

DOI:10.19651/j.cnki.emt.2108586

面向精确定位的钢包运输车车号罐号识别算法*

张继凯 梁勇 周亚辉 柴轶凡
(内蒙古科技大学 包头 014010)

摘要: 为了提高基于计算机视觉的钢包运输车车号罐号检测定位的准确性,降低在污损情况下的检测误差,减少罐号面积较小导致的漏检问题以及提升检测速度,提出一种基于改进 YOLOv5 网络的车号罐号检测识别方法。通过在特征提取网络中加入注意力机制,增强模型的特征提取能力;通过将骨干网络升级为轻量级的 GhostBottleNeck 加快了模型的推理速度;通过对目标字符进行仿射变换,将扭曲变形字符转换为接近正面视角,进而利用改进的 ResNet 网络进行单字符识别。结果表明,改进后的网络在钢包车号定位的精度达到了 90.3%,召回率为 87.3%,最终号码识别准确率为 97.7%,说明该方法可有效实现钢包运输车车号罐号的精确定位与识别,为智能化管理提供可靠的数据支持。

关键词: 号码识别;精确定位;注意力机制;YOLOv5

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.4050

An algorithm to identify the vehicle number and tank number of ladle truck for accurate location

Zhang Jikai Liang Yong Zhou Yahui Chai Yifan
(Inner Mongolia University of Science & Technology, Baotou 014010, China)

Abstract: In order to improve the accuracy of the detection and positioning of the number of ladle carrier tank based on computer vision, reduce the detection error in the case of contamination, reduce the missing detection problem caused by the small area of the number of tank and improve the detection speed, a detection and recognition method of the number of ladle carrier tank based on improved YOLOv5 network was proposed. The feature extraction capability of the model was enhanced by adding attention mechanism into the feature extraction network. By upgrading the backbone network to lightweight GhostBottleNeck, the reasoning speed of the model is accelerated. By performing Affine Transformation on the target character, the distorted character is converted into a near-positive perspective, and then the improved ResNet network is used for single-character recognition. The results show that the accuracy of the improved network is 90.3%, the recall rate is 87.3%, and the final number identification accuracy is 97.7%, indicating that the method can effectively achieve the accurate location and identification of the number of the ladle carrier tank, and provide reliable data support for intelligent management.

Keywords: car number identification; accurate positioning; attention mechanism; YOLOv5

0 引言

人工智能的发展推动着行业的变革,智能制造技术逐渐应用到传统的生产环节中^[1-3]。在钢包智能定位的解决方案中,基于计算机视觉的目标检测技术因无需为车辆加装额外易损装置(如 RFID^[4]技术或车载传感器^[5]),只需通过喷涂或焊接简单的数字编码就能实现目标信息的无接触

自动记录,成为目前钢包物流管控的首选方案。但工厂环境存在背景杂乱、光照条件差、车号罐号喷涂不规范以及因相机无合适安装角度造成的号码偏斜以及过小等问题,使得从工厂监控视频中准确检索钢包罐号车号具有一定的挑战性。

车号罐号的检测识别主要分为号码检测和号码识别两个步骤。其中号码检测属于场景文本检测(scene text

收稿日期:2021-12-13

* 基金项目:国家自然科学基金(51904161)、内蒙古自治区自然科学基金(2019BS06005)、内蒙古自治区高等学校科学研究项目(NJZY20095)、内蒙古自治区科技计划项目(2019GG138)资助

detection, STD)的范畴,目前主要的实现方式为CTPN^[6]、SegLink^[7]和EAST^[8]等方法。其中文献[9]提出一种改进的CTPN算法,通过融合多尺度的特征图和困难样本挖掘的方法,提升了车辆文本定位效果;对于长文本的检测,游洋彪等^[10]提出了基于短边回归的文本检测网络,采用先分离再组合的方式,实验表明该方法对于长文本的检测效果较好;为了解决背景干扰的问题,谢斌红等^[11]提出的主动中心轮廓场景文本检测网络,提高了多尺度文本检测的准确度;文献[12]提出的挖掘文本框位置特性的Anchor-free的自然场景文本检测方法,将多层特征进行融合后预测文本框的几何特征和位置特征,最后进行过滤得到最终检测的文本框,提升了文本检测的准确率。

号码识别属于场景文本识别的范畴,目前主要的实现方式为光学字符识别(optical character recognition, OCR)。朱斌等^[13]针对定位精度低,提出一种改进卷积网络的训练方法。文献[14]提出一种半监督的文本识别方法,将强化学习、编辑奖励和嵌入奖励方法相结合评估生成的字符串的质量;对于不规则方向文本的识别,Lin等^[15]提出尺度自适应方向注意网络用于任意方向的场景文字识别,网络由Transformer和序列识别网络组成,在多个数据集上提升了文字识别的准确度;Wang等^[16]提出了一种基于卷积长短期记忆(long-short term memory, LSTM)的场景文本识别器,其在识别大边缘曲线文本图像的性能上优于现有的方法。

为了减少复杂背景对车号和罐号区域的定位的干扰,本文首先尝试了单发文本检测器(single shot text detector, SSTD)^[17]和TextBoxes^[18]方法,实施后发现其容易受到非号码文本信息(背景信息)的干扰,且对于小目标检测效果不好。因此改为先采用目标检测的方式进行车号

和罐号区域的定位,得到剪裁的号码区域再单独对号码区域进行识别,以此缓解复杂背景下号码检测效果不佳的情况。由于车号和罐号在原始图像中的占比较小,面积占比约为0.5%,属于小目标范畴。文献[19]提出一种改进的YOLOv3的小目标检测算法,对于小目标的检测精度有了较大提升;王纪武等^[20]提出基于Faster-RCNN的多尺度融合与在线难例挖掘的小目标检测算法;王建军等^[21]基于YOLOv3检测算法,重构了骨干网络,增强了特征提取能力,通过加入RFB结构增大特征图的感受野,即增强了特征图的语义信息,实现了对于遥感小目标图像的较好检测。文献[22]提出了hgSSD模型,在SSD模型进行改进,实现了复杂场景下的行人检测;针对红外小目标检测的问题,孙泽军等^[23]首先根据红外目标的成像特点,得到红外目标的显著图消除部分背景干扰,然后使用图像分割提取感兴趣区域,在感兴趣区域中计算显著度,最终完成对于小目标检测;针对红外弱小目标的检测,史晓刚等^[24]首先对红外图像进行分割,再进行图像融合,最后使用多帧积累检测和形态学运算完成弱小目标检测。

针对以上所提出的小目标、复杂背景和环境干扰下检测效果不佳等问题,本文在YOLOv5网络的基础上,通过在特征提取网络中加入注意力模块来提升网络的特征表示能力,将部分卷积进行了轻量化处理以加快模型推理速度,然后使用改进的ResNet网络进行字符识别,最终使字符识别精度达到了97.7%。

1 基于YOLOv5的车号罐号区域定位

1.1 YOLOv5 简介

YOLOv5目标检测模型是在YOLO^[25-27]系列的基础上经过改进和优化的目标检测模型,其网络模型如图1所示。

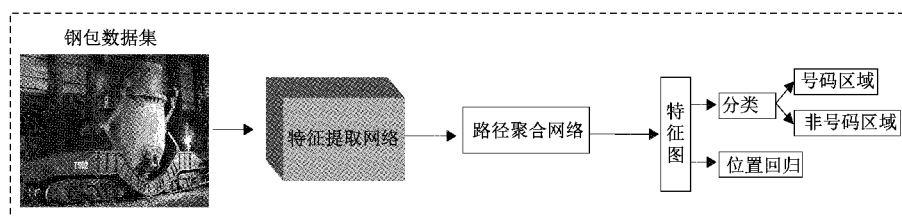


图1 简化的网络结构

基于YOLOv5的钢包运输车车号罐号区域定位步骤如下。

1)将图片送入到特征提取网络,经过卷积、池化和激活函数的作用之后生成图片的高层语义信息特征图。

2)路径聚合子网络将得到的不同层特征图进行融合,可在高层的特征图中增加较多的底层细节信息,从而增加最终识别的精度。

3)分类和回归子网络将融合后的特征图进行全连接等处理,判断所属的类别以及边界框的信息,即字符区域和非字符区域,生成精确的位置信息并进行裁剪,送入到识别网络中进行识别。

1.2 特征提取子网络

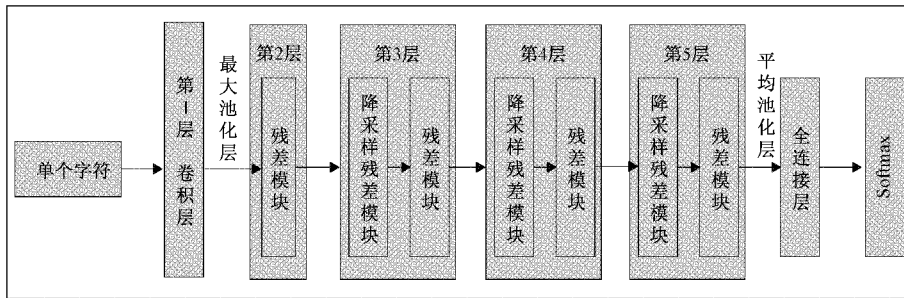
特征提取网络作为YOLOv5的基础,用于提取输入图片的语义信息,特征提取的效果直接影响着网络最终的性能。常见的特征提取网络有VGG、ResNet等,YOLOv5使用的特征提取网络类似于ResNet,将前一层的部分特征叠加到经过卷积后的特征图上,通过部分复制的方式抵消一部分在卷积过程中损失的细节信息,然后将其余的特征图经过卷积生成新的特征图,为了进一步优化卷积层的计算量,在特征提取网络的基础上进行了改进。

1.3 PAN 子网络

路径聚合网络(PAN)在特征金字塔网络(FPN)的基础上进行改进,用于融合特征图。首先根据特征提取网络各阶段的输出,将高层的特征图进行上采样,然后和上一层的输出进行相加,融合得到分辨率较大的特征图,将特征提取网络各层的输出依次进行处理,得到 3 个分辨率不同的特征图(FPN 的输出结构),然后将分辨率最大的特征图再进行下采样,和上一层的特征图进行融合,最终生成 PAN 的 3 个输出,此时输出的特征图包含有较多的底层信息,对于最终的检测效果有较大提升。

2 基于 ResNet 的车号罐号字符识别

残差神经网络(ResNet)由何恺明等于 2015 年提出,获得当年 ILSVRC 竞赛的冠军,其主要贡献是发现了神经网络的退化现象,并且通过短路连接对该现象进行有效缓解。该网络主要功能是进行特征提取,通过添加不同的输出端就可进行不同类型的任务,本文中对其全连接层参数进行修改,即可实现对于钢包字符的识别,使用 ResNet 进行字符识别的网络结构如图 2 所示。



(Softmax 为激活函数)

图 2 ResNet 网络结构

3 网络分析及改进

YOLOv5 网络在许多领域已广泛应用,尤其在通用目标检测领域表现优异,对于大物体的检测有较好的检测精度和速度,在部分移动端能够实现实时的检测效果。但是直接将其应用到钢包运输车车号和罐号检测时准确率并

不高,究其原因主要是车号罐号在复杂环境下表现出的字号偏小、字体模糊、角度倾斜等问题。因此本文针对这些问题对 YOLOv5 网络(号码定位网络)和 ResNet 网络(号码识别网络)进行改进,改进后的网络结构如图 3 所示。其中号码定位网络的输出为 4 个类别,分别是 Pot(罐车)、Pot_num(罐号)、Car_num(车号)和 Num(单个数字),其

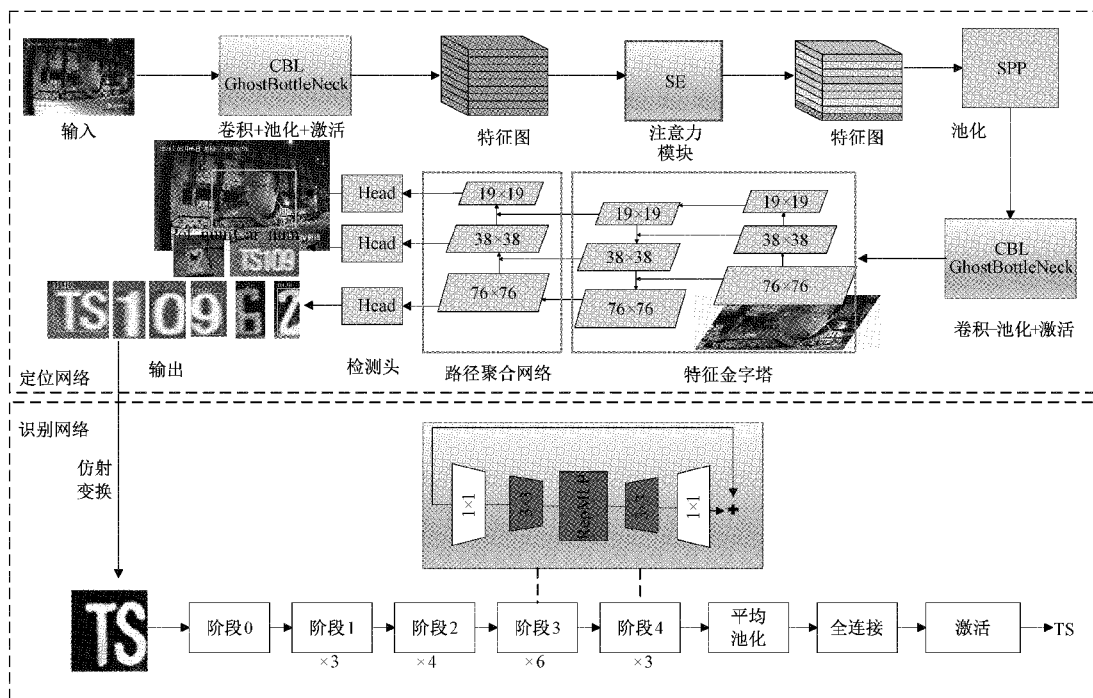


图 3 本文网络框架

中 Num 包含了数字 0~9 以及字母 TS 等 11 个类别,TS 字母仅会出现在车辆上,数字在车辆和罐号上都会出现。在监控画面中会出现日期等其他数字信息,因此使用 Pot_num 和 Car_num 降低画面中其他数字对于识别精度的影响,使用坐标信息对识别的结果进行约束,排除异常识别的数字信息。定位网络输出的 Num 类别会经过放射变换之后送入到识别网络进行号码的识别,最后将识别到的号码信息按照坐标位置进行拼接,产生最终的识别结果,将识别结果绘制到原图中。

3.1 Anchor 的分析及改进

YOLOv5 网络在钢包运输车号码定位等方面应用较少,而通用数据集上的 Anchor 比例(YOLOv5 网络自带 9 组 Anchor)直接应用容易导致号码定位准确性差,因此本文使用 kmeans++ 算法对所用数据集进行统计,计算出合适的 Anchor 比例,以提高定位的精确度,其标记样本(部分)的长宽分布统计如图 4 所示。

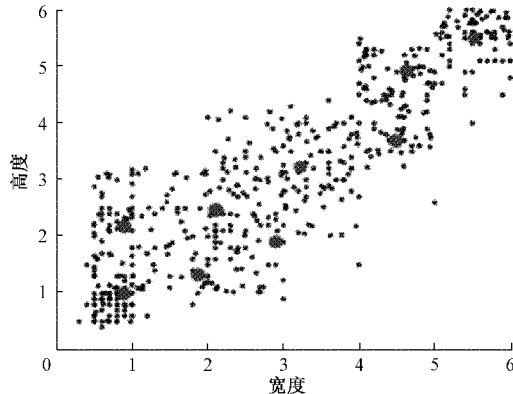


图 4 数据集信息统计图

标准的 kmeans 算法以欧氏距离作为度量,在本文中只关心锚框与真实标注框的交并比(IOUS),因此使用交并比作为相似性的度量。改进前、后 Anchor 的可视化结果如图 5(a)、(b)所示,图中矩形框代表 Anchor 位置。

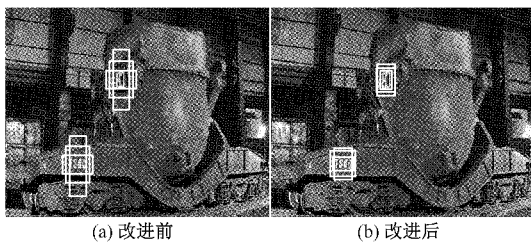


图 5 Anchor 的可视化结果

采用 kmeans++ 聚类的方式替换原始的锚框,使得网络的候选框更贴合号码区域,进一步提升网络的定位精度。kmeans++ 首先会将数据集划分为多个簇,然后判断每个样本与簇中心的距离,并且经过不断的迭代,更新簇中心,直至簇中心不再发生变化。

3.2 特征提取网络的分析与改进

1) 骨干网络卷积层的分析及改进

为了加快实际应用中的模型推理速度,将网络中的部分 BottleNeckCSP 模块替换为了更轻量级的 ghost 瓶颈网络(GhostBottleNeck,GBN)模块。Ghostnet^[28]的核心思想是设计分阶段的卷积计算模块,在得到少量特征图的基础上,使用一次线性卷积得到更多的特征图,新得到的特征图被称为“ghost”,能够通过消除冗余特征获得轻量级模型。传统的非线性卷积指的是卷积、批归一化以及非线性激活等过程的组合,如图 6 所示。

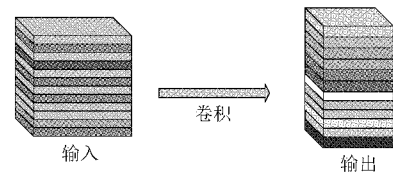


图 6 普通卷积层

线性操作则是直接进行卷积而不包含归一化处理和线性激活过程。在 Ghost 模块中,先进行一次非线性变换输出较少的特征图,如图 7 虚线矩形框所示,然后再进行线性变换,以此减少计算量进而提升网络推理速度。

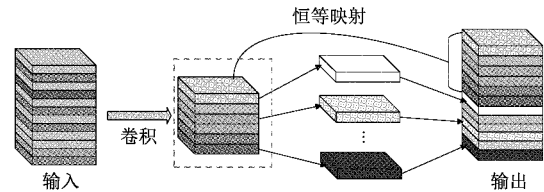
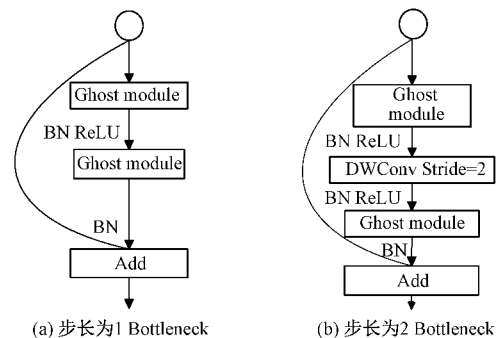


图 7 Ghost 模块

GhostBottleNeck 由两个 Ghost 模块组成,第 1 个 Ghost 层能增加特征的通道数。第 2 个 Ghost 模块能减少通道数,具体结构如图 8 所示。



(BN 为 Batch Normalization,ReLU 为激活函数, DWConv 为深度可分离卷积)

图 8 Bottleneck 结构

2) 基于注意力机制的特征图权重分配

特征图的各通道往往具有不同的重要性,因此在损失计算过程中应对各通道分配不同的权重值,即增大重要通道的权重,降低影响力较小通道的权重,以提升损失计算

的合理性,增强网络的特征提取能力。考虑到人类在进行观察时只关注自己感兴趣的部分,并在一定程度上忽视其他部分,因此,本文在特征提取过程中引入注意力机制,首先将大小为 $W \times H \times C$ 的输入特征图进行“压缩”操作,使其大小变为 $1 \times 1 \times C$ (其中 W, H, C 分别表示特征图的宽、高以及通道数),进而通过“激励”操作计算各通道的权重值,最后将该通道值与输入特征图的对对应通道进行乘积,得到 $(W \times H \times C)$ 的输出特征图,整体流程如图 9 所示。

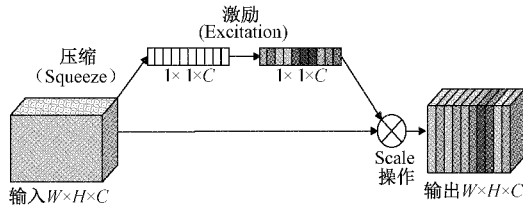


图 9 SE 模块

上述“压缩”操作是指利用全局平均池化操作将原特征图的长和宽“压缩”至 1,仅保留通道维度不变,如图 10 所示。

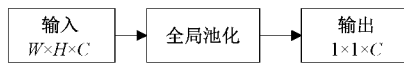


图 10 压缩操作示意图

“激励”操作是在“压缩”之后通过两个全连接层预测通道权重,如图 11 所示。

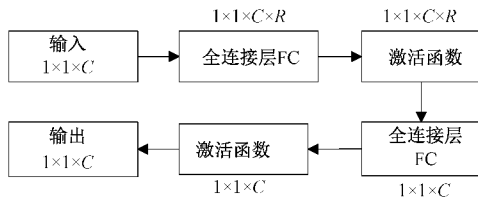


图 11 激励操作示意图

其中 R 为比例系数,通过系数 R 能减少通道的数量,进而降低参数量,第 1 个全连接层有 $R \times C$ 个神经元,第 2 个全连接层有 C 个神经元,通过比例系数将输出通道个数上升到 C 个。在本文中,上一层神经网络的输出 $C=128$,所以 SE 模块的输入维度为 128,压缩的比例系数为 16,参数 R 的取值为 $1/16$,第 1 个全连接层神经元个数就为 $1 \times 1 \times 128 \times 1/16=8$ 个,在经过激活函数的作用后,再次送到 $1 \times 1 \times 128$ 的全连接层进行维度的提升,然后再进行激活,最终输出的维度为 $1 \times 1 \times 128$ 。在得到 $1 \times 1 \times 128$ 的向量之后,就可以进行 Scale 操作,即将得到的各通道权重值分别与原始特征图对应通道相乘,得到经过注意力模块处理的同尺寸特征图,注意力模块的应用提高了号码区域定位的精度。

3.3 号码识别网络分析及改进

为了提高号码识别的准确率,本文分别从图像预处理

和网络结构方面进行优化,先使用号码定位网络将号码区域从原图中精确分割,将分割后的号码区域送入号码识别网络进行识别。

1) 区域图像预处理

在实际应用中,受炼钢厂环境限制可能无法选择合适位置和角度安装工业相机,导致车号罐号图像出现不同程度的倾斜。为了降低号码识别过程中因此造成的干扰,本文在号码区域图像预处理阶段加入图像逆仿射变换操作,与因相机问题造成的仿射变换影响进行抵消。仿射变换过程可以用式(1)进行描述:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} R_{00} & R_{01} & T_x \\ R_{10} & R_{11} & T_y \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

其中, (x, y) 表示原始像素坐标, (x', y') 表示该像素经过变换后的坐标, T_x, T_y 表示平移量,参数 R 代表了线性变换,包含了图像的旋转、缩放等信息。本文中 R 参数代表了旋转角度 θ 信息, θ 为 Y 轴与最小外接矩形小于 90° 的夹角。 R_{00} 为 $\cos\theta$, R_{01} 为 $\sin\theta$, R_{10} 为 $-\sin\theta$, R_{11} 为 $\cos\theta$ 。在应用中需要计算出参数 R 和 T 的值,进而得到上述变换矩阵的逆矩阵,即可复原像素原始坐标,如图 12 所示。

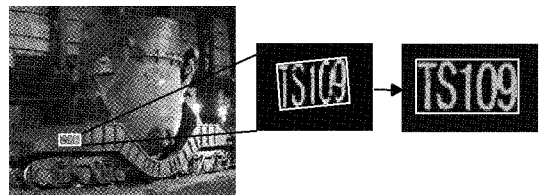
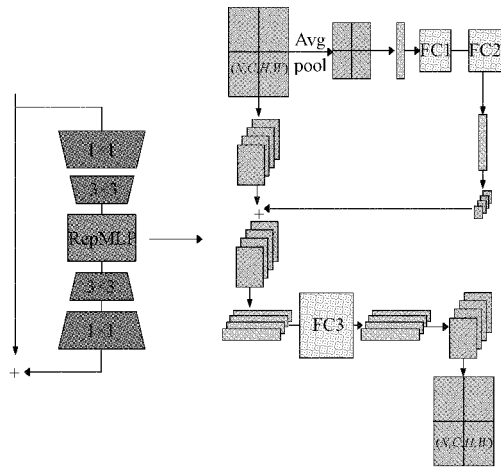


图 12 仿射变换效果图

仿射变换的主要思想就是先将图像进行二值化,得到前景和背景的像素点集合,然后获得图像的最小外接矩形,此处的最小外接矩形是指最小面积外接矩形,根据外接矩形可以得到图像偏斜的角度 θ ,由此可以知道线性变换矩阵 R 的参数值,并且经过统计,数据集中的号码区域倾斜角度均小于 30° ,所以在本文中对小角度进行变换。最后将角度和大小信息转变成变换矩阵的参数实现原始图像的仿射变换。实验表明,仿射变换可以增加识别网路对于号码识别的准确率。

2) 号码识别网络结构与改进

本文选择 ResNet50 作为号码识别基础网络,其简单精巧的网络结构使之在图像分类领域不但能达到较高的精度,还可拥有较好的推理速度。为了进一步提升识别网络对号码识别的准确率,本文将 ResNet50 中的部分卷积层进行了改进,将 ResNet50 的 $c3$ 和 $c4$ 中的 3×3 卷积块改进为 RepMLP 模块,RepMLP 模块在全连接层加入局部的先验信息使其具备强大的图像识别能力,推理时的 RepMLP 的结构如图 13 所示。



(Avg pool为平均池化, FC为全连接层)

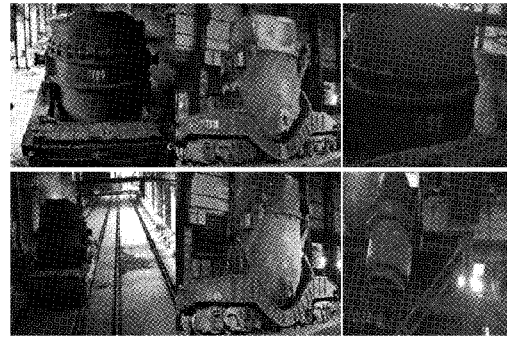
图13 RepMLP 结构

在本文中将 ResNet 的 c3 和 c4 层中的 3×3 卷积改进为了 RepMLP 模块,结果表明,对于卷积层的改进有效提高了检测的准确率。

4 实验结果分析

本文实验环境如下:硬件系统为 3.0GHz CPU,16 G 内存,NVIDIA GeForce 1080Ti 显卡 $\times 2$ 张,显存为 11 G;软件环境为 Ubuntu20.04 操作系统,Pytorch 深度学习框

架,开发语言为 Python,CUDA 10.1,CUDNN 8.0.2。部署测试环境包括移动端平台 joston nano 以及 2.2 GHz CPU,32 G 内存,T4 推理卡 16 GB 显存。实验中数据集包括 3 类,分别来自不同钢厂的出铁口运输线(记为数据集 A)、过跨线的采样平台(记为数据集 B)以及整备、热修、烘烤等静止位(记为数据集 C)。数据集包含了倾斜、模糊、污损情况下的图片。最终数据集包含 3 600 张图片,数据集示例如图 14 所示。



(a) 数据集A (b) 数据集B (c) 数据集C

图14 数据集示例图

数据增强部分使用了图片翻转、对比度增强和随机颜色等数据增强方式,可在一定程度上模拟工厂内复杂的灯光变化,有助于增强模型的鲁棒性,部分的数据增强效果如图 15 所示。

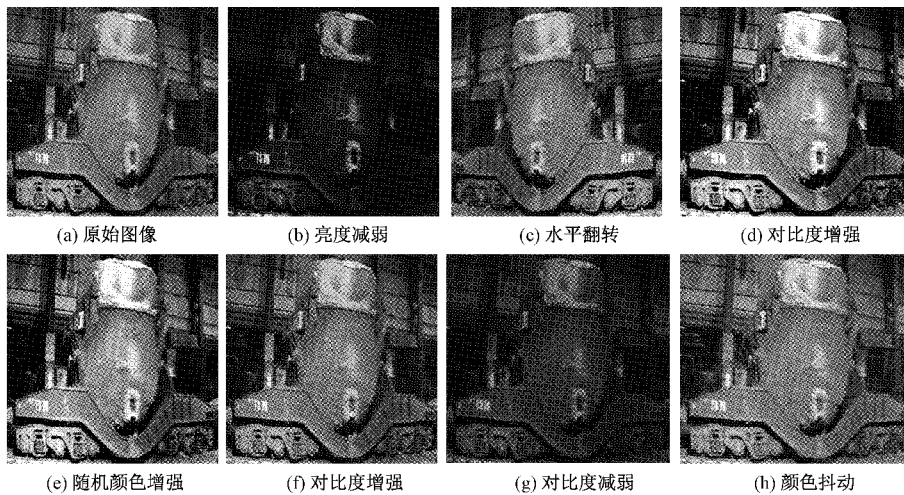


图15 数据增强效果图

4.1 车号罐号定位实验设计

为了验证本文算法有效性,采用目标检测领域通用的准确率(Acc)和召回率(Rec)作为评价标准,其计算公式如式(2)、(3)所示。

$$Acc = \frac{\text{定位正确的车号数量}}{\text{算法判定为车号的个数}} \quad (2)$$

$$Rec = \frac{\text{定位正确的车号个数}}{\text{测试集中车号个数}} \quad (3)$$

一般随着设定阈值的提升,准确率会上升,召回率会

下降,因此为对本文算法进行全面分析,还采用 F1 评估值综合衡量指标,其计算公式如式(4)所示。

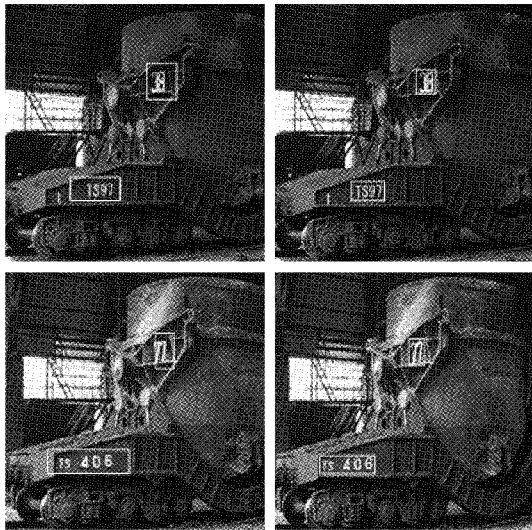
$$F1 = \frac{2 \times Acc \times Rec}{Acc + Rec} \quad (4)$$

本文通过在自制的 3 套数据集上分别进行网络训练来验证算法改进效果。针对号码定位网络,本文做出两项改进,加入 RepMLP 模块和仿射变换(affine transformation,AT)。为了验证改进效果,设计 4 项实验,其实验结果如表 1 所示。

表 1 定位结果对比

方法	准确率	FPS
ResNet	0.967	102
ResNet+AT	0.970	98
ResNet+RepMLP	0.975	86
ResNet+AT+RepMLP	0.980	81

由表 1 可知,实验 2、3 对于号码定位的精度均优于实验 1,说明针对网络的两项改进均有利于号码区域的定位;由于实验 2 加入了注意力模块,导致模型参数数量的提升,造成其在精度上升的同时处理速度略微降低;实验 3 对于精度的影响较小,但是提升了模型的处理速度,有利于同时处理多路视频或在性能较低的移动端进行部署;实验 4 是实验 2、3 的综合效果,在准确率提高的同时,处理速度也得到了提升,相比于原始网络,改进后的网络定位准确率在数据集上提升了 5.1%,召回率提升了 6.3%,整体的 F1 值提升了 5.7%,总体上实现了比较精准的号码定位。定位效果可视化如图 16 所示。



(a) YOLOv5定位效果图 (b) 本文算法

图 16 号码定位效果图

图 16(a)为原始的定位效果,可以看出网络对于号码区域的覆盖面积比较大,即定位较粗糙;16(b)为改进后的定位效果图,对于号码区域的覆盖面积较小,即实现了较精准的号码定位。

4.2 号码识别算法实验分析

针对本文在号码识别过程做出的改进与优化,设 4 组实验进行验证,实验结果如表 2 所示。

由表 2 可知,原始未改进(实验 1)的 ResNet 对于号码识别的准确率为 96.7%,通过加入仿射变换(实验 2),对于号码识别的准确率提升了 0.3%,说明经过图像矫正能够在一定程度上提高识别的准确率。通过替换部分的卷积层(实验 3),对号码识别的准确率达到 97.5%,提升了

表 2 字符识别结果对比

方法	Acc	Rec	F1	FPS
YOLOv5	0.852	0.810	0.830	45
YOLOv5+SE	0.891	0.821	0.854	43
YOLOv5+GBN	0.856	0.811	0.832	56
YOLOv5+SE+GBN	0.903	0.873	0.887	53

0.8%,相比于未改进的网络处理速度下降了 26 FPS,这是由于 RepMLP 中含有较多的全连接层,在准确率上升的同时网络参数有所增加,所以导致了处理速度下降。在综合了仿射变换和替换卷积层之后,总体的识别准确率达到 98%,相比原始的网络有 1.3%的准确率提升。

对 ResNet 的最后输出进行修改,本文分类的类别为 12 个类别,因此将最后的输出节点个数改为了 12,每个类别的识别的准确率如表 3 所示。

表 3 字符识别准确率结果

字符	准确率
0	0.97
1	0.97
2	0.98
3	0.95
4	1
5	0.97
6	0.98
7	0.98
8	1
9	0.96
T	0.98
TS	1
mAP	0.98

在单个字符的识别中,字符 TS、4 和 8 的识别准确率达到了 100%,分析是 TS 在数据集中的占比较多,4 和 8 字符印刷较清晰,因此对这 3 个字符的识别准确率较高,字符 3 的识别精度较低,分析后发现字符 3 的图像数据存在污损面积较大的情况,所以导致精度不高。

整体车号和罐号的识别准确率如表 4 所示。

表 4 号码整体识别准确率

模型	准确率	召回率	F1
Origin-ResNet50	0.968	0.831	0.89
Ours	0.977	0.85	0.90

图 17 为整体识别的可视化效果图,与原始的 ResNet 网络相比准确率提升了 0.90%,实现了比较准确的号码识别。

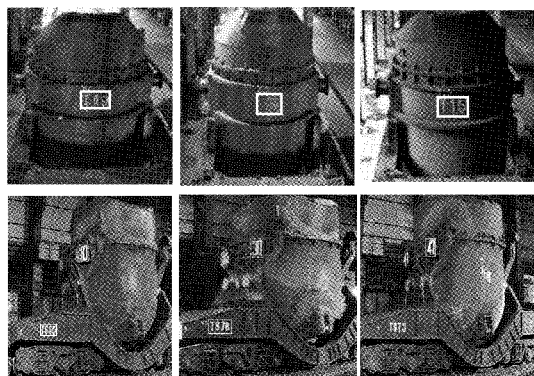


图 17 识别效果图

5 系统部署与性能分析

为了实现工厂环境下的多路视频的实时识别,使用 DeepStream 平台进行部署,实验所用的硬件包括 T4 推理卡和 jeson nano 移动开发套件。通过在 jeson nano 平台 (MAX well GPU) 上进行测试,本模型的处理速度能达到 9 FPS。对于钢包运输车来说,运行速度较慢,因此每一帧图像都进行检测不仅实用性不强,而且还会浪费计算资源,因为在实际的应用中一台机器会处理多路的视频输入,所以每隔 3 帧进行检测,中间使用跟踪的方式(3 种跟踪方式 NvDCF、KLT、IOU)进行拟合,通过中间帧跟踪能提高检测效率,对于多路视频和不同的跟踪方式在视频流上的处理速度做了多组实验,结果如表 5 所示。

表 5 不同跟踪方式下对多路视频的处理速度

多路 视频	T4			jeson nano		
	NvDCF	KLT	IOU	NvDCF	KLT	IOU
2	290	310	298	14	16	15
4	148	156	150	7	8	8
8	75	80	76	×	×	×
16	38	44	40	×	×	×
32	20	25	22	×	×	×

由表 5 可知,使用不同的跟踪方式推理速度有较明显的变化,当输入视频达到 16 路时,在 T4 推理卡上系统尚能达到实时处理的能力。而 jeson nano 平台作为低能耗边缘设备,算力有限,当输入 2 路视频时已难以实现实时处理。

6 结 论

本文使用网络的级联实现对于钢包运输车车号和罐号识别,将 YOLOv5 网络应用到了车号和罐号的定位任务上,对网络的结构进行了优化,通过在定位网络中加入 GBN 模块和注意力模块提升模型的定位精度和检测的速度,降低网络的漏检率;车号的识别阶段使用改进的

ResNet50 网络对获得的的车号和罐号识别,通过加入仿射变换,增强了模型对于形变字符的识别精度,将 RepMLP 模块加入到网络中,虽然模型的参数略有增加,但是也提高了识别的准确率。在工厂环境下当输入为 16 路视频时,可以达到实时检测的效果,并且车号和罐号检测的准确率达到了 97.7%,达到了实际场景中使用的要求。但在光线变化较剧烈(如明暗变化大、有反射情况和车号罐号遮挡严重等)时,识别精度不高且效果不佳,因此,未来将进一步做研究实验,以解决在光线变化时识别精度不高的问题。

参考文献

- [1] 卢笑,曹意宏,周炫余,等. 基于深度强化学习的两阶段显著性目标检测[J]. 电子测量与仪器学报, 2021, 35(6): 34-42.
- [2] 李澄非,蔡嘉伦,邱世汉,等. 基于改进 YOLOv4 算法的 PCB 缺陷检测[J]. 电子测量技术, 2021, 44(17): 146-153.
- [3] 李娅,王兴路,雷蕾,等. 基于极性检测和改进 CNN 框架的车牌识别方法[J]. 电子测量技术, 2021, 44(14): 26-32.
- [4] 曹立波,陈峥,颜凌波,等. 基于 RFID,视觉和 UWB 的车辆定位系统[J]. 汽车工程, 2017, 39(2): 225-231.
- [5] 蒋荣,陈柏. 基于 UPF 的车辆紧耦合定位算法[J]. 机电工程, 2018, 35(4): 420-424.
- [6] 宋彭彭,曾祥进,郑安义,等. 基于注意力机制的自然场景文本检测[J]. 电子测量技术, 2021, 44(14): 122-127.
- [7] SHI B, BAI X, BELONGIE S. Detecting oriented text in natural images by linking segments[J]. IEEE Computer Society, 2017, DOI: 10.1109/CVPR.2017.371.
- [8] ZHOU X, YAO C, WEN H, et al. EAST: An efficient and accurate scene text detector[J]. IEEE, 2017, DOI: 10.1109/CVPR.2017.283.
- [9] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist Text proposal network[J]. CoRR, 2016, DOI: 10.1007/978-3-319-46484-8_4.
- [10] 游洋彪,石繁槐. 短边顶点回归网络:新型自然场景文本检测器[J]. 哈尔滨工业大学学报, 2021, 53(12): 89-97.
- [11] 谢斌红,秦耀龙,张英俊. 基于学习主动中心轮廓模型的场景文本检测[J]. 计算机工程, 2022, 48(3): 244-252, 262.
- [12] 卢利琼,吴东,吴涛,等. 挖掘文本框位置特性的 anchor-free 自然场景文本检测[J]. 计算机应用研究, 2021, 38(8): 2556-2560.

- [13] 朱斌, 陈磊, 邬金萍. 基于改进卷积神经网络结构的机器视觉室内定位算法[J]. 国外电子测量技术, 2021, 40(1): 58-64.
- [14] GAO Y, CHEN Y, WANG J, et al. Semi-supervised scene text recognition [J]. IEEE Transactions on Image Processing, 2021, 30: 3005-3016.
- [15] LIN Q, LUO C, JIN L, et al. STAN: A sequential transformation attention-based network for scene text recognition [J]. Pattern Recognition, 2021, 111: 107692.
- [16] WANG Q, HUANG Y, JIA W, et al. FACLSTM: ConvLSTM with focused attention for scene text recognition[J]. 中国科学: 信息科学(英文版), 2020(2): 31-44.
- [17] CHEN Y, CHEN Z, CONG G, et al. SSTD: A distributed system on streaming spatio-textual data[J]. Proceedings of the VLDB Endowment, 2020, 13(12): 2284-2296.
- [18] LIAO M, SHI B, BAI X, et al. TextBoxes: A fast text detector with a single deep neural network[J]. CoRR, 2016, DOI: 10.11499/sicjil.58.637.
- [19] 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 253-260.
- [20] 王纪武, 罗海保, 鱼鹏飞, 等. 基于 Faster R-CNN 的无人机航拍图像小目标检测[J]. 测试科学与仪器, 2020, 11(1): 11-16.
- [21] 王建军, 魏江, 梅少辉, 等. 面向遥感图像小目标检测的改进 YOLOv3 算法[J]. 计算机工程与应用, 2021, 57(20): 133-141.
- [22] 张思宇, 张轶. 基于多尺度特征融合的小目标行人检测[J]. 计算机工程与科学, 2019, 41(9): 1627-1634.
- [23] 孙泽军, 丁萌, 王洁, 等. 基于频域显著性分析的红外小目标检测算法[J]. 电光与控制, 2017, 24(10): 1-6.
- [24] 史晓刚, 白晓东, 李丽娟, 等. 一种基于小波域的双色红外弱小目标检测算法(英文)[J]. 红外技术, 2015, 37(12): 1027-1031.
- [25] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [J]. IEEE, 2016, DOI: 10.1109/CVPR.2016.91.
- [26] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [J]. IEEE, 2017, DOI: 10.1109/CVPR.2017.690.
- [27] HIROKI N, HARUYOSHI Y, TOMOYA F, et al. A lightweight YOLOv2[J]. Field-Programmable Gate Arrays, 2018, DOI: 10.1145/3174243.3174266.
- [28] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, DOI: 10.1109/CVPR42600.2020.00165.

作者简介

张继凯, 工学博士, 讲师, 研究生导师, 主要研究方向为数据可视化技术和增强现实技术, 研究空间结构识别、自然交互、环境感知以及大规模科学计算数据可视化, 包括三维空间标量场、矢量场, 大规模地形网络构建等。

E-mail: jkzhang0314@imust.edu.cn

梁勇, 硕士研究生, 主要研究方向为目标检测和图像实例分割处理等。

E-mail: 13088906116@163.com

周亚辉, 硕士研究生, 主要研究方向为目标检测。

E-mail: 13026954365@163.com

柴轶凡, 工学博士, 讲师, 研究生导师, 主要研究方向为高炉炼铁原燃料优化、复杂共生矿高效冶炼新技术和冶金资源能源综合利用。

E-mail: chaiyifan@live.cn