

基于多模态的缺陷绝缘子图像的多标签分类

周 景, 王满意, 田兆星

(华北电力大学控制与计算机工程学院, 北京 102206)

摘 要: 对巡检图像中绝缘子缺陷准确分类是输电线路自动巡检领域中的关键技术之一。针对传统深度学习的分类方法对文本信息利用不够充分以及绝缘子图像分类标签较为单一的问题, 该文首次提出了一种基于多模态的缺陷绝缘子图像的多标签分类方法。首先, 采用一种多模态联合数据增强方法, 实现了绝缘子图像和标签文本间跨模态的数据增强。然后, 使用 Vision Transformer 网络提取图像的特征信息和 BERT 网络提取标签文本的特征信息, 充分利用图像和标签文本的特征信息, 从不同模态获取全面的信息, 提高了网络的分类能力。最后, 通过对比学习的方式将图像和文本的特征信息关联, 增强网络分类的可靠性的同时, 又为分类结果提供了良好的可解释性。实验结果表明, 该方法的分类总体准确率达到 93.87%, 在同一数据集中对比其他模型, 分类性能具有明显优势, 为多模态技术在电网领域的应用提供了较好的基础。

关键词: 绝缘子图像; 多标签分类; 多模态; 对比学习; 数据增强

Multi-label Classification of Defective Insulator Images Based on Multimodality

ZHOU Jing, WANG Manyi, TIAN Zhaoxing

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

Abstract: Accurate classification of insulator defects in inspection images is one of the key technologies in the field of automatic inspection of transmission lines. To address the issue of the insufficient utilization of textual information by traditional deep learning classification methods and the issue of relatively simplistic insulator image classification labels, this paper proposes for the first time a multi-label classification method for defective insulator images based on a multi-modal approach. Firstly, a multimodal joint data augmentation method is employed, achieving cross-modal data enhancement between insulator images and label texts. Then, the Vision Transformer network is utilized to extract features from images, and the BERT network is used to extract features from label texts, fully leveraging the feature information from both images and label texts to obtain comprehensive information from different modalities, thereby enhancing the network's classification capabilities. Finally, through correlating the feature information of images and texts via contrastive learning, the reliability of network classification is enhanced, while also providing good interpretability for the classification results. The experimental results demonstrate that this method achieves an overall accuracy rate of 93.87%, showing a significant advantage in classification performance over other models on the same dataset.

Key words: insulator images; multi-label classification; multi-modal; contrastive learning; data augmentation

0 引言

绝缘子作为电路线路主要的组成部分, 对于电网的稳定运行有重要作用^[1]。绝缘子长期在野外及强电场环境中会产生多种故障, 导致其性能降低, 进而直接威胁电力系统的安全运行^[2]。绝缘子可分为玻璃和陶瓷材质。玻璃绝缘子透明度高, 便于检查内部裂纹, 但易破损; 陶瓷绝缘子耐高温, 耐腐

蚀, 但自洁性较差。两者常见故障包括破损、闪络和污渍, 分别由内部应力、表面污染或积水引起。针对这些故障, 采取适宜的维护和预防措施对提高绝缘子性能至关重要。同时, 在实际应用中, 每个图像有多个标签还可以为后续数据处理和分析提供更多信息。因此, 对缺陷绝缘子进行细粒度化的多标签分类的方法, 具有非常重要的研究意义^[3]。

通过计算机视觉技术对缺陷绝缘子图像进行分类的方法可以分为传统图像分类方法和深度学习图像分类方法。传统图像分类方法一般包括底层特征学习、特征编码、空间约束、分类器设计等阶段。

基金资助项目: 国家电网公司科技项目 (5108-202218280A-2-400-XG)。
Project supported by Science and Technology Project of SGCC (5108-202218280A-2-400-XG).

文献[4]使用 Otsu 算法和形态学滤波进行图像预处理。采用免疫遗传蛇算法对绝缘子串凹区边界进行优化, 使绝缘子串轮廓更加完整。然后使用椭圆拟合和连通区域方法计算绝缘子的数量, 从而确定是否有绝缘子缺失。文献[5]首先利用分割图像块的颜色统计以及连通域的形状和边界链编码特征来识别玻璃绝缘子所在区域。然后, 采用滑动窗口直方图统计和直方图匹配判定方法对玻璃绝缘子破损区域进行识别。文献[6]提出了一种基于红蓝色差和改进 K-means 算法的无人机检测绝缘子准确识别方法。文献[7]提出了一种输电线路绝缘子故障诊断方法, 将二值支持向量机分类器与贝叶斯优化相结合, 对绝缘子闪络红外光谱进行分类识别。综上所述, 传统图像分类方法需要手动设计特征提取算法, 且难以处理复杂、多样化的图像数据。

深度学习的图像分类方法主要基于深度神经网络, 通过自动学习图像中的特征来实现分类任务。文献[8]利用 VGGNet^[9]框架神经网络算法对样本进行分类训练和状态预测评估, 最终达到判断绝缘子是否劣化的目的。文献[10]采用重构的 ResNet50^[11]架构提高了对绝缘子缺陷程度数据集的特征提取能力, 并进行多尺度特征融合, 能够完成不同程度绝缘子缺陷的识别。文献[12]通过添加自注意力与卷积混合注意力机制更加关注特征中的细节, 进一步区分不同的绝缘子缺陷。文献[13]提出一种基于层次多任务深度学习的绝缘子自爆缺陷检测模型, 使用专用的卷积神经网络区分自爆绝缘子和正常绝缘子, 并结合多任务学习和特征融合方法提高分类准确率。

目前, 绝缘子图像分类方法都是采用单一分类策略, 即将绝缘子图像分为正常和缺陷两类, 或者只针对某种特定缺陷类型进行标注。针对缺陷的绝缘子图像, 这种单一分类策略无法提供更加精细、全面的图像分类结果, 不利于后续处理和维修。对每个绝缘子图像引入多种标签, 包括缺陷类型、材质等, 这样既可以更加准确、全面地描述缺陷绝缘子图像的缺陷信息, 又能够更好地服务于电力系统后续的数据处理和故障检测与维修。

同时, 传统的深度学习图像分类方法对标签文本信息利用不够充分。在深度学习方法中, 模型训练所使用的数据集包含两种不同的模态信息: 缺陷绝缘子的图像信息和对应的标签文本信息, 它们以图像-文本对的形式存在于数据集中且互相匹配。目

前, 传统的绝缘子图像分类方法只关注单一模态图像上的特征信息, 因此被归类为单模态方法。通常情况下, 标签文本包含了与缺陷绝缘子密切相关的重要信息, 如材质、缺陷类型等。然而传统的单模态方法没有提取并利用标签文本上的特征信息, 导致模型在训练数据集上的信息利用不足。另外, 单模态方法在处理噪声、光照变化和视角变化等方面的鲁棒性也相对较差, 容易受到外界因素的干扰。此外, 单模态方法无法提供更深入和具体的解释说明, 使得分类结果的可解释性较差。

多模态技术的发展和已证明了综合利用多种模态信息的优势。在绝缘子图像缺陷分类任务中, 图像和文本属于不同的模态, 采用多模态方法有三大优势: 首先, 多模态方法能够综合考虑绝缘子图像和标签文本的特征, 从不同角度获取全面的信息, 从而提高分类的准确性和鲁棒性。其次, 通过比较图像特征和标签文本特征的一致性, 可以降低误判的概率, 提高分类结果的可靠性。最后, 多模态方法能够提供良好的可解释性, 因为标签文本信息可以增加对于绝缘子分类结果的解释和说明。综上所述, 本文提出了一种基于多模态^[14]的绝缘子图像缺陷分类方法。本文的主要贡献如下所述。

1) 本文提出了一个基于多模态的缺陷绝缘子图像多标签分类方法。在该方法中, 采用 Vision Transformer^[15]网络提取图像特征, 采用 BERT^[16]网络提取标签文本特征, 通过对比学习计算两种模态的特征信息的相似度, 进而完成对缺陷绝缘子图像的多标签分类任务。

2) 在将数据集集中的图像-文本对输入特征提取网络之前, 使用一种多模态联合数据增强方法对其进行处理, 既保留了图像-文本对内的语义关系, 又丰富了数据集的内容, 提高了数据的使用效率。最后在相同数据集上与其他分类方法进行对比实验和消融实验, 结果表明本文方法有显著优势。

1 基于多模态的缺陷绝缘子图像的多标签分类模型

模型的总体结构如图 1 所示, 其结构可以分为文本编码器、图像编码器、对比学习和混合生成 4 部分。模型的输入是若干个图像-文本对, 图像编码器采用 Vision Transformer 网络模型, 负责提取图像的特征信息。文本编码器采用 BERT 网络模型, 图像对应的文本包含多个标签, 需要将其按转换为一

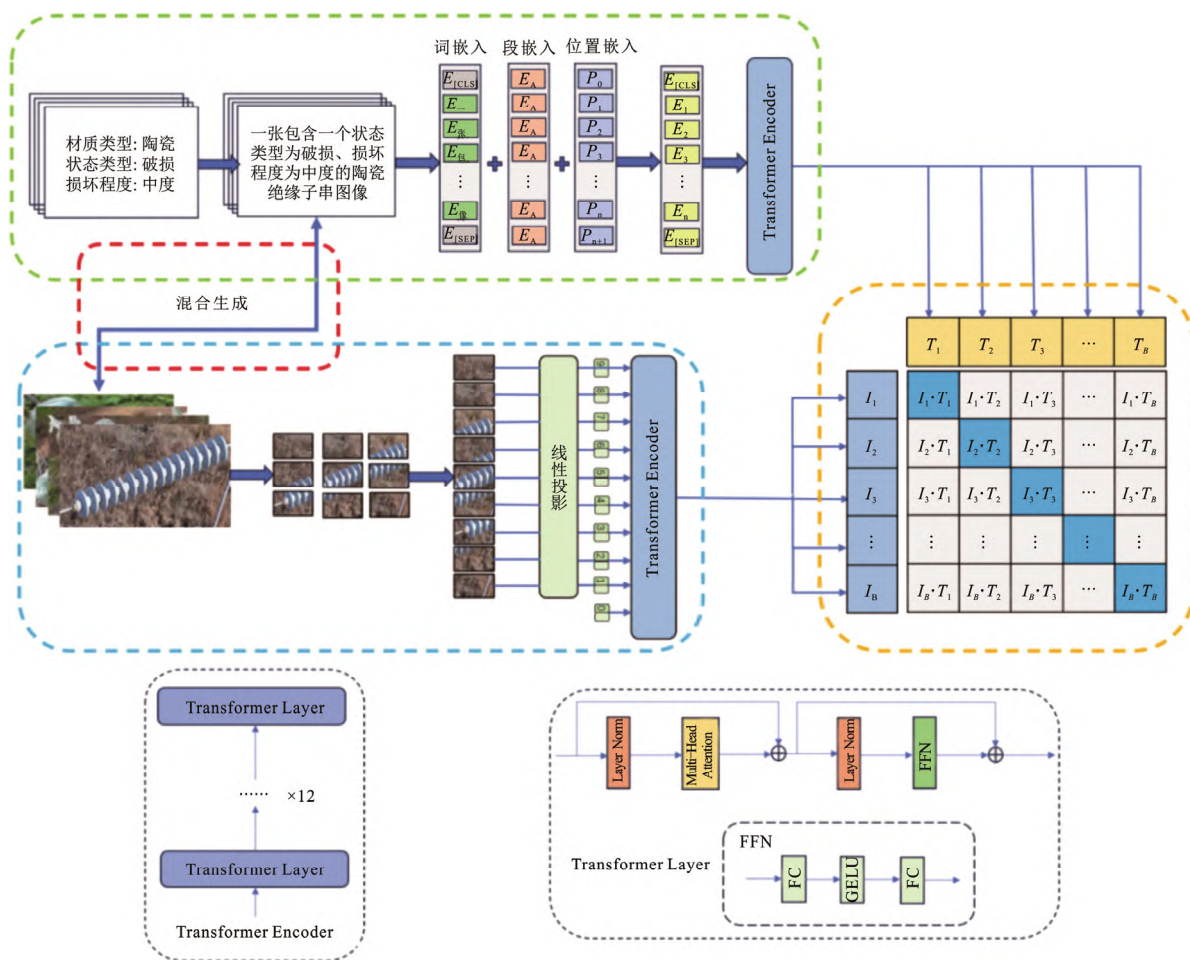


图1 模型总体结构

Fig.1 Overall structure of the model

段文本, 然后由 BERT 网络提取其特征信息。最后, 通过对比学习的方式来实现最大化正样本对的相似度、最小化负样本对的相似度。其中, 配对的图像和文本为正样本对(即对角线部分), 没有配对的图像和文本为负样本对(除对角线以外的部分)。混合生成是将两个不同的图像-文本对融合并生成新的图像-文本对。若每个 batch 中包含 B 个图像-文本对, 经过文本编码器和图像编码器处理则可以得到 B 个图像特征 I_1-I_B 和 B 个文本特征 T_1-T_B , 并产生 B 个正样本对和 B^2-B 个负样本对。

1.1 图像编码器

图像编码器采用 Vision Transformer 网络, Vision Transformer 网络结构可分为两个部分: Embedding 层、Transformer Encoder。

在 Embedding 层中, 输入的绝缘子图像为 $X \in \mathbf{R}^{H \times W \times C}$, 其中, H 、 W 、 C 分别表示输入图像的高、宽和通道数。首先将输入图像分为 N 个大小相同的图像块(x_1, x_2, \dots, x_N), 使用 Vision Transformer

网络的图像块编码器将每个图像块线性投影成维度为 D 的特征向量 $E(x_i)$, 然后, 在输入序列中加入可学习的分类向量 x_{cls} 用于表示绝缘子图像经过编码后的全局特征; 最后加入包含空间信息的位置编码 $P \in \mathbf{R}^{N \times D}$, 得到序列如下:

$$Z_0 = [E(x_{cls}); E(x_1); \dots; E(x_N)] + P \quad (1)$$

在 Transformer Encoder 层中, 输入为 Embedding 层的最终输出序列 Z_0 , 经过 12 个 Transformer Layer 模块处理后得到最终序列。

1.2 文本编码器

文本编码器采用的是 BERT 网络模型。首先将绝缘子图像对应的多标签转换成一段可描述的文本, 然后对这段文本进行预处理操作, 如分词、添加句子分类标记[CLS]和句子结尾标记[SEP]。接着将处理后的文本送入到 BERT 模型, 在词向量的基础上添加段向量和位置向量得到嵌入向量, 并将嵌入向量直接送入 Transformer Encoder 模块中处

理,得到标签文本最终对应的特征向量。其中,BERT模型中使用的 Transformer Encoder 模块与 Vision Transformer 模型使用的 Transformer Encoder 模块相同。

1.3 对比学习

图像编码器提取的图像特征经过处理得到 $I(I_1, I_2, I_3, \dots, I_N)$, 文本编码器提取的文本特征经过处理得到 $T(T_1, T_2, T_3, \dots, T_N)$ 。然后通过构造一个对称的 contrastive loss 计算图像特征和文本特征的相似度, 损失函数如下:

$$\ell_i^{(I \rightarrow T)} = -\ln \frac{\exp(\text{sim}(I_i, T_i) / \tau)}{\sum_{k=1}^N \exp(\text{sim}(I_i, T_k) / \tau)} \quad (2)$$

$$\ell_i^{(T \rightarrow I)} = -\ln \frac{\exp(\text{sim}(T_i, I_i) / \tau)}{\sum_{k=1}^N \exp(\text{sim}(T_i, I_k) / \tau)} \quad (3)$$

$$\ell_i = \ell_i^{(I \rightarrow T)} + \ell_i^{(T \rightarrow I)} \quad (4)$$

式(2)表示图像的损失; 式(3)表示文本的损失; 式(4)表示总损失; 其中式(2)、式(3)中的温度系数 τ 是可学习的, 并不是超参数; sim 表示余弦相似度。仅当 I 和 T 下标相同时, 表示图像和文本的特征配对, 否则表示不配对。使用对称式对比训练的优点在于在一轮循环中, 实现了两次损失计算, 明显提高了单轮次的训练效率, 有效提升了模型对不同故障类别的特征表达能力和聚类性能。

1.4 采用混合生成方法进行数据增强

为了增强模型的泛化能力, 除缩放、翻转等数据增强方法外, 本文使用了一种多模态联合数据增强^[17]方法: 混合生成。如图 2 所示, 混合生成方法通过对两个现有的图像-文本对中的图像进行线性插值, 同时连接来自这两个图像对应的文本, 从而生成新的训练样本。在新生成的样本中, 绝大多数的对象和场景布局都保留在混合图像中, 文本信息则完全保留, 所以新生成的图像-文本对内的语义关系是匹配的。

在每个 batch 的图像-文本对中, 图像和文本分别表示为 V 和 D , 给定两个图像-文本对 (V_i, D_i) 和 (V_j, D_j) , 其中 $i, j \in \{1, \dots, N\}, i \neq j$, 一个新的训练样本 (V_k, D_k) 可以通过以下方法得到:

$$V_k = \lambda \cdot V_i + (1 - \lambda) \cdot V_j \quad (5)$$

$$D_k = \text{concat}(D_i, D_j) \quad (6)$$

式中: λ 为介于 0 和 1 之间的超参数, 表示两个图像原始像素之间的线性插值。concat 运算符直接连

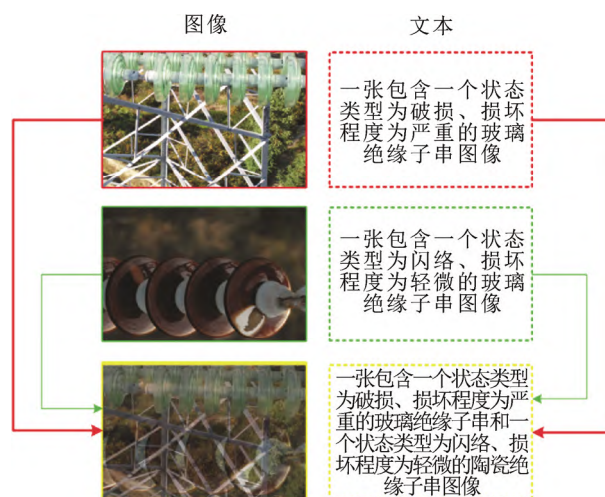


图 2 混合生成

Fig.2 Hybrid generation

接两个文本序列 D_i 和 D_j 以最好地保留原始信息。在包含 N 个图像-文本对的 batch 中, 混合生成会将前 M 个训练样本替换为新的生成样本。因此, batch 大小、总训练迭代次数保持不变。

2 数据集与实验设置

2.1 数据集

本研究使用的实验数据集包括 4 个部分: 中国电力线绝缘子数据集(CPLID)、绝缘子缺陷图像数据集(IDID)、自行采集的航拍图像, 以及基于这 3 个数据集利用合成技术制作的图像。这些数据集综合考虑了输电线路中绝缘子的实际应用场景, 并按照 3 大类标签进行分类: 材质类型、状态类型和损坏程度。其中, 材质类型分为玻璃和陶瓷; 状态类型包括正常、破损、闪络和污渍; 损坏程度则分为完好、轻微(如小范围污渍)、中度(如多处闪络)和严重(如掉串), 这些分类主要基于人工主观判断。整个数据集共包含 2 877 张绝缘子图像, 其具体类型及数量分布见表 1, 部分图像及其对应标签展示于图 3。在材质和状态类型方面, 样本数量相对均衡。对于损坏程度的标签, “完好” 和 “正常” 属于对照关系, 而 “轻微”、“中度” 和 “严重” 3 种标签受主观判断影响较大, 因此使用合成技术对其比例进行平衡。为增强模型的学习能力并确保其充分收敛, 对原始图像进行了缩放、裁剪等数据增强处理。实验中, 数据集按 7:3 比例随机分为训练集和测试集。

2.2 评价指标

任何用于单标签分类的度量都适用于多标签分类, 可采用宏观和微观两种计算方式计算所有标

签的这些度量^[18-19]。本文采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数 4 个指标来评价网络的分类性能。宏观计算方式是从对象标签的角度评价网络性能,即平均每类准确率(class accuracy, CA, 用 C_{CA} 表示)、平均每类精确率(class precision, CP, 用 C_{CP} 表示)、平均每类召回率(class recall, CR, 用 C_{CR} 表示)和平均每类 F1 分数(class F1, CF1, 用 C_{CF1} 表示)。微观计算方式是从总体样本的角度评价网络性能,即总体准确率(overall accuracy, OA, 用 O_{OA} 表示)、总体精确率(overall precision, OP, 用 O_{OP} 表示)、总体召回率(overall recall, OR, 用 O_{OR} 表示)和总体 F1 分数(overall F1, OF1, 用 O_{OF1} 表示)。 T_P 、 T_N 、 F_P 和 F_N 分别表示由单标签分类器评估的真阳性、真阴性、假阳性和假阴性的数量。

$$C_{CA} = \frac{1}{q} \sum_{i=1}^q \frac{T_{P_i} + T_{N_i}}{T_{P_i} + F_{N_i} + F_{P_i} + T_{N_i}} \quad (7)$$

$$C_{CP} = \frac{1}{q} \sum_{i=1}^q \frac{T_{P_i}}{T_{P_i} + F_{P_i}} \quad (8)$$

$$C_{CR} = \frac{1}{q} \sum_{i=1}^q \frac{T_{P_i}}{T_{P_i} + F_{N_i}} \quad (9)$$

$$C_{CF1} = \frac{1}{q} \sum_{i=1}^q \frac{2R_i P_i}{R_i + P_i} \quad (10)$$

式中: q 为标签个数; R_i 和 P_i 为第 i 个标签的召回率与精确度。

$$O_{OA} = \frac{\sum_{i=1}^q T_{P_i} + \sum_{i=1}^q T_{N_i}}{\sum_{i=1}^q T_{P_i} + \sum_{i=1}^q F_{N_i} + \sum_{i=1}^q F_{P_i} + \sum_{i=1}^q T_{N_i}} \quad (11)$$

$$O_{OP} = \frac{\sum_{i=1}^q T_{P_i}}{\sum_{i=1}^q T_{P_i} + \sum_{i=1}^q F_{P_i}} \quad (12)$$

$$O_{OR} = \frac{\sum_{i=1}^q T_{P_i}}{\sum_{i=1}^q T_{P_i} + \sum_{i=1}^q F_{N_i}} \quad (13)$$

$$O_{OF1} = \frac{2O_{OP}O_{OR}}{O_{OP} + O_{OR}} \quad (14)$$

2.3 参数设置

本文采用 Python3.8、Pytorch 框架进行编程,在 16 GB 内存、NVIDIA GeForce RTX 2080Ti、Intel

Xeon Gold 5120 CPU 处理器和 Window10 系统上进行网络的训练与测试,其他详细参数设置如表 2。

3 实验结果与分析

3.1 对比试验

表 3 展示了不同绝缘子标签及对应指标数据。观察表中数据可发现,“玻璃”和“陶瓷”标签数据较突出。这主要是由于在固定数量的图像样本中,这两种标签的样本量相对较多,因此识别准确度也相对更高。此外,由于“正常”和“完好”两个标签在图像标注过程中存在一一对应关系,它们的数

表 1 绝缘子类型及其数量

Table 1 Types and quantities of insulators	
绝缘子类型	数量/张
正常的玻璃绝缘子	360
有污渍的玻璃绝缘子	356
有破损的玻璃绝缘子	371
有闪络的玻璃绝缘子	348
正常的陶瓷绝缘子	357
有污渍的陶瓷绝缘子	353
有破损的陶瓷绝缘子	369
有闪络的陶瓷绝缘子	363

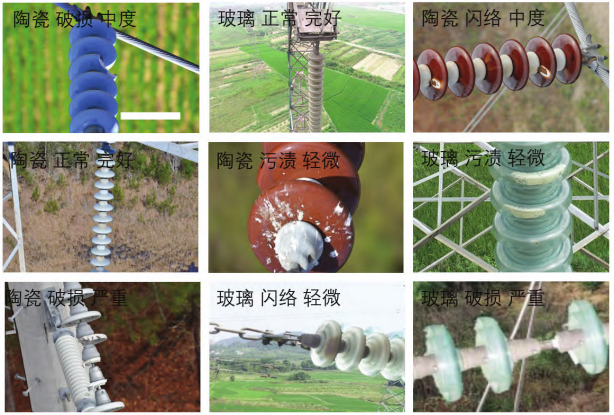


图 3 部分图像及其标签

Fig.3 Images and their labels

表 2 实验参数设置

Table 2 Experimental parameter settings	
参数	数值或名称
Train epochs	300
Batch size	8
Data enhancement	$\lambda=0.5$, $M=2$
momentum	0.9
optimizer	AdamW
Learning rate	10^{-4}
Weight decay	10^{-2}

据也显示出高度的相似性。与此同时,“闪络”和“污渍”标签的指标数据则明显低于其他类别。通过图像分析可发现,这两类标签所对应的图像区域在视觉上具有较高的相似度,且在整个图像中所占比例较小,这是影响其识别准确率的主要因素。

为验证本文方法的有效性,分别将本文方法与图像分类领域中的6个模型:VGG16、ResNet50、Mobile-V2^[20]、ITMMB^[21]、ViT-Base^[15]和CMIF^[22]在数据集上的分类性能进行对比,其中VGG16、ResNet50和Mobile-V2是基于CNN的模型,ViT-Base是基于Transformer的模型,ITMMB和CMIF是利用图像和文本两种特征信息进行分类的模型。各项分类指标如表4所示。

表4结果显示,本研究提出的方法在两种不同计算方式下均展现了卓越的分类性能。在宏观计算方式下,该方法能较好地反映模型在不同类别上的平均性能,其中分类准确率达到93.37%。而在微观计算方式下,该方法更加准确地展示了模型在整个数据集上的总体表现,准确率高达93.87%。相较于传统的单模态分类模型,本文提出的多模态方法结合了绝缘子图像和标签文本两种模态的特征。这种方法能够有效利用文本中包含的关于绝缘子图像的材质、状态等信息。通过将图像特征与文本特征进行相互关联,本方法不仅提高了模型的分类能力,还增强了分类结果的可解释性。例如:在对一个闪络的玻璃绝缘子图像进行分类时,模型会将图像的特征与候选标签文本的特征进行相似度比较。这样,图像最终会与包含“闪络”和“玻璃”等文本的标签相匹配,实现了既准确又具有解释性的分类。鉴于本团队更加重视模型的总体表现,后续实验的指标将主要采用微观计算方式。

3.2 混合生成中 λ 和 M 的取值

为了确保两张图片对于新生成图片具有相同的影响,将 λ 的默认值设为0.5。图4展示了在混合生成的数据增强方法中,当 λ 为0.5时,不同的 M 取值对分类性能的影响。由于批次大小为8,所以 M 的最大取值被限定为8。当 M 为0时,表示没有引入混合生成方法;当 M 取值为2时,表示每个批次中的前两张图片将被混合生成的图片所替代;当 M 取值为8时,意味着每个批次中的所有图片都采用混合生成方法。从图8可以看出,当 M 取值为2时,模型的准确率提升最显著;而当 M 超过4时,则会降低模型的准确率。因此,当 M 取值为2时,

混合生成方法的效果最佳。

如图5所示,展示了在混合生成数据增强方法

表3 各标签分类指标

Table 3 Classification metrics for each label

标签	Accuracy	Precision	Recall	F1
玻璃	96.28	95.79	94.42	95.10
陶瓷	95.38	94.50	94.92	94.71
正常	94.88	94.86	93.49	94.17
破损	92.79	93.81	94.14	93.97
闪络	89.67	91.62	89.67	90.63
污渍	90.61	90.38	88.73	89.55
完好	94.41	94.47	93.95	94.21
轻微	92.59	93.12	92.13	92.62
中度	93.72	92.27	91.93	92.10
严重	93.30	93.27	92.82	93.04

表4 不同模型分类指标的对比

Table 4 Comparison of different models

模型	C_{CA}	C_{CP}	C_{CR}	C_{CFI}	O_{OA}	O_{OP}	O_{OR}	O_{OFI}
VGG16	79.44	79.31	80.23	80.07	80.91	80.22	79.54	79.88
ResNet50	85.39	84.27	85.91	85.31	86.33	85.87	85.76	85.81
Mobile-V2	87.57	87.82	86.48	87.94	88.19	88.47	88.02	88.24
ITMMB	89.66	90.58	89.34	90.53	89.92	91.34	90.80	91.07
ViT-base	91.43	90.86	92.07	92.14	90.76	91.92	91.53	91.72
CMIF	92.69	92.14	91.88	92.74	92.94	93.21	92.42	92.81
本文模型	93.37	93.41	92.62	93.01	93.87	93.66	94.04	93.84

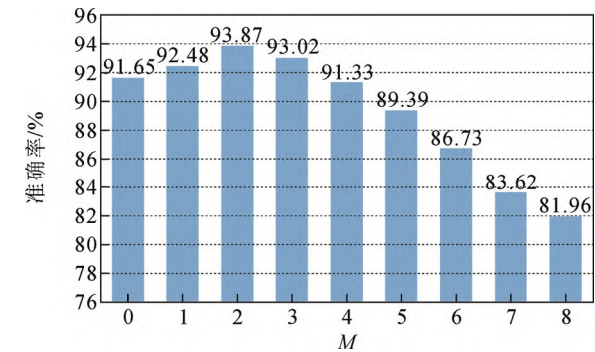


图4 M 与准确率的关系

Fig.4 Relationship between M and accuracy

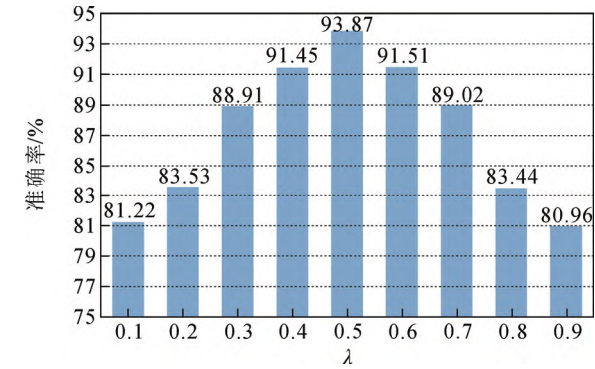


图5 λ 与准确率的关系

Fig.5 Relationship between λ and accuracy

中, 参数 M 设定为 2 时, 不同 λ 值对分类性能的具体影响。在该混合生成方法中, 新图像-文本对是通过将两个图像及其相应的标签文本进行叠加合成实现。实验结果表明, 当 λ 取值为 0.1 或 0.9 时, 由于文本内容与图像内容之间的偏差较大, 模型的准确率受影响, 呈现下降趋势。然而, 当 λ 值接近 0.5 时, 模型的准确率开始逐渐提高。特别是在 λ 值等于 0.5 时, 模型达到了最高的准确率, 此时两张原始图像对新生成图像的影响等同。

3.3 消融实验

如表 5 所示, M1 表示原始的多模态分类方法, M2 表示在多模态方法中引入了混合生成的数据增强方法。由表 5 可知, M1 的准确率为 91.65%, M2 的准确率为 93.87%, 表明通过生成并替换数据集中的部分图片, 能提高数据的使用效率, 进而提升模型的准确率。

3.4 可视化

在进行定性测试后, 使用可视化技术来更深入地理解图像中模型, 认为在将文本与图像关联时最重要的部分分别是图像整体的类激活图和两种标签对应的类激活图。

整体的类激活图如图 6 所示。整体的类激活图通过不同颜色的渐变来表示模型对绝缘子图像不同区域的分类重要程度。深色区域代表模型更加关注的区域, 而浅色区域则表示模型对该区域的关注较少。通过观察整体的类激活图, 有助于理解模型的决策过程, 并揭示模型在分类任务中的关注点和特

表 5 消融实验结果

Table 5 Results of ablation experiments				
模型	Accuracy	Precision	Recall	F1
M1	91.65	92.27	91.85	92.06
M2	93.87	93.66	94.04	93.84

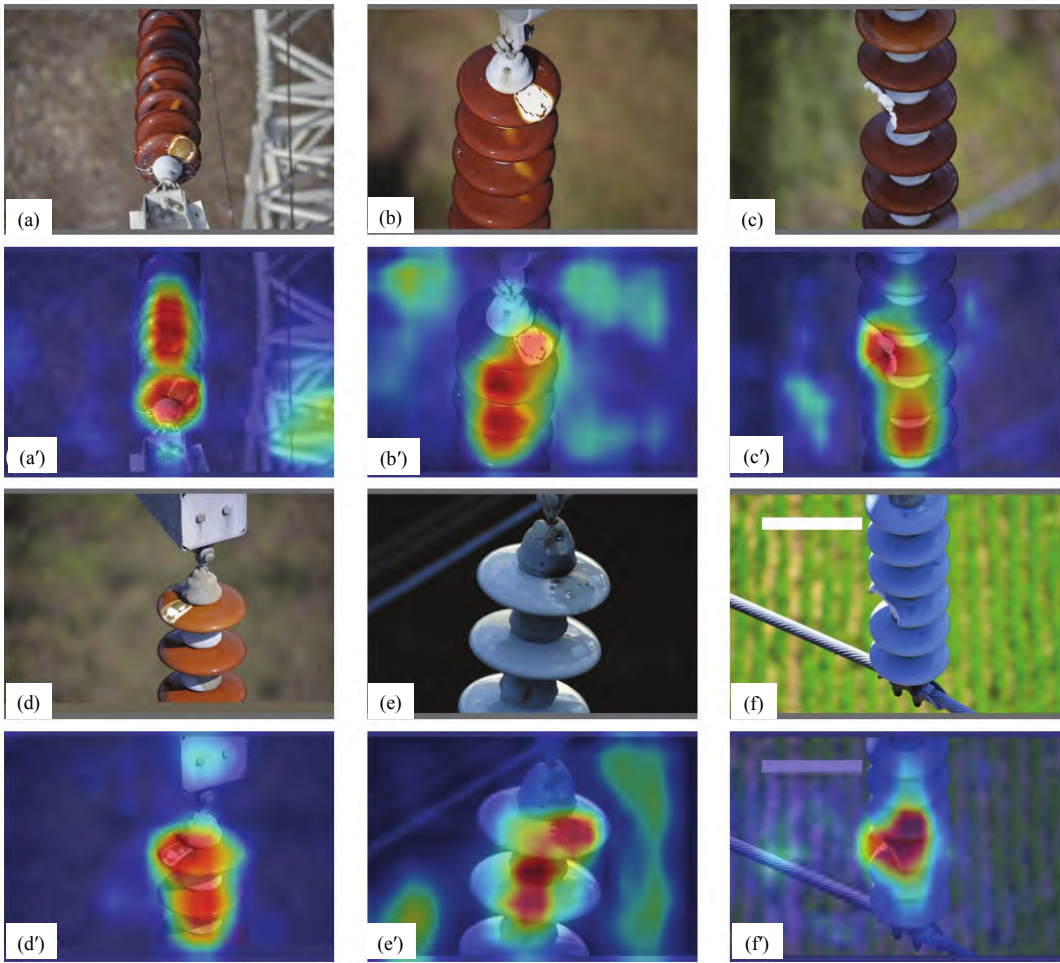


图 6 整体类激活图
Fig.6 Overall class activation map

征区域。图 6 中, 第 1 行和第 3 行的(a)、(b)、(c)、(d)、(e)和(f)表示待检测的原图, 第 2 行和第 4 行的(a')、(b')、(c')、(d')、(e')和(f')则表示模型对原图进行检测时所生成的对应热力图。可以看到, 深色区域更集中于绝缘子主体结构以及有缺陷部分的区域, 而这些区域正是目标的关键特征所在。这表明模型能够有效捕捉目标的关键信息, 具有较强的目标特征提取能力。

标签对应的类激活图如图 7 所示。对于每个标签, 生成一个独立的类激活图, 用以展示模型在该标签下的分类依据和关键区域。通过这些部分类激活图, 可详细观察模型在不同标签下对图像的关注点, 并进一步分析模型对于不同类型的识别依据和特征区域。图 7 中, (a)、(b)和(c)表示待检测的原图像, (a')、(b')和(c')是在材质标签条件下对原图进行检测得到的对应热力图, (a'')、(b'')和(c'')是状态类型标签条件下对原图进行检测得到的对应热力图。由图可知, 对于绝缘子的材质, 模型更加关注绝缘子整体, 而对于状态类型, 则更加关注绝缘子局部

内容。这表明了该模型具备较强的识别和聚焦特定目标的能力, 同时也可为相应的分类结果提供良好的可解释性。

3.5 测试结果

图 8 展示了部分图像测试结果。图像 a、b、d、e 的分类标签均正确无误, 但图像 c、f 的分类标签出现部分错误。具体来说, 图像 c 实际上的材质类型应为陶瓷, 损坏程度为轻微; 而图像 f 的正确状态类型应标为污渍。根据表 3 中的数据, 模型整体准确率高达 93.87%, 在分类错误的标签中, 材质类型的错误较少, 而在状态类型的分类中, “闪络”和“污渍”标签的错误较多。通过分析图像测试结果可看出, 在绝缘子图像中, 材质特征在视觉上占据重要地位, 且各种材料之间具有显著的外观差异。因此, 即使某些图像被错误分类, 通常是因为这些图像的材质特征与被误分配的标签所代表的材质外观之间存在较高的相似度。然而, 对于状态类型的标签, 特别是闪络和污渍, 由于它们在图像中所占的比例较小且特征差异不明显, 模型在识别这些状

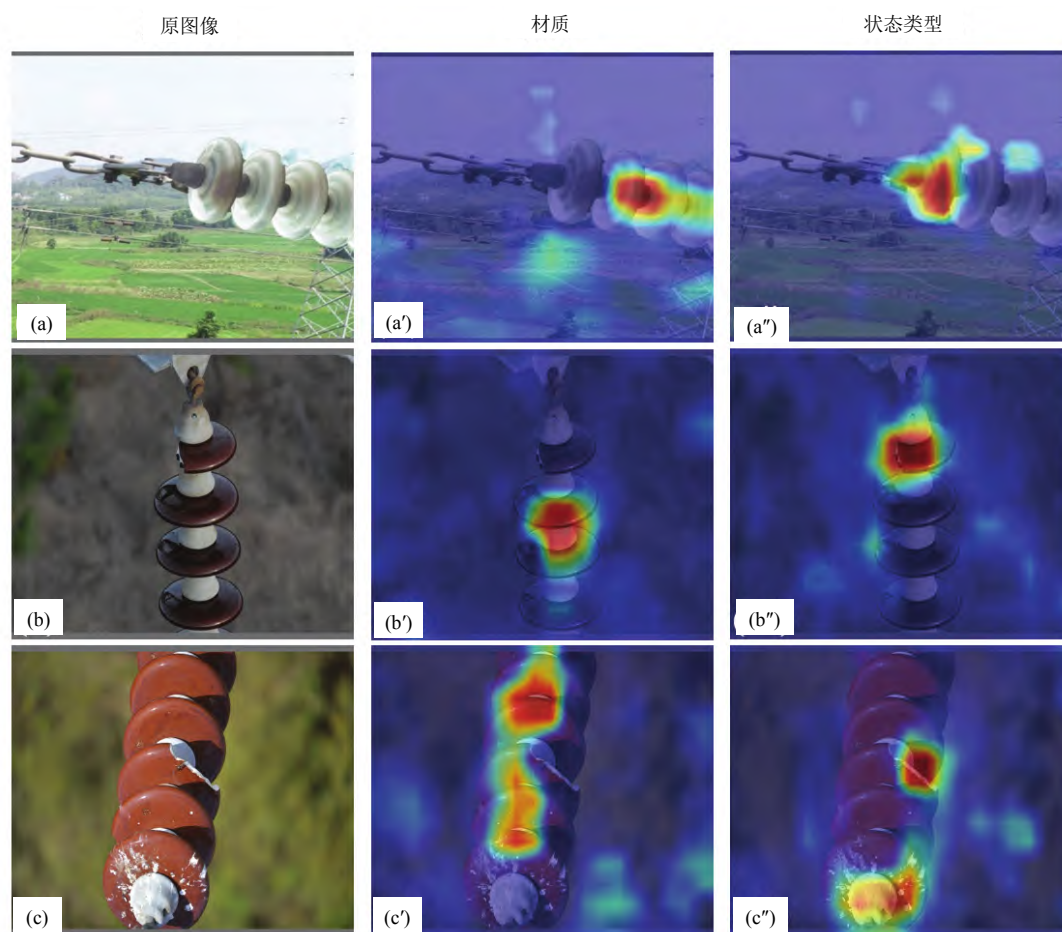


图 7 标签对应的类激活图

Fig.7 Class activation maps for labels



图8 部分测试图像

Fig.8 Some test pictures

态时的误判率相对较高。总体而言,该模型在测试绝缘子图像过程中展现了出色的性能。

4 结论

1) 本文首次提出了一种基于多模态的缺陷绝缘子图像的多标签分类方法。这种方法解决了传统深度学习分类方法在文本信息利用上的不足,并应对了绝缘子图像分类标签的单一性问题。

2) 本文提出的方法采用多模态联合数据增强方法,实现了跨模态数据增强。通过对比学习方式,使图像和文本的特征信息产生了有效关联,不仅增强了网络分类的可靠性,还提供了良好的可解释性。实验结果表明,该方法在数据集上的分类总体准确率为93.87%。相比其他分类方法,本文提出的方法在分类性能上具有明显优势,为多模态技术在电网领域的应用提供了较好的基础。

3) 多模态技术需要提取不同模态的特征,这导致其所需模型体量较大。未来的研究方向将集中在模型的轻量化上,以保证检测精度的同时提高模型的检测速度。此外,还可以进一步完善实验数据集,从而增强模型的泛化能力和鲁棒性。

参考文献 References

- [1] 张嘉伟, 叶子帆, 王 倩, 等. 基于 F-P 光纤泄漏电流传感器的绝缘子状态监测[J]. 高电压技术, 2022, 48(8): 2915-2923.
ZHANG Jiawei, YE Zifan, WANG Qian, et al. Insulator condition monitoring based on F-P optical fiber leakage current sensor[J]. High Voltage Engineering, 2022, 48(8): 2915-2923.
- [2] 郝 帅, 张 旭, 马 旭, 等. 基于 PKAMNet 的输电线路小目标故障检测方法[J]. 高电压技术, 2023, 49(8): 3385-3394.
HAO Shuai, ZHANG Xu, MA Xu, et al. Small target fault detection

method for transmission lines based on PKAMNet[J]. High Voltage Engineering, 2023, 49(8): 3385-3394.

- [3] 王 波, 张迎晨, 齐冬莲, 等. 数字化安全管控视角下的全息影像: 定义、基本框架及关键技术[J]. 高电压技术, 2023, 49(8): 3335-3345.
WANG Bo, ZHANG Yingchen, QI Donglian, et al. Holographic images for digital security control: definition, basic framework and key technologies[J]. High Voltage Engineering, 2023, 49(8): 3335-3345.
- [4] 方 挺, 董 冲, 胡兴柳, 等. 航拍图像中绝缘子串的轮廓提取和故障检测[J]. 上海交通大学学报, 2013, 47(12): 1818-1822.
FANG Ting, DONG Chong, HU Xingliu, et al. Contour extraction and fault detection of insulator strings in aerial images[J]. Journal of Shanghai Jiaotong University, 2013, 47(12): 1818-1822.
- [5] 林聚财, 韩 军, 陈昉明, 等. 基于彩色图像的玻璃绝缘子缺陷诊断[J]. 电网技术, 2011, 35(1): 127-133.
LIN Jucai, HAN Jun, CHEN Fangming, et al. Defects detection of glass insulator based on color image[J]. Power System Technology, 2011, 35(1): 127-133.
- [6] 黄新波, 刘新慧, 张 烨, 等. 基于红蓝色差和改进 K-means 算法的航拍绝缘子分类识别方法[J]. 高电压技术, 2018, 44(5): 1528-1534.
HUANG Xinbo, LIU Xinhui, ZHANG Ye, et al. Classification recognition method of insulator in aerial image based on the red-blue difference and developed K-means algorithm[J]. High Voltage Engineering, 2018, 44(5): 1528-1534.
- [7] 裴少通, 刘云鹏, 陈同凡, 等. 基于 BOA-SVM 的劣化绝缘子红外图谱诊断方法[J]. 电测与仪表, 2018, 55(24): 11-16.
PEI Shaotong, LIU Yunpeng, CHEN Tongfan, et al. Infrared spectrum diagnosis method of deteriorated insulators based on BOA-SVM[J]. Electrical Measurement & Instrumentation, 2018, 55(24): 11-16.
- [8] 马立新, 豆晨飞, 宋晨灿, 等. 基于 VGGNet 算法的绝缘子无损检测[J]. 光电工程, 2021, 48(1): 200072.
MA Lixin, DOU Chenfei, SONG Chencan, et al. Insulator nondestructive testing based on VGGNet algorithm[J]. Opto-Electronic Engineering, 2021, 48(1): 200072.
- [9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]// 3rd International Conference on Learning Representations(ICLR 2015). San Diego, USA: s. n., 2015: 1-14.
- [10] 郝 帅, 马瑞泽, 赵新生, 等. 基于超分辨深度残差网络的玻璃绝缘子自爆故障检测算法[J]. 高电压技术, 2022, 48(5): 1817-1825.
HAO Shuai, MA Ruize, ZHAO Xinsheng, et al. Self-explosion fault detection algorithm for glass insulator based on super-resolution deep residual network[J]. High Voltage Engineering, 2022, 48(5): 1817-1825.
- [11] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770-778.
- [12] 亢 洁, 王 勃, 刘文波, 等. 融合 CAT-BiFPN 与注意力机制的航拍绝缘子多缺陷检测网络[J]. 高电压技术, 2023, 49(8): 3361-3372.
KANG Jie, WANG Qing, LIU Wenbo, et al. Detection model of multi-defect of aerial photo insulator by integrating CAT-BiFPN and attention mechanism[J]. High Voltage Engineering, 2023, 49(8): 3361-3372.
- [13] 徐建军, 黄立达, 闫丽梅, 等. 基于层次多任务深度学习的绝缘子自爆缺陷检测[J]. 电工技术学报, 2021, 36(7): 1407-1415.
XU Jianjun, HUANG Lida, YAN Limei, et al. Insulator self-explosion defect detection based on hierarchical multi-task deep learning[J].

- Transactions of China Electrotechnical Society, 2021, 36(7): 1407-1415.
- [14] 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. [S.l.]: PMLR, 2021: 8748-8763.
- [15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C]// Proceedings of the 9th International Conference on Learning Representations. Seoul, Korea: IEEE Press, 2021: 10-20.
- [16] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, USA: Association for Computational Linguistics, 2018: 4171-4186.
- [17] HAO X S, ZHU Y, APPALARAJU S, et al. MixGen: a new multi-modal data augmentation[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops. Waikoloa, USA: IEEE, 2023: 379-389.
- [18] 武红鑫, 韩 萌, 陈志强, 等. 监督和半监督学习下的多标签分类综述[J]. 计算机科学, 2022, 49(8): 12-25.
WU Hongxin, HAN Meng, CHEN Zhiqiang, et al. Survey of multi-label classification based on supervised and semi-supervised learning[J]. Computer Science, 2022, 49(8): 12-25.
- [19] 朱旭东, 熊 贲. 基于多层次注意力与图模型的图像多标签分类算法[J]. 计算机工程, 2022, 48(4): 173-178, 190.
ZHU Xudong, XIONG Yun. Multi-label image classification algorithm based on multi-scale attention and graph model[J]. Computer Engineering, 2022, 48(4): 173-178, 190.
- [20] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 4510-4520.
- [21] 李佳欣, 苏曙光. 基于 BERT 的图像和文本多模态融合分类模型[J]. 计算机应用, 2023, 43(增刊 1): 39-44.
LI Jiaxin, SU Shuguang. Image and text multi-modality fusion classification model based on BERT[J]. Journal of Computer Applications, 2023, 43(Supplement 1): 39-44.
- [22] 李象贤, 郑裕泽, 马浩凯, 等. 基于跨模态特权信息增强的图像分类方法[J/OL]. 软件学报, 2024: 1-17[2024-11-19]. <https://doi.org/10.13328/j.cnki.jos.007052>.
LI Xiangxian, ZHENG Yuze, MA Haokai, et al. Image classification method based on cross-modal privileged information enhancement[J/OL]. Journal of Software, 2024: 1-17[2024-11-19]. <https://doi.org/10.13328/j.cnki.jos.007052>.



ZHOU Jing

Ph.D.

Associate professor

Corresponding author



WANG Manyi



TIAN Zhaoxing

周 景(通信作者)

1978—, 男, 博士, 副教授, 硕导

从事人工智能、机器学习和电力大数据分析方面的研究工作

E-mail: zhoujing108@ncepu.edu.cn

王满意

2000—, 男, 硕士生

从事计算机视觉与电力故障检测方面研究工作

E-mail: wmy@ncepu.edu.cn

田兆星

1998—, 男, 硕士生

从事计算机视觉与电力巡检方面的研究工作

E-mail: 120222227370@ncepu.edu.cn

收稿日期 2023-12-18 修回日期 2024-12-14 编辑 卫李静