

大模型数据扩增技术 在输电线路智能监控中的应用研究

Research on the Application of Data Augmentation with Large Model in Visual
Monitoring of Transmission Lines in Complex Scenarios

杨 勇¹,万超伦¹,马建友¹,赵文杰¹,董振江²(1. 江苏讯汇科技股份有限公司,江苏 南京 211316;2. 南京邮电大学,江苏 南京 210003)

Yang Yong¹, Wan Chaolun¹, Ma Jianyou¹, Zhao Wenjie¹, Dong Zhenjiang² (1. Jiangsu Xunhui Technology Co., Ltd., Nanjing 211316, China; 2. Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

摘要:

针对复杂场景下输电线路可视化监控中目标识别准确率不高、误检误报率较高的问题,提出了一种结合大模型的数据扩增和YOLOv10优化算法来提升复杂场景下目标识别准确率的实现方法。首先,针对样本数据数量短缺的问题,基于Stable-diffusion大模型进行数据扩增,丰富并增加了样本数量。其次,针对训练样本数量有限的情况,对YOLOv10算法进行改进,进一步强化图片特征的提取算法,优化目标识别算法,提升复杂自然场景下的目标识别准确率及性能。最终的实验结果表明,与现有的实现方法相比,针对复杂场景的输电线路的可视化监控,对目标识别的准确率从原有的52.6%提升至54.3%。

关键词:

数据扩增;目标识别;YOLO

doi:10.12045/j.issn.1007-3043.2024.09.006

文章编号:1007-3043(2024)09-0031-07

中图分类号:TN915.5

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

Aiming at the low object recognition accuracy and high rates of false detections and false alarms are common in the visual monitoring of transmission lines under complex scenarios, an combined method that leverages the data augmentation by large model with optimized YOLOv10 is proposed. Firstly, data augmentation is performed based on the Stable-diffusion large model to address the shortage of available sample data, which can enlarge the number of samples. Secondly, based on the limited number of training samples, the YOLOv10 algorithm is optimized by further enhancing the image feature extraction algorithm and optimizing the object recognition algorithm to improve its accuracy and performance. The final experimental results show that, compared to existing method, the proposed approach significantly improves the target recognition accuracy for visual monitoring of transmission lines in complex scenarios, from an original 52.6% to 54.3%.

Keywords:

Data augmentation; Object recognition; YOLO

引用格式:杨勇,万超伦,马建友,等. 大模型数据扩增技术在输电线路智能监控中的应用研究[J]. 邮电设计技术,2024(9):31-37.

1 概述

输电线路的可视化监控是保障电网安全运行的重要手段。通过在电力杆塔上架装智能摄像头,对输电线路的环境安全、电力作业安全等进行实时、远程和智能的监控,这种方式已经在电网运营企业中进行

了大规模的应用,并取得了不错的效果,可以显著提高输电线路的安全性和运维效率。然而,由于自然场景的复杂性,现有方案存在目标识别的准确性不高和及时性不够的问题,误检、漏检现象时有发生。例如,在复杂自然场景下,挖掘机在输电走廊附近进行作业,受树木、建筑等物体的遮挡,或者摄像头拍摄角度的影响,无法及时识别出挖掘机对输电线路安全运行造成的潜在危害。产生这种问题的原因在于,这种智

收稿日期:2024-07-12

能视觉实现方案通常基于YOLO等深度学习模型,通过大量的标注图片进行模型训练,形成最终的推理模型和算法。

国内外的许多学者对基于深度学习的电力线路监测技术都有过深入的研究,文献[1]提出了新模型YOLO-2MCS用于输电线路走廊隐患目标检测,并使用混合数据增强策略对数据集进行有效扩充;文献[2]提出一种基于改进型目标检测算法(YOLO v3)的绝缘子异物检测方法,在实现中首先使用视频分割算法和背景替换的方式对样本数据进行了扩增;文献[3]从近10年来基于深度学习的输电线路视觉检测方法的研究进展出发,首先概述适用于输电线路巡检的深度卷积神经网络,接着重点阐述基于深度卷积神经网络的输电线路巡检图像数据的目标检测,最后指出基于深度学习的输电线路巡检图像数据视觉检测方法目前存在样本均衡性、小目标检测等方面的问题。文献[4]对基于计算机视觉和深度学习的电力线路巡检方案及数据来源问题进行了总结和回顾,提出了一种新的基于自主视觉的电力线路巡检概念。综上,在基于深度学习的电力线路监控和巡检应用中,研究人员始终面临着如下挑战:样本数据获取困难,由于电网运营安全性要求很高,管理措施也比较严格,加之这种影响电网安全的作业行为并不经常发生,所以很难获取到大量的训练样本数据来训练模型;复杂场景下的目标识别问题,包括复杂的自然场景下电力线路危害监测和预警,以及电网小目标识别问题,如绝缘子粉化、小金具锈蚀、导线断股、M销松动等。

针对训练样本数据短缺的问题,现有方案通常采用数据扩增的方法来改进^[5-12]。文献[5]采用光照变换、噪声变换、旋转变换、平移变换等传统图像处理方法来对原始样本数据集进行扩充,文献[6]采用生成对抗网络技术来进行数据扩增,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[9-10]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。文献[11]使用去噪扩散概率模型提取出域外数据中的背景知识,将电力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图像。

以上这些数据扩增的实现方法,总体上可以分为2类,一类是基于传统图像处理的数据扩增技术,另一类是基于深度学习网络的数据库扩增。本文将针对电力复杂场景下的目标识别问题,提出一种基于大模

型内容生成技术的数据扩增方法,同时,针对这些包含真实样本数据和扩增生成数据的复杂场景下的目标识别,改进YOLOv10算法,从而提升了目标识别的准确性。

2 基于大模型的数据扩增方法

基于大模型的数据扩增方法涉及到文本-图像预训练模型和稳定扩散模型等关键技术,本章将对这些技术进行简单介绍。

2.1 vae变分自编码器

vae变分自编码器^[13]包含如下2个部分。

a) 编码器(Encoder)。它将图像作为输入,并将其转换为深层潜在特征空间的低维度表示。

b) 解码器(Decoder)。它将低维度特征表示作为输入,并将低维度的特征表示进行解码,还原成图像进行输出。

整体过程如图1所示。

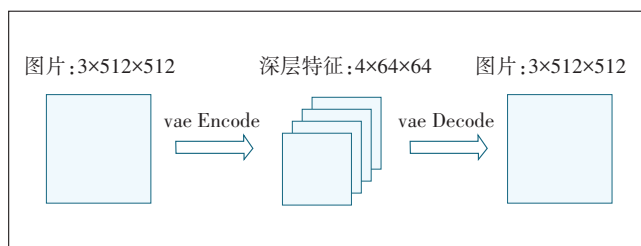


图1 变分自编码器流程

vae变分自编码器可以高效地减少高分辨率图像生成的计算时间,扩散模型算法^[14]无需直接生成目标图片,只需生成低维度的潜在特征即可,再通过变分自编码器的Decode部分,将潜在特征还原成目标图像。在该过程中,送入扩散模型算法的潜在特征是一个低维度特征,降低了扩散模型算法的推理时间。

2.2 CLIP文本-图像预训练模型

CLIP是OpenAI在2021年初发布的,是一个采用了4亿对图像文本对和对比学习算法训练而成的,用于匹配图像和文本的预训练神经网络模型。它能够将图像和文本同时映射到一个共享的向量空间中,使得模型能够理解图像与文本之间的语义关系。

CLIP由一个图像编码器(image_encoder)^[15]与一个文本编码器(text_encoder)构成,将图像和文本分别输入,从而得到图像和文本的向量表示image_features和text_features。然后将特征向量映射至一个联合多模态空间内部,再分别计算图像向量与文本向量之间

的余弦相似度。

CLIP 的训练十分灵活,采用自监督的方案,在图像特征与文本特征构造成的联合多模态空间中,将对角线部分设置为正样本,其余所有联合点被设定为负样本,正负样本被构造后,模型就可以正常训练了。CLIP 的结构如图 2 所示。

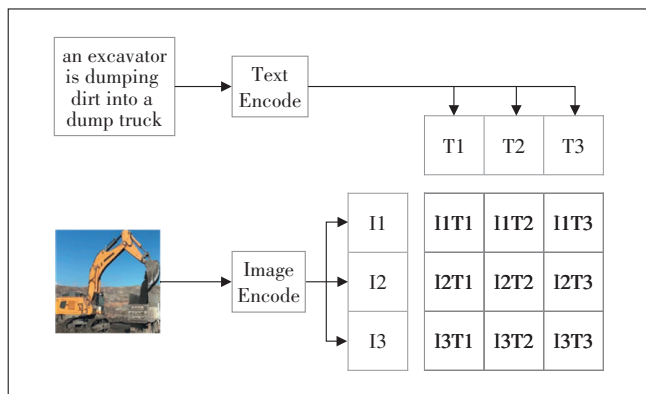


图 2 CLIP 模型结构

2.3 基于稳定扩散模型的目标图像生成

稳定扩散模型是基于 2022 年发表于 CVPR 的潜在扩散模型算法演进^[16]而来,被用于图像生成工作。与原有的生成模型相比,比如 GAN,稳定扩散模型在训练稳定性和生成多样性上都有了很大的提升。

稳定扩散模型的生成所需目标图片的过程可以分为训练过程与推理过程 2 个部分。

a) 训练过程。

(a) 采用 vae 变分自编码器的编码器对输入图片进行推理,得到图片的潜在特征 x_0 。

(b) 采用 CLIP 模型对输入图片对应的图片描述进行推理,得到相对应的文本特征 t_0 。

(c) 对步骤(a)中的潜在特征 x_0 ,分 t 次添加满足标准高斯分布的噪声,直至图像被破坏为随机噪声 x_t 。其中,加噪公式满足:

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} z_t \quad (1)$$

其中, x_t 代表当前时刻加噪后的数据; $\alpha_t = 1 - \beta_t$, 代表加噪的权重系数; β_t 是超参数,从 0.000 1~0.002 呈线性衰减; x_{t-1} 代表上一时刻加噪后的数据; z_t 代表加入的服从标准高斯分布的噪声,均值为 0,方差为 1, $z_t \sim N(0, 1)$ 。由 0 时刻推算 t 时刻数据公式为:

$$x_t = \sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t} z_t \quad (2)$$

(d) 针对步骤(c)中添加的噪声,采用 Unet 网络对

潜在特征 x_0 添加的噪声进行预测,构建噪声预测网络。然后训练网络,拟合噪声 z_t ,模型的输入为加噪之后的特征 x_t 、时刻 t 和 CLIP 模型得到的文本特征 t_0 。

b) 推理过程。

(a) 采用 CLIP 模型对输入图片对应的图片描述进行推理,得到相对应的文本特征 t_0 。

(b) 随机生成一组噪声图片 x_t ,将其与 CLIP 模型得到的文本特征、时刻 t 输入 Unet 噪声预测模型中去,预测出噪声,同时推导出 $t-1$ 时刻的数据,迭代 t 次。由 x_t 推理 x_{t-1} 的去噪公式如下:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} z_t \right) + C \quad (3)$$

其中, z_t 表示当前输入被添加的噪声,以此来预测 x_{t-1} ; C 是常数项。

(c) 通过 t 次的噪声预测,并去噪还原特征后,将去噪之后的数据作为潜在特征,送入预训练 vae 变分自编码器的解码器中,生成目标图片。

3 基于图像扩增的复杂场景电网安全监测算法

3.1 基于大模型的图像扩增

针对电力复杂场景下的目标识别工程中的检测样本稀少导致检测精度低下的问题,样本扩增能有效地提升模型的训练效果。如前所述,现有的研究主要集中于传统图像处理或基于深度学习的图像扩增方法。

本文提出了一种基于大模型的数据扩增技术,算法的实现过程如图 3 所示。

算法的实现包括如下 2 个阶段。

a) 训练阶段,即潜在图像特征生成阶段,先通过变分自编码器(vae)大模型的自动编码器提取数据的潜在特征,再通过文本-图像预训练模型(CLIP)提取文本描述特征,使用扩散模型算法,微调噪声预测网络。

b) 推理阶段,基于第 1 阶段的模型,将随机噪声和文本特征输入网络进行噪声预测,之后使用扩散模型算法中的 Unet 模型^[17],生成潜在图像特征,再通过解码器生成高质量的目标图片。

稳定扩散模型生成图片的过程如图 4 所示。

本文使用收集的工程机械图片来微调稳定扩散模型,工程机械图片数据集来自人工收集监控摄像头采集的输电线路沿线包含的工程机械图片。以工程

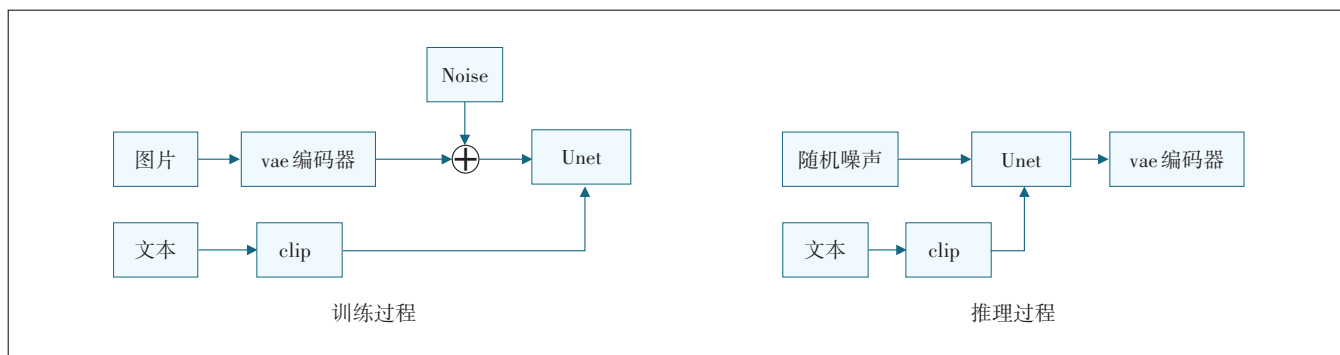


图3 基于大模型的数据扩增方法

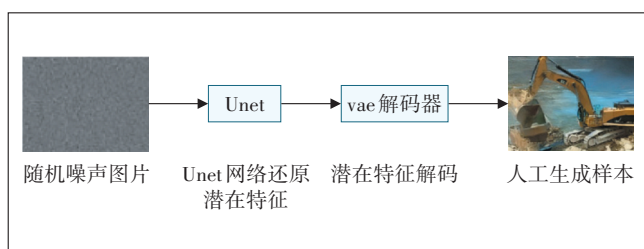


图4 图片生成过程

机械中的挖掘机为例,针对包含挖掘机的图片,裁剪出指定目标,并对每一张指定目标图片采用预训练的文字描述生成算法(BLIP)生成对应文字描述。构建图片-文本数据集。采集的图片不参与后续目标检测模型的测试。挖掘机图片-文本数据集构建过程如图5所示。



图5 挖掘机图片-文本数据集构建过程

针对稳定扩散模型生成的挖掘机图片,还需要经过一些处理,从而变成可以用来作为目标检测任务数据扩增的数据。

首先,收集多张无目标图片,作为实际图片的背景部分。

其次,将稳定扩散模型生成的挖掘机图片,在先验知识的约束下进行尺度上的随机缩放,使其更具有随机性,再将其粘贴至背景图片下半部分的随机位

置,从而构建目标检测数据集。

在先验知识的约束下,将生成的前景挖掘机图片与真实背景图片进行合成,即可实现对不同工程机械设备的数据自动扩增。图片扩增过程如图6所示。



图6 图片扩增过程

3.2 一种改进的YOLOv10网络

基于图像扩增技术生成的复杂场景下的作业样本图像,由于画面内容比较复杂,加之采用大模型生成的扩增数据,包含了较多的小目标特征,现有YOLOv10^[18]算法的准确性还有待提升。

本文在YOLOv10原有模型的基础上,添加了SE(Squeeze-and-Excitation)^[19]通道注意力模块,旨在让神经网络更有效地聚焦关键信息,学习权重特征,引导模型关注输入序列中最重要的部分,从而显著提升模型的性能。

SE通道注意力模块主要有Squeeze和Excitation 2个操作。

a) Squeeze操作主要是由一个全局平均池化操作完成的,对于一个大小为 $H \times W \times C$ 维度的特征,对于每个通道执行全局平均池化操作,会得到一个 $1 \times 1 \times C$ 维度的特征向量,该向量上的每一个元素都表示对应通道上的平均值,因此,得到的该特征向量也可以被理解为整个特征映射在通道上的描述符。特征向量的每一个元素可由式(4)表示:

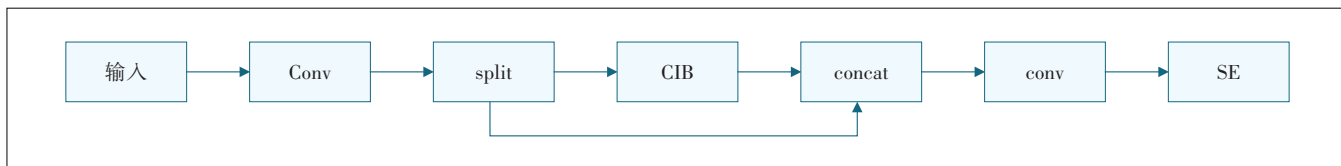


图9 C2fCiBAttention模块

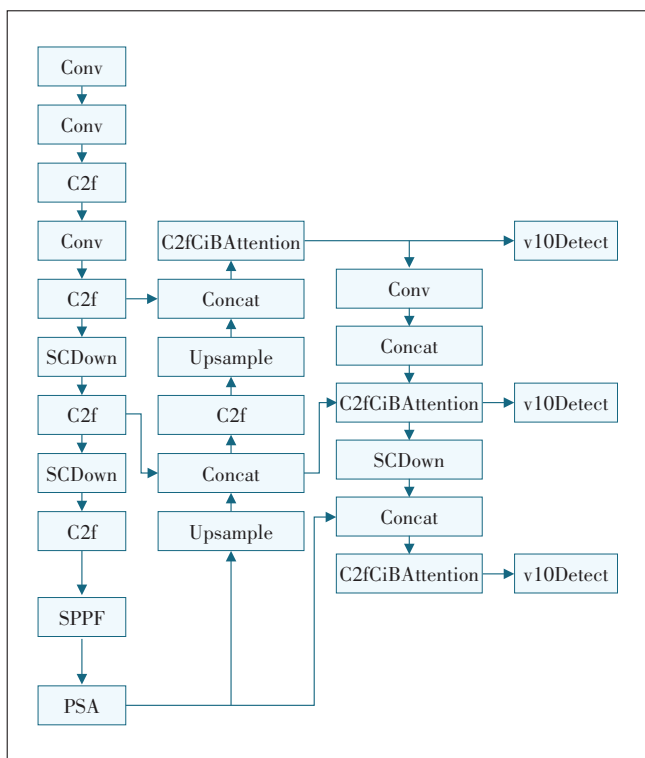


图10 在颈部和头部处增加C2fCiBAttention模块的改进YOLOv10模型

表1 训练平台配置参数

操作系统	ubuntu 20.04
显卡配置	NVIDIA GeForce RTX 3090
开发环境	python 3.8
框架版本	pytorch 1.13.1
加速库版本	CUDA 12.0

本文选取了一类工程机械(挖掘机)作为实验对象,其中原始数据训练集和验证集样本图片分别为6 026张和677张,在数据集扩增阶段,利用本文提出的基于大模型的样本生成方案,对挖机数据进行训练集扩增(见表2),在数据扩增后,训练集图像数量为7 137张,验证集图像数量不变。

为了验证本文YOLOv10改进方案的有效性,本文设计了一组消融实验。同时,为了验证本文提出的基

表2 数据集详细情况

工程机械类别	原始训练集数量/张	原始验证集数量/张	扩增训练集数量/张	扩增验证集数量/张
挖掘机	6 026	677	7 137	677

于大模型的数据扩增有效性,根据数据集的划分,将实验分为原始数据集与扩增后数据集2组。共设置3组实验内容,模型A是YOLOv10基线模型,模型B是在颈部与头部之间加入了通道注意力机制的YOLOv10模型,模型C是在骨干网络中加入了通道注意力机制的YOLOv10模型。模型的训练轮次(epoch)为100,输入图片大小为640×640,训练时的批处理(batch size)参数为16。表3给出了消融实验结果。

表3 消融实验结果(单位:%)

模型	数据扩增	P	R	AP50	mAP50-95
A	无	87.132	70.193	80.774	52.607
A	有	85.563	72.817	81.725	53.954
B	无	83.245	73.379	81.246	53.188
B	有	88.595	74.236	83.583	54.379
C	无	85.272	66.525	78.303	51.000
C	有	84.898	73.548	81.504	53.408

从表3可以看出,本文提出的方案能有效地提升模型的检测精度。

在模型A YOLOv10的基线模型上,与采用原始训练集训练相比,使用了本文提出的基于大模型生成的样本进行训练之后,mAP50-95与AP50分别提高了1.347%与0.951%。

在模型B采用的方案中,与基线模型A相比,使用原始数据集进行训练,mAP50-95高出0.581%,使用扩增数据进行训练,mAP50-95比同条件训练出的基线模型高出0.425%。

在模型C采用的方案中,与采用原始训练集训练相比,使用扩增数据进行训练,mAP50-95提高了2.408%。但使用原始数据集进行训练,mAP50-95相比基线模型A反而有所下降。

由此可见,基于本文提出的数据集扩增方案进行

模型训练,可以实现比原始数据更高的准确性。同时,本文提出的在颈部和头部之间引入通道注意力机制的YOLOv10优化方案比在骨干网络中加入了通道注意力机制的YOLOv10方案能取得更好的效果。

5 结论

针对输电线路沿线施工机械存在样本不足的问题,本文提出了基于大模型的数据扩增方法与改进的基于YOLOv10的施工机械检测方法。首先基于大模型的潜在特征扩散模型提取工程机械的内在特征,生成大量的工程机械样本,再根据先验知识,将生成的大量工程机械样本与真实背景进行融合,生成图片数据,从而解决了输电线路沿线物体识别中的工程机械样本不足的问题;然后将通道注意力机制分别放入YOLOv10的骨干网络、模型颈部与头部之间,以提高模型的特征提取能力,使其更专注于数据中的重要区域,提高模型准确率。实验结果表明,本文提出的基于大模型的数据扩增方案,能有效解决样本不足的问题,同时,本文改进的模型能有效提高输电线路沿线的施工机械识别的检测精度。

本文提出的样本生成方案,为输电线路沿线的稀少样本扩增提供了新的思路。本文提出的改进模型也具有较高的检测精度,下一步将把研究重点放在无监督学习上,通过少量真实样本与人造样本提高未标注数据的检测精度。

参考文献:

[1] 郑含博,胡思佳,梁炎桑,等. 基于YOLO-2MCS的输电线路走廊隐患目标检测方法[J]. 电工技术学报,2024,39(13):4164-4175.

[2] 张焕坤,李军毅,张斌. 基于改进型YOLOv3的绝缘子异物检测方法[J]. 中国电力,2020,53(2):49-55.

[3] 刘传洋,吴一全. 基于深度学习的输电线路视觉检测方法研究进展[J]. 中国电机工程学报,2023,43(19):7423-7446.

[4] NGUYEN V N, JENSSEN R, ROVERSO D. Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning[J]. International Journal of Electrical Power & Energy Systems, 2018, 99: 107-120.

[5] 冯庆贺,江铭凯,郝巧红,等. 一种改进的YOLOv8太阳能电池板缺陷识别方法[J]. 广西电力,2024(4):46-52.

[6] 李昆,朱卫纲. 基于MDGAN网络的数据集扩增方法[J]. 雷达科学与技,2020(2):211-217.

[7] HAN S, YANG F, JIANG H, 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 H B, CUI Y H, YANG W Q, et al. An infrared image detec-

tion method of substation equipment combining irectgroup structure and CenterNet[J]. IEEE Transactions on Power Delivery, 2022, 37(6):4757-4765.

[9] 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):5345-5355.

[10] OU J H, WANG J G, XUE J, 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.

[11] 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.

[12] 夏正勋,杨一帆,罗圣美,等. 生成技术在人工智能平台中的应用探索[J]. 大数据,2020,6(6):119-128.

[13] KINGMA D P, WELING M. Auto-encoding variational bayes [EB/OL]. [2024-01-26]. <https://arxiv.org/abs/1312.6114>.

[14] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models [C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2020:6840-6851.

[15] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision [C]//Proceedings of the 38th International Conference on Machine Learning. New York: PMLR, 2021:8748-8763.

[16] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022:10674-10685.

[17] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Cham:Springer, 2015:234-241.

[18] WANG A, CHEN H, LIU L H, et al. Yolov10: real-time end-to-end object detection [EB/OL]. [2024-01-26]. <https://arxiv.org/abs/2405.14458>.

[19] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 42(8):2011-2023.

作者简介:

杨勇,毕业于东南大学,高级工程师,博士,主要研究方向为多媒体通信和计算机视觉;万超伦,硕士,主要从事计算机视觉方面的技术研究等工作;马建友,硕士,主要从事计算机视觉方面的技术研究等工作;赵文杰,工程师,学士,主要从事电网可视化监控技术研究和相关产品研发工作;董振江,南京邮电大学教授,博士,主要从事大数据和人工智能相关的教学科研工作。