

DOI:10.19651/j.cnki.emt.2314607

复杂背景条件下的电气设备图像实例分割算法^{*}

张志君^{1,2} 张惊雷^{1,2} 贾鑫³

(1.天津理工大学电气工程与自动化学院 天津 300384; 2.天津市复杂系统控制理论及应用重点实验室 天津 300384;
3.天津理工大学工程训练中心 天津 300384)

摘要:变电站巡检拍摄的电气设备可见光图像存在背景杂乱、目标轮廓不规则等特点,造成设备分割精度不高,影响智能巡检系统设备识别效果。基于此,提出一种改进的 YOLACT++模型,实现设备目标精确实例分割。首先,设计了电气设备特征提取主干网络 DAGNet,提升了网络对复杂背景下重要特征的关注度;同时在原型网络分支引入3D注意力模块 SimAM,降低混乱背景对目标分割的干扰。使用某市8个区域58座110 kV变电站和86座35 kV变电站巡检所得避雷器、断路器等6类电气设备的1 730张可见光图像的标记数据集对该模型进行验证,实验结果表明,改进YOLACT++模型分割的AP_{all}指标为84.1%,相较原模型提高了4.4%,与YOLACT、Mask R-CNN和YOLOv8模型相比分别高出4.0%、9.3%、1.6%,较好地实现了6类电气设备的识别,可满足电力巡检中准确性和快速性的要求。

关键词:电气设备识别;实例分割;YOLACT++;可见光图像

中图分类号: TM755 文献标识码: A 国家标准学科分类代码: 520.6040

Instance segmentation algorithm for electrical equipment images under complex background conditions

Zhang Zhijun^{1,2} Zhang Jinglei^{1,2} Jia Xin³

(1. School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, China;
2. Tianjin Key Laboratory for Control Theory and Applications in Complicated Systems, Tianjin 300384, China;
3. Engineering Training Center, Tianjin University of Technology, Tianjin 300384, China)

Abstract: The visible light images of electrical equipment in substation inspection are characterized by background clutter and irregular target contours, causing poor equipment segmentation accuracy and affecting the equipment recognition effect of intelligent inspection systems. This paper proposes an improved YOLACT++ model to realize accurate instance segmentation of equipment targets. First, the electrical equipment feature extraction backbone network DAGNet is designed to improve the network's attention to important features in the complex background. Simultaneously, the 3D attention module SimAM is introduced in the prototype network branch to reduce the interference of the chaotic background on target segmentation. The model is validated using a labeled dataset of 1 730 visible images of six types of electrical equipment, including surge arresters and circuit breakers, obtained from inspections of 58 110 kV substations and 86 35 kV substations in eight regions of a city. The experimental results show that the AP_{all} index of the improved YOLACT++ model segmentation is 84.1%. It is 4.4% higher, and with YOLACT, Mask R-CNN, and YOLOv8 models, it is 4.0%, 9.3%, and 1.6% higher, better realizing the recognition of the six types of electrical equipment. The accuracy and rapidity of electric power inspection are met.

Keywords: electrical equipment detection;instance segmentation;YOLACT++;visible image

0 引言

变电站是电网系统中的关键环节,及时掌握变电站中

电力设备的绝缘水平和运行状态对电力系统的安全稳定运行至关重要。目前变电站设备巡检主要为人工巡检,不仅耗时费力,且不能及时准确地反映关键设备的运行状况。

收稿日期:2023-09-18

* 基金项目:国家自然科学基金青年项目(62302335)资助

基于现场图像的故障自动检测系统可有效地保证变电站的稳定运行。随着人工智能和计算机视觉技术的不断发展,视觉检测在变电站设备巡检和维保的应用越发普及。目前基于人工智能方法的变电站电气设备识别方法主要分为传统机器学习类方法和深度学习类方法^[1]。前者主要针对电气设备的形状和纹理特征进行识别分类,优点是模型复杂度低,对计算资源的要求不高,缺点是对目标的拍摄角度等要求较高且需要人工提取数据中的特征;深度学习类方法通过卷积神经网络自动提取图像中的特征,可实现端到端的识别,并获得了更加令人印象深刻的结果。

深度学习类电气设备识别算法主要分为单阶段与双阶段两种方式,前者以追求快速检测的 YOLO(you only look once)系列为代表,后者以强调精确识别的 Faster R-CNN (towards real-time object detection with region proposal networks) 系列为代表。研究者们根据特定设备检测需要,在算法上做相应改进。双阶段的算法因高检测精度受到研究者的关注。针对绝缘子背景复杂导致巡检精度低等问题,柳方圆等^[2]提出一种改进的 Faster R-CNN 算法,采用双分支的 ResNet50 代替单分支的 ResNet50 且引入可变形卷积,并根据绝缘子的形状重新设计锚框参数,有效改善了复杂背景下小目标漏检、误检现象;Wang 等^[3]提出一种对红外绝缘子图像进行实例分割和温度分析的自动诊断方法,利用迁移学习和动态学习率的方法训练 Mask R-CNN 模型,实现对绝缘子的故障分析;苟军年等^[4]对 Mask R-CNN 做出改进,通过在特征提取网络中引入注意力机制加强语义信息提取,使用全局交并比(generalized intersection over union, GIoU)计算目标间的相似度,以及使用 Tversky 损失函数计算掩膜分支损失,提升绝缘子缺陷检测小目标的检测精度。

单阶段 YOLO 类算法以简洁的网络结构和较高的准确度受到广泛关注,其中 YOLACT++ 网络是单阶段实例分割算法的代表作,所提出的原型网络可有效地区分前景与背景,特别适用于复杂场景的工业现场设备识别任务。为解决煤矿井下巷道中无人驾驶电机车障碍物识别问题,童佳乐^[5]在 YOLACT++ 中引入自下而上的路径增强方法(bottom-up path augmentation, BPA) 和 Focal Loss 函数,利用底层细节信息提升模型的学习能力;针对绝缘子的遮挡及背景复杂等问题,高熠等^[6]对 anchor 超参数以及掩膜后处理做了优化,使用路径增强思想优化特征金字塔网络(feature pyramid network, FPN),实验表明该算法可以实现输电线路绝缘子的有效检测。上述研究分割的物体轮廓规则,未强化网络的特征提取能力,且由于分割场景较简单,也未加强对原型掩膜的辨识力。

基于视觉的电力系统设备巡检分基于可见光图像和基于红外热图像两类方法。相比于红外热图,可见光巡检图像不仅背景更混乱,且目标边缘也更复杂^[7]。因此现有的实例分割方法并不能直接应用于电气设备可见光图像的分

割任务。

本文基于 YOLACT++(better real-time instance segmentation) 网络,提出了一种改进 YOLACT++ 的电气设备可见光图像实例分割网络(visible light image of electrical equipment based on improved YOLACT++),命名为 Visible YOLACT。主要工作如下:在设备特征提取方面,设计了一个应用于可见光图像复杂背景下提取电气设备特征的主干网络 DAGNet(adaptive significant feature extraction network),提升对电气设备复杂轮廓的特征抽取能力;在抑制复杂背景方面,将无参注意力机制(simple attention mechanism, SimAM)融入原型网络分支,强化有用前景信息,弱化嘈杂的背景信息;同时在自建的电气设备可见光数据集上进行了大量实验评估所提出的方法,实验结果表明,本文方法较好地实现了 6 类电气设备的检测和实例分割。

1 YOLACT++ 检测框架

YOLACT++ 模型^[8]是 BOLYA 等基于一阶段型目标检测模型提出的实时实例分割模型,其将分割任务拆解为两个并行的分支。原型网络分支生成一系列独立于单一实例的原型掩码;预测头部分支在目标检测分支上添加额外的头去预测掩膜系数,用于编码一个实例在原型掩膜的空间表示,最后合并处理两个分支的输出,用系数与原型掩膜做简单的线性组合,即可得到每个实例的最终掩膜,基本框架如图 1 所示。YOLACT++ 模型架构简洁、高效且易于训练,可实时对变电站电力设备进行检测,然而由于全卷积网络无差别的特征提取,致使网络对上下文信息利用不充分且无法自适应地关注有效特征,导致精度不高。为提高对电力设备的检测精度,本文替换主干网络并对原型网络做了改进,以让其更适应于变电站的电力巡检任务。

2 Visible YOLACT 检测框架

变电站中电气设备密集排布导致背景复杂混乱,对目标电气设备的分割造成很大的困难。为实现电气设备的准确分割,本文利用自建的变电站电气设备数据集,提出一种 Visible YOLACT 算法对各类设备进行分类识别、目标定位和实例分割。1)现有对电气设备的识别多集中于红外图像,而可见光图像的色彩、纹理更加丰富,因此轮廓也显得更为复杂,对此本文设计了一个用于可见光图像电气设备特征提取的主干网络 DAGNet;2)原型网络分支会提取出特征图 P3 中不同物体或位置的原型掩码,将捕捉神经元重要性的 SimAM 注意力机制融合进去,所得到的各个原型掩码区分度会更高,网络会更不容易受到复杂背景等噪声因素的干扰。图 2 展示了 Visible YOLACT 的网络结构图。

2.1 深度特征提取主干网络 DAGNet

YOLACT++ 采用基于 ResNet101 的主干网络,由于

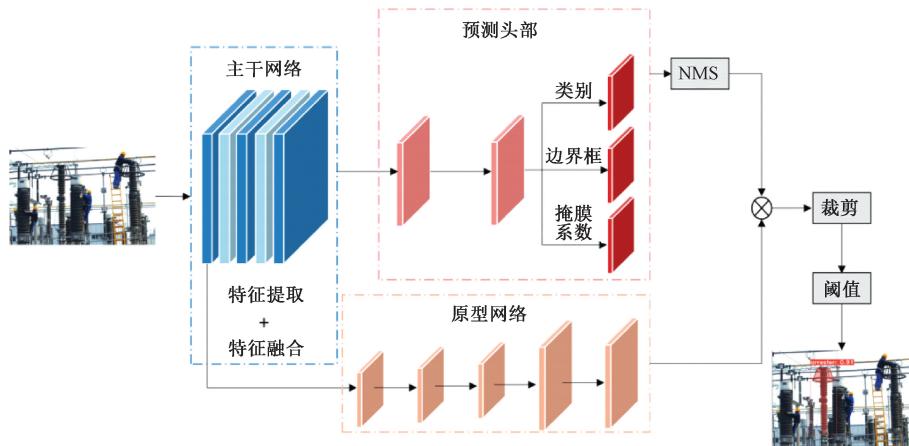


图 1 YOLACT++ 检测框架

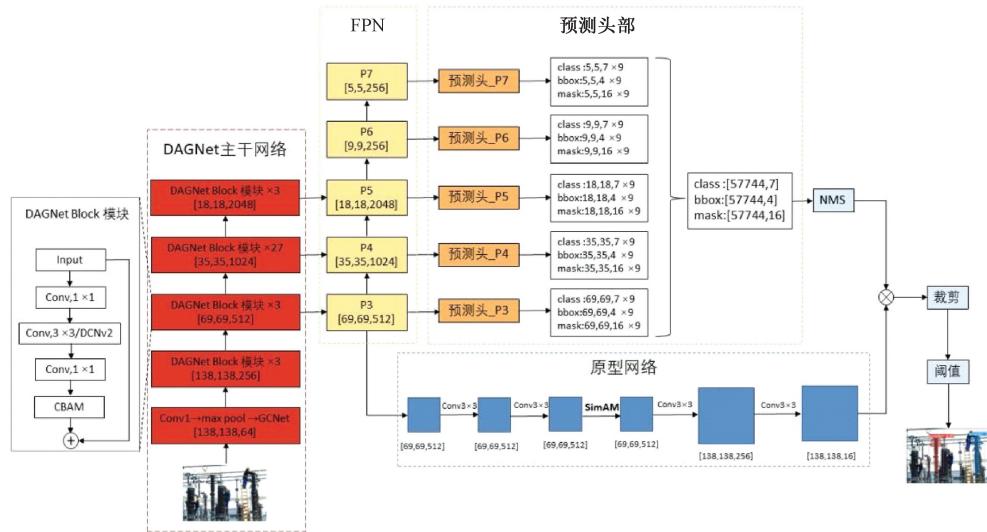


图 2 Visible YOLACT 检测框架

缺少目标在空间布局以及几何形状之间的联系建立,导致特征提取不充分;此外,变电站可见光图像所蕴含的色彩、纹理更加丰富,更需要关注特征的细节捕捉。因此,本文设计了一个用于可见光图像电气设备特征提取的主干网络 DAGNet。DAGNet 与 ResNet101 的结构对比如表 1 所示。

1) 残差堆叠设计。变电站中不同种类电力设备外观相似度高,为避免误检等现象的发生,需要增强网络提取特征信息的能力。ResNet101 网络通过引入残差连接保留更多原始特征,提高模型的精度与泛化能力,而且在学习输入、输出的差别过程中,朝恒等映射的方向收敛,避免了因网络层数加深而导致的梯度消失与梯度爆炸问题。因此 DAGNet 采用与 ResNet 一样的残差结构与四阶段的分布式设计,但在四个阶段 block 的堆叠比例上采用 ConvNext 网络^[9][1 : 1 : 9 : 1]的设计策略,兼容下游任务,提高实例分割的准确率。

2) 浅层网络设计。注意力机制能够以高权重聚焦重要信息,以低权重忽略不相关信息,通过不断调整权重,使

网络选取重要的信息。变电站巡检所拍摄的图像往往质量较差,存在图像模糊或目标倾斜等问题。对于此类特殊图像,检测时期望网络能对特定区域进行高分辨率处理,同时更加关注纹理、细节等信息的学习,以实现更加精确的定位。全局上下文网络 (global context network, GCNet)^[10]可以提高网络对长距离依赖特征提取能力,避免了普通卷积在应对关系密切的目标距离卷积核所在的位置较远,导致该卷积只能观察到其卷积范围内的部分图像,无法建立长距离依赖关系的缺陷,其结构如图 3 所示。

注意力模块 GCNet 首先将经过特征提取后的特征图 $\mathbf{X} \in \mathbf{R}^{C \times W \times H}$ 通过全局上下文建模中 1×1 卷积层以及非线性激活函数 Softmax 得到包含全局上下文信息的权重矩阵,然后与 reshape 后的输入特征矩阵相乘,先得到具有全局上下文信息的特征矩阵 (\mathbf{Z}_1);再利用通道权重变换模块计算每个通道的重要程度,得到大小为 $C \times 1 \times 1$ 的特征矩阵 \mathbf{Z}_2 ;最后将具有不同通道权重的特征矩阵 \mathbf{Z}_2 与输入特征图 $\mathbf{X} \in \mathbf{R}^{C \times W \times H}$ 相乘,即可得到含有不同权重信息的特征

图。3个过程的计算式如下:

$$\mathbf{Z}_1 = \sum_{j=1}^{N_p} \frac{e^{\mathbf{w}_k \mathbf{x}_j}}{\sum_{m=1}^{N_p} e^{\mathbf{w}_k \mathbf{x}_m}} \quad (1)$$

$$\mathbf{Z}_2 = W_{v2} \text{ReLU}(LN(W_{v1} \mathbf{Z}_1)) \quad (2)$$

$$\mathbf{Z} = \mathbf{X} \times \mathbf{Z}_2 \quad (3)$$

式中: $\sum_{j=1}^{N_p} \frac{e^{\mathbf{w}_k \mathbf{x}_j}}{\sum_{m=1}^{N_p} e^{\mathbf{w}_k \mathbf{x}_m}}$ 表示全局注意力池化的权值, \mathbf{W}_k 表示线性转换矩阵, $\mathbf{X}_j, \mathbf{X}_m$ 表示特征图中任一位置信息, N_p 表示特征图中位置信息数量, ReLU 为非线性激活函数, LN 表示层归一化, W_{v1}, W_{v2} 分别为两个卷积层的权重参数。由于浅层的特征图包含较多的纹理、细节等信息, 在主干网络的开头的第一个 7×7 的卷积层与最大池化层之后融合 GCNet, 建立电气设备各细节长距离的依赖, 充分利用深层语义信息, 提升网络应对目标倾斜或图像模糊等特殊情形的能力。

表 1 DAGNet 与 ResNet-101 网络对比

ResNet-101	DAGNet
$7 \times 7, 64, \text{stride } 2$	$7 \times 7, 64, \text{stride } 2$
$3 \times 3 \text{ max pool, stride } 2$	$3 \times 3 \text{ max pool, stride } 2$
$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \\ CBAM \end{bmatrix} \times 3$
$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3/DCN, 128 \\ 1 \times 1, 512 \\ CBAM \end{bmatrix} \times 3$
$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3/DCN, 256 \\ 1 \times 1, 1024 \\ CBAM \end{bmatrix} \times 27$
$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3/DCN, 512 \\ 1 \times 1, 2048 \\ CBAM \end{bmatrix} \times 3$

3)瓶颈架构设计。由于电气设备边缘轮廓不规则, 采用普通的卷积神经网络进行特征采样往往存在一定缺陷, 该缺陷主要来自卷积神经网络固有的几何结构: 卷积核对输入特征图的固定位置进行采样, 池化以固定的比例进行池化。本文算法采用可变形卷积(deformable convolution, DCN), 对采样点学习一个偏移, 从而适应目标形状的变化, 提高分割的鲁棒性。普通卷积采样和可变形卷积采样的对比如图 4 所示。

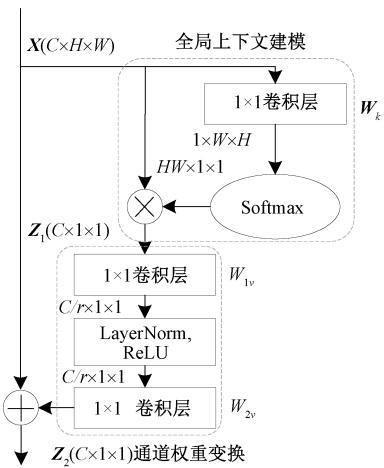


图 3 GCNet 自注意机制模块

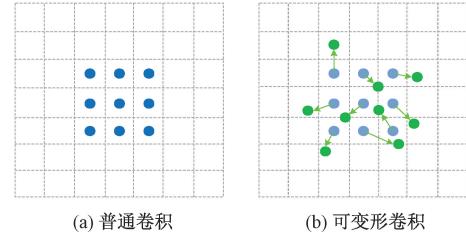


图 4 普通卷积与可变形卷积对比

在标准的二维卷积过程中, 对于输出特征图中的每一个位置 P_0 , 其特征值 $y(P_0)$ 计算式为:

$$y(P_0) = \sum_{P_n \in R} W(P_n) x(P_0 + P_n) \quad (4)$$

式中: $W(P_n)$ 是该采样位置的卷积核权重, $x(P_0 + P_n)$ 是采样位置的输入特征值, P_n 是感受野区域中所有采样位置, R 是感受野区域。而在可变形卷积中

$$y(P_0) = \sum_{P_n \in R} W(P_n) x(P_0 + P_n + \Delta P_n) \quad (5)$$

式中: $x(P_0 + P_n + \Delta P_n)$ 为采样偏移位置的输入特征值, ΔP_n 为采样点位置的偏移量。可见, 可变形卷积是在传统卷积操作上加入了一个采样点位置偏移量, 使普通卷积变为可变形卷积, 从而拥有更大、更灵活的感受野。由于可变形卷积引入偏移模块, 导致过多的无关上下文, 因此可变形卷积 v2^[11] 通过控制权重有效降低过多上下文信息的影响。其计算公式变为:

$$y(P_0) = \sum_{P_m \in R} W(P_m) x(P_0 + P_m + \Delta P_m) \cdot \Delta m_m \quad (6)$$

可变形卷积在自适应地提取特征信息的过程中会增加一定的参数和计算, 如果全部使用可变形卷积设计网络, 会导致参数量巨大、网络难以训练。因此在瓶颈结构中有间隔地引入 DCNv2, 并选择效果最好的每隔 4 个 block 用可变形卷积 v2 替换传统的 3×3 卷积, 其中可变形卷积的使用选择如表 2 所示。同时在瓶颈结构中引入 CBAM^[12] 注意力模块, 进一步对通道特征与空间特征进行

加权,抑制无效信息,使网络关注更多目标区域特征,实现网络对目标边缘更加精细的分割。

表 2 使用可变形卷积的不同选择

模型	分割平均精度 / %		
	AP _{all}	AP ₅₀	AP ₇₅
inter=3	81.7	91.7	89.2
inter=4	83.7	93.3	90.6
inter=5	81.2	91.2	88.3

2.2 融合注意力机制的原型网络

原型网络是 YOLACT++ 网络的特有结构,用于生成一系列的原型掩码,有的原型掩码可以分割出目标的轮廓,有的会分割出背景。具体过程如下:当网络确定一个位置为所要分割的目标像素点时,对该位置的响应很大,因此网络最后一层使用非线性激活函数 relu,得到最终的原型掩膜。然而变电站电气设备排布密集且多有重叠,容易将背景误分割。因此在原型网络中引入可以查找每个神经元重要性的能量函数,强化像素点的辨识能力,得到更准确的原型掩码,有效滤除干扰分割的背景信息。

SimAM^[13]是一种无参化的 3D 注意力模块,其结构如图 5 所示。该模块通过能量函数的解析计算提高注意力权值的计算速度,相比于现有的通道、空间注意力模块具有轻量性和灵活性的优势。

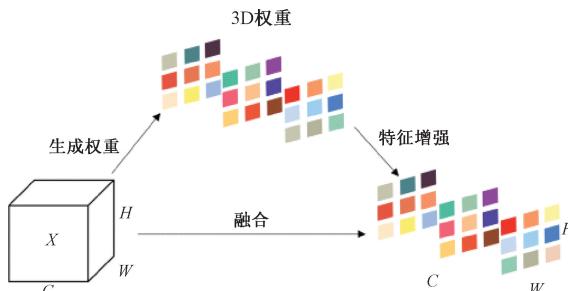


图 5 SimAM 结构

由于此注意力模块是通过度量每个神经元的线性可分性的方法来评估每隔神经元之间的的重要性。因此提出能量函数,经过一系列简化,最小能量函数表达式如式(7)所示。

$$e_i^* = \frac{4(\sigma^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (7)$$

其中, t 表示神经元, λ 为超参数, $\hat{\mu}$ 和 $\hat{\sigma}^2$ 的表达式如式(8)和(9)所示。

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i \quad (8)$$

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2 \quad (9)$$

其中, x 为神经元 t 的相邻神经元。由式(7)可知,能量越低,神经元与周围神经元的区别越大,重要性则越高,

因此,神经元的重要性可由 $\frac{1}{e_i^*}$ 获得。最后通过式(10)的方式对特征进行增强处理。

$$\tilde{X} = \text{Hardsigmoid}\left(\frac{1}{E}\right) \odot X \quad (10)$$

与常见的注意力机制不同,SimAM 能够对独立神经元进行重要性评估,可以在不提高模型参数量的同时对重要的神经元赋予更高的权重。

3 实验结果分析

3.1 数据集

本文所使用的数据集是实验室私有的数据集。数据来源于某网省公司 8 个区域的 58 座 110 kV 变电站和 86 座 35 kV 变电站采集的超 2TB 的设备巡检图像,建立了包含避雷器、断路器、电流互感器、绝缘子、电抗器、隔离开关等六类电气设备共 1 730 张图像的数据集。其中图像的尺寸多为 1 600×1 200 像素与 640×480 像素。

3.2 实验环境及参数

本文实验环境基于浪潮英信服务器 NF5280M4, GPU 为 NVIDIA GeForce GTX1080Ti, 操作系统为 Ubuntu 20.04.5。实验采用的深度学习框架为 Pytorch 1.11, 编程语言为 Python 3.8, cuda 版本为 11.3。

3.3 评价指标

本文选用平均精度(average precision, AP)评价模型的性能,这是实例分割领域中最常用的评价指标。由于 IoU 阈值的选取会影响到精确度和召回率的大小,故选用 AP_{all}、AP₅₀ 和 AP₇₅ 作为预测框和实例分割掩膜主要的评价指标。其中 AP_{all} 表示计算以 0.05 为步长, IoU 在 0.5~0.95 之间变化的 AP 平均值, AP₅₀ 和 AP₇₅ 分别表示 IoU 为 0.50 和 0.75 时计算所得 AP。IoU 是模型对某一类别预测结果和真实值的交集与并集的比值,即模型产生的预测区域与原标注区域的重叠度。

3.4 消融实验结果与分析

本文选用变电站电力设备数据集,选择目标与背景颜色相近、分割场景复杂等几种情况的图像进行测试,并通过消融实验验证改进算法的有效性。表 3 为消融实验的测试结果,可见主干网络更换为 DAGNet 后,与原模型相比,预测框的 AP_{all}、AP₅₀、AP₇₅ 分别提升了 4.0%、4.6%、4.3%,达到了 83.7%、93.3%、90.6%。采用 DAGNet 并将 SimAM 注意力机制融入原型网络分支后,改进后的模型不仅在测试集的评价指标上有一定的提升,在可视化的结果图上表现更为明显。如图 6 可视化结果所示,虽然 3 个断路器与其他相似电气设备发生紧密重叠且所摄图像也较为模糊,但改进后的模型依然可以完整地分割出图像中的 3 个断路器。

此外,由训练的损失函数(图 7)可知,原 YOLACT++ 模型与改进后的 Visible YOLACT 均在迭代到第 147 个

表3 消融实验测试结果

模型	边界框平均精度/%			分割平均精度/%		
	AP_{all}	AP_{50}	AP_{75}	AP_{all}	AP_{50}	AP_{75}
YOLACT++	76.4	89.7	85.8	79.7	88.7	86.3
YOLACT++ +DAGNet	77.1	93.5	88.3	83.7	93.3	90.6
YOLACT++ +DAGNet +SimAM	77.9	94.7	89.0	84.1	94.3	92.3



(a) YOLACT++分割效果



(b) YOLACT++ +DAGNet分割效果



(c) YOLACT++ +DAGNet +SimAM分割效果

图6 消融实验可视化结果对比

epoch 开始收敛,最终稳定在 0.794,收敛的情况也验证了本文提出模型的训练效果较理想。

设备之间的相似度也较高,这对检测具有一定的影响。如图 6 中要分割的避雷器所处场景复杂,由于干扰因素较多,采用原模型出现漏检现象;又由于不同设备之间的相似度较高,对模型的判断产生了误导,从而造成检测精度低以及漏检的现象。改进后的模型可避免复杂背景的干扰,实现电气设备的有效检测,尤其是更换了有着更强几何细节信息和语义信息表征能力的 DAGNet 主干网络,加强了图像中不同目标空间和语义的上下文联系,极大地提升了模型在复杂背景下多目标的识别能力。

3.5 不同算法检测结果比较

为进一步验证本文算法的检测性能,将 Visible YOLACT 与 YOLACT^[14]、YOLACT++、Mask R-CNN^[15] 和 YOLOv8 进行了对比,其中 Mask R-CNN 常应用于电气设备红外图像分割任务,YOLOv8 是 2023 年提出的可用于目标检测和实例分割等任务的热门网络。各算法的评价指标如表 4 所示,可视化结果比较如图 8。为更好地验证模型的泛化能力,从网上搜集了俩张特殊场景的图像进行对比

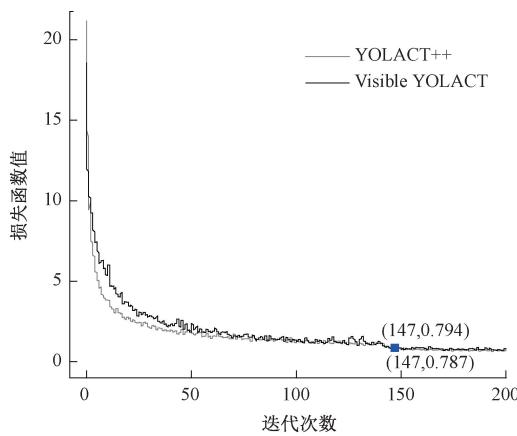


图7 训练损失曲线

结合评价指标和检测效果图进一步分析得到以下结论:电气设备通常处于较为复杂的环境,且分布密集,不同

分析:1)所要识别的目标较多且存在多个小目标。此时, YOLACT++、Mask R-CNN 以及 YOLOv8 仅能识别出图中未被遮挡的电抗器,漏检了俩个绝缘子,YOLACT 虽然检测出了其中的一个绝缘子,但对电抗器的分割却比较粗糙,而本文算法不仅对检测出的电抗器分割较为精细,且完整地分割出了全部的小目标绝缘子;2)目标倾斜。此时, YOLACT 算法出现了误检,Mask R-CNN 与 YOLOv8 只能检测到其中的一个电流互感器,YOLACT++虽然识别到了全部的电流互感器,在精度上却略低于 Visible YOLACT。

可见 Visible YOLACT 具有较强的泛化能力。综上所述,较高的精度值表明本文算法可实现准确的检测与分割,可视化结果则体现了检测结果与实际目标的贴合性,且算法可有效滤除背景干扰,即便是在复杂场景下仍可捕捉到有效的特征信息,完成设备的精确分割。在平均检测帧率上 Visible YOLACT 较其他的一阶段算法检测速度变慢,这是由于 DAGNet 引入了一部分的参数量。综合实验结果来看,Visible YOALCT 对检测速度和检测精度有更好的平衡,能够应用于对检测精度有更高要求的变电站巡检场景中。

表 4 5 种分割模型在测试集中的测试结果

模型	边界框平均精度/%			分割平均精度/%			FPS
	AP_{all}	AP_{50}	AP_{75}	AP_{all}	AP_{50}	AP_{75}	
YOLACT	76.1	91.8	85.1	80.1	91.6	89.2	32.3
YOLACT++	76.4	89.7	85.8	79.7	88.7	86.3	26.1
Mask R-CNN	80.3	91.3	88.4	74.8	91.2	87.5	11.6
YOLOv8	77.0	93.9		82.5	93.7		29.8
Visible YOLACT	77.9	94.7	89.0	84.1	94.3	92.3	15.1

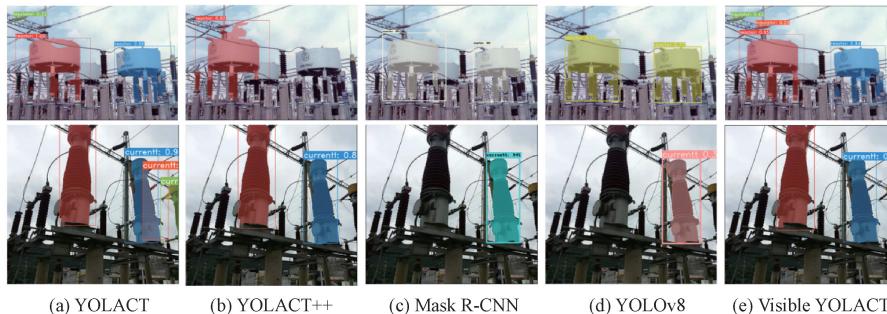


图 8 各模型的分割效果

为进一步验证模型复杂度,本文与其他几个单阶段、双阶段的实例分割算法在参数量上做了对比。由表 5 参数量对比可知,Visible YOLACT 虽然采用融合了注意力机制 DAGNet 主干网络,但在模型的复杂度上仍较其他网络具有优势,这是由于 Visible YOLACT 继承了原网络通过矩阵乘法生成掩膜的方式,最大程度地简化了掩膜生成的过程,大大避免了参数量的引入。可见本文算法在模型简洁度上仍保持着令人满意的效果。

表 5 不同算法的参数量对比

模型	主干网络	参数量/M
MS R-CNN ^[16]	ResNet-101	208.6
BMask ^[17]	ResNet-101	195.4
SOLOV2 ^[18]	ResNet-101	65.5
Visible YOLACT	DAGNet	55.7

4 结论

本文针对变电站巡检图像背景复杂、设备轮廓不规则等问题,在经典 YOLACT++算法的基础上,提出一种基于电气设备可见光图像的实例分割网络 Visible YOLACT。主要结论如下:

在特征提取阶段采用 DAGNet 替换原基于 ResNet101 的主干网络,丰富网络的特征表达能力,其分割 AP_{all} 与之前相比提高了 4.0%。

在掩码分支网络加入 SimAM 注意力机制,强化各原型掩码的表现力,降低混乱背景对目标分割的干扰。优化后的模型可视化结果明显提升,改善了因背景复杂而导致的漏检问题。

综上,通过上述两种改进,使得 Visible YOLACT 对电气设备的分割 AP_{all} 达到了 84.1%,且模型的参数量只有 55.7 M,在保证模型简洁性的同时实现了较高的分割精度。

参考文献

- [1] 唐文虎,牛哲文,赵柏宁,等.数据驱动的人工智能技术在电力设备状态分析中的研究与应用[J].高电压技术,2020,46(9):2985-99.
- [2] 柳方圆,任东,王露,等.高分辨率的航拍输电线路绝缘子检测[J].电子测量技术,2023,46(13):102-9.
- [3] WANG B, DONG M, REN M, et al. Automatic fault diagnosis of infrared insulator images based on image instance segmentation and temperature analysis [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(8): 5345-55.
- [4] 苟军年,杜愫愫,刘力.基于改进掩膜区域卷积神经网络的输电线路绝缘子自爆检测[J].电工技术学报,2023,38(1):47-59.
- [5] 童佳乐.基于改进实例分割的煤矿电机车障碍物检测技术研究[D].淮南:安徽理工大学,2023.
- [6] 高熠,田联房,杜启亮.基于Mask R-CNN的复合绝缘子过热缺陷检测[J].中国电力,2021,54(1):135-41.
- [7] 蒲天骄,乔骥,韩笑,等.人工智能技术在电力设备运维检修中的研究及应用[J].高电压技术,2020,46(2):369-383.
- [8] BOLYAI D, ZHOU C, XIAO F, et al. YOLACT++ better real-time instance segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(2): 1108-21.
- [9] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022, 11976-11986.
- [10] CAO Y, XU J, LIN S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond [C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop(ICCVW), 2020, 1971-1980.
- [11] ZHU X, HU H, LIN S, et al. Deformable ConvNets V2: more deformable, better results[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2019:9300-9308.
- [12] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]. Computer Vision-ECCV 2018, Munich, Germany, 2018: 3-19.
- [13] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]. International Conference on Machine Learning, PMLR, 2021.
- [14] BOLYAI D, ZHOU C, XIAO F, et al. Yolact: Real-time instance segmentation [C]. Proceedings of the IEEE International Conference on Computer Vision, 2019:9157-9166.
- [15] HE K M, GKIOXARI G, DOLLAR P, et al. Mask r-cnn [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017:2961-2969.
- [16] HUANG Z, HUANG L, GONG Y, et al. Mask scoring R-CNN [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019:6409-6418.
- [17] CHENG T, WANG X, HUANG L et al. Boundary-preserving mask R-CNN[C]. Computer Vision-ECCV 2020. Berlin, German: Springer, 2020: 660-676.
- [18] WANG X, ZHANG R, KONG T, et al. SOLOv2: Dynamic and fast instance segmentation[EB/OL]. (2020-03-23)[2021-09-30]. <https://arxiv.org/abs/2003.10152>.

作者简介

张志君,硕士研究生,主要从事图像处理、实例分割方面的研究。

E-mail:877816185@qq.com

张惊雷(通信作者),教授,硕士生导师,博士,主要从事模式识别、图像处理方面的研究。

E-mail:17335032830@163.com

贾鑫,博士,主要从事红外图像识别方面的研究。

E-mail:1026311742@qq.com