

# 基于视频的轨道车辆自主定位方法研究

沈拓<sup>1</sup>, 谢远翔<sup>1</sup>, 盛峰<sup>2,3</sup>, 谢兰欣<sup>4</sup>, 张颖<sup>1</sup>, 安雪晖<sup>2</sup>, 曾小清<sup>1</sup>

(1. 同济大学上海市轨道交通结构耐久与系统安全重点实验室, 上海 201804; 2. 清华大学土木水利学院, 北京 100084; 3. 中电建路桥集团有限公司, 北京 100070; 4. 上海泽高电子工程技术股份有限公司, 上海 201900)

**摘要:** 针对轨道施工车辆自主定位需求, 提出一种基于车载前视相机拍摄百米标视频的轨道车辆自主定位方法。该方法首先对 YOLOX-s 网络进行改进并构建了百米标的目标检测模型, 完成对百米标的目标检测; 其次, 当检测到百米标后, 结合图像预处理和卷积循环神经网络(CRNN)网络构建百米标数字文本识别模型, 提取百米标的数字文本信息, 从而实现轨道施工车辆的定位。经实验验证该方法能够快速准确定位轨道施工车辆的位置信息。

**关键词:** 轨道车辆定位; 机器视觉; 目标检测; 文本识别

中图分类号: U239.5; TP391.41

文献标志码: A

## Vision-Based Absolute Position Extraction Method for Rail Vehicles

SHEN Tuo<sup>1</sup>, XIE Yuanxiang<sup>1</sup>, SHENG Feng<sup>2,3</sup>, XIE Lanxin<sup>4</sup>, ZHANG Ying<sup>1</sup>, AN Xuehui<sup>2</sup>, ZENG Xiaqing<sup>1</sup>

(1. Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, Shanghai 201804, China; 2. School of Civil Engineering, Tsinghua University, Beijing 100084, China; 3. Powerchina Roadbridge Group Co., Ltd., Beijing 100070, China; 4. Shanghai Zegao Electrical Engineering Technology Co., Ltd., Shanghai 201900, China)

**Abstract:** Aiming at the demand of autonomous positioning of rail construction vehicles, this paper proposes a method of absolute position extraction of rail vehicles based on the video of 100-metre markers captured by the on-board forward-looking camera. The method first improves the YOLOX-s network and constructs the target detection model of the 100-metre marker to complete the target detection of the 100-metre marker. Then, when the 100-metre marker is detected, it combines image preprocessing with convolutional recurrent neural network (CRNN) network to construct

the 100-metre marker digital text recognition model to extract the digital text information of the 100-metre marker, so as to achieve the absolute position location of rail construction vehicles. The method is verified to be able to quickly and accurately extract the absolute position information of rail construction vehicles.

**Keywords:** rail vehicle positioning; machine vision; target detection; text recognition

车辆定位技术作为列车运行控制系统中的一项关键技术, 为高速铁路列车运行控制系统实时提供可靠的速度和位置信息, 其定位精度将影响列车运行效率、安全性和服务质量。将车辆定位技术用于高速铁路施工车辆定位, 现有轨道车辆定位方式, 主要是以“车载+轨旁”测量方式来实现“误差累积+复位”。该定位模式强依赖于轨旁设备和信号系统, 不仅存在轨道交通线路建设难度大, 维修成本高等问题, 而且列车长期运行会出现因轨旁设备或信号系统故障, 导致传输过程中列车定位失效, 从而影响轨道交通的运营秩序。此外, 地铁的运行环境多为地下长隧道, 因此全球导航卫星系统(global navigation satellite system, GNSS)等精准定位方法不具有适用性。现行轨道车辆定位方法主要有轨道电路定位方法、查询应答器/信标定位方法、无线扩频定位方法、全球卫星导航定位和地图匹配方法, 但这些方案依赖信号系统传输数据, 易受干扰和控制, 且无法避免特殊环境如大量长隧道运行环境下卫星信号较差的局限性。随着下一代高速铁路列控系统, 中的列车测速定位系统精确性、实时性、可靠性等应用需求越来越高, “车载中心化”的列车自主定位方

收稿日期: 2023-10-24

基金项目: 国家重点研发计划(2022YFB4300501); 上海市科委课题(23DZ2204900); 校企战略合作专项高速铁路绿色智能施工关键技术研究(kh0160020230946\_LQKY2022-01-1)

第一作者: 沈拓, 工程师, 博士生, 硕士生导师, 主要研究方向为轨道交通控制与安全。E-mail: st8250@163.com

通信作者: 盛峰, 教授级高级工程师, 博士生, 主要研究方向为土木水利工程、轨道交通控制。

E-mail: s18663726777f@126.com



论文  
拓展  
介绍

式已成为新的研究发展方向。

基于视觉的前方目标定位测速测距方法因具有远量程、非接触和高精度等特点,很多学者致力于使用单幅图像来实现车辆路径的保持和跟踪,车辆导航和定位等。轨道交通的视觉定位研究主要有基于人工标志的定位<sup>[1-3]</sup>、实时定位与建图<sup>[4-5]</sup>(simultaneous localization and mapping, SLAM)和基于局部运动估计的定位<sup>[6-7]</sup>三种类型,具有远量程、非接触等优点,但是易受光照影响。随着图像处理技术的不断发展,视觉传感器等硬件成本的不断降低,其环境适应性和感知准确性都得到较大幅度提升,将为车辆定位研究提供更加精确可靠的基础性能保障。

为适应全自动运行列车技术的发展和推广,并能够满足负责调车作业和抢险救援等的轨道工程车的自主定位需求,本研究提出一种基于列车前视相机拍摄百米标视频的轨道车辆自主定位方法。百米标是设置在轨道线路旁的一类标识,表示正线每百米距离该线路起点的长度。因此,利用视频识别到轨旁百米标数字编号即可确定百米标的位置信息,从而实现对线路运行列车定位。该方法分别构建了百米标目标检测模型和百米标数字文本识别模型。首先针对轨道交通环境复杂、轨旁标识多样、百米标成像小和定位性能要求,使用单阶段类型的YOLOX目标检测算法并进行改进,实现复杂环境下的百米标检测;其次,在检测到百米标之后,为解决百米标文本倾斜模糊、光线不均等问题,提出一种结合图像预处理的改进卷积循环神经网络(convolutional recurrent neural network, CRNN)百米标文本识别算法,实现对百米标的线路位置信息提取,从而完成对轨道车辆的定位。最后通过实验验证了算法的有效性和环境适应性。

## 1 基于改进YOLOX的百米标目标检测

YOLO算法是由旷视科技于2021年提出的新一代YOLO系列目标检测算法。YOLOX网络在提升检测速度的同时兼具相当的检测精度,正被广泛应用于各个领域对的目标检测任务中。YOLOX网络主要由4个模块构成,分别是输入模块Inputs、网络主干模块Backbone、特征增强模块Neck以及预测模块Prediction。

### 1.1 YOLOX算法改进

针对前文所述的百米标检测存在的问题,本研究选取综合性能最好的YOLOX-s网络进行改进并构建百米标目标检测模型,来完成百米标检测的任务。在进一步提高识别准确率的同时,保持原模型的轻量化和检测速度。首先在Mixup和Mosaic图像增强的基础上继续增加随机改变亮度、对比度、旋转以及加入高斯噪声等方式完成数据增强,以适应轨道交通光线变化范围广的复杂环境。然后,改进原YOLOX-s网络结构以提升检测速率和精度,具体方法为:基于原YOLOX-s框架,在主干网络CSPDarknet的CSPLayer结构中引入注意力机制(efficient channel attention, ECA),增强重要特征;引入自适应空间特征融合(adaptive spatial feature fusion, ASFF),使得提取后的三个特征层自适应地学习各尺度特征,增强特征融合效果;最后对损失函数进行优化,以提高模型精度和加快收敛速度。

#### 1.1.1 CSPLayer引入ECA注意力机制

注意力机制的原理可简单地描述成通过某些网络来学习特征权重,得出各特征图重要程度,以及基于上述计算的重要程度,给各特征图赋予权重值。以便于让神经网络把注意力集中在重要的特征图上,通过增强重要特征向量且削弱不重要特征向量来优化模型<sup>[8]</sup>。本研究将ECA注意机制引入到YOLOX的主干网络的CSPLayer结构中,网络将更加关注重要特征的提取。具体步骤如下:

(1)首先输入特征图 $x$ ,维度是 $H*W*C$ ;

(2)对输入特征图进行空间特征压缩,使用全局平均池化(global average pooling, GAP)方法,得到 $1*1*C$ 的特征图;

(3)对压缩后的特征图进行通道特征学习;采用自适应内核尺寸(设卷积核大小为 $k$ )的一维卷积方式实现局部跨通道的信息交互,通过Sigmoid激活函数,此时输出的维度还是 $1*1*C$ ;

卷积和自适应函数定义如下:

$$k = \phi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (1)$$

式中: $\lfloor \cdot \rfloor_{\text{odd}}$ 表示距离 $t$ 最近的奇数; $\gamma$ 和 $b$ 是线性参数,分别取值为2和1。

(4)完成通道注意力结合,将通道注意力的特征图 $1*1*C$ 、原始输入特征图 $H*W*C$ 进行逐通道乘,最终输出具有通道注意力的特征图 $\bar{x}$ ,维度是 $H*W*C$ 。

### 1.1.2 Neck网络引入ASFF机制

特征金字塔网络(feature pyramid network, FPN)结构会导致检测结果受不同层的特征影响,如果目标在某一层中被判断为正类后,其他层会将其判断为负类。然而当特征融合时,判别为负类的无用信息可能也会被融合,从而产生大量冗余信息。因此本研究在路径聚合特征金字塔网络(path aggregation feature pyramid network, PAFPN)尾部添加自适应空间特征融合ASFF机制<sup>[9]</sup>,能够自适应地学习不同尺寸特征的融合权重,而不是基于元素和抑或级联方式的特征融合。该方法可以缓解特征金字塔中不同尺度特征的不一致性缺点,达到最优融合效果。

将PAFPN中的三个不同尺度的特征图通过ASFF机制,学习融合权重,理解不同特征尺度在预测特征图中的贡献程度,最后融合成对应尺度的三个特征图。以ASFF3为例,经过PAFPN得到的三个不同尺度的特征层Level 1~3,通过分别与权重参数 $\alpha$ 、 $\beta$ 、 $\gamma$ 相乘,然后相加,就得到了ASFF3,上述过程可描述如下:

$$y_{ij}^{(l)} = \alpha_{ij}^{(l)} \cdot x_{ij}^{(1 \rightarrow l)} + \beta_{ij}^{(l)} \cdot x_{ij}^{(2 \rightarrow l)} + \gamma_{ij}^{(l)} \cdot x_{ij}^{(3 \rightarrow l)} \quad (2)$$

式中: $y_{ij}^{(l)}$ 为通过ASFF特征融合得到的新特征图; $x_{ij}^{(1 \rightarrow l)}$ 、 $x_{ij}^{(2 \rightarrow l)}$ 、 $x_{ij}^{(3 \rightarrow l)}$ 分别为第1、2、3层到第 $l$ 层特征层的特征向量; $\alpha_{ij}^{(l)}$ 、 $\beta_{ij}^{(l)}$ 、 $\gamma_{ij}^{(l)}$ 分别为三个不同特征层的权重参数。

由于求和公式要求各层特征维度保持一致,因此还需要通过下采样且调整通道数的方式对各层特征进行降维。对于权重参数 $\alpha$ 、 $\beta$ 、 $\gamma$ ,经过拼接之后通过softmax函数使其范围都在 $[0, 1]$ 内,并且和为1,上述过程可由式(3)来描述( $\lambda^{(l)}$ 表示与网络的第 $l$ 层相关的一个超参数,用于调整softmax函数中指数的尺度):

$$\alpha_{ij}^{(l)} = \frac{e^{\lambda^{(l)} \alpha_{ij}}}{e^{\lambda^{(l)} \alpha_{ij}} + e^{\lambda^{(l)} \beta_{ij}} + e^{\lambda^{(l)} \gamma_{ij}}} \quad (3)$$

### 1.1.3 损失函数优化

YOLOX-s网络会生成大量检测帧,大多数的检测帧都是负样本。损失函数中负样本数量过多会导致正样本权重较小,从而对网络最终训练效果造成影响。对此,本研究对于损失函数的改进方法如下:

#### (1)对象置信度损失函数改进

对象损失函数利用置信度来判断边界框内的目标是否为正负样本,若超过置信度阈值,则判定为正样本,若小于置信度阈值,则判定为负样本(即背

景)。由于百米标检测场景下,被检测目标较小,会存在大量负样本,若使用二元交叉熵损失函数 $BCE_{loss}$ 对目标置信度进行回归,很难克服YOLOX-s因负样本过多而导致权重不平衡的问题。因此,本研究试图对YOLOX-s算法中的置信度损失函数进行改进,把 $BCE_{loss}$ 函数改进为 $Focal_{loss}$ 函数,该损失函数能够降低负样本所占权重,以确保网络训练过程更加专注于正样本。 $Focal_{loss}$ 损失函数公式如下:

$$F_L = -\alpha_p (1 - p_t)^\gamma \log(p_t) \quad (4)$$

式中: $\alpha_p$ 为平衡参数,用于解决正负样本数量不均的问题,将其设置为0.35; $p_t$ 为预测目标的置信度; $\gamma$ 是调制因子,将其设置为2。

关于对象类型的确定,仍沿用 $BCE_{loss}$ 计算方法。计算公式如下:

$$B_L = -p_t \log(p_t) - (1 - p_t) \log(1 - p_t) \quad (5)$$

#### (2)边界框回归损失函数改进

YOLOX-s算法采用 $IoU_{loss}$ 函数对检测目标边界框进行回归。 $IoU$ 是计算框(prediction box, P)和真实框(ground truth box, G)的交并比,用于衡量P和G之间的距离, $IoU_{loss}$ 的计算公式为

$$I_L = 1 - \frac{G \cap P}{G \cup P} \quad (6)$$

如图1a所示,当所述预测框P和所述真实框G没有重叠时, $IoU=0$ ,这时无法反映出两者的真实距离,此时若将 $IoU$ 作为回归损失对模型进行回归,由于 $IoU$ 为0,梯度无法回传更新。此外,如图1b所示, $IoU$ 值相等,但这2种情况下两者的真实距离不等。而 $IoU$ 不能反映两者的回归效果。

而除 $IoU_{loss}$ 函数外,常用的还有 $GIoU_{loss}$ 、 $DIoU_{loss}$ 、 $CIoU_{loss}$ 等损失函数。

$GIoU$ ,首先计算P与G的最小封闭区域面积 $A^c$ ,然后再采用 $IoU$ 减去封闭区域中不含两个框的区域在封闭区域中的占比, $GIoU_{loss}$ 损失函数计算公式为

$$G_L = 1 - IoU(G, P) + \frac{A^c - (G \cap P)}{A^c} \quad (7)$$

$GIoU$ 清楚反映出两框的重合程度。但 $GIoU$ 也存在问题:当P完全包含G时G在P内的分布位置不同或者当P完全被包含G时,P在G内的分布位置不同, $GIoU_{loss}$ 仍然相同,此时也无法反映出各自回归效果。

$DIoU$ ,通过约束这个最小封闭区域面积和预测框、真实框中心点的位置,让网络学习移动P向G靠近。但 $DIoU_{loss}$ 还存在一种问题:没有考虑P与G

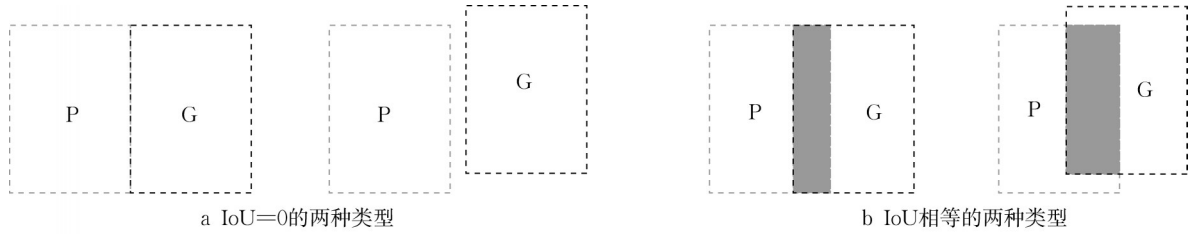


图1 两种IoU交并比计算情况

Fig. 1 Calculation of two IoU cross-combination ratios

的长宽比。DIoU<sub>loss</sub> 损失函数公式为:

$$D_L = 1 - \text{IoU}(G, P) + \frac{\rho(b, b^{GT})}{c^2} \quad (8)$$

式中: $\rho$ 表示预测框与目标框中心点之间的欧几里得距离; $c$ 为能同时包含预测框和真实框的最小矩形对角线长度; $(b, b^{GT})$ 代表了预测框和真实框的中心点。

因此,本研究在DIoU基础上把长宽比考虑进来,得到CIoU,即同时考虑了覆盖面积、中心点距离和长宽比三种情形。CIoU计算公式为

$$C = \text{IoU}(G, P) - \frac{\rho(b, b^{GT})}{c^2} + \alpha v \quad (9)$$

式中: $v$ 为P与G长宽比( $h$ 为框长, $w$ 为框宽)的距离,计算公式为

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{GT}}{h^{GT}} - \arctan \frac{w}{h} \right) \quad (10)$$

$\alpha$ 是一个权重系数,计算公式为

$$\alpha = \frac{v}{(1 - \text{IoU}(G, P)) + v} \quad (11)$$

计算DCIoU<sub>loss</sub> 损失函数的公式为

$$D_{Cl} = 1 - C \quad (12)$$

理论上说,CIoU效果最好,因此,本研究将原YOLOX-s算法边界框回归损失函数使用的IoU<sub>loss</sub>函数改为CIoU<sub>loss</sub>函数。

## 1.2 百米标目标检测

### 1.2.1 数据集制作

轨道交通线路上有设置指示线路状态和位置的线路标识,包括公里标、百米标、曲线标、坡度标等;也有设置指示司机完成相关作业的信号标识,包括警冲标、站界标、驾驶员鸣笛标、停车位置标等,这些线路标志的外形十分相似。因此,为防止百米标误检、漏检,需要把其他标识考虑在目标检测范围内。鉴于已有公开数据集并没有可以有效囊括轨道工况的数据,无法满足轨道环境检测应用需求。为适应本研究的百米标检测需要,采集某线路上列车前视相机拍摄的各类天气全天候运行视频,截取视频中

高架、隧道、阴天、夜间等场景的视频帧制成图像数据集。

### (1)图像数据增强

由于人工采集数据工作量大,并且数量仍然不充分。对此首先增加了大量基于实测轨道为背景的正、负样本数据;并且通过图像增强方法对数据集进行扩充,包括图像翻转、旋转、裁剪、缩放、颜色变换等,使用扩充后的数据集训练模型能够增强模型的鲁棒性和泛化能力。

### (2)图像数据标注

本研究使用labelImg工具标注出被探测目标所属类别和目标区域,标签设为5类常见标识:hundred、speedlimit、stationname、curve、warning,分别代表百米标、限速标、站牌、曲线标、警冲标。最后按照8:1:1的比例将数据集划分为训练集、测试集和验证集。

### 1.2.2 模型训练

本研究将图像尺寸统一为640\*640,Epoch迭代次数设置为1000,batch\_size设置为118,初始学习率为0.01,选择余弦退火衰减(Cosine annealing scheduler)动态学习策略,最低学习率为0.0001,IoU阈值设置为0.5。训练过程中Loss损失如图2和图3所示。

## 2 基于改进CRNN的百米表文本识别

检测到图像中含有百米标后,继续对百米标进行文本识别,以提取百米标的线路位置信息。文本识别的算法主要分为传统光学字符识别(optical character recognition, OCR)方法和基于深度学习OCR算法两种。传统OCR处理流程包括图像预处理、基于连通域的文本行检测、字符分割并提取特征、基于分类器的单字符识别和后处理输出5个步骤。在背景简单的前提下能够精准识别文本序列,

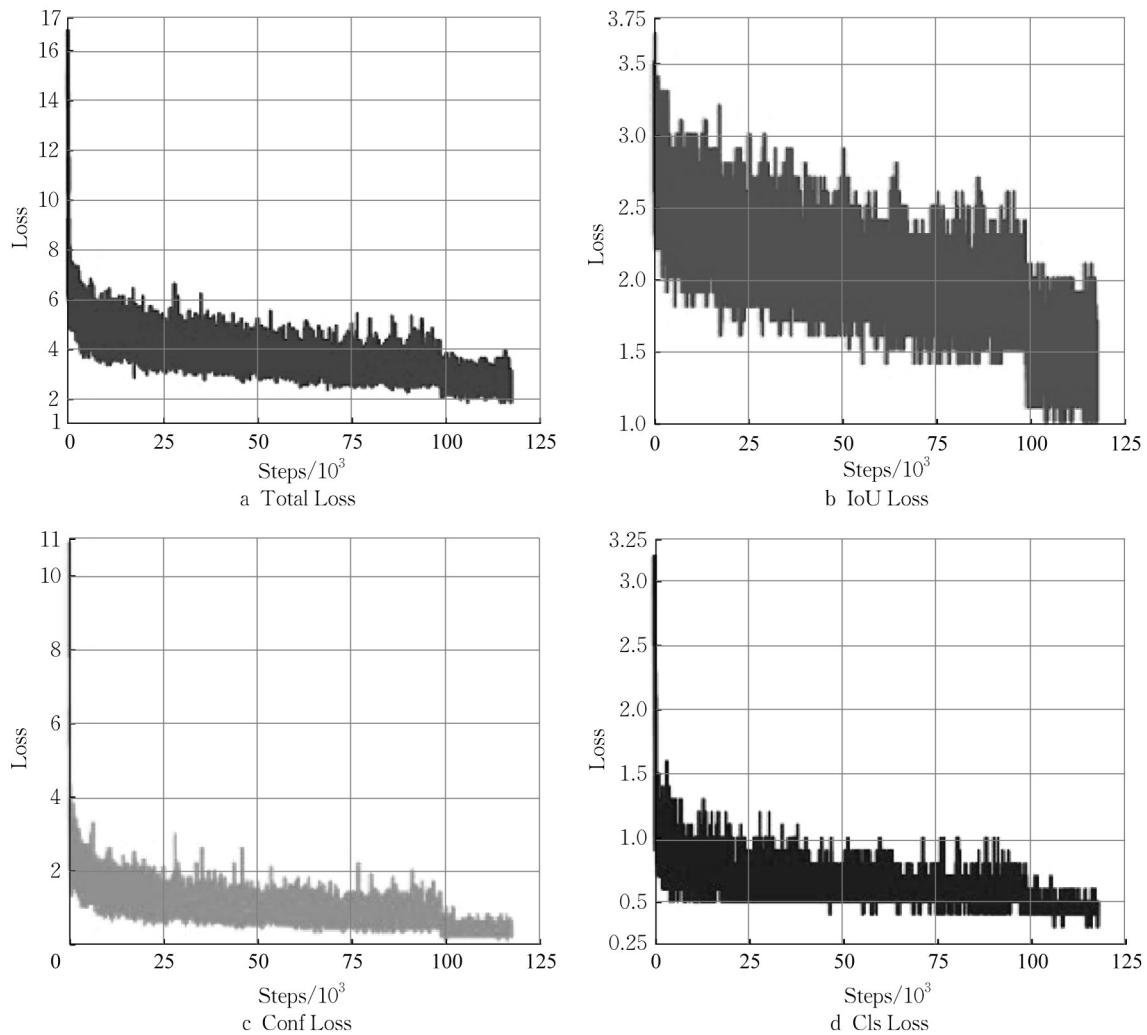


图2 YOLOX-s训练过程Loss曲线

Fig. 2 Loss curves of YOLOX-s training process

但在复杂环境下文本识别效果较差。

本研究需要识别的百米标数字序列存在运动模糊、噪声点多、目标小和文本倾斜等问题,因此本研究选择实际应用中效果更好的基于深度学习的OCR算法。其中,CRNN+CTC(connectionist temporal classification,连接主义时序分类)识别算法具有可以识别任意长度的文本,识别速度快,准确率高的优点,广泛应用于各个领域的识别任务中。因此本研究针对轨道交通特殊环境,使用图像预处理+改进CRNN+CTC方法,来完成百米标的文本识别任务,提取百米标的位置信息。首先针对百米标进行图像预处理,便于提高后续CRNN网络识别速率和准确率;然后使用CRNN+CTC文本识别方法对百米标文本进行识别,提取出百米标的线路位置信息。

## 2.1 图像预处理

由于图像检测到的百米标文本序列存在运动模糊、目标小和文本倾斜等问题。因此,在使用CRNN网络进行文本识别之前,先对带有噪声的百米标图像进行处理。文本预处理流程如图4所示,主要包括:图像裁剪、灰度化、二值化、倾斜检测与校正、图像平滑与图像规范化。

## 2.2 CRNN算法与改进

本研究选择的文本识别CRNN网络依次使用了CNN结构和RNN结构。图5是CRNN网络结构,前两部分卷积层(CNN)和循环层(RNN)都采用了基于结构相似性的特征提取方法,后者转录层(CTC)采用的是基于概率统计特性的特征抽取算法。

原CRNN算法的特征提取的主干网络为VGG16网络,会存在梯度消失以及多层卷积之后信

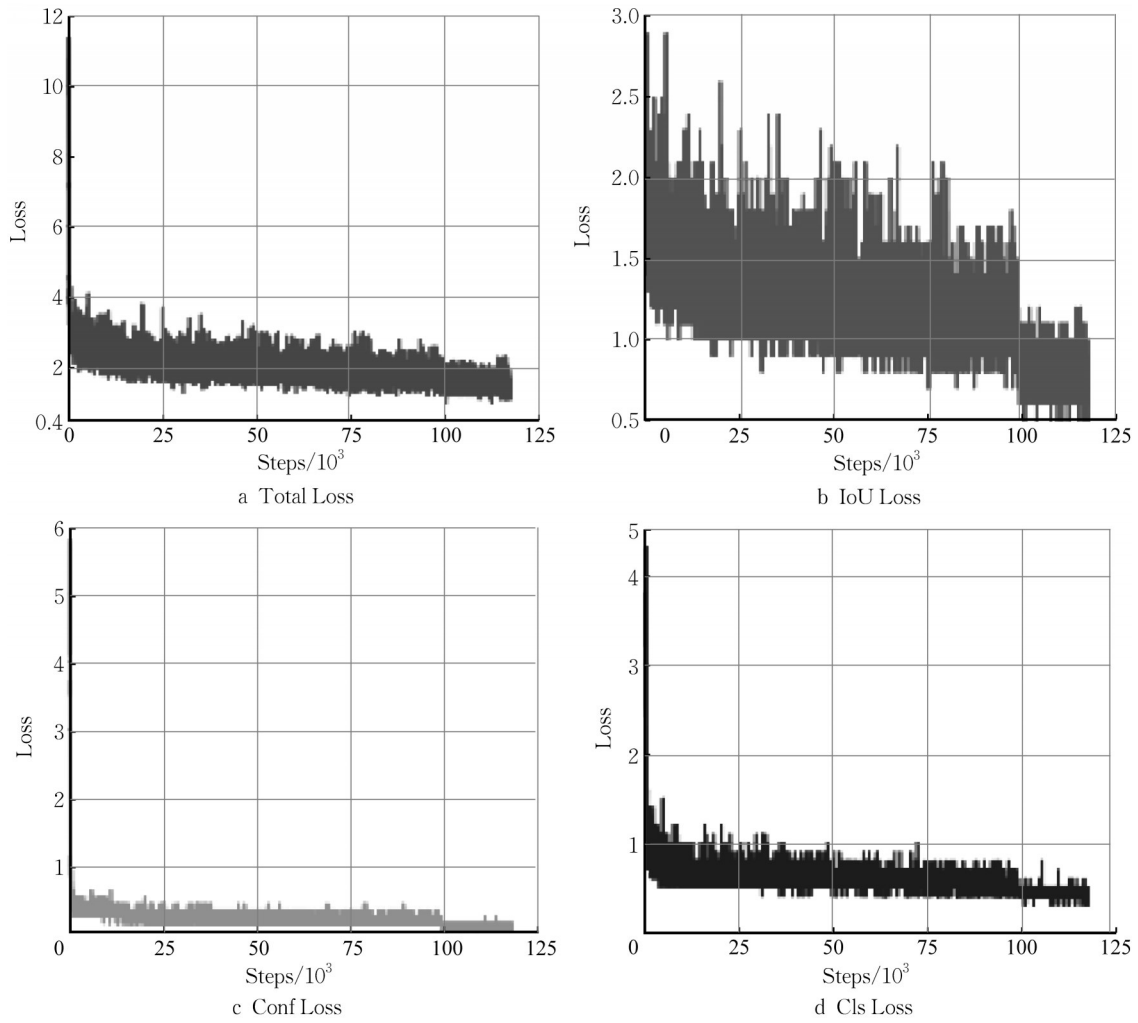


图 3 改进 YOLOX-s 训练过程 Loss 曲线

Fig. 3 Loss curves of improvement of YOLOX-s training process

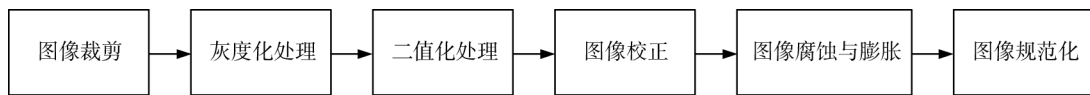


图 4 文本图像预处理流程

Fig. 4 Preprocessing process of text image

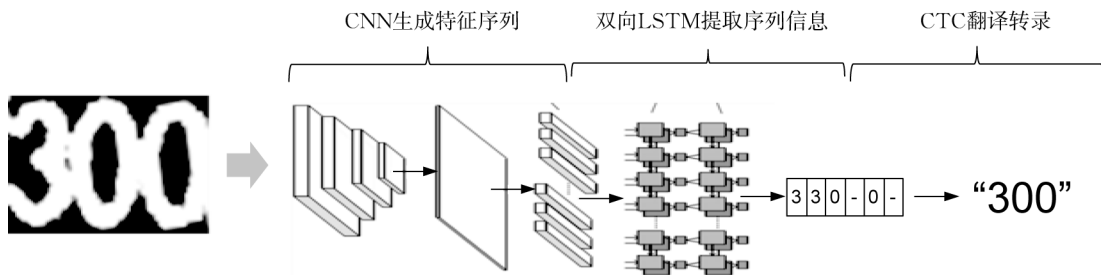


图 5 CRNN 网络结构

Fig. 5 Structure of CRNN network

息丢失严重的问题,因此,本研究在 VGG16 网络的第 3 层和第 4 层加入如图 6b 所示的残差结构,选择

残差卷积网络作为特征提取的主干网络。该网络以跨层连接(shortcut connection)的方式连接各层网

络,能够提高网络计算效率和性能。

残差学习模块有两种形式:两层结构称为 BasicBlock (图 6a),一般适用于 Resnet18 和 Resnet34。weight layer 卷积层为用于学习输入特征

的权重,relu为激活函数;三层结构称为 Bottleneck (图 6b),是对 BasicBlock 的优化,旨在缩小特征图的大小,由此减少参数量。

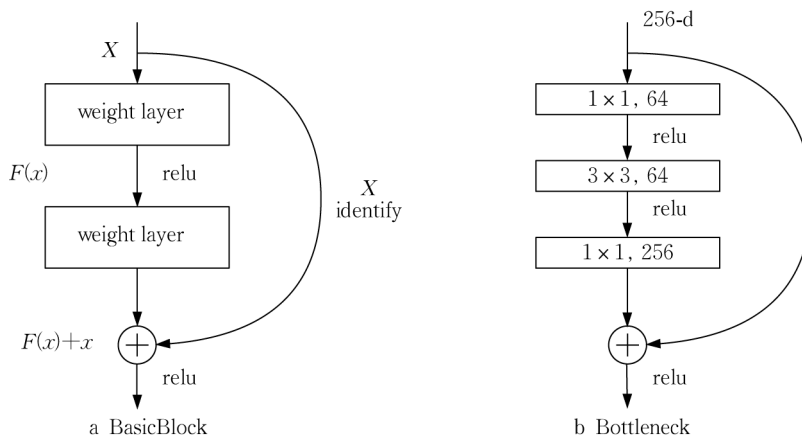


图 6 残差单元结构

Fig. 6 Structure of residual cell

此种残差单元结构将  $F(x)$  学习的内容由  $x$  变为  $x$  的残差,由于残差的值是接近 0 的,故模型学习比 VGG 每层堆叠更加高效,并且能解决训练集随着网络深度的增加误差反而增大的退化问题。

ResNet 中任意一层的计算公式为

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, w_i) \quad (13)$$

式中:  $x_L$  为  $L$  层的特征信息;  $x_l$  为比  $L$  层浅一层的特征信息;  $F(x_i, w_i)$  为残差函数。

### 2.3 百米标文本识别

#### 2.3.1 文本图像预处理

百米标文本图像的预处理流程及对应的效果如图 7 所示。

可以看到,此时的百米标不再倾斜,文本更加清晰,极大地降低了后续文本识别算法的计算量。在依次进行了图像裁剪、灰度化、二值化、倾斜检测与校正、图像平滑与图像规范化之后,生成的标准百米标图像作为后续 CRNN 网络的输入。

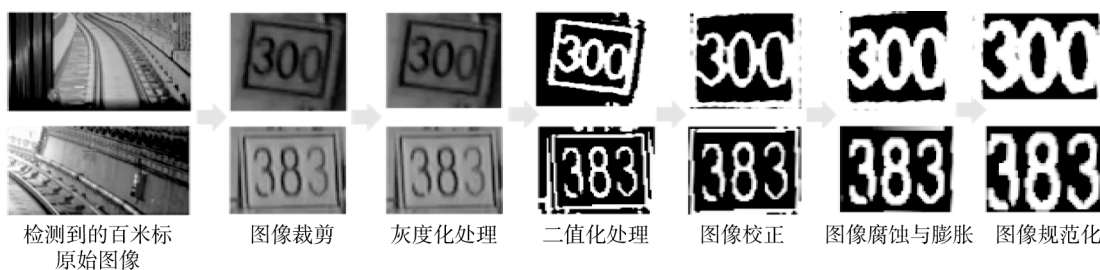


图 7 百米标文本图像预处理流程

Fig. 7 Preprocessing process of 100-metre sign text image

#### 2.3.2 数据集制作

本研究参照 ICDAR2015 文本识别数据集格式,制作百米标文本识别数据集。标注文件如图 8 所示。其标注格式为

$$\{[x_1, y_1], [x_2, y_2], [x_3, y_3], [x_4, y_4], 'XXX'\}$$

其中,  $[x_1, y_1], [x_2, y_2], [x_3, y_3], [x_4, y_4]$  表示的是按顺时针方向标注文本位置的 4 个顶点坐标,

“XXX”代表文本内容。

#### 2.3.3 模型训练

由于数字文本数量较少,可适当降低模型训练迭代次数,节省训练时间和消耗。本研究将模型参数 batch\_size 设置为 32,学习率设置为 0.001,权重衰减系数为 0.001,epoch 设置为 300。CRNN 网络训练的 Loss 损失过程如图 9 所示。

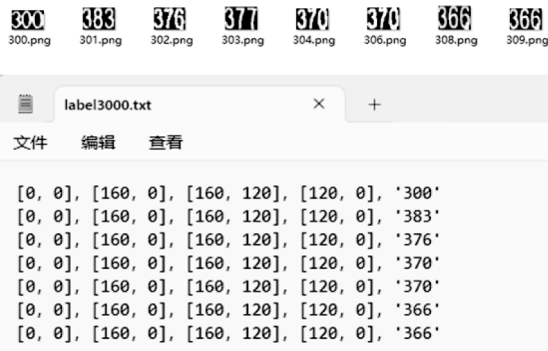


图 8 图像文本标注示例

Fig. 8 Annotation example of image text

### 3 实验结果与分析

#### 3.1 实验环境

在 PC 机中基于 Pytorch 完成百米标线路位置信

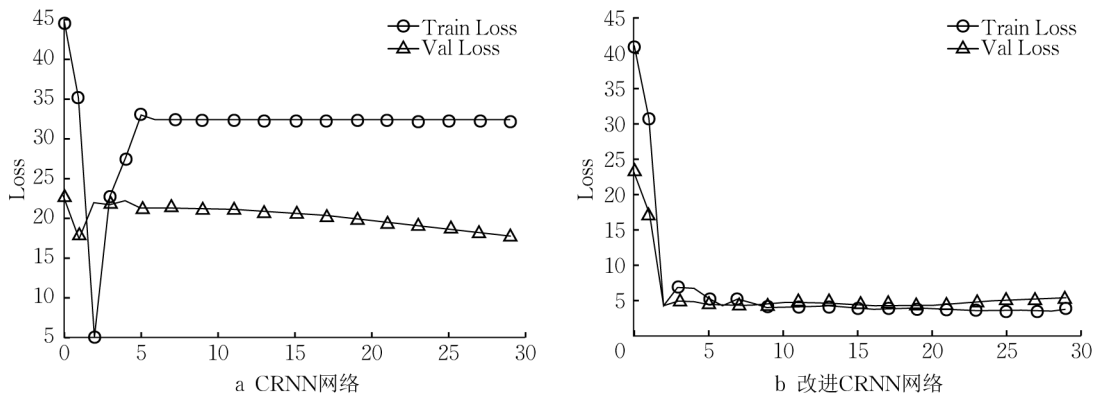


图 9 文本识别网络训练过程的 Loss 曲线

Fig. 9 Loss curves of text recognition network training process

表 1 实验环境配置

Tab.1 Configuration of experimental environment

	硬件配置		软件配置
操作系统	64 位 WIN11	CUDA	CUDA 11. 0; CUDNN 8. 0. 4
处理器	2. 4 GHz Intel Core i5-11350	框架	Pytorch 1. 7. 1
显卡	NVIDIA GeForce RTX 3060	编程环境	Python 3. 7
内存	16 GB	软件	VS Code

表 2 改进 YOLOX-s 算法检测效果对比

Tab. 2 Comparison of detection performance of improved YOLOX-s algorithm

算法	AP / %						mAP / %	FPS / (帧 · s <sup>-1</sup> )
	百米标	曲线标	坡度标	站名标	警冲标	限速标		
YOLOX-s	98. 21	80. 00	98. 95	99. 98	97. 75	93. 61	94. 24	51
改进 YOLOX-s	99. 12	79. 73	99. 01	99. 98	97. 99	96. 75	96. 60	52

图 10a 为改进 YOLOX 算法在高架上检测百米标效果,图 10b 为改进 YOLOX 算法在隧道里检测百米标效果。可以看出,当百米标距车辆较近时能被有效检测出,当小目标百米标距车辆较远且在高架

息提取所需的模型训练和测试,该框架可通过 GPU 加速优化,结构清晰,并且可移植性强,可以在嵌入式设备上使用。PC 机具体配置见表 1。

#### 3.2 百米标目标检测实验结果与分析

本研究在相同的实验环境下以同样的方式进行训练,对改进 YOLOX-s 算法和原始 YOLOX-s 算法的百米标检测效果进行对比。结果如表 2 所示,改进 YOLOX-s 算法的 mAP (mean average precision, 平均精度均值) 值为 96. 60 % ,FPS (frames per second) 为 52 帧 · s<sup>-1</sup>;原始 YOLOX-s 算法的 mAP 值为 94. 24 % ,FPS 为 51 帧 · s<sup>-1</sup>;平均精确率提升了约 2 % ,百米标精确率提升 0. 91 % ,并且检测速度也有所提高。因此,改进 YOLOX-s 算法的百米标检测的准确率能够满足系统功能要求。

环境里仍能被有效检测出。

本研究设置检测百米标是否有效的 AP 判断阈值为 80 % ,当被检测百米标 AP 值大于等于 80 % ,则判断该目标为有效目标,送入文本识别模型中。





图10 改进YOLOX-s算法百米标检测效果

Fig. 10 Improvement of YOLOX-s algorithm for detecting 100-metre markers

### 3.3 百米表文本识别实验结果与分析

本研究在相同的实验环境下以相同的训练方式进行训练,对CRNN算法和改进CRNN算法的百米标文本识别效果进行对比。结果如表3所示,在相同环境下,改进CRNN模型的准确率(PR)和召回率

(RR)比CRNN模型分别提高了3.52%和3.48%,并且检测速度也有所提高。可以看出,改进后的网络模型可以很好地满足系统在实际应用中的功能需求。

表3 改进CRNN文本识别算法检测效果对比

Tab. 3 Comparison of detection performance of improved CRNN text recognition algorithm

算法	PR/%	RR/%	F1值	测试时间/s
CRNN(VGG)	93.24	93.20	0.932	12.5
改进CRNN(ResNet)	96.76	96.68	0.966	12.3

CRNN文本识别网络的输出结果由6种元素组成,其输出格式为:  $([x_1, y_1], [x_2, y_2], [x_3, y_3], [x_4, y_4], 'XXX', \text{score})$ 。其中,前4个元素为文本框的顶点坐标信息,'XXX'为分类结果,score为文本框置信度分数。

图11a为改进CRNN算法的百米标文本识别结果,图11b为目标检测后的有效百米标通过图像预处理和改进CRNN算法的文本识别结果的可视化展示。以图11b为例,该有效百米标的文本检测结果为321,则提取出百米标的线路位置信息为M321。



a 改进CRNN网络文本识别结果

b 文本识别可视化

图11 改进CRNN文本识别结果

Fig. 11 Text recognition results of improved CRNN

## 4 结语

本研究提出了一种适用于轨道交通复杂环境的基于百米标视频的轨道施工车辆自主定位方法。介

绍以下流程:首先分析轨道交通复杂的轨旁环境特点和百米标成像特点;然后,针对轨道复杂环境、小目标检测和定位性能要求,选择YOLOX-s网络做改进并构建百米标目标检测算法,完成百米标的目标

检测;若检测出图像含有百米标,则针对百米标文本倾斜模糊、光线不均等特点,结合图像预处理方法和改进CRNN算法构建百米标数字文本识别模型,实现对百米标的文本识别,输出百米标的线路位置信息,从而实现对轨道车辆的定位。实验结果表明该方法能够快速准确定位轨道施工车辆的位置信息。

#### 作者贡献声明:

沈拓:建模,初稿撰写。

谢远翔:数据分析,修订论文。

盛峰:方法思路,需求分析。

谢兰欣:数据搜集。

张颖:算法测试。

安雪晖:框架建议。

曾小清:技术应用分析。

#### 参考文献:

- [1] 曹海燕,张大维.基于机器学习算法的交通标志图像智能识别[J].微型电脑应用,2021,37(1):19.  
CAO Haiyan, ZHANG Dawei. Intelligent recognition of traffic sign images based on machine learning algorithm [J]. Microcomputer Applications, 2021,37(1):19.
- [2] WELZEL A, REISDORF P, WANIELIK G. Improving urban vehicle localization with traffic sign recognition [C]//2015 IEEE 18th International Conference on Intelligent Transportation Systems. [S.l.]:IEEE, 2015: 2728-2732.
- [3] 张善彬.基于标准路牌的单目视觉车辆自定位[D].青岛:山东大学,2020.  
ZHANG Shanbin. Monocular vision vehicle self-localisation based on standard road signs [D]. Qiaodao: Shandong University, 2020.
- [4] 郭保青.基于图像特征的列车自主定位方法研究[D].北京:北京交通大学,2008.  
GUO Baoqing. Research on autonomous train positioning method based on image features [D]. Beijing: Beijing Jiaotong University, 2008.
- [5] 曹蜜,胡凌燕,熊彭文,等.基于PROSAC算法与ORB-SLAM2的RGB-D相机室内视觉定位研究[J].传感技术学报,2019,32(11):1706.  
CAO Mi, HU Lingyan, XIONG Pengwen, *et al.* Research on indoor visual localisation of RGB-D camera based on PROSAC algorithm and ORB-SLAM2 [J]. Journal of Sensing Technology, 2019,32(11):1706.
- [6] 谢家浩,刘延飞,姚二亮.基于视觉SLAM的拾取目标定位方法研究[J].电子设计工程,2019,27(23):189.  
XIE Jiahao, LIU Yanfei, YAO Erliang. Research on pickup target localisation method based on visual SLAM [J]. Electronic Design Engineering, 2019,27(23):189.
- [7] POLLEFEYS M, GOOL L V, VERGAUWEN M, *et al.* Visual modeling with a hand-held camera [J]. International Journal of Computer Vision, 2004,59(3):207.
- [8] HAN G, LI T, LI Q, *et al.* Improved Algorithm for insulator and its defect detection based on YOLOX [J]. Sensors, 2022, 22(16):6186.
- [9] HU J, QIAO P, LV H, *et al.* High speed railway fastener defect detection by using improved YoLoX-Nano model [J]. Sensors, 2022,22(21):8399.