

DOI: 10.19650/j.cnki.cjsi.J2514407

基于跨尺度特征融合的光伏板故障诊断方法研究*

何明月¹, 范思远¹, 曹生现¹, 张艳辉²

(1. 东北电力大学自动化工程学院 吉林 132012; 2. 中国科学院深圳先进技术研究院 深圳 518055)

摘要:光伏发电是我国能源转型与“双碳”目标实现的核心支撑,其系统的高效安全运行至关重要。然而,光伏板故障会直接导致发电量骤降、安全风险攀升及运维成本增加,已成为制约光伏产业高质量发展的关键瓶颈。为此,提出了基于跨尺度特征融合的光伏板故障诊断方法,该方法借助光伏场站已有的安防设备,在不额外增加硬件成本的前提下,可实现光伏板正常、积灰附着、鸟粪附着、电气损失、物理损伤、积雪覆盖这 6 类运行状态的精准检测。具体而言,所提方法首先构建了 Transformer 与卷积神经网络(CNN)的跨尺度特征融合框架,其中 Transformer 利用自注意力机制捕获光伏板图像的全局上下文特征,为 CNN 提取局部细节特征提供全局引导;其次,在卷积分支中设计密集连接机制,通过特征图的跨层连接,提升不同层级特征的传播与复用能力;同时,针对性开发了涵盖不同光照条件、天气状况及故障类型的光伏板 6 类运行状态可见光数据集。与其他模型相比,所提方法综合性能更优,Top-1 准确率达 93.33%,Top-3 准确率达 100%,且模型参数规模相对轻量,对硬件算力要求较低。此外,为全面评估方法的工程应用价值,构建了“模型精度-使用效率-应用规模”三维综合评价指标,进一步验证了所提方法在光伏现场实际部署的可行性。

关键词: 光伏板;故障诊断;跨尺度;特征融合;卷积神经网络;Transformer

中图分类号: TH7 TM914.4 **文献标识码:** A **国家标准学科分类代码:** 480.6070

Research on fault diagnosis method for photovoltaic panels based on cross-scale feature fusion

He Mingyue¹, Fan Siyuan¹, Cao Shengxian¹, Zhang Yanhui²

(1. School of Automation Engineering, Northeast Electric Power University, Jilin 132012, China;

2. Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract: Photovoltaic (PV) power generation serves as a core pillar for China's energy transition and the achievement of the dual carbon goals, and the efficient and safe operation of its systems is of vital importance. However, PV panel faults can directly lead to a sharp drop in power generation, a rise in safety risks, and an increase in operation and maintenance costs, which have become a key bottleneck restricting the high-quality development of the PV industry. To address this issue, a PV panel fault diagnosis method based on cross-scale feature fusion is proposed. By leveraging the existing security equipment in PV power stations, this method can accurately detect six operational states of PV panels, including normal, dust accumulation, bird droppings adhesion, electrical loss, physical damage, and snow coverage, without incurring additional hardware costs. Specifically, this method firstly constructs a cross-scale feature fusion framework combining the Transformer and the convolutional neural network (CNN). The Transformer captures the global contextual features of PV panel images through its self-attention mechanism, providing global guidance for the CNN to extract local detailed features. Secondly, a dense connection mechanism is designed in the CNN branch, which enhances the propagation and reuse capabilities of features at different levels through cross-layer connections of feature maps. Meanwhile, a visible light dataset for six operating states of PV panels is developed in a targeted manner, covering different lighting conditions, weather situations, and fault types. Compared with other models, this method exhibits superior comprehensive performance, with a Top-1 accuracy of 93.33% and a Top-3 accuracy of 100%. Additionally, the model has a relatively lightweight parameter scale and low hardware computing power

收稿日期:2025-08-22 Received Date: 2025-08-22

* 基金项目:国家自然科学基金(62303105)、吉林省科技发展计划(20240304193SF)项目资助

requirements. Furthermore, to comprehensively evaluate the engineering application value of the method, a three-dimensional comprehensive evaluation index system, model accuracy-usage efficiency-application scale, is established, which further verifies the feasibility of the method for practical deployment in PV fields.

Keywords: photovoltaic panel; fault diagnosis; cross-scale; feature fusion; convolutional neural network; Transformer

0 引 言

光伏发电作为我国能源结构转型与实现双碳目标的重要支撑,其系统高效安全运行的重要性愈发凸显^[1]。光伏板作为能量转换的核心部分,其运行状态直接影响了整个系统的发电效率^[2]。然而,在长期运行过程中,光伏板会受到自然环境与外部因素等多重影响,容易发生各类故障,通常可分为表面污染类和功能性损伤类,如:鸟粪附着、灰尘覆盖、积雪堆积、电气性能退化、物理组件损坏等。这些故障将导致系统发电效率下降 10% ~ 50%,甚至更高^[3-4]。受光伏场站运维成本限制,新增专用传感设备往往存在经济性不足的问题。因此,充分利用现有安防监控设备,结合人工智能技术,实现对光伏板故障的及时、精准诊断,对于保障系统稳定运行、提升发电效率具有重要意义。

近年来,基于深度学习的光伏板故障诊断方法已成为研究热点^[5]。其中,卷积神经网络(convolutional neural network, CNN)凭借其强大的局部特征提取能力,在光伏板故障诊断领域取得了良好的结果^[6]。研究前期多采用 LeNet^[7]、AlexNet^[8]等网络结构,通过对光伏板图像的特征提取与分类实现故障诊断;随着人工智能的发展, VGG^[9]、ResNet^[10]等模型被相继提出,这些模型通过增加网络深度、引入残差连接等方法,进一步提升了特征提取和模型泛化能力。针对基于光伏板图像缺陷诊断问题,冯秀坤等^[11]提出了一种基于 YOLOX 算法的光伏发电板异物监测系统,解决了西北荒漠地区高速公路服务区光伏发电板异物遮挡问题。吕游等^[12]基于 EfficientNet 网络构建故障识别模型,在模型中引入双通道注意力模块,提升了模型对空间特征信息的提取能力,增加模型的识别准确率。上述模型本质上依赖局部感受野进行的特征建模,但却难以有效捕捉图像中的全局上下文信息与长距离依赖关系,在处理复杂背景干扰及多尺度缺陷时,存在特征表达能力不足的问题。

Transformer 模型借助自注意力机制可有效捕捉全局依赖关系,在处理长距离依赖问题上具有优势,研究人员尝试将其应用于光伏板故障诊断任务^[13]。Ramadan 等^[14]基于 Transformer 开发了一种用于光伏板红外热像图自动故障检测的深度学习模型,旨在提高光伏故障检测和分类的准确性。Khan 等^[15]基于 Transformer 开发了一种多标签分类模型,用于识别光伏系统中重叠及并发

故障,有效提升了复杂故障场景下的检测准确率。可以看出,对于电气损失这类可能涉及多区域关联的故障而言,仅依靠局部特征往往难以提供有效的特征信息。相比之下,Transformer 模型在这类需要全局信息的场景中已展现出一定优势。然而,单一的 Transformer 模型在提取局部细节特征方面仍存在不足,特别是在处理光伏板表面少量鸟粪、积灰等小尺度信息时,其对局部特征的捕捉能力显得较为薄弱。

为克服单一模型的局限性,研究人员开始探索将 CNN 与 Transformer 融合的方法,将 CNN 作为特征提取器,将其提取的特征输入 Transformer 进行全局特征学习与分类。Guo 等^[16]针对遥感图像中光伏组件结构多样性带来的语义分割难题,基于 U-Net 与视觉 Transformer (vision Transformer, ViT) 构建了 TransPV 模型,有效提升了光伏系统识别精度与结构适应能力。Mahboob 等^[17]基于 SegFormer 构建光伏组件制造缺陷检测框架,实现了裂纹、栅线断裂等 5 类缺陷的高精度语义分割与可视化标注。Wang 等^[18]提出光伏缺陷检测模型 PDFormer,通过引入基于注意力机制的特征重加权模块结合多层改进策略,有效提升了模型在小样本数据下的检测精度。然而,现有特征融合方法在机制设计上仍有优化空间,尚未充分挖掘卷积模型与 Transformer 的协同效能。部分研究仅采用简单拼接策略,未能通过全局语义信息对局部特征提取过程形成有效引导与约束,导致诊断性能提升有限。

针对上述问题,提出了一种基于跨尺度特征融合的光伏板故障诊断方法。该方法依托光伏场站现有的安防设备,在不增加额外硬件成本的前提下,实现了对正常、积灰附着、鸟粪附着、电气损失、物理损伤和积雪覆盖 6 类光伏板状态的精准检测。方法通过卷积分支提取局部特征,Transformer 分支捕捉全局信息,并利用全局信息辅助卷积网络进行特征提取,从而提升故障诊断的准确性。同时,卷积分支中引入密集连接网络,增强特征的传播与复用能力,进一步提升局部特征提取效果。此外,开发 6 类可见光光伏板故障数据集,并构建跨尺度特征融合的故障诊断模型,为光伏板故障诊断的深入研究提供基础数据支撑。

1 光伏故障诊断方法

1.1 整体框架

提出的基于 Transformer 与卷积双分支融合的光伏板

故障诊断方法整体框架,如图1所示。该框架采用双分支并行结构,由Transformer通道(上分支,负责全局上下文建模)与卷积通道(下分支,负责局部细粒度特征提取)组成,

通过跨尺度特征融合机制实现不同层级特征的互补。最终融合特征经全局平均池化与多层感知机(multilayer perceptron, MLP)分类层输出6类光伏板状态的预测结果。

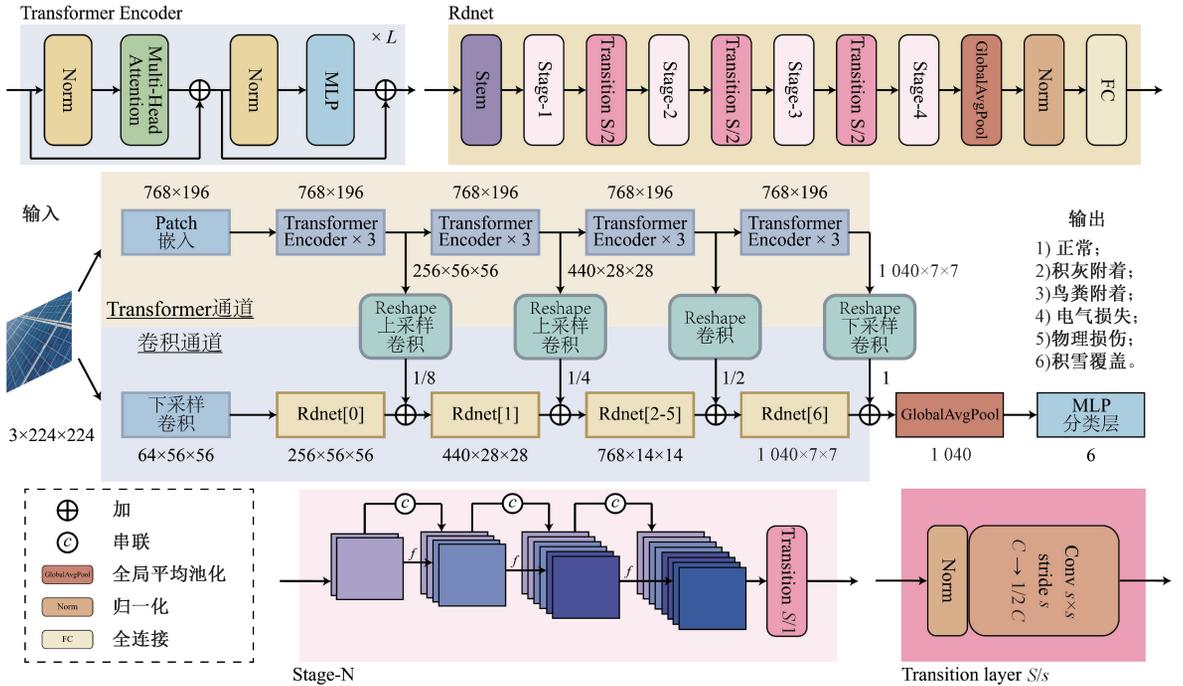


图1 Transformer与卷积神经网络跨尺度特征融合框架

Fig. 1 A cross-scale feature fusion framework of Transformer and convolutional neural networks

具体而言,Transformer通道首先将输入图像划分为 $16 \text{ pixels} \times 16 \text{ pixels}$ 的Patch并嵌入至768维空间,随后通过3组Transformer Encoder堆叠进行全局特征编码;卷积通道则以 7×7 的卷积核进行下采样卷积,将输入图像压缩至 $64 \times 56 \times 56$ 维度后,接入基于密集连接的RdnNet模块逐步提升特征表达能力。跨尺度融合阶段,Transformer通道的特征经上采样卷积匹配卷积通道的空间维度,通过残差连接实现全局信息对局部特征的动态引导;最终融合特征通过全局平均池化压缩为1040维向量,输入多层感知机(MLP)分类层完成6类光伏板运行状态的预测输出。

1.2 Transformer通道

光伏板的电气损伤、积灰、积雪等大面积覆盖故障,通常会表现为长距离空间关联特征,Transformer的自注意力机制能够有效捕捉此类全局依赖关系。为此,设计了基于Transformer的通道结构以匹配此类故障特性。

1) Patch嵌入

输入的光伏板可见光图像 $I \in \mathbb{R}^{3 \times 224 \times 224}$,被划分为不重叠的Patch,尺寸 P 为16。那么,经分块操作后能够得到 $N = (224/P) \times (224/P) = 196$ 个Patch,每个Patch通过展平与线性变换嵌入至 $D = 768$ 维高维空间,得到嵌入

特征 $X_p \in \mathbb{R}^{N \times D}$ 。此操作将光伏板的局部纹理编码为高维向量,同时保留Patch间的空间位置信息。

2) Transformer Encoder堆叠

随后,嵌入特征 $X_p \in \mathbb{R}^{N \times D}$ 将进入Transformer Encoder模块。该模块内部由层归一化、多头自注意力(multi-head attention, MHA)与多层感知机组成,采用残差方式连接的方式,缓解梯度消失的问题。多头注意力通过并行计算多组注意力头,捕捉光伏板图像中远距离空间依赖关系,其数学表达为:

$$\text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \cdot \mathbf{W}^O \quad (1)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q} \cdot \mathbf{W}_i^Q, \mathbf{K} \cdot \mathbf{W}_i^K, \mathbf{V} \cdot \mathbf{W}_i^V) \quad (2)$$

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V} \quad (3)$$

其中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别表示查询、键、值矩阵; $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V$ 表示投影矩阵; d_k 为键的维度; head_i 为头数; \mathbf{W}^O 为输出投影矩阵。

1.3 卷积通道

针对鸟粪、细微物理损伤等光伏板局部故障,卷积神经网络的局部感受野与层级化特征提取机制,适用于此类细粒度特征提取等需求。为此,将用卷积通道提取光伏图像的局部细节信息。

1) 下采样卷积

输入的光伏板图像 $I \in \mathbb{R}^{3 \times 224 \times 224}$, 首先经过下采样卷积层, 将空间维度从 224×224 压缩至 56×56 , 卷积核尺寸为 7×7 , 步幅为 4, 填充为 3, 同时将通道数从 3 扩展至 64, 该操作模拟人类视觉系统对光伏板表面进行的局部观察, 优先捕捉小故障的边缘与纹理特征, 其过程能够表示为:

$$\mathbf{X}_{cd} = \text{Conv}(I; \mathbf{W}_d, b_d) \quad (4)$$

其中, $\mathbf{W}_d \in \mathbb{R}^{64 \times 3 \times 7 \times 7}$ 为卷积核权重, $b_d \in \mathbb{R}^{64}$ 为偏置。

2) RdnNet 模块

RdnNet 基于密集连接^[19]设计, 通过特征的跨层复用, 提升光伏板多尺度局部特征的提取能力, 其内部由多个 Stage 组成, 每个 Stage 则由 Dense Block 与 Transition 层组成。Dense Block 的第 k 层的输入为前面所有层输出的通道维度拼接, 输出为:

$$\mathbf{x}_k = H_k([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{k-1}]) \quad (5)$$

其中, H_k 表示卷积操作, 具体为批量归一化、ReLU 激活和卷积; $[\cdot]$ 为通道维度拼接。

Transition 层则用于空间下采样与通道压缩, 平衡特征表达能力与计算资源。内部由 1×1 卷积与 2×2 平均池化构成, 能够表示为:

$$\mathbf{X}_{\text{Trans}} = \text{AvgPool}(\text{Conv}(\mathbf{X}_{\text{dense}}; \mathbf{W}_{1 \times 1}, b_{1 \times 1})) \quad (6)$$

此操作在降低特征图分辨率的同时, 保留关键局部特征, 为后续跨尺度融合提供多尺度局部信息。

1.4 跨尺度特征融合

光伏板故障具有多尺度特性, 如小面积鸟粪、细微物理损伤等局部特征, 和积灰、积雪覆盖、电气损失等全局特征。单分支难以有效应对多尺度特征的场景, 因此, 将设计跨尺度特征融合机制, 通过 Transformer 的全局信息引导卷积通道的局部特征提取, 实现多尺度故障特征的互补表达。

1) 序列到空间特征维度匹配

Transformer 通道的特征需与卷积通道对应尺度的特征融合, 通过上/下采样卷积匹配空间维度。以第 1 个融合节点为例, Transformer Encoder 输出 $\mathbf{X}_{tr} \in \mathbb{R}^{196 \times 768}$ 经 Reshape 转换为空间特征 $\mathbf{X}_{tr_s} \in \mathbb{R}^{768 \times 14 \times 14}$, 通过上采样卷积将空间维度扩展至 56×56 , 通道数调整为 256。

2) 全局信息引导局部特征

上采样后的 Transformer 特征 $\mathbf{X}_{tr_{up}} \in \mathbb{R}^{256 \times 56 \times 56}$ 与卷积通道 RdnNet[0] 的输出 $\mathbf{X}_{rdn} \in \mathbb{R}^{256 \times 56 \times 56}$ 进行元素相加, 具体为:

$$\mathbf{X}_{\text{fused}} = \mathbf{X}_{rdn} + \mathbf{X}_{tr_{up}} \quad (7)$$

该操作使全局信息动态调制局部特征提取过程, 当卷积通道聚焦于局部特征边缘时, Transformer 所捕获的全局信息能够有效区分那些局部特征相似但全局分布模

式存在显著差异的区域; 反之, 由 Transformer 提取的跨区域故障关联信息, 则可引导卷积通道强化局部区域的特征提取, 从而降低误判率。

3) 多尺度特征融合分类

将多尺度特征聚合为低维向量, 消除空间位置对分类的干扰, 同时保留故障的判别性模式, 融合后的特征 $\mathbf{X}_{\text{final}} \in \mathbb{R}^{1040 \times 7 \times 7}$ 经全局平均池化压缩空间维度, 得到特征向量, 即:

$$\mathbf{X}_{\text{gap}} = \frac{1}{7 \times 7} \sum_{i=1}^7 \sum_{j=1}^7 \mathbf{X}_{\text{final}}[i, j], \quad \mathbf{X}_{\text{gap}} \in \mathbb{R}^{1040} \quad (8)$$

随后, 输入多层感知机分类层, 输出 6 类光伏板状态的预测分数。

2 实 验

为验证所提跨尺度特征融合方法在光伏板故障诊断中的有效性, 构建一套光伏板多类型故障的可见光图像数据集, 涵盖实际光伏场站的典型故障场景, 确保对光伏现场复杂环境的适配性。

2.1 数据集

数据集的样本主要来源于我国东北、西北光伏场站的安防监控图像, 以及互联网公开资源。所有样本均为自然场景下的无干预采集, 完整保留光伏场站的复杂环境特征, 不同的光照条件(晴天正午、阴天、日出/日落)、气象条件及组件类型(单晶硅、多晶硅)。目前, 该数据集已开源至网站 (<https://github.com/fansiyuan520/PV-images>)。数据集中各类别样本数量及占比, 如表 1 所示。

表 1 数据集样本数量及占比

Table 1 The number and proportion of samples in the dataset

类别	训练集	验证集	总计	占比/%
正常(清洁)	183	10	193	21.8
鸟粪附着	197	10	207	23.4
积灰覆盖	180	10	190	21.5
电气损失	93	10	103	11.6
物理损伤	59	10	69	7.8
积雪覆盖	113	10	123	13.9
总计	825	60	885	100.0

2.2 数据增强

光伏板故障诊断的实际应用时, 同一故障会呈现多样化视觉特征, 为提升模型对环境差异的鲁棒性, 提出一种训练集在线数据增强策略, 仅在模型训练阶段动态应

用,不改变原始数据集数据。

1) 随机几何变换

(1) 镜像操作:模拟光伏板因安装朝向不同导致的图像差异,主要是针对积灰、积雪等无方向性的故障,其表达式为:

$$I'(x, y) = I(x, H - y) \quad (9)$$

其中, I 为原始图像, I' 为翻转后图像, (x, y) 为像素坐标, H 为图像高度。

(2) 旋转操作:模拟安防摄像头产生的角度偏移,避免模型对固定角度的过拟合。设旋转角度为 $\theta \in [-10^\circ, 10^\circ]$, 旋转后像素坐标通过映射,如式(10)所示。

$$\begin{cases} x' = x \cos \theta - y \sin \theta \\ y' = x \sin \theta + y \cos \theta \end{cases} \quad (10)$$

其中, (x', y') 为旋转后像素坐标,采用双线性插值进行像素值重计算。

(3) 随机裁剪:从 224×224 图像中随机裁剪 200×200 区域后变换回原尺寸,提升对小尺度故障的局部特征提取能力。

2) 随机光照/噪声调整

(1) 亮度调整:模拟晴天与阴天的光照强度差异,面向积雪覆盖场景,调整范围是原始亮度的 0.8~1.2 倍;

(2) 对比度调整:模拟采集图像的灰度差异,调整范围同样是原始的 0.8~1.2 倍;

(3) 噪声叠加:用于模拟低光照条件下安防摄像头的传感器噪声,提升模型对含噪图像识别的能力,具体公式为:

$$I'(x, y) = I(x, y) + N(0, \sigma^2) \quad (11)$$

其中, $N(0, \sigma^2)$ 为零均值高斯噪声, $\sigma \in [0, 0.02]$ 为噪声标准差。

2.3 评价指标

采用模型参数数量 (parameters, Param.)、计算复杂度 (giga floating-point operations per second, GFLOPs)、Top-1 准确率 (top-1 accuracy, Top-1 AC) 与 Top-3 准确率 (top-3 accuracy, Top-3 AC) 作为评价指标。

模型参数数量用于衡量模型的存储成本与边缘部署可行性,为网络中所有可训练参数的总和,即:

$$Param. = \sum Param_{conv} + \sum Param_{attn} + \sum Param_{fc} \quad (12)$$

计算复杂度用于评估模型的实时推理效率,为完成一次前向推理所需的浮点运算总数为:

$$GFLOP_s = \sum GFLOP_{conv} + \sum GFLOP_{attn} + \sum GFLOP_{fc} \quad (13)$$

准确率为:

$$Top-1 AC = \frac{N_{correct,1}}{N_{total}} \times 100\% \quad (14)$$

$$Top-3 AC = \frac{N_{correct,3}}{N_{total}} \times 100\% \quad (15)$$

其中, $N_{correct,1}$ 为最高置信度预测正确的样本数, N_{total} 为测试集总样本数, $N_{correct,3}$ 为真实标签落入前 3 预测类别的样本数。

3 结果分析

3.1 实验设置

所有实验均在统一工作环境展开,硬件配置的中央处理器 CPU 为 Intel Core i7-13700K,图形处理器 GPU 为 NVIDIA RTX 4090,内存 RAM 为 64 GB;软件环境是基于 Windows 操作系统搭建,采用 PyTorch 深度学习框架,编程语言为 Python 3.8。模型训练的优化器采用随机梯度下降法,初始学习率为 1×10^{-3} ,动量参数为 0.9,权重衰减系数为 1×10^{-4} ;训练总轮数为 100,批次大小为 32。损失函数采用交叉熵损失,其计算公式为:

$$L = - \frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_{n,c} \log(p_{n,c}) \quad (16)$$

其中, N 表示批次样本数量; C 表示对应 6 种检测的光伏板状态; $y_{n,c}$ 表示第 n 个样本的真实标签,采用 one-hot 编码,属于第 c 类时为 1,反之为 0; $p_{n,c}$ 表示模型通过 softmax 函数输出的第 n 个样本属于第 c 类的概率。

3.2 损失曲线对比分析

为验证所提跨尺度特征融合网络的性能优势,选取了 ResNet、VGG、ViT 和 Swin-Transformer^[20] 这 4 类主流架构作为对比基准。图 2 给出了各类方法在训练和验证时的损失曲线。其中,图 2 中①~④分别表示 ResNet 系列、VGG 系列、ViT 系列和 Swin-Transformer 系列模型的训练过程损失曲线;图 2 中⑤~⑧分别表示 ResNet 系列、VGG 系列、ViT 系列和 Swin-Transformer 系列模型的验证过程损失曲线。

通过图 2 可以看出,各模型在训练集与验证集上的损失值随训练轮数的变化趋势。具体而言,随着训练轮次增加,损失值呈下降趋势,表明模型逐步学习到有效特征。不同模型在收敛速度、最终损失表现上存在差异,部分模型在训练集损失快速降低,但验证集损失下降至一定程度后出现波动,而表现更优的模型在训练集与验证集损失同步下降且差距较小。

进一步对比各系列模型与提出模型的训练和验证损失曲线,如图 3 所示。其中,图 3 上图展示了训练过程中各模型的损失变化,图 3 下图展示了验证过程中各模型的损失变化。

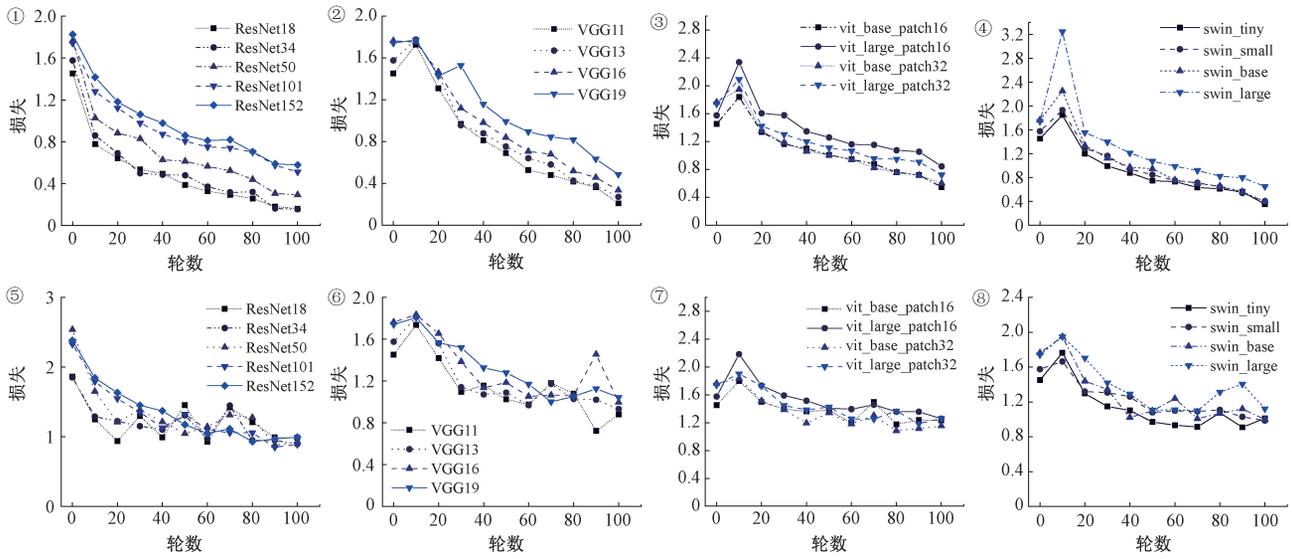


图 2 各模型训练和验证的损失曲线

Fig. 2 The loss curves for the training and validation of each model

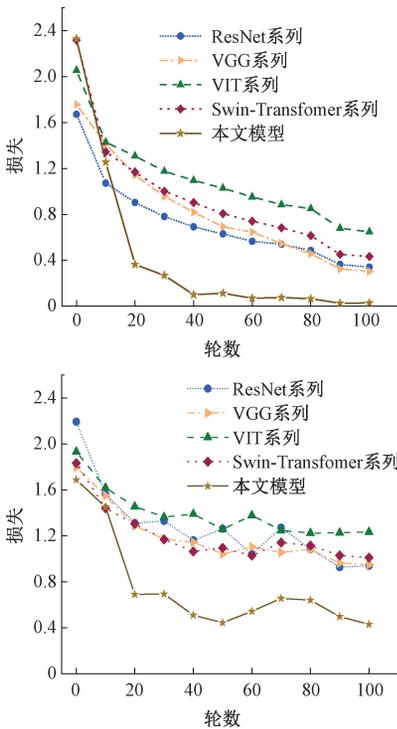


图 3 系列模型与提出模型的训练和验证损失曲线

Fig. 3 The training and validation loss curves of the series model and the proposed model

由图 3 可知,提出模型在训练和验证过程中,损失值下降迅速且最终稳定在较低水平。训练损失从初始轮数的 2.329 0 快速下降,到第 99 个轮数时降至 0.027 9,这表明模型对训练数据的拟合能力持续增强。验证损失从初始的 1.688 5 波动下降至 0.430 6,整体呈

现收敛趋势。

3.3 性能指标对比分析

为了分析各模型在光伏板故障诊断任务中的表现,将从参数量 (Param.)、计算复杂度 (GFLOPs)、Top-1 准确率 (Top-1 AC) 和 Top-3 准确率 (Top-3 AC) 这 4 个关键指标进行对比分析,详细结果如表 2 所示。

由表 2 可知,与其他模型相比,提出模型参数量为 125.59 M,GFLOPs 为 32.19,在保持较高准确率的同时复杂度处于中等水平。其中,VGG11 的参数量为 128.79 M,GFLOPs 为 7.61,虽然参数量与提出模型相近,但 Top-1 准确率远低于提出模型,未能在模型复杂度和性能之间取得良好的平衡。此外,提出模型在 Top-1 准确率上达到了 93.33%,远高于 ResNet34 模型 (准确率 80%)、VGG11 模型 (准确率 81.67%)、vit_base_patch32 模型 (准确率 63.33%) 和 swin_small 模型 (准确率 70%)。而在 Top-3 准确率上,提出模型的准确率达到 100%,表明所有测试样本的真实故障类型均落入前 3 预测类别。相比之下,ResNet34、VGG11 等模型的 Top-3 准确率虽接近 95%,但 Top-1 准确率普遍低于 82%,在单类别诊断方面存在不足。

为了更直观地展示不同模型在光伏板故障诊断任务中的性能差异,对测试集中的样本进行了 Top-1 准确率和 Top-3 准确率的可视化分析,结果如图 4、5 所示。

如图 4 右侧柱状图可知,提出模型的平均 Top-1 准确率达 99.52%,远超 VGG 系列的 89.64%、ResNet 系列的 88.17%、Swin-Transformer 系列的 85.48% 和 VIT 系列的 77.45%。

表2 结果分布对比分析
Table 2 Comparison analysis of result distribution

方法	结构	Param.	GFLOPs	Top-1 AC/%	Top-3 AC/%
ResNet 系列	ResNet18	11.18	1.82	80.00	86.67
	ResNet34	21.29	3.68	80.00	93.33
	ResNet50	23.52	4.13	78.33	95.00
	ResNet101	42.51	7.86	76.67	93.33
	ResNet152	58.16	11.60	70.00	90.00
VGG 系列	VGG11	128.79	7.61	81.67	95.00
	VGG13	128.98	11.30	78.33	91.67
	VGG16	134.29	15.45	76.67	90.00
	VGG19	139.59	19.63	73.33	86.67
ViT 系列	vit_base_patch16	85.65	16.86	61.67	90.00
	vit_large_patch16	303.11	59.69	61.67	80.00
	vit_base_patch32	87.42	4.37	63.33	88.33
	vit_large_patch32	305.46	15.26	61.67	85.00
Swin-Transformer 系列	swin_tiny	27.50	4.37	68.33	91.67
	swin_small	48.79	8.54	70.00	93.33
	swin_base	86.69	15.17	65.00	95.00
	swin_large	194.91	34.08	68.33	86.67
本文模型	ctfnet_base	125.59	32.19	96.67	100.00

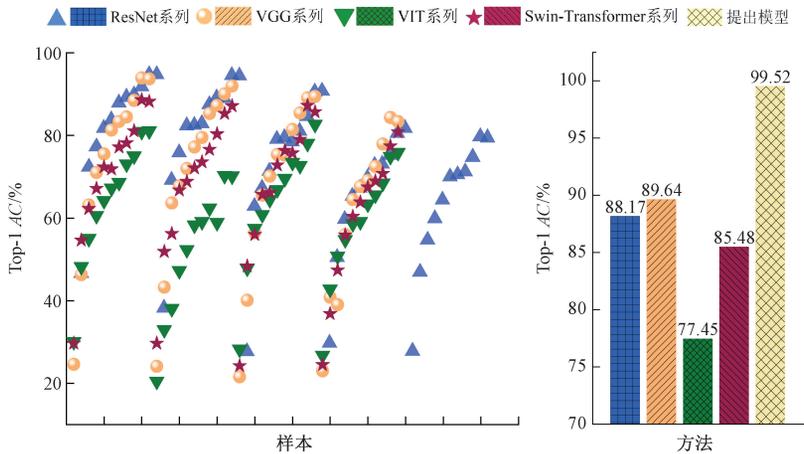


图4 各模型 Top-1 准确率对比

Fig. 4 Comparison of Top-1 accuracy for each model

如图5所示,在 Top-3 准确率方面,提出模型同样表现优异,平均 Top-3 准确率高达 99.88%。相比之下,其他系列模型 Top-3 平均准确率较低,进一步验证了提出模型在故障诊断中优异性能。

3.4 超参数性能分析

为了深入分析超参数设置对模型性能的影响,将开展两方面内容研究:一方面调整 Transformer Encoder 层的

堆叠数量;另一方面调整特征融合模块中局部与全局特征的融合比例。

1) Encoder 层堆叠数量的影响

模型的参数量、计算复杂度随 Encoder 层堆叠数量的增加呈线性增长。为了分析不同 Encoder 层堆叠数量对模型性能的影响,设置堆叠数量 L 分别取 1、2、3、4、5 开展实验,结果如图6所示。

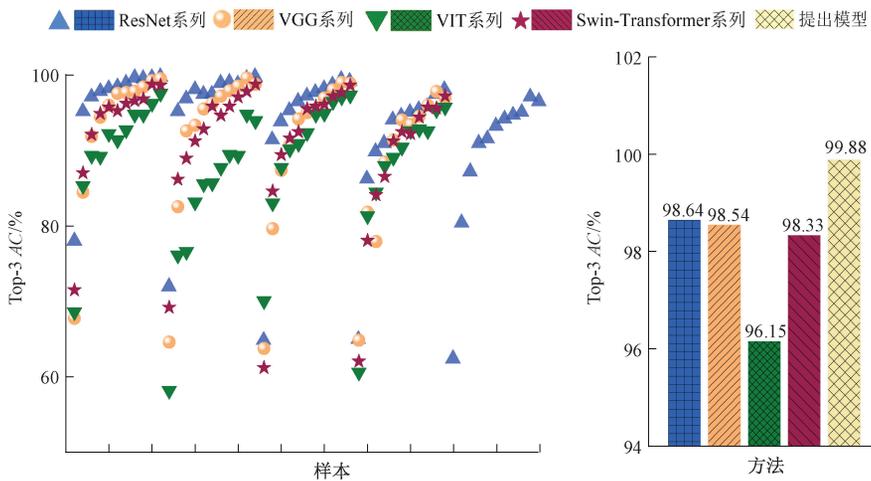


图 5 各模型 Top-3 准确率对比

Fig. 5 Comparison of Top-3 accuracy for each model

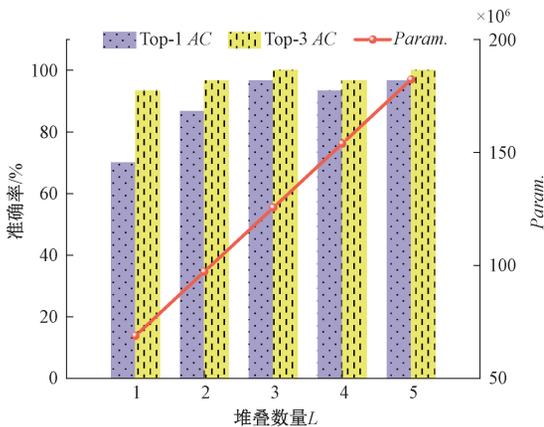


图 6 不同堆叠层数对模型性能影响结果

Fig. 6 The results of the impact of different stacking layer numbers on model performance

通过图 6 可以看出,模型参数量与堆叠数量 L 之间是对数增长的关系。当 L 较小时,模型的参数量较少,对图像全局信息的学习能力有限,整体性能表现较差;随着 L 增加,模型的参数量相应增多,对图像全局信息的捕捉能力得到显著提升,性能也随之改善。然而,当 $L > 3$ 时,模型的精度增长逐渐趋于平缓并达到稳定状态,表明模型已充分学习到图像的全局信息,进一步增加 L 对精度的提升效果并不明显,反而会大幅增加模型的参数量和计算复杂度。

2) 特征融合比例

在特征融合模块中,融合比例决定了局部特征与全局特征在模型中的贡献程度。为了探究不同融合比例对模型性能的影响,设计了多组对比实验,结果如表 3 所示。

表 3 不同融合比例对模型性能影响

Table 3 The impact of different fusion ratios on model performance

场景	融合位置 1	融合位置 2	融合位置 3	融合位置 4	Top-1 AC/%	Top-3 AC/%
场景 1	1/8	1/8	1/8	1/8	91.667	96.667
场景 2	1/3	1/3	1/3	1/3	93.333	100.000
场景 3	1/8	1/4	1/2	1	96.667	100.000
场景 4	1	1/2	1/4	1/8	78.333	93.333
场景 5	1	1	1	1	70.000	88.333

从表 3 中可以看出,当各位置的融合权重较小时(场景 1),全局信息对局部特征提取的指导作用有限,模型性能的提升幅度较小。当融合权重增加到 1/3 时(场景 2),全局信息开始在模型中发挥重要作用,模型性能得到提升。然而,不同的特征图位置应该赋予不同的指导权重,充分利用局部特征和全局特征的优势。在场景 4 中,

卷积层前面的层主要学习低级特征,如线条、简单纹理等,这些特征并不包含目标的整体信息。如果在这一层使用较大的全局信息权重进行引导,会破坏原始的低级语义信息,导致模型性能下降。而在场景 3(提出模型)中,前面的层使用较小的权重进行引导,后面的层则通过全局信息引导局部特征的学习,这种方式能够有效地提

高模型性能。场景5中所有位置的融合权重均为1,同样面临着破坏低级特征学习的问题,导致模型性能较差。

3.5 注意力可视化分析

为直观验证跨尺度特征融合机制的有效性,采用 Grad-CAM 方法对模型的注意力分布进行可视化,重点对比提出模型与 ResNet34、VGG11、ViT-Base32、Swin-Small 模型在故障区域定位能力上的差异。典型光伏板故障注意力热力图如图7所示,其中,深色阴影区域表示模型重点关注的特征区域,颜色越深代表注意力权重越高。

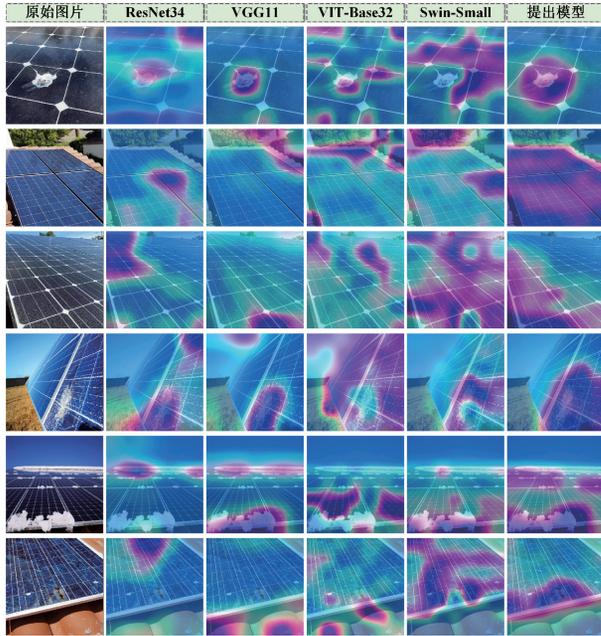


图7 典型光伏板故障注意力热力图

Fig. 7 Typical fault attention heat map of PV panels

通过图7可视化结果能够观察到,提出模型热力图高度聚焦于真实故障区域如鸟粪斑块、积雪覆盖区,且边界清晰;相比之下,ResNet34与VGG11的热力图存在明显的注意力不聚焦现象。在第1行鸟粪图像中,ResNet34的热力图覆盖了光伏板的非故障区域,而提出模型热力图严格限定在鸟粪斑块内。对于多尺度故障,提出模型热力图同时覆盖了裂纹和积灰区域,体现了跨尺度特征融合的优势;而ViT-Base32与Swin-Small的热力图仅聚焦于单一尺度,导致多尺度故障存在漏检风险。在第5行积雪覆盖图像中,存在复杂的天空背景,提出模型热力图能有效抑制背景噪声,仅对光伏板表面的积雪区域产生响应,而其他模型的热力图存在背景误识别现象,影响故障定位的准确性。

3.6 现场部署适配性分析

为了保证模型在光伏电站现场使用的性能,提出现场部署适配性“模型精度-使用效率-应用规模”三维指标,主要面向高精度、低计算复杂度、小模型参数的需求。

其中,模型精度是硬性前提,若 Top-1 准确率低于 90%,模型将无法满足基本诊断要求,然后,计算模型使用效率得分 (efficiency score, ES) 与应用规模得分 (scale score, SS),定义其取值范围为 0~100,数值越大表示计算效率越高、模型规模越小,具体为:

$$ES = 100 \times \frac{GFLOPs_{max} - GFLOPs}{GFLOPs_{max} - GFLOPs_{min}} \quad (17)$$

$$SS = 100 \times \frac{Param._{max} - Param.}{Param._{max} - Param._{min}} \quad (18)$$

其中, $GFLOPs_{max}$ 、 $GFLOPs_{min}$ 表示模型的最大、最小计算复杂度, $Param._{max}$ 、 $Param._{min}$ 表示模型的最大、最小参数数量。

各模型在光伏板故障诊断现场部署中的三维适配性散点图,如图8所示。其中, x 和 y 轴分别表示计算效率得分和应用规模得分, z 轴表示 Top-1 准确率,灰色半透明平面 ($z=90$) 上方为现场部署可行区,代表光伏场站对模型精度的要求。

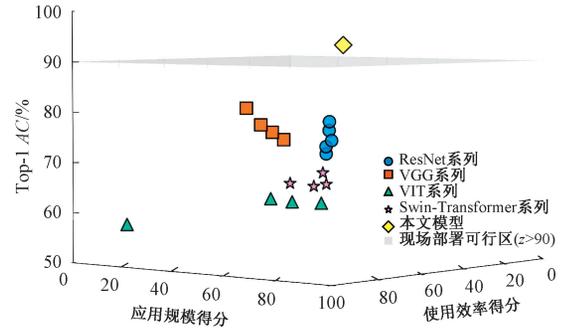


图8 三维适配性散点图

Fig. 8 Three-dimensional adaptive scatter plot

从图8中的散点分布可知,ResNet系列和Swin-Transformer系列模型在计算效率得分和应用规模得分表现较好,散点大多分布在80~100范围,但Top-1准确率普遍低于80,因未达可行区被排除;VGG系列模型在计算效率得分和应用规模得分表现中等,而Top-1准确率最高为81.67%,仍不满足要求;ViT系列模型得分相对较低,三维得分均无优势。提出模型位于可行区平面的上方,虽然在计算效率得分和应用规模得分低于ResNet等模型,但通过精度优先的平衡,成为唯一满足现场需求的模型,体现了其在光伏场站现场部署中的优势。

3.7 泛化性能分析

1) 未知故障类别

为适应未知的新类别,结合分类模型的输出特性,额外增设“其他”类别,保障模型对未知类别的可识别性,避免因类别缺失导致的误判问题。模型输出端针对6个已知类别,通过Softmax函数对输出概率进行归一化处理,最终将概率值最大的类别作为模型的预测

结果。当输入未知新类别样本时,由于其特征与已知 6 个类别的特征存在显著差异,模型输出端 6 个分类神经元的相似度得分均较低。经 Softmax 归一化后,输出概率会相对均匀地分布到 6 个已知类别上,不会出现某一类概率明显偏高的情况。基于上述先验知识,在模型输出 6 个已知类别的概率后,额外增加一层逻辑判断。若某样本对应的 6 类概率中,最大概率值低于 0.3,则直接判定该样本不属于现有 6 个已知类别,将其归类至“其他”类别。

为验证该方法的可行性,额外选取 70 个实验样本,覆盖 7 个类别,每类各 10 张图像。其中,“其他”类别的样本涵盖树叶遮挡、表面水渍、边框变形等小概率光伏板表面状态,详细结果如表 4 所示。

由表 4 可知,光伏板正常、积灰附着、物理损伤、积雪覆盖四类的识别准确率最高,均达到 100%,鸟粪附着与电气损失两类各有 1 张图像识别错误,准确率为 90%。“其他”类别整体识别效果良好,仅 2 张因目标尺寸过小误识别为“正常(清洁)”,其余 8 张均实现正确归类。综合来看,新增“其他”类别的策略未对原有 6 类的准确率产生干扰,且在泛化能力上能够有效判断未知类别,整体具有较高的适用性。

表 4 新增分类准确率对比结果

Table 4 Comparison results of accuracy rates for newly added categories

类别名	样本数	正确数	准确率/%
正常(清洁)	10	10	100.00
鸟粪附着	10	9	90.00
积灰覆盖	10	10	100.00
电气损失	10	9	90.00
物理损伤	10	10	100.00
积雪覆盖	10	10	100.00
其他	10	8	80.00
平均准确率	—	—	94.29

2) 鲁棒性分析

为进一步验证所提模型在复杂场景下的鲁棒性,将从图像尺寸变化、光照条件差异、环境干扰类型这 3 个维度设计实验,每个类包含基准组与干扰组,确保单一变量控制。同时,延续上文 7 类光伏板表面状态类别,每个基准和干扰样本分别采集 20 张图像,详细的实验设计与结果如表 5 所示。

表 5 模型鲁棒性验证实验设计与结果

Table 5 Experimental design and results for model robustness verification

干扰类别	干扰条件	正常(清洁)	鸟粪附着	积灰覆盖	电气损失	物理损伤	积雪覆盖	其他	平均准确率/%	偏差/%
图像尺寸	640×640(基准)	100	90	100	90	100	100	85	95.00	—
	320×320(缩小)	100	80	95	80	100	100	70	89.29	-5.71
	1 024×1 024(放大)	100	90	100	90	100	100	85	95.00	0.00
光照条件	正常光照(基准)	100	90	100	90	100	100	85	95.00	—
	强光(正午反光)	100	85	95	85	100	100	80	92.14	-2.86
	弱光(傍晚低光)	100	90	90	90	100	100	85	93.57	-1.43
天气干扰	正常环境(基准)	100	90	100	90	100	100	85	95.00	—
	雾天(能见度低)	100	85	100	85	100	95	80	92.14	-2.86
	雨天(湿滑反光)	100	85	95	85	100	95	80	91.43	-3.57

通过表 5 可以看出,模型在图像尺寸放大场景下准确率无波动,证明其对大尺寸图像的特征提取能力稳定;在尺寸缩小场景下,偏差为-5.71%,主要因小尺寸图像细节信息损失导致鸟粪附着、电气损失等依赖局部特征的类别准确率下降。强光与弱光场景下,模型平均准确率仍保持在 92.14% 以上,最大偏差为-2.86%，“其他”类别因目标特征模糊出现误判。雾天、雨天等复杂环境下,模型平均准确率不低于 91.43%。综上,提出模型在无干扰场景下平均准确率

达 95.00%,且在多场景干扰下具备较强鲁棒性,仅在极端干扰下出现可控的准确率偏差,整体满足光伏电站复杂运维场景的需求。

3) 泛化能力分析

为了分析提出模型的泛化能力,验证其在非光伏板场景下的工业视觉检测适配性,选取 SDNET2018 公开数据集作为实验测试样本。该数据集涵盖桥梁、道路与墙面 3 类典型工业检测场景,包含各场景下有无裂缝共 6 类图像样本,具体实验结果如表 6 所示。

表6 SDNET2018数据集结果对比分析

Table 6 Comparison analysis of results for SDNET2018 dataset

方法	Param.	GFLOPs	Top-1 AC/%
ResNet34	21.29	3.68	88.35
ResNet50	23.52	4.13	89.73
VGG11	128.79	7.61	86.56
VGG13	128.98	11.30	87.92
vit_base_patch16	85.65	16.86	78.36
vit_large_patch32	305.46	15.26	76.53
swin_small	48.79	8.54	83.37
swin_base	86.69	15.17	82.21
提出模型	125.59	32.19	91.86

通过表6可以看出,提出模型以91.86%的Top-1准确率达到测试最优,较ResNet50、swin_base分别提升2.13和9.65个百分点。效率方面,提出模型的参数量低于VGG、ViT系列,更适配工业部署。因此,提出模型通过跨尺度特征融合方法,既能精准捕捉细节特征,又能高效利用全局信息,验证了其在跨工业场景检测任务中能力。

4 结 论

提出了一种跨尺度特征融合网络,通过卷积分支提取局部细节特征和Transformer分支提取全局关联信息的双分支架构,实现了光伏板正常、积灰、鸟粪、积雪、物理损伤、遮挡这6类状态的精准诊断。该方法能够利用光伏场站现有的安防设备,无需增加硬件成本,通过与ResNet、VGG、ViT、Swin-Transformer系列主流模型对比,与现场部署适配性指标分析,能够得出的结论为:

1) 提出的Transformer与卷积双分支融合框架,能够实现全局信息对局部特征提取的有效辅助,并结合两者在局部和全局特征提取上的优势,在卷积分支引入密集连接方法,增强了特征的传播和复用,提升了局部特征提取效果;

2) 提出模型在Top-1准确率达93.33%,Top-3准确率达100%,分别较次优模型提升11.66%,能确保真实故障类型落入前3预测候选,有效降低光伏运维中的漏检风险;

3) 提出模型的参数数量为125.59M与GFLOPs为32.19,仅为vit_large_patch16等大模型的41.4%和54.0%,适配光伏现场的部署需求。

尽管所提模型在光伏板故障诊断中表现优异,但实际现场的故障类型更为复杂,后续将构建覆盖多地域、多气候条件的大规模光伏故障数据集;此外,将利用知识蒸馏技术设计更轻量化的网络架构,使其适配算力较低的终端设备。

参考文献

- [1] XIAO W B, DONG H F, WU H M, et al. Wireless photovoltaic fault monitoring system[J]. Instrumentation, 2025, 12(2): 23-35.
- [2] 范思远, 王煜, 曹生现, 等. 积灰对光伏组件输出特性影响建模与分析[J]. 仪器仪表学报, 2021, 42(4): 83-91.
- [3] FAN S Y, WANG Y, CAO SH X, et al. Effect modeling and analysis of dust accumulation on output characteristics of photovoltaic modules[J]. Chinese Journal of Scientific Instrument, 2021, 42(4): 83-91.
- [4] ZHANG J L, YANG W J, CHEN Y M, et al. Fast object detection of anomaly photovoltaic (PV) cells using deep neural networks[J]. Applied Energy, 2024, 372: 123759.
- [5] WANG J J, BI L, MA X D, et al. An efficient YOLOX-based method for photovoltaic cell defect detection[J]. Instrumentation, 2024, 11(2): 83-95.
- [6] YOUSIF H, AL-MILAJI Z. Fault detection from PV images using hybrid deep learning model[J]. Solar Energy, 2024, 267: 112207.
- [7] SU B, CHEN H, ZHOU ZH. BAF-detector: An efficient CNN-based detector for photovoltaic cell defect detection[J]. IEEE Transactions on Industrial Electronics, 2021, 69(3): 3161-3171.
- [8] MAITY R, SHAMAUN A M, PATI A. An approach for detection of dust on solar panels using CNN from RGB dust image to predict power loss[M]. Cognitive Computing in Human Cognition: Perspectives and Applications, Cham: Springer International Publishing, 2020: 41-48.
- [9] SRIDHARAN N V, VAITHIYANATHAN S, AGHAEI M. Voting based ensemble for detecting visual faults in

- photovoltaic modules using AlexNet features[J]. Energy Reports, 2024, 11: 3889-3901.
- [9] LIU B, CHEN L, SUN K, et al. A hot spot identification approach for photovoltaic module based on enhanced U-net with squeeze-and-excitation and VGG19[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 3379412.
- [10] ABDELSATTAR M, ABDELMOETY A, EMAD-ELDEEN A. ResNet-based image processing approach for precise detection of cracks in photovoltaic panels[J]. Scientific Reports, 2025, 15(1): 24356.
- [11] 冯秀坤, 刘国通, 李占峰, 等. 基于 YOLOX 算法的西北荒漠地区光伏发电板异物监测系统设计与[J]. 工业仪表与自动化装置, 2025(3): 8-14.
FENG X K, LIU G T, LI ZH F, et al. Design and application of foreign object monitoring system for photovoltaic panels based on YOLOX algorithm[J]. Industrial Instrumentation & Automation, 2025(3): 8-14.
- [12] 吕游, 郑茜, 齐欣宇, 等. 基于改进 EfficientNet 的红外图像光伏组件故障识别研究[J]. 仪器仪表学报, 2024, 45(4): 175-184.
LYU Y, ZHENG X, QI X Y, et al. A study on fault recognition of photovoltaic module with infrared images based on improved efficientnet[J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 175-184.
- [13] DU T F, LI Q, REN J X, et al. Photovoltaic system fault diagnosis method based on physics-constrained causal discovery and causal perception graph neural network[J]. Energy, 2025, 340: 139248.
- [14] RAMADAN E A, ABOUZALM B A, EL-BANBY G M, et al. An innovative transformer neural network for fault detection and classification for photovoltaic modules[J]. Energy Conversion and Management, 2024, 314: 118718.
- [15] KHAN U, KHAN S F, KHALIL I U, et al. Deep learning based vision transformer approach for detecting overlap PV faults using multi labeling[J]. Measurement, 2025, Part E: 118537.
- [16] GUO ZH L, LU J Y, CHEN Q, et al. TransPV: Refining photovoltaic panel detection accuracy through a vision transformer-based deep learning model[J]. Applied Energy, 2024, 355: 122282.
- [17] MAHBOOB Z, KHAN M A, LODHI E, et al. Using segformer for effective semantic cell segmentation for fault detection in photovoltaic arrays[J]. IEEE Journal of Photovoltaics, 2024, 15(2): 320-331.
- [18] WANG J J, DU H, ZENG Y Y. PDFFormer: Efficient vision transformer for photovoltaic defect detection[J]. IEEE Transactions on Consumer Electronics, 2025, 71(2): 6602-6611.
- [19] ZHANG Y L, TIAN Y P, KONG Y, et al. Residual dense network for image super-resolution[C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2472-2481.
- [20] XU SH J, MA H, EKANAYAKE C, et al. Swin Transformer-based transferable PV forecasting for new PV sites with insufficient PV generation data[J]. Renewable Energy, 2025, 246: 122824.

作者简介



何明月, 2019 年于吉林化工大学获得学士学位, 2022 年于东北电力大学获得硕士学位, 现为东北电力大学博士研究生, 主要研究方向为光伏系统故障诊断。

E-mail: 1202500136@neepu.edu.cn

He Mingyue received her B. Sc. degree from Jilin University of Chemical Technology in 2019, and her M. Sc. degree from Northeast Electric Power University in 2022. She is currently a Ph. D. candidate at Northeast Electric Power University. Her main research interest includes fault diagnosis of photovoltaic systems.



范思远(通信作者), 2014 年于东北电力大学获得学士学位, 2018 年于东北电力大学获得硕士学位, 2022 年于东北电力大学获得博士学位, 现为东北电力大学副教授, 主要研究方向为光伏积灰检测和功率预测。

E-mail: fans@neepu.edu.cn

Fan Siyuan (Corresponding author) received his B. Sc., M. Sc., and Ph. D. degrees all from Northeast Electric Power University in 2014, 2018, and 2022, respectively. He is currently an associate professor at Northeast Electric Power University. His main research interests include photovoltaic dust accumulation detection and power prediction.



曹生现,1997年于东北电力大学获得学士学位,2002年于东北电力大学获得硕士学位,2009年于华北电力大学获得博士学位。现为东北电力大学教授。主要研究方向为新能源发电测控技术。

E-mail: csxlb_jl@163.com

Cao Shengxian received his B. Sc. and M. Sc. degrees from Northeast Electric Power University in 1997 and 2002, respectively, and his Ph. D. degree from North China Electric Power University in 2009. He is currently a professor at Northeast Electric Power University. His main research interest includes renewable energy power generation detection and control technology.



张艳辉,2008年于青岛科技大学获得学士学位,2011年于东北电力大学获得硕士学位,2014年于中国科学院大学获得博士学位,现为中国科学院深圳先进技术研究院副研究员,主要研究方向为电池储能和检测算法。

E-mail: zhangyh@siat.ac.cn

Zhang Yanhui received his B. Sc. degree from Qingdao University of Science and Technology in 2008, his M. Sc. degree from Northeast Electric Power University in 2011, and his Ph. D. degree from University of Chinese Academy of Sciences in 2014. He is currently an associate researcher at the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. His main research interests include battery energy storage and detection algorithms.