

DOI: 10.19783/j.cnki.pspc.221785

改进 YOLOv5 在电力生产违规穿戴检测中的应用

杨乐¹, 郭一鸣¹, 霍勇博¹, 任晓龙¹, 林平远², 张志宏²

(1. 国网陕西省信息通信有限公司, 陕西 西安 710065; 2. 厦门大学信息学院, 福建 厦门 361000)

摘要: 为了解决现有目标检测系统在电力现场识别中存在的环境复杂、检测物体形状方差过大以及视觉特征辨识度不佳等问题, 提出了一种适用于电力现场穿戴识别的目标检测模型。首先, 通过在 YOLOv5 特征提取网络中嵌入非对称卷积模块, 从而得到更加具备辨识度及鲁棒性的视觉特征。其次, 为了能够在全局背景噪声的影响下自适应地关注与检测物体特征相关性更强的区域, 采用全局注意力机制进行上下文信息的建模, 改进了视觉信息处理的效率与准确性。最后, 通过对比现有的目标检测算法, 证明了所提针对 YOLOv5 改进算法的有效性和优越性。同时, 通过消融实验证明了所改进的模块在目标检测模型中的有效性。

关键词: 非对称卷积网络; 注意力机制; 目标检测; 违规穿戴检测; YOLOv5

Application of improved YOLOv5 for illegal wearing detection in electric power construction

YANG Le¹, GUO Yiming¹, HUO Yongbo¹, REN Xiaolong¹, LIN Pingyuan², ZHANG Zhihong²

(1. State Grid Shaanxi Electric Power Company Information and Communication Co., Ltd., Xi'an 710065, China;

2. School of Informatics, Xiamen University, Xiamen 361000, China)

Abstract: There are problems in existing object detection systems caused by the complicated detection environment of electric power construction, large variance of detected object shape and poor recognition of visual features. This paper proposes an object detection model for illegal wearing detection in electric power construction. First, an asymmetric convolutional group is added to the feature extraction backbone network of YOLOv5 to obtain more discriminative and robust visual features. Second, in order to be able to adaptively focus on feature regions that are more relevant to the detected object under the influence of visual noise, a transformer-based global attention mechanism for modeling contextual information is used to improve the efficiency and accuracy of visual information processing. Finally, the effectiveness and superiority of the improved YOLOv5 algorithm in this paper are demonstrated by comparing with existing object detection algorithms. Meanwhile, the effectiveness of the improved modules in the target detection model is demonstrated by ablation experiments.

This work is supported by the National Natural Science Foundation of China (No. 62176227 and No. U2066213).

Key words: asymmetric convolutional network; attention mechanism; object detection; illegal wearing detection; YOLOv5

0 引言

随着智慧电网项目的全面推进, 如何保证电力现场的施工安全成为了备受关注的问题。由于施工人员常在高危环境下进行作业, 因此存在大量的安

全隐患。在此背景下国家电网提出了多种施工规范, 通过要求施工人员规范穿戴安全保障设备来降低安全隐患。然而, 许多施工人员由于安全意识薄弱而未穿戴安全防护器具, 这个现象成为了电力施工现场发生事故的严重导火索。因此, 研究如何展开智能化监管, 从而提高施工人员安全意识并减少施工事故, 对于电力公司的可持续发展具有重要意义^[1]。

目前绝大多数违规行为的发现依赖于现场安全员以及远程的实时监测, 而人力监管一方面会造成

基金项目: 国家自然科学基金资助项目资助(62176227, U2066213); 中央高校基本科研业务费资助(20720210047); 国家电网陕西省电力公司科技项目资助(SGSNXT00GCJS2200106)

人力资源的浪费,增大企业成本;另一方面长时间工作容易视觉疲劳导致遗漏。因此,为了尽可能降低人为干扰因素,提高监管响应的速度以及准确率,急需引入人工智能技术以实现智能自动化监控与响应,其中涉及到多种计算机视觉算法,如图像识别^[2]与目标检测^[3]方法。与现有的电力现场视频监控系统结合,无需额外改造成本就可以降低企业人力成本,提高企业效益。

目标检测方法是解决电力现场违规穿戴简单且有效的工具,按照特征提取方法的不同可分为两类。1) 基于数字图像处理的方法。这类方法通过设计手工特征对头部、面部和肤色等区域进行特征提取,进而利用分类算法对特征进行识别。文献[4]利用 Harris 检测和方向梯度直方图特征描述图像,然后利用支持向量机对目标进行检测。文献[5]选取尺度不变性特征、变换角点特征和颜色统计特征的方法对目标物体进行检测。虽然这些方法能够初步应用于简单背景下的目标检测,但随着应用场景的变化其准确度也会有相应的波动,泛化能力较差。随着电力公司规模迅速扩大,这类方法已无法适应电网发展与数字化变革的要求。2) 基于深度学习的方法。近年来,随着人工智能技术的高速发展,采用深度学习方法进行目标检测取得了阶段性的成果。文献[6]利用在线困难实例挖掘方法对目标检测模型进行优化,实现了边缘终端的实时运算。文献[7]利用人体关键点检测模型提取人员图像,此后 YOLO^[8] (you only look once)能够实现电力现场保障设备的实时佩戴检测。文献[9]基于 Faster R-CNN^[10]算法,融合多尺度特征提取方案与特征增强模块,通过剔除无关背景区域来提高检测精度。然而,由于实际场景中要检测的物体种类以及数量的快速增长,尤其是针对密集及不规则形状物体的检测难以取得令人满意的效果。

为了实现电力现场违规穿戴的准确检测,本文以 YOLOv5 算法作为基线模型进行改进,提出了一种改进 YOLOv5(improved YOLO, IYOLOv5)算法。该方法引入非对称卷积网络^[11](asymmetric convolutional network, ACNet)中的非对称卷积结构(asymmetric convolutional block, ACB),利用中心增强机制提高视觉特征在不规则物体上的表现力与鲁棒性。为了降低电力现场中环境背景噪声的影响,该方法将自注意力机制^[12]嵌入到模型检测中,通过获取图像的全局依赖信息,进而改进视觉信息处理的效率与精准程度。通过上述改进,本文所提出的方法泛化能力与检测精度均得到有效提升。

1 相关工作与技术背景

1.1 目标检测

目标检测可以根据是否包含兴趣区域搜索分为一阶段方法和二阶段方法。二阶段方法将目标检测任务分解为两个不同的阶段,其中第一个阶段主要是对图片中的候选区域进行筛选,随后的第二阶段则是对筛选出的兴趣区域物体进行分类。作为二阶段目标检测的开创性工作, R-CNN^[13]通过选择性搜索^[14]获得区域建议,然后通过提取卷积神经网络的机器特征对结果进行细化。为了提高 R-CNN 的速度, Fast R-CNN^[15]从整幅图像中提取特征,然后分别通过空间金字塔池和兴趣区域池生成区域特征,最后根据区域特征对预测结果进行细化。Faster R-CNN^[10]可以完全端到端训练,它通过区域建议网络给出预测结果。Cascade R-CNN^[16]通过多级细化改进 Faster R-CNN,可以帮助网络实现更加准确的预测。

与二阶段检测方法相比,一阶段方法效率更高,但精确度较低。SSD^[17]在多尺度特征上扩展锚盒,采用 CNN 直接预测类概率和锚偏移量。EfficientDet^[18]使用网络架构搜索算法自适应构建最优模型。而上文提及的 YOLO^[8]是第一个将目标检测重新定义为一个简单的回归问题的,它可以直接从图像像素中获得边界框坐标和类概率。

1.2 注意力机制

注意力机制的本质是根据不同的输入自适应对特征的权重进行更新,将与检测物体更加相关的特征权重加强,而忽略一些相关性较低的区域特征。该机制来自于人类注意力视觉系统的启发,通过相关性对视觉特征建立动态权重参数组合,可以有效地帮助模型理解复杂的场景^[12]。

注意力机制已广泛用于目标检测, DERT^[19]能够有效地建模图像中的远程依赖关系,并使得目标检测算法完全摒弃了以往算法中依赖人工先验的后处理步骤,构造了一个完全端到端的检测框架。文献[20]构建的方法与本文方法类似,区别之处在于文献[20]所使用的空间注意力^[21]与本文所引入的基于 Transformer 的自注意力机制有所不同。

2 建模方法

2.1 YOLOv5 概述

YOLOv5 主要包括特征提取主干网络 BackBone、特征聚合网络 Neck 以及检测头 Head。其中采用 CSPDarknet53^[22]作为特征提取主干网络,深度为 53 层,在主干网络输入阶段的数据操作与 YOLOv3 相

同, 都利用了 Focus 层来对输入数据进行维度上的切片, 从而对通道进行扩充, 相较于 YOLOv3 所采用 DarkNet53, 其采用的 CSP 结构通过 ResNet^[23] 残差模块的叠加来扩大局部特征的感受野, 同时采用 SiLU 激活函数来替代 YOLOv3 所使用的 Leaky ReLU, 增加了梯度的平滑程度, 使其效果更好。同时在主干网络的最后使用空间金字塔池化层^[24]来获取不同特征尺度下的空间语义信息, 进而提升模型的鲁棒性。

然而, 在电力现场中普遍存在的拍摄角度、环境影响、检测对象不规则等不利因素, 使得 YOLOv5 的表现结果并不理想。本文针对性地引入两个结构, 以挖掘更具有辨识性的语义特征信息来提高模型的识别能力。

2.2 非对称卷积网络

卷积神经网络(convolutional neural network, CNN)由于其局部感受野和权值共享的特性, 具有较好的平移不变性, 在计算机视觉领域倍受欢迎。然而, 在面对图像背景复杂且目标物体不规则的情况下, 普通卷积提取的视觉特征鲁棒性不佳。文献[25]尝试将标准方形卷积核 $d \times d$ 分解为 $1 \times d$ 和 $d \times 1$ 两个非对称卷积核, 在保证模型性能不剧烈降低的同时减少了 $1/3$ 的参数。其结果间接地说明了, 在标准卷积神经网络中, 起最主要作用的是其方形卷积核中间骨架的权重。与此同时, 本文发现标准卷积在面对电力现场情况下, 其学习的视觉特征辨识度不足, 鲁棒性不佳。为了解决这个问题, 本文在视觉特征主干网络中引入了 ACNet^[11] 中的 ACB 结构, 即在标准卷积核 ($d \times d$) 中额外添加水平 ($1 \times d$) 和垂直 ($d \times 1$) 两个非对称卷积组, 显式地增强了方形卷积核中心骨架位置的权重, 在目标检测中能使矩形框中与物体相关性更高的中心十字区域得到了更多的关注, 加强了视觉特征的辨识力与鲁棒性。

图 1 以 3×3 的卷积核为例, 在训练阶段分别经过所构建的 3 组不同的卷积核后, 再将 3 组计算后的视觉特征融合, 可表示为

$$I \otimes K^{(1)} + I \otimes K^{(2)} + I \otimes K^{(3)} \quad (1)$$

式中: I 表示输入卷积的图像特征; $K^{(1)}$ 、 $K^{(2)}$ 和 $K^{(3)}$ 分别表示 $d \times d$ 、 $1 \times d$ 和 $d \times 1$ 卷积核组; \otimes 表示卷积操作。在训练阶段利用不同形式的额外卷积操作扩充特征空间, 自适应地学习不同长宽比目标。相反地, 添加额外卷积操作不可避免地增加了额外的参数量。但是根据卷积运算的可加性, 利用重参数技巧, 可以将这 3 组卷积操作融合为原始标准卷积大小, 降低推理阶段的计算开销, 3×3 非对称卷积操作的推理阶段如图 2 所示。其通用形式可以表

示为

$$I \otimes (K^{(1)} + K^{(2)} + K^{(3)}) = I \otimes K^{(1)} + I \otimes K^{(2)} + I \otimes K^{(3)} \quad (2)$$

式中, $(K^{(1)} + K^{(2)} + K^{(3)})$ 表示将卷积核中对应位置的权重相加。这意味着, 在推理阶段非对称卷积组被归一成了一个标准卷积, 在不额外增加计算量和参数的同时, 提高了检测模型的效果。

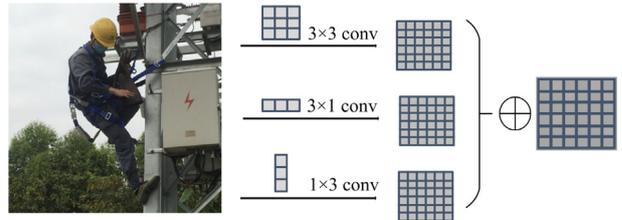


图 1 ACB 训练阶段示意图

Fig. 1 Schematic diagram of ACB module in the training stage

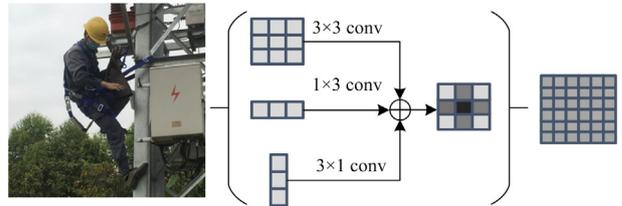


图 2 ACB 推理阶段示意图

Fig. 2 Schematic diagram of ACB module in the inference stage

2.3 自注意力 Transformer 结构

Transformer 结构最早在自然语言处理(natural language processing, NLP)领域中被提出^[12]。该结构利用全局自注意力机制自适应地捕获全局上下文消息, 在多项 NLP 任务中取得卓越成果。而在计算机视觉领域中, 视觉 Transformer 模型成为近年来学者们的热门研究对象^[26-27]。本文在 YOLOv5 模型中添加了文献[27]中的视觉 Transformer 模块, 增强了模型在电力现场检测任务下的长距离上下文建模能力, 进而缓解由物体远近不一致、场景复杂度等噪声干扰所带来的影响。

视觉 Transformer 模块架构如图 3 所示, 针对下采样阶段输出的阶段性特征, 采取与 VIT 中相同的数据处理, 对输入特征进行 Patch Embedding 操作, 即采取图像切分的方式将语义图片从二维结构信息转化为一维序列信息, 以降低计算的复杂度, 同时通过添加位置信息编码以保留图像中各个切片像素的相对位置, 本文在具体实现中将图像特征划分为 14×14 的 Patch 大小。后续的结构包含多头注意力层(multi-head self attention, MSA), 标准化层(layer norm, LN), 多层感知机层(multi-layer perceptron,

MLP)以及残差恒等映射,并在训练过程中添加 Dropout 正则化层来防止网络过拟合。

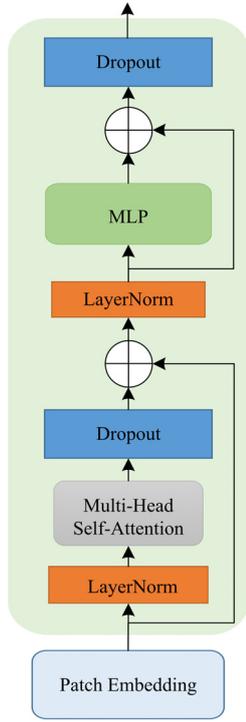


图3 Transformer 模块架构

Fig. 3 Architecture of the Transformer module

多头注意力模块是视觉 Transformer 的核心结构,由多个自注意力机制组成,各个注意力分别关注不同粒度的上下文信息,聚合得到多尺度的注意力信息。自注意力机制由 3 个关键矩阵 Q 、 K 和 V 组成,其获得方式可表示为

$$Q = F^{(l-1)}W^q \quad (3)$$

$$K = F^{(l-1)}W^k \quad (4)$$

$$V = F^{(l-1)}W^v \quad (5)$$

式中: l 为网络深度; $F^{(l-1)}$ 为前一层的特征输出结果; W^q 、 W^k 和 W^v 为可学习的权重矩阵,其目的为对 $F^{(l-1)}$ 进行仿射变换得到 Q 、 K 和 V 。后续利用 Q 和 K 的相关性得到注意力信息,公式可表示为

$$G(Q, K, V) = \sigma\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (6)$$

式中: $G(\cdot)$ 为自注意力机制函数; $\sigma(\cdot)$ 为非线性激活函数 Softmax; d 为通道维度, $1/\sqrt{d}$ 为缩放操作以防止梯度消失。上述介绍了自注意力机制的过程,而多头注意力机制聚合若干个注意力的结果,每个自注意力头可表示为

$$h_i = G(F^{(l-1)}W_i^q, F^{(l-1)}W_i^k, F^{(l-1)}W_i^v) \quad (7)$$

式中,下标 i 表示自注意力头的序号。多头注意力

可表示为

$$O = (h_1 \oplus h_2 \oplus \dots \oplus h_n)W^o \quad (8)$$

式中: \oplus 为特征聚合操作; W^o 为输出层; O 代表输出的特征;下标 n 表示注意力头的总数,本文将 n 设置为 2。

MLP 层由两个全连接层组成, LN 层能够调整输入分布,残差恒等映射构建输入输出的链接关系,防止网络层数过深、梯度消失导致的模型退化。正如文献[26]提及的,Transformer 具有优秀的特征建模能力。本文希望利用 Transformer 的全局注意力与卷积建模领域像素的局部能力相结合互补,将 Transformer 作为卷积网络的补充,以提高整体网络的特征建模关注能力^[28]。

2.4 IYOLOv5

IYOLOv5 的整体模型结构如图 4 所示。本文针对复杂电力场景中可能存在的不规则物体检测以及背景噪声影响,基于 YOLOv5 的网络结构做出了以下两方面的改进。

首先,本文将特征提取主干网络中的标准卷积模块替换为 ACB 结构。通过 ACB 的中心增强机制将权重进行倾斜,针对不规则物体得到更具有辨识性和鲁棒性的视觉特征。该机制能够在推理阶段不额外增加计算开销的情况下,更好地服务于后续的下流任务,以构建更适合不同长宽比物体的检测模型。

其次,本文将 Transformer 结构融于 YOLOv5 中,形成卷积局部关注和全局上下文信息的互补,在保留 CNN 的归纳偏置的同时,也避免了全 Transformer 结构在数据量偏小情况下易出现的过拟合现象。具体实现的细节包括,在特征主干网络的最后输出层添加了 6 层 Transformer 结构,用于对之前所提取的不同尺度特征进行融合,利用大尺度特征中的兴趣区域对小尺度特征图进行特征重构。Neck 层学习的目的是更好地利用 Backbone 提取的多尺度视觉特征,从相邻尺度特征之间捕捉细节部位。因此在特征聚合网络 Neck 中也同样应用了 Transformer 结构,为了避免之前获取的金字塔特征因为连续的自注意力机制出现过拟合现象,在 Neck 融合中的下采样阶段并未应用注意力机制,而是将其应用于后续的上采样层中,利用其对于全局上下文信息的优秀建模能力,自适应地关注与检测物体更加匹配的特征模式,更好地筛选出与检测对象高度相关的特征区域,并通过着重考虑这些区域来提高最终的检测效果。

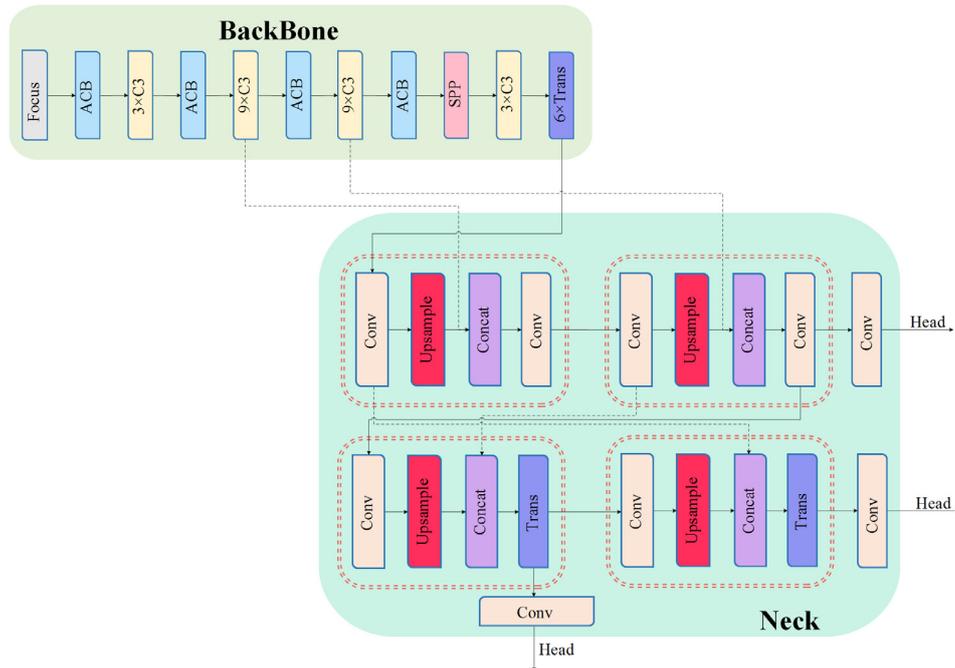


图 4 IYOLOv5 模型结构

Fig. 4 Architecture of the IYOLOv5

IYOLOv5 在网络结构上与原始 YOLOv5 不同，但其训练时损失函数与原始 YOLOv5 保持一致。本文参考 YOLOv5 的参数设置对 IYOLOv5 进行端到端的训练。

3 实验验证

3.1 数据集与实现细节

本文所使用的电力现场穿戴检测数据集来自阿里天池大赛广东电网智能现场运营挑战赛，其检测对象总共有 6 种，包括施工人员、佩戴安全袖章、佩戴绝缘手套、未佩戴绝缘手套、手持操作栏和手持验电笔。该数据集总共有 2561 张图像，其中的标签数目分布如表 1 所示。本文将数据标注格式处理为 Pascal VOC^[29] 的组成格式，并将数据集随机划分成两组，其中 90% 的样本用于训练，而剩余 10% 的样本则作为测试集用于验证模型性能。

本文采用 Pytorch1.7.0 深度学习框架搭建 IYOLOv5 模型，并在 NVIDIA RTX 2080Ti 上完成模型训练。输入图像的大小为 1280×1280×3，设定非极大值抑制和交并比(interaction-over-union, IoU) 的阈值分别为 0.3 和 0.5。优化策略使用 Adam 优化器进行梯度下降，设定批次大小为 2，轮次为 150，初始学习率为 0.001。除此之外，采用余弦模拟退火算法实现对学习率的动态衰减，并在最后一个轮次时将学习率固定为 0.0003。

表 1 数据集标签个数

Table 1 Number of categories in the dataset

标签	个数
佩戴绝缘手套	5094
未佩戴绝缘手套	1871
佩戴安全袖章	796
手持操作杠	863
手持验电笔	801
施工人员	4732

3.2 评估指标

为了对模型性能进行更好的定性评估，本文选取平均精确度均值(mean average precision, MAP)来衡量 IYOLOv5 对电力现场不同类别的总体识别效果。首先需要计算每个类别的平均精确度值(average precision, AP)。AP 的计算方式又依赖于精确率 P 和召回率 R 两项指标， P 和 R 可表示为

$$P = \frac{T_p}{T_p + F_p} \quad (9)$$

$$R = \frac{T_p}{T_p + F_N} \quad (10)$$

式中： T_p 表示预测为正类且实际也为正类的样本个数； F_p 表示预测为正类但实际为负类的样本个数； F_N 表示预测为负类而实际为正类的样本个数。 P 和 R 在某些情况下是互相矛盾的，因此通常计算 $P-R$

曲线下的面积作为 AP 指标来衡量目标检测的性能, 可表示为

$$A_p = \int_0^1 P(R)dR \quad (11)$$

M_{AP} 为多个类别的 AP 值的均值, 可表示为

$$M_{AP} = \frac{\sum_{i=1}^N A_{p_i}}{N} \quad (12)$$

式中: 下标 i 表示类别序号; N 表示数据集中存在的总类别数目。本文在实验过程中所采用的评估指标为 $M_{AP}(0.5)$ 和 $M_{AP}(0.5:0.95)$ 。其中 $M_{AP}(0.5)$ 表示当预测结果与真实目标的 IoU 阈值大于等于 0.5 标注为检测成功; $M_{AP}(0.5:0.95)$ 则表示在不同 IoU 阈值下(从 0.5 至 0.95, 步长为 0.05)的平均值。

3.3 对比实验

本文通过对比现有的目标检测算法, 来验证其在电力现场检测下的优良性能。所对比的方法包括一阶段方法(YOLOv5、SSD^[17]、CenterNet^[30])和二阶段方法(Faster R-CNN^[15])。在本实验中, 根据对比方法的推荐训练参数在本训练集上进行训练。

为了对各个不同的方法进行公平的比较, 首先需要保证每一个方法在训练阶段达到收敛, 不同方法的训练损失曲线如图 5 所示, 观察图 5 可得 SSD、CenterNet、YOLOv5 以及 IYOLOv5 在 20 个 epoch

之后损失曲线都逐渐趋于平稳, Faster R-CNN 在 20 个 epoch 后虽然整体损失曲线有所波动, 但经过平滑处理后整体还是趋于平缓, 从训练曲线中也不难看出, 基于 YOLOv5 的方法在训练中收敛速度也相对较快。测试集 M_{AP} 对比结果如表 2 所示。首先将本文的 IYOLOv5 与改进前的基线模型 YOLOv5 进行对比, 其网络性能好于对比的基线模型, 分别在两项指标上提高了约 15%和 13%, 由此证明了本文所提出的基于 YOLOv5 的改进方法 IYOLOv5 在电力现场环境下的优越性与有效性。对比其他方法, IYOLOv5 的性能也表现得更好, 其主要原因是由于 YOLOv5 模型的良好设计, 以及本文所提出的两项针对电力现场环境的特殊改进。

为了更加直观地表现 IYOLOv5 在电力现场环境下检测的优越性, 随机选取两幅测试图像并分别采用表 2 中的对比方法进行测试, 检测结果如图 6 所示。由图 6 可以发现, IYOLOv5 的表现更好, 尤其是面对远景下的不规则物体, 如第二行的操作杆, IYOLOv5 能够很好地检测识别。综上所述, 在电力现场检测环境下, IYOLOv5 依赖其 ACB 和 Transformer 模块能够更好地建模视觉特征, 在忽视视觉噪声的同时能够更聚焦于检测物体上, 达到更高的检测精度和识别效果。IYOLOv5 更多的可视化结果如图 6 示。

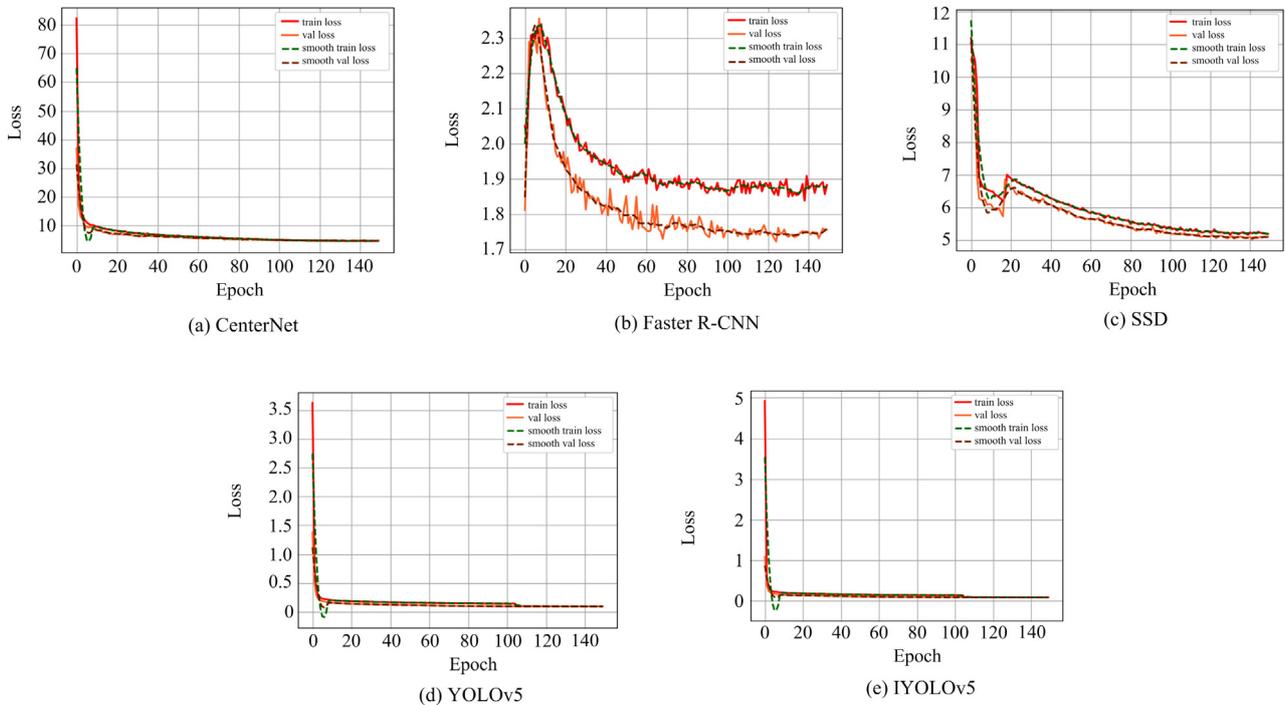


图 5 不同方法训练损失对比

Fig. 5 Training loss comparison results of different algorithms



图6 不同算法检测结果对比

Fig. 6 Detection results comparison of different algorithms

表 2 违规穿戴数据集上不同物体检测器的比较
Table 2 Comparison of different object detectors on the violation of wear dataset

对比方法	$M_{AP} (0.5)$	$M_{AP} (0.5:0.95)$
YOLOv5	51.5	26.2
Faster R-CNN	35.5	16.8
SSD	36.5	19.9
CenterNet	50.2	24.7
IYOLOv5	59.2	29.7

3.4 消融实验

消融实验将分别从两个方面对 IYOLOv5 进行设置, 分别仅保留 ACB 以及 Transformer 模块来训练网络, 以验证每个模块的作用。其结果如表 3 所示。实验表明, 本文所添加的 ACB 与 Transformer 模块皆对电力现场检测有积极作用。其中, ACB 对

表 3 消融实验结果

Table 3 Experiment results of the ablation study

网络设置		$M_{AP} (0.5)$	$M_{AP} (0.5:0.95)$
ACB	Transformer		
×	×	51.5	26.2
√	×	57.4	28.7
×	√	53.5	27.3
√	√	59.2	29.7

网络的提升效果最明显, 鲁棒的视觉特征是目标检测效果好坏的基石。而一旦移除 Transformer 模块, 模型缺乏捕捉全局信息的能力将会导致结果变差。

4 结论

本文以 YOLOv5 为基础, 提出了改进算法 IYOLOv5, 并在电力现场环境下的目标检测任务上取得了优越的结果。首先为了得到更具有辨识性与鲁棒性的视觉特征, 将非对称卷积组添加进 YOLOv5 的特征提取主干网络。然后, 利用 Transformer 的全局注意力机制建模上下文信息, 进而缓解由视觉噪声干扰所带来的负面影响, 能够自适应地关注检测物体更相关的特征区域。实验结果表明, 本文所提出的 IYOLOv5 在电力现场检测任务中表现卓越, 尤其是针对复杂环境和物体下具有更强的辨识能力与抗干扰能力, 表现出良好的泛化能力, 为电网的智能化改革提供了解决方案。

参考文献

[1] 吕学宾, 李岩, 李英, 等. 变电站施工作业人员安全管理及评价系统研究[J]. 电力系统保护与控制, 2021, 49(4): 21-27.
LÜ Xuebin, LI Yan, LI Ying, et al. Research on safety management and an evaluation system for substation

- construction personnel[J]. *Power System Protection and Control*, 2021, 49(4): 21-27.
- [2] HU Zhuangli, HE Tong, ZENG Yihui, et al. Fast image recognition of transmission tower based on big data[J]. *Protection and Control of Modern Power Systems*, 2018, 3(2): 149-158.
- [3] 魏贤哲, 卢武, 赵文彬, 等. 基于改进 Mask R-CNN 的输电线路防外破目标检测方法研究[J]. *电力系统保护与控制*, 2021, 49(23): 155-162.
WEI Xianzhe, LU Wu, ZHAO Wenbin, et al. Target detection method for external damage of a transmission line based on an improved Mask R-CNN algorithm[J]. *Power System Protection and Control*, 2021, 49(23): 155-162.
- [4] 林晓林, 孙俊. 基于机器学习的小目标检测与追踪的算法研究[J]. *计算机应用研究*, 2018, 35(11): 3450-3453.
LIN Xiaolin, SUN Jun. Research on small object detection and tracking algorithm based on machine learning[J]. *Application Research of Computers*, 2018, 35(11): 3450-3453.
- [5] 冯国臣, 陈艳艳, 陈宁, 等. 基于机器视觉的安全帽自动识别技术研究[J]. *机械设计与制造工程*, 2015, 44(10): 39-42.
FENG Guochen, CHEN Yanyan, CHEN Ning, et al. Research on automatic recognition technology of safety helmet based on machine vision[J]. *Machine Design and Manufacturing Engineering*. 2015, 44(10): 39-42.
- [6] 吴晖, 李铭钧, 杨英仪, 等. 面向变电站视频监控终端的目标检测方法及其优化[J]. *广东电力*, 2019, 32(9): 62-68.
WU Hui, LI Mingjun, YANG Yingyi, et al. An object detection method and its optimization for substation video surveillance terminals[J]. *Guangdong Electric Power*, 2019, 32(9): 62-68.
- [7] 曾纪钧, 温柏坚, 梁哲恒. 基于深度学习的安全帽识别算法研究与模型训练[J]. *电力系统保护与控制*, 2021, 49(21): 107-112.
ZENG Jijun, WEN Bojian, LIANG Zheheng. Research and model training of helmet recognition algorithm based on deep learning[J]. *Power System Protection and Control*, 2021, 49(21): 107-112.
- [8] 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*, June 27-30, 2016, Las Vegas, Nevada, USA: 779-788.
- [9] 王纪武, 罗海保, 鱼鹏飞, 等. 基于 Faster R-CNN 的多尺度高压塔鸟巢检测[J]. *北京交通大学学报*, 2019, 43(5): 37-43.
WANG Jiwu, LUO Haibao, YU Pengfei, et al. Bird's nest detection in multi-scale of high-voltage tower based on faster R-CNN[J]. *Journal of Beijing Jiaotong University*, 2019, 43(5): 37-43.
- [10] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(6): 1137-1149.
- [11] DING Xiaohan, GUO Yuchen, DING Guiguang, et al. Acnet: strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks[C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, June 16-20, 2019, Long Beach, California, USA: 1911-1920.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C] // *31st Annual Conference on Neural Information Processing Systems*, December 4-9, 2017, Long Beach, California, USA: 6000-6010.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 23-28, 2014, Columbus, Ohio, USA: 580-587.
- [14] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. *International Journal of Computer Vision*, 2013, 104(2): 154-171.
- [15] GIRSHICK R. Fast R-CNN[C] // *Proceedings of the IEEE International Conference on Computer Vision*, December 7-13, 2015, Santiago, Chile: 1440-1448.
- [16] CAI Zhaowei, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 18-22, 2018, Salt Lake City, Utah, USA: 6154-6162.
- [17] LIU Wei, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C] // *European Conference on Computer Vision*, October 8-16, 2016, Amsterdam, Holland: 21-37.
- [18] TAN Mingxing, PANG Ruoming, LE Q V. Efficientdet: scalable and efficient object detection[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 14-19, 2020, Seattle, Washington, USA: 10781-10790.
- [19] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C] // *European Conference on Computer Vision*, August 23-28, 2020, Glasgow, US: 213-229.

[20] 施保华, 井任月, 杨超, 等. 用于变电站保护压板状态识别的增强 YOLO 网络[J]. 电力系统保护与控制, 2021, 49(23): 163-170.
SHI Baohua, JING Renyue, YANG Chao, et al. Enhanced YOLO network for status recognition of a substation protection plate[J]. Power System Protection and Control, 2021, 49(23): 163-170.

[21] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[C] // Advances in Neural Information Processing Systems, December 7-12, 2015, Montreal, Canada: 2017-2025.

[22] WANG Chienyao, LIAO Hongyuan, WU Yuehua, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, June 14-19, 2020, Seattle, Washington, USA: 390-391.

[23] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, Nevada, USA: 770-778.

[24] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[25] YANG Changchun, GAO Fei. EDA-Net: dense aggregation of deep and shallow information achieves quantitative photoacoustic blood oxygenation imaging deep in human breast[C] // International Conference on Medical Image Computing and Computer-Assisted Intervention, October 13-17, 2019, Shenzhen, China: 246-254.

[26] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: hierarchical vision transformer using shifted windows[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, June 19-25, 2021, Montreal, QC, Canada: 10012-10022.

[27] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C] // 9th International Conference on Learning Representations, May 6-9, 2021, New Orleans, USA: 1-21.

[28] 田永林, 王雨桐, 王建功, 等. 视觉 Transformer 研究的关键问题: 现状及展望[J]. 自动化报, 2022, 48(4): 957-979.
TIAN Yonglin, WANG Yutong, WANG Jianguo, et al. Key problems and progress of vision Transformers: the state of the art and prospects[J]. Acta Automatica Sinica, 2022, 48(4): 957-979.

[29] EVERINGHAM M, VAN GOOL L, WILLIAMS CKI, et al. The Pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[30] DUAN Kaiwen, BAI Song, XIE Lingxi, et al. Centernet: Keypoint triplets for object detection[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, June 16-20, 2019, Long Beach, California, USA: 6569-6578.

收稿日期: 2022-11-09; 修回日期: 2023-02-09

作者简介:

杨乐(1985—), 男, 本科, 高级工程师, 研究方向为电力物联网、云计算、人工智能;

郭一鸣(1994—), 男, 硕士研究生, 工程师, 研究方向为电力物联网、云计算、人工智能;

张志宏(1986—), 男, 通信作者, 博士生导师, 研究方向为人工智能、复杂网络分析和以新能源为主体的新型电力系统。E-mail: zhihong@xmu.edu.cn

(编辑 周金梅)