京:北京交通大学,2017.

[31] 黄智, 何颖, 刘象祎,等. 考虑过街行人运动不确定性的人-车碰撞伤害评估[J]. 汽车工程, 2018(1):28-33.

[32] Chen Z , Yung N . Improved multi-level pedestrian behavior prediction based on matching with classified motion patterns[C]// International IEEE Conference on Intelligent Transportation Systems. IEEE, 2009.

[33] 车领. 智能车辆对行人的运动感知和过街意图预测研究[D]. 西安:西安理工大学,2019.

[34] 广州大学.基于HMM算法的行人过马路的预测方法和系统:CN201810861624.0[P].2018-12-21.

[35] Likun W , Lu L , Dameng Y , et al. Pedestrian trajectory prediction in crossing scenario using fuzzy logic and switching Kalman filter[J]. The Journal of China Universities of Posts and Telecommunications, 2018, 25(06):35-47.

[36] 北京交通大学.基于马尔可夫的行人行走行为预测方法:CN201710895401.1[P].2018-02-02.

[37] Zhang R , Li F , Zhou J , et al. A Review on Pedestrian Crossing Detection and Behavior Analysis: In Intelligent Transportation System[J]. Recent Patents on Computer Science, 2015, 8(1):2-12.

[38] 闫旭, 曹一夫. On Application of Path Selection and Pedestrian Traffic Analysis in Public Transportation Hubs%路径选择与行人流量分析-在公共交通枢纽中的应用[J]. 河北建筑工程学院学报, 2019, 037(002):65-69.

[39] Ren G , Zhou Z , Wang W , et al. Crossing Behaviors of Pedestrians at Signalized Intersections : Observational Study and Survey in China[J]. Transportation Research Record, 2011, 2264(-1):65-73.

[40] 杨燕, 孙剑, 李克平. 行人违章过街行为心理成因分析及预测[J]. 中国安全科学学报, 2011(04):22-27.

[41] Hinton, G, E, et al. Reducing the Dimensionality of Data with Neural Networks.[J]. Science, 2006.

[42] 陈硕. 深度学习神经网络在语音识别中的应用研究[D]. 广州:华南理工大学, 2013.

[43] Mcculloch W S , Pitts W . A Logical Calculus of the Ideas Immanent in Nervous Activity[J]. Bulletin of Mathematical Biology, 1943, 52(1–2):99-115.

[44] Ranzato M, Mnih V, Susskind J M, et al. Modeling natural images using gated MRFs[J]. IEEE Trans Pattern Anal Mach Intell. 2013, 35(9): 2206-2222.

[45] Klinker M , Marante C , Argenti L , et al. Electron Correlation in the Ionization Continuum of Molecules: Photoionization of N 2 in the Vicinity of the Hopfield Series of Autoionizing States[J]. Journal of Physical Chemistry Letters, 2018:acs.jpclett.7b03220.

[46] Fukushima K . Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological Cybernetics, 1980, 36(4):193-202.

[47] Rumelhart D E , Hinton G E , Williams R J . Learning Representations by Back Propagating Errors[J]. Nature, 1986, 323(6088):533-536.

[48] Cybenko G . Approximation by superpositions of a sigmoidal function[J]. Mathematics of Control Signals & Systems, 1989, 2(4):303-314.

[49] Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural computation, 1997, 9(8):1735-1780.

[50] Hinton G E , Osindero S , Teh Y W . A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527-1554.

[51] Glorot X , Bengio Y . Understanding the difficulty of training deep feedforward neural networks[J]. Journal of Machine Learning Research, 2010, 9:249-256.

[52] Glorot X , Bordes A , Bengio Y . Deep Sparse Rectifier Neural Networks[J]. Journal of Machine Learning Research, 2011, 15:315-323.

[53] Wang S H , Xie S , Chen X , et al. Alcoholism Identification Based on an AlexNet Transfer Learning Model[J]. Frontiers in Psychiatry, 2019, 10.

[54] Girshick R . Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2016.

[55] Han J , Moraga C . The Influence of the Sigmoid Function Parameters on the Speed of Backpropagation Learning[C]// International Workshop on from Natural to Artificial Neural Computation. Springer-Verlag, 1995.

[56] Malfliet W , Hereman W . The tanh method: I. Exact solutions of nonlinear evolution and wave equations[J]. Physica Scripta, 1996, 54(6):563-568.

[57] Hahnloser R H R , Sarpeshkar R , Mahowald M A , et al. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit[J]. Nature, 2000, 405(6789):947-951.

[58] 周晓东. 深度学习分类网络研究及其在计算机视觉中的应用[D]. 河南:郑州大学, 2018.

[59] Boureau Y L , Bach F , Lecun Y , et al. Learning mid-level features for recognition[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010.

[60] Shi W , Caballero J , Ferenc Huszár, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient, Sub-Pixel Convolutional Neural Network[C]// CVPR 2016. IEEE, 2016.

[61] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer ence, 2014.

致 谢

硕士生活即将结束，两年的时间转瞬即逝。在这两年里，我不论是在生活上还是在学习上都收获了很多，也成长了很多。回首过去的两年时光，我有很多的感谢要说。

首先，感谢我的导师祝老师在研究生期间对我认真负责的指导，祝老师从一开始就详细了解每一位学生的个人情况，针对每个人的情况以及个人的需求制定合理的研究课题。祝老师一直都教导我要摆脱本科时候的学习方法，不断激发自己的创新思维，将书本知识与自己的想法进行结合。学习上有什么不懂得问题，祝老师也会积极为我解答，耐心为我寻找与课题相关的资料。在研究生期间，老师一直都教导我们要努力认真，不断地激励鼓舞我们前进。不仅在学习上对我们很关心，在生活上祝老师也展现了细致的关心。祝老师一直以来认真负责的工作作风也为我树立了榜样，指导我在以后的生活中工作中都努力严谨。

其次，还要感谢我的同学和舍友在两年研究生生活中的陪伴。生活中总会有不快，他们会在我失意的时候，为我开解，陪伴我左右，帮助我一起解决问题。使得我在研究生期间过的更加充实开心，收获了美好的友谊。我还要感谢我的家人。在学习的过程中一直做我坚强的后盾，给予我源源不断的力量。让我一直感觉踏实，无论何时，他们都会在我背后支持我。

最后，感谢辛苦评阅我论文的各位老师专家以及论文中涉及的参考文献作者。

大连理工大学学位论文版权使用授权书

本人完全了解学校有关学位论文知识产权的规定，在校攻读学位期间论文工作的知识产权属于大连理工大学，允许论文被查阅和借阅。学校有权保留论文并向国家有关部门或机构送交论文的复印件和电子版，可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

学位论文题目：

作 者 签 名 ： 日期： 年 月 日

导 师 签 名 ： 日期： 年 月 日