Computer Science and Application 计算机科学与应用, 2021, 11(11), 2804-2816
Published Online November 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.1111285

Hans汉斯

基于CRNN的车牌识别方法

刘高洪[1,2]，孙博洋[3]，刘宗伟[1]，叶　剑[1,2*]

[1]临沂中科人工智能创新研究院，山东 临沂
[2]中国科学院计算技术研究所，北京
[3]北京建筑大学，北京

收稿日期：2021年10月25日；录用日期：2021年11月22日；发布日期：2021年11月29日

摘　要

车牌识别是道路交通、智慧城市建设的重要组成部分，传统的车牌识别需要先检测出车牌位置，然后通过像素映射等方法分割出单个字符，最后利用模板匹配等方法进行识别。整个过程不仅速度慢，而且操作繁琐，分割或识别的效果也很难令人满意。本文基于YOLOv4-tiny和卷积循环神经网络(Convolution Recurrent Neural Network, CRNN)提出了一种端到端的方法。该方法利用注意力机制与YOLO4-tiny的融合，有效且快速的检测车牌位置，然后利用空间变换网络(Spatial Transformer Networks, STN)、残差学习(Residual Learning)以及注意力机制(Attention)与CRNN的融合高效的识别车牌信息。本文使用平均精度(Average Precision, AP)和识别准确率(Accuracy)作为检测和识别结果的主要评估指标。实验结果表明，车牌检测模型在交并比(Intersection-over-Union, IoU)为0.5的前提下AP值达到了93.60%，并且识别模型在蓝牌、绿牌的混合车牌下达到了92.15%左右的识别准确率。该方法相比于之前的车牌识别模型，不但识别准确率更高，而且能够直接通过该模型识别混合车牌，大大减少了现实情况下车牌识别的复杂度。

关键词

车牌检测，车牌识别，YOLOv4-Tiny，CRNN，STN，残差学习，注意力机制

CRNN-Based License Plate Recognition Method

Gaohong Liu[1,2], Boyang Sun[3], Zongwei Liu[1], Jian Ye[1,2*]

[1]Linyi Artificial Intelligence Innovation Research Institute, Linyi Shandong
[2]Institute of Computing Technology, Chinese Academy of Sciences, Beijing
[3]Beijing University of Civil Engineering and Architecture, Beijing

Received: Oct. 25th, 2021; accepted: Nov. 22nd, 2021; published: Nov. 29th, 2021

*通讯作者。

文章引用: 刘高洪, 孙博洋, 刘宗伟, 叶剑. 基于 CRNN 的车牌识别方法[J]. 计算机科学与应用, 2021, 11(11): 2804-2816.
DOI: 10.12677/csa.2021.1111285

Abstract

License plate recognition is an important part of road traffic and smart city construction. Traditional license plate recognition needs to detect the position of the license plate first, then segment a single character by pixel mapping, and finally use template matching and other methods for recognition. The whole process is not only slow, but also cumbersome to operate, and the effect of segmentation or recognition is difficult to be satisfied. This paper proposes an end-to-end method based on YOLOv4-tiny and Convolution Recurrent Neural Network (CRNN). This method uses the fusion of the attention mechanism and YOLO4-tiny to effectively and quickly detect the position of the license plate, and then uses the spatial transformation network (STN), residual learning, attention mechanism and CRNN to efficiently recognition of license plate information. This article uses Average Precision (AP) and Recognition Accuracy as the main evaluation indicators for detection and recognition results. The experimental results show that the AP value of the license plate detection model reaches 93.60% under the premise that the Intersection-over-Union (IoU) is 0.5, and the recognition accuracy reaches about 92.15% under the mixed license plate of blue and green plates. Compared with the previous license plate recognition model, this method not only has higher recognition accuracy, but also can directly recognize mixed license plates through the model, which greatly reduces the complexity of license plate recognition in real situations.

Keywords

License Plate Detection, License Plate Recognition, YOLOv4-Tiny, CRNN, STN, Residual Learning, Attention Mechanism

1. 引言

近年来随着我国车辆不断的增多，城市路面交通及日常汽车管理的压力也越来越大，伴随着智慧城市建设的不断完善，在复杂环境下对车牌检测及车牌识别的需求也逐步增加。自 2012 年 Hinton 和他的学生 Alex Krizhevsky 提出 AlexNet 以来，卷积神经网络飞速发展，多年来学者们在深度和结构上不断优化，在目标识别、目标检测取得了巨大的突破。通过将卷积神经网络应用到车牌检测及车牌识别中，使得无论是速度还是精度都取得了非常好的效果。

现阶段主要的目标检测算法大致分为两大类，第一类是基于候选框区域的一阶段方法，例如 YOLO [1]、SSD [2]等，这类方法将图像送入网络，在产生候选区域的同时进行物体的类别和位置预测，其特点是速度快而精度低。第二类是基于回归的二阶段方法，Fast R-CNN [3]为具有代表性的一种，该方法首先生成侯选区域，再对候选区域进行类别和位置预测，其特点往往是精度很高但是速度较慢。考虑到实际应用，二阶段的方法无论是内存还是计算量消耗都十分巨大，以至于在使用设备的选择上具有极大的局限性，不适合商用产品实际落地，因此在检测时选择一阶段方法，能够在速度和精度上取得一个较好的平衡。

在检测出车牌位置后，下一步的操作是进行车牌识别，传统的车牌识别分为字符分割和字符识别两

部分，字符分割算法包括基于模板匹配的字符分割方法[4]、基于垂直投影的字符分割方法[5]、基于连通域的字符分割方法[6]以及基于聚类分析的分割方法[7]。通过以上方法将车牌进行分割后下一步对其进行识别，传统的字符识别算法包括基于模板匹配的字符识别算法[8]、基于特征统计的字符识别算法[9]以及基于机器学习的字符识别方法[10]。

但以上方法都极其复杂繁琐，随着深度学习的发展，大量的车牌识别网络孕育而生，Zhao 等人[11]提出使用卷积神经网络(Convolution Neural Network, CNN)利用改进后的 LeNet-5 来自适应的学习特征训练模型从而识别车牌字符；Jain 等人[12]使用 CNN 模型从整张车牌图像提取特征，通过 11 个全连接层将其解码为 11 个定长序列，然后依次分类来得到指定位置的字符；Li 等人[13]尝试通过单个神经网络来同时解决车牌检测与车牌识别的问题；Sergey 等人[14]采用全卷积来取代长短期记忆网络(Long Short-Term Memory, BLSTM)，提出了第一个不使用循环神经网络(Recurrent Neural Network, RNN)的实时车牌识别网络。在本文中，我们将介绍基于深度学习的端到端的车牌检测与识别模型，该模型：1) 有效的将字符识别中的分割与识别合成一步进行。2) 自然的处理任意长度的车牌，包括蓝牌和绿牌。3) 快速且准确的输出车牌框位置信息及车牌号码。通过该方法，使得在实际应用中减少了大量的停车等待时间，同时有效的识别了混合车牌，避免了因混合车牌识别错误而造成的不必要的经济损失。

2. 车牌检测与识别算法

2.1. 基于 YOLOv4-Tiny 的车牌检测方法

2.1.1. YOLOv4-Tiny 介绍

本 2020 年 4 月在 Alexey 等人[15]的研究下新的目标检测算法 YOLOv4 被提出，同年 6 月份，作者推出了 YOLOv4 简化版的模型 YOLOv4-tiny，作为简化版的 YOLOv4，YOLOv4-tiny 整体的网络结构降低了参数所以相对简单，使之成为在移动和嵌入式开发中可行的算法之一。

YOLOv4-tiny 实际使用两个 YOLO 头取代了 YOLOv4 中的三个头，并且 YOLOv4-tiny 在预训练卷积层上使用了 29 层进行训练，相比较 YOLOv4 的 137 个预训练卷积层减少了 108 层。减少了网络的复杂度，在 COCO 数据集上进行测试时，精度相比较 YOLOv4 降低了 1/3，但是随之而来的是 YOLOv4-tiny 中每秒帧数(Frames Per Second, FPS)大约是 YOLOv4-tiny 的八倍。YOLOv4-tiny 模型在 RTX 2080Ti 上以 443FPS 的速度实现了 22% AP，而通过使用 TensorRT、batch size = 4 和 FP16-precision，YOLOv4-tiny 实现了 1774 FPS。

针对本文中涉及的车牌识别的实时性，与 YOLOv4 相比，YOLOv4-tiny 是更好的选择，因为相比较精度或者是准确度来说，更快的检测时间是需要达成的首要目标。

由图 1 所示，YOLOv4-tiny 核心网络(Backbone)中两次 CBL 块主要是通过步长为 2 的卷积对输入进来的图像进行两次下采样，CSP 模块将特征按照通道划分为两组，其中一组在进行正常的卷积操作的同时采用大残差边和小残差边进行合并，这样进行特征提取的好处在于可以通过两个残差边使网络得到更多的语义信息，进而在减少内存开销的同时可以提高网络整体的检测准确率。其中 CBL 是由卷积层、批量归一化(Bath-Norm)和非线性激活函数组成。通过 CBL 避免梯度消失和梯度下降的同时，还可以减少权重初始化的影响，提高网络整体的泛化能力，加快网络训练。

2.1.2. 与注意力机制融合后的车牌检测算法

YOLOv4-tiny 算法在多个尺度的融合特征图上分别独立做检测，在 Pascal VOC、COCO 数据集上检测结果得到很大提升，但在本文车牌检测应用中仍具备优化的潜力，需要对 YOLOv4-tiny 的算法进行改进来适应特定的检测。

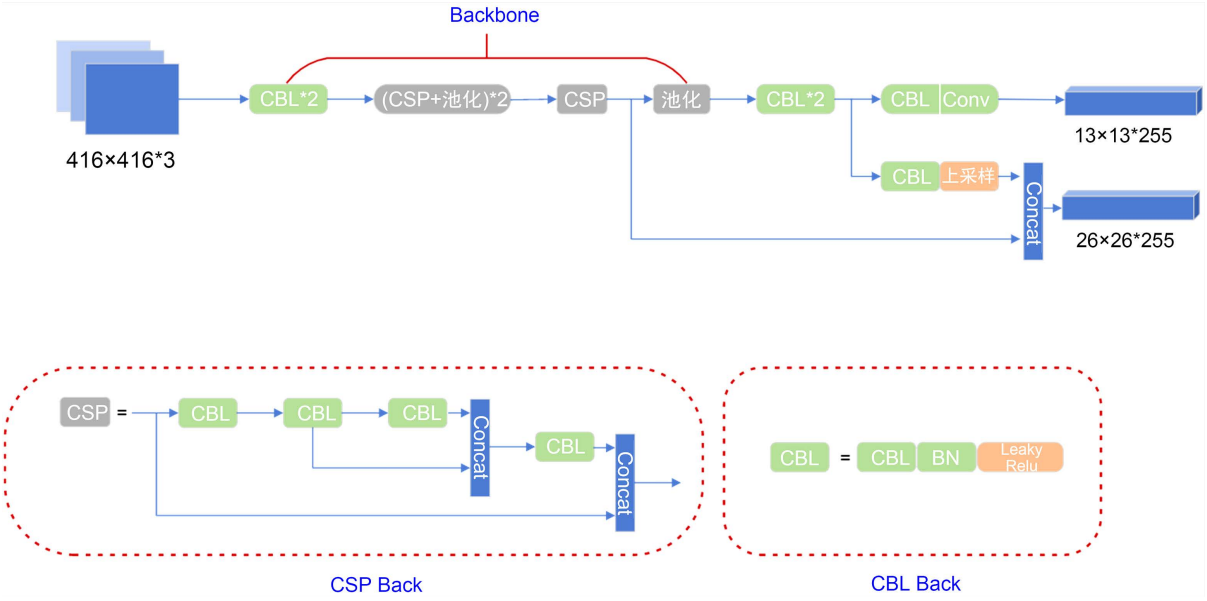

Figure 1. YOLOv4-tiny network diagram
图 1. YOLOv4-tiny 网络图

1) 改进初始框参数

YOLOv4-tiny 可以将输入的 416 × 416 大小的图像分成若干个网络，其中每一个网络都对应着三个先验框，最终将这三个先验框通过网络参数进行调整得到预测的目标框。因此这三个先验框在实际使用中越接近真实框的大小，那么网络整体的误差就越小，收敛也会更快。在 YOLOv4-tiny 中，先验框作者是根据 COCO 数据集，以 IOU 作为距离，通过 K-means [16]算法计算得来的，取值为 10，14，23，27，37，58，81，82，135，169，344，319。

如图 2 所示，对于本文的车牌数据来说，原始的锚框参数由公共数据集 COCO 聚类而成。但是 COCO 数据集类别十分丰富，实际的尺寸太大，所以对应的先验框参数值具有普遍性，不适用于本文的实验要求，所以需要重新进行维度聚类，从而使得本文的模型进行更好的车牌预测。

Figure 2. Target box display
图 2. 目标框展示

由于 YOLO 系列自 v3 开始加入了类似于 FPN 上采样和融合的做法，融合了不同尺度的特征图，YOLOv4 对此又进一步改进了 PAN 层使得特征图尺寸发生了变化。但是本次实验采用的数据集中拍摄的图片距离相近，导致车牌的宽高十分相近，在使用 K-means 聚类后的候选框尺寸集中，为了使得多尺度输出发挥更好的作用，对聚类后的候选框进一步做了线性尺度放缩的方式，将锚框尺寸向两边拉伸，具体拉伸依据如下：

$$x_1' = \alpha x_1 \tag{1}$$

$$x_6' = \beta x_6 \tag{2}$$

$$x_i' = \frac{x_i - x_1}{x_6 - x_1}\left(x_6' - x_1'\right) + x_1' \tag{3}$$

$$y_i' = x_1' \frac{y_i}{x_i} \tag{4}$$

在完成 K-means 聚类和线性拉伸后，先验框为 50，10，168，33，266，53，375，71，498，90，696，123。

2) 引入注意力机制

为了使得网络专注于目标特征，并且忽略非目标特征，研究者在神经网络中加入注意力机制，目前注意力机制依靠其优势已经被广泛应用于自然语言处理、图像处理等任务中，其优势在于只增加少量的计算量的情况下，可以使网络能自动学习到图像或文字中需要注意的地方。注意力机制主要分为：空间注意力机制、通道注意力机制、空间和通道混合注意力机制。针对本文数据多为复杂场景下的车牌检测，注意力机制能够有效的学习车牌中的目标特征，并且抑制其他非目标特征，强调车牌信息，抑制背景信息，提高检验精度。

Woo 等人[17]于 2018 年提出卷积块注意力模型(Convolution Block Attention Module, CBAM)，他们认为目前大多数神经网络主要针对三个因素进行性能提升的方向，分别是深度、宽度和基数，但他们关注的是另一个方面：注意力。它是人类视觉系统的一个有趣的方面。他们将学习通道注意力和空间注意力的过程进行分解，使得注意力生成过程具有更少的计算量和模型参数。因此可以作为已存在的基础卷积架构的即插即用模块。同时，Hu 等人[18]与 Woo 等人的工作非常接近，但是 Hu 等人使用全局池化特征来计算通道上的注意力，并且他们忽略空间注意力的影响。Woo 等人提出空间注意力在决定注意力的位置方面起着重要的作用，在 Woo 等人提出的 CBAM 中，他们基于一种有效的架构同时利用了空间注意力和通道注意力，并且最后效果优于使用全局池化特征的只计算通道注意力的架构。这里我们选择代表空间和通道混合注意力机制的轻量级的通用模块 CBAM。其结构如下图 3 所示：

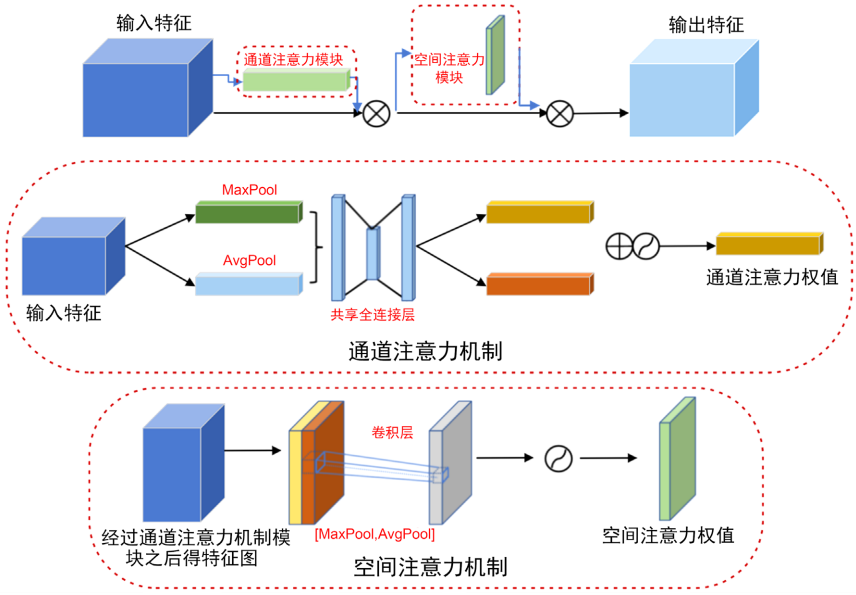

Figure 3. CBAM schematic
图 3. CBAM 示意图

CBAM 模块是一个轻量级和通用的模块，所以 CBAM 模块可以插入到整个网络的卷积模块中，实现端到端的同步训练。如图 3 通道注意力模块所示，特征输入后，分别经过最大池化操作(Max Pooling)和平均池化(Average pooling)后共享全连接层，将输出特征进行加和操作，再经过非线性激活，最终形成通道注意力模块。将通道注意力机制输出结果和原输入的特征图做乘法操作，形成空间注意力模块需要的输入特征。

从另一个角度来看通道注意力机制，其实是将特征图(feature map)在两个维度上做压缩，压缩成一维矢量后再进行后续的卷积等操作。此处卷积使用了最大池化和平均池化两种。两种池化的使用可以解决特征映射不聚合的问题，其中平均池化对特征图中每一个特征点都有相应的反馈，而最大池化在进行反向传播的时候，只响应特征图中最大的梯度。通道注意力机制可以用表达式表达为：

$$\begin{aligned} M_c(F) &= \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}(F)\right)+\mathrm{MLP}\left(\mathrm{MaxPool}(F)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(F_{avg}^c\right)\right)+W_1\left(W_0\left(F_{\max}^c\right)\right)\right) \end{aligned} \tag{5}$$

如上图 3 所示空间注意力机制。在处理完通道注意力模块后将通道注意力模块的输出作为空间注意力模块的输入。与通道注意力机制不同的是，空间注意力机制先做一个基于通道的最大池化和平均池化，再将结果基于通道做一个连接(Contact)操作，然后通过卷积将其降为一个通道后继续经过非线性激活生成空间注意力模块。最后将该输出结果与通道注意力机制的输出进行乘法运算，最终生成想要的特征。

$$\begin{aligned} M_c(F) &= \sigma\left(f^{7\times7}\left(\left[\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)\right]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F_{avg}^s;F_{\max}^s\right]\right)\right) \end{aligned} \tag{6}$$

其中，σ 为 sigmoid 操作，7×7 表示卷积核的大小，7×7 的卷积核比 3×3 的卷积核效果更好。所以 CBAM 的整个处理过程如下式所示：

$$\begin{aligned} F' &= M_c(F)\otimes F, \\ F'' &= M_s(F')\otimes F' \end{aligned} \tag{7}$$

其中，$\otimes$ 表示对应元素逐个相乘。即上述所说输入特征图先进入通道注意力模块生成通道注意力权重的特征图后并与原始输入特征图对应元素相乘得到新的特征图 F'，新的特征图 F' 送入空间注意力模块生成空间注意力权重的特征图 M_s，最终与 F' 逐元素相乘得到特征图 F'' 完成注意力模块。

改进前 CSP 结构如图 4 所示：

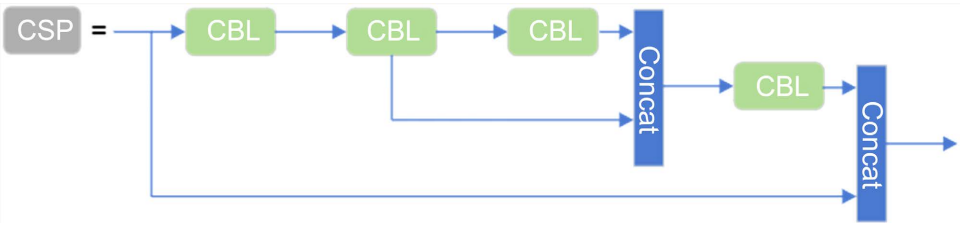

Figure 4. CSP structure
图 4. CSP 结构

改进后 CSP 结构如图 5 所示：

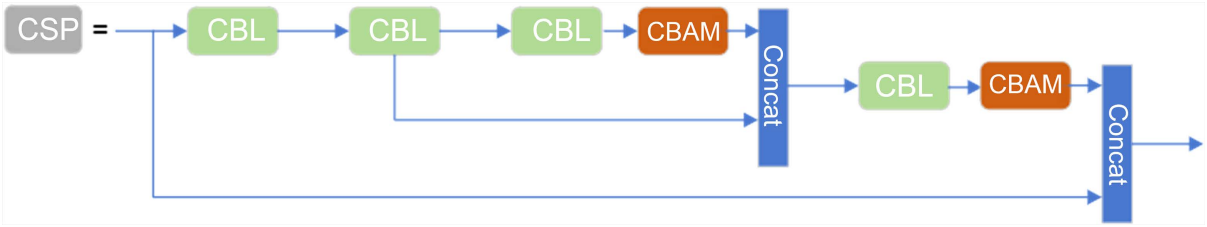

Figure 5. Improved structure of CSP
图 5. CSP 改进后结构

2.2. 基于 CRNN 的车牌识别方法

2.2.1. CRNN 介绍

Shi 等人[19]在 2015 年提出卷积循环神经网络(Convolution Recurrent Neural Network, CRNN)，该方法提出了一种将特征提取、序列建模和转录整合到统一框架中的新型神经网络。该网络结构基于端到端训练；可以自然的处理任意长度的序列，不涉及字符分割或水平尺度归一化；不限于任何预定义的词汇，在无词典和基于词典的场景文本识别任务中都取得显著的表现；产生一个有效且小的多的模型，对于现实场景更为使用。CRNN 由深度卷积神经网络(Deep Convolution Neural Network, DCNN)和循环神经网络(Recurrent Neural Network, RNN)组成。在 CRNN 的底部，卷积层自动从每个输入图像中提取特征序列，然后在循环层使用 stack 形深层双向长短期记忆网络(Bidirectional Long Short-Term Memory, BLSTM)对特征序列的每一帧进行预测，即输入图像在通过 LSTM 后将文字的时序序列进行编码，然后将该编码送入连接时序分类算法(Connectionist Temporal Classification, CTC) [20]进行解码。CTC 作为经典的 OCR 解码算法，假设 CRNN 的循环层的输出维度是 $T \times n$，令 $T = 25$，表示该字符串有 25 个时序；n 表示每个时序有 n 种可能的字符类型结果，其中包含 *blank* 字符(用 ε 表示)，对输出结果的每个时序进行归一化指数函数(Soft Max)操作从而得到该时序概率最大的字符；最后在得到的序列中去除重复的字符以及 *blank* 字符。假设最终序列字符为 $\varepsilon stt\varepsilon ate$，在去重合并后所得到的字符序列就是 *state*。通过应用该算法我们能够有效的解决车牌字符中相邻位置相同字符的识别问题，不会产生车牌字符缺漏的情况。

2.2.2. 与注意力机制融合后的车牌识别算法

在将车牌送入识别模型前，由于识别过程中可能遇到不同类型的车牌，包含蓝牌、绿牌、黄牌等。首先将车牌原始图像调整为 1 × 32 × 100 的灰度图，这样在统一车牌图像尺寸的同时，去掉车牌中的颜色信息，只保留字符信息，有利于在特征提取的过程中不会因为颜色的影响而改变卷积层本该提取的车牌字符区域特征。值得注意的是，通过不断的完善我们的数据集，本文提出的模型能够识别变长车牌，这对于车牌识别的实际应用无疑是迈出了一大步。

1) 引入 STN

这里将 CRNN 作为车牌识别的基本框架，仍有些不足，首先是 CRNN 在识别中需要将特征图送入长短期记忆网络(Long Short-Term Memory, LSTM)进行上下文的时序关联，但在通过卷积层进行特征提取时，输入车牌由于拍摄角度以及拍摄距离等因素的影响，容易出现车牌倾斜，当车牌倾斜后在送入特征提取器中容易得到不准确的特征，为了输出完美的特征，这里引入空间变换网络[21] (Spatial Transformer Networks, STN)网络，其主要由三部分组成，首先通过定位网络生成输入与输出特征的映射关系 θ；再由网格生成器利用映射关系进行相应的空间变换，生成输入图像对应于输出图像的每个像素值；最后通过采样器将输入图像中的像素值复制到输出图像中，从而得到不改变尺寸的矫正图像。

2) 引入残差学习与注意力机制

其次在于 CRNN 的卷积层过于简单，在卷积过程中不能充分的提取所需要的特征。从经验上来说，卷积网络的深度对于提高模型的性能是相当重要的，当增加网络层数后，网络可以进行更加复杂的特征提取。但是单纯的增加深度又会产生另一个问题，即当网络的深度增加时，网络的准确度出现饱和，更甚者会下降，产生这些情况的原因便是出现了梯度消失或者爆炸。He 等人[22]在 2015 年提出残差学习，他们认为深层网络即使什么都不做，只单纯的复制之前层的特征，那么得到的特征图就算不会变好，也至少不会变差，这样深层网络的性能便能保证至少和浅层网络性能一样，这也称作恒等映射。那么怎么才能实现恒等映射呢？He 等人提出当初始输入特征为 x 时，记下一层学习之后得到的特征为 $H(x)$，现在我们不直接学习 $H(x)$，而是通过学习两者的残差 $F(x)=H(x)-x$，然后将初始特征与残差相加，则得到我们想要的特征 $H(x)$。通过这样的方式我们每步学习到的特征即使在最坏的情况下(残差为 0)也是上一层的特征，即网络性能始终不会下降。这里我们使用的残差模块如下图 6 所示：

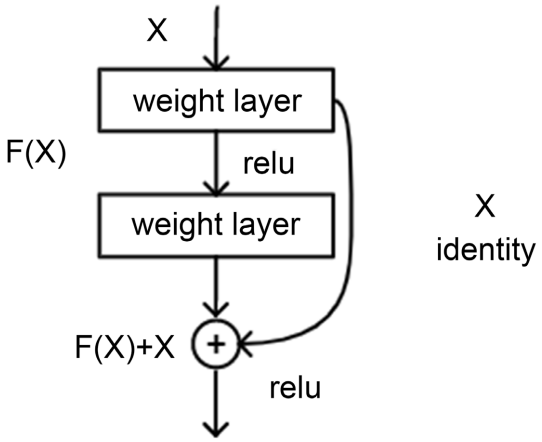

Figure 6. Residual learning: A building block
图 6. 残差学习：构建模块

通过引入残差模块，我们解决了深度问题，但只有深度往往并不能有效的提取特征，这里我们考虑用检测阶段所使用的空间和通道的混合注意力机制与残差模块进行融合来作为本文中的特征提取层。这里的输入特征是经过 STN 后的大小不变的图像 $F \in R^{1\times 32\times 100}$，首先将 F 进行一次卷积核为 3，填充(padding)和步长(stride)为 1，通道数变化的卷积层得到特征图 F'，利用 F' 生成权重 M_c，然后将 M_c 与特征图 F' 进行相乘后得到特征图 F''，在学习的过程中通过通道注意力机制让模型在通道的层面知道应该关注哪里，同时也不断的提高感兴趣区域的权重，最后得到通道上最优模型。然后利用空间注意力机制再次学习权重 M_s，同样的将 M_s 与特征图 F'' 相乘得到新的特征图 F'''，在经过通道和空间的混合注意力机制后我们的模型已经能够自动的学习特征图中重要的区域并且不断的提高这个区域的权重。在注意力机制后再加上一个卷积层得到 F''''，该卷积层的卷积核、步长与填充与上一个卷积层相同，不会更改特征图的大小，不同的是这里的输出通道数与输入通道数相同。最后将残差 F'''' 与 F 相加得到新的特征图，从而完成一轮残差学习。融合后的模块如下图 7 所示。

在本文中的特征提取阶段总共使用了 8 次残差模块，每隔一次残差模块输出通道数翻倍，通过每两次残差模块后进行的最大池化操作，将所得到的特征图尺寸进行缩放，由于残差模块中特征图尺寸并不改变，所以最后得到的特征图尺寸完全由最大池化层的个数决定。该模型特征提取阶段的卷积操作具体如下图 8 所示。

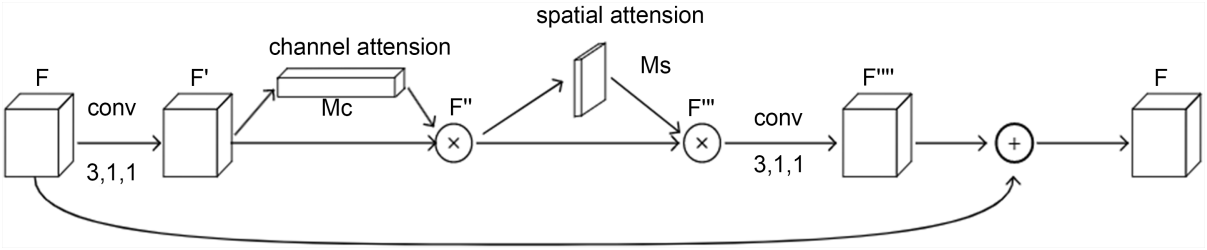

Figure 7. The convolution model based on CBAM and RES block
图 7. 基于残差模块与注意力机制的卷积模型

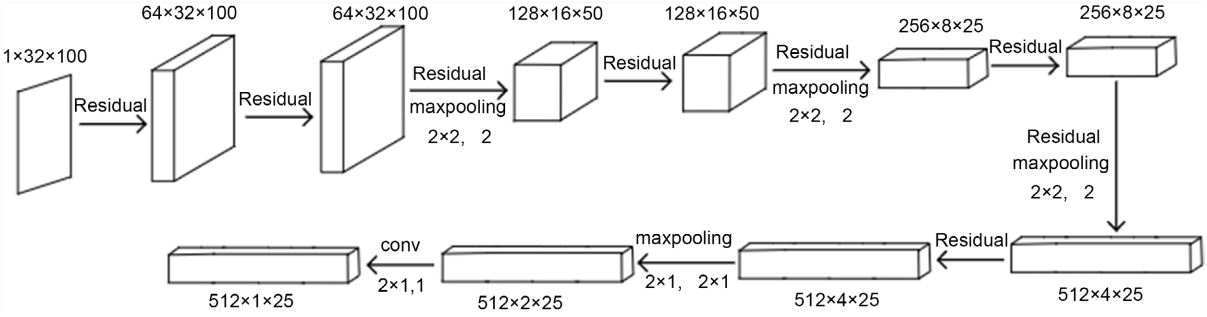

Figure 8. Convolution model of license plate recognition
图 8. 车牌识别中卷积模型

在经过残差模块和注意力卷积模块的共同作用下，车牌图像变成了 $F \in R^{512\times1\times25}$ 的特征图，随后将其送入双向长短期记忆网络(Bidirectional Long Short-Term Memory, BLSTM)中，在经过两个隐藏层为 512 层的双向 BLSTM 后预测出了具体的车牌字符，然后利用连接时序分类算法(Connectionist Temporal Classification, CTC)进行函数的拟合。传统的解码操作是每个时序对应一个字符，如果出现相同的字符，在解码过程中只留下一个字符。这样的坏处是一个单词中可能存在相邻的两个或多个相同的字符，如果按照传统的对齐操作，最后的结果是该单词中相邻的多个字符只会留下一个，结果是出现了字符丢失的现象。连接时序分类算法引入一个特殊的占位符 ε，不对应任何字符，如果一个单词中存在相邻的多个相同字符，则在他们间加入 ε 进行分隔，最后仍然将相同字符去掉，这样得到的就是一个完整的单词，具体的转录过程如下图 9 所示：

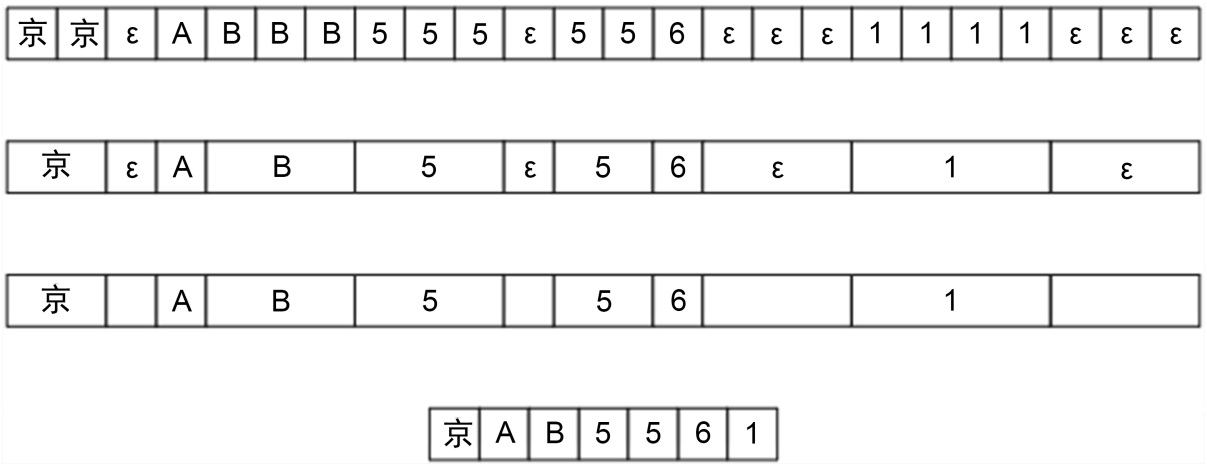

Figure 9. The process of CTC decoding the license plate characters
图 9. CTC 对车牌字符进行解码的过程

3. 实验结果与分析

3.1. 数据集及其标注

本文的车牌数据由两部分组成，一部分来自于 CCPD 数据集[23]，其余部分来自网络车牌。由于检测阶段不需要考虑数据分布不均匀的问题，因此检测阶段所有数据皆使用已经完全标注好的 CCPD 数据集共计 11,776 张；在识别阶段的数据处理上，由于 CCPD 数据集的车牌基本都是安徽车牌，省份比较单一，容易造成数据分布不均衡，车牌省份识别不准确。所以这里只取用 2961 张，其他大部分车牌数据来自于网络车牌图片，基本涵盖各个省份，共计 4502 张。但由于网络车牌并未标注，在取到这部分车牌后首先需要将其进行手工标注，与 CCPD 数据集中标注不同的是我们的标注信息只有一个，即将车牌号码标注成图像名，然后利用已经训练好的检测模型将车牌框信息进行识别，最后将车牌框裁剪出来形成只有单个车牌不含后景图像的车牌。对于 CCPD 数据集中的 2961 张车牌同样进行裁剪留下只包含车牌框的前景图像，这样就得到识别阶段所使用的车牌数据集 7463 张。将这些数据集按照 7:3 的比例分成训练集和测试集，其中训练集包括 5220 张，测试集包括 2243 张。

3.2. 模型训练与测试

本文采用基于 pytorch 的深度学习框架进行训练和测试，并通过上文所示的网络结构对车牌进行检测与识别，在 GPU 的选择方面，采用 NVIDIA GeForce RTX 1080Ti，CUDA10，内存为 16 GB，使用的编程语言为 python，在训练之前，需要对训练超参数进行初始化，具体超参数设置如下表 1 所示。

Table 1. Hyper parameter setting during the training
表 1. 训练过程中超参数设置

超参数名	检测超参数	识别超参数
Epoch	100	150
Batch size	8	64
Learning rate	1e−5	0.1
Weight decay	0.001	0.001

在训练时，我们根据平均精度(Average Precision, AP)，识别准确率(Accuracy)，训练参数量(Params)以及内存访问成本(Memory access cost, MACs)来判断模型性能。其中平均精度表示检测阶段 PR 曲线(Precision-Recall)上的 Precision 值取平均值；识别准确率表示识别阶段该模型对训练集和测试集数据上正确识别的车牌号码的比例；而训练参数量表示该模型中需要训练的参数总和，其计算公式为：

$$\text{Params} = C_{out}\left(h \cdot w \cdot C_{in} + 1\right) \tag{8}$$

这里的 h 和 w 表示输入图像的高度和宽度，C_{in} 和 C_{out} 分别表示输入图像的通道数以及经过该网络层输出图像的通道数；内存访问成本表示该模型的计算量大小，其计算公式为：

$$\text{MACs} = h \cdot w\left(C_{in} + C_{out}\right) + K \cdot C_{in} \cdot C_{out} \tag{9}$$

这里的 h, w, C_{in}, C_{out} 所表示的和上式相同，唯一不同的是 K 表示卷积核的大小。

3.3. 实验结果与分析

本次实验在车牌检测时将多种注意力机制通过聚类、放缩等做了对比实验，具体结果如下表 2 所示：

Table 2. Comparison of the influence of different modules on the test results
表 2. 不同模块对检测结果的影响对比

模型	AP	Params (M)	MACs (G)
YOLOv4-tiny	92.92%	5.874116	3.410797
YOLOv4-tiny + K-Means + SE	91.61%	5.917124	3.411187
YOLOv4-tiny + SE	93.01%	5.917124	3.411187
YOLOv4-tiny + CBAM	93.60%	5.960426	3.411465
YOLOv4-tiny + ECA	93.05%	5.874131	3.411149
YOLOv4-tiny + K-Means	88.83%	5.874116	3.410797
YOLOv4-tiny + K-Means + 线性变化	91.54%	5.874116	3.410797

通过观察及分析实验结果发现，在本次实验中加入的 K-means 并没有得到应得的结果，本文在算法改进阶段分析了 K-means 起反作用的原因，并加入了线性变化，将框进行了线性双方向的延伸，结果比单纯的使用 K-means 效果要好，对于本次实验来说，车牌大小几乎固定，即便是线性变化后的框也没有得到最优的结果。但是证明将先验框进行线性变换是有效的，日后是值得其他实验进行参考的。通过对比各个模型的结果，我们发现在 Yolo-tiny 中加入 CBAM 注意力机制得到的检测结果是最优的，AP 值达到了 93.60%。

在识别时，我们通过消融实验来对各个模块所带来的影响进行分析。这里将测试分为三类，第一类不包含 STN 网络，只经过残差学习和注意力机制的卷积模块提取特征，然后送入 BLSTM 和 CTC 进行识别；第二类不使用特殊卷积层，在经过 STN 后将特征图送入普通卷积，然后直接送入 BLSTM 和 CTC 中识别；第三类将使用所有的模块进行训练。最终得到的结果如下表 3 所示：

Table 3. Recognition accuracy and parameter comparison under the ablation experiment
表 3. 消融实验下识别准确率及参数对比

模型	Accuracy (train)	Accuracy (test)	MACs (G)	Params (M)
Residual + Attention + CRNN	100%	92.38%	2.1340	14.9408
STN + CRNN	81.49%	59.52%	0.7080	8.8145
STN + Residual + Attention + CRNN	100%	92.15%	2.1476	15.0131

同时将训练过程中的各个损失进行记录，结果如下图 10 所示：

我们在图 10 中可以清晰的看到残差学习与注意力机制对于识别过程中损失的下降具有明显的加速作用，当采用普通的卷积层时，损失在前面几轮中下降速度较慢，效果不好。而在本次实验中是否采用 STN 对于损失的下降影响不大，然后通过上表 3 所示，我们可以看到 STN 对于最后结果中的识别准确率甚至更好，我们分析认为是因为本文的数据集不够完善，对于倾斜车牌的数量不够多，数据集中大部分车牌都是平整的，甚至不需要空间变换，因此 STN 对于整个网络的效果可能出现了副作用。但是考虑到之后我们将继续完善数据，待完善后再次进行分析观察是否需要 STN，并且由表 3 中的参数量可知添加 STN 后的网络参数基本不变，不会对该模型的计算量造成过大的影响，因此我们这里暂时加上空间变换网络。同时，由表 3 所示，我们明显看到残差学习和注意力机制相结合的卷积模块能够大幅度的提升车牌识别的准确率，对于车牌识别的效果最是明显。

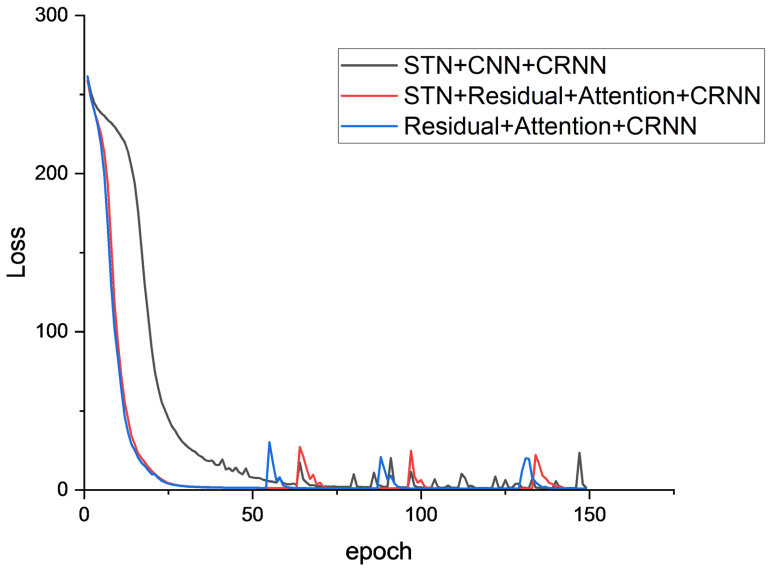

Figure 10. Comparison of ablation experiment on the loss
图 10. 消融实验损失对比图

4. 结论

在车牌检测阶段本文在 YOLOv4-tiny 算法的基础上，通过精简网络结构优化参数解决了现实场景下的车牌定位困难等问题，此外，通过实验分析了先验框 K-means 和线性的拉伸未发挥出相应优势的具体原因，最后通过引入注意力机制使得检测结果的 AP 值达到 93.60%，极大的提高了检测精度。在车牌识别阶段通过 STN 和残差模块以及注意力机制的融合，使得车牌识别不用再依靠繁琐的字符分割与字符识别，而是直接进行端到端的识别。最主要的一点是该方法能够进行混合车牌的识别，同时识别准确率达到了 92.15%，有效的解决了混合车牌下需要不同算法的现实问题，使得车牌识别在实际项目中得以更好的应用。但由于车牌数据较难采集，若要更加准确的识别车牌省份的汉字则还需要大量的数据进行训练，而我们这里搜集的数据还不够完善。我们下一步计划是：1) 进一步完善我们的数据集，通过收集更多的全国省份的车牌来提高汉字识别率，并采集多种场景下的数据，比如雨天、雾霾以及严重的灰尘覆盖车牌和在这些场景下的倾斜以及远距离和近距离的车牌数据集。借此来提高在实际生活我们进行车牌识别的鲁棒性以及泛化性；2) 继续优化模型内的结构，优化识别网络中所用到的循环层，寻找一种比 BLSTM 与 CTC 更为高效的结构来进行识别；3) 验证 STN 在全面的数据集上是否有作用。

基金项目

本文得到山东省重大科技创新工程项目(No. 2019JZZY020102)；江苏省重点研发计划项目(No. BE2018084)；2019 年工信部工业互联网创新发展工程项目——工业互联网标识解析二级节点平台项目(综合型应用服务平台)分包 1 项目(No. TC190A3X8-1)；2019 年工信部项目：国产数据库数据与应用迁移支撑验证关键技术攻关；2021 年工业互联网创新发展工程——标识解析数据网关项目(No. TC210A02M)；2021 年工业互联网创新发展工程——工业实时数据库(No. TC210804D)的资助。。

参考文献

[1] Redmon, J., Kumar Divvala, S., Girshick, R., *et al*. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[2] Liu, W., *et al*. (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV* 2016. *Lecture Notes in Computer Science*, Springer, Cham, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[3] Ren, S.Q., He, K.M., Girshick, R., *et al*. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[4] 洛雪超, 刘桂雄, 冯云庆. 一种基于车牌特征信息的车牌识别方法[J]. 华南理工大学学报, 2003, 31(4): 70-73.

[5] Xia, H. and Liao, D. (2011) The Study of License Plate Character Segmentation Algorithm Based on Vetical Projection. 2011 *International Conference on Consumer Electronics*, *Communications and Networks*, Xianning, 16-18 April 2011, 4583-4586. https://doi.org/10.1109/CECNET.2011.5768714

[6] Anagnostopoulos, C.N.E. (2006) A License Plant-Recognition Algorithm for Intelligent Transportation System Applications. *IEEE Transactions on Intelligent Transportation System*, **7**, 377-392. https://doi.org/10.1109/TITS.2006.880641

[7] Li, C. (2002) Cluster-Based Method of Characters Segmentation of License Plate. *Computer Engineering & Application*, **6**, 221-222.

[8] Wang, X.H., Yu, J.J., Miao, Z.H., *et al*. (2014) License Plate Recognition Based on Pulse Coupled Neutral Networks and Template Matching. *Proceedings of the* 33*rd Chinese Control Conference*, Nanjing, 28-30 July 2014, 5086-5090. https://doi.org/10.1109/ChiCC.2014.6895805

[9] 肖秀春, 吴伟鹏. 基于深度学习的车牌字符识别的设计与实现[J]. 电子技术与软件工程, 2018(16): 65-66.

[10] Chen, L.C., Papandreou, G., Kokkinos, I., *et al*. (2016) DeepLab: Semantic image Segmentation with Deep Convolution Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **40**, 834-848.

[11] 赵志宏, 杨绍普, 马增强. 基于神经网络 LeNet-5 的车牌字符识别研究[J]. 系统仿真学报, 2010(3): 638-641.

[12] Jain, V., Sasindran, Z., Rajagopal, A., Biswas, S., Bharadwaj, H.S. and Ramakrishnan, K.R. (2016) Deep Automatic License Plate Recognition System. *Proceedings of the Tenth Indian Conference on Computer Vision*, *Graphics and Image Processing*, New York, 18-22 December 2016, 1-8. https://doi.org/10.1145/3009977.3010052

[13] Li, H., Wang, P. and Shen, C. (2019) Towards End-to-End Car License Plates Detection and Recognition with Deep Neural Networks. *IEEE Transactions on Intelligent Transportation Systems*, **20**, 1126-1136.

[14] Zherzdev, S. and Gruzdev, A. (2018) LPRnet: License Plate Recognition via Deep Neural Network. https://arxiv.org/abs/1806.10447

[15] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. https://arxiv.org/pdf/2004.10934.pdf

[16] 杨俊闯, 赵超. K-Means 聚类算法研究综述[J]. 计算机工程与应用, 2019, 55(23): 7-14+63.

[17] Woo, S.H., Park, J.C., Lee, J.Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss Y., Eds., *Computer Vision—ECCV* 2018. *Lecture Notes in Computer Science*, Springer, Cham, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[18] Hu, J., Shen, L. and Sun, G. (2017) Squeeze-and-Excitation Networks. 2018 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[19] Shi, B.G., Bai, X. and Yao, C. (2015) An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2298-2304.

[20] Graves, A., Fernandez, S., Gomez, F. and Schmidhu, J. (2006) Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Network. *Proceedings of the* 23*rd International Conference on Machine Learning*, Pittsburgh, 25-29 June 2006, 369-376. https://doi.org/10.1145/1143844.1143891

[21] Jaderberg, M., Simonyan, K., Zisserman, A. and Kavukcuoglu, K. (2016) Spatial Transformer Networks. https://arxiv.org/abs/1506.02025

[22] He, K.M., Zhang, X.Y., Ren, X.Q. and Sun, J. (2015) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[23] Xu, Z., Yang, W., Meng, A., *et al*. (2018) Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss Y., Eds., *Computer Vision—ECCV* 2018. *Lecture Notes in Computer Science*, Springer, Cham, 261-277. https://doi.org/10.1007/978-3-030-01261-8_16