

基于3D卷积的图像序列特征提取与 自注意力的车牌识别方法

曾淦雄^{1,2}, 柯道^{1,2}

(1. 福州大学数学与计算机科学学院, 福建 福州 350116;
2. 福建省网络计算与智能信息处理重点实验室, 福建 福州 350116)

摘要: 近年来, 基于自注意力机制的神经网络在计算机视觉任务中得到广泛的应用。随着智能交通系统的广泛应用, 面对复杂多变的交通场景, 车牌识别任务的难度不断提高, 准确识别的需求更加迫切。因此提出一个基于自注意力的免矫正的车牌识别方法 T-LPR。首先对图像进行切片和序列化, 并使用 3D 卷积对切片序列进行特征提取, 从而得到图像的嵌入向量序列。然后将嵌入向量序列输入基于 Transformer Encoder 的编码器中, 学习各个嵌入向量之间的关系并输出最终的编码结果。最后使用分类器进行分类。在多个公共数据集上的实验结果表明, 所提方法对各类困难场景下的车牌识别都非常有效。

关键词: 车牌识别; 图像嵌入向量; 自注意力; 免矫正

中图分类号: TP183, TP39

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202128

3D convolution-based image sequence feature extraction and self-attention for license plate recognition method

ZENG Ganxiong^{1,2}, KE Xiao^{1,2}

1. College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China
2. Fujian Provincial Key Laboratory of Networking Computing and Intelligent Information Processing, Fuzhou 350116, China

Abstract: In recent years, neural networks based on self-attentive mechanism have been widely used in computer vision tasks. As the intelligent transportation system is widely used, the task difficulty of license plate recognition is increasing and the need for correct recognition is getting more pressing in the face of complex and changing traffic scenes. Therefore, a rectification-free license plate recognition method T-LPR based on self-attention was proposed. Firstly, the images were sliced and sequenced, and 3D convolution was used for feature extraction of the sliced sequences to obtain a sequence of image embedding vectors. Secondly, the sequence of embedding vectors was fed into an encoder based on Transformer Encoder, which learned the relationship between the individual embedding vectors and outputs the final encoding result. Finally, the final encoding result was classified by a classifier. Experimental results on several public datasets show that T-LPR proposed is very effective for recognizing license plates in all kinds of difficult scenarios.

Key words: license plate recognition, image embedding vector, self-attention, rectification-free

收稿日期: 2021-05-28; 修回日期: 2021-07-05

通信作者: 柯道, kex@fzu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61972097); 福建省科技引导性项目 (No.2017H0015); 福建省自然科学基金资助项目 (No.2018J01798)

Foundation Items: The National Natural Science Foundation of China (No.61972097), The Technology Guidance Project of Fujian Province (No.2017H0015), The Natural Science Foundation of Fujian Province (No.2018J01798)

1 引言

随着城市智能管理的进一步发展,智能交通系统^[1]变得越来越重要。车牌识别作为智能交通系统中的关键一环,识别效果的好坏将会在很大程度上影响整个系统的运行。根据以往的经验,车牌识别系统的应用场景较为受限,对识别的图像要求高,如特定的拍摄角度、拍摄距离、环境光等。随着车牌识别需求的进一步提升,在非受限的场景下进行无感的车牌识别变得愈加重要,即对识别算法提出了更高的要求,需要对任意的车牌图像进行准确的识别。

在非受限的开放场景下获取的车牌图像存在许多问题,主要体现在曝光条件不一致,如夜间、白天、逆光拍摄等;天气环境不一致,如下雨导致的模糊、下雪导致的部分遮挡;拍摄角度不一致导致的倾斜、旋转等,以及其他问题导致的车牌污损、扭曲变形、模糊等。图1展示了部分典型的困难车牌样本。在过去的很长一段时间里,车牌识别由传统的图像处理技术完成^[2-3],而在卷积神经网络^[4]成为视觉任务的主流方法之后,许多车牌识别方法开始基于卷积神经网络来实现^[5-6]。

车牌识别任务可以简单分为两大类:基于字符分割的方法^[7-9]和基于免分割的方法^[10-12]。在基于字符分割的方法中,第一步是对车牌图像进行处理,直到获取到图像中的每个字符图像,随后逐一对单个的字符图像完成识别,从而完成整体的识别过程。在获取单个字符图像的过程中有许多可行的技术路线可供选择,如使用语义分割^[13-14]、目标检测^[15-16]等。对单个字符的识别方法有支持向量机(support vector machine, SVM)^[3]、反向传播(back prop-

agation, BP)神经网络^[17]、LeNet^[18]等。这类方法的最大限制来源于第一步获取的单个字符图像,在复杂的开放场景下的车牌图像通常难以获取完整的清晰可用的单个字符并用于后续的分类。基于免分割的方法将车牌识别视为一个文本序列进行识别,技术路线上主要有基于卷积神经网络与序列识别模型的方法^[5,12,19]以及基于卷积神经网络与循环神经网络的识别方法^[11,18]。通常这类方法在识别上更加有效,但是面对复杂的开放场景下的车牌图像,也出现了明显的性能下降。此外,该类方法通常需要一个前置的矫正器,用于完成旋转、倾斜、扭曲车牌的矫正任务,使得车牌处于一个正面的视角。除此之外,自然场景文本识别算法^[20-21]也可以被应用到车牌识别中。但是,自然场景文本识别需要考虑文本的上下文语义联系,而在车牌序列中并无上下文联系,每个字符是随机的,不包含语义关系。

自注意力(self-attention)结构^[22]的神经网络被应用于各式各样的任务中,其中以Transformer为代表的一系列方法^[22-23]在自然语言处理中的应用较早被人所熟知,在多个自然语言处理任务中达到了较佳的表现效果。在计算机视觉任务中,人们越来越意识到自注意力网络也可以大有作为,在典型的图像分类^[24]、目标检测^[25]、语义分割^[26]、视频处理^[27]等场景,自注意力模型有非常亮眼的表现。

本文提出了一个基于自注意力的免矫正的车牌识别方法 T-LPR (Transformer-based license plate recognition),可以实现在非受限的开放场景下的车牌图像识别。该方法首先对图像进行切片,形成图像的切片序列。接着通过3D卷积学习每个子图像



图1 部分典型的困难车牌样本

的图像嵌入向量。在切片序列中,本文采用高度优先,使得在切片序列中属于同一字符的切片尽可能靠近,从而3D卷积能很好地获取字符的局部和全局两种特征用于生成图像嵌入向量。然后通过自注意力网络的编码器学习每个嵌入向量之间的关系,从而完成字符的编码过程。最后使用一个简单的多类别分类器进行分类以获取概率矩阵,再进行最终的解码即可完成变长的车牌号码识别。可以简单地认为该方法首先学习字符的局部特征和全局特征,随后通过自注意力的形式学习更加细致的局部与局部之间的联系,构建字符的整体特征,从而使用简单的多分类器就可以完成识别任务。本文方法可以在不使用矫正器的基础上以很高的识别准确率(以下简称识别率)完成对旋转、倾斜、扭曲车牌的识别;同时对部分遮挡的污渍车牌、雨雪雾等天气下的车牌具有较高的识别率,也不需要使用复杂的循环神经网络(recurrent neural network, RNN)或长短期记忆(long short-term memory, LSTM)网络的解码器(这类解码器通常难以并行化计算)。

本文在多个公共数据集(CCPD^[7]、PKUData^[28]、CLPD^[11])上进行了充分的实验。实验结果表明,本文的识别方法比之前的识别方法具有明显的优势。由于PKUData数据集规模较小,本文还分析了采用大规模数据集CCPD进行预训练后的实验结果。本文还对流行的图像分类网络视觉注意力模型ViT(Vision Transformer)^[24]进行了修改,并与T-LPR方法在车牌识别上进行了对比。

本文的主要贡献总结如下。

(1) 本文提出了基于自注意力的免矫正的车牌识别方法T-LPR。该方法可以被应用于非受限场景的车牌识别任务,具有很高的泛化性和鲁棒性。

(2) 本文引入了基于3D卷积神经网络(convolutional neural network, CNN)的图像嵌入向量方法,用于对图像切片进行序列化,同时保留了图像切片的上下文语义关系。

(3) 本文方法在多个公共数据集上取得了最佳的识别表现,包括大规模数据集CCPD和包括多省份地区的车牌数据的PKUData以及CLPD数据集。

2 相关工作

随着深度学习的进一步发展,近些年来以深度学习为技术路线的车牌识别方案被广泛应用并取

得了很好的效果。本节将简要介绍利用深度学习技术实现车牌识别的一些相关工作。此外,车牌号码识别可以被认为是场景文本识别的特定实现,因此本节也将简单介绍场景文本识别的部分相关工作。

2.1 基于字符分割的方法

基于字符分割实现车牌识别的基本流程为通过对车牌图像进行预处理来获取单个字符图像^[9,29-30]或者通过目标检测的方法获取单个字符图像^[8,10],前者需要使用分类器对单个字符进行识别;后者可以通过分类器对检测到的字符进行识别,也可以将检测与识别融合,同时完成检测与识别任务。

Björklund T等人^[31]利用卷积神经网络进行车牌检测以获取车牌框,随后进行车牌矫正,对矫正后的车牌图像再次使用卷积神经网络进行字符检测,获取字符框并进行分类,识别对应的字符并处理形成最终的识别结果。该方法在开放场景中存在较大的局限,矫正与分割识别大大地影响了车牌识别系统的识别速度。

Silva S M等人^[10]观察到提高待识别车牌图像的质量对于在非受限场景下完成高精度的车牌识别效果非常必要。一个正面、无旋转、变形小的车牌是易于识别的,因此他们首先通过YOLOv2检测器检测开放场景下的车牌图像;然后设计一个专门的矫正网络WPOD-NET实现对不同形态车牌的统一矫正,使得用于后续识别的车牌处于正面的、易于检测字符的状态;最后使用简单的YOLOv2^[14]检测器完成字符的目标检测。该方法不仅受限于矫正网络的输出,还受限于字符的目标检测。该方法对数据集的标注等提出了较高的要求,同时该矫正网络的处理速度也是一个瓶颈。Gou C等人^[30]提出了一种基于极值区域和受限玻尔兹曼机的车牌识别算法。首先,使用边缘检测和图像过滤对车牌进行粗略的检测;然后使用极值区域来提取字符区域,反过来,极值区域又被用来细化车牌区域;最后使用受限玻尔兹曼机完成字符的逐个识别。Gao F等人^[32]提出了一个基于编码器-解码器框架的车牌识别方法。该方法通过编码器两阶段地获取精细的字符框,随后将字符框序列送入seq2seq网络进行字符的过滤或排序,以获取最终的车牌号码序列。

显而易见的是,基于字符分割的方法的识别率

依赖于车牌字符的分割结果,而字符分割或语义分割任务本身是极具挑战性的,在复杂的环境条件(包括雨、雪、雾),以及污渍、旋转、倾斜等各种极端的车牌下,分割结果的精确性难以保证。再者,分割任务与多个字符的识别会导致整个识别过程变得冗长,降低车牌识别系统的实时性。而基于免分割的方法可以在一定程度上解决这些问题。

2.2 基于免分割的方法

基于免分割的方法一般将整个车牌图像作为输入,通过CNN、多分类器、RNN等形式直接获取完整的车牌号码。这类方法的普遍优点是对数据集的标注要求低、处理速度较快、处理流程较短。

得益于CNN在特征提取方面的发展以及RNN在解决序列问题方面的发展,许多研究者^[5,11,19]开始利用CNN进行特征提取形成特征序列,然后使用RNN进行序列解码来完成车牌号码的识别任务。

Xu Z B等人^[7]提出基于端到端训练的检测与识别一体的车牌识别方法RPNNet。该方法首先检测图像中感兴趣的车牌区域,然后在多个尺度的特征图中获取特征并进行融合,最后使用固定的多个分类器进行识别。显然,该方法在训练策略上首先需要完成检测任务的训练,随后加入分类器继续训练,存在一定的局限性;其次需要固定的分类器个数,针对不同类型的车牌需要调整模型。

Li H等人^[5]提出通过滑动窗口和CNN在车牌图像上提取特征序列,随后将特征序列送入双向长短期记忆(bidirectional long short-term memory, BLSTM)网络^[33]进行序列识别,同时使用连结时间分类(connectionist temporal classification, CTC)^[34]完成对BLSTM的输出的最终解码。CTC不要求网络输出的预测序列与目标序列对齐,就能完成模型的训练与解码。

Li H等人^[19]提出了端到端的车牌识别方法。该方法与Xu Z B等人^[7]提出的方法类似,可以实现检测与识别同步完成。首先使用感兴趣区域(region of interest, RoI)提取器获取可能的车牌区域,随后将RoI提取的特征分别送入车牌检测器完成背景与车牌区域的分类、车牌区域的定位和车牌号码的识别。该识别器由双向循环神经网络(bidirectional recurrent neural network, BRNN)和一个线性分类

器构成。

在神经网络中,注意力机制发展迅速且非常有效,Zhang L J等人^[11]提出基于注意力机制的车牌识别网络。该方法利用Xception^[35]网络进行基本的特征提取以形成一个特征向量和2D注意力权重向量,使用LSTM^[36]模型实现对特征向量的解码任务,使用交叉熵损失作为损失函数进行训练。此外,Jain V等人^[6]提出使用CNN配合多个分类器实现定长序列的车牌识别方法;Zherzdev S等人^[12]提出基于CNN和CTC^[34]的变长车牌识别方法,该方法还将空间变换器作为前置模块,以提高识别效果。

这类以CNN、分类器以及RNN为基础的识别模型在一定程度上解决了基于语义分割的方法中车牌字符的难分割问题,在带来性能提升的同时,大大简化了车牌识别方法的流程。但需要注意的是,RNN难以并行化,在计算效率上不及CNN,为了进一步提升识别性能,需要引入矫正模型进行前置处理,这会进一步使模型复杂化,同时矫正结果对识别结果有一定的影响。

2.3 场景文本识别方法

自然场景文本识别是一项具有较高挑战性的计算机视觉任务,其主要目标是从图像中获取文本位置与文本内容,与车牌识别任务具有很高的相识性,因此在识别端也可以使用场景文本识别的网络来实现车牌识别。

在相对规则的文本图像上,Shi B G等人^[37]提出了端到端训练的免分割的卷积循环神经网络(convolutional recurrent neural network, CRNN)。使用CNN获取一维的特征向量序列,对该序列使用RNN进行解码以获取序列概率向量。Zhang Y P等人^[38]在可变长度的场景文本识别任务中采用了一种无监督的固定长度领域适应方法,该方法基于注意力编码器-解码器架构。在与开放场景下的车牌识别任务更相似的、更困难的不规则文本场景识别上,许多学者关注到文本矫正对识别率有很大的影响。Shi B G等人^[37]通过引入空间变换网络(spatial transformer network, STN)^[39]来实现对不规则图像的识别,在车牌识别任务中通常也需要嵌入STN来实现更高的识别率。Luo C J等人^[20]提出一个多目标矫正网络(multi-object rectified attention network, MORAN),包括一个矫正器MORN(multi-object rectified network)和一个基于注意力机制的识别器

ASRN (attention-based sequence recognition network)。Yang M K 等人^[40]提出一个基于对称性约束的矫正网络 ScRN (symmetry-constrained rectification network)，用于高度不规则文本的矫正。Lu N 等人^[41]基于自注意力提出 MASTER (multi-aspect non-local network for scene text recognition) 方法，基于全局语义模块 (GC block) 提出基于多头注意力机制的 GCAttention 模块，进一步提高了不规则场景文本的识别率。尽管使用场景文本识别网络完成车牌识别任务是一个不错的选项，但是基于车牌识别任务的特征 (如车牌字符通常在同一直线上、车牌字符间没有上下文的语义联系等) 提出一个更加高效的车牌识别器是必需的，本文提出基于图像切片嵌入向量和使用自注意力编码器对局部字符特征和整体字符特征进行学习，既不用考虑复杂的矫正网络，也避免了全局的语义信息导致的数据集偏见，进而避免识别错误，同时 Transformer 的编码器可以完全并行计算，使得计算效率得到进一步提升。

3 方法

本节详细介绍本文提出的车牌识别方法 T-LPR。如图 2 所示，T-LPR 是一个基于 3D CNN 和 Transformer Encoder 的端到端训练的车牌识别方法，包括图像嵌入向量层、向量编码层、解码层。图像嵌入向量层通过 3D CNN 获取图像嵌入向量。并添加位置嵌入向量信息。向量编码层使用 Transformer Encoder 对图像嵌入向量序列进行编码。解码层使用多层感知机 (multilayer perception,

MLP) 和 Softmax 函数对编码层的输出向量进行分类学习，获得概率矩阵；使用贪心算法对概率矩阵进行解码，获得最终的识别结果。模型将一张未经任何前置处理的灰度车牌图像作为输入，解码后输出对应的车牌号码。

3.1 图像嵌入向量

将二维的图像转化为有效的一维特征序列是首要工作。本文设定输入图像的向量大小为 (C, H, W) ，其中， C 表示图像的通道数， H 表示图像的高度， W 表示图像的宽度。对图像进行切片处理，得到大小为 $(S, C, P1, P2)$ 的向量，其中， S 表示图像被切片后的切片数，且有 $S = \frac{H}{P1} \times \frac{W}{P2}$ ， $P1$ 、 $P2$ 分别表示图像切片的高和宽，本文选定的典型切片大小为 $P1 = P2 = 16$ 。需要注意的是，在对图像切片进行排序时以高度优先，符合字符主体从上至下的结构特征，这使得在序列化后的序列中，相同字符的图像切片尽可能靠近。

这样得到的每个图像切片可被定义为一个图片单词。引入自然语言处理任务中常用的词嵌入概念。词嵌入将一个单词通过嵌入层转化为一个连续的一维向量，用于表示该单词的含义。本文将 3D CNN 作为图片单词的嵌入层，为每个图像切片计算对应的图像嵌入向量。使用 3D CNN 对切片序列进行特征提取后，每个图像嵌入向量实际上都包含了相邻图像切片的特征，使得学习到的嵌入向量包含更加完整的切片语义。定义输入图像为 $y \in R^{C \times H \times W}$ ，通过图像切片后为 $y_p \in R^{C \times S \times P1 \times P2}$ 。对 y_p 进行 3D 卷积操作得到输出 $y'_p \in R^{C_{out} \times S \times \frac{P1}{stride} \times \frac{P2}{stride}}$ ，其中，stride 表

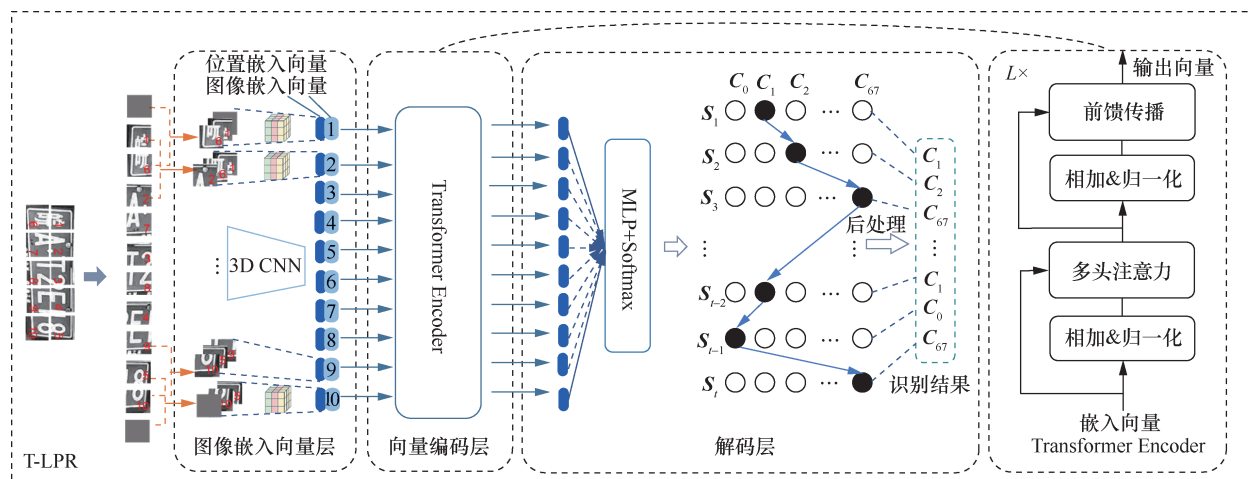


图2 T-LPR 模型结构

示 CNN 的下采样步长, C_{out} 表示 CNN 的输出通道数, 同时也是图像嵌入向量的大小。随后通过平均池化, 得到输出 $y_{pe} \in R^{C_{out} \times S \times 1 \times 1}$ 。同时, 对输入 $y_p \in R^{C \times S \times P1 \times P2}$ 做卷积核大小为 $1 \times 1 \times 1$ 、步长为 $(1, 1, 1)$ 的 3D 卷积, 并通过均值池化得到 $y_p'' \in R^{C_{out} \times S \times 1 \times 1}$ 。对 y_p' 与 y_p'' 做残差连接, 得到最终的嵌入向量。定义 3D 卷积特征提取操作为 F_{3D} , 卷积核大小为 $1 \times 1 \times 1$ 的卷积操作为 F'_{3D} , 均值池化操作为 F_{avg} , 则对于输入 y , 有:

$$y_p = \text{Reshape}(y), y_p \in R^{C \times S \times P1 \times P2}, y \in R^{C \times H \times W} \quad (1)$$

$$y_{pe} = F_{avg}(F_{3D}(y_p)) + F_{avg}(F'_{3D}(y_p)), y_{pe} \in R^{C_{out} \times S \times 1 \times 1} \quad (2)$$

3.2 自注意力编码网络

完成第 3.1 节的图像嵌入向量处理后, 二维图像被转化为一维的序列向量。这些序列向量将被输入自注意力编码网络 SA 进行编码。在编码网络的结构上, 本文使用与 Transformer Encoder 一致的结构。这里简要回顾 SA 的基本结构。对于一个长度为 L_e 、维度为 D 的输入序列 $x \in R^{L_e \times D}$, 通过线性映射函数 F_{linear} 得到 3 个权重向量, 分别为 q, k, v , 对于序列中的每个元素, 计算所有 v 的加权和并将其作为输出, 其中加权系数 w_{ij} 由各个元素通过线性映射学习得到的 q^i 和 k^j 的对偶相似度决定。其中, 隐藏层的维度为 D_{hidden} 。则该过程可表示为:

$$[q, k, v] = x F_{linear}, F_{linear} \in R^{D \times 3D_{hidden}} \quad (3)$$

$$A = \text{Softmax}\left(\frac{qk^T}{\sqrt{D_k}}\right) \quad (4)$$

$$SA(x) = Av \quad (5)$$

其中, A 表示对偶相似度矩阵, D_k 表示 k 的维度, $SA(x)$ 表示输入 x 经过自注意力网络后的输出。

使用并行计算来构建自注意力模型, 可以得到一个多头自注意力 (multi-head self-attention, MSA) 模型。定义注意力头的数量为 h , 则有:

$$MSA(z) = [SA_1(x); SA_2(x); \dots; SA_h(x)] F_{linear}, \quad (6)$$

$$F_{linear} \in R^{h \times D_{hidden} \times D}$$

标准的 Transformer^[22]通过堆叠 MSA 模块和 MLP 来完成, 使用 LN (layer-norm)^[42]进行层归一化。本文定义 L 为多层注意力的层数, D_h 为隐藏层

的大小。($P1, P2$) 表示当前使用的切片大小, 切片大小决定了输入自注意力网络的序列长度。小切片尺寸可以更加细致地学习局部信息, 但是计算需求更大; 大切片尺寸可使得待学习的序列长度变短, 计算更快。对于编码器的输入 y_{pe} , 有:

$$y^0 = y_{pe} \quad (7)$$

$$y^l = \text{MSA}(\text{LN}(y^{l-1})) + y^{l-1}, l = 1, 2, \dots, L \quad (8)$$

$$y_o = \text{MLP}(\text{LN}(y^l)) + y^{l-1}, l = 1, 2, \dots, L \quad (9)$$

3.3 分类解码器

一维特征序列向量通过基于自注意力的编码层后形成可用于分类的特征向量。本文使用一个多层感知机对这些特征进行解码, 得到每个位置的分类结果。即对于编码器的输出 $y_o = [y_o^1; y_o^2; \dots; y_o^s]$, $y_o \in R^{S \times D}$, 将其输入共享的多层感知机 MLP, 得到:

$$y = \text{Softmax}([MLP(y_o^1); MLP(y_o^2); \dots; MLP(y_o^s)]), \quad (10)$$

$$MLP \in R^{D \times C}$$

其中, S 表示特征向量的序列长度, 也表示允许的最长车牌号码长度, D 表示图像嵌入向量大小, C 表示字符类别数量, 这里使用 68 个车牌字符 (包括 31 个省 (自治区、直辖市), 26 个字母, 10 个数字, 以及一个分隔符 “-”)。 $y = [y_1; y_2; \dots; y_s]$ 表示序列长度为 S 的输入 y_o 对应的概率矩阵, 其中 y_i 表示第 i 个字符的概率分布。对概率矩阵进行解码, 就可以得到一个可能的识别序列 l 。这个可能的识别序列还可能存在重复字符和分割符 “-”, 需要去除相邻的重复字符和分割符。例如 $l = \text{“闽闽 AA-11-2-3-44- 5---”}$, 经过处理后得到的 $l = \text{“闽 A12345”}$ 即最终的车牌识别结果。本文采用贪心搜索算法对概率矩阵进行解码, 即将每个字符的概率分布中概率最大的作为最终的识别结果。

3.4 训练网络

本文基于 PyTorch 框架构建网络模型。模型在 16 GB RAM 的 AMAX XG-24201GP Tesla P100 上进行训练。本文使用 CTC 损失函数进行训练, 将 Adam 作为模型优化器。对于大规模数据集 CCPD, 批大小设置为 128, 学习率为 4×10^{-5} 。对于小数据集 PKUData, 批大小为 16, 学习率为 4×10^{-4} , 并使用余弦衰减函数, 将学习率衰减至最小, 即 1×10^{-6} 。

4 实验结果与分析

本节对本文提出的方法在 3 个公共数据集 (CCPD、PKUData、CLPD) 上的实验结果进行分析。对于 3 个数据集中的每个车牌而言, 当且仅当车牌的所有号码均被正确识别时, 才能将该车牌标记为正确识别样本。因此准确率 (accuracy) 等于所有正确识别的车牌数除以待测的车牌总数, 即 $ACC = \frac{TP}{TP + FP}$, 其中 TP 表示正确识别的车牌数量, FP 表示错误识别的车牌数量。为了进一步评估在小数据集上由于训练数据省份分布不均衡导致的问题, 针对 PKUData 和 CLPD 还提供了去除省份字符后的准确率。本文在各个数据集上都与当前主流的深度学习方法进行了对比, 包括基于注意力机制的车牌识别方法和部分自然场景文本识别的方法。几种数据集的基本情况见表 1。

4.1 数据集介绍

本节介绍 3 个公共数据集的基本情况及使用细节。

CCPD^[7]: CCPD 是目前已知公开可获取的最大开放场景的车牌检测与识别数据集, 包含超过 28 万张车牌图像。根据车牌图像的特征, CCPD 包括 7 个子数据集, 分别为 Base (基础数据集), 包括多种场景的正面车牌图像, 识别难度较低; DB, 包括过曝或暗光环境下的车牌图像; FN,

包括远距离或很近距离拍摄的车牌图像; Rotate 和 Tilt, 包括旋转或倾斜的车牌图像; Weather, 包括雨、雪、雾等天气条件下的车牌图像; Challenge, 包括各类充满挑战性的困难车牌图像。需要注意的是, 该数据集的省份分布过于集中, 超过 95% 的车牌来自安徽省。本文使用与 Xu Z B 等人^[7]相同的方法, 将 Base 的一半数据取出作为训练集, 另外一半数据作为测试集。其他所有子集均用于测试。

PKUData^[28]: PKUData 数据集由 Yuan Y L 等人^[28]提出, 包括不同场景 (如高速路、城市道路、白天、黑夜等) 下的车辆图像。原数据集不包括车牌号标签, Zhang L J 等人^[11]标注了 2 253 张相应的车牌号标签用于车牌识别训练与测试。本文使用完整的数据集进行训练与测试。在训练集与测试集的划分上, 本文使用与 Zhang L J 等人^[11]一致的比例进行划分, 测试样本占比 40%, 训练样本占比 60%。

CLPD^[11]: CLPD 数据集由 Zhang L J 等人^[11]提出, 共有 1 195 张车牌图像, 包括我国的 31 个省 (自治区、直辖市) 及不同拍摄条件下的车牌图像。该数据集作为独立的测试集使用, 不参与训练, 以验证模型的泛化性与在实际场景下的鲁棒性。

4.2 CCPD 的实验结果分析

在 CCPD 数据集上将本文提出的方法与其他方法进行充分的对比, 结果见表 2, 其中, HC 表示 Holistic-CNN^[43]。除了使用额外合成数据集进行训

表 1 数据集基本情况汇总

数据集	基本描述	数量/张 (总数/训练/测试)
CCPD	大规模的开放场景车牌数据集	280 000/100 000/180 000
Base	简单的正面车牌图像	200 000/100 000/100 000
DB	过曝或暗光环境下的车牌图像	20 000/0/20 000
FN	远距离或很近距离拍摄的车牌图像	20 000/0/20 000
Rotate	旋转或倾斜的车牌图像	10 000/0/10 000
Tilt	旋转或倾斜的车牌图像	10 000/0/10 000
Weather	雨、雪、雾等天气条件下的车牌图像	10 000/0/10 000
Challenge	困难车牌图像	10 000/0/10 000
PKUData	包括较完整省份的正常车牌图像	4 210/2 526/1 684
CLPD	包括完整省份和多种拍摄环境的车牌图像	1 195/0/1 195

表2 不同方法在CCPD数据集上的车牌识别率

方法	All	Base	DB	FN	Rotate	Tilt	Weather	Challenge
Faster RCNN ^[43] +HC	92.8%	97.2%	94.4%	90.9%	82.9%	87.3%	85.5%	76.3%
YOLO9000 ^[14] +HC	93.7%	98.1%	96%	88.2%	84.5%	88.5%	87.0%	80.5%
SSD300 ^[45] +HC	95.2%	98.3%	96.6%	95.9%	88.4%	91.5%	87.3%	83.8%
TE2E ^[19]	94.4%	97.8%	94.8%	94.5%	87.9%	92.1%	86.8%	81.2%
RPnet ^[7]	95.5%	98.5%	96.9%	94.3%	90.8%	92.5%	87.9%	85.1%
DAN ^[21]	96.6%	98.9%	96.1%	96.4%	91.9%	93.7%	95.4%	83.1%
MORAN ^[20]	98.3%	99.5%	98.1%	98.6%	98.1%	98.6%	97.6%	86.5%
MTLPR ^[18]	98.8%	—	—	—	—	—	—	—
LPRNet ^[12]	93.0%	97.8%	92.2%	91.9%	79.4%	85.8%	92.0%	69.8%
ANet (real data only) ^[11]	98.5%	99.6%	98.8%	98.8%	96.4%	97.6%	98.5%	88.9%
ANet (real+synthetic data) ^[11]	98.9%	99.8%	99.2%	99.1%	98.1%	98.8%	98.6%	89.7%
T-LPR	99.0%	99.6%	99.1%	99.2%	98.1%	98.6%	98.9%	93.1%

训练的ANet (real+synthetic data)^[11], 本文的T-LPR在同等的训练数据集下都取得了最佳的性能表现。在CCPD数据集中, Rotate、Tilt数据集用于评估识别算法面对旋转、倾斜车牌图像的识别性能, Weather用于评估识别方法在不同天气情况下的性能, Challenge则用于评估识别方法在各类较高难度场景下的性能, 包括但不限于倾斜、扭曲、局部遮挡、环境光复杂等。这4个子集能够有效评估方法的识别性能。从Rotate、Tilt数据集上的结果来看, 与两阶段的方法, 如Faster RCNN^[44]+HC^[43]、YOLO9000^[14]+HC等方法相比, 端到端的识别方法(RPnet、DAN、MORAN、LPRNet、ANet、T-LPR)具有明显的优势。没有使用前置矫正器的方法(DAN^[21]、TE2E^[7]等)相较于带有矫正模块的MORAN^[20]而言性能差距明显, 而本文提出的同样没有矫正的识别方法则达到了与MORAN同等的性能表现, 这说明了本文方法可以很好地实现对未矫正车牌图像的号码识别。在Weather数据集上, T-LPR与MORAN相比性能提升了1.3%, 与基于注意力模型的只有真实数据的ANet(real data only)相比也有0.4%的提升, 与使用了额外合成数据集的ANet (real+synthetic data)相比有0.3%的提升, 这说明了T-LPR同样可以实现不同天气的拍摄条件下的车牌号码识别。最大的提升体现在Challenge

数据集上, 即使与ANet (real+synthetic data)进行比较, T-LPR依然有很大的性能优势, 具有超过3%的性能提升, 识别率达到了93.1%。这表明在复杂的开放场景下, 本文的识别方法对难度更高的车牌号码的识别具有更高的可用性。

4.3 PKUData的实验结果分析

Yuan Y L等人^[28]提出的PKUData数据集最早用于车牌检测。Zhang L J等人^[11]对PKUData数据集的部分车牌进行车牌号码标注, 并用于评估车牌识别方法。本文使用了全部的车牌数据集。采用的数据集划分和训练策略与Zhang L J等人^[11]所提方法相同。此外, 由于Sighthounds^[46]提供的公开的API不支持中文的省份识别, 因此在评估时提供的是去除中文省份信息的识别率。PKUData和CLPD数据集上的实验结果见表3。由表3可知, 对于PKUData数据集, 本文方法在包括省份信息的车牌号码识别上, 相比之前的方法有明显的提升, 比使用相同训练数据集的TE2E方法提升了17.1%, 比ANet (real data only)提升了9.9%, 比ANet (real+synthetic data)提升了6.5%, 达到了94.7%的识别率。在不包括中文省份信息的识别率(即ACC w/o)上, 本文方法达到了97.5%, 比ANet (real+synthetic data)提升了7%。这说明本文方法在小数据集上具有明显的优势, 同时比其他方法具有更好的泛化性。

表 3

PKUData 和 CLPD 数据集上的实验结果

方法	PKUData: ACC	PKUData: ACC w/o	CLPD: ACC	CLPD: ACC w/o
Sighthound (2017) ^[46]	—	89.3%	—	85.2%
TE2E ^[7]	77.6%	78.4%	66.5%	78.9%
ANet (real data only) ^[11]	84.8%	86.5%	70.8%	86.1%
ANet (real+synthetic data) ^[11]	88.2%	90.5%	76.8%	87.6%
T-LPR	94.7%	97.5%	81.4%	92.1%

4.4 CLPD 的实验结果分析

CLPD 数据集是由 Zhang L J 等人^[11]提出的不包括训练集的独立测试集。独立测试集的引入可以很好地评估车牌识别方法的泛化性与鲁棒性。测试模型的权重使用 CCPD 的训练集训练得到。测试结果见表 3。由表 3 可以发现, 本文方法比其他方法具有更好的泛化性。在包含全部车牌字符的准确率上, 本文方法达到了 81.4%, 比 ANet (real data only) 提升了 10.6%, 比 ANet (real+synthetic data) 提升了 4.6%。在不包括中文省份信息的准确率上, 本文方法达到了 92.1%, 比之前最好的模型高 4.5%。包括中文省份信息的识别率比不包括中文省份信息的

识别率低 10.7%, 这主要是 CCPD 训练集的省份字符分布不均衡导致的, CCPD 的车牌数据大多采集于同一个省份。

部分典型困难的车牌识别样例如图 3 所示。第一行到第四行展示了困难条件下识别成功的样例, 第五行展示了一些典型的识别错误样例, 如过曝、遮挡、扭曲导致的识别错误。图 3 表明本文提出的 T-LPR 在许多极端的车牌样例中依然能实现准确的号码识别, 同时也表明该方法可以被应用于真实的非受限场景, 极大地扩大了车牌号码识别的应用范围, 降低了对拍摄设备与拍摄条件的要求。



图 3 典型困难车牌的识别样例

4.5 预训练模型

预训练对 Transformer 方法有明显的影响, PKUData 中的车牌图像只有 4 210 张, 因此本文使用 CCPD 的训练集对模型进行预训练, 随后在 PKUData 上进行微调, 以观察本文方法在使用大规模数据集预训练后的性能表现, 实验结果见表 4。经过 CCPD 训练集预训练之后, 预训练模型相比未预训练模型在 ACC 上提高了 3.2%, 达到了 97.9%, 在不包括中文省份信息的准确率上则提高到 98.9%。

表 4 预训练模型的影响

方法	ACC	ACC w/o
未预训练的 T-LPR	94.7%	97.5%
预训练后的 T-LPR	97.9%	98.9%

4.6 与 ViT 的对比

在图像分类领域, ViT^[24]取得了可与许多卷积神经网络相媲美的结果, 且它对训练所需的计算资源要求更少。本文对 ViT 的分类层进行修改, 以输出一个定长的概率矩阵, 并采用 CTC 进行训练, 以完成车牌识别任务。将方法与本文方法进行对比。在参数上, 图像切片大小与 T-LPR 一致, 宽和高均为 16。在编码层数和注意力头的数量上也保持一致。实验对比结果见表 5。从表 5 可知, 在局部遮挡、污渍车牌图像较多的 Weather 数据集上, T-LPR 和 ViT 取得了相同的结果, 这表明基于 Transformer 的车牌识别方法可以有效处理局部遮挡、污渍车牌图像。而在旋转、倾斜车牌图像较多的 Rotate、Tilt 数据集上, 与 ViT 相比, T-LPR 表现出了明显的性能优势, 分别提高了 1.3%和 0.8%, 这表明 T-LPR 可以更加有效地处理旋转、倾斜车牌图像。在 PKUData 和 CLPD 数据集上, T-LPR 展现出了比 ViT 更强的泛化性与鲁棒性, 在 PKUData 上提高了 6.3%, 在 CLPD 上则提高了 20.1%, 这也表明在相同训练集的情况下, T-LPR 在独立测试集上更加有效。与 ViT 相比, T-LPR 在特征提取上采取 3D 卷积, 很好地解决了对图像序列化后导致的图像切片之间的语义分割问题。通过 3D 卷积可以很好地聚合图像切片的上下文, 获得更佳的图像嵌入向量。这也体现在 T-LPR 对局部遮挡、扭曲车牌图像具备更高的识别率上。而 T-LPR 在小数据集上的表现, 则进一步其说明其在泛化性与实际应用上具有更加明显的性能优势。

表 5 T-LPR 与 ViT 的对比实验结果

数据集	ViT ^[24]	T-LPR
Base	99.4%	99.6%
DB	99.0%	99.1%
FN	99.2%	99.2%
Rotate	96.8%	98.1%
Tilt	97.8%	98.6%
Weather	98.9%	98.9%
Challenge	92.2%	93.1%
PKUData	88.4%	94.7%
CLPD	61.3%	81.4%

5 结束语

本文提出了一个基于自注意力的免矫正的车牌识别方法 T-LPR。利用 3D 卷积对图像切片进行向量化, 以获取更加完整的切片信息。基于 Transformer 编码器完成对图像嵌入向量的编码, 使用分类器获取车牌字符的概率矩阵, 最后使用贪心算法对概率矩阵进行求解, 从而获得最终的车牌号码识别结果, 解决了在非受限的应用场景下车牌背景复杂、位置形状多变的识别问题。在大规模数据集和独立测试集上的测试结果表明, T-LPR 相比之前的相关工作在识别率和鲁棒性上都得到了进一步的提升。同时, 笔者也意识到 3D 卷积在计算效率存上在一定的局限性, 以及大规模数据集的预训练对 Transformer 的性能表现是非常重要的。因此在未来的工作中, 笔者将继续关注如何对图像的局部信息和全局信息进行更加有效的序列化与向量化, 并着重思考如何将预训练模型引入车牌识别, 这将使得只拥有少量训练数据集的研究者也可以使用性能优越的车牌识别方法。

参考文献:

- [1] 刘晓波, 蒋阳升, 唐优华, 等. 综合交通大数据应用技术的发展展望[J]. 大数据, 2019, 5(3): 55-68.
LIU X B, JIANG Y S, TANG Y H, et al. Development prospect of integrated transportation big data application technology[J]. Big Data Research, 2019, 5(3): 55-68.
- [2] KHAN M A, SHARIF M, JAVED M Y, et al. License number plate recognition system using entropy-based features selection approach with SVM[J]. IET Image Processing, 2018, 12(2): 200-209.
- [3] 薛磊, 杨晓敏, 吴炜, 等. 一种基于 KNN 与改进 SVM 的车牌字符识别算法[J]. 四川大学学报(自然科学版), 2006, 43(5): 1031-1036.
XUE L, YANG X M, WU W, et al. An algorithm based on KNN and improved SVM for licence plate recognition[J]. Journal of Sichuan

- University (Natural Science Edition), 2006, 43(5): 1031-1036.
- [4] 张钹. 人工智能进入后深度学习时代[J]. 智能科学与技术学报, 2019, 1(1): 4-6.
ZHANG B. Artificial intelligence is entering the post deep-learning era[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 4-6.
 - [5] LI H, WANG P, YOU M Y, et al. Reading car license plates using deep neural networks[J]. Image and Vision Computing, 2018, 72: 14-23.
 - [6] JAIN V, SASINDRAN Z, RAJAGOPAL A, et al. Deep automatic license plate recognition system[C]//Proceedings of the 10th Indian Conference on Computer Vision, Graphics and Image Processing. New York: ACM Press, 2016: 1-8.
 - [7] XU Z B, YANG W, MENG A J, et al. Towards end-to-end license plate detection and recognition: a large dataset and baseline[C]//Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 261-277.
 - [8] LAROCA R, SEVERO E, ZANLORENSI L A, et al. A robust real-time automatic license plate recognition based on the YOLO detector[C]//Proceedings of 2018 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2018: 1-10.
 - [9] ZHUANG J F, HOU S H, WANG Z L, et al. Towards human-level license plate recognition[C]//Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 314-329.
 - [10] SILVA S M, JUNG C R. License plate detection and recognition in unconstrained scenarios[C]//Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 593-609.
 - [11] ZHANG L J, WANG P, LI H, et al. A robust attentional framework for license plate recognition in the wild[J]. IEEE Transactions on Intelligent Transportation Systems (Early Access), 2020: 1-10.
 - [12] ZHERZDEV S, GRUZDEV A. LPRNet: license plate recognition via deep neural networks[J]. arXiv preprint, 2018, arXiv:1806.10447.
 - [13] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. arXiv preprint, 2016, arXiv:1606.00915.
 - [14] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2980-2988.
 - [15] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv preprint, 2018, arXiv:1804.02767.
 - [16] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 6517-6525.
 - [17] ZHANG Z G, WANG C. The research of vehicle plate recognition technical based on BP neural network[J]. AASRI Procedia, 2012, 1: 74-81.
 - [18] TAO X Z, LI L, LU L. A lightweight convolutional neural network for license plate character recognition[C]//Proceedings of 2019 International Symposium on Intelligence Computation and Applications. Heidelberg: Springer, 2019: 379-387.
 - [19] LI H, WANG P, SHEN C H. 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.
 - [20] LUO C J, JIN L W, SUN Z H. MORAN: a multi-object rectified attention network for scene text recognition[J]. Pattern Recognition, 2019, 90: 109-118.
 - [21] WANG T W, ZHU Y Z, JIN L W, et al. Decoupled attention network for text recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12216-12224.
 - [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv preprint, 2017, arXiv:1706.03762.
 - [23] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, 2019, arXiv:1810.04805.
 - [24] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[J]. arXiv preprint, 2020, arXiv:2010.11929.
 - [25] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the European conference on Computer Vision. Heidelberg: Springer, 2020: 213-229.
 - [26] ZHENG S X, LU J C, ZHAO H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[J]. arXiv preprint, 2020, arXiv:2012.15840.
 - [27] LI X Y, ZHANG Y Y, LIU C H, et al. VidTr: video transformer without convolutions[J]. arXiv preprint, 2021, arXiv:2104.11746.
 - [28] YUAN Y L, ZOU W B, ZHAO Y, et al. A robust and efficient approach to license plate detection[J]. IEEE Transactions on Image Processing, 2017, 26(3): 1102-1114.
 - [29] LI P Y, NGUYEN M, YAN W Q. Rotation correction for license plate recognition[C]//Proceedings of 2018 4th International Conference on Control, Automation and Robotics. Piscataway: IEEE Press, 2018: 400-404.
 - [30] GOU C, WANG K F, YAO Y J, et al. Vehicle license plate recognition based on extremal regions and restricted boltzmann machines[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4): 1096-1107.
 - [31] BJÖRKLUND T, FIANDROTTI A, ANNARUMMA M, et al. Robust license plate recognition using neural networks trained on synthetic images[J]. Pattern Recognition, 2019, 93: 134-146.
 - [32] GAO F, CAI Y C, GE Y S, et al. EDF-LPR: a new encoder-decoder framework for license plate recognition[J]. IET Intelligent Transport Systems, 2020, 14(8): 959-969.
 - [33] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681.
 - [34] GRAVES A, FERNÁNDEZ S, GOMEZ F, et al. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks[C]//Proceedings of the 23rd International Conference on Machine Learning. New York: ACM Press, 2006: 369-376.
 - [35] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1800-1807.
 - [36] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
 - [37] SHI B G, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2298-2304.
 - [38] ZHANG Y P, NIE S, LIU W J, et al. Sequence-to-sequence domain adaptation network for robust text image recognition[C]//Proceedings

- of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 2735-2744.
- [39] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks[J]. arXiv preprint, 2016, arXiv:1506.02025.
- [40] YANG M K, GUAN Y S, LIAO M H, et al. Symmetry-constrained rectification network for scene text recognition[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9146-9155.
- [41] LU N, YU W W, QI X B, et al. MASTER: multi-aspect non-local network for scene text recognition[J]. Pattern Recognition, 2021, 117: 107980.
- [42] BA J L, KIROUS J R, HINTON G E. Layer normalization[J]. arXiv preprint, 2016, arXiv:1607.06450.
- [43] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. arXiv preprint, 2015, arXiv:1506.01497.
- [44] ŠPAÑHEL J, SOCHOR J, JURÁNEK R, et al. Holistic recognition of low quality license plates by CNN using track annotated data[C]//Proceedings of 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Piscataway: IEEE Press, 2017: 1-6.
- [45] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European conference on Computer Vi-

sion. Heidelberg: Springer, 2016: 21-37.

- [46] MASOOD S Z, SHU G, DEHGHAN A, et al. License plate detection and recognition using deeply learned convolutional neural networks[J]. arXiv preprint, 2017, arXiv:1703.07330.

[作者简介]



曾淦雄（1996—），男，福州大学数学与计算机科学学院硕士生，主要研究方向为计算机视觉、模式识别等。



柯道（1983—），男，博士，福州大学数学与计算机科学学院副教授，主要研究方向为计算机视觉、机器学习与模式识别等。