

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2024)07-2011-13

论文引用格式: Bai Y F, Wang L B, Gao W D and Ma Y L. 2024. Multi-modal hierarchical classification for power equipment defect detection. Journal of Image and Graphics, 29(07):2011-2023(白艳峰, 王立彪, 高卫东, 马应龙. 2024. 面向电力设备缺陷检测的多模态层次化分类. 中国图象图形学报, 29(07):2011-2023)[DOI:10.11834/jig.230269]

## 面向电力设备缺陷检测的多模态层次化分类

白艳峰, 王立彪, 高卫东, 马应龙\*

华北电力大学控制与计算机工程学院, 北京 102206

**摘要:** 目的 电力设备的状态检测和故障维护是保障电力系统正常运行的重要基础。针对目前多数变电站存在电力设备缺陷类型复杂且现有的单分类缺陷检测方法无法满足电力设备的多标签分类缺陷检测需求的问题, 提出一种面向电力设备缺陷检测的多模态层次化分类方法。方法 首先采集来自多个变电站的电力设备缺陷图像并进行人工标注、数据增强及归一化等预处理, 构建了一个具有层次标签结构的电力设备缺陷图像数据集。然后提出一种基于多模态特征融合的层次化分类模型, 采用 ResNet50 网络对图像进行特征提取, 利用区域生成网络对目标进行定位以及前景、背景预测; 为避免对区域生成网络生成的位置坐标进行量化时引入误差, 进一步采用 ROI Align (region of interest align) 方法连续操作, 生成位置坐标。最后采用层次化分类, 将父类别标签嵌入到当前层目标特征表示进行逐层缺陷分类, 最后一层得到最终的缺陷检测结果。结果 在电力设备缺陷数据集和基准数据集上, 与多标签分类电力设备缺陷检测方法和流行的常用目标检测算法进行对比实验。实验结果表明, 模型对绝大部分设备缺陷类别的检测准确率最高, 平均检测准确率达到 86.4%, 相比性能第 2 的模型, 准确率提升了 5.1%, 并且在基准数据集上的平均检测准确率也提高了 1.1%~3%。结论 提出的电力设备缺陷检测方法充分利用设备缺陷标签的语义信息、层次结构和设备缺陷数据的图像特征, 通过多模态层次化分类模型, 能够提升电力设备缺陷检测的准确率。

**关键词:** 缺陷检测; 图像识别; 层次化分类; 多模态特征融合; 标签嵌入; 区域特征聚集

### Multi-modal hierarchical classification for power equipment defect detection

Bai Yanfeng, Wang Libiao, Gao Weidong, Ma Yinglong\*

School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

**Abstract: Objective** Safety state detection of power equipment is a fundamental task to ensure the safe operation of power systems. The state detection and fault maintenance of power equipment are the basic prerequisites for ensuring the normal operation of the power system. With the growing diversities and complexity of defects in substations, the current defect recognition and power detection has increasingly been required to handle multi-label classification tasks based on a large number of closely related defect labels. However, due to the complex types of power equipment defects in most substations, most existing approaches for power equipment defect detection are inefficient at multi-label defect detection because the

收稿日期: 2023-05-29; 修回日期: 2023-10-27; 预印本日期: 2023-11-03

\* 通信作者: 马应龙 yinglongma@ncepu.edu.cn

基金项目: 国家自然科学基金项目(62072450); 国家电网公司科技项目(SGGSXT00XMJS2250023)

Supported by: National Natural Science Foundation of China (62072450); State Grid of Corporation of China Science and Technology Program (SGGSXT00XMJS2250023)

defect category labels often have different granularities in their semantic concepts and are often closely related with each other. All these problems cause existing defect detection methods to have difficulty meeting the requirements of multi-label classification-based defect detection tasks of power equipment. To address these problems, this paper proposes a multi-modal hierarchical classification for power equipment defect detection, which is suitable for defect detection in complex power equipment environments. **Method** We propose a multi-modal hierarchical classification method, which fuses the feature information of defect images, hierarchical structure information, and the semantic information of category labels. First, defect images of power equipment from multiple substations are collected and preprocessed with manual annotation, data enhancement, and normalization to construct a power equipment defect image dataset with a hierarchical label structure. Then, a hierarchical classification model based on multi-modal feature fusion and hierarchical fine-tuning techniques is proposed, which uses the ResNet50 network to extract features from images, and a region proposal network to locate object and predict the foreground and background. The region of interest align (ROI Align) method is further used to continuously generate the position coordinates to avoid introducing errors in quantifying the position coordinates generated by the region proposal network. Finally, the hierarchical structure of power equipment to be detected is used to embed the parent category labels into the current layer's object feature representation for layer-by-layer defect classification. The final defect detection result is obtained in the final layer. **Result** Comparative experiments are conducted on the real-world power equipment defect dataset and the PASCAL VOC2012 benchmark dataset against the current multi-label classification-based power equipment defect detection methods and the popularly used object detection algorithms. Experimental results show that the proposed method achieved the best detection accuracy for most equipment defect categories, with a mean average precision of 86.4%. Compared with the second-best performing model, the accuracy improved by 5.1%, and the mean average precision on the benchmark dataset increased by 1.1% to 3%. The proposed method can be executed in a relatively shorter time than the compared methods. **Conclusion** Our method achieves superior detection accuracy performance against the compared methods while maintaining a lower computational cost. It can improve the accuracy of power equipment defect detection through a hierarchical classification model based on multi-modal feature fusion by fully utilizing the semantic relationship between equipment defect labels.

**Key words:** defect detection; image recognition; hierarchical classification; multi-modal feature fusion; label embedding; regional feature aggregation

## 0 引言

电力设备是电力系统的重要组成部分,电力系统的可靠性要求需要保证输变电等过程中的电力设备保持安全、稳定的运行状态。因此,全面、准确地检测电力设备状态具有重要意义,提高电力设备缺陷检测准确率能够提升电力系统的自动化和智能化水平,减少电力设备维护的人力投入。

基于深度学习的目标检测技术是计算机视觉领域的重要研究方向(曹家乐等,2022),近年来被广泛应用到电力设备的缺陷检测。一方面,部分学者针对变电站中的具体一类设备进行单分类缺陷检测。例如,戚银城等人(2021)针对输电线路上的螺栓进行精细的缺陷检测,提出一种双注意力机制方法,利用不同尺度特征图计算得到的注意力图和相邻层注意力图的差异性,增强了小部件的精细特征,

并通过空间注意力图增大了目标与背景的差异度,从而提高了螺栓的检测效果;赵文清等人(2021)针对绝缘子缺陷,提出一种基于改进多尺度网络的绝缘子识别模型,该模型通过扩大小尺度特征来丰富绝缘子的语义信息,并改进了锚点框的参数,提升了绝缘子的识别准确率。另一方面,部分学者针对变电站中的多类设备进行多分类缺陷检测。毛进伟(2021)提出一种样本偏移网络,通过构建候选框偏移模块,来计算分类回归任务的偏移量,以此得到最佳的候选框,同时在传统实例分割网络的基础上对特征融合、特征提取和损失函数进行改进,提高了网络的分割检测性能。Ying等人(2020)基于Faster R-CNN(region convolutional neural network)算法,根据实际应用场景提出了类间采样、中心指导NMS(non-maximum suppression)和类别自适应阈值等改进策略,解决了样本分布不均匀、边界框冗余以及漏检或误检等问题。

然而,由于电力设备种类越来越丰富,设备缺陷

也越来越复杂,现有的只对具体一类设备进行单分类缺陷检测的方法以及对多类设备进行缺陷检测的方法存在无法充分利用设备缺陷标签之间语义关系的问题,导致变电站内各类电力设备存在的复杂缺陷无法得到全面准确的检测。

针对上述问题,提出了一种基于层次化多标签分类的电力设备缺陷检测方法。首先利用设备缺陷标签之间的语义关系和先验知识;然后利用父类别标签嵌入和多模态特征融合(Huang等,2021)方法增强缺陷语义信息,逐层进行缺陷分类,通过两个全连接层进行最终的缺陷目标边界回归和类别预测。

## 1 电力缺陷数据集问题分析及预处理

### 1.1 问题分析

本文方法主要针对呼吸器、硅胶及硅胶桶、表计、箱门、油封、高压线、开关柜、大型设备的零部件、绝缘子等维护频率较高的设备进行检测。电力设备缺陷数据集来自国网某公司,包含两级层次化缺陷类别标签,经过数据采集和清洗、分类和数据增强,构造的电力设备缺陷图像数据集的各类标签的层次结构、标签名称、数量如表1所示。

电力设备缺陷数据集有以下几个方面的特点。首先,从表1可以发现,表计、绝缘子、呼吸器3类目标类别的图像数据中,既存在有缺陷的目标,又存在没有缺陷的目标,因此将表计正常、绝缘子正常、硅胶正常、油封正常作为正样本类别。其次,由于同一设备可能包含多种缺陷,因此同一缺陷图像可能同时属于多个细粒度类别。最后,缺陷图像数据在第2层的细粒度类别上样本分布很不均衡。例如,油封破损的样本占比(3.1%)远小于金属锈蚀的样本占比(17%)。即使在具有相同父类别的细粒度类别的样本占比也差异很大,如油封破损和硅胶变色两个类别。

针对电力设备缺陷数据数据不均衡、多标签分类的特点,本文利用电力数据的层次化类别结构,提出基于多特征融合的层次化多标签分类方法以实现电力设备缺陷的多目标检测。该方法利用上一层父分类器的分类结果,来指导和优化下一层的更细粒度的分类任务。同时,为了进一步提升缺陷特征表示能力,将图像样本特征、样本的父类别的位置特征以及父类别名称标签的文本语义特征等进行融合,

表1 类别标签层次结构

Table 1 Hierarchy of category label

根结点 (目标总数)	第1层标签(占比)	第2层标签(占比)
电力设备缺陷 (15 361)	表计类(19%)	表计破损(5.7%)
		表盘模糊(6.4%)
		表计正常(6.8%)
	绝缘子类(11%)	绝缘子破裂(5.9%)
		绝缘子正常(5%)
	呼吸器类(23%)	硅胶变色(12.6%)
		硅胶正常(4.2%)
		油封破损(3.1%)
		油封正常(2.6%)
	异物类(11%)	挂空悬浮物(4.3%)
		鸟巢(7%)
	渗漏油类(19%)	表面油污(10.4%)
		地面油污(9%)
	金属锈蚀类(17%)	金属锈蚀(17%)

以提升基于多模态特征融合的电力设备缺陷层次化分类性能。

### 1.2 数据增强

采用Mosaic数据增强方法(Bochkovskiy等,2020),首先读取4幅图像并绘制画布和随机点(坐标为 $(x_i, y_i)$ ),接着对4幅图像分别进行随机缩放、裁剪和色域变化等,并按照矩形框的4个角分别摆放,最后进行图像组合以及边界框的结合。

图像左上角拼图、右上角拼图的拼接过程如图1所示,其中蓝色框表示画布,橙色框表示图像, $S_{\text{mosaic}}$ 是增强数据中表示画布尺寸的变量。左上角拼图中,情况1表示图像超出画布,情况2表示图像不超出画布,这时画布的摆放区域可表示为

$$(x_1, y_1, x_2, y_2) = \left\{ \max(x_i - w, 0), \max(y_i - h, 0), x_i, y_i \right\} \quad (1)$$

图像的区域可表示为

$$(x'_1, y'_1, x'_2, y'_2) = \left\{ w - (x_2 - x_1), h - (y_2 - y_1), w, h \right\} \quad (2)$$

同理,可实现右上角拼图、左下角拼图和右下角拼图。

增强效果如图2所示。Mosaic数据增强方式丰富了数据集,特别是随机缩放以及多幅图像的拼接

等操作增加了小目标的数量,更利于模型检测。并且模型训练时可以直接计算4幅图像,使 batch\_size

不需要很大就能够得到较好的训练效果,因此节省了计算资源。

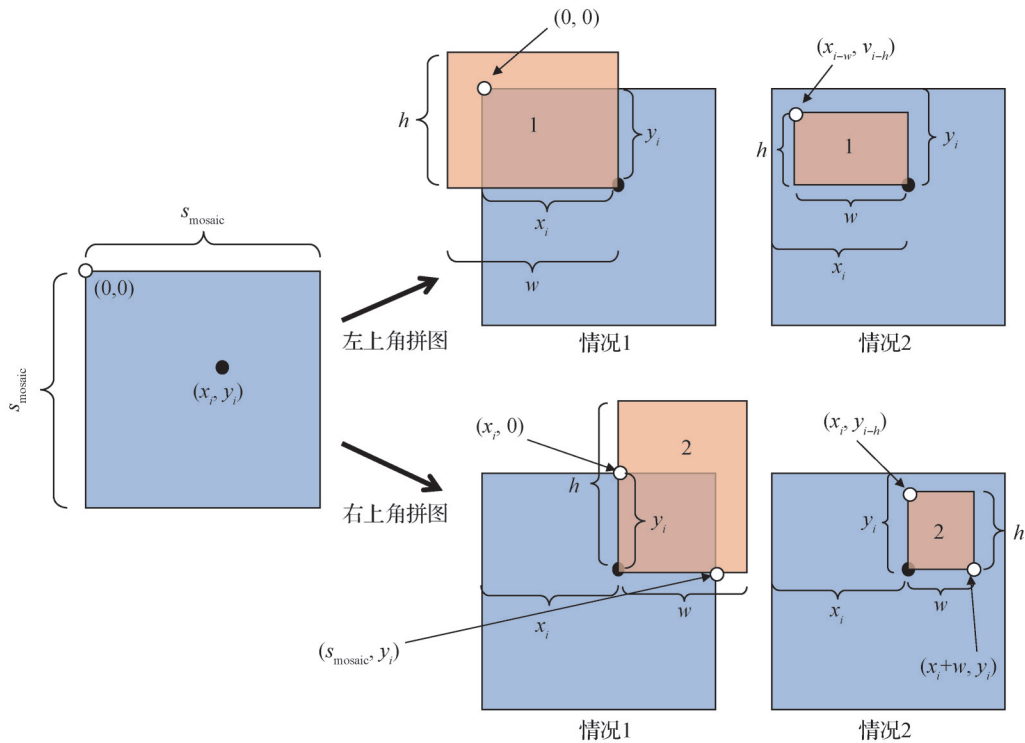


图1 图像拼接过程

Fig. 1 Process of image stitching

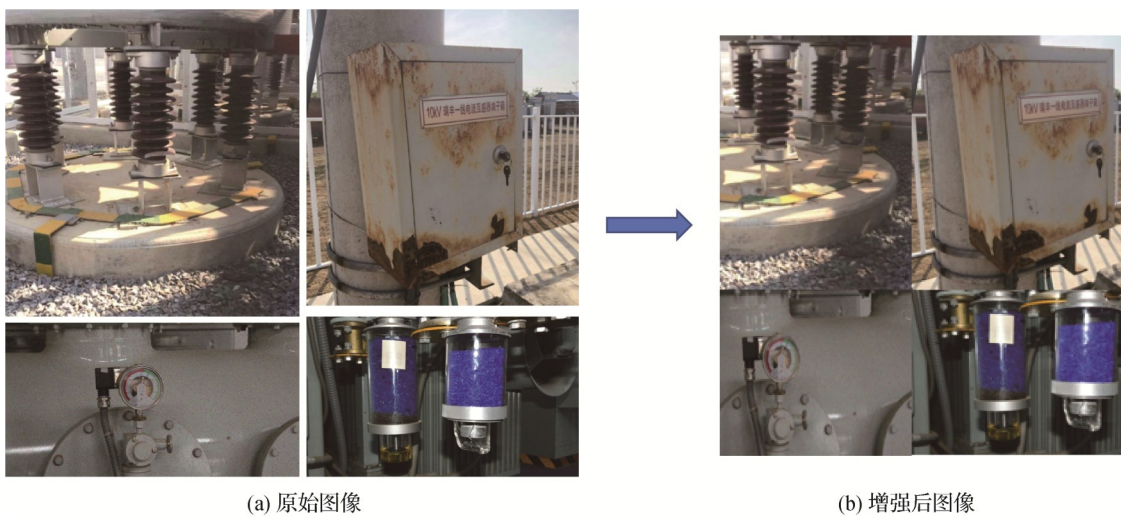


图2 图像增强效果

Fig. 2 Effect of image enhancement ((a) original images; (b) enhanced images)

### 1.3 图像归一化处理

图像预处理对图像进行归一化处理,并调整图像尺寸。输入神经网络的图像特征值越大,模型在反向传播过程中传递到输入层的梯度就越大,这会使学习率受到输入层参数大小的严格限制,此时通

常需要降低学习率,否则模型会跳过最优解。归一化处理将整幅图像的特征值调整到相近的范围,有利于训练过程中动态地调整学习率,使模型更快收敛。调整图像尺寸是为了将尺寸差别较大的图像缩放到统一的范围,并对齐同一训练批次中的所有

图像,使神经网络对同一批次图像的操作能够统一。

## 2 基于层次化分类的缺陷检测方法

层次化结构包括两层,每一层步骤如下:

- 1) 主干网络特征提取。
- 2) 区域生成网络(region proposal network, RPN) (Ren 等,2017)生成建议框。
- 3) 区域特征聚集。
- 4) 多模态特征融合及层次化分类。第 1 层分类

网络的训练执行以下步骤:将图像特征向量依次输入分类网络的两个全连接层,得到本层预测标签和边界回归参数,并保存预测标签;第 2 层分类网络的训练执行以下步骤:首先,将上一层保存的预测标签的向量化表示  $w_{i-1}$  与本层图像特征向量联合嵌入,得到联合向量;然后,将联合向量分别输入分类网络的两个全连接层,得到本层预测标签的向量化表示  $w_i$ 。

5) 分别计算 RPN 损失和分类网络损失,并进行反向传播,更新 RPN 和分类网络参数。

算法每一层的流程示意如图 3 所示。

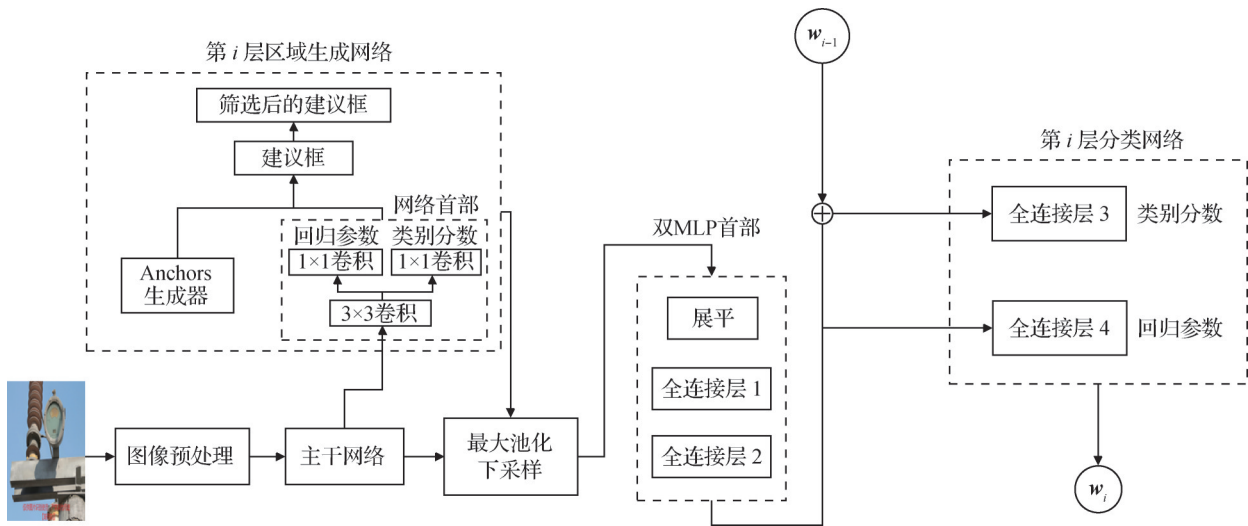


图 3 方法每一层的操作流程

Fig. 3 Process of each layer of the method

### 2.1 主干网络

将 ResNet50(He 等,2016)作为主干网络,对图像进行特征提取,并使用特征金字塔网络(feature pyramid networks, FPN)(Lin 等,2017)对特征图进行多尺度训练和预测。ResNet50 网络及特征金字塔网络的结构如图 4 所示。

其中,ResNet50 网络有 50 层,共分为 5 个块。Conv1(BN+ReLU)块包含 1 个卷积核,之后进行批量归一化和 ReLU 激活函数处理;Conv2\_x 块包含 1 个池化层和 3 个残差块,每个残差块的主分支包含 3 个卷积核,捷径分支包含 1 个卷积核,其中残差块的结构如图 5 所示;Conv3\_x 包含 4 个残差块,每个残差块的主分支包含 3 个卷积核;Conv4\_x 包含 6 个残差块,每个残差块的主分支包含 3 个卷积核;Conv5\_x 包含 3 个残差块,每个残差块的主分支包含 3 个卷积

核。Upsample 为一个上采样块,Maxpool 是一个最大池化层。图像首先从 Conv1 向前传播,每经过一个卷积块会输出特定尺度的特征图,并通过一个大小为  $1 \times 1$  的卷积核分别调整各个尺度特征图的通道数;接着从 C5 得到的特征图开始,自顶向下对较小尺寸的特征图进行上采样后与相邻的较大尺寸特征图进行融合,并将融合后的特征图进行上采样后继续与相邻的较大尺寸特征图进行融合,直到全部特征图融合完成;最后,分别通过大小为  $3 \times 3$  的卷积核后输出上一步每次融合得到的特征图,作为最后的多尺度特征进行后续的预测。

特征金字塔网络将不同尺度的特征图进行融合,并在每个尺度的特征图上进行预测,因此特征金字塔网络能够提高对各种尺寸物体的识别精度,尤其提高了对小目标的检测准确度。

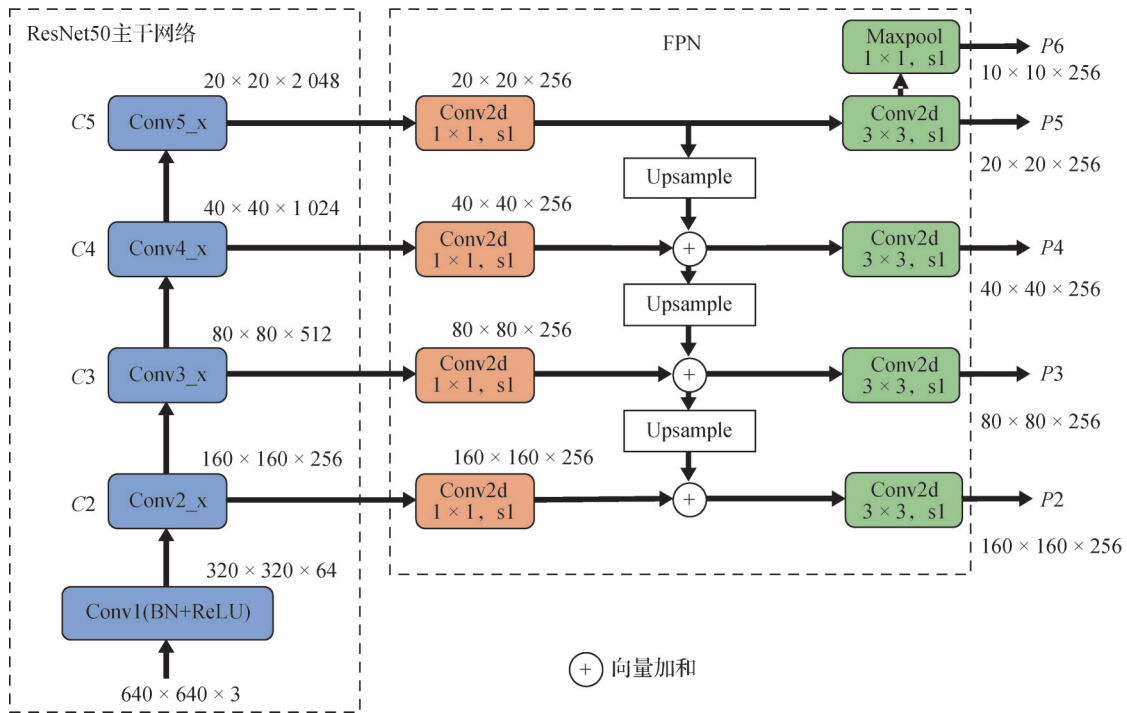


图4 主干网络与特征金字塔网络结构

Fig. 4 Structure of backbone and feature pyramid network

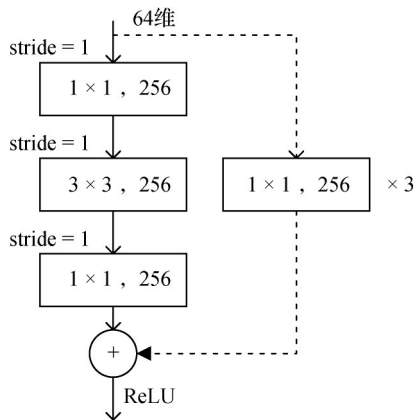


图5 ResNet50-Conv2\_x 结构

Fig. 5 Structure of ResNet50-Conv2\_x

## 2.2 区域生成网络

经过主干网络的特征提取后,采用区域生成网络生成建议框(proposals)。RPN的输入是图像不同尺度的特征图,输出是一系列的建议框。

RPN在每一尺度的特征图上进行操作的具体步骤如下:

1) 利用滑动窗口在特征图上生成  $k$  个锚框(anchor boxes),并将每个锚框分别输入两个全连接层得到  $2k$  个类别概率分数和  $4k$  个边界回归参数,RPN生成锚框的过程如图6所示;

2) 利用边界回归参数将锚框调整为建议框,利

用非极大值抑制算法进行正、负样本划分,其中与实际目标边界重叠度大于0.8的建议框为正样本,反之,与目标边界重叠度小于0.2的建议框为负样本,然后对正样本进行筛选,并随机选取一定数量的负样本;

3) 利用正、负样本分别计算分类损失和边界框定位损失,并通过反向传播调整网络参数。

区域生成网络进行分类的作用是识别出图像中可能存在目标的区域,以便后续的网络进行类别预测,因此RPN输出的类别概率表示该目标为前景或

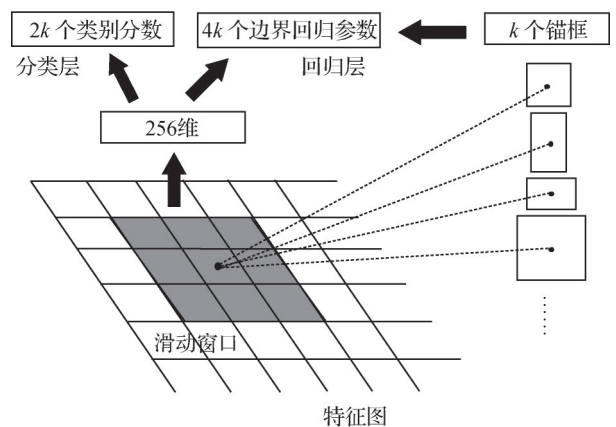


图6 RPN生成锚框过程

Fig. 6 Process of generating anchor boxes in RPN

背景的概率,而后续分类网络输出的类别概率则表示该目标为具体目标类别的概率。

RPN从图像的特征图中选取候选区域,因此相比传统的选择性搜索算法,滑动窗口的计算量大大降低;由于RPN可以和ROI(region of interest) Pooling共用基础的神经网络,实现参数共享,因此RPN可以大大地减少参数量和预测时间。同时,RPN在特征空间生成候选框,可以使网络学习到高层语义的图像特征,这使RPN生成的候选区域具备更高的可靠度。

### 2.3 区域特征聚集

在区域生成网络生成一系列候选框后,采用ROI Align(He等,2017)进行区域特征聚集,其作用是映射RPN生成的候选框位置坐标并池化特征图,将特征图输入后续的分类网络进行分类任务和边界框回归任务。常见的两阶段目标检测算法通常采用ROI Pooling进行区域特征聚集,但是,由于RPN回归得到的候选框位置坐标一般为浮点数,而池化后特征图的尺寸要求为整数,故ROI Pooling需要进行两次量化操作:将候选框边界坐标量化为整数坐标值,以及分割量化后的边界区域,对分割后的每一个单元的边界再进行一次量化。然而,上述两次量化产生的偏差会影响检测的准确度。

为解决上述ROI Pooling的缺点,ROI Align取消量化操作,将整个特征聚集过程转化为连续的操作。而且ROI Align不是将候选区域边界上的坐标点进行补充后池化,而是重新以如下的流程进行操作:

1)对每一个候选区域,保持边界的浮点数值不做量化;

2)将候选区域平均分割成若干个单元,单元边界不做量化;

3)在每个单元中选取固定的4个坐标位置进行最大池化。

### 2.4 多模态特征融合

目前常用的多模态特征融合方法有3种,第1种是特征拼接(Huang等,2021);第2种是跨模态注意(Zhang等,2023);第3种是条件批量归一化(Liao等,2022)。由于模型所涉及的文本数据为电力设备缺陷的父类别标签,其中包含的语义信息有限,因此为了降低计算复杂度,提出的层次化分类方法采用基于特征拼接的多模态特征融合方法。特征融合过程如图7所示。

首先,通过特征提取、区域生成以及特征聚集等

一系列操作得到图像特征向量,然后将此特征向量展开得到图像特征向量 $P_l$ ;接着利用Word2Vec进行父类别标签的词向量表示,得到父类别标签的词向量 $W_{l-1}$ ;然后将展开后的图像特征向量 $P_l$ 和父类别标签对应的词向量 $W_{l-1}$ 在第二维度上进行拼接,得到联合向量 $V_l$ ,具体为

$$V_l = P_l \oplus W_{l-1} \quad (3)$$

式中,符号 $\oplus$ 表示向量拼接运算。最后使用一个全连接层和一个softmax层,输出最终的目标类别分数预测结果,计算过程为

$$\begin{cases} V'_l = \text{ReLU}(K_1 V_l + B_1) \\ W_l = \text{softmax}(K_2 V'_l + B_2) \end{cases} \quad (4)$$

式中, $K$ 为权重, $B$ 为偏置项, $V'_l$ 为经过ReLU激活函数变换的结果, $W_l$ 为第 $l$ 层分类网络输出的分类结果。每层分类网络均进行上述操作,直到得出最后一层的分类结果。

由于电力设备数据集的标签层级结构共两层,因此本文方法的结构也共有两层,每一层的分类网络中包含两个全连接层,其中一个全连接层预测目标的类别分数,另外一个全连接层预测目标的边界框回归参数。为了不使父类别标签的特征影响边界框回归参数的预测,将融合后的特征 $V_l$ 只输入到预测目标类别分数的全连接层,而将单一的图像特征 $P_l$ 输入预测目标边界框回归参数的全连接层。

### 2.5 层次化分类网络

目前,在文本分类和图像分类等领域中,层次化分类算法可分为两类:全局算法和局部算法。全局算法(Aly等,2019)采用一个分类器,输出样本在所有层次类别上的分类结果。局部算法大致有3种,第1种算法针对每个结点,训练一个二分类器,判断样本是否属于该结点表示的标签;第2种算法针对每个父节点,训练一个多分类器,下一层分类器的分类结果则由该父节点所属类别的子类别产生,依此类推;第3种算法为每一层训练一个多分类器,每个分类器负责训练当前层的所有类别。全局算法不会产生层级之间的错误传递,且模型的参数量较少,但难以有效地学习到各个层级内部的信息。局部算法的计算成本较高,而且会产生层级之间的错误传递。

提出的层次化分类方法为局部算法,其中的层次化分类网络共有两个分支,分别是父类别预测分

支和子类别预测分支。父类别预测分支的预测结果嵌入到子类别预测分支,以辅助子类别的分类。层次化分类网络结构如图8所示。首先,针对电力设备缺陷目标父标签类别差异大、子标签类别差异小的特点,增大父分类器的学习率参数,减小子分类器

的学习率参数,加快模型的收敛速度。其次,利用父分类器隐藏层训练参数初始化子分类器隐藏层参数,并对子分类器网络参数进行微调,提升模型的检测速度,图中 $W$ 表示子类别分类器使用父类别分类器权重矩阵。

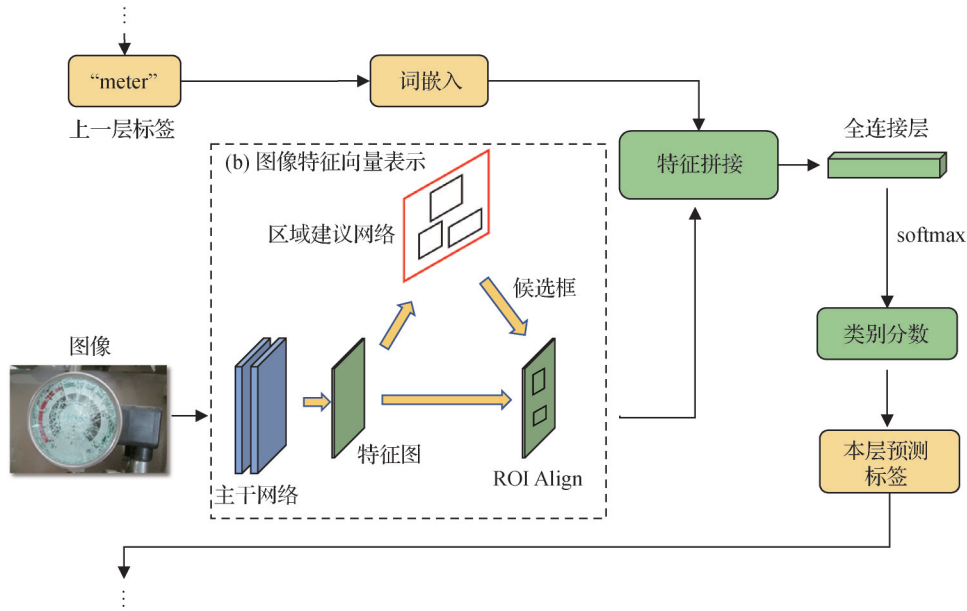


图7 多模态特征融合

Fig. 7 Multi-modal feature fusion

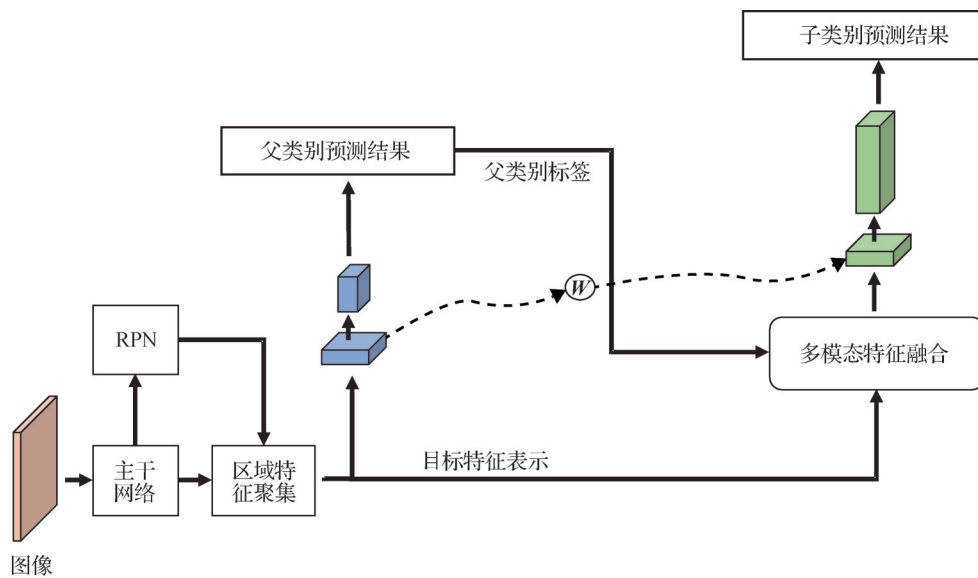


图8 层次化分类网络结构

Fig. 8 Architecture of hierarchical classification network

### 3 实验

#### 3.1 参数设置

实验相关的参数设置如表2所示。其中,缺陷

图像数据集的训练集、验证集划分比例为8:2, VOC2012(visual object classes)数据集的划分比例按照默认设置;实验采用动态学习率的训练策略,每3轮调整一次学习率。图像裁剪为图像预处理阶段进行的操作。

表 2 实验参数设置  
Table 2 Experimental parameter settings

参数名称	值
训练集与验证集划分比例	8:2
batch_size, epoch	8, 20
初始学习率, 调整倍数	0.005, 0.3
RPN 进行 NMS 处理的 IoU 阈值	0.8
RPN 生成的 anchor 尺寸	{32 <sup>2</sup> , 64 <sup>2</sup> , 128 <sup>2</sup> , 256 <sup>2</sup> , 512 <sup>2</sup> }
每个 anchor 的长、宽比例	0.5, 1.0, 2.0
图像裁剪后的最小分辨率/像素	800 × 800
图像裁剪后的最大分辨率/像素	1 333 × 1 333

由于需要保证将图像的训练顺序打乱以获得更好的泛化能力, 同时还需要保证本文的层次化分类网络能够将上一层预测的标签与下一层网络中对应的图像特征图进行联合嵌入, 因此从以下两个方面对实验过程进行优化:

1) 由于实验采用单 GPU 训练, 因此设置随机种子 (seed) 来保证每一层分类网络在训练时, 每个 epoch 的图像训练顺序是不同的, 但是不同层分类网络相对应 epoch 的图像训练顺序是一致的。

2) 将每幅图像生成的候选框数量固定为 500。

### 3.2 评价指标

采用平均精度均值 (mean average precision, mAP) 作为性能评价指标, 该指标综合了模型预测的精准率 (precision) 和召回率 (recall), 表示该数据集所有类别的平均检测准确率。

通过 PR (precision-recall) 曲线在 (0, 1) 的积分来反映模型对于该类别目标的检测精度, 该积分表示为 AP。

### 3.3 对比算法

基于电力设备缺陷检测任务, 本文不仅与最新的电力设备缺陷检测算法 (毛进伟, 2021) 进行了对比, 还与目前较为流行的常用目标检测算法 (Zhang 等, 2022; Zhao 等, 2019; Carion 等, 2020; Tan 等, 2020) 进行了比较。其中, Zhao 等人 (2019) 整合了 MLFPN 和 SSD, 能够融合主干网络中的多层特征图作为基本特征; Tan 等人 (2020) 引入更加快速的双向多尺度和加权特征融合; Carion 等人 (2020) 利用自注意力机制将图像特征图和两个维度的位置编码进行结合并对图像特征图进行全局分析; Zhang 等人 (2022) 为了降低计算复杂度, 在 Carion 等人

(2020) 的基础上利用 Siamese Network 使 object query 更轻松地聚焦到特定区域, 降低了匹配的难度。

### 3.4 实验数据集

本文采用电力设备缺陷图像数据集和基准数据集 PASCAL (pattern analysis, statistical modeling and computational learning) VOC2012。电力设备缺陷数据集各类标签的名称、数量及标签结构如表 1 所示。缺陷图像数据集包括了 11 911 幅图像文件, 其中包含的目标总数为 15 361 个, 各目标数量占比如表 1 所示。每幅图像对应的子类别标签标注文件和父类别标签标注文件大小为 1 KB。缺陷数据集总大小为 4 GB, 每幅图像为 50 KB~1.2 MB 不等, 每幅图像的分辨率为 540 × 960 像素~1 280 × 848 像素不等。其中, 表计类、呼吸器类、绝缘子类和异物类目标尺寸较小, 渗漏油类、金属锈蚀类目标尺寸大小不一; 表计类、呼吸器类和绝缘子类子类别缺陷特征相似度较高, 异物类和渗漏油类子类别缺陷特征相似度较低。

在 VOC2012 数据集上的实验旨在验证本文方法的通用性, 该数据集具备丰富的目标类别和良好的图像标注, 并且类别标签之间也具有一定的层次关系, 如表 3 所示。其中第 1 行为第 1 层标签, 即 L1 层; 第 1 层每个标签所在列的其他标签为该标签对应的第 2 层标签, 即 L2 层。

表 3 VOC2012 数据集的目标类别标签层次关系

Table 3 Hierarchical relationship of object category label in VOC2012 dataset

层次	类别标签			
L1 层	Vehicles	Household	Animals	其他
	Aero plane	Bottle	Bird	Person
	Train	Chair	Cat	
L2 层	Car	Table	Horse	
	Bus	Potted plant	Sheep	
	Motorbike	Sofa	Dog	
	Boat	Monitor	Cow	
	Bicycle			

### 3.5 实验结果与分析

#### 3.5.1 消融实验

为验证本文方法中特征提取网络、多尺度训练模块以及层次化分类结合多模态特征融合方法的有效性, 本文设计了 4 种不同的模型: 1) 以 Mobilenet\_

v3 (Howard 等, 2019) 作为主干网络的模型; 2) 以 ResNet50 作为主干网络的模型; 3) 以 ResNet50 作为主干网络并使用特征金字塔网络进行多尺度训练的模型; 4) 以 ResNet50 作为主干网络、使用特征金字塔网络进行多尺度训练并使用层次化分类结合多模态特征融合的模型。各模型在电力设备缺陷图像每一层标签上的 mAP 如表 4 所示。表中 mAP(L1 层) 表示模型在第 1 层标签检测的 mAP 值, mAP(L2 层) 表示模型在第 2 层标签检测的 mAP 值。从表 4 可以看出, ResNet50 网络对于目标特征的表达能力更强, 因为该网络结构中卷积层数量较多且引入了残差结构, 保证梯度正常传播。FPN 能够实现不同尺度特征图的融合, 并且随着特征的融合, 较浅层的特征图同时具有深层特征图和浅层特征图的信息。这样不仅实现了在不同尺度特征图上的预测, 还有效地提升了模型对小目标的检测能力, 解决了缺陷数据集中表计、呼吸器等小目标难以检测的问题。另一方面, 第 1 层标签的分类平均准确率达到 93.6%, 表明模型很好地学习到了父类别的特征, 将第 1 层预测的父类别标签与第 2 层的图像特征进行联合嵌入, 父类别标签的语义特征有助于增强样本在第 2 层的特征表示能力和分类性能, 模型也能够学习父类别标签与子类别标签的关系, 通过层次化分类, 提高了子类别的分类准确率。

表 4 缺陷图像数据集检测准确率  
Table 4 Detection accuracy (mAP) of defect image dataset

模型	mAP/%	
	L1 层	L2 层
Mobilenet_v3	87.6	79.3
ResNet50	91.3	82.8
ResNet50+FPN	93.5	84.7
ResNet50+FPN+多模态特征融合层次化分类	<b>93.6</b>	<b>86.4</b>

注: 加粗字体表示各列最优结果。

### 3.5.2 准确率对比实验

本文分别在电力设备缺陷数据集和 VOC2012 数据集上进行了准确率对比实验, 其中模型在电力设备缺陷数据集上的准确率对比实验结果如表 5 所示, 模型在 VOC2012 基准数据集上的准确率对比实验结果如表 6 所示。

从实验结果可知, 模型在电力设备缺陷数据集的准确率达到 86.4%, 与流行的缺陷检测方法 (毛进伟, 2021) 和常用的目标检测模型 (Zhao 等, 2019; Tan 等, 2020; Zhang 等, 2022; Carion 等, 2020) 相比, 准确率分别提升了 5.9%, 5.1%, 6.4%, 6.2%, 7.6%。尤其对于表计、呼吸器等小目标, 本文算法的检测精度较高, 且优于其他算法。由于父类别标签特征的辅助, 表计、呼吸器等父类别所对应的子类别的准确率都得到了提升, 而异物缺陷中的“鸟巢”和“挂空悬浮物”两个类别的目标特征相关性较低, 导致父类别特征无法得到有效的表示, 子类别“鸟巢”和“挂空悬浮物”在没有充分利用父类别特征的情况下, 分类准确率相对较低。

相比其他算法, 本文算法在基准数据集上也表现良好。但是, 本文算法和对比算法在 VOC2012 数据集上的准确率均低于在缺陷数据集上的准确率。相较于电力设备缺陷图像数据, VOC2012 基准数据集中图像的类别不具有明显的层次结构, 这极大地削弱了本文算法的优势。但是在训练过程中, 父类别标签与图像特征进行联合嵌入能够丰富图像特征, 使原本难分的样本更易区分, 因此相较于其他方法, 本文提出的方法依然具有一定的优势。

### 3.5.3 检测时间对比实验

接下来计算各个模型在数据测试集上的平均检测时间, 即测试集上所有样本检测时间的平均值, 其实验结果如图 9 所示。结合上述实验结果可以看出, 本文算法作为一种两阶段的目标检测方法, 与同类的两阶段检测对比算法 (毛进伟, 2021; Zhao 等, 2019) 相比, 在提升检测准确率基础上, 其平均检测时间维持在较低水平。另外, 本文方法在平均检测时间上略高于 Tan 等人 (2020) 的一阶段检测方法以及基于注意力机制的对比方法 (Zhang 等, 2022; Carion 等, 2020)。

### 3.5.4 可视化效果

本文算法以及其他基线算法在电力设备缺陷数据集上的实验结果的可视化效果如图 10 所示。从图中可以看出, 本文算法能够精确地定位到目标位置并将目标准确分类, 其他算法已经能够较为准确地检测出缺陷所属的类别, 但是对于缺陷目标的定位存在偏差, Zhang 等人 (2022)、Zhao 等人 (2019) 的 Tan 等人 (2020) 的算法存在重复检测的情况, 毛进伟 (2021) 的算法甚至存在误检的情况。

表 5 不同算法在电力设备缺陷数据集上的检测准确率结果

Table 5 Detection accuracy results of different algorithms on power equipment defect dataset

设备缺陷	毛进伟(2021)	Zhao 等人(2019)	Tan 等人(2020)	Zhang 等人(2022)	Carion 等人(2020)	本文
表计破损	85.7	86.1	84.5	83.2	82.3	<b>92.7</b>
表盘模糊	85.5	84.3	82.1	82.6	81.5	<b>89.8</b>
表计正常	84.2	85.7	85.2	83.3	83.7	<b>90.2</b>
绝缘子破裂	79.4	80.2	78.3	82.2	80.5	<b>87.4</b>
绝缘子正常	80.6	81.1	79.9	81.5	78.3	<b>85.3</b>
硅胶变色	82.3	79.4	80.7	83.7	82.1	<b>86.7</b>
硅胶正常	82.6	81.2	80.6	77.6	73.6	<b>87.9</b>
油封破损	81.3	82.3	80.8	76.5	74.5	<b>86.4</b>
油封正常	80.6	81.8	81.3	78.2	78.4	<b>86.8</b>
挂空悬浮物	81.6	<b>83.5</b>	80.5	74.1	73.8	82.4
鸟巢	77.7	80.3	78.3	80.3	76.7	<b>82.1</b>
部件表面油污	73.8	75.8	75.2	78.2	77.8	<b>81.9</b>
地面油污	74.4	77.4	76.4	79.3	78.3	<b>84.3</b>
金属锈蚀	77.3	79.1	78.1	83.2	81.7	<b>85.8</b>
mAP	80.5	81.3	80	80.2	78.8	<b>86.4</b>

注:加粗字体表示各行最优结果。

表 6 VOC2012数据集上的检测结果

Table 6 Detection results (mAP) of different algorithms on VOC2012 dataset

模型	mAP(L1层)	mAP(L2层)
毛进伟(2021)	-	81.2
Zhao 等人(2019)	-	81.4
Tan 等人(2020)	-	80.2
Zhang 等人(2022)	-	80.8
Carion 等人(2020)	-	79.5
本文	92.6	<b>82.5</b>

注:加粗字体表示最优结果,“-”表示模型不支持进行相关实验。

各算法在 VOC2012 数据集上的实验结果可视化效果如图 11 所示。从图中可以看出,相较于其他算法,本文算法能够精确地定位到目标的位置并将目标准确分类。类似地,毛进伟(2021)的算法存在误检的情况,而 Carion 等人(2020)和 Zhang 等人(2022)的算法存在重复检测的情况。

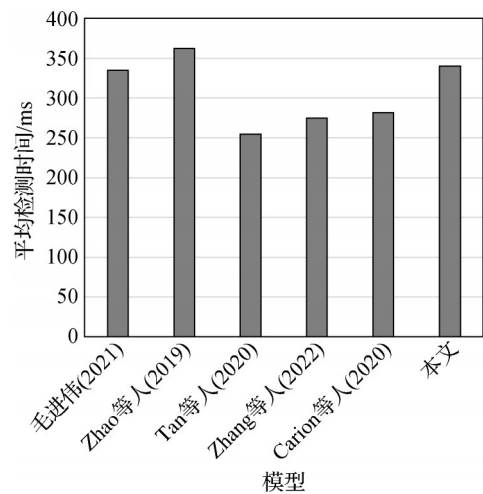


图 9 各算法在设备缺陷数据集上的检测时间结果  
Fig. 9 Detection time results of various algorithms on the equipment defect dataset

## 4 结 论

为提高变电站中复杂电力设备环境下的缺陷检

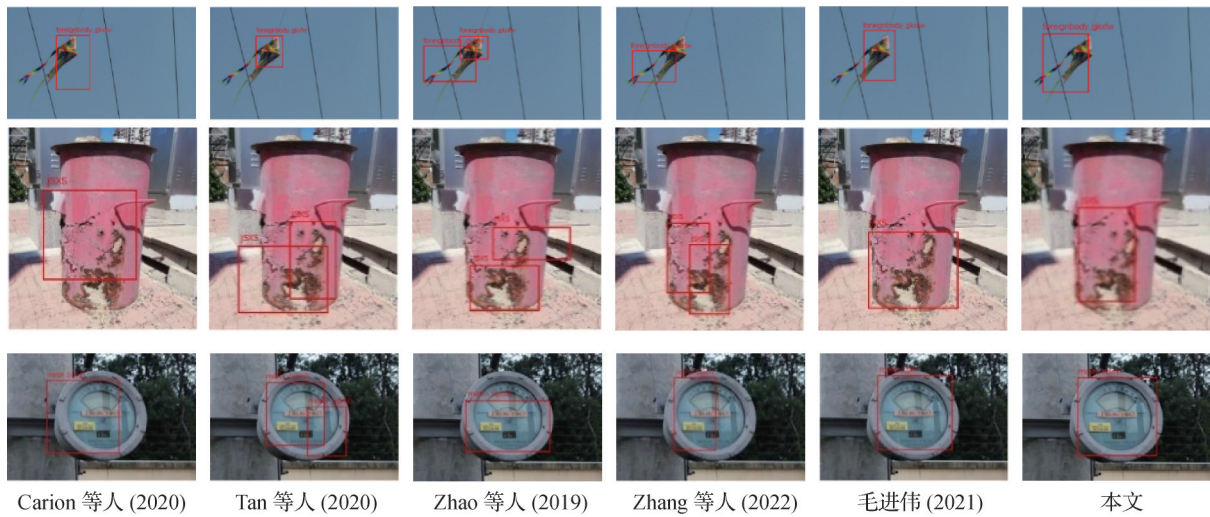


图10 电力设备缺陷数据集检测结果的可视化效果

Fig. 10 Visualization of detection results in the power equipment defect dataset

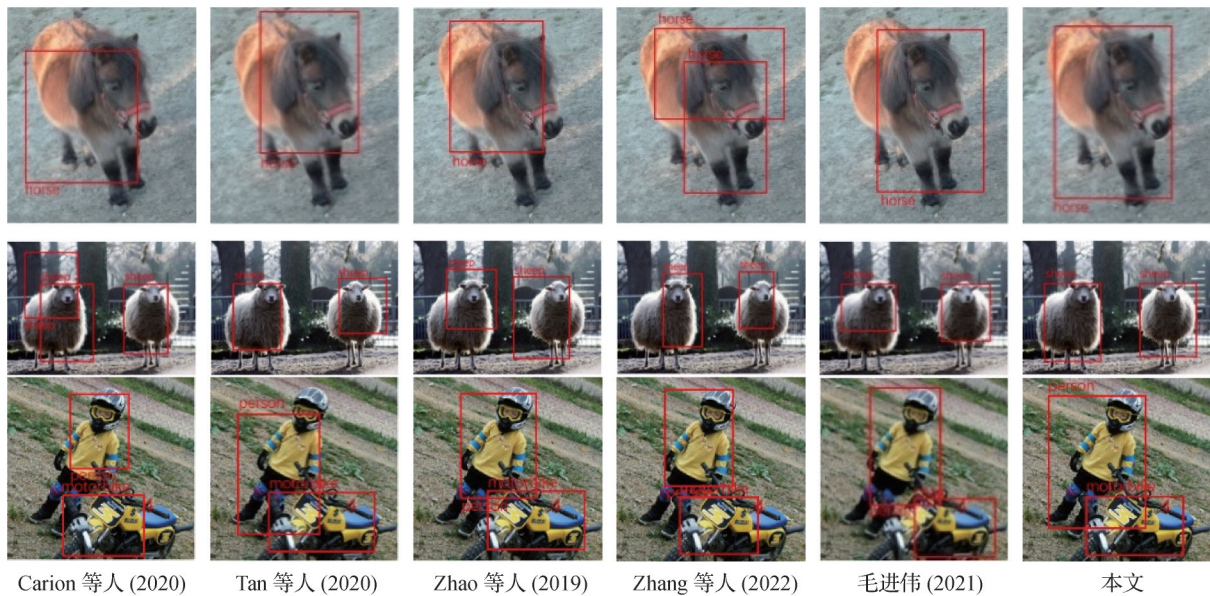


图11 VOC2012数据集检测结果的可视化效果

Fig. 11 Visualization of detection results in the VOC2012 dataset

测准确率,提出一种多模态层次化分类方法用于电力设备缺陷检测。通过分析电力设备缺陷标签之间的语义关系,完善了具有层次化标签结构的电力设备缺陷图像数据集。针对提出的基于多模态层次化分类网络模型,在缺陷数据集和基准数据集上进行了实验验证,实验结果显示本文方法优于当前流行的缺陷检测方法,获得了较高的准确率,尤其对于存在明显层次关系类别的准确率更高。本文方法为变电站设备故障的全面、准确诊断提供了一种有效的技术支撑。

由于词向量嵌入对父类别标签的操作引入了额外耗时,导致本文方法的计算耗时还有待降低,因此

将来考虑采用统一区域生成网络及分类网络以缩短检测时间。另外,考虑到图像类别极其不均衡情形下的设备缺陷高效检测,进一步研究基于长尾分布的图像分类方法。

## 参考文献 (References)

- Aly R, Remus S and Biemann C. 2019. Hierarchical multi-label classification of text with capsule networks//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. Florence, Italy: ACL: 323-330 [DOI: 10.18653/v1/P19-2045]
- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal

- speed and accuracy of object detection [EB/OL]. [2023-05-29].  
<https://arxiv.org/pdf/2004.10934.pdf>
- Cao J L, Li Y L, Sun H Q, Xie J, Huang K Q and Pang Y W. 2022. A survey on deep learning based visual object detection. *Journal of Image and Graphics*, 27(6): 1697-1722 (曹家乐, 李亚利, 孙汉卿, 谢今, 黄凯奇, 庞彦伟. 2022. 基于深度学习的视觉目标检测技术综述. *中国图象图形学报*, 27(6): 1697-1722) [DOI: 10.11834/jig.220069]
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 213-229
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Howard A, Sandler M, Chen B, Wang W J, Chen L C, Tan M X, Chu G, Vasudevan V, Zhu Y K, Pang R M, Adam H and Le Q. 2019. Searching for MobileNetV3//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 1314-1324 [DOI: 10.1109/ICCV.2019.00140]
- Huang Z C, Zeng Z Y, Huang Y P, Liu B, Fu D M and Fu J L. 2021. Seeing out of the box: end-to-end pre-training for vision-language representation learning//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 12971-12980 [DOI: 10.1109/CVPR46437.2021.01278]
- Liao W T, Hu K, Yang M Y and Rosenhahn B. 2022. Text to image generation with semantic-spatial aware GAN//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 18166-18175 [DOI: 10.1109/CVPR52688.2022.01765]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Mao J W. 2021. Research on Transformer Image Analysis Method Based on Deep Learning. Nanjing: Nanjing Normal university (毛进伟. 2021. 基于深度学习的变电图像分析方法研究. 南京: 南京师范大学)
- Qi Y C, Wu X L, Zhao Z B, Shi B Q and Nie L Q. 2021. Bolt defect detection for aerial transmission lines using Faster R-CNN with an embedded dual attention mechanism. *Journal of Image and Graphics*, 26(11): 2594-2604 (戚银城, 武学良, 赵振兵, 史博强, 聂礼强. 2021. 嵌入双注意力机制的Faster R-CNN航拍输电线路螺栓缺陷检测. *中国图象图形学报*, 26(11): 2594-2604) [DOI: 10.11834/jig.200793]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10778-10787 [DOI: 10.1109/CVPR42600.2020.01079]
- Ying Y, Wang Y Z, Yan Y F, Dong Z K, Qi D L and Li C Y. 2020. An improved defect detection method for substation equipment//Proceedings of the 39th Chinese Control Conference. Shenyang, China: IEEE: 6318-6323 [DOI: 10.23919/CCC50068.2020.9189042]
- Zhang G J, Luo Z P, Yu Y C, Cui K W and Lu S J. 2022. Accelerating DETR convergence via semantic-aligned matching//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 939-948 [DOI: 10.1109/CVPR52688.2022.00102]
- Zhang J M, Liu H Y, Yang K L, Hu X X, Liu R P and Stiefelhagen R. 2023. CMX: cross-modal fusion for RGB-X semantic segmentation with transformers. *IEEE Transactions on Intelligent Transportation Systems*, 24(12): 14679-14694 [DOI: 10.1109/TITS.2023.3300537]
- Zhao Q J, Sheng T, Wang Y T, Tang Z, Chen Y, Cai L and Ling H B. 2019. M2Det: a single-shot object detector based on multi-level feature pyramid network//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI Press: 9259-9266 [DOI: 10.1609/aaai.v33i01.33019259]
- Zhao W Q, Zhang H M and Xu M F. 2021. Insulator recognition based on an improved scale-transferrable network. *Journal of Image and Graphics*, 26(11): 2561-2570 (赵文清, 张海明, 徐敏夫. 2021. 面向改进尺度缩放网络的绝缘子识别. *中国图象图形学报*, 26(11): 2561-2570) [DOI: 10.11834/jig.200697]

## 作者简介

白艳峰,男,硕士研究生,主要研究方向为人工智能、目标检测。E-mail: yanfengbai@ncepu.edu.cn

马应龙,通信作者,男,教授,主要研究方向为人工智能与知识工程、大数据分析和软件工程。

E-mail: yinglongma@ncepu.edu.cn

王立彪,男,硕士研究生,主要研究方向为人工智能、目标检测。E-mail: libiaowang@ncepu.edu.cn

高卫东,男,硕士研究生,主要研究方向为人工智能、目标检测。E-mail: 1033876073@qq.com