

AI 设计下的智能驾驶场景文本识别技术

梁敏¹, 秦海波¹, 覃京燕¹, 殷绪成^{1,2}

(1.北京科技大学, 北京 100083; 2.北京科技大学顺德研究生院, 佛山 528399)

摘要: **目的** 交通标志识别作为智能驾驶、交通系统研究中的一项重要内容, 具有较高的理论价值和应用前景。尤其是文本型交通标志, 其含有丰富的高层语义信息, 能够提供极其丰富的道路信息。因此通过设计并实现一套新的端到端交通标志文本识别系统, 达到有效缓解交通拥堵、提高道路安全的目的。**方法** 系统主要包括文本区域检测和文字识别两个视觉任务, 并基于卷积神经网络的深度学习技术实现。首先以 ResNet-50 为骨干网络提取特征, 并采用类 FPN 结构进行多层特征融合, 将融合后的特征作为文本检测和识别的共享特征。文本检测定位文本区域并输出候选文本框的坐标, 文字识别输出词条对应的文本字符串。**结果** 通过实验验证, 系统在 Traffic Guide Panel Dataset 上取得了令人满意的结果, 行识别准确率为 71.08%。**结论** 端到端交通标志文本识别非常具有现实意义。通过卷积神经网络的深度学习技术, 提出了一套端到端交通标志文本识别系统, 并在开源的 Traffic Guide Panel Dataset 上证明了该系统的优越性。

关键词: 智能驾驶; 交通标志; 文本检测; 文字识别; 深度学习

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2021)06-0013-07

DOI: 10.19554/j.cnki.1001-3563.2021.06.003

Text Recognition Technology of Intelligent Driving Scene Based on AI Design

LIANG Min¹, QIN Hai-bo¹, QIN Jing-yan¹, YIN Xu-cheng^{1,2}

(1.University of Science and Technology Beijing, Beijing 100083, China;

2.Shunde Graduate School, University of Science and Technology Beijing, Foshan 528399, China)

ABSTRACT: As an important content in the research of intelligent driving and traffic system, traffic signs recognition system has great theoretical value and application prospects. Especially, the text-based traffic signs contain rich high-level semantic information, which can provide rich road information. Therefore, according to the design and implementation of a new end-to-end text-based traffic signs recognition system, it can effectively alleviate traffic congestion and improve road safety. The system mainly includes two visual tasks: text area detection and character recognition. The system is implemented based on deep learning technology of convolutional neural network. First, resnet-50 is used as the backbone network for feature extraction, and the FPN-like structure is used for multi-layer feature fusion. The fused features are used as shared features for text detection and recognition. The text detection block locates text areas and outputs coordinates of candidate bounding boxes. The character recognition block outputs the text strings corresponding to the entry. The experimental results show that the system achieves great performance on the Traffic Guide Panel Dataset, and the F1-score is 71.08%. End-to-end text-based traffic signs recognition system is of great practical significance. According to the deep learning technology of convolutional neural network, we propose an end-to-end text-based traffic signs recognition system. Extensive experiments on publicly available dataset demonstrate the state-of-the-art performance of our method.

KEY WORDS: intelligent driving; traffic signs; text detection; character recognition; deep learning

收稿日期: 2020-12-10

基金项目: 长江学者奖励项目 (FRF-TP-18-010C1); 国家重大专项课题 (2018YFB0704301); 北科大顺德研究生项目 (BK19AE011)

作者简介: 梁敏 (1994—), 女, 山西人, 北京科技大学博士生, 主要研究方向为场景图像文本检测。

通信作者: 殷绪成 (1977—), 男, 湖南人, 博士, 北京科技大学、北京科技大学顺德研究生院教授、博士生导师, 主要研究方向为模式识别与计算机视觉、文字识别及文档图像分析与识别、信息检索与自然语言处理、人工智能芯片技术。

随着科学技术的发展和社会经济的进步,智能驾驶、智能交通的出现大大改变了人们的生活方式,同时吸引了大批的研究工作者投入到这一研究领域。其中,交通标志检测与识别是智能交通系统中最关键的子系统之一。交通标志是一种用图形和文本来传递特定信息的公共标识,用以管理交通、指示方向等,从而保证道路畅通与行车安全。充分理解每个交通标志的含义非常重要,尤其是文本型交通标志,其在辅助驾驶、智能导航、自动驾驶、交通标志维护等领域都具有广泛的应用。本文的研究内容是文本型交通标志的检测与识别任务。常用的方法有两种:一种是基于传统的方法,将人工特征提取与机器学习相结合,人工提取特征容易理解,但不适用于复杂环境;另一种是基于深度学习的方法,利用卷积神经网络来提取特征,深度学习模拟了人脑的信息处理过程,其层次结构比前者能更好地处理环境信息,具有更强的鲁棒性和准确度。因此,本文提出一套基于深度学习的端到端系统,用来检测和识别文本型交通标志,并通过实验证明了系统的有效性。

1 文本检测与识别相关工作

1.1 文本检测

基于传统的文本检测方法主要可以分为两类:基于滑动窗口的文本检测方法^[1-2]和基于连通域分析的文本检测方法^[3-4]。基于滑动窗口的文本检测方法,首先利用设计的滑动窗口依次扫描图像,并将每个窗口覆盖的图像区域视为文本候选区域,然后提取文本候选区域的特征,再根据特征分类器得出文本置信度值,最后通过阈值过滤实现文本定位和背景区域分割。基于连通域分析的文本检测方法,首先采用不同的连通域分析方法,将提取到的许多连通域作为字符候选,然后连接这些字符候选,从而实现文本定位。

随着深度学习的发展,基于深度学习的文本检测算法逐渐受到人们的关注。目前,文本检测领域基于深度学习的算法主要有两大主流方法:基于锚(Anchor-based)的文本检测方法和基于像素(Pixel-based)的文本检测方法。

锚一般指锚框(Anchor Box),可以将其视作预设的具有不同尺寸和长宽比的“滑动窗口”,最早出自Faster R-CNN^[5]。其通过卷积神经网络提取特征,利用锚框的尺寸和长宽比分别表示目标的大小和形状,并根据锚框中点和特征图上与原图目标标注框中点(Ground Truth Bounding Box)对应位置的点的偏移值来定位目标的位置。在文本检测领域,文本具有不同于物体的特性,如文本本长宽比变化大、字体大小变化大、方向多变等。为了使锚框的方法适用于文本检测,Tian Zhi等人^[6]将文本切分成了一个一个小竖条,设计了不同高度、等宽的锚框来表示这些小竖条,最后

将小竖条连接得到文本检测框。Liao Ming-hui等人^[7-8]设计了更多长宽比的锚框,TextBoxes只能检测水平文本,而TextBoxes++在此基础上又回归了锚框到文本框四个点的距离,用以检测多方向文本。然而,基于锚点的文本检测方法需要人工设计繁多的锚框来适应文本,大大制约了检测性能。

基于像素的文本检测方法,一类方法根据深度神经网络提取的特征,进行像素级分类预测和回归,即对每一个像素点不仅进行分类预测,还需回归整个文本框^[9-10]。这类方法抛弃了锚框,使网络更加简单快速,但也存在一些问题,如正负样本严重不平衡,以及回归值域的变化大,使得长文本回归不全。另一类方法将文本检测任务视为实例分割任务,输入图像预测像素级标签,然后通过阈值过滤、聚类等后处理操作,将同一文本实例的像素聚合起来,输出最终的文本框。Deng Dan等人^[11]在进行像素级文本/非文本分类的同时,能预测相邻像素是否连接进而构成文本框。这类基于分割的方法对文本形状、文本长度等不敏感,效果更加鲁棒,但是后处理过程往往非常繁琐和耗时。

1.2 文字识别

OCR在扫描文档图像的识别中得到了很好的发展,但由于受到复杂的背景、多样的字体及不完善的拍摄条件等因素的影响,自然场景下的文字识别仍然面临着诸多挑战。在文字识别算法发展的过程中,有基于传统的方法和基于深度学习的方法。

基于传统的方法通常采用自底向上的方式,先对字符进行检测和分类,然后利用启发式规则、语言模型或词汇表将其分组成一个单词或文本行。此类方法采用人工设计的算子进行特征提取,然后使用这些特征训练分类器,例如SVM分类器。ESTR模型^[12]使用多尺度的滑动窗口提取HOG特征并分类。Strokelets模型^[13]通过对字符的某一部分学习,提取出中间特征,然后利用中间特征进行字符分类。传统的方法需要手工设计各种特征算子,这些算子无法高效表示高层次的特征。

基于深度学习的方法大多采用自上而下的方式,直接识别单词或文本行。Jaderberg Max等人^[14]将整个单词视为一个类,将单词识别问题转化为图像分类问题,并设计了神经网络对90 000个英文单词图片进行分类识别。目前,大多数研究均将文字识别问题视为序列预测问题,认为图片上具有空间位置关系的字符具有某种时序关系,这种上下文的时序信息在预测字符时能明显地提高识别准确率。基于序列的算法大致可以分为基于时间联结主义分类(CTC)解码和基于注意力(Attention)解码两种。CTC首先在语音识别领域被提出,其可以在输入和输出无需对齐的情况下,端到端地训练神经网络,之后便被引入到文字

识别领域。CRNN 模型^[15]提出使用 CNN 和 RNN 去编码序列特征,并将 CTC 损失用于字符对齐。现有方法更多采用的是,基于注意力机制的编码器和解码器的框架。注意力机制首先应用于机器翻译领域,后续被广泛地应用到各种序列任务中去,将其用于文字识别可以在解码时重点关注当前字符区域的视觉信息,从而提升文字识别的准确率。R2AM^[16]通过 CNN 提取图像视觉特征,RNN 用于建模序列上下文信息,并采用基于注意力的解码器预测输出。Cheng Zhan-zhan 等人^[17]首次提出注意力解码模型存在注意力漂移的问题,并提出了一种聚焦注意力,并且取得了更好的效果。以上的方法都默认文本行是水平排列的,难以处理那些弯曲或者仿射的文本图片,一些研究在常规文字识别方法的基础上进行了改进,以适用于任意形状的文字识别。ASTER 模型^[18]提出先基于空间变换网络 STN^[19]对文本进行校正,再将其作为水平文本进行处理。SAR 模型^[20]通过引入二维空间上的注意力机制,从另一种角度解决了不规则文本的识别问题。

1.3 端到端文本识别

近年来,端到端文本检测与识别算法取得了不错的效果,成为研究的一大热点。一方面文本检测和文本识别可以共享底层特征,这降低了检测到识别过程的运算参数;另一方面通过反向传播算法,文本识别的损失能够进一步优化底层特征的提取和文本检测的结果,这使得端到端训练方式成为现在很多算法,用来进一步提升文本检测性能的手段。端到端文本检测与识别算法的一般步骤为首先利用文本检测网络得到文本检测候选框;然后对文本检测候选框区域进行校正并输入后续的识别模型;最终输出检测和识别的结果。目前,已经出现了许多优秀的端到端文本检测与识别方法。Qiao Liang 等人^[21]提出了一种方法用于端到端检测、任意形状文本识别,该方法的检测模型采用的是基于分割的文本检测方法,其识别模型采用的是 CRNN,其识别模型的误差可以回传给检测模型用于检测模型的进一步优化。CRAFTS^[22]生成字符级标注的数据集并进行半监督训练,首先检测模型会得到字符级预测结果,然后通过修正网络迭代进行修正,并将得到的字符注意力特征输入双向 LSTM 网络进行序列建模,再输入基于注意力的解码器进行文本识别。将文本检测和文本识别融合到同一个框架中已经成为一种趋势,也更能满足人们的实际需求。然而,端到端文本检测与识别方法仍然存在一些如训练困难等问题,这些问题值得人们更加深入的研究和探索。

2 端到端交通标志文本识别系统设计

端到端交通标志文本识别系统,面向车载环境,

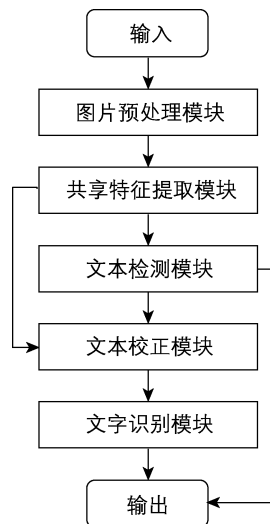


图 1 系统结构

Fig.1 System structure

以车载摄像头拍摄的图片为输入,主要包括五个模块:图片预处理模块、共享特征提取模块、文本检测模块、文本校正模块和文字识别模块。系统接收到图像,先将其转换成系统内部可用的数据格式,然后进行归一化等一系列预处理操作,之后输入到共享特征提取模块,利用深度学习技术提取得到特征图,特征图作为文本检测和文本识别的依据,检测部分得到候选文本框,经过几何变换等校正之后,识别候选文本框中的文本内容,最终文本检测候选框和对应的文本识别结果作为系统的输出结果,用于为驾驶员提供准确的道路信息。系统结构见图 1。

2.1 图片预处理模块

自然场景图片中的光照、模糊、尺度等因素对文本检测和识别有较大影响。该模块负责先将外部数据类型转化为系统内部可用的数据类型,然后对输入图片进行一系列处理,如规范尺寸和维度、采样、归一化、仿射变换、特征提升等,以提高文本检测和识别性能。

2.2 共享特征提取模块

端到端系统将文本检测和文字识别统一在同一个框架中,其中,特征提取模块是文本检测模块和文字识别模块共享的。该模块利用深度神经网络提取图片特征,输出多维的特征图,用于文本区域定位和文字识别。

2.3 文本检测模块

该模块负责利用提取的特征进行是否为文本区域的判断,进而实现文本区域的定位,输出候选文本框的坐标,用于后续的文字识别。

2.4 文本校正模块

该模块负责利用文本检测模块得到的候选文本

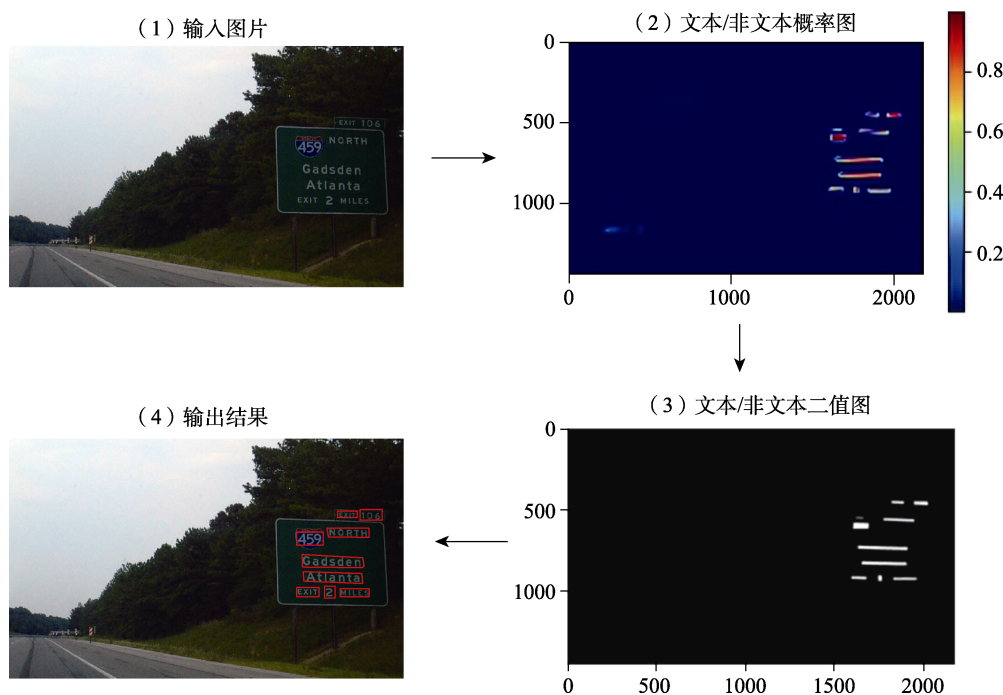


图2 文本检测模块示意
Fig.2 Schematic of text detection module

框坐标和共享特征,提取模块得到的共享特征,进行文本区域特征提取和校正,之后将校正后的特征输入到文字识别模块。这是因为自然场景中存在不规则形状的文本词条,直接识别弯曲或仿射的文本往往不能取得较好的效果,经过校正之后的水平排列的文本更有利于识别。

2.5 文字识别模块

该模块负责识别文本校正模块输出的区域特征,并输出词条对应的文本字符串。

3 端到端交通标志文本识别系统实现

3.1 图片预处理模块

图片预处理模块拟针对自然场景图片存在光照、模糊、尺度等问题,对数据进行相应的预处理。先将外部输入数据转换为系统内部可用的数据格式,训练时对图片进行翻转、裁剪、旋转、归一化、尺寸规范和维度规范处理,测试时直接对输入图片进行归一化、尺寸规范和维度规范处理。

3.2 共享特征提取模块

共享特征提取模块通过卷积神经网络提取图片特征。以 ResNet-50 为骨干网络,提取多层特征图。其中,低层的特征图包含更多的位置、细节信息,而高层特征图包含更多的语义信息,为了进一步使特征丰富化,该模块采用类 FPN 结构对提取的多层特征进行特征融合,并将融合后的特征图作为该模块的输出。



图3 校正模块示意
Fig.3 Schematic of rectification module

3.3 文本检测模块

文本检测模块采用基于分割的文本检测方法。依次对共享特征提取模块输出的特征进行扫描,判断每一个像素是否属于文本,输出一张概率图,值域区间为 $[0,1]$,代表每一个像素点属于文本的概率值。同时,该模块通过卷积神经网络预测像素级阈值,不同于人工设计固定阈值进行二值化的方法,该模块能自适应学习每个像素位置的阈值,然后通过二值化、聚类、扩张等一系列后处理操作,得到最终的候选文本框。该模块不仅大大简化了后处理过程,而且能够自适应学习,更加高效和有效。文本检测模块示意图见图2。

3.4 文本校正模块

文本校正模块首先根据文本检测模块得到的候选文本框,将共享特征上的文本区域的特征提取出来,并进行采样以规范其尺寸和维度,之后对该模块使用薄板样条插值(TPS)进行进一步校正。薄板样条插值是一种插值算法,常用于图像变形,通过少量控制点就可以驱动图像变化。当输入一张单词区域时,就提供其字符区域映射和链接映射,它们包含了文本区域的几何信息,使用薄板样条插值变换之后就可以得到想要的结果。校正模块示意图见图3。

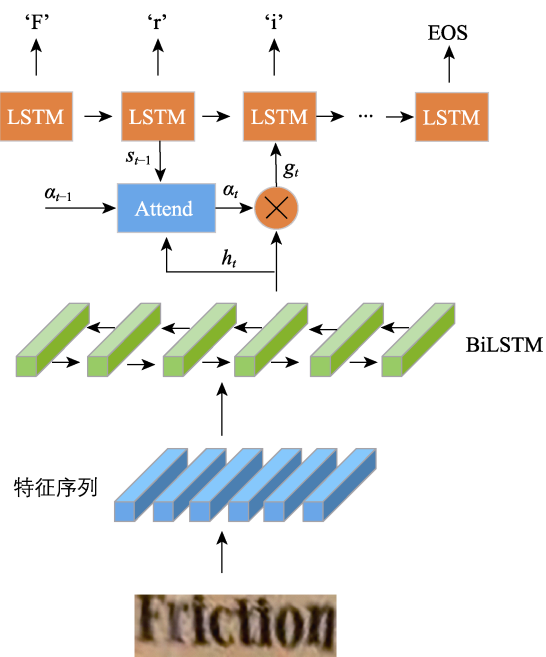


图 4 识别模块结构

Fig.4 Structure of recognition module

3.5 文字识别模块

文字识别模块采用基于注意力机制的文本识别方法。编码器特征采用校正过后的共享视觉特征，每个文本区域表示一个特征向量序列，同时采用一个双向多层的 LSTM 进行序列信息的建模以获得更大的感受野信息。解码器端同样使用 LSTM 进行循环解码，在解码时结合注意力机制，将每一步解码关注的重点集中在图像上相关的字符区域，提升文字识别的准确率。同时为了缓解注意力机制本身带来的注意力漂移的问题，引入历史时刻的注意力关注权重，作为先验指导当前时刻注意力权重的计算。识别模块结构见图 4。

注意力机制与人类在阅读文字时的机制很类似，一般不会将词条当成一个整体去看，而是在解码每个字时，倾向于根据所需选择性地获取词条中对应部分的信息。基于注意力机制的文本识别方法也是目前性能最好的一类方法。

4 实验验证

4.1 数据集

交通标志开源的数据集大部分是图形和文本混杂的数据集，其中更多针对的是图形交通标志的识别，针对文本型交通标志的数据集甚少，开源的数据集几乎没有。本文选择在 Traffic Guide Panel Dataset (TGPd) [23] 上进行实验验证。TGPd 是一个交通指示牌识别的基准数据集，由车载摄像头拍摄得到，包含数千个英文文本的高速公路指示牌。然而，仅发布了 404 张测试图像及其相应的标注，训练图像不可用。

表 1 实验数据
Tab.1 The experimental data

方法	检测			端到端	
	召回率	准确率	F 值	F1 值	训练集
DF ^[24]	71%	59%	64%	—	TGPd
CLNet ^[23]	64%	73%	68%	—	TGPd
AAM ^[25]	78.92%	70.28%	75.06%	—	MLT
Ours	83.49%	74.17%	78.55%	71.08%	MLT

4.2 评测指标

现有文本检测性能评测指标主要有三个，分别为召回率 (Recall)、准确率 (Precision)、F 值 (F-measure)。召回率衡量的是文本目标检测的查全率，是指在实际为正的样本中被预测为正样本的概率。准确率衡量的是文本目标检测的查准率，是指在所有被预测为正的样本中实际为正样本的概率。F 值是召回率和准确率的加权调和平均，由于文本检测召回率和文本检测准确率的相互影响，所以在两个指标都要求取高或者两个指标发生冲突时，一般用 F 值来衡量，该值是评价文本检测方法性能的综合指标。若记正确检测的文本目标数量为 TP (True Positive)，未检测到的文本目标数量为 FN (False Negative)，非文本目标检测为文本的数量为 FP (False Positive)，则召回率、准确率和 F 值的公式可写成：

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (3)$$

现有文本端到端检测与识别的性能评价指标为 F1 值 (F1-score)，F1 值是同时根据文本检测结果和文本识别内容决定的。文本检测结果是指检测网络预测为文本且与真实文本相匹配的候选框，文本识别内容是指将预测文本内容与真实文本内容直接比较，若完全一致，则认为检测正确。

4.3 实验验证

通过使用本文提出的系统对文本型交通标志进行检测和识别，在 TGPd 上进行实验，验证系统性能。实验数据见表 1，实验结果见图 5。

RONG Xue-jian 等人 [23] 提出 TGPd，并用该数据集进行训练。由于 TGPd 仅开源了其测试集，没有公开的训练集，HOU Jie-bo 等人 [25] 采用了 ICDAR2017 MLT 数据集进行训练。本文同时使用 ICDAR2015 数据集和 ICDAR2017 MLT 数据集共 8 200 张图像及其标注来进行训练。由表 1 的数据可以看出，本文所述方法无论是在检测性能还是端到端性能都有更优的表现。因此可以得到结论，本文提出的端到端交通标志文本识别系统在 TGPd 上是有效的。



图 5 实验结果

Fig.5 Experimental results

5 结语

在社会经济不断发展的同时,交通安全和交通堵塞随之成为了日益严重的社会问题,智能驾驶、智能交通系统应运而生,文本型交通标志识别是智能交通系统中的重要一环,也是难度较大的自然场景文本识别问题之一。本文提出了一套完整的端到端交通标志文本识别系统,为文本型交通标志识别提出了一种新的解决方案,并在 TGPLD 上通过实验验证了该系统的有效性。该系统能在车载环境下识别文本型交通标志,可以为驾驶员提供准确的交通信息,实现辅助驾驶、智能导航、自动驾驶等功能,是智能驾驶、智能交通系统中必不可少的部分。未来,系统会在实时性方面进行研究,并在准确性方面进一步优化,以达到实际应用的需求。

参考文献:

- [1] LEE J J, LEE P H, LEE S W, et al. AdaBoost for Text Detection in Natural Scene[C]. Beijing: International Conference on Document Analysis & Recognition, 2011.
- [2] YIN X C, ZUO Z Y, TIAN S, et al. Text Detection, Trac-

king and Recognition in Video: a Comprehensive Survey[J]. IEEE Transactions on Image Processing, 2016, 25(6): 2752-2773.

- [3] YI C, TIAN Y. Text Extraction from Scene Images by Character Appearance and Structure Modeling[J]. Comput Vis Image Underst, 2013, 117(2): 182-194.
- [4] YIN X C, YIN X, HUANG K, et al. Robust Text Detection in Natural Scene Images[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 36(5): 970-983.
- [5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards Real-time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137-1149.
- [6] TIAN Z, HUANG W, HE T, et al. Detecting Text in Natural Image with Connectionist Text Proposal Network[C]. Amsterdam: Proc. ECCV, 2016.
- [7] LIAO M, SHI B, BAI X, et al. TextBoxes: a Fast Text Detector with a Single Deep Neural Network[C]. San Francisco: Proc. AAAI, 2017.
- [8] LIAO M, SHI B, XIANG B. TextBoxes++: a Single-shot Oriented Scene Text Detector[J]. IEEE Transactions on Image Processing, 2018, 27(8): 3676-3690.

- [9] ZHOU X, YAO C, WEN H, et al. EAST: an Efficient and Accurate Scene Text Detector[C]. Honolulu: Proc. CVPR, 2017.
- [10] HE W, ZHANG X Y, YIN F, et al. Deep Direct Regression for Multi-oriented Scene Text Detection[C]. Venice: 2017 IEEE International Conference on Computer Vision (ICCV), 2017.
- [11] DENG D, LIU H, LI X, et al. PixelLink: Detecting Scene Text via Instance Segmentation[J]. 2018.
- [12] WANG K, BABENKO B, BELONGIE S. End-to-end Scene Text Recognition[C]. Barcelona: IEEE International Conference on Computer Vision, 2012.
- [13] YAO C, BAI X, SHI B, et al. Strokelets: a Learned Multi-scale Representation for Scene Text Recognition[C]. Columbus: IEEE Conference on Computer Vision and Pattern Recognition, 2014.
- [14] JADERBERG M, SIMONYAN K, VEDALDI A, et al. Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition[J]. 2014.
- [15] SHI B, 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.
- [16] WOJNA Z, GORBAN A N, LEE D S, et al. Attention-based Extraction of Structured Information from Street View Imagery[C]. Kyoto: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017.
- [17] CHENG Z, BAI F, XU Y, et al. Focusing Attention: towards Accurate Text Recognition in Natural Images[C]. Venice: Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [18] SHI B, YANG M, WANG X, et al. ASTER: an Attentional Scene Text Recognizer with Flexible Rectification[J]. Pattern Anal, 2019: 2035-2048.
- [19] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial Transformer Networks[C]. Montreal: In Advances in Neural Information Processing Systems, 2015.
- [20] ZHAN F, LU S. Esir: End-to-end Scene Text Recognition via Iterative Image Rectification[C]. Long Beach: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
- [21] QIAO L, TANG S, CHENG Z, et al. Text Perceptron: towards End-to-End Arbitrary-Shaped Text Spotting[C]. New York: Proc. AAAI, 2020.
- [22] BAEK Y, SHIN S, BAEK J, et al. Character Region Attention for Text Spotting[C]. Glasgow: Proc. ECCV, 2020.
- [23] RONG X, YI C, TIAN Y. Recognizing Text-based Traffic Guide Panels with Cascaded Localization Network[C]. Amsterdam: Proc. ECCV, 2016.
- [24] JADERBERG M, VEDALDI A, ZISSERMAN A. Deep Features for Text Spotting[C]. Zurich: Proc. ECCV, 2014.
- [25] HOU J B, ZHU X B, LIU C, et al. Detecting Text in Scene and Traffic Guide Panels with Attention Anchor Mechanism[J]. IEEE Transactions on Intelligent Transportation Systems, 2020.