

DOI:10.3969/j.issn.1003-5060.2024.11.004

# 基于改进 YOLOv5 算法的票据检测

扈 静, 贺竞娇, 龚 宇, 汪俊峰

(合肥工业大学 机械工程学院, 安徽 合肥 230009)

**摘 要:**在票据全流程识别过程中,针对不同类型票据的目标检测是关键步骤,为实现日常报销过程中不同类型票据快速检测其类型和位置,文章提出一种基于改进 YOLOv5 算法的票据检测方法。对于原始数据集进行预处理,模拟票据检测中可能出现的干扰信息,提高训练模型的性能;利用 CSPDarkNet53 网络进行特征提取,采用基于重合面积、中心点距离、长宽比和角度 4 个几何参数的 SIOU 目标位置损失函数,对原损失函数进行改进,提升 YOLOv5 票据检测的精度,构建票据检测方法。最后通过自建实际拍摄的票据数据集来验证基于改进 YOLOv5 算法的票据检测的有效性。结果表明,该方法拥有较高的平均精度均值(99.20%)和检测速度(51 帧/s),可以满足实际应用场景的要求。

**关键词:**票据;目标检测;识别;改进 YOLOv5;损失函数

**中图分类号:**TH693.9;TP274.5

**文献标志码:**A

**文章编号:**1003-5060(2024)11-1459-06

## Bill detection based on improved YOLOv5 algorithm

HU Jing, HE Jingjiao, GONG Yu, WANG Junfeng

(School of Mechanical Engineering, Hefei University of Technology, Hefei 230009, China)

**Abstract:**In the whole process of bill recognition, object detection for different types of bills is a key step. In order to realize the rapid detection of the types and positions of different types of bills in the daily reimbursement process, this paper proposes a bill detection method based on the improved YOLOv5 algorithm. The collected images are preprocessed to simulate the possible interference information in the bill detection and improve the robustness of the training model. With the help of CSPDarkNet53 network for feature extraction, the SIOU target position loss function based on the four geometric parameters of coincidence area, center point distance, aspect ratio and angle is used to improve the original loss function, so as to improve the accuracy of YOLOv5 bill detection, and build a bill detection model. Finally, the effectiveness of the bill detection based on the improved YOLOv5 algorithm is verified through the actual test set. The experimental results show that the method has high mean average precision(99.20%) and detection speed(51 frames per second), and can meet the requirements of practical application scenarios.

**Key words:**bill; object detection; recognition; improved YOLOv5; loss function

## 0 引 言

票据是会计核算重要的原始凭证,是加强会计监督及内控建设的重要方面,是财政、税务、审计等部门进行监督检查的依据<sup>[1]</sup>。目前大多数企

业、单位等票据报销流程拥有统一的标准,主要有将票据粘贴在费用报销单上、提交费用报销单、审核、审批等步骤。而这些步骤中,基础且最为重要的一步是将粘贴在报销单上的不同票据定位并识别出来,此步骤完成后才能继续进行之后的真伪

收稿日期:2023-02-09;修回日期:2023-03-16

基金项目:安徽省科技攻关计划资助项目(JZ2016AKKG0837)

作者简介:扈 静(1976—),女,安徽六安人,博士,合肥工业大学副教授,硕士生导师。

验证、信息提取<sup>[2]</sup>等步骤。

传统的票据报销流程中,票据目标检测步骤通常由人工完成,但对于大型数据集来说,手动检测极其耗时且费力、成本高昂、流程繁琐,而这些问题也是当下迫切需要被解决的。

在深度学习出现之前,传统票据检测方法主要以票据图片中的方框顶点以及票据本身自带的颜色特征等作为实现位置检测的基准。文献[3]采用角点特征检测法定位票据特征位置,实现在可见光环境下的票据识别目的;文献[4]首先对待处理图像进行对比度增强处理,再根据最小二乘法拟合出纸币边缘,并完成边缘校正,最终实现纸币位置区域检测。与人工视觉检测方法相比,这类方法具有在复杂生产环境中长期运行的优点,并且有效提高了检测精度和检测效率。然而这种方法存在很大的局限性,人工设计的特征提取方案通常需要丰富的专家经验和大量的实验,这一过程的高度复杂性导致成本高和时间消耗,并且不能保证长期的有效性和泛化性,因此这种方式难以在实际应用中得到长足的发展。

随着深度学习技术的快速发展,各类基于深度学习的检测方法已广泛应用在各行各业,并成为主流方法。其可以从底层数据中学习特征,对复杂结构的表征能力更强,从而取得更优异的效果。关于票据分类任务,文献[5]提出一种基于谷歌二维码识别开源项目 Zxing 的发票二维码检测技术,可以通过模糊发票的 QR(quick response)二维码对发票进行分类和识别;文献[6]提出一种基于改进的 AlexNet 网络进行发票分类的方法,准确率可达 98%;文献[7]提出一种基于残差网络的发票分类模型,对于增值税发票、火车票和出租车发票的分类准确率达到 98.81%。但以上方法仅对于票据进行了分类,只能输出类别信息而无法输出位置信息,在实际的报销单检测流程中存在局限性,故提出针对报销单的票据检测方法迫在眉睫。

基于深度学习的目标检测代表性的两类算法分别为基于区域候选的目标检测方法<sup>[8]</sup>和基于回归方法的目标检测方法<sup>[9]</sup>,也称 Two-stage 方法和 One-stage 方法。Two-stage 方法以区域卷积神经网络(region convolutional neural networks, R-CNN)<sup>[10]</sup>系列为代表,通过选取候选框,对候选框进行分类或回归两步骤来完成目标检测;而 One-stage 方法以 YOLO<sup>[11]</sup>系列为代表,主要通过采用不同的尺寸对图像进行遍历抽样,然后利

用 CNN 提取特征后直接进行回归。两类方法各有利弊,总体上来讲,Two-stage 方法更侧重于检测精度,One-stage 方法更侧重于检测速度。故在离线检测系统中采用 Two-stage 方法情况较多,而实时性检测系统中使用 One-stage 方法较多。

报销单涉及到的票据目标较为明显,目标相对于整张图片所占比例适中,属于常规目标检测,而对于票据检测问题,保证其检测速度是提升报销过程效率的关键,要在保证速度的基础上提升其检测精度,因此,本文提出以 YOLOv5 为基础的票据目标检测方法,在保证检测速度的情况下,对其改进以提升检测精度,并通过与其他方法进行对比,验证本方法的实时性和准确性。

## 1 基于 YOLOv5 的票据检测方法

### 1.1 YOLOv5 网络原理

YOLOv5 的网络结构分为输入端、Backbone、Neck、输出端 4 个部分。输入端中首先采用随机缩放、随机裁剪、随机排布的方式对 4 张输入图像进行拼接,称为 Mosaic 数据增强,作用是扩充数据集并减少对图形处理器的数量要求。YOLOv5 可以针对不同类型的数据集在训练时进行锚框的自适应计算,这一步骤可以减少算法的整体计算量,也可提升检测速度。

Backbone 模块主要是由 Focus、通信顺序进程(communication sequential process, CSP)结构构成<sup>[12]</sup>。原始的  $608 \times 608 \times 3$  图像输入 Backbone 模块中的 Focus 结构之后,经过切片操作变为  $304 \times 304 \times 12$  的特征图,该特征图经过一次卷积核为 32 卷积操作之后,成为一个  $304 \times 304 \times 32$  的特征图。此外,Backbone 主干网络中的 CSP 结构为 CSP1-X,其主要作用是对特征图进行特征提取,从图像中提取丰富的信息。

Neck 模块采用特征金字塔(feature pyramid network, FPN)+个人局域网(personal area network, PAN)结构,并新增了 CSP2-X 结构加强网络特征融合的能力。

Prediction 模块包括边界框坐标损失函数和非极大值抑制。YOLOv5 中边界框坐标损失函数采用 GIoU 损失函数计算损失值。非极大值抑制可以消除多余的框,找到最佳的物体检测位置。

### 1.2 损失函数

本文检测方法的误差损失函数由 3 部分组成,分别为置信度损失、分类损失和预测边界框的

坐标损失误差。

在票据检测的过程中,首先对于合理的候选区域进行选择并预测,预测结果根据置信度的高低来进行选择,保留置信度高的预测框,并对这些预测框进行非极大值抑制处理,最终得到多组边界框后导出位置参数。本文的票据检测方法中使用二元交叉熵函数计算目标置信度得分和类别概率的损失。

关于边界框的坐标损失,YOLOv5 原本采用 GIoU,主要依赖于边界框回归指标的聚合,即预测框与真实框的距离、重叠区域和纵横比。GIoU 损失函数计算公式为:

$$L_{\text{GIoU}} = 1 - L_{\text{IoU}} + |C/(B \cup B^{\text{GT}})| / |C| \quad (1)$$

其中: $B$  为预测框; $B^{\text{GT}}$  为真实框; $C$  为包含预测框  $B$  与真实框  $B^{\text{GT}}$  的最小区域; $L_{\text{IoU}}$  损失为预测框  $B$  与真实框  $B^{\text{GT}}$  的交并比。 $L_{\text{IoU}}$  的计算公式为:

$$L_{\text{IoU}} = |B \cap B^{\text{GT}}| / |B \cup B^{\text{GT}}| \quad (2)$$

### 1.3 损失函数改进

相比 IoU,虽然 GIoU 损失函数精度和速度有一定程度的提高,解决了部分边界框不重合的问题,但在实际检测过程中,需要进行较多次数的迭代才能收敛,而如果出现了一个框包含另一框的情况时,GIoU 损失就会退化为 IoU 损失,检测效果将大打折扣。

为解决 GIoU 损失函数收敛速度较慢问题,并提高回归精度,本文提出了一种改进的 YOLOv5 算法的票据检测方法,将原本使用的 GIoU 损失函数修改为 SIoU 损失函数<sup>[13]</sup>,SIoU 损失函数引入了角度、重合面积、中心点距离、长宽比这 4 个几何参数,相比 GIoU 损失函数的重合面积一个几何参数有了较大的改善。SIoU 损失函数的计算公式为:

$$L_{\text{SIoU}} = 1 - L_{\text{IoU}} + (\Delta + \Omega) / 2 \quad (3)$$

其中, $\Delta$  为距离损失,与真实框和预测框的最小外接矩形有关,计算公式为:

$$\Delta = \sum_{t=x,y} \{1 - \exp[-(2 - \Delta)\rho_t]\} \quad (4)$$

$\rho_x$  和  $\rho_y$  计算公式为:

$$\rho_x = \left(\frac{b_x^{\text{GT}} - b_x}{C_w}\right)^2 \quad (5)$$

$$\rho_y = \left(\frac{b_y^{\text{GT}} - b_y}{C_h}\right)^2 \quad (6)$$

其中: $C_w$  和  $C_h$  分别为真实框与预测框最小外接矩形的宽和高; $b_x$  和  $b_y$  分别为预测框中心点横、纵坐标; $b_x^{\text{GT}}$  和  $b_y^{\text{GT}}$  分别为真实框中心点横、纵坐

标。距离损失示意图如图 1 所示,图 1 中, $B$ 、 $B^{\text{GT}}$  分别表示预测框和真实框。

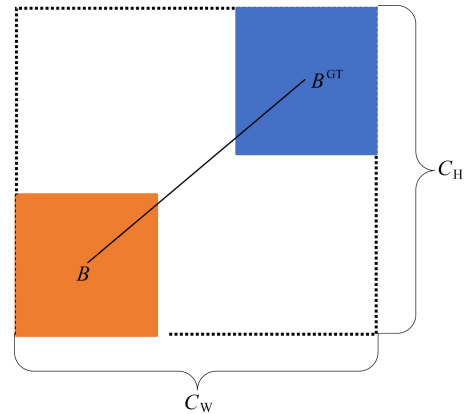


图 1 距离损失示意图

$\Delta$  代表角度损失,计算公式为:

$$\Delta = 1 - 2\sin^2\left(\arcsin\frac{C_h}{\sigma} - \frac{\pi}{4}\right) \quad (7)$$

其中: $C_h$  为真实框与预测框中心点的高度差; $\sigma$  为真实框与预测框中心点的距离。角度损失示意图如图 2 所示,图 2 中  $B$ 、 $B^{\text{GT}}$  分别表示预测框和真实框。

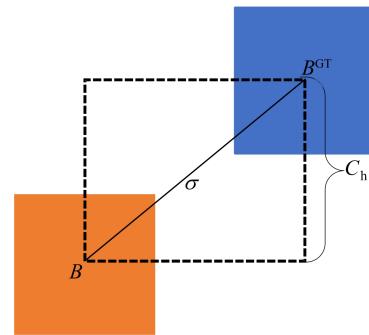


图 2 角度损失示意图

$\Omega$  表示形状损失,计算公式为:

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (8)$$

其中, $\theta$  为控制对形状损失的关注程度。为避免过于关注形状损失而降低对预测框的移动,在本文中取 4<sup>[13]</sup>。

其中  $\omega_w$  和  $\omega_h$  计算公式为:

$$\omega_w = \frac{|W - W^{\text{GT}}|}{\max(W, W^{\text{GT}})} \quad (9)$$

$$\omega_h = \frac{|H - H^{\text{GT}}|}{\max(H, H^{\text{GT}})} \quad (10)$$

其中: $W$  和  $H$  分别为预测框的宽和高; $W^{\text{GT}}$  和  $H^{\text{GT}}$  分别为真实框的宽和高。形状损失示意图如

图 3 所示。

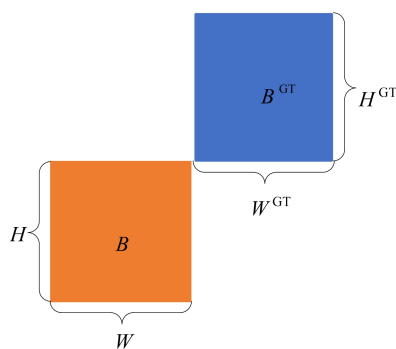


图 3 形状损失示意图

## 2 网络模型训练

### 2.1 数据集制作

在开始票据目标检测之前,需要准备票据数据集,本文主要针对火车票、定额发票、增值税发票这 3 类票据进行研究。

市面上增值税发票可分为增值税普通发票和增值税专用发票,但因票版相差不大,故在本文中可看作一类增值税发票来检测与识别。火车票主要为自助取票机打印蓝色圆角车票,但也存在少量人工窗口打印的粉色直角车票,2 类火车票在数据集中均需要体现。目前使用的定额发票多存在于停车费报销等场合,均为统一制式,印章为国税局统一印章。

经过网络爬虫、实地拍摄等方法,共获取火车票图片 134 张,包含自动取票机(蓝色)和人工窗口取票/上车补票(粉色)2 类火车票,增值税发票图片 120 张,定额发票 149 张,票据数据集图像如图 4 所示。



图 4 票据数据集图像

由于目标检测所用的 YOLO 网络需要 xml 形式的数据集,因此需要使用 LabelImg 进行标

注,最终得到与图片一一对应的 xml 文件,其中选取 80% 用于训练,20% 用于验证。使用 134 张实拍报销单(包含 50 张普通发票,42 张定额发票,42 张火车票)图像进行测试,每张图片里含有 1 个或多个目标,以更好地验证模型的性能。

### 2.2 数据集预处理

由于票据自身的特性,在进行票据数据集的预处理时与其他目标检测数据集预处理方法会有所不同,常见目标检测任务对于数据集的预处理,往往只需要放大、缩小、旋转、镜像、拉伸、裁剪等方式,但由于发票目标检测是作为之后的真伪检测、信息提取的基础,与一般的图像样本预处理方式会有所不同,而且发票图像的数据增强手段要求更加丰富。例如,由于拍摄设备的分辨率等不同,需要对拍摄的票据图像进行模糊、加噪等降低图像质量的操作;此外,实际生活场景中,票据由于其特性,在进行识别时可能会出现折痕、噪点、污渍等对识别不利的情况,因此需要增加一部分的干扰信息来模拟这些不利因素。这些方法既可以扩充数据集,也可提高训练模型的鲁棒性。

考虑到 YOLOv5 网络中已存在 Mosaic 数据增强步骤,本文对于原始数据集采取如下数据增强方法:① 旋转。对初始数据集中的图片进行旋转,旋转角度为  $90^\circ$ 、 $180^\circ$ 、 $270^\circ$  随机取其一,以模拟不同的票据拍摄角度。② 添加噪声。随机对数据集中的图片添加高斯噪声或椒盐噪声。③ 添加直线。随机在数据集图片中绘制白色直线,以模拟可能存在的票据折痕。

经过以上方法,数据集由最初的 262 张扩充至 786 张,较为充分,可满足数量及质量要求。

### 2.3 模型训练

实验以采集的 3 种票据图像为检测对象。本实验的硬件环境为: Intel(R) Xeon(R) Silver 4210R CPU @ 2.40 GHz 的处理器, GeForce RTX3090 的显卡;软件环境主要为:基于 Ubuntu 18.04.5 LTS 操作系统,基于 PyTorch 深度学习框架,基于 CUDA11.1 及 PyCharm 集成的开发环境。

为使模型更好地适用本数据集,模型的超参设置为默认参数,初始学习率为 0.001,批处理数量为 64,设置待训练模型运行迭代轮数为 400。

### 2.4 评价标准

本文使用平均精度均值(mean average precision, mAP)作为模型精确度的评价标准,即所有类别 AP 的平均值,每一种类别票据的平均精度

的计算公式为:

$$P_{AP} = \frac{1}{11} \sum_{i=1}^{11} \rho \quad (11)$$

其中,  $\rho$  为在这一类别的  $P$ - $R$  曲线(查准率-查全率曲线)上的点的纵坐标。本文使用的是 11 点插值法,即在  $P$ - $R$  曲线中取出 11 个数值,求出平均值。平均精度均值为每个类别的平均精度的均值,计算公式为:

$$P_{mAP} = \frac{1}{n} \sum_{i=1}^n P_i \quad (12)$$

其中,  $n$  为类别数。同时,本文采用每秒内可以处理的图片数量值帧率(frame per second, FPS)作为检测速度的评价标准。该指标越大,表示每秒内可处理图片数量越多,检测速度越快。

### 3 实验验证

网络性能评价主要依据训练好的网络是否具备良好的鲁棒性,即网络在验证集中的表现。本文对所提出的方法设计了 2 组实验,验证改进损失函数方法以及改进数据集预处理方法的可行性和准确性。所用到的验证集由未经训练的 134 张报销单实拍图像构成,包含普通发票、定额发票和火车票。

#### 3.1 消融实验

本实验将未经改进的普通 YOLOv5 检测方

法与使用 SIOU 损失函数改进后的 YOLOv5 方法进行对比,消融实验结果见表 1 所列。其中:方法 1 使用原始 GIoU 损失函数的 YOLOv5;方法 2 在方法 1 的基础上,添加数据集预处理;方法 3 将 YOLOv5 中的损失函数改为 SIOU;方法 4 在方法 3 的基础上添加数据集预处理。

表 1 消融实验结果 %

| 方法   | mAP  | AP   |      |      |
|------|------|------|------|------|
|      |      | 火车票  | 普通发票 | 定额发票 |
| 方法 1 | 95.6 | 99.3 | 92.3 | 95.1 |
| 方法 2 | 98.9 | 99.4 | 98.4 | 99.0 |
| 方法 3 | 98.1 | 99.2 | 96.2 | 98.9 |
| 方法 4 | 99.2 | 99.5 | 99.2 | 98.0 |

从表 1 可以看出:修改 YOLOv5 中的 GIoU 损失函数为 SIOU 损失函数后,整体的平均精度均值提高了 2.5%,这表明本文改进损失函数为 SIOU 方法对于票据检测精度有较好的提升作用;改进数据集处理方式后,使用 GIoU 损失函数时,整体的平均精度均值提高了 3.3%,使用 SIOU 损失函数时,整体的平均精度均值提高了 1.1%。这表明本文改进数据集预处理方法对于票据检测方法精度有较好的提升作用。

4 种不同方法在普通发票、定额发票、火车票上的可视化结果如图 5 所示。



图 5 4 种算法的可视化结果

#### 3.2 对比实验

为进一步评价本文方法的性能,选取当前市面上较为主流的 2 种目标检测算法 YOLOv3<sup>[14]</sup>

和 Faster-RCNN<sup>[15]</sup>进行对比实验,实验结果见表 2 所列,所有检测方法均在相同的训练集、验证集中训练和验证。

从表 2 可以看出,本文方法在同一验证集的验证下平均精度均值最高,为 99.20%,分别较 YOLOv3、Faster-RCNN 方法的检测平均精度均值高了 3.19%、2.44%,这是由于本文方法采用 CSPnet 结构,具有更好的网络特征融合能力,以及改进了损失函数之后,具有更高的准确性。而

在检测速度方面,本文方法的检测速度可以达到 51 帧/s,高于 YOLOv3 的 44 帧/s 以及 Faster-RCNN 的 28 帧/s。

由此可知,在 3 种实验方法中,本文的改进 YOLOv5 方法平均精度均值最大,检测速度最快,作为票据检测方法在性能上达到了预期。

表 2 对比实验结果

| 方法          | 检测速度<br>/(帧/s) | mAP/% | AP/%   |       |       |
|-------------|----------------|-------|--------|-------|-------|
|             |                |       | 火车票    | 普通发票  | 定额发票  |
| Faster-RCNN | 28             | 96.76 | 100.00 | 99.92 | 90.36 |
| YOLOv3      | 44             | 96.01 | 98.71  | 93.84 | 95.48 |
| 本文方法        | 51             | 99.20 | 99.50  | 99.20 | 99.00 |

综上所述,本文采用的改进 YOLOv5 检测方法在准确率上优于其与对比方法,检测速度优于其他检测方法,能保证大多数情况下的检测需求。由于实际应用中的票据存在折痕、污渍的特点,改进了数据集预处理方法,该方法比未经数据集预处理方法展示了更好的训练效果;同时,将原本的预测边界框的坐标损失函数 GIoU 改进为 SIoU 之后,各类别的平均精度值都有所提升。

#### 4 结 论

本文研究了票据目标检测问题,通过对报销单图像进行检测从而判断出该图像中票据的类型及其位置,为进一步信息提取奠定了基础。本文提出了一种基于改进 YOLOv5 的票据目标检测方法:将 YOLOv5 的损失函数改进为 SIoU,相比原 GIoU 损失函数多考虑了中心点距离、长宽比和角度;对于票据数据集提出一种新的预处理方法,有针对性地对票据数据集进行扩充和数据增强。而本实验研究的出发点是为了提供一种准确且高效的票据检测方法,来代替人工的低效率方式。实验结果表明,本文提出的方法基本满足了这一需求。

#### [参 考 文 献]

- [1] 杨佳丽. 智能识别模式在电子发票会计核算中的应用[J]. 财会通讯, 2021(9): 119-122.
- [2] 刘桂雄, 刘思洋, 吴俊芳, 等. 基于深度学习的机器视觉目标检测算法及在票据检测中应用[J]. 中国测试, 2019, 45(5): 1-9.
- [3] YOUNG P, SEUNG K, TUYEN P, et al. A high performance banknote recognition system based on a one-dimensional visible light line sensor[J]. Sensors, 2015, 15(6): 14093-14115.
- [4] LIU X W, LIU C Y. Paper currency CIS image fuzzy en-

hancement and boundary detection[J]. Applied Mechanics and Materials, 2014, 651: 356-2361.

- [5] 徐爱鹏, 朱杰, 陈叶明. 基于模糊发票 QR 二维码的识别技术研究[J]. 现代商贸工业, 2019(30): 212-213.
- [6] 成思齐, 宋晓茹, 高嵩, 等. 基于卷积神经网络进行发票分类[J]. 电子设计工程, 2020, 28(22): 170-174.
- [7] 王润. 基于深度学习的发票分类研究与智能财务报销系统的实现[D]. 南京: 南京邮电大学, 2021.
- [8] 吴帅, 徐勇, 赵东宁. 基于深度卷积网络的目标检测综述[J]. 模式识别与人工智能, 2018, 31(4): 335-346.
- [9] 吴加莹, 杨赛, 堵俊, 等. 自底向上的显著性目标检测研究综述[J]. 计算机科学, 2019, 46(3): 48-52.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S. l.]: IEEE, 2014: 580-587.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S. l.]: IEEE, 2016: 779-788.
- [12] ZHOU F, ZHAO H, ZHEN N. Safety helmet detection based on YOLOv5 [C]//2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA). [S. l.]: IEEE, 2021: 1-6.
- [13] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression [EB/OL]. (2022-05-25) [2022-12-01]. <https://arxiv.org/pdf/1902.09630.pdf>.
- [14] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. (2018-08-04) [2022-11-29]. <http://arxiv.org/pdf/1804.02767.pdf>.
- [15] 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, 2017, 39(6): 1137-1149.

(责任编辑 吴 亮)