

DOI:10.19651/j.cnki.emt.2209494

## 基于深度学习的交通标志文字信息检测与识别方法\*

王立刚<sup>1</sup> 张志佳<sup>1</sup> 贺继昌<sup>1</sup> 张长海<sup>1</sup> 赵永茂<sup>2</sup>

(1. 沈阳工业大学人工智能学院 沈阳 110870; 2. 沈阳美行科技有限公司 沈阳 110167)

**摘要:** 为解决无人驾驶汽车外界环境感知系统对交通标识文字信息检测问题,提出一种在自动驾驶场景下对交通标识的文本信息进行检测并识别的两阶段方法,实现了自动驾驶信息精细化采集。首先使用 YOLO 检测器检测交通标识,同时使用本文改进的 DB 检测网络对场景内文本进行检测,将交通标识检测结果与场景文本检测结果进行交集运算得到待识别文本区域;最后使用轻量化 CRNN 网络对待识别区域文本进行识别。使用 CSCT-1600 数据集和 MTWI-2018 数据集分别进行训练和测试。实验结果表明,交通标识信息定位算法在召回率为 92.98 时精确度为 94.95%,交通标识信息识别算法在 F1 为 77.2% 时识别速度为 25 帧。

**关键词:** 交通标识;YOLO;文本检测;文本识别

**中图分类号:** TP391.41 **文献标识码:** A **国家标准学科分类代码:** 510.40

## Traffic sign information extraction combined with YOLO detection and text detection

Wang Ligang<sup>1</sup> Zhang Zhijia<sup>1</sup> He Jichang<sup>1</sup> Zhang Changhai<sup>1</sup> Zhao Yongmao<sup>2</sup>

(1. School of Artificial Intelligence, Shenyang University of Technology, Shenyang 110870, China;

2. MXNAVI Co., Ltd., Shenyang 110167, China)

**Abstract:** In order to solve the problem of detecting the text information of traffic signs by the external environment perception system of unmanned vehicles, a two-stage method for detecting and recognizing the text information of traffic signs in an autonomous driving scenario is proposed, which realizes the refined collection of autonomous driving information. First, use the YOLO detector to detect traffic signs. At the same time, use the improved DB detection network in this article to detect the text in the scene. The intersection of the traffic sign detection results and the scene text detection results to get the text area to be recognized; finally, the lightweight CRNN network is used to treat Recognize area text for recognition. Use CSCT-1600 data set and MTWI-2018 data set for training and testing respectively. The experimental results show that the accuracy of the traffic sign information positioning algorithm is 94.95% when the recall rate is 92.98, and the recognition speed of the traffic sign information recognition algorithm is 25 frames when the F1 is 77.2%.

**Keywords:** traffic signs;YOLO;text detection;text recognition

## 0 引言

近年来,深度学习技术在自动驾驶等许多场景中得到广泛应用<sup>[1-3]</sup>。目前对交通标识的研究主要集中在符号交通标识的检测与识别,对于含有中文文字信息的交通标识的研究相对较少。带有文字的交通标识牌通常设计为颜色鲜艳区别于周围环境样式,其中的文字颜色与交通标识牌颜色对比明显以起到醒目作用。但在自动驾驶场景下,诸多不确定因素,如天气、光照、遮挡以及交通标识角度等因

素都会给检测带来不利影响。同交通标识牌等目标相比,交通标识上的文字大小、高度、宽度以及排版等因素也会对检测带来巨大挑战。

现阶段交通标识检测与识别主要以使用卷积神经网络算法为主要研究方向。文献[4]使用 EIOU 损失函数代替 YOLOv5 算法所使用的 GIOU 损失函数来优化训练模型,提高算法的精度,实现对目标更快速的识别,然后使用加权 Cluster 非极大值抑制(NMS)改进 YOLOv5 本身所使用的加权 NMS 算法,提高生成检测框的准确率比原始的

收稿日期:2022-04-02

\* 基金项目:国家自然科学基金(61540069)、教育部创新基金(2020HYA02007)项目资助

YOLOv5 算法提高了 6.23%。文献[5]针对智能驾驶中出现的交通场景多目标检测与分割效率低、鲁棒性差等问题,提出一种改进的 Mask R-CNN 交通场景多目标快速检测与分割方法,改进网络能够准确实现多目标的检测与分割,平均检测精度可达 85.2%。文献[6]使用替换特征提取网络为 VGG16 与 ResNet-50 网络的 Mask R-CNN 算法端到端的对交通标识进行检测与识别,其优化后的算法在德国交通标志数据集<sup>[7]</sup>有低于 3% 的错误率。文献[8]提出一种交通标识检测网络语义嵌入的方法。其从符号领域知识创建嵌入,引入语义距离的概念:测量语义空间置信度的实值分数。在交通标志分类器上评估了此分数,其实现了接近最先进的性能,同时比其他置信度分数的计算速度要快得多。文献[9]提出了一种新的数据库生成方法,只需要 1)任意自然图像,即不需要来自感兴趣领域的真实图像,以及 2)交通标志的模板,即合成创建的模板来说明交通标志类别的外观。生成的训练数据库在德国交通标志上使用 Faster R-CNN 平均达到 95.66% 的 mAP。文献[10]提出了一种用于交通标志识别的新型轻量化的卷积神经网络架构。在提出的架构中,每个卷积层包含少于 50 个特征,即使没有图形处理单元的帮助,卷积神经网络也能快速训练。在德国交通标志识别基准和比利时交通标志分类数据集,所提出的架构优于最先进的交通标志方法的性能,在单个端到端网络中用于训练的参数至少少五倍。

现阶段交通标识上的文字检测主要以自然场景文本检测算法为主要研究方向。文献[11]使用全卷积网络 (fully convolutional networks, FCN) 并融合多尺度文字特征图,结合语义分割的方法分割文字候选区域,利用分割得到的文字候选区域直接获取文字候选检测框并进行扩大补偿处理,对文字候选检测框进行后处理得到最终检测结果。其使用方法未对弧形排列文字、栏杆或条纹窗户等非文字负类进行抑制检测存在误检现象。文献[12]使用改进的 DenseEAST 网络与 CRNN-X 对车载屏幕信息进行检测与识别,其检测与识别准确率在其自建数据库可达良好的检测效果。检测准确率为 97.3%,识别准确率在 97.6%。此方法应用于车内场景,在自然场景下检测效果会由于光照等因素降低。文献[13]使用 ResNet-50 为骨干网络提取特征,并采用类 FPN(feature pyramid networks)<sup>[14]</sup> 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,文本检测区域定位文本位置并由文本识别网络进行识别。该方法与端到端方法相比具有明显优势。

以上所介绍使用文本检测与文本识别方法的应用在其自建或公开数据集效果显著,但还存在以下两点不足:

1)目前的交通标志检测与识别主要以神经网络算法对交通标志进行检测定位并对检测到的交通标志进行分类,对于交通标志上更精细化的信息未能做到更精确的提取;

2)在自动驾驶场景下面对外界环境干扰与文本自身特点更快更准的检测到目标信息成为巨大的挑战。

针对以上问题,本文设计一个两阶段交通标志文本信息检测系统实现交通信息精细化采集,对 DBNet<sup>[15]</sup> 进行感受野与特征融合增强两方面改进,针对场景中的非感兴趣文本信息例如街道广告文本、商铺牌匾等信息使用 YOLO 检测器检测场景中的交通标志同时与交通标志文本区域做交集处理。将检测到的文本信息使用 CRNN<sup>[16]</sup> 进行识别后输出识别信息。算法能对复杂场景下的交通标志信息进行提取。

## 1 系统构成

本文设计了一种两阶段系统对交通标识信息提取与识别,其中交通标志检测模块将输入图像中的交通标志检测出,同时文本检测模块中的检测网络检测出输入图片中的文本框。经过 ROI 筛选后将交通标识上的文本框信息使用文本识别模块算法进行识别信息提取,系统框架如图 1 所示。

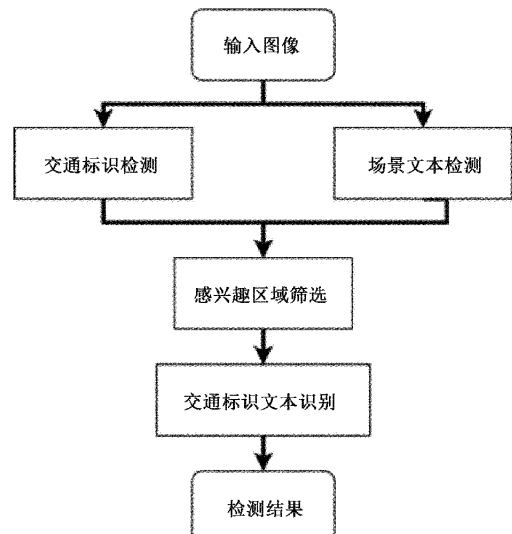


图 1 系统结构

### 1.1 交通标志区域定位

YOLO 检测算法通过单步回归坐标框完成检测目标区域的预测。YOLO v5 算法使用 3 种不同大小特征图单步回归完成对目标预测,与两阶段检测算法相比存在检测速度快但小目标检测能力弱等特点。但在自动驾驶场景下,交通标识所占图像比例较大,可忽略小目标检测能力弱这一特点。

由于交通标识上的文本信息所占图像比例较少且由于文本属于非刚体,与通用目标相比字体区域非连贯且文本之间带有语义关联性,因此交通标识上文本检测应先定位到交通标识然后再定位到交通标识文本。

YOLO v5 包括 4 种不同的结构:YOLO v5s、YOLO v5m、YOLO v5l、YOLO v5x。四种网络结构宽度与深度参数各不相同。由于本文先只对占图像比例较大的交通标识进行定位,因此使用 YOLO v5s 结构。如图 2 所示,与其

他 3 种结构相比其拥有最快的速度,但检测精度稍低,对于本文使用场景满足需求。

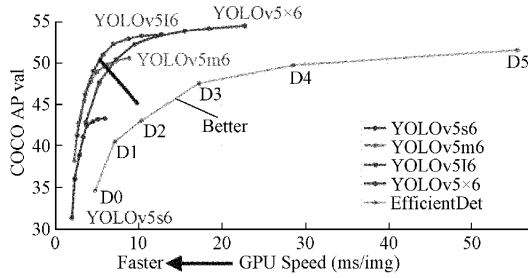


图 2 YOLOv5 性能指标

### 1.2 文本信息区域定位

本文通过语义分割精细化定位文本信息区域。并根据交通标识检测任务需求,通过结合 SPP (spatial pyramid pooling)<sup>[17]</sup> 结构与 DA<sup>[18]</sup> 结构增加 DBNet 检测精度。针对检测精度增加带来的速度降低问题使用网络量化加速文本检测算法。

#### 1) 文本语义分割

基于语义分割的文本检测网络分为特征提取与融合和后处理两阶段。在特征提取部分通过卷积与池化操作减少特征图尺寸,得到感兴趣特征后通过上采样特征图实现还原特征图分辨率。将得到的特征图通过后处理将特征图上每一个像素进行分类,通过及连等操作得到文本区域。

#### 2) 感受野增强的文本分割定位算法

基于分割的文本检测算法流程如图 3 所示,主要由主干网络、特征融合结构检测头以及后处理四部分组成。主干网络使用了移除了全连接层与分类层的 ResNet50<sup>[19]</sup> 网络;特征融合结构使用 FPN 结构将主干网络提取的不同等级特征图通过反卷积操作进行上采样并进行融合;将融合后的特征图使用检测头对特征图上逐像素进行分类;分类后的像素使用后处理进行像素级联操作形成文本区域最后输出文本区域预测框。

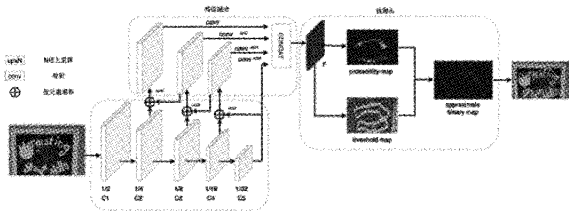


图 3 基于分割的文本检测

假设第  $i - 1$  层特征的感受野  $RF_{i-1}$ , 第  $i$  层感受野可通过式(1)计算得出:

$$RF_i = RF_{i-1} + (k_i - 1) \prod_{j=1}^{i-1} s_j \quad (1)$$

其中,  $k_i$  为第  $i$  层卷积核或池化核尺寸,  $s_j$  为第  $j$  层的卷积或池化步长。通过计算可得出 ResNet50 的感受野为  $427 \times 427$ 。输入图像尺度大于该尺寸时,图像边缘大于其

感受野部分特征信息存在使用不充分的不足。

池化通过设置池化核函数大小以达到对图像进行降维,去除特征图冗余信息的作用。将多个不同尺度的池化组合即可得到空间金字塔池化结构,如图 4 所示。结合多个不同尺度的池化操作可将任意尺度转换成相同尺度特征图。与卷积相比,池化具有计算量小的特点。与单个池化核函数相比,池化金字塔具有在单个特征图上提取多尺度特征的特点。

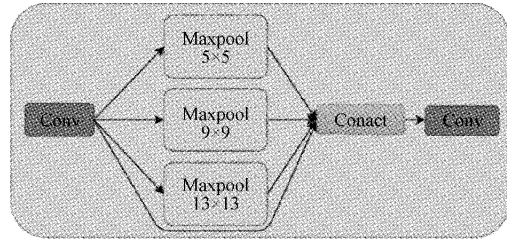


图 4 空间金字塔池化结构

基于空间金字塔池化结构特点,本文针对主干网络 ResNet50 网络感受野受限特点提出使用空间金字塔池化结构增加 ResNet50 网络感受野以缓解输入图像尺度大于感受野带来的限制。

#### 3) 特征融合增强的文本分割定位算法

主干网络 ResNet50 的 Conv\_stage 层输出输入图像  $1/4, 1/8, 1/16, 1/32$  大小特征图。单层特征图检测结构使用  $1/32$  下采样特征图即顶层特征进行目标预测,如图 5(a) 所示。但经过卷积网络的顶层特征图虽然拥有丰富的语义特征,但目标位置相对粗略;底层特征图虽然语义信息较少,但其目标位置相对准确。FPN 结构通过将顶层特征图进行上采样并结合上一等级特征图,得到的特征图既包含顶层特征语义信息,又包含底层特征位置信息,如图 5(b) 所示。

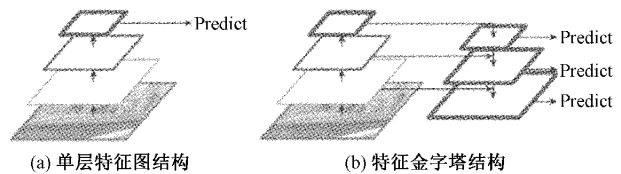


图 5 特征检测结构

虽然 FPN 结构能够融合顶层与底层信息,但是对于输入图像中不同目标没有明显侧重,对于想要增强的信息缺少语义加强。本文通过在 FPN 结构后增加 DA 注意力结构动态增加不同尺度语义特征在特征图上的位置权重以达到提高检测精度的目的。DA 结构和增加感受野与注意力的文本定位算法如图 6~7 所示。

### 1.3 文本语义信息提取

YOLO 检测器做区域限制下的文本信息定位算法可从输入图像上提取出精确位置的交通标识文本预测框,使用文本网络对精确定位的交通标志文本行区域进行识别。本文使用结合 CTC 损失函数的 CRNN 文本识别网络进行

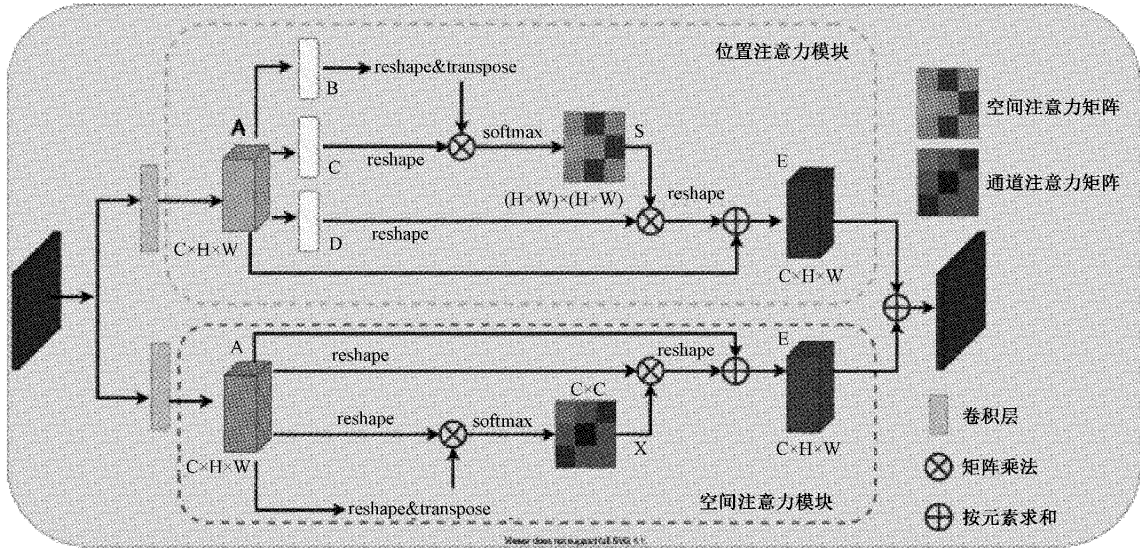


图 6 DA 结构

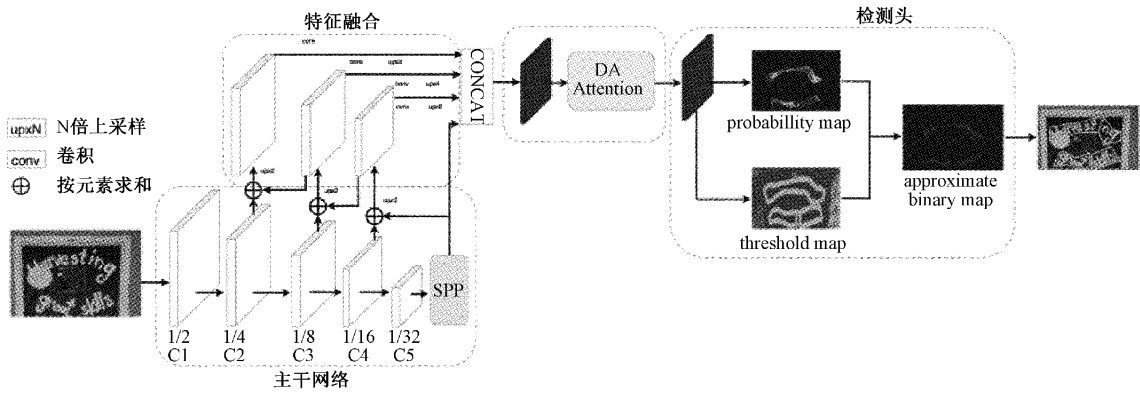


图 7 增加感受野与注意力的文本定位算法

文本图像语义转换输出,并将识别结果在图像上进行展示。文本语义特征提取流程如图 8 所示。



图 8 文本图像语义提取

CRNN 网络结构如图 9 所示。输入图像大小为 (32, 100, 3), 经过卷积网络 VGG 获得大小为 (1, 25, 512) 的卷积特征图。将所获得卷积特征图使用深层双向 LSTM 网络继续提取文字序列特征。获得每帧的预测经过翻译层输出最后的预测序列。

循环层使用的 Softmax 损失函数,每一列输出都需要对应一个字符,在训练时需要标出每一个字符的位置,再通过卷积网络感受野对其每一列对应的标签才能进行训练,如图 10 所示。在实际使用中,标记对其样本难度较高,工作量大。CTC 损失函数不要求训练数据对齐和一标注,直接输出不定长的序列结果。因此本文使用 CTC 损失函数代替原有损失函数。

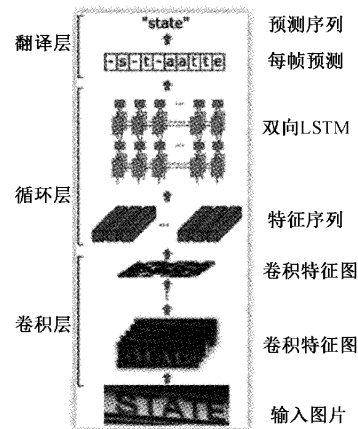


图 9 CRNN 网络结构

### 1.4 模型量化加速

网络训练后得到的模型为 FP32(单精度浮点数),为了优化检测速度,本文对 32 位模型进行 16 位量化。与 FP32 相比,FP16 的访存消耗仅为 1/2,因此 FP16 是更适合在计算能力较弱的移动终端侧进行 AI 计算的数据格式。在训

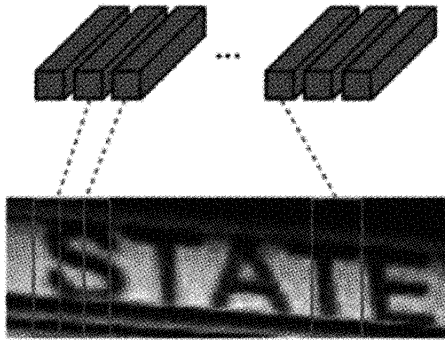


图10 感受野与标签对应关系

练模型时,首先使用 FP16 的输入,进行前向传播时使用 FP32 格式初始化权重,将恢复的 checkpoint(断点)转换成 FP16 进行前向传播。损失函数计算时涉及到指数、对数预算,使用 FP16 存在数据溢出不稳定现象,因此使用 FP32。在线量化优化算法已集成于 PyTorch 框架于 TensorRT 工具中,本文使用 PyTorch-1.6 以及 TensorRT-7.0 工具对模型进行训练于推理量化加速。

## 2 实验结果与分析

### 2.1 数据库介绍

BDD100K<sup>[20]</sup>:BDD100K 是由伯克利大学 AI 实验室 (BAIR)发布的目前最大规模、内容最具多样性的公开驾驶数据集。其中天气场景主要以晴天为主的多云、阴天、下雨、下雪、雾天等 6 种天气;以城市街道为主包括住宅区、公路、城市街道、停车场、加油站、隧道 6 种场景。本文所使用交通标志有 34 377 张图片。

CTW1500<sup>[21]</sup>:CTW1500 为 1 000 张训练图片和 500 张测试图片组成的文本检测数据集。该数据集为文本行标注的纯英文弯曲数据集。

MLT-2017<sup>[22]</sup>:MLT-2017 是一个多语言文本检测数据集。数据集中有 7 200 张训练图像、1 800 张验证图像和 9 000 张测试图像。该数据集包括代表 6 种不同文字的 9 种语言。

CTST-1600<sup>[23]</sup>:CTST-1600 是一个交通标志文本数据集。数据集有 1 600 张图像,其中训练集包含 1 320 张图像,测试集包含 280 张图像。数据集主要以背景复杂多样且拍摄于黄昏、夜晚等光线条件不佳的街景交通标志图片为主的街景交通标识图片,这些图片的,且包含部分图片,提高了检测难度。CTST-1600 共包含 1 600 张图片,其中,测试集包含 280 张图片,训练集包含 1 320 张图片。

### 2.2 实验参数设置

实验平台操作系统为 Ubuntu18.04 LTS,深度学习框架为 PyTorch。硬件环境 CPU 为 Intel(R) Core(R) I7-8700K,内存为 32 G, GPU 为 Nvidia (R) GTX (R) 1080Ti GPU。

### 2.3 交通标志文本区域定位实验

本文的实验数据来源于公开数据集,其中使用 BDD100K 数据集训练 YOLO v5s 网络;使用 CTW1500 验证改进的 DBNet,最后使用 MLT-2017 训练改进后 DBNet。使用 CTST-1600 验证由 YOLO v5s 与 MLT-2017 组成交通标志文本区域定位算法。

#### 1) 文本定位算法改进

文本定位通过对场景内文本区域像素进行预测,通过像素级联后处理成文本区域进行预测,因此定位算法自身分割准确度会对文本定位准确度造成影响。因此,将本文提出的结合感受野结构与注意力结构的分割算法进行对比实验。

本文对原始分割定位算法进行实验,实验结果如表 1 第一行数据所示。在原始分割网络基础上,本文针对感受野不足问题通过添加增感受野结构以达到提高网络性能表现,如表 2 第二行数据所示,增加感受野结构后的召回率提升 5.6%,准确率提升 0.5%,F1 得分提升 3.2%;针对 FPN 只融合各尺度特征但不能加强感兴趣特征的不足,本文使用 DA 注意力结构提升融合效果,如表 2 第 3 行所示,增加注意力结构后的召回率提升 2%,准确率提升 0.9%,F1 得分提升 2.1%。最后将两种结合分割算法实验如表 2 第 4 行所示,召回率提升 3.1%,准确率提升 4%,F1 得分提升 3.6%。

表 1 在 MSRA-TD500 和 CTW1500 数据集上对比实验结果

Backbone	SPP block	DA block	CTW1500		
			P	R	F
ResNet50	O	O	86.9	80.2	83.4
ResNet50	P	O	87.4	85.8	86.6
ResNet50	O	P	87.8	83.2	85.5
ResNet50	P	P	<b>90.0</b>	<b>84.2</b>	<b>87.0</b>

本文文本定位算法针对场景内文本区域进行自适应阈值分割文本,然后对阈值后的文本区域进行标注预测框输出。由于场景内的背景复杂,光线干扰等因素影响,存在分割不准确的问题。在实际场景中,文本区域分割如图 11 所示,其中左侧白色连通区域为文本分割结果,右侧为自动驾驶场景文本定位结果。

#### 2) 文本区域定位算法实验

使用 DBDD00K 数据集训练的 YOLO v5s 网络模型与改进后的 DBNet 模型组成的交通标志文本区域定位模型在 CTST-1600 交通标志文本数据集上进行算法验证,结果如表 2 所示,本文方法在 CTST-1600 数据集上可达 93.97% 的 F1 得分。

### 2.4 文本区域识别算法实验

本文使用结合 CTC 损失函数的 CRNN 文本识别算法

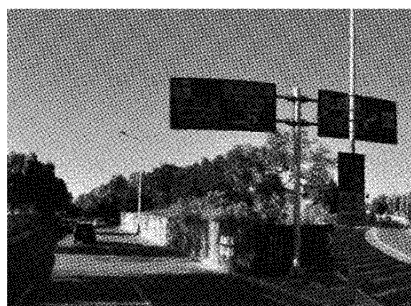


图 11 交通标识文本定位示例

表 2 文本区域定位结果

算法	P	R	F
改进 DBNet+YOLO v5s	94.95	92.98	93.97

在 MTWI-2018 数据集上训练 300 个 epoch 得到模型并保存,使用前文所述改进 DBNet 与 YOLO v5s 组成的文本行定位算法得到精细化定位的文本图像后,使用保存的文本识别算法模型对所得到的图像精细化文本图像进行预测,由于 CTST-1600 数据集为文本行级标注,因此手动统计识别所得结果。如表 3 所示,本文算法在 CSTST-1600 数据集在召回率为 82.6 的同时识别准确率可达 77.2。

表 3 文本识别结果

算法	P	R	F
CRNN	72.5	82.6	77.2

### 2.5 量 化

为更快的进行算法实时计算,将本文所使用 YOLO v5s 与改进 DBNet 组成的交通标志文本区域定位算法与文本识别算法进行量化推理,在输入图像大小为  $640 \times 640$

时,推理速度由未使用加速前的 90 ms 提升至 40 ms。加速后的整体算法可达到更快的推理速度。

表 4 量化后性能对比

输入图像大小	Float32/ms	Int8/ms
$640 \times 640$	90	40

### 3 结 论

本文基于结合 YOLO 检测算法和改进的 DBNet 算法与 CRNN 算法设计了一种两阶段的交通标志文本信息识别系统,实现了交通标志文本信息的精细化检测与提取,识别综合得分在 77.2% 时推理速度可达 25 帧,算法实现了以下两点创新:

设计精细化交通标志文本检测模块,结合了 YOLO 检测算法与改进的 DBNet 算法在检测交通标志文本区域具有能加强特征细粒度具有更好的鲁棒性;

所提出交通标志文本信息识别系统,为智能驾驶场景信息感知项目提供了关键的技术手段。

文中提出的交通标志文本信息识别技术可有效的获取高密度多聚集的自动驾驶场景下的环境信息,提高自动驾驶系统的稳定性。

### 参考文献

- [1] HE P, HUANG W L, HE T, et al. Single shot text detector with regional attention [C]. IEEE International Conference on Computer Vision, IEEE Computer Society, 2017: 3066-3074.
- [2] 王晓东,魏成保,冯海荣,等.基于 SVM 的变电站保护室数显仪表数字识别方法[J].电子测量技术,2019,42(2):92-95.
- [3] 刘素行,吴媛,张军军.基于 YOLO v3 的交通场景目标检测方法[J].国外电子测量技术,2021,40(2):116-120,DOI:10.19652/j.cnki.femt.2002271.
- [4] 吕禾丰,陆华才.基于 YOLOv5 算法的交通标志识别技术研究[J].电子测量与仪器学报,2021,35(10):137-144,DOI:10.13382/j.jemi.B2104449.
- [5] 伍锡如,邱涛涛,王耀南.改进 Mask R-CNN 的交通场景多目标快速检测与分割[J].仪器仪表学报,2021,42(7):242-249,DOI:10.19650/j.cnki.cjsi.J2107749.
- [6] TABERNIK D, SKOAJ D. Deep learning for large-scale traffic-sign Detection and recognition[J]. IEEE Transactions on Intelligent Transportation Systems, 2019(99): 1-14.
- [7] STALLKAMP J, SCHLIPSING M, SALMEN J, et al. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition [J]. Neural Netw, 2012, 32:323-332.

- [8] BRUNNER T, DIEHL F, LE M T, et al. Leveraging semantic embeddings for safety-critical applications [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019.
- [9] TORRES L T, PAIXO T M, BERRIEL R F, et al. Effortless deep training for traffic sign detection using templates and arbitrary natural images [C]. 2019 international joint conference on neural networks (IJCNN), IEEE, 2019; 1-7.
- [10] HAQUE W A, AREFIN S, SHIHAVUDDIN A, et al. DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements[J]. Expert Systems with Applications, 2021, DOI: 10.1016/j.eswa.2020.114481.
- [11] 崔家瑞, 纪建才, 胡广大, 等. 基于图像的换流站用避雷器仪表识别[J]. 电工技术学报, 2015, 30(S1): 377-382.
- [12] 刘进博, 朱新新, 伍越, 等. 基于神经网络和支持向量机的多机动车车牌在线检测方法[J]. 自动化学报, 2021, 47(2): 316-326, DOI: 10.16383/j.aas.c180753.
- [13] 白睿, 徐友春, 李永乐, 等. 智能车道路场景数字字符识别技术[J]. 激光与光电子学进展, 2019, 56(19): 211-219.
- [14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), IEEE Computer Society, 2017.
- [15] MINGHUI L, ZHAOYI W, CONG Y, et al. Real-time scene text detection with differentiable binarization[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7), DOI: 10.48550/arXiv.1911.08947.
- [16] SHI B, XIANG B, CONG Y. 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 & Machine Intelligence, 2016, 39(11): 2298-2304.
- [17] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [C]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014.
- [18] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2020.
- [19] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [20] YU F, CHEN H, WANG X, et al. BDD100K: A diverse driving dataset for heterogeneous multitask learning [C]. 2020 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2020.
- [21] YAO C, BAI X, LIU W, et al. Detecting texts of arbitrary orientations in natural images [C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012: 1083-1090.
- [22] NAYEF N, FEI Y, BIZID I, et al. ICDAR 2017 robust reading challenge on multi-lingual scene text detection and script identification-RRC-MLT [C]. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2017.
- [23] HE X, R WANG, LI X, et al. HTSTL: Head-and-tail search network with scale-transfer layer for traffic sign text detection[J]. IEEE Access, 2019(99), DOI: 10.1109/ACCESS.2019.2936540.

### 作者简介

王立刚, 硕士, 主要从事图像处理与模式识别、计算机视觉方面的研究。

E-mail: ganggang233@hotmail.com

张志佳, 教授, 博士生导师, 博士, 主要从事机器视觉检测、人工智能与模式识别方面的研究。

E-mail: zzsut@126.com

贺继昌(通信作者), 硕士。

E-mail: hejichang10@163.com