

〈图像处理与仿真〉

基于改进 YOLOv7-Tiny 的变电设备红外图像识别

邓长征¹, 刘明泽¹, 付 添², 弓萌庆¹, 骆冰洁¹

(1. 三峡大学 电气与新能源学院, 湖北 宜昌 443002; 2. 湖北交投科技发展有限公司, 湖北 武汉 430000)

摘要: 针对复杂背景下变电设备红外图像目标识别精度不高、识别速度慢的问题, 本文提出一种基于改进 YOLOv7-Tiny 的变电设备红外图像识别算法。首先, 引入轻量级瓶颈结构 GhostNetV2 BottleNeck 替换部分 CBS 模块构建轻量级高效聚合网络 L-ELAN (Lightweight-Efficient Layer Aggregation Network), 同时在特征提取阶段嵌入 CA (Coordinate Attention) 注意力机制, 在降低网络参数数量的同时加强网络对目标关键特征的提取, 提升检测精度; 将网络坐标损失函数替换为 SIoU Loss, 以提升锚框定位精度和网络收敛速度; 在变电设备红外数据集上进行测试, 结果表明, 改进后网络的精确率达到 96.28%, 检测速率达到 26.42 frame/s, 模型大小降低至 7.82 M。与 YOLOv7-Tiny 原算法相比较, 本文算法在提升识别精度的同时将检测速率提升 21.69%, 模型大小减少 36.89%, 可以满足变电站设备的精准实时识别要求, 为后续的变电站设备故障诊断奠定基础。

关键词: 变电设备; 红外图像识别; YOLOv7-Tiny; 注意力机制; 轻量化

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1001-8891(2025)01-0044-08

Infrared Image Recognition of Substation Equipment Based on Improved YOLOv7-Tiny Algorithm

DENG Changzheng¹, LIU Mingze¹, FU Tian², GONG Mengqing¹, LUO Bingjie¹

(1. College of Electrical and New Energy, China Three Gorges University, Yichang 443002, China;

2. Hubei Communications Investment Technology Development Co., LTD., Wuhan 430000, China)

Abstract: To address the problem of low accuracy and slow recognition speed of infrared (IR) image target recognition of substation equipment in complex backgrounds, this study proposes an IR image recognition algorithm for substation equipment based on the improved YOLOv7-Tiny. First, the lightweight bottleneck structure GhostNetV2 bottleneck is introduced to replace a part of the CBS module and build a lightweight and efficient aggregation network known as a lightweight-efficient layer aggregation network. Simultaneously, a coordinate attention mechanism is embedded in the feature extraction stage to reduce the number of network parameters while strengthening the network's extraction of key features of the target and improving detection accuracy. The network coordinate loss function is replaced by SIoU_Loss to improve the anchor frame positioning accuracy and network convergence speed. The results show that the accuracy of the improved network is 96.28%, the detection rate is 26.42 frames/s, and the model size is reduced to 7.82 M. Compared with the original YOLOv7-Tiny algorithm, the detection rate is increased by 21.69%, the identification accuracy is improved, and the model size is reduced by 36.89%. These results meet the requirements of accurate real-time identification of substation equipment and lay a foundation for subsequent substation equipment fault diagnosis.

Key words: substation equipment, infrared image recognition, Yolov7-tiny, attention mechanism, lightweight

0 引言

变电站作为电力系统中重要的组成部分,对变电设备进行实时监测对于保障变电站的安全稳定运行至关重要。通过实时检测,可以对变电设备的运行状态进行实时监测和分析,及时发现设备的故障隐患,有利于提高变电站的运行稳定性并降低其运维成本^[1]。因此,快速且准确地识别变电设备类型对智能电网建设起着至关重要的作用^[2]。

近年来,红外检测因其高准确性、无接触测温 and 不停电作业等优点而被广泛应用于变电设备监测中。变电设备在故障的初期往往伴有局部异常发热点的现象。红外检测正利用这种特点获取设备温度信息进行分析识别,为后续电力设备故障诊断提供依据。变电站中拍摄到的变电设备红外图像有以下特点:

- 1) 图像中目标尺度变化较大,导致远处的目标在红外图像上包含的特征信息有限;
- 2) 待检测目标设备的背景可能较为复杂,容易被遮挡。

使用深度学习的方法对变电设备红外图像进行识别可以极大提高后续的设备故障诊断效率。文献[3]在无锚点算法 CenterNet^[4]的基础上,引入高斯核函数,精简网络结构的同时提高变电站目标检测方法的识别定位精度。文献[5]构建以深度可分离卷积为基本特征提取单元的轻量级特征提取结构,分别提取可见光和红外图像中的特征信息,再通过双支路特征交叉融合结构对特征进行融合互补,在保障精度与效率的前提下实现对不同尺度目标的高精度识别与定位。文献[6]在YOLOv5^[7]目标检测算法的基础上,在骨干网络中引入 Ghost 卷积^[8]替换标准卷积,实现网络轻量化。并通过添加 ECA (efficient channel attention)^[9]注意力模块与改进 NMS (non-maximum suppression) 和损失函数的方法增强特征捕捉能力,提高网络精度,加速网络收敛。文献[10]提出了一种基于自适应残差网络的诊断模型。首先建立包含大量变电站红外图像的多分类数据集,再优化残差网络 (residual network, Resnet) 的基础架构,提升模型的识别准确率。并在此基础上对该模型利用改进贝叶斯优化算法获取超参数的最优解,在提高准确率的同时获取性能最佳的诊断模型。

1 改进YOLOv7-Tiny网络框架

本文提出一种基于改进YOLOv7-Tiny^[11]的变电设备红外图像识别算法,能够更准确快速地对变电设备红外图像进行识别。首先,在特征提取模块中引入

轻量级瓶颈结构 GhostNetV2 BottleNeck^[12],在保证原有精度的同时建立轻量化目标检测网络,以达到对变电设备红外图像进行实时监测的目的。此外,针对变电设备红外图像中远距离小尺度目标特征信息较少的问题,在特征提取模块中引入 CA^[13]注意力机制,在保留图像中的位置信息的同时充分捕获图像微小特征信息,增强网络对目标的精准定位能力,在提高精度的同时,以更少的计算量捕获更多的特征。最后,将坐标损失函数改为 SIoU Loss^[14-15]以加速函数收敛并提升锚框定位精度,在一定程度上解决了错误检测、多次检测和漏检的问题。

本文网络分为 Input、Backbone、Neck 和 Head 四个部分,网络框架如图1所示。Input 主要实现对输入图片的预处理,经过 Mosaic 数据增强,自适应锚框计算和自适应图片放缩等方法使图像满足特征提取网络需求。Backbone 即特征提取网络,由 CBS 模块、L-ELAN 模块、CA 注意力机制模块与 MP 模块组成,CBS 模块提取图像原始特征,L-ELAN 模块学习原始特征,再通过 CA 模块提取关键信息,MP 模块将不同特征拼接到一起,共同实现图像目标特征提取;Neck 即特征融合网络,采用 SPPCSPC 模块与 PANet 结构融合各层特征以检测不同尺度的目标。其中,SPPCSPC 模块由金字塔池化结构 (Spatial pyramid pooling, SPP) 和跨阶段区域网络 (Cross stage partial, CSP) 组成。CSP 将特征图分成两个部分,其中一个特征经过 SPP 池化操作后与另一个特征进行融合,以获取不同尺寸的感受野。Head 即预测网络,部分使用 Rep 层与 CBM 模块,Rep 层的作用是调整通道数,CBM 模块由 Conv 层、BN 层与 Sigmoid 函数构成。在 Head 层阶段,将输出图片通过 3 个 Rep 层以及 CBM 层输出 3 个不同尺度预测结果。

1.1 GhostNetV2 BottleNeck 模块

为将模型搭载至边缘设备以达到实时识别的目的,本文引入轻量级瓶颈结构 GhostNetV2 BottleNeck 替换 ELAN 中部分 CBS 模块构建轻量级高效聚合网络 L-ELAN。GhostNetV2 BottleNeck 主要由 Ghost 模块和解耦全连接注意力机制 (decoupled fully connected attention, DFC) 模块共同构成,在降低参数数量的同时捕捉长距离空间位置中像素之间的相关性。GhostNetV2 Bottleneck 结构如图2所示。

Ghost 模块将普通卷积分解成两个步骤,首先是严格控制数量的传统卷积计算,然后进行一系列简单的线性变换,生成原有特征图的 Ghost 特征图,最后将其拼接到一起生成最终的特征图,以此消除特征图冗余。Ghost 模块如图3所示。

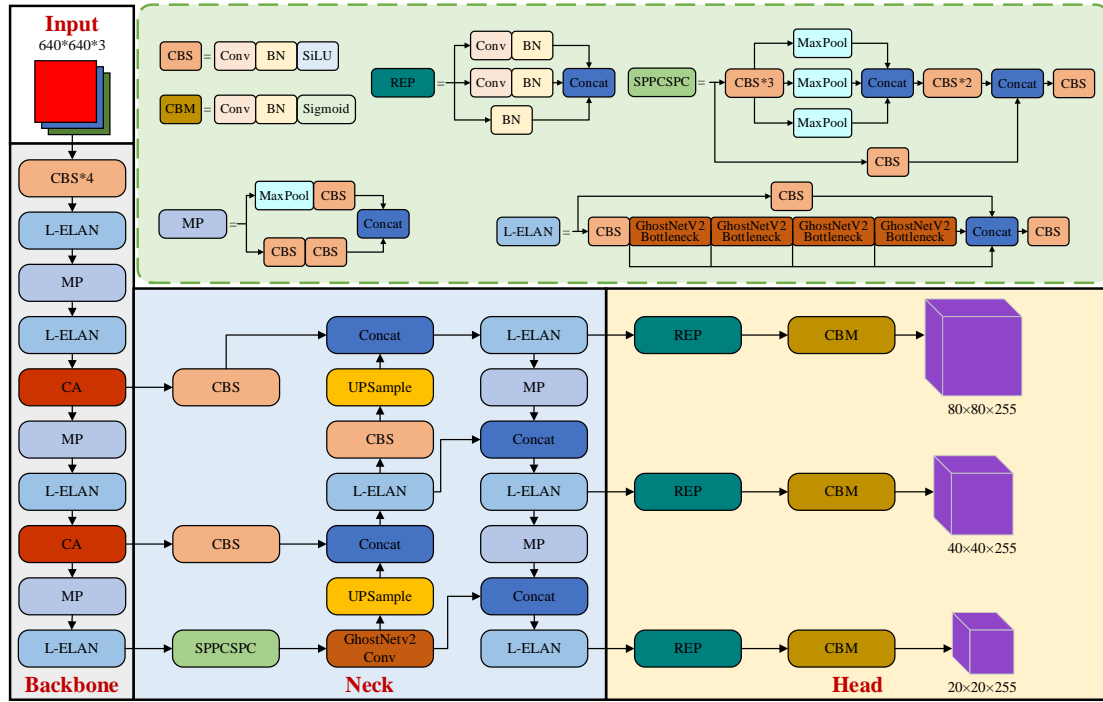
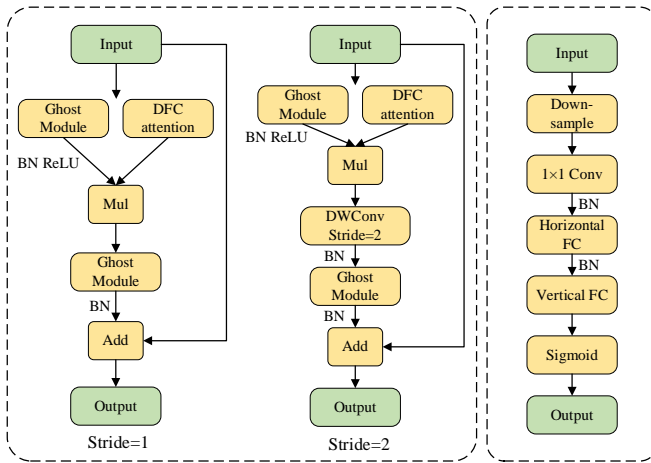


图 1 本文改进后的算法结构

Fig.1 The improved algorithm structure of this article



(a) GhostNetV2 Bottleneck (b) DFC attention

图 2 GhostNetV2 Bottleneck 结构

Fig.2 The structure of GhostNetV2 Bottleneck

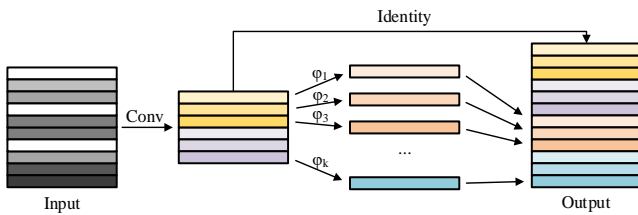


图 3 Ghost 模块示意图

Fig.3 Schematic diagram of Ghost module

假设输入特征图的尺寸为 $h \times w \times c$ ，其中通道数为 c ，输入特征图的高和宽分别为 h 和 w ，输出特征

图的尺寸为 $h' \times w' \times n$ ，输出特征图的通道数为 n ，卷积核大小为 $k \times k$ ，线性运算核大小为 $d \times d$ ，线性运算次数为 s ，且 s 远小于 c ，则：

采用普通卷积的计算量为：

$$\text{FLOPS} = n \times h' \times w' \times c \times k \times k \quad (1)$$

Ghost 模块的计算量为：

$$\text{FLOPS} = \frac{n}{s} \times h' \times w' \times c \times k \times k \times (c + s - 1) \quad (2)$$

相比于普通卷积，Ghost 模块的压缩比为：

$$r_c = \frac{n \times h' \times w' \times c \times k \times k}{\frac{n}{s} \times h' \times w' \times c \times k \times k \times (c + s - 1)} = \frac{s \times c}{c + s - 1} \approx s \quad (3)$$

由公式(3)知，在其他条件相同的情况下，Ghost 模块的计算量是传统卷积计算量的 $1/s$ ，在网络中使用 Ghost 模块替换标准卷积能大大减少模型的计算成本。

DFC 注意力机制基于完全连接层 (fully connected layer)，既能在普通硬件上快速执行，又能捕获远距离像素间的依赖关系。它将完全连接层分解为水平 FC 和垂直 FC 两个方向，这两个方向的完全连接层沿各自的方向建模长距离的空间信息，结合到一起就可以得到全局的感受野。加入 DFC 模块之后，GhostNetV2 Bottleneck 可以在降低参数数量的同时，在精度和速度之间实现更好的权衡。DFC 注意力机制的原理如图 4 所示。

1.2 CA 注意力机制

针对远距离变电设备特征信息较少，识别难度大的问题，本文在特征提取阶段引入轻量级注意力机制 CA，在不额外增加模型计算量的同时充分捕获图像微小特征信息。

CA 注意力机制将通道注意力分解为两个一维特征编码过程，分别沿两个空间方向进行特征聚合。通过这种方式，CA 注意力机制可以在保留精确的位置信息的同时捕获远程依赖关系。相比于其他注意力机制，CA 模块的优势在于足够灵活轻量，能够简单地插入移动网络的核心结构中。

CA 注意力机制结构如图 5 所示，其实现过程为：

1) 首先用尺寸为 $(H, 1)$ 和 $(1, W)$ 的一维全局池化核沿水平和垂直两个方向对每个通道进行编码，聚合为两个单独的方向注意力特征图。高度为 h 的第 c 个通道的输出如式(4)所示：

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (4)$$

宽度为 w 的第 c 个通道的输出如式(5)所示：

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (5)$$

式中： H 和 W 分别为输入特征的高度和宽度； $x_c(h, i)$ 为第 i 行特征向量； $x_c(j, w)$ 为第 j 列特征向量。

2) 然后将获得全局感受野的两个方向的特征图进行拼接，使用 1×1 的卷积进行 F_1 变换操作，输入非线性激活函数得到对空间信息在水平方向和垂直方向的中间特征图 f ，其尺寸为 $1 \times (W+H) \times C/r$ ，如

公式(6)所示：

$$f = \delta(F_1([z^h, z^w])) \quad (6)$$

式中： δ 为非线性激活函数； F_1 为 1×1 卷积变换。

3) 然后将中间特征图 f 划分为两个单独张量，按照原来的高度和宽度进行 1×1 的卷积得到与输入特征图相同通道数的特征图 f^h 和 f^w ，经过 sigmoid 激活函数后分别得到特征图在高度和宽度上的注意力权重 g^h 和 g^w ，如公式(7)、(8)所示：

$$g^h = \sigma(F_h(f^h)) \quad (7)$$

$$g^w = \sigma(F_w(f^w)) \quad (8)$$

式中： σ 为 Sigmoid 激活函数； F_h 、 F_w 为 1×1 卷积变换。

4) 最后在原始特征图上通过乘法加权计算，得到带有注意力权重的输出特征图 $Y = [y_1, y_2, \dots, y_c]$ 。如公式(9)所示：

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (9)$$

式中： $x_c(i, j)$ 为输入特征图； $g_c^h(i)$ 、 $g_c^w(j)$ 分别为得到的水平和竖直方向注意力图。

1.3 损失函数改进

YOLOv7-Tiny 的损失函数包括坐标损失 (box_loss)、置信度损失 (obj_loss)、分类损失 (cls_loss) 三部分。为达到提高模型检测精度的目的，本文将 YOLOv7-Tiny 的坐标损失函数改为 SIOU Loss 以加速函数收敛并提升锚框定位精度，在一定程度上解决了错误检测、多次检测和漏检的问题。

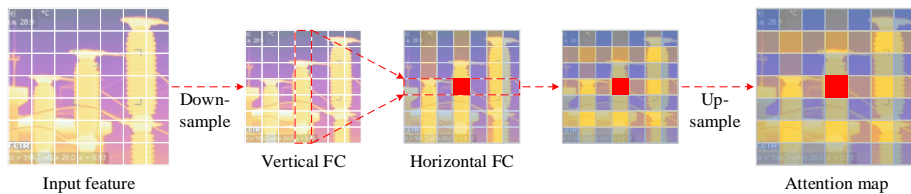


图4 DFC 注意力机制原理

Fig.4 The principle of decoupled fully connected attention

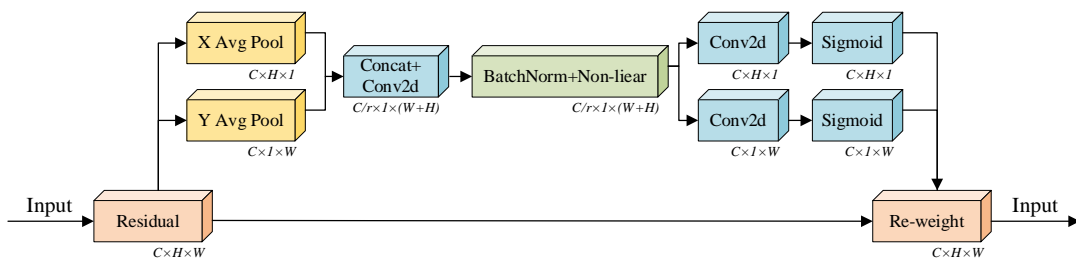


图5 CA 注意力机制结构原理

Fig.5 Structural principle of CA attention mechanism

传统的坐标损失函数依赖于边界框回归的聚合，而忽略了期望的真实框和预测框之间方向不匹配的问题，这个缺陷可能会导致在训练过程中预测框无法快速定位的问题，从而导致收敛速度较慢且效率较低。相比之下，SIoU Loss 除考虑到预测框之间的距离 σ 、重叠面积和宽高比之外，还进一步考虑了真实框 B^{GT} 和预测框 B 之间的向量角度 α 、 β ，加入了角度惩罚项，重新定义相关损失函数。SIoU Loss 示意图如图 6 所示。

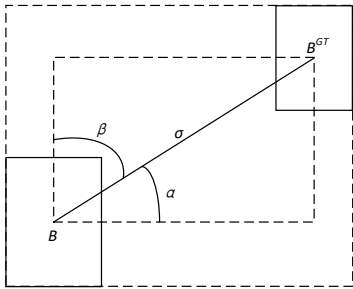


图 6 SIoU Loss 示意图
Fig.6 Schematic diagram of SIoU loss function

2 实验结果及分析

本实验在 Pytorch 1.12.0 深度学习框架下进行，软件和硬件平台配置参数如表 1 所示。

表 1 软件和硬件平台配置参数

Table 1 Software and hardware platform configuration parameters

Operating system	Windows 10,16 G, 64 bit
GPU	NVIDIA GeForce RTX 2070 Max-Q
CPU	Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz
CUDA	10.2
OpenCV	4.7.0

2.1 数据采集与处理

本文所采用数据集来自某国网公司提供的变电站红外巡检图像，从中选取具有代表性的变电设备包括电压互感器、高压套管、悬式绝缘子、避雷器 4 种电气设备共 1000 幅，数据集构成如表 2 所示。为提升模型的泛化能力，将数据集分别进行随机旋转、随机平移、随机缩放等数据扩充方法，使样本数量达到 5000 幅，实验图像尺寸统一为 640×512 ，使用 labeling 软件对数据集进行标记。将训练集、验证集、测试集按 6:2:2 的比例进行划分。

2.2 评价指标

为客观评价本文算法对于变电站设备红外图像的识别效果，采用精确率 (precision, P)、召回率 (recall, R)、平均精度均值 (mean Average Precesion, mAP)、

每秒检测帧数 (Frames Per Second, FPS)、模型大小 (Params) 作为评价指标。本文以 IoU 阈值取 0.5 时的 mAP 值为标准。

表 2 数据集构成

Table 2 Composition of the dataset

Target device	Quantity	Proportion/%
Potential transformer	146	14.6
High pressure bushing	437	43.7
Suspension insulator	291	29.1
Lightning protector	126	12.6

精确率 P 定义为所有检测到的目标的正确检测率，公式如(10)所示：

$$P = \frac{TP}{TP+FP}$$
 (10)

式中：TP 表示正确识别出的目标数量；FP 表示目标误检和漏检的数量。

召回率 R 定义为所有检测目标的正确检出率，公式如(11)所示：

$$R = \frac{TP}{TP+FN}$$
 (11)

式中：FN 表示目标误检的数量。

平均精度均值 mAP 定义为各类别 AP 的平均值， N 为目标类型的个数，公式如(12)所示：

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$
 (12)

式中：平均准确率 AP 为单一类别的识别平均准确率，是准确率与召回率的积分，公式如(13)所示：

$$AP = \int_0^1 P(R) dR$$
 (13)

每秒检测帧数 FPS 反映模型的推理速度，如式(14)所示：

$$FPS = \frac{frameNum}{elapsedTime}$$
 (14)

式中：frameNum 表示检测图片总数；elapsedTime 表示检测所花费的总时间。

2.3 网络参数设置

损失函数曲线能够反映出深度学习模型的学习质量^[16]，对损失曲线的收敛情况进行分析，可以更好地评估模型性能。本文按照表 3 的网络参数配置在红外变电设备数据集上进行训练，训练过程中的损失函数变化如图 7 所示。

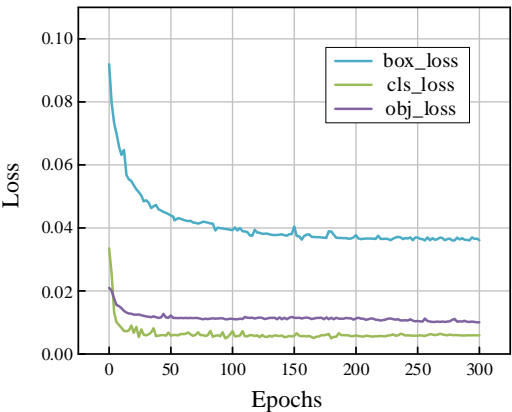
由图 7(a)可知，训练过程中网络并未发生过拟合、欠拟合的现象，均在迭代结束之前完成收敛。图 7(b)为改进前后 box_loss 对比，由图可以看出改进后的损失函数相较于原函数收敛速度更快，loss 值更小，最

后稳定在 0.036，改进后的算法相较于原算法收敛效果得到明显提升。

表 3 网络参数设置

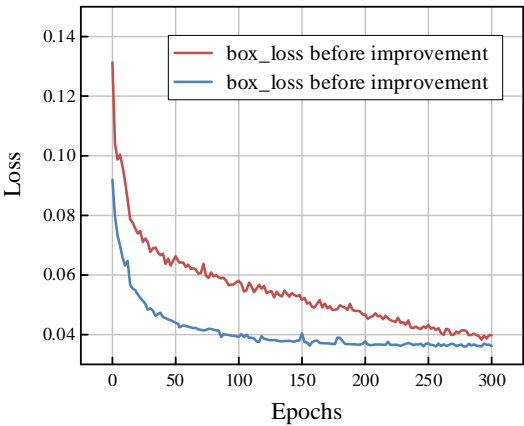
Table 3 Network parameter settings

Parameters	Value
Epoch	300
Size of batch	16
Learning rate	1×10^{-3}
Weight decay	5×10^{-4}
Image size	640×640
Momentum	0.9



(a) 训练过程损失函数变化

(a) Training process loss function variation



(b) 改进前后 box_loss 变化

(b) Box_loss variation before and after improvement

图 7 损失函数变化

Fig.7 Variation of loss function curves

表 4 列举了本文算法在红外变电设备数据集上对各类变电设备目标的识别结果，从表中数据可以看出本文算法在各类目标设备的识别中均表现优秀，即使在样本较少的避雷器上也能达到较高的精确率。在迭代过程中，各项指标变化趋势如图 8 所示。

图 9 展示了本文算法与 YOLOv7-Tiny 算法的精确率和 mAP0.5 在迭代过程中的具体变化趋势。从图

中可以看出，本文算法相较于 YOLOv7-Tiny 收敛速度更快，达到收敛后的指标表现更优秀。

表 4 各类目标识别结果

Table 4 Recognition results of various targets

Equipment type	Precision/%	Recall/%	mAP0.5/%
Potential transformer	95.5	90.7	94.2
High pressure bushing	96.9	87.2	96.0
Suspension insulator	94.1	95.6	88.1
Lightning protector	98.3	80.8	88.9

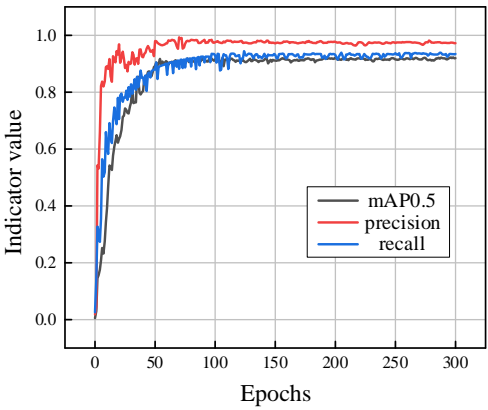
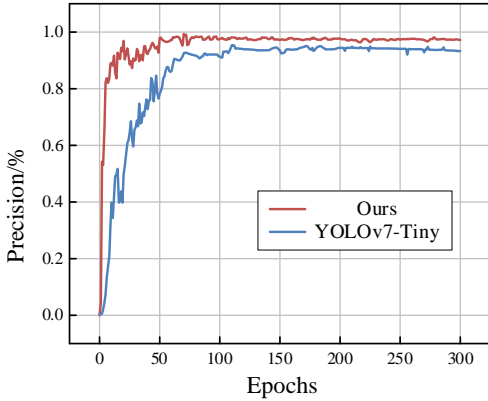
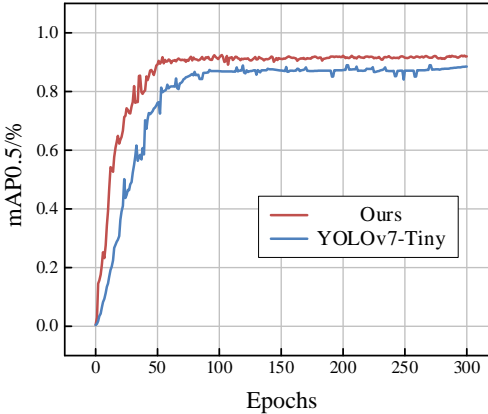


图 8 各指标变化趋势

Fig.8 Trends in each indicator value



(a) Precision



(b) mAP0.5

图 9 评价指标对比曲线

Fig.9 Comparison curves of evaluation indicators

2.4 消融实验

为验证本文所提改进算法的有效性，本文以 YOLOv7-Tiny 为基础进行消融实验，以评价所作改进对算法性能的影响。各改进策略依次为轻量化特征提取网络、加入 CA 注意力机制以及修改损失函数。实验结果如表 5 所示。

由消融实验结果可以看出，各项改进措施对模型性能均有一定程度的提升作用。本文的轻量化结构能够将模型大小从 12.39 M 压缩至 7.82 M，有效地节约了模型存储空间。通过添加 CA 注意力模块和将坐标

损失函数改为 SIoU Loss 的方法将模型精确率提升 6.0%，更有利于完成识别任务。综上所述，本文所做出的各项改进将平均精确率和 mAP0.5 分别提高了 6.0%、1.9%，参数量降低 36.9%，在保证模型良好识别性能的同时减少存储空间占用。

2.5 对比实验

为验证本文算法的优越性，选取 5 种不同场景下的变电站电力设备红外图像进行对比实验，对比算法分别为 SSD、Faster R-CNN、YOLOv7 和 YOLOv7-tiny，其结果如图 10 所示。

表 5 不同改进策略的消融实验

Optimization strategy				Evaluation			
GhostNetV2				P	R	mAP	Model size/M
Ghost module	DFC	CA	SIoU Loss				
YOLOv7-Tiny				90.2	90.4	88.9	12.39
√				87.6	92.8	85.8	7.63
√	√			90.8	91.5	89.2	7.69
√	√	√		93.5	92.2	90.1	7.75
√	√		√	93.7	91.4	90.8	7.72
√	√	√	√	96.2	93.2	91.8	7.82

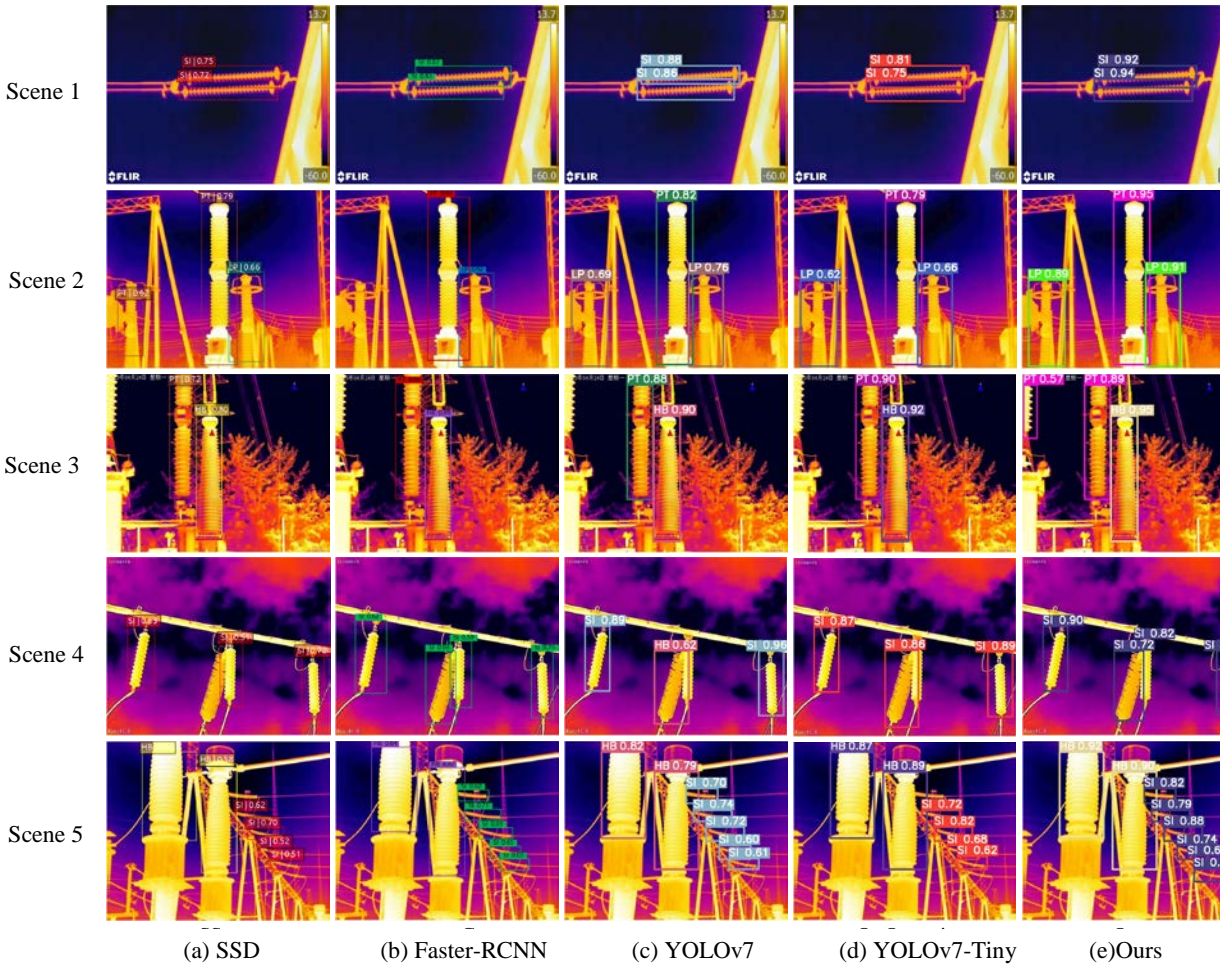


图 10 各算法识别结果

Fig.10 Recognition results of each algorithm

由场景一可以看出，在背景清晰的情况下，所有识别方法均能检测出目标，但本文方法有最高的置信度。由场景二可以看出，在多目标场景下，SSD将左下角避雷器误检为电压互感器，Faster-RCNN没有检测出左下角复杂环境中的避雷器，YOLO系列算法和本文方法都可以检测出目标，但本文方法有更高的置信度。由场景三可以看出，除本方法外的其他所有方法都没有检测出左上角的部分电压互感器。由场景四可以看出，在目标被遮挡的情况下，只有本文方法识别出所有的目标，其他方法均有漏检情况存在。由场景五可以看出，本文在YOLOv7-Tiny的基础上引入CA注意力机制后，对小目标特征信息提取更加完善，对远距离小目标的检测更加精准，可以检测出图像中所有的远距离小目标。实验结果如表6所示。

表6 不同网络的对比实验

Table 6 Comparative experiments on different networks				
Methods	P/%	mAP0.5/%	FPS	Model size/M
SSD	91.78	91.31	14.27	92.17
Faster R-CNN	93.54	91.58	9.46	119.59
YOLOv7	94.46	91.43	18.49	24.25
YOLOv7-Tiny	91.72	89.75	21.71	12.39
Ours	96.28	91.84	26.42	7.82

Note: The optimal data is displayed in bold.

由表6可知，本算法在变电站设备红外数据集上的精确率达到96.28%，检测速率达到26.42 frame/s，模型大小降低至7.82 M，优于另外4种方法。与YOLOv7-Tiny原算法相比较，本算法在提升识别精度的同时将检测速率提升21.69%，模型大小仅有原算法的63.11%。

3 结论

为解决变电设备的精准快速识别问题，本文提出一种基于YOLOv7-Tiny的变电设备红外图像识别方法。通过实验分析，得出以下结论：

- 1) 在变电设备红外图像数据集上进行实验，本文算法在不同目标识别任务中都有出色表现，平均精度均值可以达到91.84%，模型大小仅有原YOLOv7-Tiny算法的63.1%。
- 2) 将本文算法与其他4种经典算法在同一实验平台进行对比，本文提出方法的精确率高于其他四种方法，且每秒检测帧数达到26.42张，可以有效解决漏检、误检的问题。
- 3) 本文提出的变电设备红外图像识别方法，既保证了网络识别的准确性，又提高识别速度以满足变电站设备实时监控的需要。

参考文献：

[1] 王晶, 姚邹静, 赵春晖. 基于红外图像时空特征的电力设备故障诊断[J]. 控制工程, 2021, **28**(8): 1683-1690.

WANG Jing, YAO Zoujing, ZHAO Chunhui. A fault diagnosis method for power equipment based on spatiotemporal features of infrared images[J]. *Control Engineering of China*, 2021, **28**(8): 1683-1690.

[2] HAN S, YANG F, YANG G, et al. Electrical equipment identification in infrared images based on ROI-selected CNN method[J]. *Electric Power Systems Research*, 2020, **188**: 106534.

[3] 吴添权, 郭竞, 苟先太, 等. 基于改进高斯卷积核的变电站设备红外图像检测方法[J]. 红外技术, 2021, **43**(3): 230-236.

WU Tianquan, GUO Jing, GOU Xiantai, et al. Method of detecting substation equipment in infrared images based on improved gaussian convolution kernel[J]. *Infrared Technology*, 2021, **43**(3): 230-236.

[4] DUAN K, BAI S, XIE L, et al. CenterNet: keypoint triplets for object detection[C]//*IEEE International Conference on Computer Vision*, 2019: 6569-6578.

[5] 邝楚文, 何望. 基于红外与可见光图像的目标检测算法[J]. 红外技术, 2022, **44**(9): 912-919.

KUANG Chuwen, HE Wang. Object detection algorithm based on infrared and visible light images[J]. *Infrared Technology*, 2022, **44**(9): 912-919.

[6] 王媛媛, 李媛媛, 段誉, 等. 基于轻量骨干网络和注意力结构的变电设备红外图像识别[J]. 电网技术, 2023, **47**(10): 4358-4369.

WANG Yuanbin, LI Yuanyuan, DUAN Yu, et al. Infrared image recognition of power transformation equipment based on light backbone network and attention structure [J]. *Power Grid Technology*, 2023, **47**(10): 4358-4369.

[7] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779-788.

[8] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 1580-1589.

[9] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11534-11542.

[10] 王有元, 李后英, 梁玄鸿, 等. 基于红外图像的变电设备热缺陷自调整残差网络诊断模型[J]. 高电压技术, 2020, **46**(9): 3000-3007.

WANG Youyuan, LI Houying, LIANG Xuanhong, et al. Thermal defect self-tuning residual network diagnosis model of substation equipment based on infrared image[J]. *High Voltage Technology*, 2020, **46**(9): 3000-3007.

[11] WANG C Y, Bochkovskiy A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7464-7475.

[12] TANG Y, HAN K, GUO J, et al. GhostNetv2: enhance cheap operation with long-range attention[J]. *Advances in Neural Information Processing Systems*, 2022, **35**: 9969-9982.

[13] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13713-13722.

[14] LIN T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2980-2988.

[15] Gevorgyan Z. SIOU loss: More powerful learning for bounding box regression[J]. *arXiv preprint arXiv:2205.12740*, 2022.

[16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J/OL]. *Advances in Neural Information Processing Systems*, 2017, <https://arxiv.org/abs/1706.03762>.