

上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

计算机视觉课程报告



题目： 传统算法与神经网络融合的车牌视觉识别

评分： _____

学生姓名： 潘文峥

学生学号： 520030910232

专 业： 计算机科学与技术（IEEE 试点班）

学院(系)： 电子信息与电气工程学院

目 录

第一章 绪论-----	1
1.1 任务要求分析-----	1
1.2 现有算法调研-----	1
第二章 传统方法与神经网络融合的车牌检识算法-----	2
2.1 算法整体架构-----	2
2.2 车牌定位-----	2
2.2.1 基于色彩特征和形态学变换的图像预提取-----	2
2.2.2 基于几何特征的区域筛选-----	3
2.2.3 基于透视投影变换的图像立体矫正-----	4
2.2.4 车牌颜色判别与尺寸重置-----	4
2.3 基于像素直方图投影的字符分割-----	5
2.3.1 基于直方图波峰的区域判定-----	5
2.3.2 边缘、分隔符与干扰点的去除-----	5
2.4 基于边缘连通域特征的字符分割-----	6
2.4.1 基于形态学变换的字符预处理-----	6
2.4.2 连通域的边缘判定与筛选-----	6
2.5 牌号识别-----	7
2.5.1 卷积神经网络的结构设计-----	7
2.5.2 数据集的搭建与划分-----	7
2.5.3 训练、测试与参数调优-----	8
第三章 测试结果与算法评价-----	9
3.1 测试结果可视化-----	9
3.2 算法评价与总结-----	12
参考文献-----	13
谢辞-----	13

第一章 绪论

车牌识别是一种图像处理技术，用于识别不同车辆的牌号，这项技术被广泛用于各种场景检测中。一般来说，进行车牌识别的理论基础是图像分割和图像识别理论：首先对含有车辆号牌的图像进行分析处理，从而确定牌照在图像中的位置，并把牌照区域提取出来，再进一步识别上面的文本字符。车牌识别过程包括图像采集、预处理、车牌定位、字符分割、字符识别、结果输出等一系列算法运算。随着计算机视觉领域的发展，车牌检测识别方法得到极大丰富。在不同场景下，传统方法和基于神经网络的深度学习方法各有优劣，本文基于不同阶段的任务特点，对两者进行了融合，在给定测试图像中取得了较好的效果。

1.1 任务要求分析

本项目需要设计算法完成对给定图像中车牌位置的检测，进而识别车牌号（包括中文部分），并通过恰当的方式对检测过程以及识别结果进行可视化。给定的测试样本被划分为三个难度等级，其中 `easy` 和 `medium` 难度图片的车牌均正对相机，区别在于车牌是否已经从环境中提取，`difficult` 难度提供的车牌则与相机有倾角，贴近实际场景。

通过分析，本项目任务中至少应考虑以下关键点：

- （1） 如何从复杂环境中检测并准确定位车牌，排除背景干扰；
- （2） 两种颜色（蓝牌、绿牌）车牌的色彩、尺寸、字符字体均有差异，如何准确进行判别；
- （3） 对于非正面拍摄的车牌照，如何进行立体矫正得到正对试图；
- （4） 如何对检测和矫正后的车牌进行准确分割，得到利于识别的单字符图像；
- （5） 对于分割好的字符，如何准确进行识别，尤其避免相似字符混淆的情形（如：“6”和“G”、“8”和“B”、“0”和“D”等）

1.2 现有算法调研

车牌检测识别是经典的计算机视觉任务，在长时间发展中产生了诸多算法流程。通过对现有的算法进行的调研和归纳，该问题解决方案可归为以下两种主流解决思路：

- （1） 基于成熟的目标检测、序列建模的神经网络模型（如 YOLOv5、RCNN 等）直接进行端到端训练，即输入不同背景的含车牌图片，直接得到识别结果的文本序列。该方法的优势是全过程端到端完成，结构较为简明，并且在数据量和算力充足的情况下准确率高；但考虑到带序列标注的数据集缺乏，且个人电脑的算力资源有限，这种纯深度学习方法在本项目中并不适用。
- （2） 基于传统计算机视觉、数字图像处理经典算法，对输入图像进行一系列像素级处理（形态学变换、色彩阈值筛选、边缘检测、连通域识别等）、矫正、分割，最后使用模板匹配或支持向量机（SVM）进行字符分类识别。这种方法是典型的传统计算机视觉任务流水线处理，对数据和算力没有特别要求，缺点是对具体样本特征的高度依赖，识别准确度、泛化能力和可扩展性较差。

第二章 传统方法与神经网络融合的车牌检测算法

2.1 算法整体架构

基于第一章对现有方法的调研和评估,本项目设计了传统算法与深度学习相结合的车牌检测和识别算法,将项目任务分为三个大模块,利用经典的视觉算法完成对图中车牌的定位、裁切、矫正和字符分割,并通过训练 CNN 卷积神经网络进行对字符内容的识别。

算法的完整框架和执行流程如下图所示:

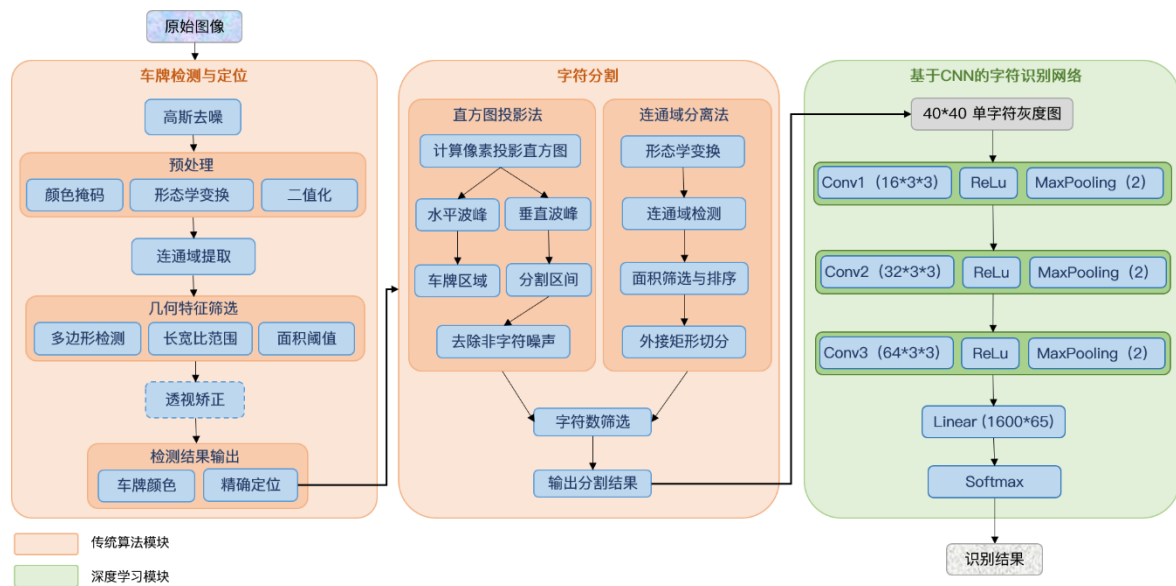


图 2-1 算法整体架构示意图

2.2 车牌定位

2.2.1 基于色彩特征和形态学变换的图像预提取

作为预处理的第一步,我们首先对输入图像进行高斯滤波,去除少量噪声点。对模糊处理后的图片,通过颜色特征对可能的车牌区域进行预筛选。考虑到现行中国车牌主要有蓝牌和绿牌两大类,我们对其色彩标准进行了调研和实际检测,决定同时采用 RGB 和 HSV 空间的阈值限定来对图像进行掩码二值化,具体限定范围如下:

蓝牌: $B \in [130, 255]$, $G \in [0, 120]$, $R \in [0, 70]$, $H \in [99, 124]$, $S \in [90, 255]$

绿牌: $B \in [80, 180]$, $G \in [140, 200]$, $R \in [0, 130]$, $H \in [35, 99]$, $S \in [35, 255]$

进一步,我们对提取出的二值图像进行形态学膨胀(dilate)操作,选用横向分量较大的核进行变换,使车牌上的字符尽可能连成整体,便于接下来的筛选操作。在理想情况下,预处理可以得到如图 2-4 展示的效果。



图 2-2 原始图像

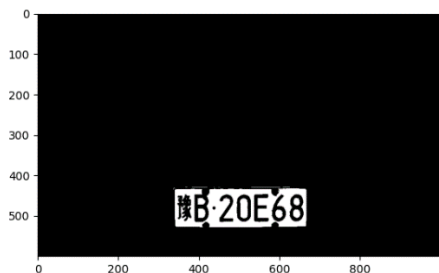


图 2-3 颜色掩码后

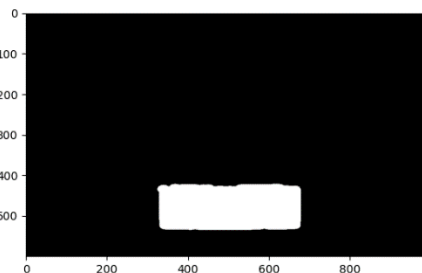


图 2-4 形态学变换后

2.2.2 基于几何特征的区域筛选

可以看到，经过预提取后，输入图像中的车牌部分已经得到基本锁定。但是，我们在测试中发现，在一些背景和光照条件复杂的输入图片中，预提取的结果仍包含相当一部分非车牌噪声，如树叶、挡风玻璃等与车牌颜色特征相近的部分被置为白色。因此，我们采取对几何特征作进一步限定的方式来过滤噪声。

具体地，我们对找到的连通域轮廓（cv2.findContours 函数实现）作如下限定：

- (1) 连通域可被多边形逼近算法拟合为面积不小于 1000 的四边形；
- (2) 四面形的长宽比应在 $[1.5, 6]$ 范围之内。



图 2-5 原始图像

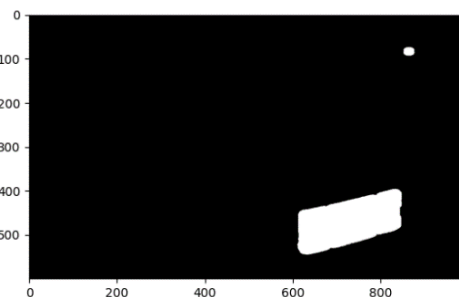


图 2-6 预提取后

如图 2-6 所示，部分图片经预处理后，还存在部分非车牌噪声（右上角白点），图 2-7 展示了经过几何特征的区域筛选后得到的车牌定位框。此时采取的框定方法是找到具有最大面积的连通域，用其最小外接矩形进行框选。



图 2-7 几何筛选后

2.2.3 基于透视投影变换的图像立体矫正

对于 difficult 难度等级的图片，在进行上述车牌定位操作后，我们得到的裁切后的车牌呈平行四边形而非矩形，这给后续进行字符切分造成了困难。因此，我们需要对车牌进行透视矫正。首先，从检测出的二值化图像中找到平行四边形的四个顶点（利用二维坐标大小关系即可），再通过透视变换算法计算出的变换矩阵对车牌区域进行校正，得到正对拍摄者效果的车牌，矫正前后如图 2-8，图 2-9 所示。



图 2-8 检测出的车牌



图 2-9 矫正后的车牌

2.2.4 车牌颜色判别与尺寸重置

考虑到两种颜色车牌的长宽比、字符位置和字体均存在一定的差异，我们对检测到的车牌图片计算其 RGB 色彩均值，以 B 通道和 G 通道均值的大小关系作为车牌颜色分类的判据。根据我国机动车牌照尺寸的国家标准，蓝牌尺寸为 440mm*140mm（有效字符 7 位），绿牌尺寸为 480mm*140mm（有效字符 8 位），我们将裁切、矫正后的车牌打上 color 标签，并按相应长宽比将蓝色和绿色车牌分别 resize 到（220,70）和（240,70）像素，以减小后续投入识别网络的字符与真实牌照上的字符之间存在的畸变。



图 2-10 Resize 后的蓝牌



图 2-11 Resize 后的绿牌

容易发现，与矫正后的车牌相比，尺寸重置后牌照上的字符更贴近现实车牌的字符形态，为后续切割和识别阶段提供了便利，增强了算法的稳定性。

2.3 基于像素直方图投影的字符分割

对于字符分割模块，在计算机视觉领域有两种主流做法，分别基于像素直方图投影和形态学变换后的连通域检测。在本项目中，为了取得更优的任务表现，我们将两种方法都进行了实现，并通过一系列校验方法进行选优后输出最终分割结果。下面首先介绍第一种方法的具体流程。

2.3.1 基于直方图波峰的区域判定

首先，传入 `segment` 函数的参数不仅有裁切好的车牌图像，还有车牌对应的颜色标签。在进行处理之前，当传入一张绿色车牌时，先对其颜色进行翻转，这是因为绿牌车的字符是黑色，比背景更暗，而蓝牌则与此相反。

接下来，我们对二值化的车牌分别逐行、逐列求和，得到水平和垂直两个方向上的像素投影直方图。对每个方向的直方图设置阈值，找出波峰和波峰宽度，其中水平方向最大的波峰被认为是车牌区域，垂直方向的波峰则作为字符分割的候选结果。

记 x_{min}, x_{avg} 分别表示水平方向直方图统计结果的最小值和平均值， y_{min}, y_{avg} 分别表示垂直方向直方图统计结果的最小值和平均值，根据经验和反复调试，我们将水平和垂直方向的阈值分别设置为

$$x_{threshold} = \frac{x_{min} + x_{avg}}{2}$$

$$y_{threshold} = \frac{y_{min} + y_{avg}}{6}$$

可以取得较好的分割效果，同时尽可能避免同一字符被分割为两个区域。



图 2-12 分割前的二值图



图 2-13 原始分割结果

2.3.2 边缘、分隔符与干扰点的去除

从图 2-13 可以明显看到，以上处理可以基本完成对字符的分割操作，但输入图像往往存在边框干扰，且车牌地区和号码之间的分隔点也被单独作为字符分割出来。为解决这个问题，我们进一步对上述操作得到的垂直波峰宽度进行筛选。

- (1) 对左右边框的判据：分割块 x 方向宽度小于最大块宽度的三分之一；
- (2) 对分隔点的判据：分割块 x 方向宽度小于最大块宽度的三分之一，且二值图中色彩均值小于 $255/4$ （白色面积小于矩形框的四分之一）。



图 2-14 去噪后的分割结果

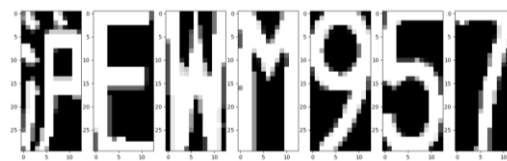


图 2-15 最终分割结果

对分割出的单字符按照标准车牌的尺寸进行 `resize`，得到最终的结果。

2.4 基于边缘连通域特征的字符分割

本节介绍前述的第二种字符分割方法，即先通过形态学变化，使得车牌中的每个字符尽可能连成一个整体，进而通过对连通域的框定实现对字符的切分。

2.4.1 基于形态学变换的字符预处理

首先，我们对输入的二值图像进行膨胀操作（图 2-17），使得车牌上每个字符加粗，并且使得具有左右结构的汉字的两部分产生粘连。

这里的关键点在于，设置恰当的核参数，使得达到上述目的的同时，尽可能避免不同字符被连成一个整体。在本项目中，经过经验和调试，我们采用的 `kernel_size` 为 (4, 9)，即纵向膨胀大于横向膨胀。



图 2-16 分割前的二值图



图 2-17 形态学膨胀后的结果

2.4.2 连通域的边缘判定与筛选

与此前相同，我们依然利用 `cv2.findContours` 函数找出膨胀后二值图像中的连通区域，再利用 `cv2.boundingRect` 函数可以找到每个连通域的最小正矩形（与 `cv2.minAreaRect` 函数不同，后者找到的最小外接矩形允许存在倾角），得到的结果如图 2-14 所示。



图 2-14 原始分割结果

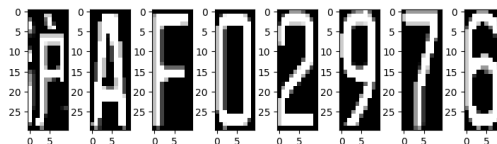


图 2-19 最终分割结果

可以看到，本例中原始分割结果中存在一些小块的错误分区，而在另一些测试样例中，还容易出现因铆钉存在而导致字符与边框相连，从而框选出整个车牌的情况。针对以上问题，我们设计如下筛选和判定方案：

- （1）有效字符框矩形面积不大于 20000（避免全域被框选）；
- （2）符合（1）的区域按面积降序排列，取前 7 个（蓝牌）或前 8（绿牌）个元素。

经筛选和排序得到的最终分割结果如图 2-19 所示，有效达到了精准分割单字符图像的目的。

需要补充说明的是，考虑到字符分割是本项目的重难点，测试中常出现分割效果不鲁棒的问题，因此，2.3 节和 2.4 节提出的两种字符分割方案都在本项目中实现并应用，如果方法一分割结果的字符数与实际颜色车牌不匹配，算法会采用方法二重新进行分割，达到选优的目的，增强算法的稳定性。

2.5 牌号识别

经过前述一系列预处理、车牌定位、透视矫正和字符切割，我们得到了给定图像中车牌的每个字符的灰度正视图。至此，我们需要解决从图像中精确识别每个字母、数字或汉字。在深度学习高度发展的今天，容易想到使用经典的卷积神经网络设计图像分类器解决问题。

2.5.1 卷积神经网络的结构设计

从 LeNet 诞生至今，CNN 已经成熟并广泛应用于各类计算机视觉任务当中，随着网络深度和参数量的上升，其准确率也得到极大飞跃。相应地，更大的网络模型对带标注数据和计算资源的需求也十分巨大。

基于这样的思考，我们将本项目的识别任务考虑成一个图像 65 分类问题，标签包括 10 个数字（0~9），24 个字母（不含 I 和 O）和 31 个省级行政区简称汉字（不含港澳台）。我们尝试过使用经典的 CNN 模型（如 ResNet18 等），但一方面由于模型参数量过于庞大，在有限数据和算力支持下难以在短时间内训练至充分收敛；另一方面，本问题的分类类别较少，输入图像较为简单且高度规范（黑色背景的单字符灰度图），我们经过尝试探索，最终选择设计了一个轻量的三层卷积神经网络，并取得了理想的识别效果（测试结果见第三章）。

模型结构如下：

表2-1 CNN网络结构和参数

Blocks	Layers with hyper-parameters
Conv_1	Conv2d(in_channels=1, out_channels=16, kernel_size=3, stride=1, padding=1), ReLU(), MaxPool2d(kernel_size=2))
Conv_2	Conv2d(in_channels=16, out_channels=32, kernel_size=3, stride=1, padding=1), ReLU(), MaxPool2d(kernel_size=2))
Conv_3	Conv2d(in_channels=32, out_channels=64, kernel_size=3, stride=1, padding=1), ReLU(), MaxPool2d(kernel_size=2))
Linear	Linear (in_features=64 * 5 * 5, out_features=65)

说明：输入图像为 1*40*40 灰度图的张量，经三层卷积和最大池化降采样后，特征图尺寸变为 64*5*5（=1600），将其展平为向量后经过全连接层输出 65 分类结果。

2.5.2 数据集的搭建与划分

本项目采用的数据集来源于百度飞桨 AI studio 的开源数据集 VehicleLicense，原数据集包含 16151 张单字符数据，根据本项目任务特点，对该数据集进行了增补和筛选，最终保留 15496 张，所有的单字符均为严格切割且都转换为黑白二值图像。

为便于训练，我们将 65 类标签从字符串转换为整数，并保存了对应映射关系，便于输出最终结果时与实际车牌内容一致。最终提交的版本中，默认将 85% 的数据（13172 张）用于训练，15% 的数据（2324 张）用于测试，训练集和测试集的划分是随机的，每次读取一个小批量（默认 batch_size=32）进行训练。

2.5.3 训练、测试与参数调优

根据经验和调试，我们选用 Adam 优化器对模型参数进行训练优化，其他超参数默认设置如表 2-2 所示，以下参数名与命令行参数名保持一致。

表2-2 训练超参数设置

Hyper-parameters	Explanation	Tuned value
train_ratio	训练集占数据集的比例	0.85
num_epochs	最大训练轮数	100
batch_size	批大小	32
lr	学习率	2×10^{-4}
wd	权重衰减(正则化参数)	5×10^{-4}
lr-period	学习率调整间隔	10
lr_decay	学习率衰减参数	0.95

在每个训练轮次（epoch）中，我们分别记录下模型在训练集和测试集的表现，当在测试集上预测的正确率出现新高时，保存当轮模型参数。由于项目任务较为简单，我们省略了惯常需要的验证集。模型在训练过程中，每个 global step 的平均 loss 值和每个 epoch 的测试集上预测正确率变化曲线如图 2-20、图 2-21 所示。

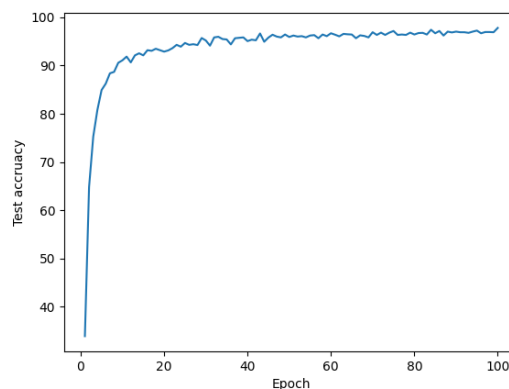


图 2-20 测试集上预测正确率曲线

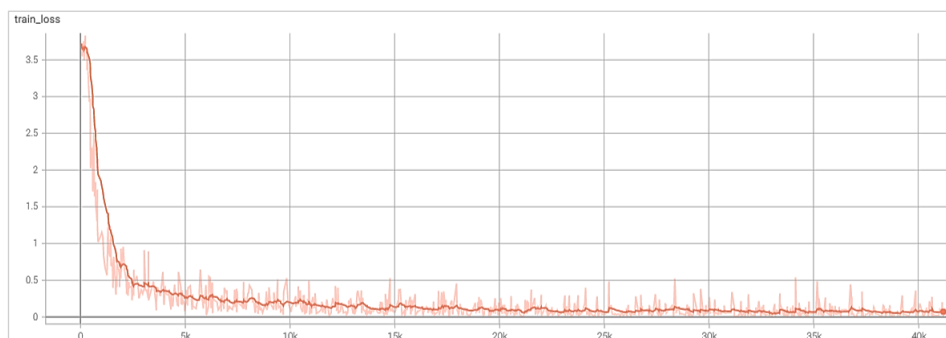


图 2-21 训练 loss 曲线

最终，我们采用测试集上正确率最高（96.6182%）的模型参数用于实验验证。

第三章 测试结果与算法评价

3.1 测试结果可视化

(1) Easy 级别:

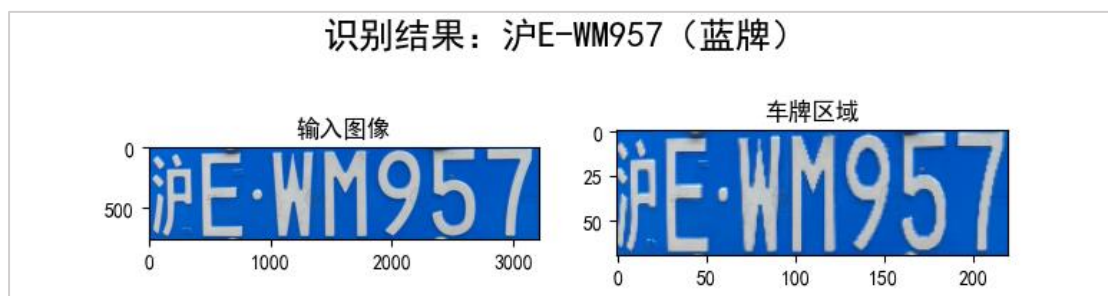


图 3-1 easy/1-1.jpg 识别结果



图 3-2 easy/1-2.jpg 识别结果

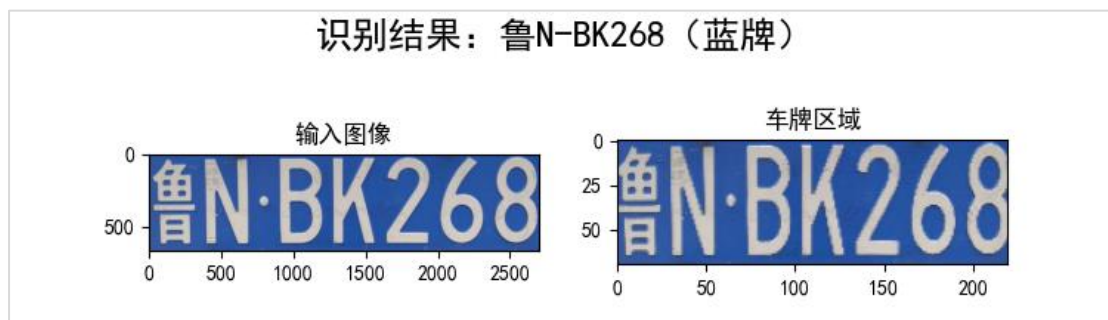


图 3-3 easy/1-3.jpg 识别结果

(2) Medium 级别:



图 3-4 medium/2-1.jpg 识别结果



图 3-5 medium/2-2.jpg 识别结果



图 3-6 medium/2-3.jpg 识别结果

(3) Difficult 级别:



图 3-7 difficult/3-1.jpg 识别结果

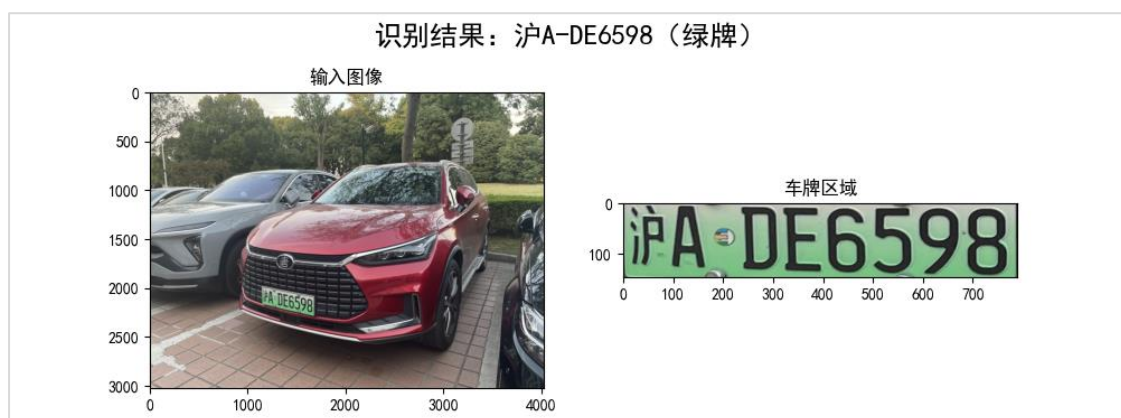


图 3-8 difficult/3-2.jpg 识别结果



图 3-9 difficult/3-3.jpg 识别结果

3.2 算法评价与总结

首先，在给定的三个难度等级 9 张样本上，我们设计的传统方法结合深度学习的车牌识别算法表现良好，整牌识别准确率为 **100%**，按要达到了项目任务的全部要求。

在字符识别的卷积神经网络的训练过程中，对于 VehicleLicense 数据集的 13172 张训练样本和 2324 张测试样本(单字符二值图)，模型表现的最高正确率分别为 **97.97%**和 **97.82%**。

除此以外，我们还探索性地使用其他与给定测试样本类似的图片对模型进行了测试，部分测试结果如下。



图 3-10 部分拓展测试结果

我们发现，在更广阔的场景下，我们提出的算法和模型仍有进一步改进的空间。

在光线、背景和拍摄角度发生变化时，模型的识别效果将受到影响，究其原因主要是传统算法对车牌的定位和提取上误差较大，因为传统方法往往只能设置固定的参数，从而难以具备足够强的自适应性和泛化能力。针对这样的事实，可以展望，在具备充足算力和数据的情况下，对车牌识别问题进行大模型的端到端训练应当能取得更为理想的表现。

至此，车牌检测与识别任务全部圆满完成。作为计算机视觉课程项目，在这一任务中，我对课程所学的理论知识有了更加深入的理解，所涉及的操作和算法包括但不限于图像灰度与二值化、多种形态学变换、边缘检测、多边形检测、透视投影变换、图像校正以及基于卷积神经网络的图像分类模型等。通过查阅文献和动手实践，极大提高了对相关算法的工程能力，在一个相对具体的实际应用场景中体会分析问题、设计方法、调整试验并最终解决问题的整体流程和思维框架，受益匪浅。

参考文献

- [1] 神经网络从 0 到 1——车牌识别, <https://www.guyuehome.com/13863>
- [2] 车牌识别之车牌定位, <https://zhuanlan.zhihu.com/p/102203294>
- [3] https://blog.csdn.net/qq_44032245/article/details/94772746?spm=1001.2014.3001.5502
- [4] VehicleLicense 车牌识别数据集: <https://aistudio.baidu.com/aistudio/datasetdetail/56280>

谢辞

通过本学期的授课和项目实践,我对计算机视觉这一极其重要的计算机科学领域有了更加深入的认识,对其中涉及的经典方法和常用思想有了更完整的把握,这为我在该方向进一步完成探索和研究打下了坚实基础。衷心感谢老师和助教在本课程中给予的指导和帮助!