

基于 CotNet 改进 YOLOv5 的接地线目标检测^①



黄 昊, 李海涛

(青岛科技大学 信息科学与技术学院, 青岛 266061)

通信作者: 黄 昊, E-mail: huanghao980101@163.com

摘 要: 在化工厂区的油罐车装卸区, 防止油罐车静电的产生和危害是避免加油车燃爆的重要手段, 通过静电接地线可以将油罐车感应的静电导走, 避免与外界物质产生跳火. 如何确保接地线在装卸车流程中正确安装不会被意外拆卸或提前拆卸是厂区急需解决的问题. 为确保在防爆区内用防爆摄像头的情况下能够对实时画面进行实时检测, 针对接地线连接角度不一, 拉伸后变细等特点提出将深度学习 YOLOv5 目标检测算法通过引入自注意力机制 CotNet 的方法. 在自制的接地线数据集上进行算法的检测速度和检测精度对比, 实验结果表明, 改进后的 YOLOv5 算法在速度稍有降低的情况下提高了 5% 的检测精度, 可以满足现场接地线检测需求.

关键词: 接地线目标检测; YOLOv5; CotNet; 自注意力机制; SIOU_Loss

引用格式: 黄昊, 李海涛. 基于 CotNet 改进 YOLOv5 的接地线目标检测. 计算机系统应用, 2023, 32(5): 283-290. <http://www.c-s-a.org.cn/1003-3254/9054.html>

CotNet-based Improved YOLOv5 for Grounding Line Target Detection

HUANG Hao, LI Hai-Tao

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: In the tanker loading and unloading area in a chemical plant area, preventing the generation and harm of static electricity in the tanker is an important means to avoid the combustion and explosion of the tanker. The static electricity induced by the tanker can be conducted away by the electrostatic grounding line to avoid sparkover with external substances. How to ensure that the grounding line is correctly installed during the loading and unloading process and will not be accidentally disassembled or disassembled in advance is an urgent problem to be solved in a plant area. To ensure that real-time images can be detected in real time when explosion-proof cameras are used in the explosion-proof area, this study gives due consideration to the characteristics, including different connection angles and thinning under stretching, of grounding lines and proposes a deep learning you only look once version 5 (YOLOv5) target detection algorithm by introducing the self-attention mechanism CotNet. The detection speed and detection accuracy of the proposed algorithm are compared on a self-made grounding line dataset. The experimental results show that the improved YOLOv5 algorithm, increasing the detection accuracy by 5% at the cost of a slight decrease in speed, can meet the needs of on-site grounding line detection.

Key words: grounding line target detection; YOLOv5; CotNet; self-attention mechanism; SIOU_Loss

石油和化学工业是我国重要的基础产业, 支柱产业, 化学品产值约占全球的 40%. 同时, 危险化学品领域

重特大事故多发, 安全生产仍处于爬坡过坎, 攻坚克难的关键时期^[1]. 在装卸区存在装卸过程中泄露的易燃, 易

① 基金项目: 农业部水产养殖数字建设试点项目 (2017-A2131-130209-K0104-004); 青岛市创新创业领军人才 (15-07-03-0030); 国家自然科学基金 (61806107)

收稿时间: 2022-09-29; 修改时间: 2022-10-27; 采用时间: 2022-11-08; csa 在线出版时间: 2023-02-17

CNKI 网络首发时间: 2023-02-17

爆的气体 and 液体, 当车辆产生的静电积累到一定值后, 就会放电引发火灾爆炸事故. 在厂区实际装卸作业中, 一是因为施工人员疏忽或偷懒不接接地线, 二是人工监管存在漏洞无法做到全程监管, 会出现接地线未正确安装的问题, 对此, 使用深度学习目标检测配合现场防爆摄像头做到接地线的实时监管, 以减少此类情况的发生.

目标检测指的是从复杂的图片背景中检索到感兴趣的目标并对定位的目标进行分类. 目标检测的发展也包含了基于传统方法和基于深度学习方法两类, 与行为识别一致. 传统的目标检测一般是通过窗口滑动的方式普遍扫描图像, 利用手工提取特征, 再将提取的特征送入 AdaBoost, 支持向量机等分类器中做分类判断. 但传统方法一般针对小数据集检测任务, 实际应用的泛化能力较弱, 而深度学习方法通过复杂的深度神经网络结构提取特征相较于传统方法能更好地学习和表达特征^[2]. 深度学习技术的成熟, 推动着目标检测技术的进步, 根据训练方式的不同又分为了双进程 (two-stage) 和单进程 (one-stage) 检测方法两类.

双进程检测方法包含了两个阶段, 一是对候选区域提议的预处理过程, 二是对提议的区域进行分类和回归并输出分类结果. 2014年 Girshick 等人^[3]提出了一种 R-CNN 网络代替传统目标检测模型, 自此开启了双阶段目标检测算法的研究道路, 衍生出许多经典的算法. Fast R-CNN^[4] 在整个图像上使用一次卷积过程, 输出感兴趣区域, 提高了 R-CNN 的运算速度, 采用 ROI 池化层对候选区域池化归一为同等大小的特征图送入完全连接网络中后利用 Softmax 进行分类. 丁志敏等人^[5]在原始 Faster R-CNN 方法的基础上将卷积网络的低层和高层特征图均送入 RPN 中实现多尺度目标检测, 并对非极大值抑制进行改进, 有效提升了输电线路挂接接地线的识别效果.

相较于双进程目标检测算法, 单进程目标检测算法的整体检测流程更具简洁性. 2015年 Redmon 等人提出 YOLOv1 算法^[6], 删减了双阶段目标检测法中提取候选区域环节, 图像输入后直接用一个深度神经网络完成提取特征, 回归候选框和分类, 相较于 Faster R-CNN, 其检测速度有大幅提升. YOLOv1 算法之后, 利用新的网络结构, 增加锚点, 跨尺度预测等调优技巧进行改进, 相继提出了 YOLOv2^[7], YOLOv3^[8]. 2020年 Bochkovskiy 等人^[9]利用感受野和参数更大的 CSP-Darknet53 作为主干网络提取特征, 并将 YOLOv3 中的 FPN 改为 PANet 完成多通道特征融合, 构建出

YOLOv4, 提高了小目标检测性能. YOLOv5^[10] 是 Ultralytics 团队在 2020年6月提出的一种典型的单进程目标检测算法, 该算法在性能上稍弱于 YOLOv4, 但其速度和灵活性远高于 YOLOv4, 从而非常适合厂区内的实时检测. 吴天舒等人^[11]首先通过设计轻量化目标检测算法完成驾驶员区域快速定位; 然后利用经过剪枝加速的语义分割模型对驾驶员区域进行分割, 得出安全带连通域; 最后通过判断安全带连通域面积检测驾驶员是否佩戴安全带. 郑福建等人^[12]提出对原生 YOLOv3 目标检测算法进行剪枝改进, 使得剪枝后的 CUT-YOLOv3 目标检测算法在与原生 YOLOv3 精度相当的情况下提升检测速度以识别安全带.

受化工厂区摄像头安装距离和防爆区防爆摄像头性能影响, 在图像中接地线部分在整个图像中所占比例大多不超过 10%. 另外, 车辆型号不同导致接地线连接位置不同, 所以接地线拉伸后呈现的状态也不同, 且拉伸越长接地线越细; 夜间接地线与背景颜色区别度不高, 都使接地线的检测难度大大增大. 因此, 本文通过引入自注意力机制 CotNet 改进 YOLOv5, 提高模型的识别精度, 最终满足实时精确检测接地线的要求.

1 算法原理介绍

1.1 YOLOv5 目标检测算法

YOLOv5 自发布以来一直在迭代升级版本, 本文采用的是 YOLOv5 的 6.0 版本. 其模型根据网络结构的深度和宽度分为 s、m、l、x 这 4 个版本, 其中 YOLOv5s 模型最小但检测精度最低, YOLOv5x 模型最大但检测精度最高. 本文基于 YOLOv5m 进行改进. 如图 1 所示 YOLOv5 的网络结构可以分为 4 个部分: 输入端, Backbone, Neck, 输出端^[13].

1.1.1 输入端

YOLOv5 的输入端采用了 Mosaic 数据增强, 随机从中筛选 4 张图片再进行随机缩放, 随机分布后进行拼接, 大大丰富了数据集. 通过随机缩放也可以得到很多小目标数据, 有助于提高网络的鲁棒性. 除此之外, 通过自适应锚框计算将计算初始锚框值的功能嵌入到代码中, 每次训练时, 自适应的计算不同训练集中的最佳锚框值. 通过自适应图片缩放减少计算量和信息冗余, 提高推理速度.

1.1.2 Backbone

本文使用的 YOLOv5 的 6.0 版本相较于之前版本, Backbone 中把原本的 Focus 模块替换为了一个使用起

来更加高效的 6×6 的卷积层. Backbone 中的 CSP 模块通过 Bottleneck 在不降低精度的前提下减少参数量, 优化计算量. 根据 Bottleneck 的不同, 含有 N 个 Bottleneck 的 CSP 可以分为 CSP1_N, CSP2_N. SPPF 模块通

过不同大小特征的融合, 提高了模型对大小不一检测目标的识别准确率, 且 SPPF 相较于 SPP, 只指定一个卷积核, 每次池化后的输出会成为下一个池化的输入, 速度提高了两倍.

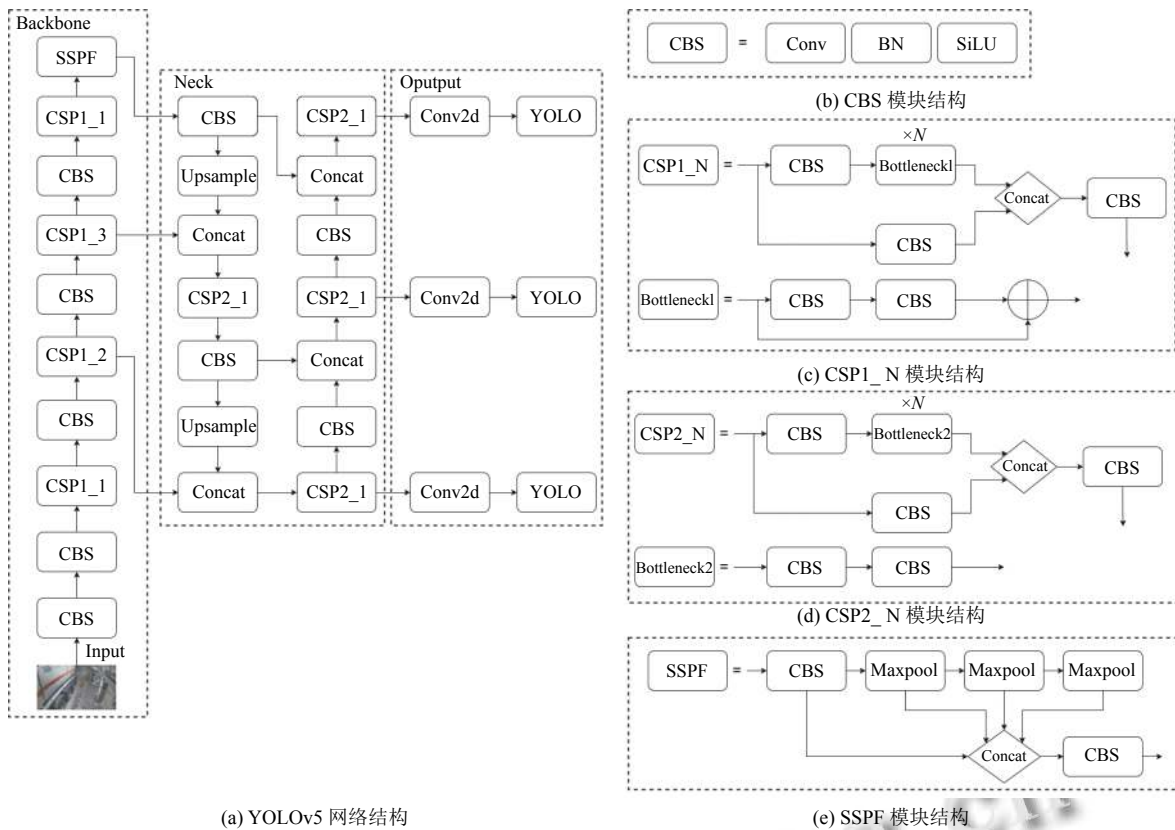


图 1 YOLOv5 网络结构

1.1.3 Neck

YOLOv5 的 Neck 使用了 FPN+PAN 结构, 该结构能够更好地处理深层特征图和浅层特征图的优缺点. FPN 能够把深层特征图携带的更强的语义特征传到浅层, 从而增强多个尺度上的语义表达, PAN 能够把浅层特征图携带的更强的定位信息传到深层, 从而增强多个尺度上的定位能力. FPN+PAN 的组合大大增强了网络不同特征层的融合能力.

1.1.4 输出端

YOLOv5 采用 CIOU_Loss 做边界框回归损失函数, 通过重叠面积, 中心点距离, 长宽比这 3 个几何因素计算真实标注框和模型预测框的吻合程度, 能够有效提高模型对重叠遮挡目标的检测效果. 且 YOLOv5 通过 NMS 非极大值抑制消除多余预测框.

2 算法改进

2.1 自注意力机制

2017 年, Transformer 算法被首先应用于自然语言处理领域 (natural language processing, NLP), 并且取得了一定的成功, 其模型的网络结果由自注意力机制组成, 不仅能够并行化训练, 而且能够包含全局信息^[14]. 目前应用于图像处理领域的自注意力机制是通过对 key, query, value 的计算来获取具有上下文信息的注意力矩阵, 自注意力机制中 key, query, value 可以理解为一个给定的 query, key 是其一个索引, 通过 query 与 key 的计算可以得出最合适的 value. 在计算机视觉中, 我们通过特征图信息与相应的参数矩阵相乘计算出 K, Q, V . 但是由于每个 key 都是单独和 query 进行计算, 从而忽视了相邻键之间的上下文信息. 因此, 传统自注

注意力机制可很好的获取图像全局信息,但卷积神经网络捕捉到的图像局部信息却被自注意力机制忽略^[15],本文使用的 CotNet 通过将两者捕捉图像全局信息和图像局部信息的能力结合起来,不再需要引入 position 模块,可以有效提高模型的特征表达能力。

2.2 CotNet 模块原理

如图 2 所示^[16],在 CotNet 中 key 采用了 $K \times K$ 的卷积操作进行编码以获取图像局部静态上下文信息 K^1 ,这里可以看作在局部信息上进行了静态的建模。然后将图像局部上下文信息与 Q 进行拼接 (concat) 处理,concat 操作在拼接时输入和输出的尺寸不变,但通道的数量进行连接。对于叠加后的结果再进行两次连续的卷积操作得到注意力矩阵。通过卷积提取局部信息的方法,充分提取了 key 内部的静态上下文信息,增强了自注意力机制。注意力矩阵的计算公式如下:

$$A = [K^1, Q]W_\theta W_\delta \quad (1)$$

其中, A 代表注意力矩阵, K^1 代表图像局部上下文信息, Q 代表 query, W_θ 和 W_δ 代表两次卷积。

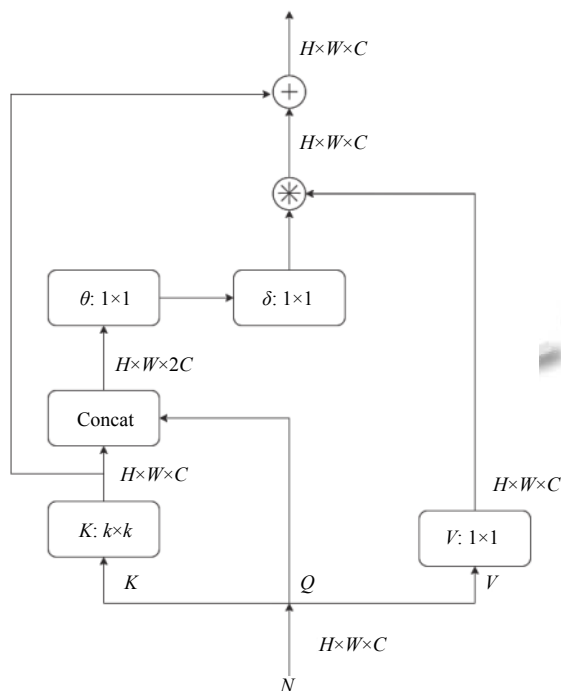


图 2 CotNet 模块结构

对于输入的特征图信息 N , value 定义为 $V = NW_V$, W_V 是 V 的权值矩阵,然后将注意力矩阵与 V 点乘得到全局动态上下文信息,计算公式如下:

$$K^2 = A \otimes V \quad (2)$$

其中, A 代表注意力矩阵, K^2 代表图像全局上下文信息, V 代表 value。

最后与图像局部上下文信息 K^1 相加输出 CotNet 的结果^[17],公式如下:

$$Y = K^1 + K^2 \quad (3)$$

其中, Y 为 CotNet 最终输出的结果, K^1 代表图像局部上下文信息, K^2 代表图像全局上下文信息。

CotNet 模块的算法如算法 1 所示。

算法 1. CotNet 模块的算法

- (1) 获取输入特征图 N 的 $[b, c, h, w]$, 即 batch, 通道数, 图像的高和宽。
- (2) 特征图 N 通过 $k \times k$ 的卷积得到局部静态上下文信息 K^1 。
- (3) 特征图 N 通过 1×1 的卷积后得到 v 。
- (4) 将 K^1 与特征图 N 通过 concat 拼接得到 y 。
- (5) y 经过连续两个 1×1 卷积得到注意力矩阵 A 。
- (6) 对 A 进行取平均降维后与 v 相乘, 得到全局动态上下文信息 K^2 。
- (7) K^1 与 K^2 相加后返回结果。

CotNet 模块通过对 key 进行卷积操作提取局部上下文信息来指导自注意力学习。如图 3 所示本文在 YOLOv5 的网络结构中引入 CotNet 自注意力机制代替 CSP1_N 部分结构,提高了算法对接地线的特征提取能力。

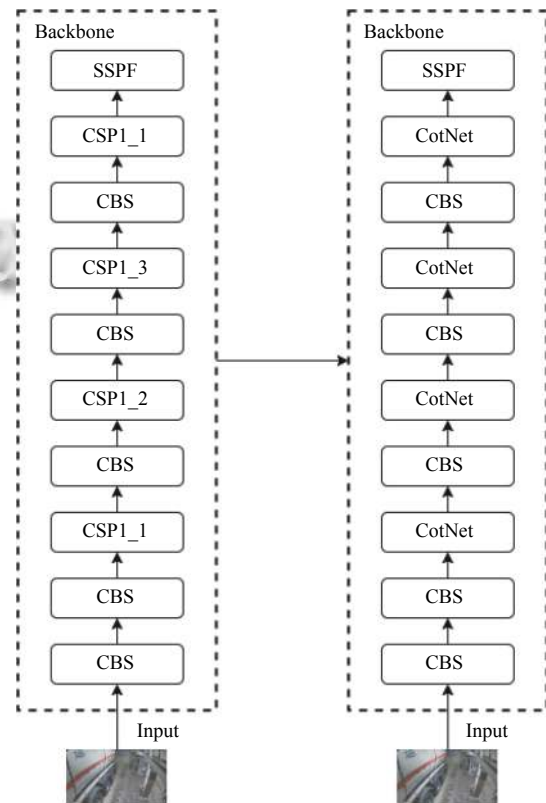


图 3 改进后的 YOLOv5 网络结构

2.3 输出端损失函数改为 SIOU_Loss

YOLOv5 原本使用的 CIOU_Loss, 在通过重叠面积, 中心点距离, 长宽比这 3 个几何因素计算真实标注框和模型预测框的吻合程度时, 没有考虑真实框和预测框之间的向量角度, 从而导致收敛速度慢且效率较低^[18]. SIOU_Loss 损失函数引入了真实标注框和模型预测框之间的向量角度, 并重新定义相关损失函数, 有效地提高了训练的速度和推理的准确性^[19].

SIOU_Loss 包含 4 个部分: 角度损失 (angle cost), 距离损失 (distance cost), 形状损失 (shape cost), IOU 损失 (IOU cost).

角度损失定义如下:

$$\begin{aligned} \Lambda &= 1 - 2 \times \sin^2 \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) \\ &= \cos \left(2 \times \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) \right) \end{aligned} \quad (4)$$

其中, Λ 为角度损失, c_h 为真实框和预测框中心点的高度差, σ 为真实框和预测框中心点的距离. 事实上 $\arcsin \left(\frac{c_h}{\sigma} \right)$ 就等于真实框中心点和预测框中心点连线与 x 轴的夹角 α . 在训练过程中, 当 $\alpha \leq \frac{\pi}{4}$ 时最小化 α , 否则最小化 β .

距离损失在定义时考虑了上面的角度损失, 定义如下:

$$\Delta = \sum_{(t=x,y)} (1 - e^{-\gamma \rho_t}) = 2 - e^{-\gamma \rho_x} - e^{-\gamma \rho_y} \quad (5)$$

其中,

$$\rho_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{c_w} \right), \rho_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{c_h} \right), \gamma = 2 - \Lambda \quad (6)$$

其中, Δ 指距离损失, c_w 和 c_h 分别指真实框和预测框最小外接矩形的宽和高, $(b_{cx}^{gt}, b_{cy}^{gt})$ 是真实框的中心点坐标, (b_{cx}, b_{cy}) 是预测框的中心点坐标. 可以看出随着夹角 α 接近 $\frac{\pi}{4}$, Δ 越来越大, 距离损失的影响也越来越大.

形状损失的定义如下:

$$\Omega = \sum (1 - e^{-w_t})^\theta = (1 - e^{-w_w})^\theta + (1 - e^{-w_h})^\theta \quad (7)$$

其中,

$$w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (8)$$

其中, Ω 指形状损失, (w, h) 指预测框的宽和高, (w^{gt}, h^{gt}) 指真实框的宽和高. θ 为对形状损失的关注度, 通

过遗传算法计算出当 θ 的值接近于 4 的时候, 可以最大程度地避免因为过于关注形状损失而降低对预测框的移动, 因此令 θ 属于 [2, 6].

IOU 损失的定义为:

$$IOU = \frac{A}{B} \quad (9)$$

其中, IOU 为 IOU 损失, A 为真实框与预测框的交集, B 为真实框与预测框的并集. 最后结合上述, SIOU_Loss 损失函数的定义为^[20]:

$$L_{\text{box}} = 1 - IOU + \frac{\Delta + \Omega}{2} \quad (10)$$

3 试验结果与分析

3.1 试验数据集

本文试验所用的数据集为化工厂中装卸车区域的防爆摄像头拍摄的数据. 在数据采集期间, 每辆作业车辆以 10 s 为间隔采集 20 张图片, 因接地线连接位置导致接地线拉长的场景采集 30 张图片, 夜间作业车辆采集 30 张图片, 共采集 2000 张图片. 受防爆摄像头性能限制, 采集的图片尺寸为 1920×1080. 对数据标注时为防止因为部分遮挡就完全无法识别, 对接地线进行分段标注, 且尽量减少标注框内背景占比, 降低背景对特征提取的影响. 根据 8:1:1 的比例随机分为训练集, 验证集, 测试集, 最终得到训练集 1600 张, 验证集 200 张, 测试集 200 张.

3.2 试验环境

本次试验是在 Windows 操作系统下, 基于 GPU, PyTorch 和 CUDA 框架完成的. 试验运行环境的 CPU 为 Intel(R) Core(TM) i7-9750H CPU @ 2.60 GHz, 显卡为英伟达 RTX2070, 显存为 8 GB, 运行内存为 16 GB.

3.3 评估指标

本研究主要采用目标检测算法中的精确度 (precision, P), 召回率 (recall, R), 平均精确度 (average precision, AP), 平均精度均值 (mean average precision, m_{AP}) 和单幅图像检测耗时作为评估指标, 其中平均精度均值 (m_{AP}) 和单幅图像检测耗时是主要评估指标, 可以有效评估算法的精度和速度.

精确度的计算如式 (11) 所示, 召回率的计算公式如式 (12) 所示:

$$P = \frac{TP}{TP + FP} \quad (11)$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

其中, TP 是接地线识别正确的样本数量, FP 是把非接地线识别成接地线的样本数量, P 代表接地线目标检测的精确度, FN 是把接地线识别成非接地线的样本数量, R 代表接地线目标检测的召回率。

平均精确度的计算公式如式 (13) 所示, m_{AP} 是对平均精确度在所有类别下取其均值, 计算公式如式 (14)。

$$X_{AP} = \int_0^1 P(R) \tag{13}$$

$$m_{AP} = \frac{\sum_{i=1}^K X_{APi}}{K} \tag{14}$$

其中, X_{AP} 代表平均精确度, X_{APi} 代表第 i 类目标检测的平均精确度, K 代表标记的类别, m_{AP} 代表平均精确度的均值。

$m_{AP@0.5}$ 的公式如式 (15) 所示, 其定义为在交并比阈值 IOU 为 0.5 的情况下的平均精确度均值。

$$m_{AP@0.5} = \frac{\sum_{i=1}^K X_{AP@0.5i}}{K} \tag{15}$$

其中, $X_{AP@0.5i}$ 代表第 i 类目标在交并比阈值为 0.5 时的平均精确度, K 代表标记的类别, $m_{AP@0.5}$ 代表交并比阈值 IOU 为 0.5 时的平均精确度均值。

3.4 模型训练

在网络模型训练阶段, 将迭代次数设置为 100 次, 将每批次训练数据量设置为 16, 将权重衰减系数设置为 0.0005, 将动量因子设置为 0.8, 学习速率决定了模型参数的更新步伐, 学习率过大可能导致网络无法收敛, 学习率过小则可能收敛过慢, 将优化器设置为 Adam, Adam 可以自动调整学习率, 将初始学习率设置为 0.01, 将置信度阈值设置为 0.5, 如表 1 所示。改进后模型的训练损失变化和 YOLOv5 的训练损失变化如图 4 所示。

表 1 训练参数

参数	数值
迭代次数	100
批次训练数据量	16
权重衰减系数	0.0005
动量因子	0.8
优化器	Adam
初始学习率	0.01
置信度阈值	0.5

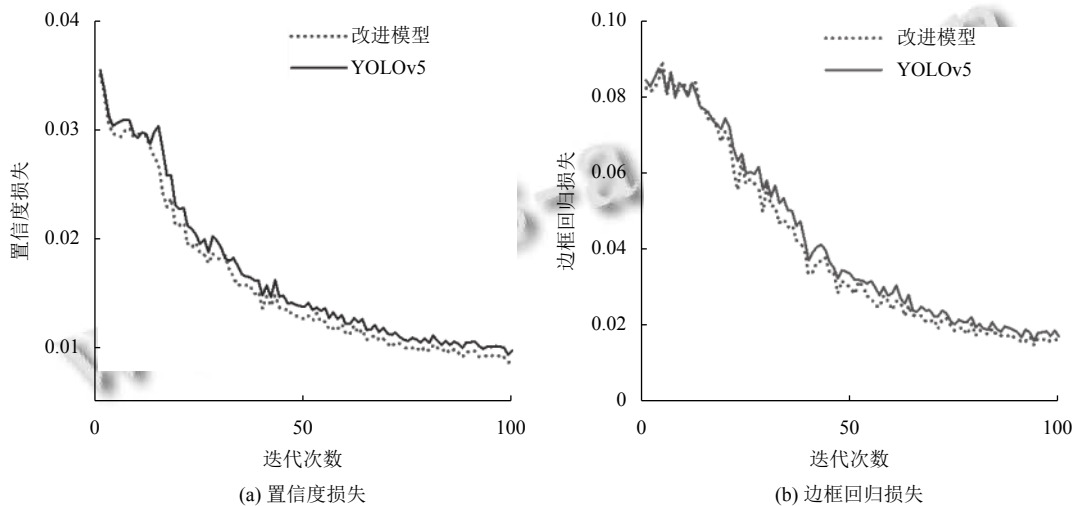


图 4 改进后模型与 YOLOv5 模型训练收敛情况

3.5 改进后模型的性能分析与消融实验

经过改进后算法对接地线的检测精确度为 94.1%, 召回率为 97.8%, $m_{AP@0.5}$ 为 97.6%, 单幅图像检测耗时为 20.2 ms。

试验结果证明, 改进后的模型对接地线取得了较高的识别精度, 且检测的耗时也属于可接受范围。为了进一步验证改进后算法相较原算法的有效性, 本文以 YOLOv5 算法为基础, 分别增加本文的两种改进方法,

再同时增加两种改进方法,共4组网络用同样的数据设计了消融实验.实验结果如表2所示.

模型	消融实验			
	精确度 (%)	召回率 (%)	$m_{AP@0.5}$ (%)	单幅图像检测耗时 (ms)
YOLOv5	91.2	97.3	92.5	19.1
YOLOv5+CotNet	94.3	97.5	97.2	23.4
YOLOv5+SIOU	94.1	96.7	96.4	20.1
改进算法	94.1	97.8	97.6	21.2
改进后与YOLOv5性能差异	+2.9	+0.5	+5.1	+3.1

由表2数据可知,在YOLOv5网络骨干中融合自注意力CotNet模块后 $m_{AP@0.5}$ 比YOLOv5提高了4.7%,单幅图像检测耗时增加了4.3 ms;将YOLOv5的回归损失函数替换为SIOU_Loss后 $m_{AP@0.5}$ 比YOLOv5提高了2.9%,单幅图像检测耗时增加了2 ms.实验结果表明,本文的两种改进方法相比YOLOv5都有不同程度的提升,其中添加CotNet模块检测精度提升效果

最好,但检测耗时增加的多.通过消融实验,证明了本文提出的两种改进方法的有效性.且本文最终改进后的算法比YOLOv5的 $m_{AP@0.5}$ 提高了5.1%,精确度提高了2.9%,召回率提高了0.5%,检测精度得到了较大的提高,虽然单幅图像检测耗时增加了3.1 ms,但仍满足厂区摄像头实时监测的要求.

为了进一步直观地验证改进算法能够更有效地识别不同环境和不同连接方式下的接地线,现选取一部分接地线图片作为验证集,分别验证原算法和改进后算法,验证结果如图5所示.根据Image1可以看出,改进后算法能够更好地识别接地线且识别精度有所提升.根据Image2可以看出,改进后算法在夜间相比原算法能够更好地识别接地线且识别精度有较大提升.根据Image3可以看出,改进后算法在识别拉长后的接地线时能更好地识别接地线且识别精度有较大提升.通过以上结果可知,改进后算法不但提高了接地线的识别精度,而且在原算法难以有效识别的场景下可以有效提高识别效果.

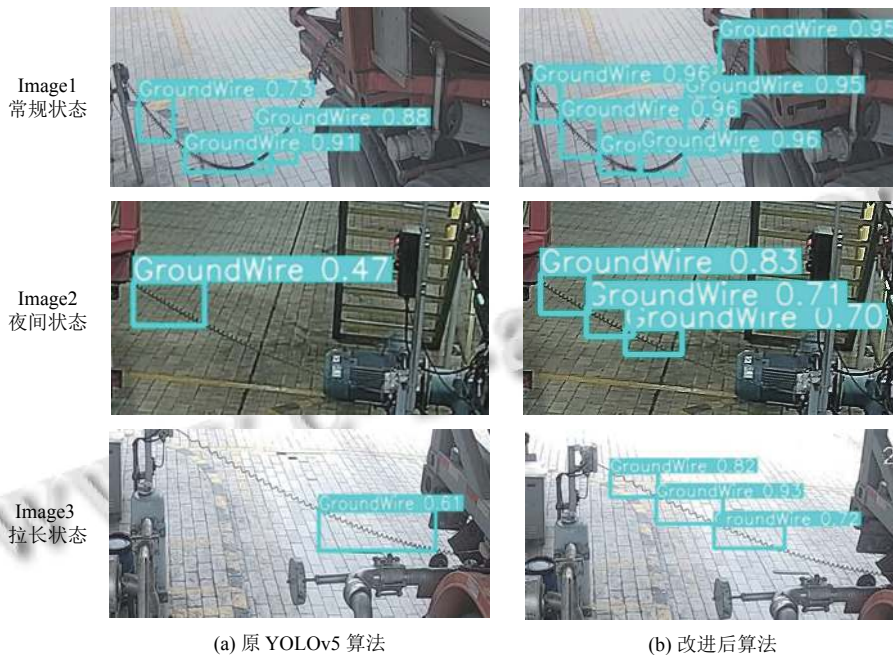


图5 验证结果

4 结论

本文提出的改进算法是在YOLOv5的基础上,首先在网络结构中引入自注意力机制CotNet模块,有效提高算法特征提取能力;再修改回归损失函数为SIOU_Loss,提高了训练的速度和推理的准确性.研究结果表

明,改进后的算法的平均精度均值达到了97.6%,单幅图像检测耗时21.2 ms,可以满足厂区装卸车接地线检测的精确度要求和实时性要求.对于拉长后的接地线和夜晚环境下的接地线也都有效提高了识别效果.未来将继续优化接地线的识别效果,考虑引入图像分

割结合目标检测提高识别效果,为化工厂区的安全监管保驾护航。

参考文献

- 1 黄志刚. 5G+工业互联网在危险化学品企业安全生产管理中的应用. 新型工业化, 2021, 11(7): 9-10. [doi: [10.19335/j.cnki.2095-6649.2021.7.004](https://doi.org/10.19335/j.cnki.2095-6649.2021.7.004)]
- 2 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述. 数据采集与处理, 2016, 31(1): 1-17. [doi: [10.16337/j.1004-9037.2016.01.001](https://doi.org/10.16337/j.1004-9037.2016.01.001)]
- 3 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580-587.
- 4 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
- 5 丁志敏, 邢晓敏, 董行, 等. 基于深度学习的输电线挂接接地线状态目标检测. 电子测量技术, 2021, 44(3): 132-137. [doi: [10.19651/j.cnki.emt.2005474](https://doi.org/10.19651/j.cnki.emt.2005474)]
- 6 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Vegas: IEEE, 2016. 779-788. [doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91)]
- 7 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 6517-6525. [doi: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690)]
- 8 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- 9 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.
- 10 Zhu XK, Lyu S, Wang X, *et al.* TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal: IEEE, 2021. 2778-2888.
- 11 吴天舒, 张志佳, 刘云鹏, 等. 结合 YOLO 检测和语义分割的驾驶员安全带检测. 计算机辅助设计与图形学学报, 2019, 31(1): 126-131.
- 12 郑福建, 汪彦, 左应祥, 等. 基于剪枝 YOLOv3 算法的高处作业人员安全带佩戴检测研究. 信息系统工程, 2021, (10): 117-119. [doi: [10.3969/j.issn.1001-2362.2021.10.037](https://doi.org/10.3969/j.issn.1001-2362.2021.10.037)]
- 13 王鹏飞, 黄汉明, 王梦琪. 改进 YOLOv5 的复杂道路目标检测算法. 计算机工程与应用, 2022, 58(17): 81-92. [doi: [10.3778/j.issn.1002-8331.2205-0158](https://doi.org/10.3778/j.issn.1002-8331.2205-0158)]
- 14 孔雅洁, 张叶. 引入高斯掩码自注意力模块的 YOLOv3 目标检测方法. 液晶与显示, 2022, 37(4): 539-548. [doi: [10.37188/CJLCD.2021-0250](https://doi.org/10.37188/CJLCD.2021-0250)]
- 15 王全增. 基于深度学习的花粉自动检测算法研究 [硕士学位论文]. 北京: 北京工业大学, 2020. [doi: [10.26935/d.cnki.gbjgu.2020.000682](https://doi.org/10.26935/d.cnki.gbjgu.2020.000682)]
- 16 Li YH, Yao T, Pan YW, *et al.* Contextual transformer networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, [doi: [10.1109/TPAMI.2022.3164083](https://doi.org/10.1109/TPAMI.2022.3164083)]
- 17 杨鹤, 柏正尧. CoT-TransUNet: 轻量化的上下文 Transformer 医学图像分割网络. 计算机工程与应用, 2023, 59(3): 218-225.
- 18 徐长凯. 基于优化边界框回归的目标检测研究 [硕士学位论文]. 青岛: 青岛科技大学, 2021. [doi: [10.27264/d.cnki.gqdhc.2021.000577](https://doi.org/10.27264/d.cnki.gqdhc.2021.000577)]
- 19 Gevorgyan Z. SIoU Loss: More powerful learning for bounding box regression. arXiv:2205.12740, 2022.
- 20 李小军, 邓月明, 陈正浩, 等. 改进 YOLOv5 的机场跑道异物目标检测算法. 计算机工程与应用, 2023, 59(2): 202-211.

(校对责编: 牛欣悦)