

DOI: 10.11992/tis.202303030

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20230802.1217.009>

融合知识迁移和改进 YOLOv6 的变电设备 热像检测方法

赵振兵^{1,2,3}, 冯烁¹, 赵文清^{2,4}, 翟永杰⁴, 王洪涛^{2,4}

(1. 华北电力大学 电子与通信工程系, 河北 保定 071003; 2. 华北电力大学 复杂能源系统智能计算教育部工程研究中心, 河北 保定 071003; 3. 华北电力大学 河北省电力物联网技术重点实验室, 河北 保定 071003; 4. 华北电力大学 控制与计算机工程学院, 河北 保定 071003)

摘要: 针对变电设备热像检测中存在复杂背景样本不足和设备定位困难的问题, 提出了融合知识迁移和改进 YOLOv6 的变电设备热像检测方法。针对复杂背景样本不足问题, 使用扩散模型从域外数据中提取背景知识来生成背景图像, 将设备样本迁移到背景图像中生成人工热像; 针对设备难以被精准定位的问题, 将多头注意力机制和显示视觉中心模块融入 YOLOv6 模型, 改善模型的特征提取能力。实验结果表明, 该方法对设备的平均检测准确率达到 86.4%, 召回率达到 89.4%, 相较于基线模型分别提升了 3.1% 和 1.5%, 为变电设备热像检测提供了新的实现方法。

关键词: 变电设备; 热红外图像; 知识迁移; 样本生成; 目标检测; 扩散模型; 数据扩增; 深度学习

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2023)06-1213-10

中文引用格式: 赵振兵, 冯烁, 赵文清, 等. 融合知识迁移和改进 YOLOv6 的变电设备热像检测方法 [J]. 智能系统学报, 2023, 18(6): 1213-1222.

英文引用格式: ZHAO Zhenbing, FENG Shuo, ZHAO Wenqing, et al. Thermd image detection method for substation equipment by incorporating knowledge migration and improved YOLOv6[J]. CAAI transactions on intelligent systems, 2023, 18(6): 1213-1222.

Thermd image detection method for substation equipment by incorporating knowledge migration and improved YOLOv6

ZHAO Zhenbing^{1,2,3}, FENG Shuo¹, ZHAO Wenqing^{2,4}, ZHAI Yongjie⁴, WANG Hongtao^{2,4}

(1. Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China; 2. Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education, North China Electric Power University, Baoding 071003, China; 3. Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding 071003, China; 4. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: To address the problems of insufficient complex background samples and difficulty in device location in substation equipment thermal image detection, a fusion knowledge transfer and improved YOLOv6 detection method are proposed. The diffusion model was used to extract background knowledge from extraterritorial data for generating background images, solving the problem of insufficient complex background samples. The device samples were then migrated to the background images to generate artificial images. The multi-head self-attention mechanism and explicit visual center module were integrated into YOLOv6 to improve its feature extraction capability, solving the issue of difficulty in detecting devices. The experiment shows that the mAP and mAR of the proposed method reach 86.4% and 89.4%, indicating an improvement of 3.1% and 1.5% compared to the baseline model, respectively. This study provides a new implementation method for thermal image detection of substation equipment.

Keywords: substation equipment; thermal infrared image; knowledge migration; sample generation; object detection; diffusion model; data augmentation; deep learning

收稿日期: 2023-03-22. 网络出版日期: 2023-08-03.

基金项目: 国家自然科学基金项目 (61871182, U21A20486);
河北省自然科学基金项目 (F2020502009, F2021-502008, F2021502013).

通信作者: 赵振兵. E-mail: zhaozhenbing@ncepu.edu.cn.

在“双碳”和“新基建”双重背景下, 智能电网在促进中国经济发展与能源绿色转型中发挥着重要作用^[1]。近年来, 快速发展的人工智能技术为

电力设备运行状态的评价提供了新的理论与关键技术支撑,如何利用人工智能等数字化技术实现对新型电力系统的高效运维这一问题亟待解决^[2]。变电站作为电力系统重要组成部分,其运行状态直接关系到整个电网的安全可靠运行,为了及时并准确获取反应变电设备运行状态的图像信息,必须完成对其图像的智能处理^[3]。红外热像检测技术因具有不接触和操作简单的优势,在变电设备温度状态监测领域中有着广泛应用。随着无人值守变电站的普及,远程图像监测系统、无人机和巡检机器人等设备已经被大量应用,这些设备在运行时会产生大量红外图像,仅依靠人工对图像进行处理,存在漏检的问题且效率低^[4-6]。基于深度学习的图像处理技术逐渐成为对变电设备红外热像进行智能分析的有力工具,为变电站的智能化巡检提供了可靠稳定的技术支持,因此对多类变电设备热像进行快速精准检测的研究具有极高价值。

目前,电力领域中绝大部分的图像数据并未被公开,已有的变电站设备检测工作^[7-12]都是基于少量私有图像完成的。为了解决样本不足问题,多种扩增图像数据的方法被引入到电力领域:文献[7-8]通过图像的镜像、翻转,以及改变原始图像的大小、亮度和模糊度来扩增训练数据集;文献[9-12]通过调整图像的饱和度、曝光度和色调来生成更丰富的样本。但以上方法仍存在以下问题:通过旋转等方法虽提升了训练图像的数量,却难以提升训练数据的信息量;在人工、无人机或巡检机器人采集图像的过程中,均力图使变电设备在图像中保持竖直、清晰的状态,而使用旋转、翻折等方法扩充的训练图像,与真实图像的风格并不相符。

目前针对变电设备红外热像检测的研究普遍涉及到对深度学习模型进行改进,通过改进深度学习模型自身的检测性能来提升对变电设备的识别准确率,进而获得变电设备在红外热像中准确的类别信息、区域信息和温度信息。文献[7]首先将变电设备热像由伪色图转换为灰度图,并将YOLOv4^[13]目标检测模型的骨干网络维度压缩为原网络的80%来减少计算量,最后使用改进的YOLOv4模型在灰度图上进行设备检测和温度信息提取。文献[8]提出使用改进的残差网络来替换CenterNet^[14]的骨干网络,并调整了骨干网络中激活函数的数量,提升了对变电设备热像的检测精度。文献[9]使用Mask R-CNN^[15]算法检测热

像中的绝缘子区域,同时利用热像中的温度信息对绝缘子的状态进行评估。文献[10]提出在浅层网络中进行特征增强改进了FSSD^[16]网络,并根据变电站中绝缘子的长宽比重新设计了原算法的锚框,最后通过在公共数据集上预训练模型实现了迁移学习,提升了对绝缘子热像的检测精度。文献[11]将Faster R-CNN^[17]模型中较深的卷积层进行了删减,提升了训练和检测速度,在模型中增加了两种新型锚框来提升对细长型变电设备的检测精度,通过对五类设备热像的检测实现对设备温度状态的监测。文献[12]提出首先使用DenseNet^[18]的网络结构替换YOLOv4模型的部分骨干网络,修改特征融合网络以改变信息流,然后利用改进的YOLOv4网络对变电设备的可见光图像进行检测,最后通过图像配准将检测结果映射到热像中,实现对变电设备的温度状态评估。

从上述工作中可以看出,对深度学习模型的结构进行改进,并基于变电设备的热像对模型进行参数调优,是较为理想的热像检测方案。但是电力领域的图像数据因保密性要求不能公开,使得现有研究面临图像样本缺乏的问题,尤其是复杂背景的热像样本。此外,不同的变电设备往往具有外形类似的主体结构,使其外观之间存在较高的视觉相似性,相比于可见光图像,这一现象在缺失细节信息的红外热像中更为显著,增大了检测热像中变电设备的难度。

本文基于以上问题进行研究,针对变电设备热像检测任务中存在的复杂背景样本不足问题,本文提出基于知识迁移的红外样本生成方法,使用去噪扩散概率模型(denoising diffusion probabilistic model, DDPM)^[19]提取出域外数据中的背景知识,将变电设备在真实样本中的位置知识迁移到背景知识,从而生成具有较复杂背景的人工热像;针对热像中设备外观相似度高和细节信息丢失的问题,本文将多头注意力(multi-head self-attention, MHSA)机制^[20]融入到YOLOv6^[21]模型的骨干网络中,增强模型对变电设备的特征提取能力,将包含多层感知机(multiple perceptron, MLP)^[22]和可学习视觉中心(learnable vision center, LVC)机制^[23]的显示视觉中心(explicit visual center, EVC)模块引入到YOLOv6模型的特征融合网络,MLP可以从最深层特征中学习全局长距离依赖关系,LVC负责捕获输入图像的局部区域信息。结合基于知识迁移的样本生成方法和改进的YOLOv6模型,显著提升了变电设备热像的检测精度。

1 基于知识迁移的样本生成方法

针对部分热红外图像背景较复杂影响设备检测精度的问题, 样本扩充能够有效提升模型训练效果, 现有研究普遍依靠变电设备图像的简单变换来扩充样本, 为实现更高质量的样本扩增, 本文提出基于知识迁移的样本生成方法, 如图 1 所示。

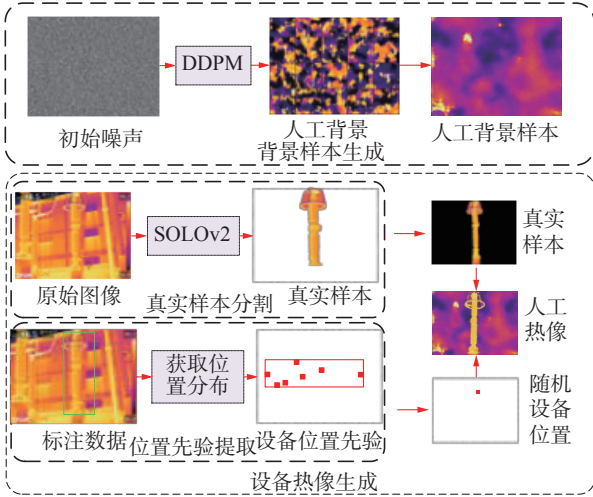


图 1 样本生成方法

Fig. 1 Method of sample generation

本文将样本生成问题分解为背景生成和设备热像生成两个部分, 两个部分均依靠知识迁移来完成: 在背景生成过程中, 使用 DDPM 模型从域外数据中提取背景知识, 模型训练完成后, 依靠背景知识的指导可生成大量背景图像, 实现了由真实热红外图像中到人工背景样本的知识迁移; 在设备热像生成过程中, 基于真实样本中设备的位置先验知识, 以知识迁移的形式将真实设备样本融合到人工背景样本中, 完成高质量的热像样本扩增。

1.1 基于 DDPM 模型的背景生成方法

去噪扩散概率模型 DDPM 于 2020 年被提出用于图像生成工作, 该模型使用随机噪声来生成图像, 该模型相比于对抗生成网络有着训练更稳定的优点, 在公共数据集上表现出较好的效果。

扩散模型分为扩散过程和去噪过程。扩散过程的原理是分 t 次为输入图像 x_0 添加高斯噪声, 直至将图像破坏为随机噪声 x_t , 如图 2 所示。

每一次添加噪声后输出的图像仅由输入图像决定, 所以扩散过程符合马尔可夫链过程。第 t 次加噪声的过程可以表示为

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t) \quad (1)$$

式中: x_{t-1} 为输入图像, x_t 为输出图像。由输入得到输出的过程 $q(x_t | x_{t-1})$ 满足以 $\sqrt{1-\beta_t}x_{t-1}$ 为均值,

β_t 为方差的高斯分布。在训练中, 因 β_t 和 x_{t-1} 等参数均不参与训练, 所以扩散过程不是一个可学习过程。从 x_{t-1} 得到 x_t 的过程表示为

$$x_t = \alpha_t x_{t-1} + \beta_t \epsilon_t, \epsilon_t \sim N(0, 1) \quad (2)$$

其中: α_t 表示 $\sqrt{1-\beta_t}$, ϵ_t 取值服从标准正态分布。训练过程中模型需要学习的是, 在由 x_{t-1} 可以得到 x_t 的前提下, 将 x_t 恢复为 x_{t-1} 。由式 (2) 可得:

$$x_{t-1} = \frac{1}{\alpha_t} (x_t - \beta_t \epsilon_\theta(x_t, t)) \quad (3)$$

其中 $\epsilon_\theta(x_t, t)$ 是网络预测出的噪声, 用来模拟在扩散过程中添加到 x_t 中的 ϵ_t 噪声, 显而易见, 当 ϵ_θ 与 ϵ_t 足够接近后, 模型就可以从一个初始噪声 x_T 经过 T 次去噪过程得到一个真实图像 x_0 。如图 2 所示, 在模型的实际训练过程中, 对 ϵ_θ 的预测由一个 U-net 网络^[24]来完成, 由 ϵ_θ 和 ϵ_t 二者之间的欧氏距离作为损失函数:

$$L = \|\epsilon - \epsilon_\theta(x_t, t)\|^2 \quad (4)$$

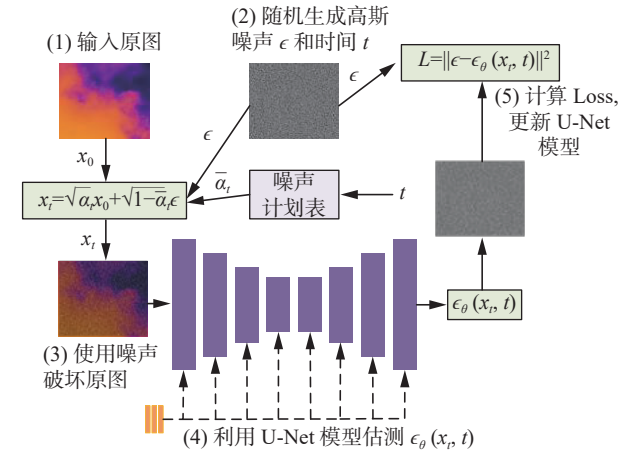


图 2 DDPM 模型结构

Fig. 2 Structure of DDPM model

本文使用课题组收集的域外数据来训练 DDPM 模型, 域外数据由 712 张绝缘子热红外图像构成, 其不参与后续目标检测模型的训练。

图 3 给出了 DDPM 模型生成红外背景图像的可视化过程, 经过训练的 DDPM 图像能够利用域外数据中的背景知识, 从初始化的随机噪声中逐步去噪, 最终生成人工背景样本。

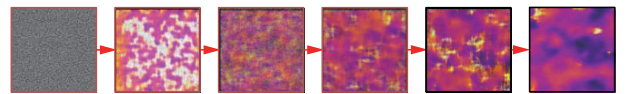


图 3 DDPM 模型生成过程可视化

Fig. 3 Visualization of the DDPM model generation process

1.2 基于知识迁移的设备样本生成方法

考虑到现有图像生成方法难以直接生成变电设备图像这一问题, 本文提出从真实样本中提取

多种变电设备的图像和位置分布先验知识, 结合 DDPM 模型生成的背景图像本, 来合成人工热像。

首先, 利用图像实例分割算法 SOLOv2^[25] 从原始图像中提取出多种设备的图像样本, 如图 4 所示, 图 4(a)~(d) 分别为分割后的电流互感器、电压互感器、避雷器和套管样本。



图 4 4 类变电设备图像

Fig. 4 Images of four types of substation equipment

其次, 对用来训练目标检测模型的数据集进行分析, 来获取不同设备在热像中的位置分布情况。该过程的原理是借助标注过程, 实现对图像中设备种类和位置信息的标识, 通过统计不同设备在数据集中的矩形框坐标, 得到每类设备在图像中出现的位置范围, 用矩形框的中心点坐标 (x, y) 表示设备在图像中出现的位置, 数据集中每张包含该设备的图像均可提供至少一组中心点坐标 (x, y) , 以避雷器为例, 遍历数据集后即可得到图像中避雷器的中心点坐标集合 L , 利用 L 中 x 与 y 的极大(小)值可以确定出一个矩形区域, 如图 1 所示。该区域的范围将作为避雷器在原始数据集集中的位置先验知识, 在生成人工样本的过程中迁移到新数据集中, 具体方法如下: 在已经得到人工背景样本和设备真实样本的情况下, 设备真实样本在人工背景样本中的中心点坐标 $N(X_N, Y_N)$ 将从位置先验中产生, 矩形区域可由 $P_1(X_M, Y_M)$ 、 $P_2(X_{MAX}, Y_{MAX})$ 表示, 其中 P_1, P_2 为矩形区域的左下角和右上角, X_M, X_{MAX} 和 Y_M, Y_{MAX} 为矩形区域中横纵坐标的极值, X_N, Y_N 将从矩形区域中随机产生:

$$X_N = \text{Rand}(P_1, P_2) = \text{Rand}(X_M, X_{MAX}) \quad (5)$$

$$Y_N = \text{Rand}(P_1, P_2) = \text{Rand}(Y_M, Y_{MAX}) \quad (6)$$

在位置先验知识约束下, 将设备真实样本融合到人工背景图像中, 即可实现对不同变电设备的样本自动扩增。

2 基于改进 YOLOv6 的设备检测方法

YOLOv6 是美团在 2022 年提出的目标检测模型, 模型由输入端(input)、骨干网络(backbones)、颈部(neck)和检测头(head)构成。如图 5 所示, YOLOv6 的骨干网络是基于 RepVGG^[26] 的轻量化

网络 EfficientRep; 颈部网络为 Rep-PAN 结构; 检测头网络为 3 个解耦检测头, 每个检测头中存在两个检测分支, 分别实现分类(Cate)和定位(Box)。

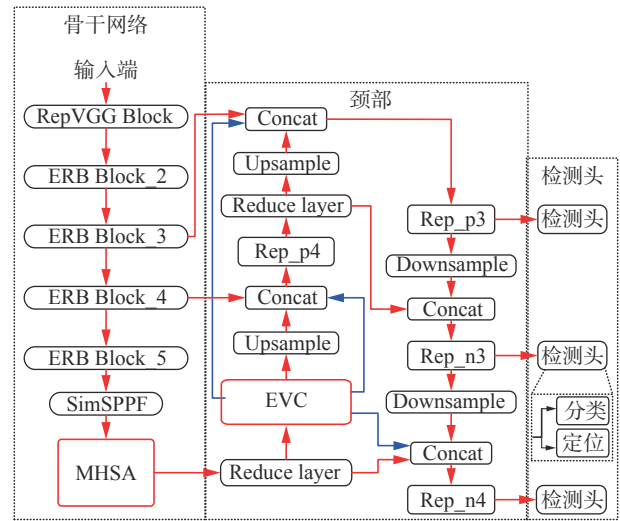


图 5 改进 YOLOv6 结构图

Fig. 5 Structure diagram of the improved YOLOv6

本文在 YOLOv6 的 EfficientRep 骨干网络上进行改进, 在其深层网络中加入了 MHSA 层来提升模型的特征提取能力; 在 Rep-PAN 颈部网络中加入了显性视觉中心模块, 从深层特征中获得视觉中心信息以改善浅层特征, 增强颈部网络的特征融合能力。

2.1 融合 MHSA 机制的骨干网络

相比于卷积神经网络(convolutional neural network, CNN) 重点关注图像的局部区域, MHSA 机制可以捕获图像数据的全局相关性, 能够表示数据之间更丰富的联系。如图 6 所示, 输入特征的高度、宽度和维度分别为 H, W 和 d , 将输入特征在 W_V, W_K 和 W_Q 中分别进行逐点卷积, 得到 Q, K, V (Q 为查询向量、 K 为键向量、 V 为值向量), 同时, 两个可学习的参数向量 R_h 和 R_w 分别表示对特征图中不同位置的高度和宽度进行位置编码, 将 R_h 和 R_w 通过广播机制扩增到 $H \times W \times d$ 的形状后进行相加, 得到相对位置编码的输出 R 。将 Q 分别与 R 和 K 进行相乘后, 得到两个宽高均为 $H \times W$ 的矩阵, 将二者相加后进行归一化(softmax)操作, 将输出结果与 V 进行相乘, 得到最终输出结果。以上过程独立重复 h 次(本文中 h 为 4), 最后将 h 次的结果进行拼接, 再进行一次线性变换得到 MHSA 模块的最终输出。

将 MHSA 机制与基于 CNN 的骨干网络相结合, 能够有效提升其特征提取能力, 在这方面较为经典的研究工作是 BoTNet^[27] 中将 MHSA 与

ResNet^[28] 相结合的做法, 受此启发, 本文将 MHSA 机制融合到 EfficientRep 网络的深层结构中。加入位置如图 5 中所示, MHSA 层的输入与输出维度相同, 均为 512。

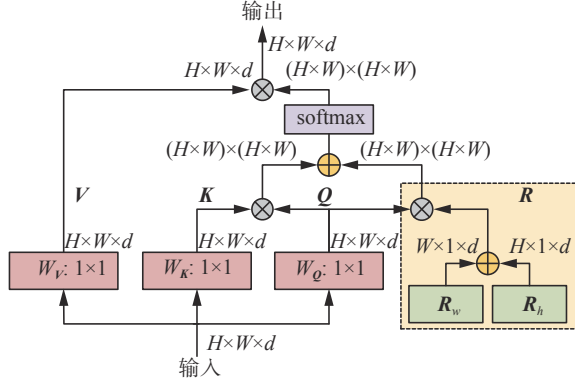


图 6 MHSA 机制原理

Fig. 6 Schematic of the MHSA mechanism

2.2 显性视觉中心模块

目前 FPN^[29] 网络已成为目标检测模型中常见的颈部网络, 用来融合不同层级的特征图, 但现有 FPN 网络的功能主要集中在实现层间特征的相互作用, 但忽略了层内特征的规则, 而这些规则已经被证明有利于视觉识别任务。针对如何利用深层特征的层内特征来改善浅层特征的问题, 文献 [23] 在 YOLOX^[30] 的颈部网络中进行了尝试, 提出使用轻量化多层感知机 (lightweight multiple perceptron, LMLP) 和可学习的视觉中心 (learnable vision center, LVC) 机制来构建显性视觉中心 (explicit visual center, EVC) 模块: 其中 LMLP 用来从特征图中计算全局长程相关性, LVC 和 LMLP 共同使用, 用来聚焦输入图像更多的局部区域, 实现对深层特征图的特征信息增强, 借助颈部网络的特征融合路径, 用增强后的深层特征改善浅层特征。

如图 7 所示, 骨干网络输入 EVC 的特征图, 首先进入 Stem Block 模块, 借助卷积核大小为 7、个数为 256 的卷积层, 将输入特征图的维度由 512 维降至 256 维, 降维后的特征图输入到 LMLP 和 LVC 模块中。

LMLP 模块主要由两个残差模块构成, 特征图输入到第一个残差模块时, 先经过群组归一化 (group normalization), 随后输入逐通道卷积 (depth-wise convolution) 层, 逐通道卷积在提升特征表征能力的同时可以降低计算量, 随后在经过通道缩放 (channel scaling) 与路径丢弃 (droppath) 后, 与原输入通过残差路径进行相加, 随后将特征图输入第二个残差模块, 在经过群组归一化后, 在特征

图上实现通道 MLP(channel MLP), 最后经过通道缩放与正则化后输出特征图。

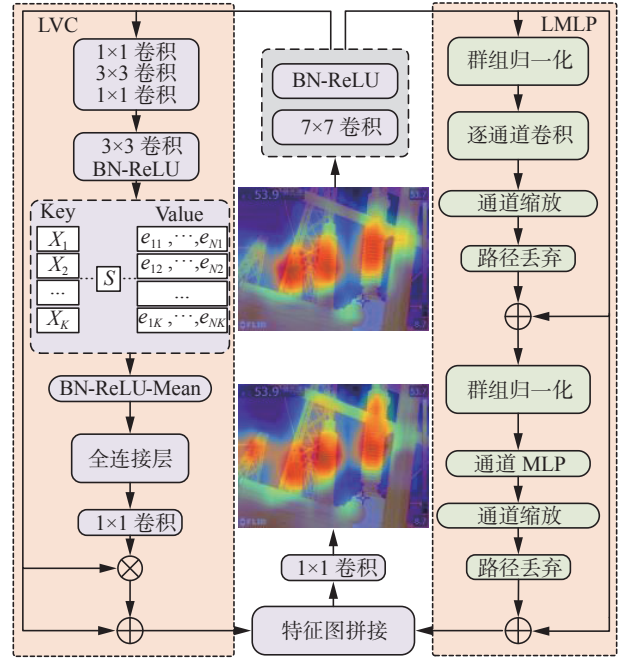


图 7 EVC 模块结构

Fig. 7 Structure of EVC module

LVC 模块首先使用一组卷积层的组合 (包括 1×1 卷积和 3×3 卷积) 对输入特征图进行编码, 随后将编码特征送入一个码本 (Codebook), 码本中包含固有码字 $B = \{b_1, b_2, \dots, b_K\}$ 和一组缩放因子 $S = \{s_1, s_2, \dots, s_K\}$, 输入的编码特征为 $X = \{X_1, X_2, \dots, X_K\}$, 通过缩放因子 s_k 使 x_i 和 b_k 映射相应的位置信息 e_k :

$$e_k = \sum_{i=1}^N \frac{e^{-s_k \|x_i - b_k\|^2}}{\sum_{j=1}^K e^{-s_k \|x_i - b_j\|^2}} (x_i - b_k) \quad (7)$$

式中: x_i 为输入特征图的第 i 个像素, b_k 是第 k 个可学习的视觉码字, s_k 是第 k 个缩放因子。 $x_i - b_k$ 是每个像素相对于码字的位置信息。 K 是视觉中心的总数。整个码本的输出 e 可以表示为

$$e = \sum_{k=1}^K \text{Relu}(e_k) \quad (8)$$

将码本的输出输入到全连接层和 1×1 卷积层来预测能够突出类别的特征, 最后将输出特征与输入 LVC 的特征图先后进行通道上的相乘和相加, 作为 LVC 模块的输出。将 LMLP 和 LVC 的输出特征图拼接之后, 利用 1×1 卷积层将特征图维度调整到 256, 与输入特征图保持一致。

如图 8 所示, 在文献 [23] 的基础上, 将显性视觉中心模块构建在了 Rep-PAN 这一双向特征融合金字塔结构上。

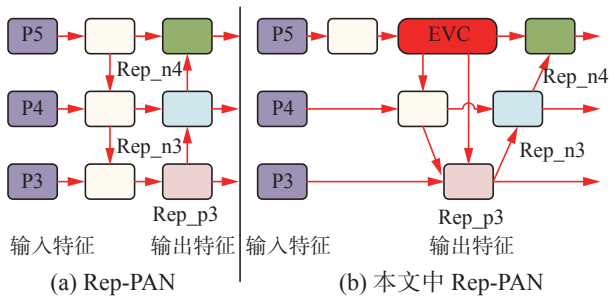


图8 颈部网络结构图对比

Fig. 8 Comparison of structural diagrams of the neck network

显式视觉中心模块利用 Rep-PAN 双向融合的特性,在改善浅层特征的同时,利用自底向上的路径改善了深层特征。

3 实验结果及分析

本文中的实验在操作系统为 Ubuntu18.04 LTS 的计算机上完成,其中 CPU 为 I9-10900X, GPU 为 NVIDIA RTX A5000。使用的深度学习框架是 Pytorch1.8.0。模型的训练轮次(epoch)为 100,训练时的批处理(batch size)参数为 16。

本文选取了 9 类变电设备的热红外图像作为实验对象,其中原始训练集和验证集样本图像分别为 903 张和 152 张,在数据集扩增阶段,利用本文提出的样本生成方法,对 4 类背景较为复杂的设备热像进行了训练集扩充,如表 1 所示,在数据扩增后,训练集图像为 1347 张,验证集图像数量不变。

表 1 数据集详细情况
Table 1 Dataset details

设备种类	原始图像数量	扩增后数量	标签名称
避雷器-1	266	342(+76)	Lightning Arrester
避雷器-2	32	—	TLT
避雷器-3	35	—	FLT
电流互感器-1	345	436(+91)	Current Transformer
电流互感器-2	52	—	TCT
电压互感器-1	137	208(+71)	Voltage Transformer
电压互感器-2	63	—	TVT
套管	100	154(+54)	Bushing
流变侧套管	25	—	LBCTG

本文使用 AP(average precision) 以及 mAP(mean average precision) 作为评价标准,AP 由召回率 $R(\text{recall})$ 、准确率 $P(\text{precision})$ 计算得到,除 AP 和 mAP 外,同时使用了 AP^{50} 和 AP^{75} 作为指标, AP^{50} 、 AP^{75} 指的是计算预测框和标注框的交

并比后,认为交并比分别在 50%、75% 以上时预测框为正确结果,并计算对应 AP 值。

3.1 消融实验

为了验证本文方法中每个模块的有效性,本文设计了两组共 8 个消融实验。为了验证本文提出的样本生成方法的有效性,依据训练模型时使用的数据集,将实验分为使用原始训练集和使用扩增后的训练集两组。在每组实验中设置了 4 个实验模型:模型 A 为 YOLOv6 基线模型;模型 B 为具有 MHSA 机制的 YOLOv6 模型;模型 C 为具有 EVC 机制的 YOLOv6 模型;模型 D 为同时具有 MHSA 和 EVC 机制的 YOLOv6 模型。表 2 给出了消融实验结果。

表 2 消融实验结果
Table 2 Ablation test results %

模型	数据扩增	mAP	AP^{75}	AP^{50}
A		83.3	94.3	96.1
A	√	84.3	96.0	97.2
B		83.9	93.1	95.7
B	√	85.5	95.3	97.1
C		83.7	92.2	96.2
C	√	84.7	94.0	95.5
D		84.3	92.4	96.6
D	√	86.4	96.1	97.8

从表 2 中可以看出,在 YOLOv6 模型中使用本文中相应的方法均能提升模型的检测准确率。与原始训练集相比,仅使用本文提出的样本生成方法来扩增数据集,在基线模型上,mAP、 AP^{75} 分别提升了 1%、1.7%;在最终改进后的 YOLOv6 模型上,mAP、 AP^{75} 分别提升了 2.1%、3.7%,均有效提升了模型的检测精度。在使用扩充训练集的前提下,与 YOLOv6 基线模型相比,分别使用 MHSA 机制和 EVC 机制能够使 mAP 提升 1.2% 和 0.4%,而同时使用两个方法能够使 mAP 提升 2.1%,因为 MHSA 和 EVC 是针对 YOLOv6 模型不同的部件进行改进的,将 MHSA 机制融入到骨干网络的最后一层,有效改善了深层特征的提取效果,EVC 机制在颈部网络中进一步增加了深层特征的显式信息,进而提升了改进模型的检测精度。

为了更明确地体现本文中样本生成方法的作用,表 3 给出了经过数据扩增的 4 种设备热像的精度指标 AP,可以看出,使用本文提出的样本生成方法来扩充训练数据集,能够有效提升模型的检测精度,在模型 A 的实验结果中,仅套管(Bushing)的检测精度低于扩增前的结果,而使用扩增

训练集进行训练的模型 D, 对 4 类设备的检测精度均超过了未进行数据扩增的模型 D。

表 3 经过数据扩增的设备热像检测精度 AP
Table 3 Device thermal image detection accuracy after data amplification %

模型	数据 扩增	Lightning Arrester	Current Transformer	Voltage Transformer	套管
A		90.2	78.2	77.8	78.1
A	√	90.9	80.0	81.7	76.4
D		90.4	77.2	78.7	70.2
D	√	91.2	82.0	80.7	79.3

表 4 给出了在实验中始终没有进行数据扩增的五类设备热像检测精度, 可以看出这五类设备虽然数据量较少, 但其因背景较为简单, 从热像中检测设备的难度较低, 在不同实验条件下, 检测精度未出现明显波动。

表 4 数据未扩增的设备热像检测精度 AP
Table 4 Device thermal image detection accuracy for non-amplified data %

模型	数据 扩增	TLT	FLT	TCT	TVT	LBCTG
A		93.8	91.5	92.0	77.6	70.3
A	√	86.0	91.3	93.4	83.0	76.2
D		91.9	93.9	92.1	86.8	77.8
D	√	89.7	93.9	89.4	84.5	86.6

从表 4 中可以看到, 随着模型改进和数据集扩增, 数据量并未发生变化的流变侧套管 (LBCTG) 的检测精度一直在上升, 模型 D 与扩增数据集的组合, 使得 FLT 和 TVT 的检测精度也高于基线模型 A。虽然对 TCT 和 TLT 的检测精度出现了下降, 但从模型 A 的角度来看, 扩增后的数据集对 TCT 的精度有提升, 对比模型 A 和模型 D 对 TLT 的检测精度可知, 仅使用模型 D 造成了对 TLT 检测精度的下降, 因此本文认为 TCT 和 TLT 的最终精度下降或由模型 D 造成。

3.2 模型检测结果可视化对比实验

为了定性分析改进模型的检测性能, 使用 GradCAM 热力图^[31]可视化方法, 对基线模型和改进模型的预测结果进行对比, 图 9 中给出了 2 个模型对多个设备热像的检测热力图, 可见改进后的模型关注区域更加准确, 而基线模型更容易关注到一些无关区域, 验证了改进方法的可行性。

图 10 中的可视化结果来自参与数据扩增的 4 类设备热像, 图 11、12 中的可视化结果来自未参与数据扩增的 5 类设备热像。

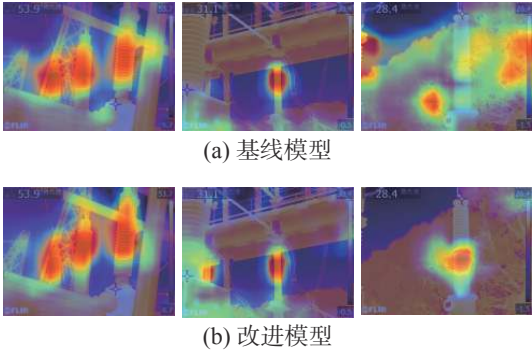


图 9 模型热力图可视化
Fig. 9 Visualization of model heat map

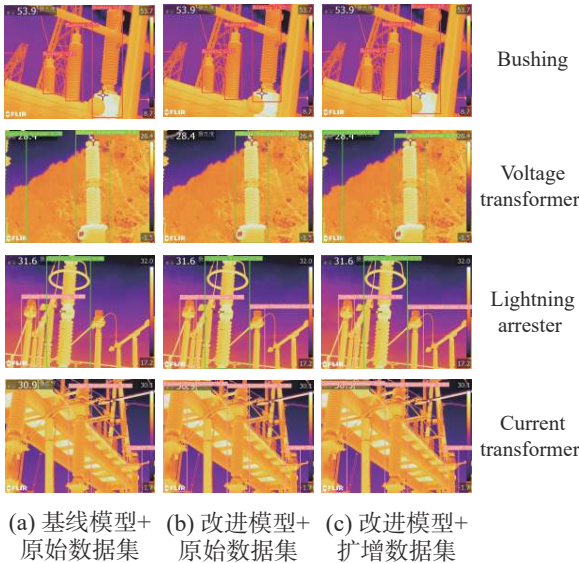


图 10 检测结果可视化
Fig. 10 Visualization of test results

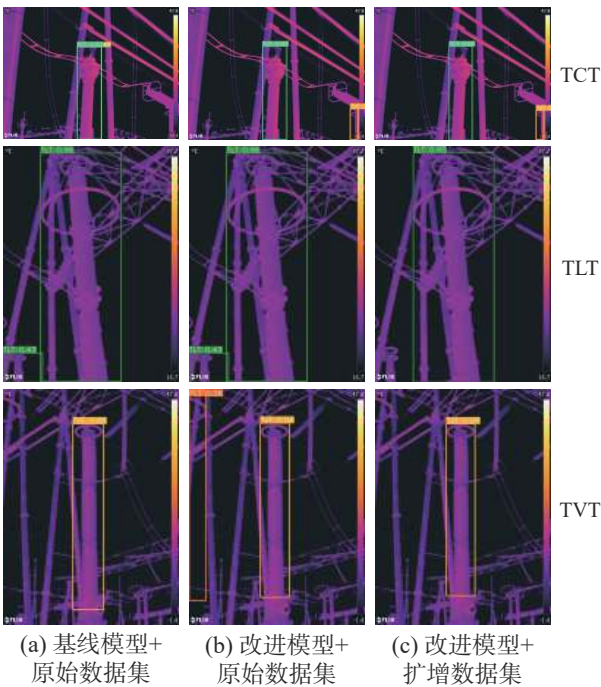


图 11 测试结果可视化
Fig. 11 Visualization of test results

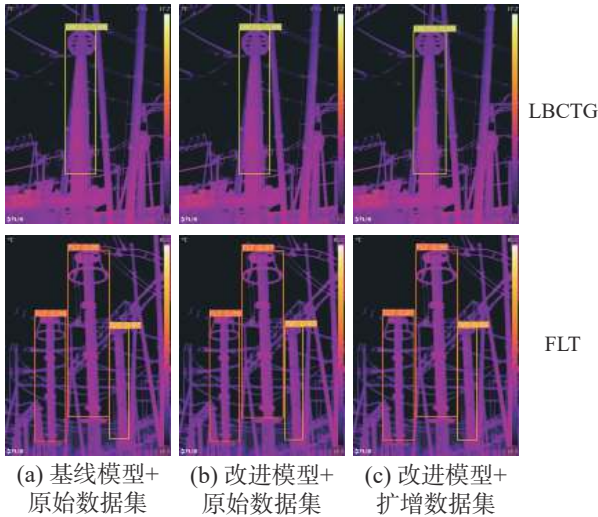


图 12 测试结果可视化
Fig. 12 Visualization of test results

图 10(c) 与 (a) 对比后可以看出, 第 3、4 行中的 Current Transformer 类, 在基线模型与原始数据集的组合中均出现了被漏检的情况, 而改进后的 YOLOv6 模型与扩增数据集的组合能够有效减少对变电设备的漏检情况。图 11(c) 与 (a)、(b) 对比后可以看出, 改进后的 YOLOv6 模型与扩增数据集的组合实现了最优的检测效果: 基线模型对 TCT 图像的检测出现了漏检, 而在改进模型的预测结果中均检测出了图像右下角的设备目标; 在使用原始数据集的情况下, 基线模型和改进模型对 TLT 图像的检测均出现了误检, 使用扩增数据集后消除了图像左下角的误检; 改进模型与原始数据集的组合, 对 TVT 图像的检测中出现了误检, 在使用扩增数据集后避免了误检。

对比图 12 中的设备检测结果, 能够看出改进模型或扩增数据集在提升部分设备检测精度的同时, 均未降低对这两类设备的检测效果, 定性地表现了本文方法的有效性与稳定性。

3.3 与先进模型对比实验

为了验证本文提出方法的有效性和先进性, 本文使用多种目标检测算法在变电设备热像数据集上进行训练, 分别与 Faster R-CNN、Cascade R-CNN^[32]、FCOS^[33]、YOLOv5、YOLOX、YOLOF^[34]、YOLOv7^[35] 和 YOLOv8 进行了比较, 各项测试指标如表 5 所示。从表 5 中可以看出, 本文模型在多个指标上与先进方法相比, 均有较明显的提升, AR 为检测模型的平均召回率, 因为数据集中变电设备的目标普遍偏大, 所以加入了 AP^l 这一指标, 来表示图像中大目标的检测精度。通过和最新方法 YOLOv8 的比较, 可以看出其检测精度

虽与本文模型较为接近, 但在检测速度这一指标上, 本文模型优势明显。

表 5 本文方法与先进目标检测模型的性能对比
Table 5 Performance comparison of improved models with advanced object detection models

模型	mAP/%	AP ⁵⁰ /%	AP ^l /%	AR/%	检测速度/(f/s)
Faster R-CNN	75.4	93.3	76.3	79.4	16.7
Cascade R-CNN	79.2	93.1	80.3	83.5	14.3
FCOS	72.0	93.4	81.2	76.0	32.5
YOLOv5	78.4	93.5	79.3	67.3	105.3
YOLOX	82.7	96.3	83.1	85.6	96.2
YOLOF	73.1	92.5	74.3	76.4	97.1
YOLOv7	81.1	94.0	81.4	78.9	108.7
YOLOv8	86.0	97.6	84.7	88.9	112.4
本文方法	86.4	97.8	87.2	89.4	192.7

4 结束语

针对变电设备热像检测中存在复杂背景样本不足和检测精度不理想的问题, 本文提出了基于知识迁移和改进 YOLOv6 的变电设备热像检测方法。首先基于 DDPM 模型提取真实热像中的背景知识, 生成人工背景样本, 基于位置分布先验知识, 将真实变电设备样本与人工背景样本进行融合, 生成了人工热像, 解决了复杂样本不足的问题; 然后在 YOLOv6 的骨干网络中实现 MHSA 机制, 提升模型特征提取能力, 并在其颈部网络中加入显示视觉中心模块, 改善深层特征的特征表示。实验结果表明, 本文提出的基于知识迁移的样本生成方法, 能够有效解决样本不足的问题, 同时, 本文模型能够有效提升对变电设备热像的检测精度, 在变电设备热像数据集上的检测效果和检测速度优于多种先进模型。

本文提出的样本生成方法, 为输变电设备热像的数据扩增提供了一种新的思路。另外本文提出的改进模型具有较高的检测精度, 下一步工作可以从小样本学习和高精度检测模型相结合的角度入手, 研究基于少量标注样本实现高质量自动标注的实现方法。

参考文献:

- [1] 周远翔, 陈健宁, 张灵, 等. “双碳”与“新基建”背景下特高压输电技术的发展机遇 [J]. 高电压技术, 2021, 47(7): 2396–2408.
ZHOU Yuanxiang, CHEN Jianning, ZHANG Ling, et al. Opportunity for developing ultra high voltage transmission technology under the emission peak, carbon neutral-

- ity and new infrastructure[J]. *High voltage engineering*, 2021, 47(7): 2396–2408.
- [2] 翟永杰, 王乾铭, 杨旭, 等. 融合外部知识的输电线路多金具解耦检测方法[J]. *智能系统学报*, 2022, 17(5): 980–989.
- ZHAI Yongjie, WANG Qianming, YANG Xu, et al. A multi-fitting decoupling detection method for transmission lines based on external knowledge[J]. *CAAI transactions on intelligent systems*, 2022, 17(5): 980–989.
- [3] 赵振兵, 孔英会, 戚银城. 面向智能输变电的图像处理技术[M]. 北京: 中国电力出版社, 2014.
- [4] 戴云峰, 冯兴明, 丁亚杰, 等. 单阶段目标检测网络的实例分割方法[J]. *应用科技*, 2023, 50(6): 42–47.
- DAI Yunfeng, FENG Xingming, DING Yajie, et al. Instance segmentation method based on single-stage object detection network[J]. *Applied science and technology*, 2023, 50(6): 42–47.
- [5] 杜景博, 姜勇. 基于YOLOv5-nS算法的绝缘子串销钉检测方法[J]. *应用科技*, 2023, 50(6): 1–6.
- DU Jingbo, JIANG Yong. Insulator string pin detection method based on YOLOv5-nS algorithm[J]. *Applied science and technology*, 2023, 50(6): 1–6.
- [6] 赵振兵, 王睿, 赵文清, 等. 基于图知识推理的输电线路缺销螺栓识别方法[J]. *智能系统学报*, 2023, 18(2): 372–380.
- ZHAO Zhenbing, WANG Rui, ZHAO Wenqing, et al. Pinmissing bolts recognition method for transmission lines based on graph knowledge reasoning[J]. *CAAI transactions on intelligent systems*, 2023, 18(2): 372–380.
- [7] HAN Sheng, YANG Fan, JIANG Hui, et al. A smart thermography camera and application in the diagnosis of electrical equipment[J]. *IEEE transactions on instrumentation and measurement*, 2021, 70: 1–8.
- [8] ZHENG Hanbo, CUI Yaohui, YANG Wenqiang, et al. An infrared image detection method of substation equipment combining iresgroup structure and CenterNet[J]. *IEEE transactions on power delivery*, 2022, 37(6): 4757–4765.
- [9] WANG Bin, DONG Ming, REN Ming, 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): 5345–5355.
- [10] ZHENG Hanbo, SUN Yonghui, LIU Xinghua, et al. Infrared image detection of substation insulators using an improved fusion single shot multibox detector[J]. *IEEE transactions on power delivery*, 2021, 36(6): 3351–3359.
- [11] OU Jianhua, WANG Jianguo, XUE Jian, et al. Infrared image target detection of substation electrical equipment using an improved faster R-CNN[J]. *IEEE transactions on power delivery*, 2023, 38(1): 387–396.
- [12] HUANG Song, SHANG Bowen, SONG Yanlou, et al. Research on real-time disconnecter state evaluation method based on multi-source images[J]. *IEEE transactions on instrumentation and measurement*, 2022, 71: 1–15.
- [13] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020–04–23)[2023–03–22]. <https://arxiv.org/abs/2004.10934>.
- [14] ZHOU Xingyi, WANG Dequan, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019–04–16)[2023–03–22]. <https://arxiv.org/abs/1904.07850>.
- [15] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980–2988.
- [16] LI Zuoxin, ZHOU Fuqiang. FSSD: feature fusion single shot multibox detector[EB/OL]. (2017–12–04)[2023–03–22]. <https://arxiv.org/abs/1712.00960>.
- [17] GIRSHICK R. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. New York: ACM, 2015: 1440–1448.
- [18] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261–2269.
- [19] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: ACM, 2020: 6840–6851.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017: 6000–6010.
- [21] LI Chuyi, LI Lulu, JIANG Hongliang, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022–09–07)[2023–03–22]. <https://arxiv.org/abs/2209.02976>.
- [22] TOLSTIKHIN I, HOULSBY N, KOLESNIKOV A, et al. MLP-mixer: an all-MLP architecture for vision[EB/OL]. (2021–05–04)[2023–03–22]. <https://arxiv.org/abs/2105.01601>.
- [23] QUAN Yu, ZHANG Dong, ZHANG Liyan, et al. Centralized feature pyramid for object detection[EB/OL]. (2022–10–05)[2023–03–22]. <https://arxiv.org/abs/2210.02093>.
- [24] HUANG Huimin, LIN Lanfen, TONG Ruofeng, et al. UNet 3: a full-scale connected UNet for medical image segmentation[C]//ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona: IEEE, 2020: 1055–1059.

- [25] WANG Xinglong, ZHANG R, KONG Tao, et al. Solov2: Dynamic and fast instance segmentation[J]. Advances in neural information processing systems, 2020, 33: 17721–17732.
- [26] DING Xiaohan, ZHANG Xiangyu, MA Ningning, et al. RepVGG: making VGG-style ConvNets great again[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13728–13737.
- [27] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 16514–16524.
- [28] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [29] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 936–944.
- [30] GE Zheng, LIU Songtao, WANG Feng, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021–07–18)[2023–03–22].<https://arxiv.org/abs/2107.08430>.
- [31] SELVARAJU R R, DAS A, VEDANTAM R, et al. Grad-CAM: why did You say that? [EB/OL]. (2016–11–22) [2023–03–22].<https://arxiv.org/abs/1611.07450>.
- [32] CAI Zhaowei, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6154–6162.
- [33] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020: 9626–9635.
- [34] CHEN Qiang, WANG Yingming, YANG Tong, et al. You only look one-level feature[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13039–13048.
- [35] 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[EB/OL]. (2022–07–06) [2023–03–22].<https://arxiv.org/abs/2207.02696>.

作者简介:



赵振兵, 教授, 博士, 主要研究方向为电力视觉。主持国家自然科学基金项目等纵向课题 10 项; 获省科技进步一等奖 2 项; 以第一完成人获得国家专利授权 16 项; 以第一作者出版专著 2 部, 发表学术论文 50 余篇。



冯烁, 硕士研究生, 主要研究方向为电力视觉。



赵文清, 教授, 博士, 主要研究方向为人工智能和图像处理。发表学术论文 80 余篇。