专 业 学 位 硕 士 学 位 论 文

基于深度学习的汽车车型识别关键问题研究

Some Key Issues of Vehicle Model Recognition Based on Deep Learning

作 者 姓 名：吴 双 敬

工 程 领 域：车辆工程

学　　号：31703169

指 导 教 师：李宝军　副教授

完 成 日 期：2019 年 6 月

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：基于深度学习的汽车车型识别关键问题研究

作 者 签 名：吴双承　　日期：2019 年 6 月 14 日

摘　要

随着深度学习等人工智能技术的迅猛进步，智能交通系统取得突破性进展。汽车车型识别（Vehicle Model Recognition）作为其最基础的一环，在智能交通、智能安防以及智能收费等若干方面发挥重要作用。汽车车型的精确识别不同于汽车品牌、类型等分类问题，属于细粒度分类范畴，如何提升识别精度和提升效率是亟待解决的难题。本文基于深度学习方法，研究了数据集属性对车型识别精度及效率的影响，提出汽车车型识别训练集角度紧致性的概念，并创建了针对中国市场的 MTV-1638s 数据集。在此基础上基于 ResNet-50 提出多阶段学习网络 MS-CNN 车型识别方法，数值实验表明该算法可进行高效车型识别。

具体研究内容如下：

（1)分析并提出了汽车车型识别训练集的角度紧致性及创建对应车型数据库。首先，针对中国汽车市场车型识别问题，创建了一个更为合理的汽车数据集 MTV-1638s，包括了中国市场常见的 1638 款车型。通过分类分析发现，训练库中单一车型样本数量及角度对识别效果影响较大。为此，本文创建了一个均匀角度采样车型数据集 ASMTV120s，在此基础上研究了汽车角度分布对识别率的影响，实验表明：训练集中单车型仅包含不多于 18 个指定角度的样本图片即可达到同等识别效果，并可减少 41.18%的训练时长。

（2）提出了加强高阶信息的多阶段神经网络 MS-CNN 车型识别方法。汽车车型识别存在类内差距大且类间差距小的问题，若能充分利用富含语义信息的高层网络特征，对于车型识别将大有裨益。本文在 ResNet-50 基础上，提出一个加强高阶信息的多阶段神经网络 MS-CNN。数值实验证明，该网络能提取到更丰富的语义特征与空间信息，与 VGG-16、ResNet-50 及 B-CNN 相比能得到更高的识别精度。

关键词：车型识别；深度学习；数据集属性；紧致性；多阶段神经网络

Some Key Issues of Vehicle Model Recognition Based on Deep Learning

Abstract

With the rapid progress of artificial intelligence technologies such as deep learning, the intelligent transportation system (ITS) has made breakthroughs. As the one of the most basic parts of the ITS, Vehicle Model Recognition (VMR) plays an important role in intelligent transportation, intelligent security and intelligent charging, et al. The accurate recognition of vehicle models is different from the classification of car brands and types, and it belongs to the category of fine-grained classification. Thus how to improve recognition accuracy and efficiency is an urgent problem to be solved. Based on the deep learning method, this paper studies the influence of dataset attributes on vehicle identification accuracy and efficiency, then we propose the concept of angle compactness of vehicle model recognition training set, and create the MTV-1638s dataset for the Chinese market. Based on ResNet-50, the multi-stage learning network MS-CNN model identification method is proposed. Numerical experiments show that the algorithm can obtain better identify results.

The research contents are as follows:

(1) Analyze and propose the angle compactness of the vehicle model recognition training set and create a corresponding VMR database. First of all, for the identification of the Chinese market vehicle models, a more reasonable car dataset MTV-1638s was created, including the 1638 models commonly found in the Chinese market. Through classification analysis, we found that the training set number and angle of samples of a single vehicle in the training set have a great influence on the recognition effect. To this end, this paper creates an average angle sampling model data set ASMTV120s, on the basis of which the influence of the vehicle angle distribution on the recognition rate is studied. The experiment shows that the single model of the training set contains only sample images of no more than 18 specified angles. Using our result, the VMR method can achieve the same recognition result and can reduce the training time by 41.18% as well.

(2) A multi-stage neural network MS-CNN model identification method for enhancing high-order information is proposed. There is a big gap between the inner-class and the small gap between inter-class. If it is possible to make full use of the high-level network features which rich in semantic information, it will be of great benefit for vehicle identification. Based on ResNet-50, this paper proposes a multi-stage neural network MS-CNN that enhances high-order information for vehicle model recognition. Numerical experiments show that the

network can extract richer semantic features and geometric information, and can obtain higher recognition accuracy than VGG-16, ResNet-50 and B-CNN, et al.

Key Words：Model Recognition; Deep Learning; Car Dataset Properties; Data Compactness; Neural Network Algorithm

目录

1 绪论

1.1 研究背景及意义

汽车数量的急剧增加，引发了交通堵塞、交通事故等问题。解决这些问题若靠人工，不仅效率低，且耗费不必要的人力与物力。因此，通过智能交通系统（Intelligent Traffic System, ITS）进行车辆的自动识别是迫在眉睫的任务。

智能交通系统可以协调交通系统的各组成部分，改善城市的交通环境，智能交通系统能实现对城市交通信息的高效采集，从而提升对突发交通事故的预判能力，能有效避免交通事故的发生，或在交通事故发生后，第一时间发现并及时做出事故处理。

由于大数据以及计算机硬件的支持，智能交通系统在过去十年一直不断发展，道路车辆的分类识别是智能交通系统中几种应用的跳板。包括交通引导，指挥调度，行车违法检测等，车辆的检测和分类，特别是汽车品牌与汽车车型识别，是对车辆的精细化研究，如图 1.1 所示。汽车品牌与车型识别得到的数据信息能促使智能交通系统能更好地服务城市交通，对“智慧城市”的建设与发展做出了突出贡献。

汽车车型
雪铁龙DS5

汽车车型
雪铁龙DS4

汽车车型
雪铁龙DS3

图 1.1 汽车车型识别示例

Fig. 1.1 Vehicle model recognition

近年来，自动驾驶技术的发展对车型识别也提出了很高要求，自动驾驶系统有三个重要的组成部分，分别为环境感知、决策与控制。环境感知作为第一环，影响甚至决定自动驾驶的后面两个环节。环境感知主要依靠雷达技术与图像识别，两者相辅相成作为汽车的“眼睛”，但是雷达存在价格昂贵，例如某些激光雷达，容易受到天气的影响的缺点。随着摄像头技术以及计算机视觉技术的发展，使得图像识别广泛应用于自动驾驶系统。在自动驾驶领域，通常使用的方法是利用摄像头采集信息，然后将采集到的信息传入计算机视觉系统，最后对采集到的信息进行处理。

除此之外，汽车车型识别技术在自动收费系统、交通统计（车辆数量、速度和流量）以及智能安防中也有广泛应用，在停车场、高速路卡口等场合，车型识别能实现根据车辆的品牌、型号进行自动收费。早前的车型识别技术对外界设施的要求很高，如需求较复杂的监控设备、对监控环境、拍摄距离等有严格要求，使得该技术的应用受到制约。如今由于智能化系统（ITS）的需求以及计算机智能领域大规模发展，研究人员已经将相关技术（如深度学习）应用到车型识别当中，使得车型识别技术不再对识别环境、识别角度等各方面都存在严格限制，同时，车型识别技术可以达到的效果也逐渐提高，不再局限于对汽车品牌或者对车辆类型（轿车、客车、货车等）的识别，而是达到细粒度的汽车精确识别，以提供更详细的信息。

汽车车型识别还可以用来辅助公安部门追踪违规车辆，维护社会治安以及行人安全。在此之前，车辆识别最常见和最古老的方法是在车牌识别（LPR）系统的帮助下捕获车辆的牌照，并通过在可用数据库中搜索牌照，然后找到分配给牌照的品牌和型号来识别车辆型号。车牌识别技术[1-4]已十分成熟。但是，应用这种方法的首先要求访问国家车辆信息数据库。此外，LPR 系统不是完全准确的，并且它们有时因性能不稳定而出错。若要改善 LPR 系统的性能，需要相当高分辨率的相机设备来应对诸如阳光反射，变化的天气和光线条件差等挑战，这使得系统操作昂贵。基于深度学习的汽车精确识别提高了车辆识别准确度，且鲁棒性强，对智能安防做出突出贡献。

1.2 相关工作

1.2.1 汽车车型识别（VMR）

互联网、云计算、大数据技术以及汽车类相关数据集的收集，促进了汽车相关任务：车辆再识别[5-8]、车辆检测[9-13]、以及车型识别[14-17]算法的发展，车型识别作为其中一个重要分支，已经有二十多年的研究历史，方法从模式识别例如 HOG[18]+SVM，演变至现在的利用深度学习方法进行识别。深度学习方法能保证在复杂环境下也具有稳定的识别效果，且能保有较好的鲁棒性，基于深度学习，有以下两大类方法可应用于车型识别问题： 一是利用深度神经网络，二是与细粒度分类算法相关的一些汽车车型识别算法。而神经网络算法的发展得益于相关汽车数据库的收集，前人已经对上述内容做了大量工作，下面将对前人工作以及该研究方向的最新进展做详细介绍。

汽车车型识别最基本的方法之一是利用神经网络，神经网络结构示意图如图 1.2 所示，利用神经网络的识别的大致流程是卷积神经网络卷积层（Convolutional layer）提取特征，低层网络提取的特征进行整合抽象输入到高层进行进一步运算，卷积层之间还有池化层进行池化降维，最后全连接层（fully-connected layer）对卷积池化得到的局部信

息进行整合，再由分类层（softmax layer）进行分类。最初神经网络用于解决分类问题是基于 AlexKrizhevsky[19]提出的模型，在此基础上神经网络不断进步，网络结构不断改进与优化，包括小卷积核的提出，利用 3*3 卷积，1*1 卷积等减少参数，此外针对网络训练的一些训练技巧，旨在减少训练误差以及提高模型的泛化能力，提出了一些正则化方法，数据增强、Dropout、对抗训练等。对模型不断优化，采用不同的参数初始化以及不同的深度学习优化算法，力求提高模型的分类识别精度。

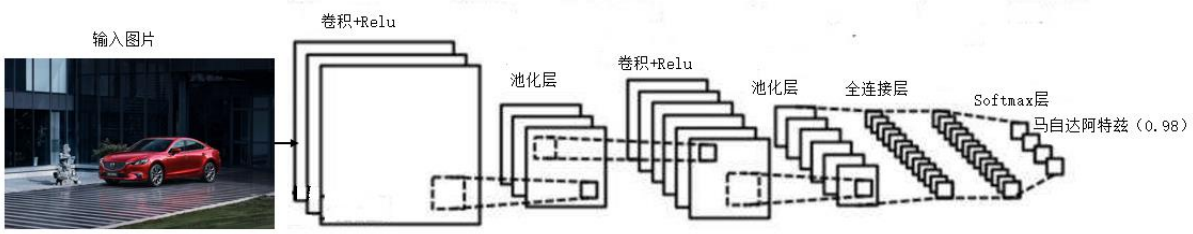


图 1.2　神经网络结构示意图

Fig. 1.2　The structure of convolutional neural network

相对于其它分类而言，汽车车型识别有其独特性，汽车的更新换代比较快，为适应市场发展以及迎合消费者需求，汽车造型也千变万化，这给车型识别带来很大挑战，不同的汽车车型之间存在类内差距大，类间差距小的问题，这不同于对汽车品牌与汽车类型的分类，该问题属于细粒度分类问题，近年来，细粒度分类算法得到迅速发展，并且也有越来越多的算法可以迁移应用在汽车数据集上，以下给出了在近几年与细粒度分类相关算法以及部分车型识别算法的研究进展。

对于细粒度分类，其中比较经典的模型是由 Lin 等[20]在文献中提出了一种简单、高效的神经网络结构：双线性卷积神经网络（Bilinear Convolutional Neural Networks, B-CNNs）。此网络使用两个卷积神经网络分别提取出图片中的特征，去掉网络最后的全连接层，经双线性池化后直接输入到分类器。在最后一层卷积层得到的特征图上进行池化操作，将最后一层卷积得到的特征用外积池化的方法进行信息整合。本此方法在汽车数据集 Stanford Cars 上，获得了 91.3%的 Top-1 识别率。

解决车型识别等细粒度分类问题，可充分利用部件级信息，Zhang 等[21]在中首先提出利用部件级信息进行细粒度分类，文中利用选择性搜索的方法提取候选框，找到部件的边界框与物体边界框。之后对边界框内的特征利用神经网络进行提取。文中使用的部件信息已经进行标注，此方法的技术创新是加入了部件级与物体之间的几何约束，此方法提高了最终分类精度。Zhang 等[17]提出一个新的深度学习网络，用不带有类别和部位

标注的图片集训练出具有判别特性的中级特征。网络首先用 RCNN 检测物体的局部特征（文中指鸟的头部）与整体（鸟的身体），并用几何关系，约束局部与整体的关系。为了在没有对象边界框的情况下从背景分割前景。文中计算置信度图，最终的对齐是在检测的局部特征以及分割前景的基础上完成的。此方法学习到一系列中级特征，与低级特征相比，具有维度低，并且对异常值具有鲁棒性等优点。Fang 等[22]在文中使用一种新的定位汽车关键区域的方法，提出一种由粗到细的方法定位关键区域，其中这些关键区域通过卷积神经网络的特征图自动检测出来。全局特征以及局部特征就是从汽车整张图片以及定位到的关键区域中提取，文中对 281 种汽车车型进行分类识别，达到了 98.29%的识别精度。本文也有一定的局限性，文中采用的数据集是基于汽车前视图进行识别。未对本文提出的多角度汽车识别作深入研究。

此外，还可利用多阶段学习的方式，Dai[23]所在的团队提出了一种新的多任务学习的深度网络结构，这个网络可以实现端到端训练部件定位和细粒度分类两个子网络，两个任务相互加强，部件定位任务可以加强分类效果。他们的多任务学习框架在汽车数据集 Stanford Cars 上取得了非常好的成果，达到了最高 93.1%的细粒度分类识别率。Fu 等[14]提出了一种网络结构 RA-CNN（recurrent attention convolutional neural network），这个网络包含 3 个结构相同的子网络，从 scale1 至 scale3 是逐渐放大局部特征。每个子网络都分为分类网络和 APN（attention proposal sub-network）网络两个部分，scale1 中 APN 网络的输出作为 scale2 的输入。在汽车数据集 Stanford Cars 上得到了最高 92.5%的分类正确率。Peng[15]等人提出了一种基于弱监督的物体-部分注意力（OPAM）模型。此网络模型避免了部件标注并且能实现端到端的训练，网络有两部分组成，分别为 object level 与 part level。最终对两部分损失的加权函数进行优化，此方法在汽车数据集 Stanford Cars 上得到了最高 92.19%的分类正确率。Kuang 等[16]提出一个多层次深度学习方法用于大规模的视觉识别。多层式是指标签分层，训练两个网络，训练多个分类器，通过最终的损失目标函数加强分类器之间的联系，有助于提高识别精度并能有效提高计算速度。Wang 等[24]采用一个非对称双流结构，其中一个分支用于对输入的整张图片识别分类，另一分支定位有利于分类正确的敏感区域，在定位分支上还有一个侧支，为了定位准确提出的，在每个分支最后都有一个损失层，计算最终损失，最后通过联合训练得到的模型，在 Stanford Cars 上面取得了 93.8%的识别精度。Zhang[25]所在的团队提出了一种嵌入了细粒度标签结构的多任务学习框架，此网络引入两个损失函数，softmax loss 与 triplet loss，最后对两个损失函数加权求和。网络训练的目标是降低目标函数大小。此方法中的 triplet loss 的计算方法属于度量学习，可以简单的理解为缩小相同类别之间的距

离而扩大不同类别之间的距离，但是度量学习的收敛速度较慢。他们的方法在 Car-333 数据集上获得了最高 89.4%的车型识别精度。

对网络结构进行修改：Zhou 和 Lin[26]在文献中提出一种双偶图（BGL），文中提出高细粒度的标签下不同类别的对象相互之间是存在联系的，而他们提出的双偶标签就是用来挖掘此类关系，模型修改了 softmax 与全连接层，最终在 Stanford Cars 和 Car-333 数据集上都获得了最高 90.5%的分类正确率。Ghassemi[27]等人利用基于 Resnet 提出一种新的网络结构，在网络中加入空间变换网络，新的网络架构能定位到多尺度汽车的敏感区域。避免了人工标记带来的劳动损耗。文中提出的方法在 Stanford Cars 上 top1 的识别精度达到了 83.36%。2019 年，Ghassemi 等[28]提出另一网络，此网络包含两部分，一部分为定位网络，应用定位具有丰富语义信息的多尺度注意力窗口。另一部分为分类网络，分类网络基于宽残差网络，代替标准残差网络，文中提出的联合汽车品牌与车型的损失函数，允许模型使用一个分类器联合预测品牌与车型，而不是每个任务分别训练。

1.2.2 汽车车型数据集

汽车识别算法的进步，得益于汽车相关数据集的收集。深度学习算法对数据有较大的依赖，数据集的好坏对识别效果有很大影响。以下给出几个深度学习常用汽车数据集。

Stanford Cars（Car-196）[29]是 2013 年被美国斯坦福大学提出的一个车辆数据集，该数据集包含 197 个车型的 16185 张汽车图片，车型按照制造商、型号、年份进行分类，如 Tesla Model S，2012 和 BMW M3 coupe，2012。数据集中汽车图片为整车图片，拍摄角度各不相同。图片来源有汽车销售商、网络等。此数据集的每个车型都有各自标签，且每张图片都已经对汽车做好包围盒（bounding box）标注。

CompCars[30]数据集在 2015 年被提出，该数据集图片来自于网络以及监控摄像头下的图片，数据集中包含 163 个汽车制造商的 1716 个汽车车型，总共 214345 张图片，其中网络来源的图片有 136726 张整车图片以及 27618 张汽车零件图，同时整车图片还带有边框和视角标记，网络来源图片有各种不同的视角，监控摄像头下的图片均为汽车前脸图。数据集还带有其他信息，包括最大速度、车门数量、座椅数量等。

Car-333[31]是 2015 年被整理的一个数据集，这个数据集中的图片均源自网络收集，包含 333 个车型类别的 157023 张汽车图片可以作为训练数据，另外还有 7840 张测试图片，全部图片都是在自然情况下正常拍摄的整车图片，图片包含各个角度。作者手动标注了所有图片的所属制造商、车型、年份信息，但没有标注边界框。

VehicleID[32]数据集是一个在 2016 年被提出的大规模数据集，此数据集内接近一半图片进行了汽车车型标注。VehicleID 数据集由中国某市中多个不同的监控摄像头拍摄

得到，总共包含对应 26267 辆车的 221763 张图片，视角均为前方或后方。数据集是针对车辆再识别、车辆检索等任务整理的，因此所有汽车图片都进行了汽车身份相关的标注。数据集中被标注上车型标签的有 90196 张图片，可供进行车型相关的研究。

BoxCars116k[33]数据集是由一个捷克的研究团队整理的大型车辆数据集，此数据集是 BoxCars21k[34]数据集的一个扩展，总共包含关于 27496 辆汽车的 116286 张图片，他们分别属于 693 个汽车车型，以及 45 个汽车制造商。这个数据集中的图片均由监控摄像头真实拍摄，包括前、后、左、右等各种不同视角，数据集的车型数据不仅包括制造商和车型，还包括车型下的子车型与上市年份。

VMMRdb (The Vehicle Make and Model Recognition dataset) [35]是一个 2017 年被提出的大型车辆数据集，该数据集在目前所有车辆数据集中车型种类、图片数量都是最多的。VMMRdb 数据集中含有 9170 个覆盖了 1950 年至 2016 年之间制造的车型（分类为车辆制造商、车型、年份），总共 291752 张图片。这些图片由多个拍摄人使用各种不同的设备从各种不同角度进行拍摄，数据集的图片也没有经过例如将汽车对齐、去除背景等操作的统一处理，保证了数据集中图片的场景范围能够更好地还原现实。本数据集中的每张图片都标注出它的汽车品牌、车型已经年份信息，但是由于此数据集中的一些汽车已经停产，因此降低了此数据集的实用性。

1.3 本文工作

由于汽车数据集规模庞大，涉及的汽车车型种类繁多，汽车的拍摄角度以及汽车拍摄环境复杂，这些都给汽车车型识别带来很大挑战，在网络训练过程中，还存在训练时间较长，训练不彻底等问题，使得误差累计，上述存在的问题都导致网络模型最终的识别精度低，汽车车型识别不准确。为此，本文的研究内容主要分为以下两个方面：

首先，在 compcars 数据集的子集 MTV-Cars 数据集进行实验，实验发现该数据集还存在一些问题，从而导致车型识别出错，该数据集的部分车型划分不准确，存在同种车型带有不同类别标签，且车型划分标准不统一，如将同种车型的两厢车与三厢车划分为不同车型等，此外该数据集的车型年份大都在 2015 年之前，降低了该数据集的时效性，因此本文在 compcars 数据集基础上，对数据集进行重建，新增加部分汽车图片，构成新的数据集 MTV-1638s，该数据集包含 1638 个汽车车型，总计 140801 张汽车图片。另外，本文在 MTV-Cars 上出对汽车数据集属性进行研究，主要对数据集的规模数量与汽车角度对车型识别精度的关系进行探究，研究发现单个车型的汽车图片数量与汽车角度对车型识别影响较大，但是由于该数据集单个车型图片数量与角度分布没有规律性，因而结果存在随机性，为此，本文创建另一个均匀采样多角度数据集 ASMTV120s，该数

据集包含 120 个车型，6000 张汽车图片，利用该数据集对汽车角度对车型识别的影响度以及数据集紧致性进行研究，数值实验证明，在使用不多于不多于 18 个指定角度的样本图片即可达到同等识别效果，并可减少 41.18%的训练时长。

另一方面，本文将解决的汽车车型精确识别问题，是针对大规模、多角度、背景复杂的汽车分类问题。汽车造型的复杂多样以及汽车拍摄角度与不同的光照条件等都给识别造成很大困难，如何提高汽车车型的识别精度是本文要解决的另一难题。为解决该问题，本文旨在提出一个新的算法，为此，本文对几个经典的深度神经网络进行实验，从而找到一个改进优化的基础模型，本文对 VGG-16、ResNet-50 以及 B-CNN 进行实验，实验证明，ResNet-50 具有更好的识别效果，因此本文在 ResNet-50 的基础上提出了一个新的多阶段神经网络（MS-CNN），此网络能实现端到端训练，浅层网络间实现卷积层共享，深层网络则采用多阶段学习的方式，有效利用网络的高阶特征，提高网络的表达能力，使用空间金字塔池化、1*1 卷积层与全局平均池化提取特征，通过可视化证实本文使用的网络能提取到更丰富的语义特征，与表现能力最好的 ResNet-50 相比能取得相对较高的识别精度。本文的技术路线图如图 1.3 所示。

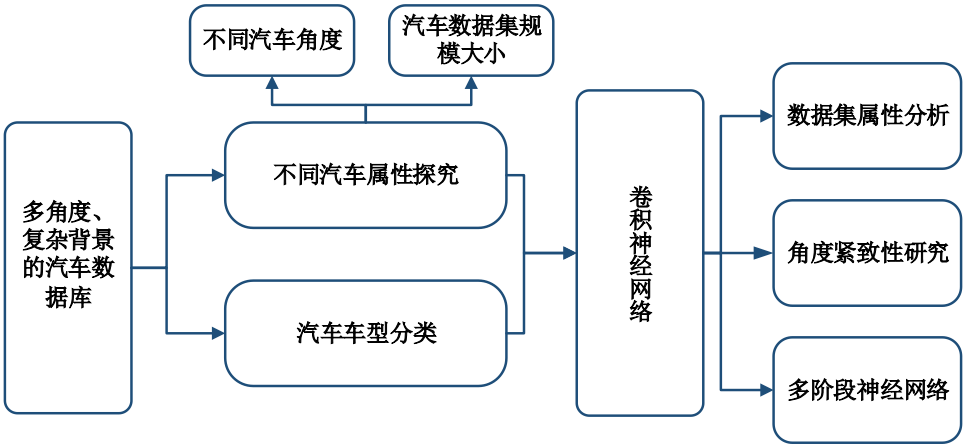


图 1.3　本文技术路线图

Fig. 1.3　Pipeline of our work

本文的章节安排如下：

第一章主要论述了本文的研究背景和意义。在智能交通飞速发展的大背景，以大数据与计算机硬件的发展为依托，衍生了本文的研究，证实了本文研究的汽车车型的精确识别问题，具有很大的现实应用价值，根据现在的研究现状与不足，验证了本文研究的可行，同时明确了本文研究的研究方向。

第二章对现有的汽车数据集进行梳理与对比，在 MTV-Cars 数据集进行实验，分析实验结果发现该数据集存在很多不足，该数据集存在数据划分不精确以及数据的“时效性”较差等问题，为了提高数据集的实用性，本文提出了一个新的数据集 MTV-1638s。此外，收集一个均匀采样多角度汽车数据集 ASMTV120s，在此数据集上对汽车的属性进行数值实验，包括汽车角度影响度与数据集紧致性进行实验探究，总结出一个构建紧致性汽车数据集的结论，有效地降低计算成本。

第三章综合阐述了本文所运用到的相关理论。包括卷积神经网络中的几个经典神经网络框架，介绍了与汽车车型识别相关的细粒度分类方法。并对几个比较常用的网络模型 VGG-16、ResNet-50 以及 B-CNN 等进行实验，确定了基础网络模型 ResNet-50。在此基础模型上，提出一个新的网络，多阶段神经网络（MS-CNN），因神经网络的深层卷积层中包含丰富的语义信息，若能充分利用深层网络信息，则能在一定程度上提高网络的表达能力。本文受多阶段学习启发，对网络结构进行改进，对浅层网络实现卷积层共享，对深层网络特征通过多阶段学习进行进一步挖掘利用，提高最终车型识别精度。

最后对本文的工作进行总结与展望，为后续工作的开展提供方向。

1.4 本章小结

本章主要介绍了在大数据、计算机计算能力飞速发展的当下，汽车车型精确识别作为其中一个重要分支，它的技术发展与进步对智能交通、自动驾驶、自动收费能都有重大贡献，说明了本文研究的重大意义。本章对神经网络的发展现状以及与汽车车型识别方法进行简单综述，对汽车数据集进行简要介绍，提出本文的研究基础，对本文的研究进行可行性分析，进而提出本文的研究技术路线。

2 汽车车型识别数据库

在深度学习界，被公认的规律是，数据越多，模型的表现就越好。但是数据增加会带来计算时间与计算资源等的成本。若能利用尽可能少的训练样本数达到很好的识别效果，就能有效降低计算成本。同时应该尽可能提高数据集的质量，数据集应具有“时效性”，对数据集进行实时更新，才能更好解决现实问题。为了更好解决以上两个问题，本文建立了两个新的数据库 MTV-1638s 与 ASMTV120s，本章在 Compcars 数据集的子集 MTV-Cars，利用 ResNet-50 网络对影响车型识别的两大数据集属性：训练样本数与汽车角度进行实验研究，并在 ASMTV120Ss 基础上总结出一个构建合理紧致数据集的有用结论。

2.1 汽车数据集

基于深度学习的汽车车型识别，是由数据驱动的，在大数据的推动下，深度学习方法得到发展，汽车类数据集主要用于汽车识别，车辆检测与车辆追踪等任务，也有数据集被作为验证算法的优越性，同样，对于本文的研究也是基于数据集的收集。在深度学习中，数据集的规模以及数据质量对研究结果有很重要影响，以下对本文使用的 Stanford cars、VMMRdb、以及 Compcars 三个数据集进行详细介绍。

2.1.1 Stanford Cars

Stanford cars 数据集是由斯坦福大学收集的数据集，其数据特点如图 2.1 所示，其中包含 197 种汽车车型，车型按照汽车品牌、型号、年份分类，如特斯拉-Model S-2012，覆盖轿车，SUV，轿跑车，敞篷车，皮卡，掀背车和旅行车。此数据及主要从网络收集，图片背景比较复杂，图片从任意角度拍摄而来，数据集中包含了 16185 张整车图片，其中 8144 张用于训练，8041 张图片用于测试。本数据集每张图片都已有各自标签，并且图片已标注包围盒，因而本数据集也可用于车辆检测等任务。

图 2.1 Stanford Cars 数据分布

Fig. 2.1 Data distribution of Stanford Cars

2.1.2 VMMRdb

VMMRdb 数据集是由路易斯维尔大学提出的一个大型数据集，该数据集的数据结构如图 2.2 左图所示，图中圆圈的大小越大，代表此车型下的图片数量越多，整个数据集总共涵盖了 9170 种 1950 年至 2016 年的整车图片，在数据集下有一个子集用来做细粒度分类任务，该子集包含 3036 类 246173 张汽车整车图片，图片分布如图 2.2 右图所示，实验时对相同车型但不同年份的汽车划为一类，同样的，此数据集大部分通过网络爬虫获得，汽车图片角度各不相同。并且为了更贴合实际应用场景，没有对数据进行对齐等处理。此数据集在数据规模上比其它数据集大，但是此数据集中收集的汽车图片年份较早。随着汽车行业的发展，汽车造型也发生了很大变化，年份较早的图片在某种程度上降低了数据集的时效性以及使用价值。

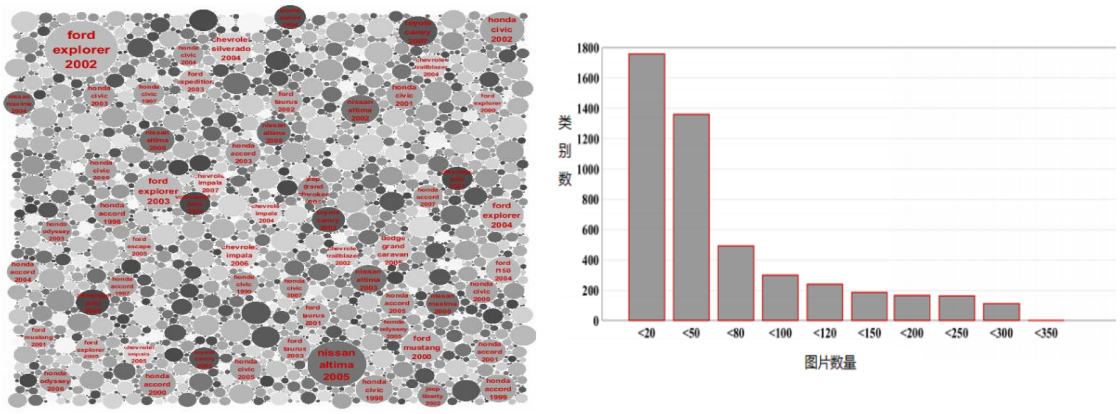


图 2.2 VMMRdb 数据结构与分布

Fig. 2.2 Data structure and distribution of VMMRdb

2.1.3 Compcars

Compcars 数据集是香港大学提出的一个汽车数据集，数据包含三大部分，汽车整车图片、监控录像下的图片以及汽车局部图片。此数据集可以用来做图像识别与图像检索任务。该数据集的整车图片、监控下的图片以及局部图可用于识别任务，在汽车整车图片上利用不同角度、不同汽车属性（座椅数、最大速度、位移等）进行分类识别，利用神经网络进行图像检索[30]。数据集包含三部分汽车图片，第一部分是汽车整车图片，其中包含 136727 张汽车图片，第二部分是汽车局部图，此部分数据包含 27618 张汽车图片，第三部分数据是 5000 张监控录像下的汽车正脸图片，本文应用的汽车整车图片包含了 163 种汽车品牌与 1716 种汽车车型，此部分图片还带有边框与视角标记，包含

了五个视觉信息，前视图、后视图、侧视图、前侧视图、后侧视图。另外还有一些其他信息，包括最大车速、座椅数、车门数量等。数据集的数据结构如图 2.3 所示：

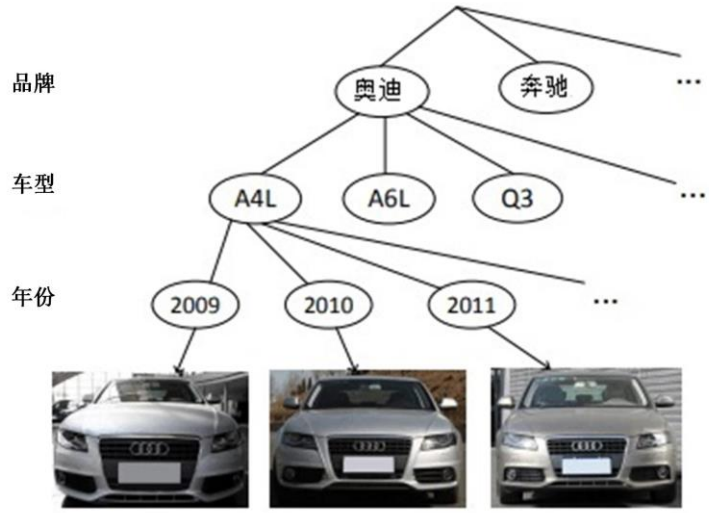


图 2.3　Compcars 数据结构

Fig. 2.3　Data structure of Compcars

2.2　MTV-1638s

怎么判断是否需要收集更多的数据？可以从以下几方面考虑，首先，模型在训练集上变现的性能是否可接受。如果模型在训练集上的表现很差，学习算法都不能在训练集上学习出良好的模型，那么就没必要收集更多的数据。这种情况下，可以选择修改网络，增加网络层或将更多的隐藏单元加入到每层中，以增加模型的规模。同时，也可以尝试调整超参数的大小，例如改变学习率、批尺寸等措施来改进学习算法。如果使用更大的模型和改进优化算法效果还是不好，那么问题可能源自训练数据的质量。数据可能含太多噪声，或者可能不包含测试输出所需的正确输入。这时候可能需要对数据集进行修改，收集更干净的数据或是收集特征更丰富的数据集。

本文通过实验证明，Compcars 数据集存在很多不足，导致最终的实验精度低。Compcars 数据集存在以下几个问题：一是本数据集中对车型的划分标准，因其对同一种车型例如雪铁龙世嘉两厢车与三厢车放置在不同的文件夹下；二是同一个车型被当作不同车型，例如名爵 MG6、上汽大众 Polo 等；三是关于数据的“时效性”问题，由于此数据集是在 2015 年收集的，因此数据集中有很多车型停在 2015 年之前，有很多车型已经停售。针对以上几个问题，对此数据集进行修改，并进一步扩大数据。对 60 个车型进行合并，对一些现实中常见的车型进行扩充，主要根据从一些销售网站的热门车型中

选择该部分汽车图片，并加入近年的一些新车型，其中对 33 个已存车型进行扩充，加入新车型 15 个，新加入车型如表 2.1 所示总计 4074 张汽车整车图片。整理后的汽车数据集 MTV-1638s，总计 140801 张整车图片，共有 1638 个车型，包含最近几年（直到 2018 年）的汽车车型，使得此数据集更具有现实应用价值。

表 2.1　MTV-1638s 新加入车型分布

Tab. 2.1　MTV-1638s new model distribution

品牌	车型	年份
一汽	森雅 R7	2016-2017
吉利	缤瑞	2018
	博越	2016-2018
东风启辰	启辰 D60	2017-2018
	启辰 T70	2015-2018
斯柯达	柯迪亚克	2017-2018
沃尔沃	S90	2017-2018
福特	撼路者	2016-2017
长安汽车	逸动 DT	2018
	凌轩	2017
	睿骋 CC	2018
雪佛兰	迈锐宝 XL	2016-2017
	科沃兹	2016-2018
雪铁龙	C4 世嘉	2016-2018
雷诺	科雷嘉	2015-2017

汽车车型识别的关键问题是在复杂场景以及复杂背景下能准确识别出汽车的品牌与型号，基于深度学习方法的汽车车型识别是通过神经网络来提取特征，最终利用 Softmax 层来分类，在计算损失时是根据预测标签与实际标签进行运算，因此，数据集收集后的整理工作也很重要，MTV-1638s 的数据来源主要来自网路，除了对图片进行车牌遮挡外，没对图片进行任何预处理，图片的尺寸大小不同，背景复杂，一张图片中可能只有一辆汽车，也可能有两辆甚至更多。数据集中只对部分数据进行边框标记，为了贴近现实，未作对齐处理，此外，本数据集对汽车车型进行标注，使用者可以更直接掌握车型信息，对数据集的使用以及之后实验结果的分析都大有裨益。此数据集可以用作车型精确识别、车辆检测与分割等任务。本文对现有的几个数据集进行对比分析，包括表格 2.2 中的七个数据集。

表 2.2　汽车数据集对比分析

Tab. 2.2　The comparision of seven car datasets

数据集	国家	图片数	车型数	图片视角	年份	标签结构	收集途径
Stanford Cars	美国	16185	196	多视角	1990-2013	品牌-车型-年份	互联网
CompCars	中国	214345	1716	多视角	至 2015	品牌-车型-年份	互联网+监控
Car-333	--	136727	333	多视角	--	品牌-车型-年份	互联网
VehicleID	中国	90196	--	多视角	--	--	监控
BoxCars116k	捷克	116286	693	多视角	--	品牌-车型-子车型-年份	监控
VMMRdb	美国	291752	9170	前、后视角	1950-2016	品牌-车型-年份	互联网
MTV-1638s	中国	140801	1638	多视角	至 2019	品牌-车型-年份	互联网

2.3　汽车数据集的属性研究

汽车数据集具有其他数据集没有的属性，因汽车自身的属性有很多是其他物体不具备的，例如 Compcars 数据集中提到的，汽车最大速度、汽车排量、车门数、座椅数等，除此之外还可对车辆颜色进行探究[36-38]，本小节主要对汽车数据集的两大属性，即数据集的数量规模大小以及车辆角度进行探究，探究训练数据大小与汽车角度对汽车车型识别的影响，为此本文在 MTV-Cars 上面进行了数值实验。

谷歌的一项研究显示，在深度模型中，视觉任务性能与训练数据量(取对数)呈线性关系，即如果不断增加数据量，则能不断提升模型的表达能力[39]。数据的增加虽然会带来识别精度的提升，但是要付出的代价也是巨大的，首先，指数级数量的数据集在收集与整理就会耗费很多人力与时间，其次，在收集到大规模的数据集后，要利用网络模型对数据进行训练，在训练过程中对计算资源与计算机计算能力要求也很高。这给实验的实施带来困难，也造成一定程度的资源浪费。若能在小的数据集上模型就能达到一个很好的性能，那不仅满足了识别精度的要求，与此同时还减少了资源的浪费，因此收集一个紧凑的数据集，来完成汽车车型的精确识别任务，具有重要意义。

跟人脸识别[40-43]相同，人脸识别效果很受拍摄角度的影响，同样的，在汽车车型识别中，与现实中人们看到的三维汽车不同，本文的视觉分类任务中，输入网络模型的是静态图片，属于二维信息，这就造成了信息的丢失。通常认为，汽车前脸[44-46]包含更多信息，不同的汽车制造商根据自己品牌的设计理念，对自己的汽车进行造型设计与改进，而不同汽车特别是同一品牌不同车型之间的造型差异，最大程度地体现在汽车前脸造型上，汽车前脸包含汽车大灯、雾灯、进气格栅以及汽车的车标，因此汽车的前脸中蕴含了丰富信息，相比于汽车前脸，汽车侧视图蕴含的信息就相对较少，因汽车侧视图体现的信息不太直观，例如有的汽车为了体现运动性，在设计时，汽车的侧面轮廓相对比较流畅，但是很多汽车车型为了体现这一点都会有这种特征，这给分类识别任务加大了难度。综合汽车的各个角度更有利于识别车辆，这也符合人类识别物体的规律。

本文实验选用的网络模型为 ResNet-50， ResNet-50 具有 50 层卷积层，深度的加深使得该网络能提取到更抽象的特征，具有更强的学习能力，在分类识别任务中有广泛应用。该网络结构将在后文进行详细介绍，这里只做简单说明，不做赘述。

本文实验用到的 MTV-Cars 数据集是 Compcars 数据集的子集，Compcars 数据集包含了 163 种不同的汽车品牌，这其中又包含了 1716 种不同的汽车车型，总计 136726 张整车图片。在实验过程中，为了减少随机性与偶然性，保证实验结果的准确性，在做汽车车型精确识别的实验时，对数据进行过滤。即设定某个值，在图片数量低于设定值时，则剔除此车型，本文设定的阈值为 20，在某一车型图片数量少于 20 张，则忽略该车型。最后本文的实验车型一共有 1448 种汽车车型，所用到的汽车整车图片数目为 133710 张，此部分数据集构成 MTV-Cars 数据集。在该数据集上进行训练时按照 6：2：2 的比例划分将该数据集为训练集、验证集与测试集。

本文从数据规模以及车辆的角度两个方面对数据集属性进行探索，首先对数据集进行随行抽样，在每个车型中按比例减少样本，样本减少的方法可按如下方式进行。

设最终训练集的样本数为 N，每个车型拥有的汽车图片数量为 n，则在原训练集上样本减少的比率（Ratio）为：

$$r = 1 - \frac{n}{N} \tag{2.1}$$

设某一车型的汽车图片数为 i，则在此车型随机删除 $i * r$ 张图片，剩余的图片参与训练。按照上述方法按照 10000 的数量级依次减少训练样本，得到 90000、80000、70000、60000、50000、40000、30000、20000、10000 的训练集，验证集与测试集采用对原始数据训练划分出的数据，因在对数据进行随机删减时是在最初的训练集上进行的，因此现在的数据设置依然保证了训练集、验证集与测试集之间没有交集。表 2.2 是不同规模训

练集对应的识别结果，并在图 2.4 中展示出来，从图中可以看出，在数据从 40000 减少到 30000 时，识别精度发生了突变，本文对 40000 的训练集以及小于 40000 数据集的数据进行进一步分析，结果发现，这部分数据集在数据分布与车辆角度等方面有很大差异。在数据分布上，40000 训练集上，单个车型拥有的图片数目 $i > m$（m 取 16，17，18）的车型分布要比小于 40000 训练集的分布广，相反，单个车型拥有的图片数目 $i < n$（n 取 5，6，7）的车型分布，40000 训练集的车型分布相对较少，如图 2.5 所示。

表 2.2　不同规模汽车图片对应车型识别精度

Tab. 2.2　Vehicle model recognition accuracy corresponding to different data scales

数据集名称	数据集规模	分类精度
MTV-Cars	原始数据 （106970）	94.66%
	90000	93.68%
	80000	91.90%
	70000	91.34%
	60000	89.10%
	50000	86.46%
	40000	82.01%
	30000	74.72%
	20000	62.03%
	10000	34.42%

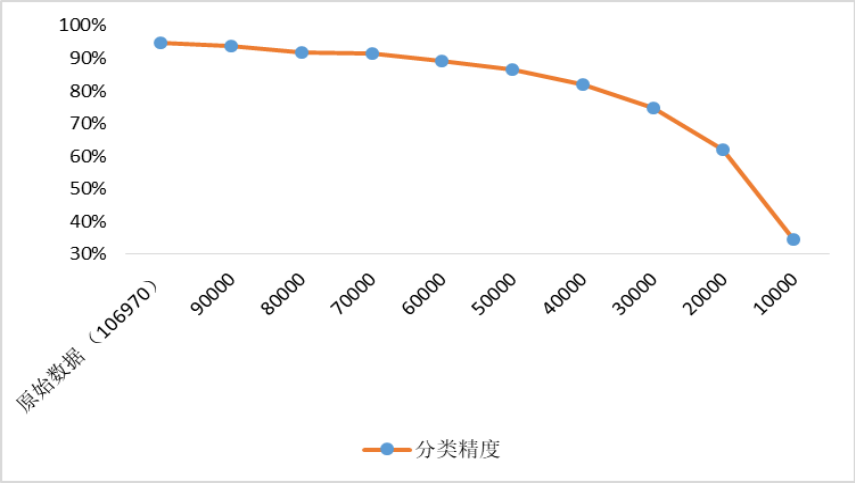


图 2.4　不同数据规模下汽车车型识别精度

Fig. 2.4　Recognition accuracy of car models under different data scales

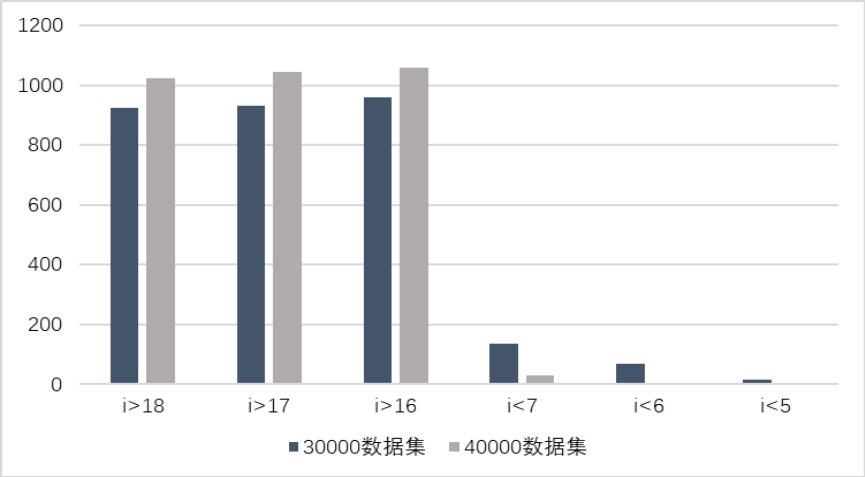


图 2.5　不同车型汽车图片数据分布图

Fig. 2.5　Data distribution of different vehicle models

表 2.3　不同汽车拍摄角度的车型识别精度

Tab. 2.3　Vehicle model recognition of different angels

汽车角度	识别精度
全部角度	94.66%
其它角度	74.68%
前视图	64.00%
其他角度（对比以上）	68.49%

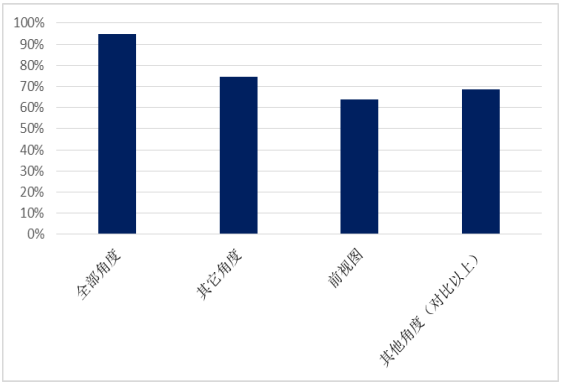


图 2.6　不同汽车拍摄角度车型识别精度对比

Fig. 2.6　Comparison of recognition accuracy between different vehicle angles

除此之外，本文发现汽车角度也会影响最终的识别效果，分析数据集，40000 训练集的汽车前后视图的比例要比 30000 训练集中的大，角度是不是影响识别的重要因素？为了验证以上猜想，本文进行了实验，在 Compcars 数据集中，挑选出前视图（FV）汽车图片，按照之前的数据处理方式，当某个车型前视图汽车图片数目少于 20 张时，则忽略该车型，并对数据集按照 6：2：2 的比例划分为训练集、验证集与测试集。最终训练集、验证集与测试集中图片数分别为 2661 张、865 张、890 张，包含 188 个汽车车型。另外，不包含前视图的其他角度图片按相同比例进行划分，训练集、验证集与测试集中图片数目分别是 69695、23223、23238 张，包含 1412 个汽车车型，若只是用这两组进行对比，还不能说明角度的作用，因为两者在图片数量与种类之间还存在很大差异，于是进行另一组对比实验，把前视图中车型所对应的汽车其他角度图片挑选出

来，此部分训练集、验证集与测试集中分别包含 13820、4627、4599 张汽车图片，同样包含 188 类汽车车型，其实验结果如表 2.3 以及图 2.6 所示：

从表中可以看出，利用各个角度的图片能提供更多有利于识别的信息，它取得了最高的识别精度，而汽车的前视图对最终的车型识别起关键作用，从柱状图的第二条与第三条可以看出，虽然其他角度的图片数量是前视图图片数目的 5.2 倍，但是识别精度只相差 4%左右，下一小结将对角度与数据规模进行更进一步探究。

通过对汽车数据集规模与角度对车型识别精度的影响实验结果分析，本文发现单个车型的汽车样本数量与拍摄角度对识别结果有较大影响。

2.4 数据集紧致性评估方法

训练样本的增加能在一定程度上增大识别率，但是数据增大，会消耗大量计算时间。例如本文在 NVIDIA GeForce GTX 1080 Ti 图形处理单元上用 ResNet-50 训练 Compcars 数据集，花费时长大约为 5 天。如果能利用相对较少的数据上取得相同的实验效果，将大大降低资源消耗。因此减少冗余数据，提出一个紧致性数据集尤为重要。

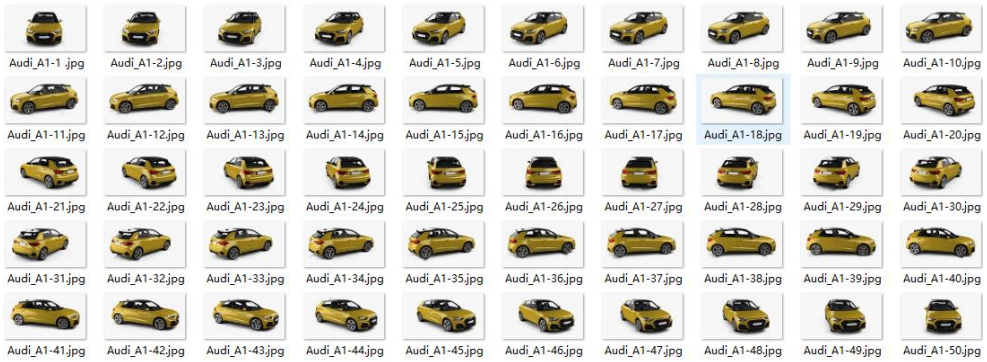


图 2.7　ASMTV120s 示例

Fig. 2.7　Examples of ASMTV120s

在 2.3 小节中，对汽车数据集属性进行实验探究，由于单个车型的样本数与角度分布不均，因此结果具有随机性，为了对数据集进一步评估，本小节将继续研究出一个评估数据集紧致性的方法，主要研究对汽车识别重要的汽车角度，以及在此角度下满足精度要求的图片数量。为此，本文收集了一个均匀采样的数据集 ASMTV120s，数据集包含 6000 张不同角度的汽车图片，其中包含 120 个汽车车型，每个车型下有 50 张汽车图片，且角度均匀分布 。图 2.7 以 Audi A1 为例，展示了数据集中图片角度的分布情况。

为了探索数据集的角度紧致性，在 ASMTV120S 上进行如下实验：

首先，让该数据集的全部角度参与实验，实验数据按照 6：2：2 的比例划分为训练集、验证集与测试集，被划分数据集中的数据随机抽取，该实验车型识别精度为 99.83%。

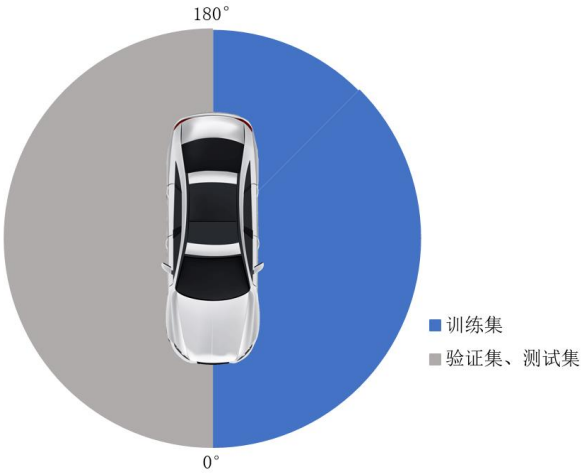


图 2.8　一半数据数据集划分方式

Fig. 2.8　Dataset division for half of training set

该数据集汽车角度中部分角度分布呈对称性，若利用一半数据就可达到全部数据的效果，就可以节约计算资源与计算时间。第二组实验利用一半的数据进行实验，数据划分方式如图 2.8 所示，其中的训练集选用蓝色区域，而验证集与测试集选择灰色区域。

接下来，为了探究角度对汽车车型识别结果的影响，按照经验角度对角度进行划分，角度分界点分别是 0 度、45 度、90 度、135 度与 180 度，进行以下四组实验：

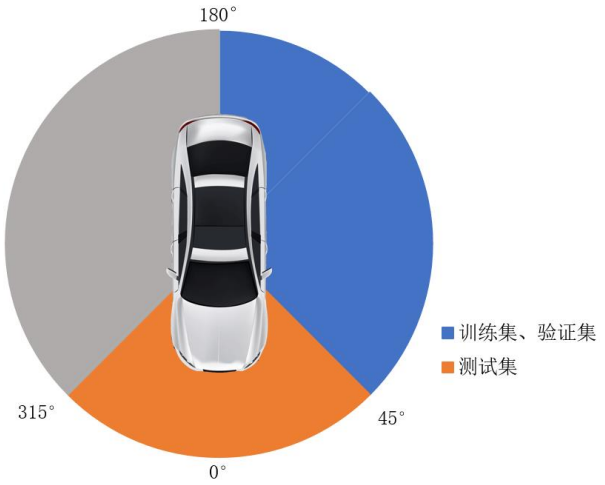


图 2.9　不包含$[0^{o}-45^{o}]$数据集划分

Fig. 2.9　Dataset division for exduclode $[0^{o}-45^{o}]$training set

第一组：在不包含[0°-45°]的数据集进行实验，数据集划分方式如图 2.9 所示，其中随机选取训练集 30%的数据进行验证。利用[0°-45°]及其对称图片进行测试。此时车型识别精度为 78.72%。

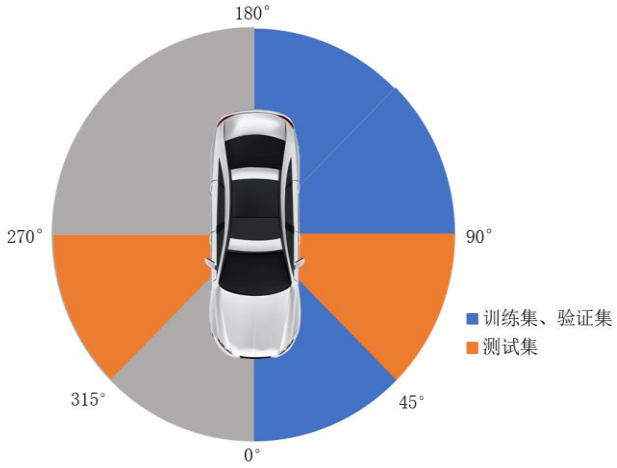


图 2.10　在不包含(45°-90°]数据集划分

Fig. 2.10　Dataset division for exclode (45°-90°] training set

第二组：在不包含(45°-90°]的数据集进行实验，数据集划分方式如图 2.10 所示，同样随机选取训练集 30%的数据进行验证，利用图中及其对称角度图片测试。此时汽车车型识别精度 98.04%。

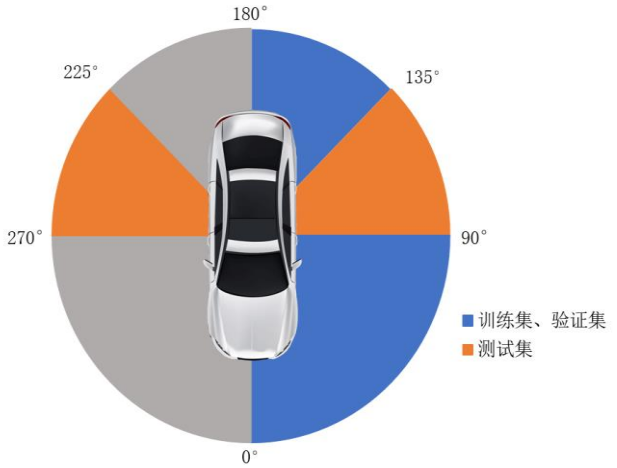


图 2.11　在不包含(90°-135°]数据集划分

Fig. 2.11　Dataset division for exclode (90°-135°] training set

第三组：在不包含 $(90^{o}-135^{o}]$ 的数据集进行实验，数据集划分方式如图 2.11 所示，同样随机选取训练集 30%的数据进行验证，利用图中橙色区域的全部数据做测试。此部分实验的车型识别精度为 99.44%。

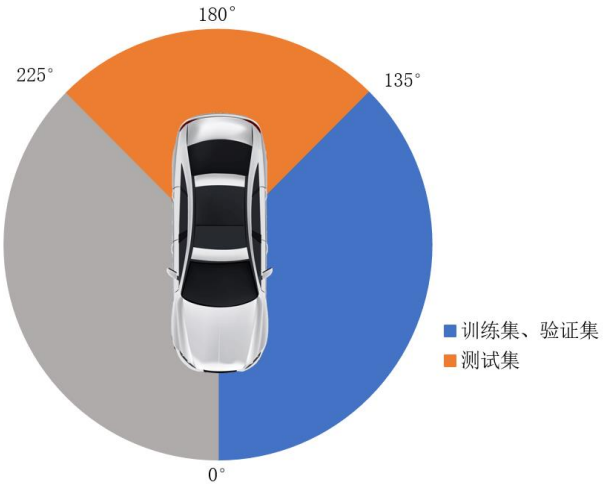


图 2.12　在不包含 $(135^{o}-180^{o}]$ 数据集集划分

Fig. 2.12　Dataset division for exclode $(135^{o}-180^{o}]$ training set

第四组：在不包含 $(135^{o}-180^{o}]$ 的数据集进行实验，数据集划分方式如图 2.12 所示，选取蓝色区域训练集 30%的数据进行验证，利用图中橙色区域的全部数据做测试。此部分实验的车型识别精度为 78.86%。

最后对以上实验进行汇总，结果如表 2.4，直观图如图 2.13 所示：

表 2.4　汽车角度对车型识别精度的影响

Tab. 2.4　Influence of vehicle angle on VMR accuracy

汽车角度	训练集图片数量	验证集图片数量	测试集图片数量	汽车车型识别精度
全部角度	3600	1200	1200	99.83%
$0^{o}-180^{o}$	3120	1030	1030	99.92%
不含 $0^{o}-45^{o}$	2280	680	1560	78.72%
不含 $45^{o}-90^{o}$	2280	680	1680	98.04%
不含 $90^{o}-135^{o}$	2400	720	1440	99.44%
不含 $135^{o}-180^{o}$	2400	720	1320	78.86%

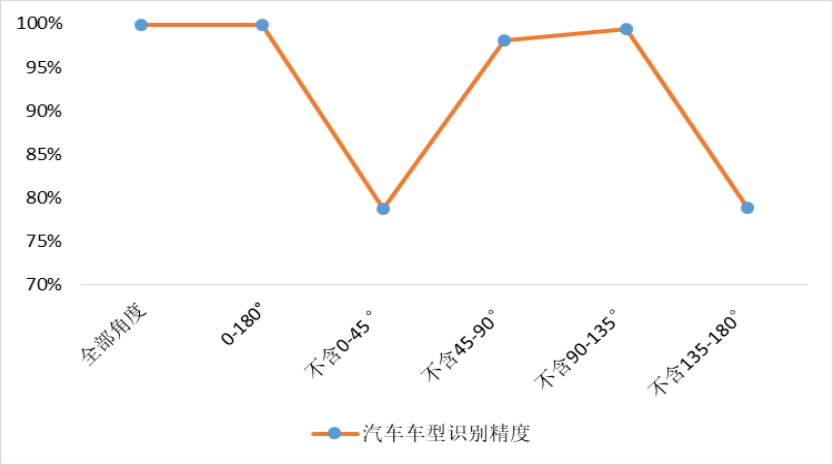


图 2.13　汽车角度对车型识别精度的影响

Fig. 2.13　Influence of vehicle angle on VMR accuracy

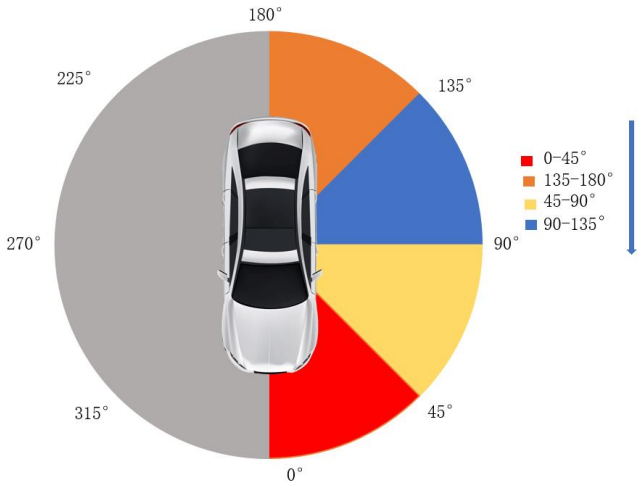


图 2.14　汽车角度影响度图

Fig. 2.14　Influence degree of vehicle angle

分析表 2.4 的结果，可知汽车的前侧视图与后侧视图对汽车车型识别有重要影响，汽车侧视图对车型识别的影响不大。并且比较上述结果发现，$[0^{o}-45^{o}]$之间的汽车拍摄角度最为重要，这或许与汽车前视图包含更丰富的语义信息有关。根据各个角度的重要程度，得到了对于车型识别的汽车角度影响度图如图 2.14 所示。为了进一步压缩数据，提高数据的紧致性，对数据进行间隔抽样，抽样原则是按照角度对车型识别精度的影响度，首先保证汽车的前侧视图与后侧视图的角度，即角度在$[0^{o}-45^{o}]$与$(135^{o}-180^{o}]$的汽车图片数量，然后减少剩余两个区间的图片，进行以下三组实验：

首先 A 组数据包含[0°-45°]与(135°-180°]两个角度区间的图片，此时单个车型包含 13 张汽车图片，数据分布如图 2.15 所示，利用该部分图片作为训练集，剩余图片作为验证集与测试集，训练集、验证集与测试集的划分比例为 6：2：2。

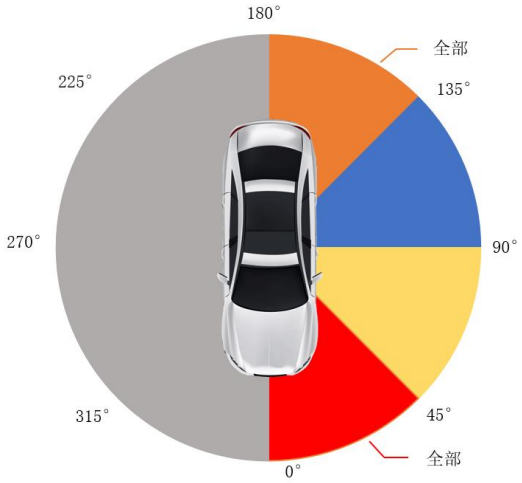


图 2.15　A 组实验数据集划分

Fig. 2.15　Dataset division for group A

B 组数据在 A 组数据的基础上加入[0°-45°]与(135°-180°]的少量图片，此时单个车型参与训练的包含 18 张汽车图片，验证集与测试集的样本从单个车型的剩余汽车图片中抽取，数据分布如图 2.16 所示。

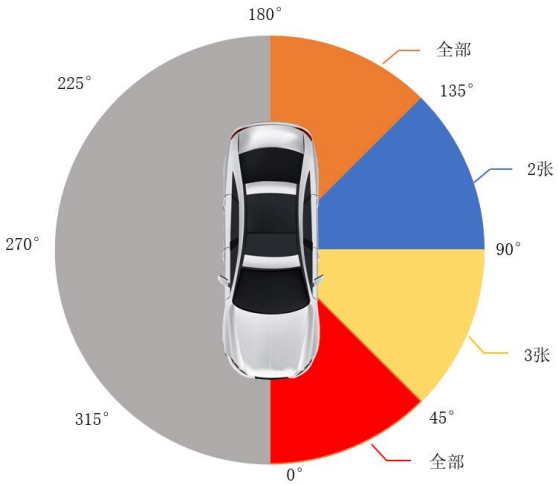


图 2.16　B 组实验数据集划分

Fig. 2.16　Dataset division for group B

C 组数据在 A 组的基础上只增加90º与135º两个经验角度，由于数据集中的角度均匀分布，因而在数据集的每个车型中抽取与以上两个角度最相近的两张汽车图片，此时单个车型包含 15 张汽车图片，B 组与 C 组的数据划分方式与 A 组相同，数据集划分方式如图 2.17 所示。

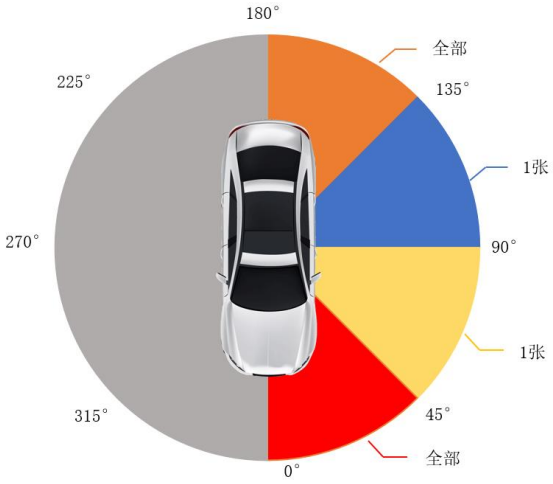


图 2.17　C 组实验数据集划分

Fig. 2.17　Dataset division for group C

实验结果汇总如表 2.5 所示。

表 2.5　数据紧致性评估实验

Tab. 2.5　Evaluation experiment on compact dataset

汽车角度	训练集图片数量	验证集图片数量	测试集图片数量	汽车车型识别精度
A 组	1560	510	510	90.62%
B 组	2160	720	720	99.86%
C 组	1800	600	600	98.78%

通过上述实验，可以发现，对于汽车车型识别来说，不需要全方位采集汽车图片，只需要采集对识别有重要影响的几个角度即可，这几个角度分别是[0º - 45º]与(135º - 180º]角度范围的汽车图片，少量前侧与后侧视图的辅助图片。本文将最初的 50 张汽车图片减少到 18 张汽车图片，18 张汽车图片的分布图如图 2.18 所示，其中在[0º - 45º]角度范围内的汽车图片不多于七张，(45º - 90º]角度范围内的汽车图片不多于 3 张，在(90º - 135º]角度范围内的汽车图片不多于 2 张，(135º - 180º]角度范围内不多于 6

张。训练集中单车型仅包含不多于这 18 个指定角度的样本图片即可达到同等识别效果，这对数据集评估与收集具有指导意义。

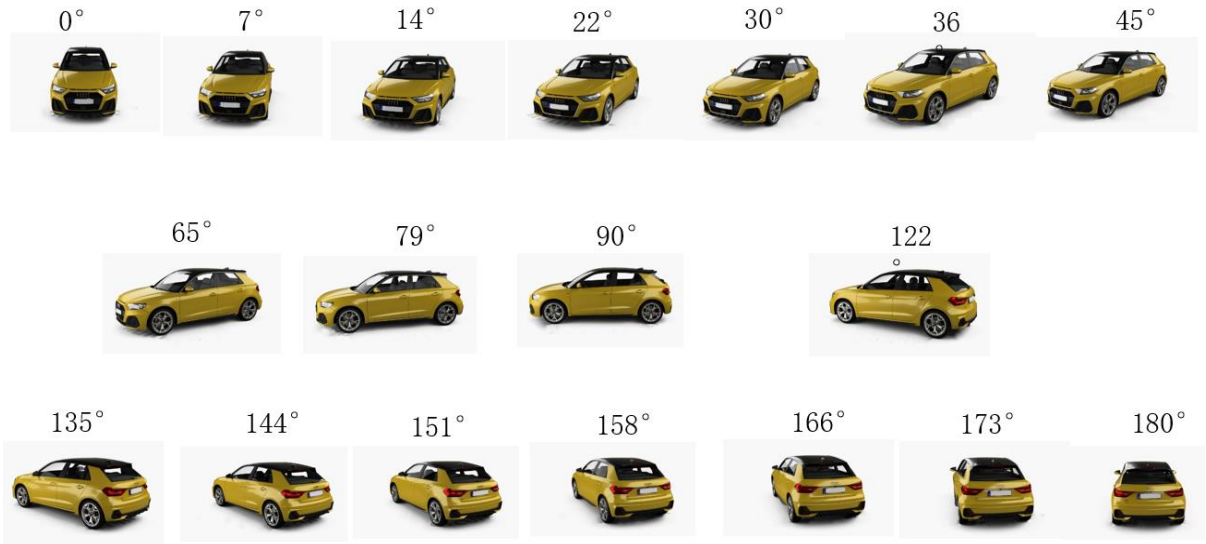


图 2.18　紧致性数据集角度分布图

Fig. 2.18　Angle distribution of Compact dataset

计算资源对于网络的训练也十分宝贵，若能减少训练样本就能节约计算时间的同时降低计算成本，本文对计算时间进行统计，经统计，在原始数据集上进行网络训练用时 59 分 32 秒，在 18 张训练图片上进行训练用时 35 分 1 秒，减少了 41.18%训练时长。

2.5　本章小结

本章对几个汽车数据集进行总结，在 Compcars 基础上进行修改，搜集整理得到两个新的数据集 MTV-Cars 与 MTV-1638s。在 MTV-Cars 上进行两组实验，探索汽车数据集不同属性，主要是角度与数据规模对汽车车型识别的影响，实验发现训练库中单一车型样本数量及角度对识别效果影响较大，但是由于数据集数据部分不均匀，结果不具有代表性，因此重新收集一个多角度均匀分布数据集 ASMTV120s，利用此数据集进行实验，发现对于汽车车型识别，只需要单个车型不多于 18 张紧致数据集就能达到较高的识别精度，减少了 41.18%训练时长,因而大大降低了由于数据集庞大带来的计算成本，这一结论对于评估一个数据集的合理性以及数据集收集都有很大的参考价值。

3 汽车车型识别算法

汽车车型的精确识别是指对于一张静态图片或者视频中的一个车辆，能正确识别出它的车型，例如输入的图片中包含宝马 X3，则输出就应该是宝马 X3，或该车型对应的标签。汽车车型精确识别属于识别分类问题，具体来说，属于细粒度分类问题。完成这一任务可以采用三大类方法，一是基于特征提取，二是基于几何估计，第三种方法是基于深度学习方法，前两种属于传统模式识别方法，本文不做详细叙述，深度学习方法主要有以下几种方案。一是利用深度卷积神经网络，此方法近几年在分类识别问题上取得了重大突破；二是利用部件级信息，综合全局信息，对应于车辆识别问题，则是利用车灯、进气格栅等部件信息加强。三是利用多任务学习的方法，用两个甚至多个神经网络提取特征，对网络的不同任务所得损失进行联合训练。四是基于注意力的方法。其中方法二到方法四又可以统一为细粒度分类方法。

本章对汽车车型识别方法进行概述，并对 VGG-16，Resnet-50 以及 B-CNN 进行实验，确定继续优化改进的基础网络 ResNet-50，在此进出网络上提出一个新的多阶段学习网络 MS-CNN，此网络能学习到更加丰富的语义信息，并且与 VGG-16、ResNet-50 以及 B-CNN 相比能得到相对较高的识别精度。

3.1 深度卷积神经网络（DCNN）

深度卷积神经网络自 2012 年的 ImageNet 图像分类竞赛，AlexKrizhevsky[19]等人利用 AlexNet 在竞赛中夺得冠军，神经网络再次登上舞台，之后提出的神经网络都是在此基础上或受此启发对网络结构的改进与优化，改进的方向主要包括网络的深度与宽度。

在对网络进行优化改进前，首先要对网络结构有深入了解，卷积神经网络主要有以下几个网络层：输入层、卷积层、池化层、全连接层以及最终用于分类的 softmax 层。对图像的像素矩阵进行一些运算。近几年比较流行的几个深度卷积神经网络包括 AlexNet[19]、VGG[47]、GoogleNet[48]、Resnet[49]以及后来的 DenseNet[50]。网络不断加深，为了减小网络加深带来的弊端，例如参数计算量增大，占用的计算时间长以及计算内存加大。神经网络在不断改进与优化，接下来将对深度卷积神经网络进行详细讲解。

3.1.1 DCNN—网络结构

深度卷积神经网络主要包含以下几个重要的网络层：卷积层、激励层、池化层、全连接层、softmax 层。对于空间中零散分布的数据，输入到网络模型中，最终由分类器进行分类，对下面就这几个网络层的原理进行简要阐述。

从数学的角度出发，卷积有如下定义：

设 $x(\alpha)$， $\omega(\alpha)$ 是 R 上的可积函数，卷积则定义为：

$$h(t)=\int x(a)\omega(t-a)d\alpha \tag{3.1}$$

记为

$$h(t)=x(t)*\omega(t) \tag{3.2}$$

但是上面的式子是针对于两个可积函数的卷积，现实中研究的变量却经常是离散型的，比如对一段声信号进行一定采样率下采样得到的数据等，因此将卷积的思想和原理推广到离散的情况，就得到了离散形式的卷积

$$h(t)=x(t)*\omega(t)=\sum_{a=-\infty}^{\infty}x(a)\omega(t-a) \tag{3.3}$$

在卷积神经网络中，其中 x 表示输入（input），参数 ω 表示核函数（kernel function）。利用卷积神经网络进行图像识别进行的运算就是离散卷积，离散卷积可以看作为矩阵乘法。神经网络进行图像识别时的输入图像以及神经网络内部的卷积核作的卷积运算就是矩阵运算。图 3.1 给出一个二维数据卷积运算的示例。在进行图像识别时，输入为三通道的图片，示例中未呈现的一个参数为图像的深度又称为通道数（channel）。

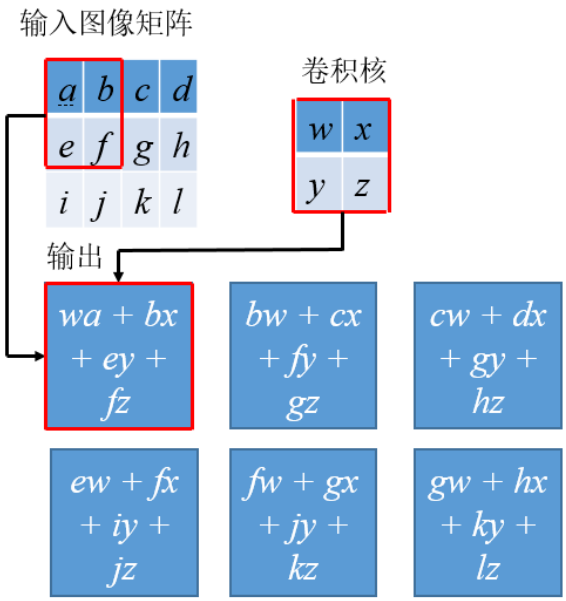
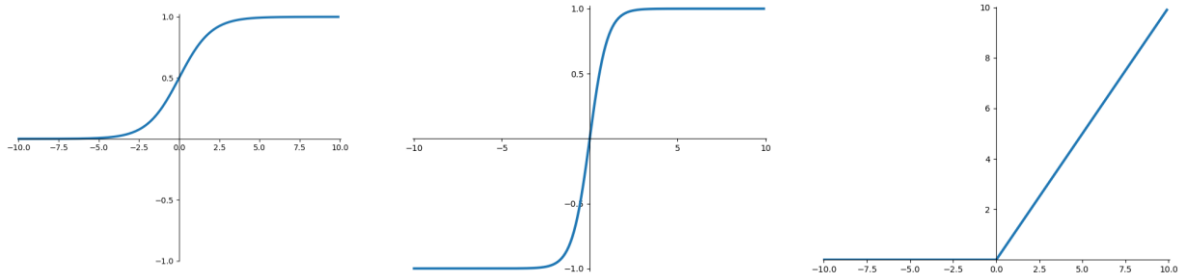


图 3.1　二维卷积原理

Fig. 3.1　Principle of two-dimensional convolution

几种经典的神经网络中网络的输入为（224*224*3）的三通道静态图像，再经过卷积核（kernel）对输入图像矩阵的卷积运算得到一个输出，此输出会作为激励层的输入，

激励层是引入非线性变换，若没有激励函数，神经网络每层只能做简单的线性变换。线性模型的表达能力不够，不能拟合很多复杂的非线性函数，因此网络中必须加入激励函数。比较通用的激励函数有以下几种 sigmoid 函数，tanh 函数，ReLU[19]函数。图 3.2 从左至右依次是 sigmoid、tanh、与 Relu 函数。其中 Relu 由 AlexKrizhevsky 等人提出，其表现优于其他的激活函数，广泛应用于神经网络中。从图中可以发现，当输入在很小或者很大的时候，前两种函数都会出现梯度消失的现象，梯度消失是指随着隐藏神经元数量的增加，梯度以指数级减小，甚至减小到零。而 Relu 函数没有 sigmoid、tanh 函数的饱和区，从而有效地避免了梯度消失的现象，使得收敛速度快，且 Relu 函数使得一部分神经元的输出指为 0，即当网络的输入小于 0 时，输出为 0，这就可以使网络变得稀疏，与此同时，减少了网络的计算参数，避免了网络的过拟合，节约了训练时间。



(a) sigmoid 函数　（b）tanh 函数　（c）relu 函数

图 3.2　激活函数

Fig. 3.2　Activation function

池化层的作用是下采样：保留显著特征、降低特征维度，增大卷积核的感受野，帮助输入的表示近似“不变”，对一些输入发生少量平移时，池化操作能使输出基本没有变化，下采样能有效减少计算参数，在一定程度上可以加快计算速度和防止过拟合。池化层中卷积核的大小影响感受野的大小，通常来说，卷积核尺寸越大，感受野越大，但是计算的参数也会增加，为此，用小卷积核的叠加使用来代替大的卷积核，在保证相同感受野的前提下，减少计算参数。池化函数包括最大池化（max-pooling）、平均池化（average pooling），NIN[51]网络中提出的全局平均池化，利用全局平均池化层代替全连接层，大幅度减少了网络计算参数，加强了特征图与最终分类类别之间的联系。GoogleNet[48]、ResNet[49]两个网络框架中都利用全局平均池化，但是最终都有一层全连接层。在[52]提出一种空间金字塔池化，神经网络需要固定的输入尺寸，是因为全连接层需要固定的输入维度。空间金字塔池化用在全连接层之前可以将任何尺度的图像经卷积得到的卷积特征

化成相同维度，这可以让网络处理任意尺度的图像，避免由于图像裁剪与缩放操作导致信息的丢失，减轻了图像预处理对图像识别带来的巨大影响。

经池化操作得到的特征图中包含一些局部特征，全连接层的作用是对局部特征进行整合，然后归一化。一般全连接层的参数最多，因此现在很多网络结构为减少计算参数数量，会用其他网络层取代全连接层。例如 NIN[51]、GoogleNet[48]、ResNet[49]利用全局平均池化代替全连接层，FCNN[53]利用 1*1 卷积层代替全连接层。虽然全连接层不能保留图像的空间结构，但是同时它减少了特征的空间位置给分类带来的影响，增强了分类的鲁棒性，但是对于位置信息很重要的任务，例如检测与分割，可以考虑替换全连接层。

3.1.2 汽车车型识别常用网络模型

A. VGGNet[47]网络结构

VGGNet 在 AlexNet 后取得了很大突破，也使得这一网络至今被用来提取特征，文中的很多创新思想例如卷积核的应用对此后网络的发展有重要影响，并且现在很多任务包括分类识别以及目标检测，都是在这个网络基础上进行改进的。VGGNet 得到一个很重要的结论：神经网络深度的增加以及小卷积核的使用对提升最终的识别分类结果有积极作用，为以后提升网络性能与网络结构优化指明了方向。

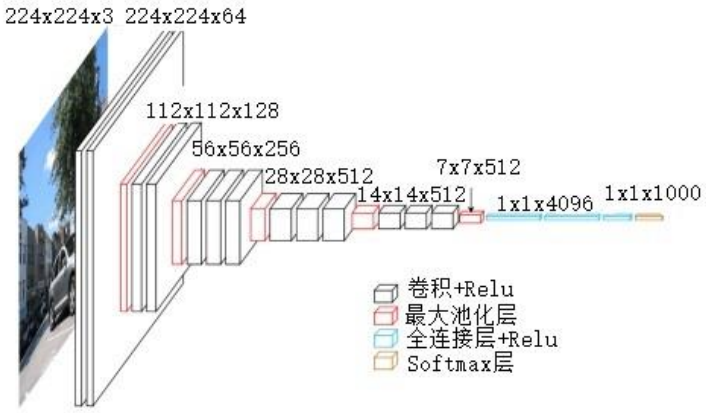


图 3.3 VGG16 结构示意图

Fig. 3.3 The architecture of VGG16

VGGNet 的经典模型 VGG-16 网络结构如图 3.3 所示，它通过堆叠 3*3 卷积层以及 2*2 最大池化层来不断加深网络，网络深度的增加并没有带来过多的参数，这也是使用小卷积核带来的好处，网络参数依然集中于网络最后的全连接层。与之前的网络结构相比，VGGNet 与之不同的是，网络卷积层的卷积核大小全部采用 3*3 的尺寸大小。原因是使用小卷积核可以减少计算参数，用一个例子来证明，两个 3*3 的卷积核叠加使用相

当于一个 5*5 卷积核的感受野大小，3 个 3*3 卷积核叠加使用相当于一个 7*7 卷积核的感受野大小。但是在计算参数上，前者的计算参数量只有后者的一半，并且前者可以进行三次非线性计算，而后者则只能进行一次非线性计算。这样使得前者对非线性特征的学习能力更强，更能够学习到更加抽象的特征。这也使得 VGGNet 网络结构更加合理。

VGGNet 摒弃了 AlexNet 中的局部响应归一化层，因为作者发现，对于 VGG 而言，加入此网络层，不但不会提升网络效果，还会加大内存消耗与训练时间。在对 VGG 网络进行训练时，对数据集进行了数据增强，采用多尺度的方法，将原始图像缩放到不同的尺寸，再将图片随机裁剪成 224*224 的大小。作者对多尺度进行了评估，相对于单一尺度，多尺度的数据增强方法能提高最终的分类精度。

B. Resnet[49]网络结构

ResNet 是继 AlexNet、VGGNet、GoogleNet 后提出的又一新型神经网络。与这些神经网络相比，此网络在结构上进一步加深。网络加深能帮助提高网络的表达能力，同时深层网络能提取更抽象的特征，但是随着网络加深也带来很多问题，作者发现，伴随着网络的加深，出现了梯度消失与梯度弥散问题，这些问题通过正则化、权重初始化等方法解决，但是随着网络层数的进一步增加出现了“网络退化”问题，即随着网络深度的加深，分类准确率相较于那些层数较少的网络，分类准确率下降。图 3.4 是两个不同深度的网络在 CIFAR-10 数据集上进行训练和测试的错误率结果，随着深度的增加，准确率趋于饱和，再经过几次迭代后，准确率突然降低，并且层数更多的网络有更高的训练误差，测试误差也更大，图中 56 层的网络效果比 20 层的更差。为了充分发挥深层网络的优势又能解决上述问题，Kaiming He 等人提出一个残差学习网络—ResNet，文中提出了几种网络改进方法，包括：残差映射、捷径连接、瓶颈结构[49]。

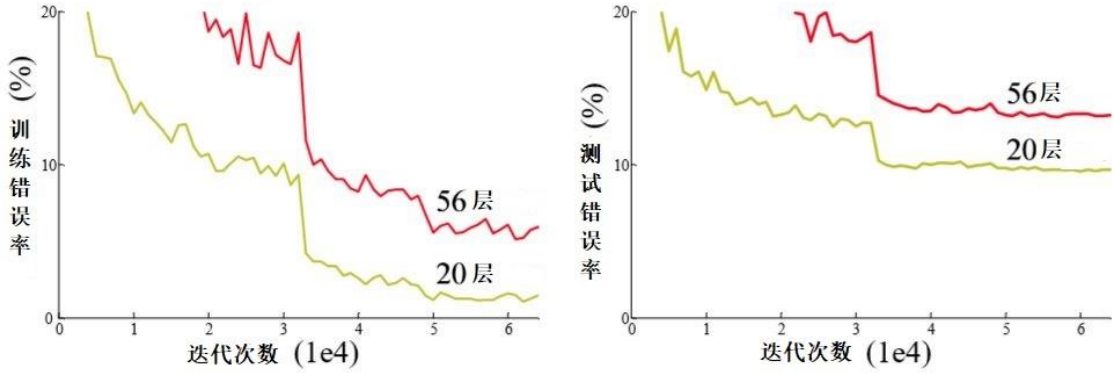


图 3.4　网络退化

Fig. 3.4　Network degradation

（1） 残差映射

由于网络的加深使得网络出现退化，并且实验证明导致此现象并不是因为发生了过拟合，因为在准确率下降不止体现在测试集上，在训练集上同样也出现这种现象。试想在一个浅层网络正确率已经达到饱和后，在这个网络后面加几层恒等映射，这样在增加网络深度的同时，也不会降低网络最终的准确度。受此启发，ResNet 神经网络引入了残差块，引入的残差块可以用图 3.5 来表示。

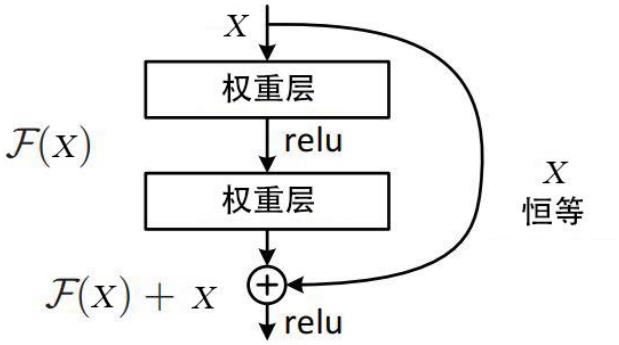


图 3.5 残差块

Fig. 3.5 Residual block

在图中，假设网络的输入为 X ，如果网络增加层为恒等映射，则输出 Y 应该满足 $Y=X$ 的关系 。残差学习的方法是，形式上，把几个堆叠层学习到的 $H(X)$ 作为所需的基本映射，假设多个非线性层的叠加能拟合复杂函数，即假设这些叠加层能拟合残差函数，用 $F(X)$ 表示。在满足上述假设的同时还需满足恒等映射，所以希望网络叠加后的输出 $H(X)$ 与输入 X 相等，因此网络最后需要拟合的函数表示为 $F(X)=H(X)+X$ 。并且网络学习的目标是使得上述残差函数趋于 0。

若采用之前网络的学习方法，网络要学习的目标函数为 $H(X)=F(X)+X$ ，但是实验证明直接拟合底层映射 $H(X)$ 的加法运算不利于优化，且很难学习到恒等映射函数，如果直接学习残差函数，随着网络加深，那么 $F(X)$ 就会逐渐趋于 0，此时就只剩下恒等映射了，理论上，此时网络不会因为网络加深而使误差增大。因此 ResNet 改变了原来的学习目标，不再直接学习它的基本映射，而是学习残差映射。

（2） 捷径连接

要实现恒等映射，在残差块中采用了捷径连接（shortcut connection）。通过捷径连接将残差块的输入和输出进行智能叠加，并且残差块的捷径连接不会给网络增加额外的参数和计算量，与此同时还可以大大加快模型的训练速度、提高训练效果 。

ResNet 对每一个堆叠层都采用残差学习的方式，那图 3.5 的输出可以用下式表示：

$$y = f(x,\{\omega_i\}) + x \tag{3.4}$$

式中 x 和 y 分别是网络残差块的输入与输出。函数 $f(x,\omega_i)$ 代表学习的残差函数。如图 2 所示残差块的叠加层是两层，则残差函数 $f(x,\{\omega_i\}) = \omega_2\sigma(\omega_1 x)$，其中 σ 表示第一层卷积运算后用 Relu 激活。式 3.4 成立的前提是假设残差映射与网络的输入维度相同，因此可以直接进行加法运算。但是若两者的维度大小不相等时，捷径连接就可以发挥作用，使残差块的输出与输入进行智能叠加。捷径连接通过 ω_s 来匹配维度，此时式 3.4 将变为如下的表达式：

$$y = f(x,\{\omega_i\}) + \omega_s x \tag{3.5}$$

需要注意的是只有在输入输出的通道数不同时才采用式 3.5。

（3） 瓶颈结构

VGG 网络采用两个 3*3 卷积层代替一个 5*5 卷积层，做到在相同感受野大小的情况下大大减少了运算参数。ResNet 中叠加层有两层或者三层，如图 3.6 所示：ResNet 沿用小卷积核，对于 ResNet-50、ResNet-101、ResNet-152 采用图右的瓶颈结构。

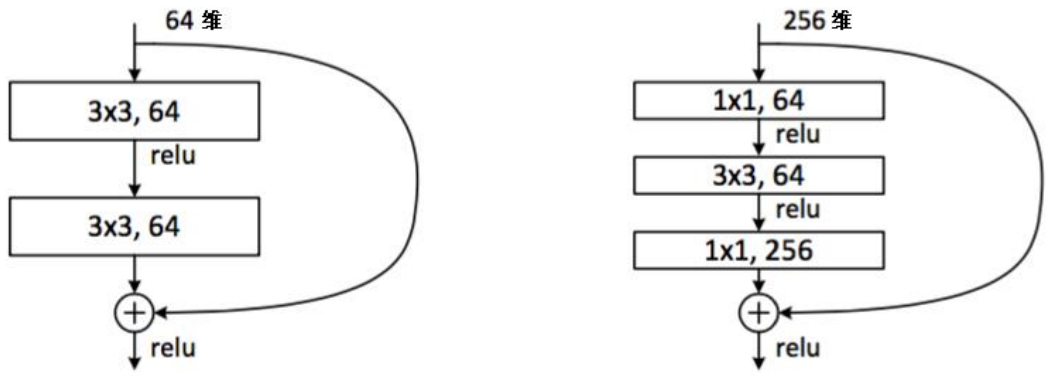


图 3.6 瓶颈结构

Fig. 3.6 Deeper bottleneck architectures

（图左 all 3x3 图右 bottleneck）

瓶颈结构是指先采用在在中间 3*3 的卷积前后分别使用 1*1 的卷积，因为残差学习需要叠加层输出与输入维度相同，采用瓶颈结构，先用 1*1 卷积层实现降维，最后再利用此卷积层进行升维。若没有 1*1 卷积层的降维再升维，实现相同的功能就只能通过两

个 3*3*256 的卷积层，这样参数计算量相较于现在的结构增加了 16 多倍。因此该结构减少了参数量，可以帮助网络拓展成更深的网络结构。

3.2 汽车细粒度分类方法

关于汽车车型的识别问题，属于汽车的细粒度分类问题。关于汽车的分类可以简单划分为粗粒度、中粒度、以及细粒度问题。粗粒度车型的识别问题可以解释为识别汽车的类型[54-57]，具体而言，是指划分汽车属于轿车、跑车、SUV、MPV、卡车等。中粒度车型识别则是指对汽车品牌的识别[58,59]，具体来说是指是识别汽车是大众、奥迪或是宝马。本文研究的汽车车型精确识别问题是指区分汽车的型号，例如区分奥迪 A6 与奥迪 A4，或者识别宝马 3 系与宝马 5 系。在细粒度分类问题上已经做了大量研究，对这些方法进行总结整理，可以概括为以下三种方法：部件级特征识别、网络集成方法、以及基于注意力的方法。

车型细粒度分类问题的同样存在类内差距大而类间差距小的问题。针对汽车车型的精确识别，为迎合消费者的喜好，很多汽车在设计风格上可能会相互借鉴，因此不同的汽车品牌之间可能会存在相似性，这更是给汽车车型的识别增加难度。且由于一些外在因素，例如外间环境的光照，不同的角度等问题都会给识别带来困难。部件级特征识别就是利用部件级的信息来帮助识别。部件级信息是指利用汽车的前照灯、进气格栅、挡风玻璃、以及它的一些局部信息不同的几何特征与纹理特征进行识别。最初提出此方法的是在[21]中，利用 R-CNN[60]中的选择性搜索（Selective Search）来提取感兴趣区域，即部件级特征，将提取的特征输入到神经网络，进行进一步识别。为了提高最终的识别精度，网络可能会加入一些约束。例如几何约束，部件对齐等方式。近几年，用深度学习方法提取部件特征上有了更大突破，主要是因为深度学习在物体检测与定位上的发展提高了部件级信息检测的精度，很多方法还可以实现端到端的训练。

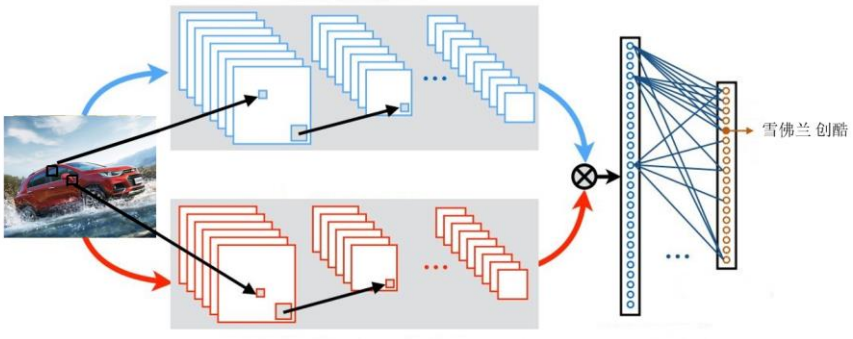


图 3.7 Bilinear CNN 结构示意图

Fig. 3.7 The architecture of Bilinear CNN

部件级特征识别因为只考虑局部信息犹如“盲人摸象”，为弥补这一缺陷，可以采用网络集成方法。文献[15]考虑利用局部与全局信息，这也综合了人认识事物的规则，会结合物体的局部信息与全局信息对事物进行分辨。目前取得很好成果是一种采用非对称网络结构模型[24]，两个网络结合局部与全局信息，为了定位到影响识别的关键区域，引入了一个跨通道池化，这样在网络的最后有三个损失函数，再对这三个损失函数联合训练。除此之外，比较经典的双线性卷积神经网络 B-CNN 也属于网络集成方法的一种，B-CNN 结构如图 3.7 所示。此网络采用两个独立的网络对同一个输入进行卷积运算，全连接层前通过外积池化，获得同一位置上的不同特征，充分利用了网络的 Mid-level 特征。在全连接层之前对最后一层卷积得到的特征图进行外积池化，此操作可以表示为：

$$bilinear(f_A, f_B) = f_A(k_A, ch_A)^T f_B(k_B, ch_B) \tag{3.6}$$

其中 $k = \omega * h$，代表不同位置的特征，ch 表示通道数，从上式可知，B-CNN 要求网络输出的特征图的大小相等，即要求式中的 k_A 与 k_B 相等。此结构可以在网络前面的层实现网络层共享，此时的两个网络模型相同，例如都采用 VGG-16，在最后一层卷积得到的特征上进行双线性池化操作。

网络集成方法还可以通过嵌入结构化标签，主要是对损失函数层进行改进，在模型中引入 Triplet loss[25]，最后利用一个影响因子与 softmax loss 一起联合训练。为了减少网络的计算参数，在网络训练时可以采用参数共享的方法，即在网络的前几层实现参数共享，在网络高层针对不同任务进行分别训练。最后再对网络的损失进行联合训练。

部件级特征识别一般需要对人工标注部件位置信息，标注需要耗费很多的人力物力，因此若有一种方法能通过学习定位到有利于识别的关键信息，将大大减少标注带来的麻烦，基于注意力的方法就能实现上述的功能，它能通过网络学习定位到有助于识别的关键位置信息以及输入识别对象的整体信息。例如采用二级注意力模型[15]，一级注意力模型用来定位图像中识别对象所在位置，二级注意力模型以一级注意力模型的输出为输入，定位识别类别之间存在差别的局部特征，即有助于准确识别的关键位置信息。利用选择性搜索的方法定位局部信息一定带来一些冗余信息，而基于注意力的方法能定位到识别的关键区域，且能定位识别整体对象位置，减少了因图片背景复杂带来的噪声。

3.3 车型识别方法数值实验与分析

本文使用的车型识别方法是基于深度学习方法。利用神经网络的卷积运算实现识别分类，神经网络的输入为二维静态图片，卷积运算实际为矩阵运算，经过一系列的卷积池化运算得到最终的损失函数，本文进行网络训练的最终目标是使损失函数降低到最小。进而得到一个较高的识别正确率。为了最小化损失函数，可以通过改善训练方式以

及选择适合的优化算法来实现。除此之外，为了增强模型的泛化能力，最好的办法是使用更多的训练数据，同样本文也采用了数据增强的方式来训练文中的网络模型。

3.3.1 神经网络优化算法

神经网络进行计算时，目标值与预测值之间的差值即为损失值，网络训练的目标为降低此损失值的大小，为了得到损失函数的最小值，模型内部参数起关键作用。这时需要用到各种优化参数与算法。优化算法可以大致分为两大类，分别是一阶优化算法与二阶优化算法，以下对几种比较常见与经常应用的算法进行对比分析。

(1) 随机梯度下降（Stochastic gradient descent，SGD）

网络训练模型参数更新的方式是梯度下降，利用梯度下降对神经网络模型进行权重更新，即沿着梯度方向更新调整网络模型参数，寻求损失函数最小值。随机梯度下降（SGD）及其演变的几种优化算法是机器学习算法中应用最多的神经网络优化算法。随机梯度下降不需要对参与训练的所有样本进行参数更新，它在每次训练时从训练集中抽取小部分样本进行参数更新。SGD 算法的一个关键参数是学习率（Learning Rate），选择一个合适大小的学习率能有效地避免损失函数落入局部最小，学习率不宜过小也不宜过大，合适大小的学习率可以通过试验与误差来选择，监测损失函数值大小随时间变化的学习曲线是最好的选择方法。SGD 算法因其参数更新速度快被广泛应用，但是它每执行一次就会进行一次参数更新，使得参数间具有高方差，这样导致损失函数会以不同的强度波动，并且容易使函数陷入局部最小值。

(2) 动量[61]（Momentum）

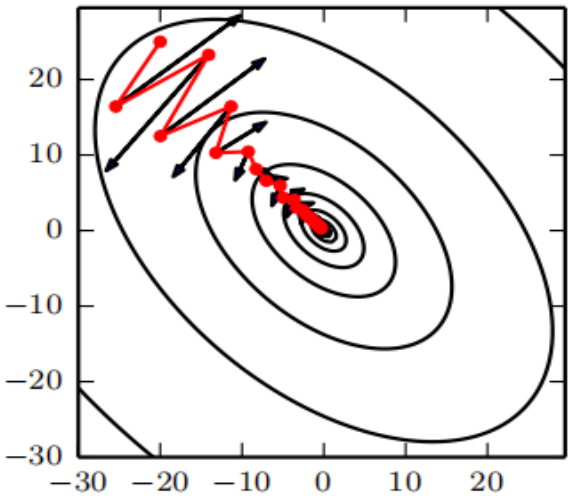


图 3.8 动量

Fig. 3.8 Momentum

随机梯度下降由于高方差振荡很难在短时间内稳定收敛，为了加快加速学习，提出了动量方法，如图 3.8 所示，红色线表示动量的学习路径，而黑色表示随机梯度下降的学习路径，可以看出动量在减小损失函数振荡，加快函数收敛方面有突出表现。此方法积累了之前梯度指数级衰减的移动平均，弱化无关方向的震荡，对于一些处理高曲率、数值小但方向一致的梯度或者带噪声的梯度，动量算法是一个合适的网络优化算法。

(3) Adam[62]算法（Adaptive Moment Estimation）

自适应学习算法是指根据梯度的方向，自动调整学习率的大小，Delta-bar-delta 算法是最早的自适应学习算法，该方法很简单，若梯度保持相同符号时，则增大学习率，反之，则减小学习率。 Adam 是应用比较广泛的自适应学习算法，不同于随机梯度下降更新权重的方式，保持学习率的不变性，该优化算法能自适应调整学习率，更新网络权重参数。与其他自适应学习率算法相比，Adam 算法加快了模型的收敛速度，因自适应调整学习率，从而提高了模型的学习能力，它还可以有效避免学习率消失、模型收敛速度慢等问题，并能大大降低了调参量。Adam 算法能实现高效计算的同时有效节省内存，它所需的内存极少，因此这个算法适合解决大规模数据与参数优化问题。此外，该算法适用于解决高噪声或梯度稀疏等问题。

本文解决的是大规模数据集的分类问题，且数据的背景复杂，因而选用自适应学习算法更能解决本文的研究问题，本文实验时使用 Adam 算法更新优化网络参数。

3.3.2 数据增强

本文所研究的内容为数据驱动，实验数据的数量以及质量对实验结果都有重要影响，在机器学习中，为了增强模型的泛化能力，最好的办法就是采用更多的实验数据对模型进行训练，但是在实验中，数据往往有限，因此为了增大数据量，需要制造一些假数据加入到训练集当中去，在深度学习中，制造这部分假数据也十分简单。

实验中比较常用的数据增强方式有以下几种：缩放变化，对图片按照某一比例进行放大或缩小，平移变换，图像按照不同方向进行平移，旋转，即对图片按照某一角度进行旋转，此外还采用色彩抖动，高斯噪声、模糊处理等方式来增加训练数据。AlexNet 在训练时就采用了数据增强的方式，从大小为 256*256 的原始图像中随机裁剪 224*224 大小的区域，除了对图片进行裁剪，还对图片进行水平镜像，经过以上操作，相当于增强了(256-224)*(256-224)*2=2048 倍的数据量。使用了数据增强后，能有效地防止过拟合，提升模型的泛化能力。而 VGGNet 中采用多尺度的方法来做数据增强，作者将图片原始尺寸首先缩放到不同大小尺寸 S，S 的取值范围是[256，512]区间，取定 S 值后，再随机裁剪成 224*224 大小的区域。这样同样也可以达到增加数据的目的。并且实验结果

证明用多尺度的方法还可以提高模型的识别精度。图 3.9-3.10 展示的是数据增强的效果图。如图 3.9 经过裁剪可以发现会裁剪掉图片的背景信息，但是也会裁剪掉对于识别对象的信息。图 3.10 是对数据进行翻转，即对图片进行镜像。

图 3.9　图片裁剪

Fig. 3.9　Image cropping

图 3.10　图片翻转

Fig. 3.10　Image flipping

由于采用随机裁剪的方式可能造成信息丢失，因此本文在训练数据时并没有采用随机裁剪的方法进行数据扩充，本文只采用了一种数据扩充方式：对图像进行水平镜像。

3.3.3 实验结果

本次实验所用的实验配置为 NVIDIA GeForce GTX 1080 Ti 图形处理单元（GPU），使用的深度学习框架为 Tensorflow，利用 CUDA8.0 和 CUDNN6.0 加速运算（以上设备都由 AutoMorpher 课题组提供）；

(1) 数据集

本章实验采用的数据集有以下三个 MTV-Cars、Stanford Cars、以及 VMMRdb，数据集介绍将在第二章已经进行了进行详细介绍，此处不再展开。

MTV-Cars 在上一章已经详细介绍，同样本次实验车型一共有 1448 种汽车车型，所用到的汽车整车图片数目为 133710 张。对 VMMRdb 数据集进行阈值过滤，即某一车型图片数量少于 20 张，则忽略该车型，最终此数据集包含 652 种汽车车型，本文的研究主要是对车型的识别问题，因此将相同车型但是不同的年份的汽车化为同一个车型，最后实验所用的图片数量为 281970 张汽车整车图片。而对于 Stanford Cars 数据集，包含了 196 种汽车车型，有 16185 张汽车整车图片，相对于前两个数据集规模较小。

实验中，实验数据被分为训练集、验证集、测试集。数据集的划分方法可以分为以下三种：留出法（留一法或者留 P 法），交叉验证法、自助法。在划分数据集时还应该注意数据集的划分比例问题。通常若测试集的数据规模越小，对模型的泛化估计就越不准确。经常使用的对训练集、测试集的划分比例为 6∶4、7∶3、8∶2，在数据集的规模很大的情况下，还可以使用 9∶1。 本文采用交叉验证对数据集进行划分，对 MTV-Cars 数据集与 VMMRdb 数据集按照 8∶1∶1 将数据集划分为训练集、验证集与测试集。验证集不参与训练，只是负责实时观察训练效果，且训练集、验证集与测试集的数据相互之间互斥，互相之间没有交集。Stanford Cars 数据集在被提出时就已经被按照 1：1 的比例划分成训练集与测试集，由于本文实验时加入了验证集，所以对数据集的测试集进行划分，数据集已经进行随机打乱（shuffle），本文用测试集的一半数据用于验证，一半数据用于测试。即对 Stanford Cars 数据集按照 2：1：1 划分为训练集、验证集、测试集。

(2) 实验模型

本次实验的实验目的是为找到适合解决汽车车型精确识别问题的网络模型，本文选择比较经典的神经网络结构 VGG-16 与 ResNet-50 以及细粒度分类经典网络模型—双线性卷积神经网络（BCNN）这三种实验模型，上述几个模型的结构在上文已经进行了详细解释，这里不再赘述。值得注意的是，本文在实验是使用的模型都是经过预训练的模

型，所谓的预训练是指：首先训练简单模型求解一个简单问题，在这之后，训练目标模型求解目标问题的方法。本文使用的 VGG-16 网络模型以及 ResNet-50 网络模型都是在 ImageNet 上训练后保存的参数，在训练自己的模型用于汽车车型的精确识别问题上，直接加载预训练保存的参数，再对模型进行微调（Finetune），因为以上提到的两个模型在对 ImageNet 数据集进行分类时的类别数是 1000，而本文的分类任务中的类别数是 1448，因此实验时要对最终的全连接层进行修改。本文实验用到的 BCNN 的两个基础网络都是 VGG-16，这样在网络的 conv5-3 之前的层实现了参数共享，这样做的好处是减少了计算参数与训练时间。在这之后再对卷积得到的特征进行双线性卷积池化操作。

(3) 实验参数设置

实验的参数设置对最终的实验结果也有很大影响，其中涉及的几个重要参数：批尺寸（Batchsize）、回合（epoch）、学习率（learning rate）。批尺寸的大小不能太大，也不能太小。若批尺寸设置太大，则训练一个 epoch 所用的迭代次数减小，那想要达到相同的精度需要的时间也更多，对参数的修正速度也变慢，并且批尺寸增大，所需要的内存也更大，很可能造成内存不够。相反，若是批尺寸设置过小，使网络模型难以收敛。本次实验的批尺寸大小设置根据不同的数据集而定，在 Stanford 数据集上设置为 64，在 Compcars 以及 VMMRdb 上使用 32 的批大小。回合是指将训练集的数据全部训练一遍，一般回合设置的越大，最后的精度也就越高，但是一般 loss 不再下降，设置大的回合数只会增加训练时间甚至发生过拟合，因此需要设置一个合适值。本文选用 ADAM 神经网络优化算法，因其更适合大数据，并且收敛速度快，学习效果更为有效，初始学习率设置为 0.0001，随后再根据训练效果调整学习率大小。以下为三个数据集在 ResNet-50 的参数设置如表 3.1 所示，

利用自适应学习率的好处之一是不用人为调整学习率，但是本文发现若采用一个固定的学习率对模型训练，在模型收敛后精度不再提高，若此时改变学习率大小，损失会继续下降。因此本文实验采用了学习率衰减因子，每次实验利用 0.1 倍的学习率训练。

表 3.1　ResNet-50 实验参数设置

Tab. 3.1　ResNet-50 parameter settings

数据集(dataset)	批尺寸（Batchsize）	初始学习率 (learning rate)	学习率衰减数（learning rate decay）	全数据迭代次数（iteration）
Stanford	64	0.0001	*0.1	40-30-30-30
VMMRdb	64	0.0001	*0.1	10-10-10-10
MTV-Cars	32	0.0001	*0.1	50-20-20-20

利用上述参数设置，基于 VGGNet-16、Resnet-50、B-CNNs 三种网络模型，在 Stanford Cars、VMMRdb、MTV-Cars 三个数据集上实验后得到的实验结果如表 3.2 所示，将结果绘制成图标，如图 3.11 所示。

表 3.2　车型识别算法准确率对比

Tab. 3.2　Accuracy comparison of three VMR algorithms

实验基准	Stanford Cars	VMMRdb	MTV-Cars
分类标签	制造商-车型-年份	制造商-车型	制造商-车型
类别数	196	652	1448
实验图片数目	16185	281970	133710
VGGNet-16	47.78%	65.76%	60.31%
ResNet-50	85.25%	89.38%	94.66%
B-CNNs	85.30%	85.74%	86.65%

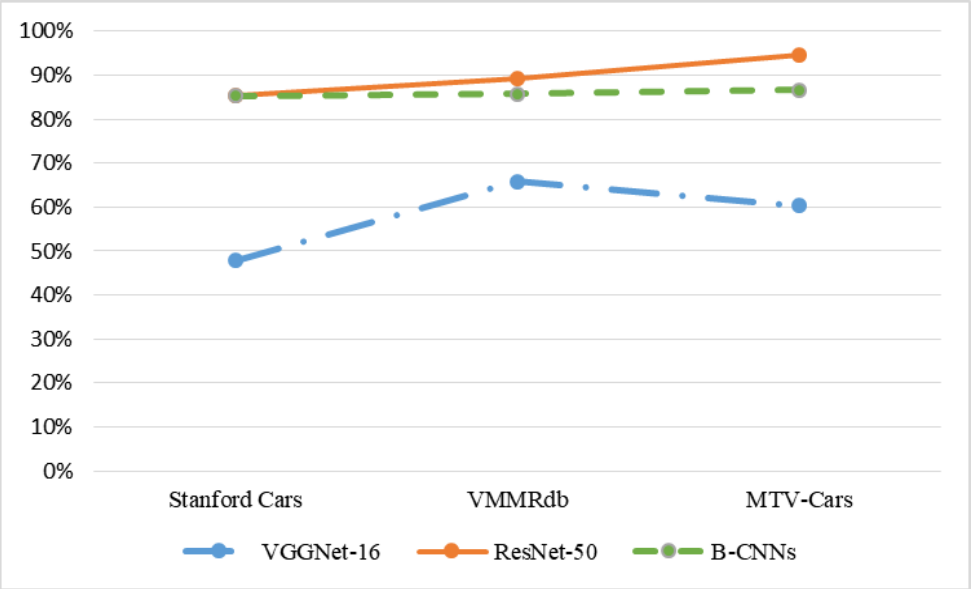


图 3.11　车型识别算法结果展示

Fig. 3.11　Vehicle Model recognition results

(4) 实验结果分析

从以上结果可以看出，ResNet-50 能很好的解决本文提出的任务，它取得了更高的精度，并且有更好的鲁棒性。分析原因可能是 ResNet-50 相较于另外两个网络，网络进一步加深，能提取到的抽象特征的更多，是否使用更深的网络会得到更高的精度？为此本文还利用另外一个深度神经网络 ResNet-101 进行实验，采用相同的参数设置与数据增强方法，然而，车型识别测试精度只有 14.47%，这说明单纯加深网络，并不会提高识

别精度，这可能需要网络模型与数据的“配合”，对于 101 层的神经网络，网络进一步加深，但是此时实验时的实验数据并没有增加，使得模型在训练集上取得很好的实验精度，但是在测试集上的精度却很低，即引起了过拟合。

为进一步明确实验结果，对实验结果进行数据分析，本文发现实验中出错的车型有两大类，第一类是比较不常见的车型，例如一些跑车与方程式赛车，此类车型的训练集中的图片数量太少，因此影响最终的识别精度；第二类识别错误较多的汽车车型是很常见车型，例如：大众、宝马、奥迪，因为在此类车型下的数据分布比较分散，在此类常见汽车品牌下拥有的汽车车型相比其他汽车品牌较多，由于家族基因或拍摄角度等的影响，不同车型之间的相似度很大，使得此类车型的出错率较高。此外还因为数据集结构问题，在第二章中已经进行讨论。

3.4 多阶段卷积神经网络（MS-CNN）

为了提高汽车车型识别的识别精度，近几年提出很多算法，这些算法中具有显著成效的，多阶段模型是其中之一，最初的多阶段模型是通过加入一个辅助网络，或者通过复杂的特征编码，从而利用高阶特征进行分类识别， 近些年，端到端的学习方法，增强了模型的高阶特征的学习能力，并且还减少了人工的参与，实验证明，ResNet-50 有很好的识别效果，本文在该网络基础上提出了多阶段神经网络（MS-CNN），旨在提高汽车车型的识别精度，并且能实现端到端的训练，本文提出的 MS-CNN 是利用一下三种卷积层，分别是空间金字塔池化（SPP）、1*1 卷积、以及全局平均池化（GAP），充分利用以上三种网络层的优势，下面将对这几种卷积层进行分析。

3.4.1 空间金字塔池化

卷积神经网络由两部分组合而成，卷积层与全连接层。卷积层可以接受任意尺度的图片，而在神经网络中，在输入网络前将图片固定成统一尺寸：224*224，这是因为网络组最后的全连接层运算要求，输入全连接层的数据需要相同的维度。但是原始图片尺寸一般很难满足上述固定尺寸，此时会对图片进行预处理，将图片进行裁剪或者变形，这样可能会造位置信息以及图片语义信息的丢失。空间金字塔池化（SPP）的作用是将任意尺度的输入固定为相同维度的输出，通常空间金字塔池化层加在最后一层卷积层之后，这样输入神经网络的图片可以是任意大小的图片，避免了信息的丢失。

空间金字塔池化的核心是空间金字塔池化层，空间金字塔池化层加在最后一层卷积层之后，代替原来的池化层，因此 SPP 是独立的结构，它并不破环原本的网络结构，假设输入网络的图片为任意尺寸，则经过卷积，在最后一层卷积后得到的特征图的大小也不同，SPP 采用多级池化，即采用多个窗口尺寸，这样才有可能得到相同的输出维度。

接下来，对 SPP 的工作原理进行进一步讲解，假设最后一层卷积得到的特征图的尺寸为 $S*Q$ ，首先确定空间卷积池化块（bins）的尺寸（$n*n$），卷积的通道数不会变化，SPP 会根据池化块的尺寸对特征图进行网格划分，每个网格的宽度为 ω， $\omega=s/n$ ，网格的高度 h， $h=Q/n$ 。例如图中池化块的尺寸大小分别是 $n=4,2,1$ 。则每个网格的宽与高分别是 $S/4$， $Q/4$ 。4*4 的网格划分最终得到 $16c$（c 为通道数）维度的特征向量。剩余两个尺寸以此类推。则最终得到的特征向量的维度为 $16c+4c+1c=21c$ 。很显然，经过空间金字塔池化后的维度特征与输入池化层的 S，Q 两个参数无关。在[52]一文中证明多尺度训练能提高分类准确度。空间金字塔结构图如 3.12 所示。

本文的多阶段神经加入空间金字塔池化，是为了提取到更丰富的空间几何信息，加快网络的收敛速度。

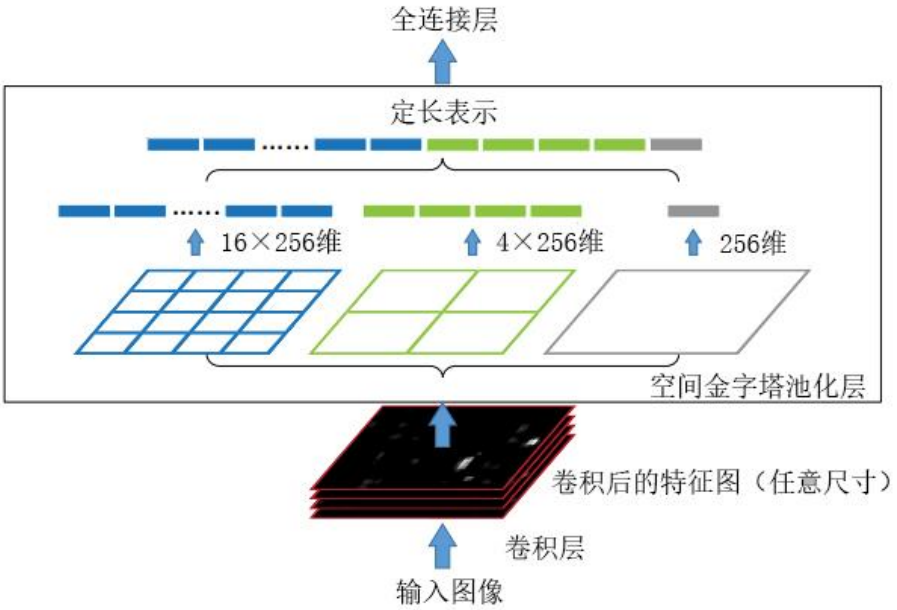


图 3.12　SPP 层结构示意图

Fig. 3.12　The network structure with a SPP layer

3.4.2　1*1 卷积层

在 Network in Network (NIN)[51]中，提出一个重要的卷积结构，1*1 卷积层，在 NIN[51]中，利用 1*1 卷积实现了对非线性函数的拟合，能提取出更加抽象的特征，在此之前，为了能提取到抽象的特征，一般采用过完备的卷积核，即尽可能采用多的卷积核，可是这样会给接下来的卷积层带来额外的负担，而 1*1 卷积能实现抽象特征的提取。之后 1*1 卷积被广泛应用，在 GoogleNet 中，利用 1*1 卷积降维，减少了运算参数，在 ResNet

中先利用 1*1 卷积进行升维，再利用相同的结构进行降维，实现了 ResNet 独特的瓶颈结构，在 FCN 中利用 1*1 卷积层代替最后的全连接层，使得网络能接受任意尺寸的输入，并且没有全连接层保留了原来的空间信息，使得网络最后在最后一层卷积层的特征图上进行上采样，实现了逐像素分类。文献[63]中用 1*1 卷积作用于不同卷积层卷积得到的特征图，得到不同卷积层的高阶信息，加强层内关系。综上所述 1*1 卷积有以下几个作用：首先，1*1 卷积能实现跨通道之间的信息整合，能够拟合更多的非线性特征，从而得到更加抽象的特征，这对最后的识别有很重要的意义。其次，1*1 卷积能实现降维，高维的特征会带来很大的计算量，并且高维特征可能都比较分散，而降维可以实现信息的压缩，与此同时还减少了参数，使网络结构更加轻量化。本文新建立的网络引入该卷积层就为了实现以上两个功能。

3.4.3 全局平均池化

全局平均池化[51]的池化方法最初在 NIN 中被提出来，全局平均池化是指在卷积得到的特征图上进行池化时，滑窗的大小正好是输出特征图的尺寸大小，例如在 ResNet-50 上，最后一层的输出大小是 7×7×2048，此时滑窗的尺寸大小应该是 7×7。在 ResNet 以及 GoogleNet 中都用到了全局平均池化，但是最终仍连接全连接层。全局平均池化主要有以下两个最作用：其一，利用全局平均池化代替全连接层，从而赋予最后一层的的特征图以特殊意义，即类别置信度，将最后一层得到的特征图直接输入到 Softmax 层进行分类，计算每个特征图属于某个类别的概率。其二，网络模型的参数集中在全连接层，而利用全局平均池化代替全连接层，就能大大减少计算参数，从而减少过拟合的可能性。

3.4.4 损失函数

损失函数是用来评估预测值与真实值误差的一个评价标准函数，网络在训练过程中将其最小化，从而找到合适的权重参数 ω，损失由两部分组成，分别是数据损失与正则化损失，网络训练过程中，通过反向传播算法得到一个梯度，来更新权值，降低损失。深度学习中解决的分类问题一般是多分类问题，针对多分类问题，在神经网络中，常见的损失函数是 softmax，假设训练样本集合 $X=\{x_1,x_2,x_3,...\}$ ，集合大小为分类个数，softmax 函数对输入进行归一化，其表达式如下：

$$soft\max(x_i)=\frac{e^{x_i}}{\sum_{i=1}^{n}e^{x_i}} \tag{3.7}$$

其中 x_i 是上一层神经元的输出，softmax 层的输入，输出为输入属于第 i 类的概率。

完成以上计算只是完成最终分类的第一步，接下来还要计算损失，softmax loss 函数计算公式为：

$$L = -\sum_{i=1}^{n} y_i \log soft\max(x_i) \tag{3.8}$$

其中 y_i 表示样本的实际标签。

3.5 MS-CNN 网络模型

3.5.1 神经网络结构

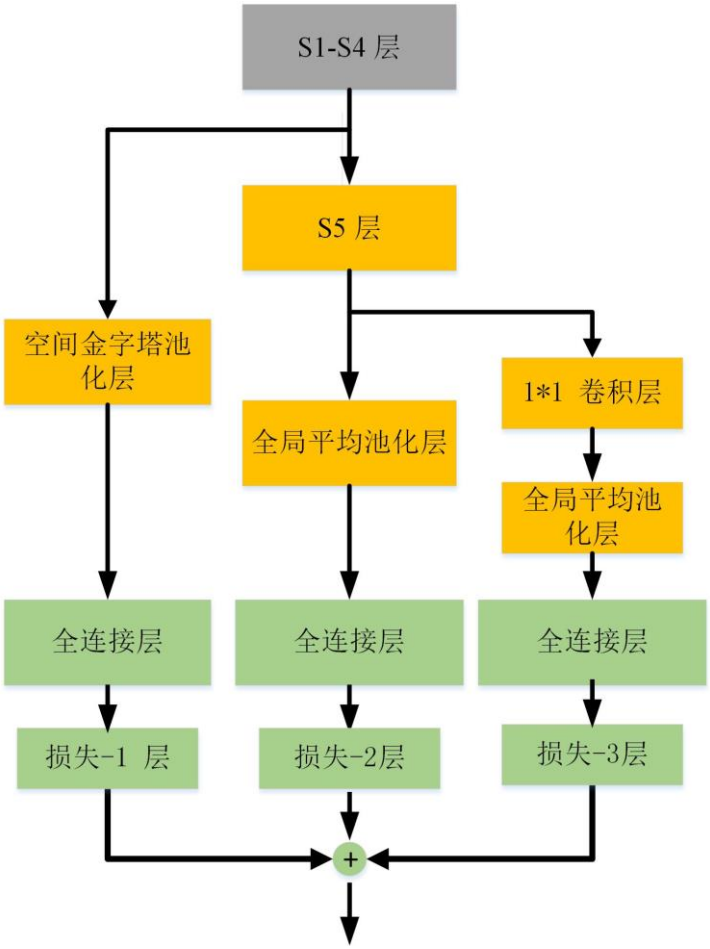


图 3.13 MS-CNN 网络结构示意图

Fig. 3.13 The structure of MS-CNN network

本文提出的多阶段卷积神经网络（MS-CNN）是在 ResNet-50 的基础上对网络的改进与优化，MS-CNN 框架如图 3.13 所示，利用 ResNet-50 的低层卷积层提取特征，并在

网络浅层实现卷积层共享，为了充分利用高阶信息，在 S4 之后连接空间金字塔池化层，共享层的选择是通过实验决定的，利用空间金字塔池化提取特征图中的多尺度信息，增强模型的鲁棒性。在网络的 S5 之后加入 1*1 卷积与全局平均池化，利用 1*1 卷积对最后一层卷积层得到的特征图得到的信息进行整合，得到更抽象的特征，本文设定新加入的 1*1 卷积层的输出维度为类别数，即本文使用的 1*1 卷积同时实现了信息整合与降维的功能，降维能对提取的特征进一步压缩，并能减少参数。此阶段的网络学习还可以用来定位学习到的关键信息，最后网络连接全连接层。本文在计算损失时引入权重参数：λ，λ 从 0.1，0.2 中选择，根据实验决定此参数的选择。

3.5.2 神经网络可视化

本文提出的 MS-CNN 是在 ResNet-50 的基础上，对网络进行改进，ResNet-50 包含 50 层网络层，其中有 49 层卷积层，用来提取特征，此外，网络包含两个池化层，在第一层卷积之后的最大池化层与全连接层之前的全局平均池化层，池化层中包含丰富的空间特征，本文对几个网络层进行可视化，并对比了 ResNet-50 网络与 MS-CNN 两个网络在特征学习上的差异，通过特征可视化理解不同的卷积层在提取特征发挥的作用。

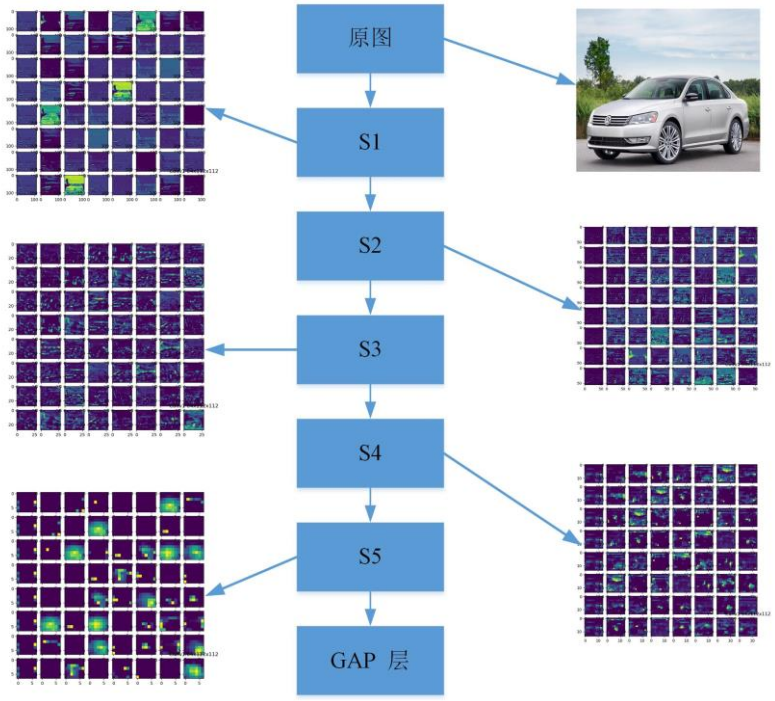


图 3.14 ResNet-50 网络层可视化

Fig. 3.14 Layer visualization of ResNet-50 network

本文的可视化是基于特征图的可视化，特征图是经过卷积池化后得到的输出，本文对 ResNet-50 几个残差块得到的特征图进行可视化，可视化如图 3.14 所示，特征图的数量与卷积核的个数相同，因此在卷积后的到的特征图不止一个，理论上，随着网络的加深，卷积核的个数会增加，只有增加卷积核个数，才能充分提取上一层卷积得到的，文中对其中的的 64 个通道的特征进行可视化。特征经激活函数（Relu）激活后，得到的特征是神经元在当前区域的最大响应，可以通过可视化体现。网络特征可以分为浅层特征，中层特征以及高层特征，本文对中层特征进行充分利用，并对高层特征进一步挖掘，随着网络的加深，网络能学习到更加抽象的特征。

MS-CNN 在 ResNet-50 基础上进行修改，加入空间金字塔池化层，网络在前几层实现了卷积层共享，因此对图片的输入尺寸有严格要求，网络的图片输入还是（224，224，3），但是加入空间金字塔池化层，能对 S4 得到的特征图进行多尺度提取，从而得到更加丰富的空间信息，提高了图像的尺度不变性，1*1 卷积加强了类别与特征的关系，本文同样对几个重要的卷积层进行可视化，对 MS-CNN 的可视化如图 3.15 所示

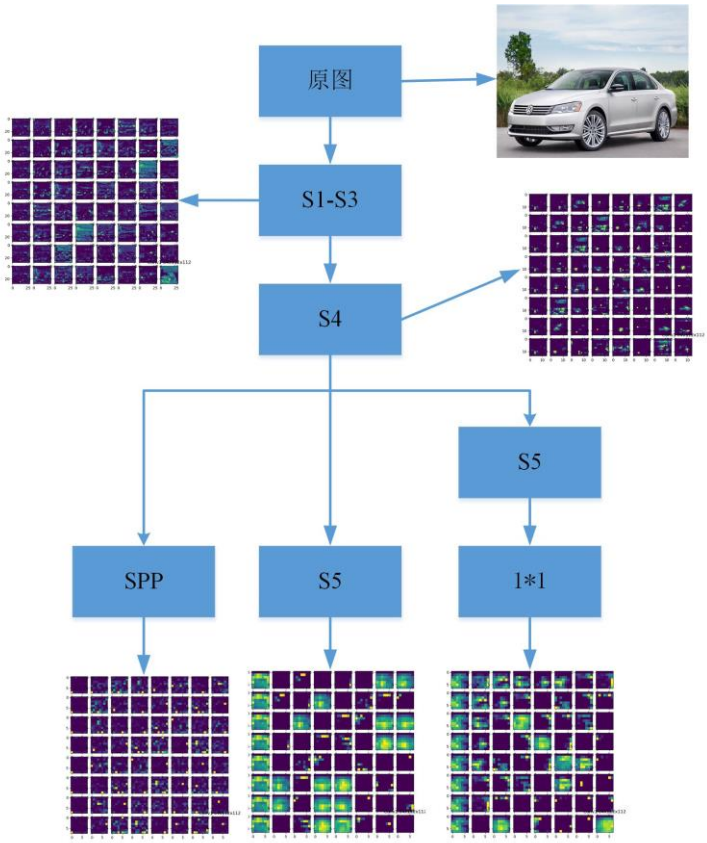


图 3.15　MS-CNN 网络层可视化

Fig. 3.15　Layer visualization of MS-CNN network

经对比发现，MS-CNN 加快了网络的收敛，通过对比两个网络模型 S4 层的特征图可视化，可以发现本文提出的 MS-CNN 能在 S4 层卷积后得到比 ResNet-50 更加抽象的特征并且网络新加入的卷积层 SPP 与 1*1 卷积是网络能提取到更加丰富的空间特征。同时本文加入的 1*1 卷积具有定位的作用，通过可视化可以看出，网络对汽车的关键位置有很大响应，新加入的左右两个分支增强了网络的表达能力，进而提升了识别精度。

3.6 MS-CNN 实验结果与分析

本文提出的算法能实现端到端的训练，低层学习比较简单的纹理信息，高层则包含丰富的语义信息，因此在网络的低层，利用网络共享的方法，从而减少多余的参数，选择共享层对最终的实验结果有重要影响，本文通过实验方法确定共享层。除此之外，网络的参数对结果也有很大影响，本文的损失由三部分组成，第一部分是由 ResNet-50 网络输出的分类损失(Class_Loss)，由空间金字塔池化输出的损失（SPP_Loss），以及由 1*1 卷积输出的损失（Part_Loss）。本文计算最终损失的的函数表示为式 3.9：

$$L = L_{cls} + \lambda L_{spp} + L \tag{3.9}$$

其中 L_{cls} 表示分类损失，L_{spp} 表示空间金字塔池化损失，L_{part} 表示 1*1 卷积损失。

为确定本文参数，以及确定框架的共享层，选择对 Stanford Cars 数据集进行实验，由于此数据集在验证算法上被广泛应用，实验结果如表格 3.3 所示

表 3.3 MS-CNN 参数设置

Tab. 3.3 MS-CNN parameter settings

数据集	SPP 共享层	λ 设置	车型识别精度
Stanford Cars	Ori	ResNet-50	85.25%
	S3	Cls+0.1SPP+Part(-fc)	85.21%
	S4	Cls+0.1SPP+Part(-fc)	86.49%
		Cls+0.2SPP+Part(-fc)	85.40%
		Cls+0.1SPP	82.67%
		Cls+Part(-fc)	85.41%
		Cls+0.1SPP+Part(+fc)	86.90%
	S5	Cls+0.1SPP+Part(-fc)	85.48%

根据实验结果可知，在 S4 层之后，接入空间金字塔池化的效果要比其它层的实验精度高，并且在 λ 取值为 0.1 时为模型表达能力比较强，在此基础上，本文尝试减少分支，发现无论去掉空间金字塔池化层或去掉 1*1 卷积层，网络的性能都会下降，因此最终确定网络的结构为，在 ResNet-50 的基础上，在 S4 之后叠加空间金字塔池化层，之

后输入全连接层计算损失，另外一个阶段学习特征则是在 S5 层之后，此部分的共享层根据经验获得，网络的加深能提取到更抽象的特征，本文受文献[51]启发，利用 1*1 卷积与全局平均池化代替全连接层。但是经试验发现，若在本阶段加入全连接层能继续提高识别精度，全连接层忽略了网络空间结构对分类结果的影响，增强了模型的鲁棒性。

根据以上实验结果，确定本文使用的网络框架结构，本文研究问题是针对多角度、大规模的数据的汽车车型识别问题，因此在本文收集整理的新汽车数据集 MTV-1638s 上进行实验，对数据集进行阈值过滤，最后参加训练的汽车车型 1377 个，按照 8：1：1 的比例将数据集划分为训练集、验证集与测试集。其中训练集包含 110340 张汽车整车图片，而验证集与测试集分别包含 13790 张汽车整车图片。同样在 Stanford Cars 与 ASMTV120s 数据集进行实验，Stanford Cars 数据集已经做好划分，对 ASMTV120s 数据集按照 6：2：2 的比例划分，训练集中包含 3600 张汽车图片，而验证集与测试集分别包含 1200 张汽车整车图片，实验结果在表格 3.4 列出。

表 3.4　汽车车型精确识别结果

Tab. 3.4　The result of vehicle model recognition

车型识别模型	Stanford Cars	MTV-1638s	ASMTV120s
ResNet-50	85.25%	95.48%	99.83%
MS-CNN	86.90%	95.51%	100%

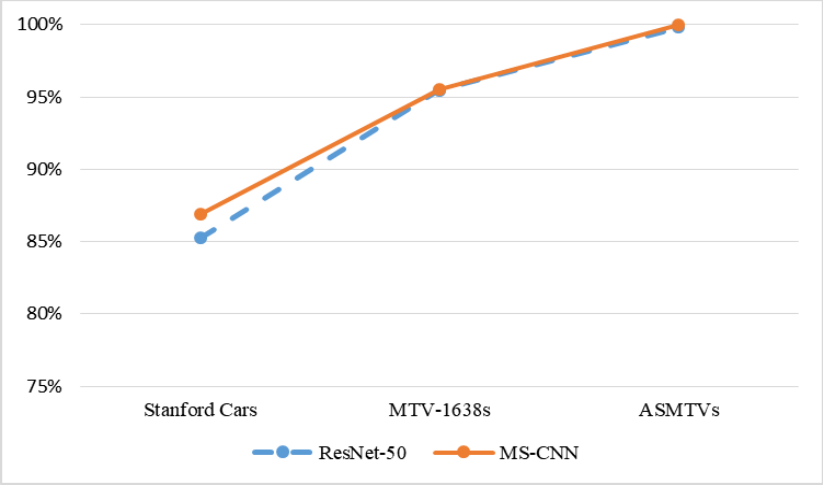


图 3.16　汽车车型精确识别结果比较

Fig. 3.16　Comparison of VMR accuracy on three dataset

从图 3.16 中可以看出，相较于 Stanford 数据集，MS-CNN 在 MTV-1638s 的提升微弱，这与数据集规模有直接关系，且 MTV-1638s 的类别输出有 1377 类，而 Stanford Cars

的输出只有 197 类，相比较而言，前者要比后者复杂的多，所以在 MTV-1638s 上的表现能力要比在 Stanford Cars 上的能力稍微逊色。而对于 ASMTV120s 数据集没有复杂背景，在 MS-CNN 网络下得到 100%的识别精度。

3.7 本章小结

本章主要讨论了用于汽车车型精确识别的两种典型的方法，主要分为两大类，第一类使用通用的深度卷积神经网络（DCNN），利用深度卷积神经网络做相关任务，例如本文的分类识别任务，要比传统模式识别方法能取得更好的效果；第二类方法就是使用细粒度分类方法，本文的车型识别属于细粒度分类范畴。因而本文选用 VGG-16 以及 ResNet-50 两个深度卷积神经网络模型，采用 B-CNN 一种经典细粒度分类方法。本文基于三个汽车类数据集 Stanford Cars、VMMRdb、以及 Compcar 数据集，保证了所选方法的优越性以及此方法在不同数据集上有很好的泛化能力，对不同的数据集上进行实验后，经过实验对比，发现 ResNet-50 在三个数据集上较其他两个网络模型表现出更高的精度与稳定性，也为后文提出新的车型识别算法提供了理论支持。本章最后引入一个新的汽车车型识别算法，多阶段学习神经网路（MS-CNN），网络在 ResNet-50 的基础上对网络进行优化，加入空间金字塔池化层以及 1*1 卷积，实验证明，网络中加入的空间金字塔池化能加快网络的收敛，并且 1*1 卷积与全局平均池化的运用有定位的功能，网络最后引入权重因子 λ，对损失进行联合训练，经过实验，实验证明此网络能提取到更丰富的语义特征与空间信息。与 VGG-16、ResNet-50 以及双线性卷积神经网络（B-CNN）相比能得到相对较高的识别精度。

结　论

本文以深度学习为基础，对汽车车型精确识别进行研究，旨在解决在复杂背景、大规模、多角度的汽车车型的精确识别问题，本文为解决这一问题，首先构建分析并提出了一个更为合理的汽车数据集 MTV-1638s，该数据集针对中国市场，其中包括了中国市场常见的 1638 款车型。在此基础上基于 ResNet-50 提出多阶段学习网络 MS-CNN 车型识别方法，数值实验表明该算法可进行高效车型识别。

数据集 MTV-1638s 是在 Comapcars 数据集基础上进行改进，该数据集由中国香港收集，数据集中包含的汽车车型比较符合中国的汽车保有情况，更具有现实利用价值，但是研究发现，此数据集还存在一些缺陷，包括数据划分不合理以及数据的时效性比较低，在数据划分上，Compcars 没有严格按照某个标准来划分车型，例如对于同一车型的两厢车与三厢车，以及对于同一车型的加长款与普通款划分为不同的车型，这种划分方式给汽车车型的识别带来很大挑战，并且数据集的很多车型停留在 2015 年前，从而降低了数据集的时效性。因而本文构建了更为合理的数据集 MTV-1638s，在此数据集对汽车属性进行实验，实验发现训练库中单一车型样本数量及角度对识别效果影响较大。为此，本文创建了一个均匀角度采样车型数据集 ASMTV120s，此数据集的角度均匀分布，且每个车型下的图片数目固定，减少了随机性因素，利用该数据集研究了汽车角度分布对识别率的影响，最后实验表明：训练集中单车型仅包含不多于 18 个指定角度的样本图片即可达到同等识别效果，并可减少 41.18%的训练时长，有效降低计算成本。

此外，本文提出了加强高阶信息的多阶段神经网络 MS-CNN 车型识别方法。本文在 ResNet-50 基础上，加入空间金字塔池化层、1*1 卷积层与全局平均池化层，提出一个加强高阶信息的多阶段神经网络 MS-CNN。此框架能得到比较丰富的几何特征，加入的空间金字塔池化层加快了网络的收敛，加入的 1*1 卷积层相当于定位功能，进一步加深网络的同时，采用降维的方式，对特征进行压缩，并加强了特征与分类之间的关系，通过可视化，验证了此网络能学习到更加丰富的语义信息，经过实验证明，该网络与 VGG-16、ResNet-50 及 B-CNN 几种经典网络相比该网络能得到更高的识别精度。

本文的研究还存在很多不足，未来研究可以从以下几个方面进一步完善：

首先，本文实验运用的数据集中的汽车图片大都包含单一车辆，或含有相同的汽车车型，这在一定程度上降低了识别难度，未来的研究可以考虑其它比较复杂的场景，比如同一张图片包含多个车辆、存在遮挡、光照等情形。

本文使用的 MTV-Cars 与 MTV-1638s 数据集数以万计，并且运用的神经网络深度也很深，计算参数量十分宏大，耗费很多的时间与计算资源，未来可以考虑在不影响识

别效果的基础上，进一步压缩网络，从而降低参数量，提高网络的计算速度。

本文收集的数据集 MTV-1638s 与 ASMTV120s 可以实现定时更新，以确保数据集的实时性，还可对数据集进行标注，从而增加它的使用价值。

……

参 考 文 献

[1] SARFRAZ M S, SHAHZAD A, ELAHI M A, et al. Real-time automatic license plate recognition for CCTV forensic applications [J]. Journal of Real-Time Image Processing, 2013, 8(3): 285–295.

[2] WANG C M, LIU J H. License plate recognition system [C]//2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015. 2016.

[3] LI H, WANG P, SHEN C. Toward End-to-End Car License Plate Detection and Recognition With Deep Neural Networks [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(3): 1126–1136.

[4] HAN J, YAO J, ZHAO J, et al. Multi-Oriented and Scale-Invariant License Plate
 Detection Based on Convolutional Neural Networks [J]. Sensors, 2019, 19(5): 1175.

[5] LIU X, B W L, MEI T, et al. A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance [J]. 2016, 9905: 869–884.

[6] ZHOU Y, LIU L, SHAO L. Vehicle Re-Identification by Deep Hidden Multi-View Inference [J]. IEEE Transactions on Image Processing, 2018, 27(7): 3275–3287.

[7] CUI C, SANG N, GAO C, et al. Vehicle re-identification by fusing multiple deep neural networks [C]//Proceedings of the 7th International Conference on Image Processing Theory, Tools and Applications, IPTA 2017.

[8] ZHU J, ZENG H, DU Y, et al. Joint feature and similarity deep learning for vehicle re-identification [J]. IEEE Access, 2018, 6: 43724–43731.

[9] SHI K, BAO H, MA N. Forward Vehicle Detection Based on Incremental Learning and Fast R-CNN [C]//13th International Conference on Computational Intelligence and Security (CIS). 2017.

[10] WEI Y, TIAN Q, GUO J, et al. Multi-vehicle detection algorithm through combining Harr and HOG features [J]. Mathematics and Computers in Simulation, 2019, 155: 130–145.

[11] ZHOU H, WEI L, LIM C P, et al. Robust vehicle detection in aerial images using bag-of-words and orientation aware scanning [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(12): 7074–7085.

[12] RUJIKIETGUMJORN S, WATCHARAPINCHAI N. Vehicle detection with sub-class training using R-CNN for the UA-DETRAC benchmark [C]//2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017. 2017.

[13] FAN Q, BROWN L, SMITH J. A closer look at Faster R-CNN for vehicle detection [C]//IEEE Intelligent Vehicles Symposium, Proceedings. 2016.

[14] ZHENG H, FU J, MEI T. Look Closer to See Better : Recurrent Attention Convolutional Neural Network for Fine-grained Image Recognition [C]//30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2017.

[15] PENG Y, HE X, ZHAO J. Object-part attention model for fine-grained image classification [J]. IEEE Transactions on Image Processing, 2018, 27(3): 1487–1500.

[16] KUANG Z, YU J, LI Z, et al. Integrating multi-level deep learning and concept ontology for large-scale visual recognition [J]. Pattern Recognition, 2018, 78: 198–214.

[17] ZHANG X, XIONG H, ZHOU W, et al. Fused One-vs-All Features with Semantic Alignments for Fine-Grained Visual Categorization [J]. IEEE Transactions on Image Processing, 2016, 25(2): 878–892.
[18] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//Conference on Computer Vision and Pattern Recognition. 2005.
[19] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks [C]//ImageNet Classification with Deep Convolutional Neural Networks. 2012.
[20] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition [J]. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2018, 40(6): 1309–1322.
[21] ZHANG N, DONAHUE J, GIRSHICK R, et al. Part-based R-CNNs for fine-grained category detection [C]//13th European Conference on Computer Vision (ECCV). 2014.
[22] FANG J, ZHOU Y, YU Y, et al. Fine-Grained Vehicle Model Recognition Using A Coarse-to-Fine Convolutional Neural Network Architecture [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(7): 1782–1792.
[23] DAI X, SOUTHALL B, TRINH N, et al. Efficient Fine-Grained Classification and Part Localization Using One Compact Network [C]//16th IEEE International Conference on Computer Vision (ICCV). 2017.
[24] WANG Y, MORARIU V I, DAVIS L S. Learning a Discriminative Filter Bank within a CNN for Fine-grained Recognition [C]//31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
[25] ZHANG X, ZHOU F, LIN Y, et al. Embedding Label Structures for Fine-Grained Feature Representation [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
[26] ZHOU F, LIN Y. Fine-Grained Image Classification by Exploring Bipartite-Graph Labels [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
[27] GHASSEMI S, FIANDROTTI A, MAGLI E, et al. Fine-grained vehicle classification using deep residual networks with multiscale attention windows [C]//19th IEEE International Workshop on Multimedia Signal Processing (MMSP). 2017.
[28] GHASSEMI S, FIANDROTTI A, CAIMOTTI E, et al. Vehicle joint make and model recognition with multiscale attention windows [J]. Signal Processing: Image Communication, Elsevier Ltd, 2019, 72(July 2018): 69–79.
[29] KRAUSE J, STARK M, DENG J, et al. 3D object representations for fine-grained categorization [C]//Proceedings of the IEEE International Conference on Computer Vision. 2013.
[30] YANG L, LUO P, LOY C C, et al. A large-scale car dataset for fine-grained categorization and verification [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2015.
[31] XIE S, YANG T, WANG X, et al. Hyper-class augmented and regularized deep learning for fine-grained image classification [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2015.

[32] LIU H, TIAN Y, WANG Y, et al. Deep Relative Distance Learning: Tell the Difference between Similar Vehicles [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
[33] SOCHOR J, SPANHEL J, HEROUT A. BoxCars: Improving Fine-Grained Recognition of Vehicles Using 3-D Bounding Boxes in Traffic Surveillance [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(1): 97–108.
[34] SOCHOR J, HEROUT A, HAVEL J. BoxCars: 3D Boxes as CNN Input for Improved Fine-Grained Vehicle Recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
[35] TAFAZZOLI F, FRIGUI H, NISHIYAMA K. A Large and Diverse Dataset for Improved Vehicle Make and Model Recognition [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2017.
[36] ZHANG Q, ZHUO L, LI J, et al. Vehicle color recognition using Multiple-Layer Feature Representations of lightweight convolutional neural network [J]. Signal Processing, Elsevier B.V., 2018, 147: 146–153.
[37] KIM K J, KIM P K, LIM K T, et al. Vehicle Color Recognition via Representative Color Region Extraction and Convolutional Neural Network [C]//2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN). IEEE, 2018.
[38] ZHUO L, ZHANG Q. High-accuracy vehicle color recognition using hierarchical fine-tuning strategy for urban surveillance videos [J]. Journal of Electronic Imaging, 2018, 27(05): 1.
[39] SUN C, SHRIVASTAVA A, SINGH S, et al. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era [C]//Proceedings of the IEEE International Conference on Computer Vision. 2017.
[40] RANJAN R, PATEL V M, CHELLAPPA R. HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(1): 121–135.
[41] ABDALMAGEED W, WU Y, RAWLS S, et al. Face recognition using deep multi-pose representations [C]//2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016. 2016.
[42] STAROVOITOV V, SAMAL D, BRILIUK D. Image enhancement for face recognition [C]//International Conference on Digital Image Computing - Techniques and Applications. IEEE, 2018.
[43] KAUR R, HIMANSHI E. Face recognition using Principal Component Analysis [C]//2015 IEEE International Advance Computing Conferen. IEEE, 2015.
[44] DONG Z, WU Y, PEI M, et al. Vehicle Type Classification Using a Semisupervised Convolutional Neural Network [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(4): 2247–2256.
[45] TANG Y, ZHANG C, GU R, et al. Vehicle detection and recognition for intelligent traffic surveillance system [J]. Multimedia Tools and Applications, 2017, 76(4): 5817–5832.

[46] BIGLARI M, SOLEIMANI A, HASSANPOUR H. A Cascaded Part-Based System for Fine-Grained Vehicle Classification [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(1): 273–283.
[47] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. arXiv preprint, 2014: 1–14.
[48] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.
[49] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [J]. arXiv preprint, 2015.
[50] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]//Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017.
[51] LIN M, CHEN Q, YAN S. Network In Network [J]. arXiv preprint, 2013: 10.
[52] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916.
[53] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2015, 39(4): 640–651.
[54] KIM P K, LIM K T. Vehicle Type Classification Using Bagging and Convolutional Neural Network on Multi View Surveillance Image [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2017.
[55] DONG Z, WU Y, PEI M, et al. Vehicle Type Classification Using a Semisupervised Convolutional Neural Network [J]. IEEE Transactions on Intelligent Transportation Systems, IEEE, 2015, 16(4): 2247–2256.
[56] ZHUO L, JIANG L, ZHU Z, et al. Vehicle classification for large-scale traffic surveillance videos using Convolutional Neural Networks [J]. Machine Vision and Applications, 2017, 28(7): 793–802.
[57] WANG X, ZHANG W, WU X, et al. Real-time vehicle type classification with deep convolutional neural networks [J]. Journal of Real-Time Image Processing, Springer Berlin Heidelberg, 2017, 16(1): 1–10.
[58] GAO Y, LEE H J. Vehicle make recognition based on convolutional neural network [C]//2015 IEEE 2nd International Conference on InformationScience and Security, ICISS 2015. IEEE, 2015.
[59] AL-MAADEED S, BOUBEZARI R, KUNHOTH S, et al. Robust feature point detectors for car make recognition [J]. Computers in Industry, Elsevier, 2018, 100(April): 129–136.
[60] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2014.
[61] POLYAK B T. Some methods of speeding up the convergence of iteration methods [J]. USSR Computational Mathematics and Mathematical Physics, 1964, 4(5): 1–17.

[62] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[C]//International Conference on Learning Representations (ICRL). 2015.

[63] CAI S, ZUO W, ZHANG L. Higher-Order Integration of Hierarchical Convolutional Activations for Fine-Grained Visual Categorization [C]//16th IEEE International Conference on Computer Vision (ICCV). 2017.

致　　谢

时光悄然流逝，又是一季花开时，研究生生活即将在此画上句点，两年的研究生生活要结束了，在这两年里，有幸得到老师们的教诲，得到同学的帮助与支持，认识志同道合的同学，一同学习与进步，我感到由衷高兴。

本论文在导师李宝军副教授的指导下完成，李老师严谨的治学态度以及追求极致的科研精神也深深触动我，让我受益匪浅。在我的生活上，李老师对问题独特的见解与深刻认知也让我醍醐灌顶，给了我很大的鼓舞与前进的力量，李老师是我们的导师，更像是我们的家人，让我们在课题组也体会到与家人相处的温暖。再次衷心感谢李老师在我读研期间给予的关怀，也祝愿李老师桃李满天下，一切顺心如意。

感谢大连理工大学这个宝贵平台给我们提供的学习环境与资源，感谢课题组博士师兄杨磊对我研究方向提出的宝贵意见与指导，感谢师兄董颖在我实验过程中提出的宝贵经验与实验指导。感谢陈峰蔚、师弟孙旭生对我的帮助。

感谢师兄赵天鹏、陶凯、王赫庭，感谢师姐武捷以及王毅、姜涛，师弟王晨、王红日。同时感谢徐晗同学，感谢教研室 523 每一位同学，对我学习以及生活上的帮助与支持，感谢我的舍友对我的照顾与包容。

最后感谢我的家人，他们是我研究生生涯的坚实后盾，谢谢他们一路默默支持我的每个决定，感谢他们对我无微不至的照顾让我能全身心投入学习，顺利完成学业。

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：基于深度学习的汽车车型识别关键问题研究

作者签名：吴双双　日期：2019 年 6 月 10 日

导师签名：　日期：2019 年 6 月 8 日