中图法分类号： 文献标识码： 文章编号：1006-8961(XXXX)XX-0001-20

论文引用格式：Guo Yurong, He Yufei, Zhang Ke, Zhang Tiefeng, Yang Hong. Dual-Branch Perceptual Enhanced Infrared and Visible Image Fusion Algorithm for Substation Equipment Detection[J/OL]. Journal of Image and Graphics, XXXX:1-20. DOI: 10.11834/jig.250581. (郭玉荣, 何雨非, 张珂, 张铁峰, 杨宏. 面向变电站设备检测的双分支感知增强红外与可见光图像融合算法[J/OL]. 中国图象图形学报, XXXX:1-20. DOI: 10.11834/jig.250581.)[DOI:10.11834/jig.250581]

面向变电站设备检测的双分支感知增强红外与可见光图像融合算法

郭玉荣[1,3,4], 何雨非[2], 张珂[1,3,4*], 张铁峰[1,3,4], 杨宏[2,3,4]

1. 华北电力大学,燕赵电力实验室,河北 保定 071003; 2. 华北电力大学,电子与通信工程系,河北 保定 071003; 3. 河北省电力物联网技术重点实验室,河北 保定071003; 4. 电力物联智慧化技术河北省工程研究中心,河北 保定 071003

摘 要：目的 红外与可见光图像融合能够为变电站设备检测提供丰富的特征信息,然而,现有融合算法所生成的图像中目标设备显著性不足、结构信息模糊、与背景区分度较低,导致检测模型难以提取有效的判别性特征,影响检测模型的准确性与鲁棒性。为此,本文提出一种面向变电站设备检测的双分支感知增强红外与可见光图像融合算法。**方法** 本文设计了一种面向变电站设备结构-细节特征的"解耦-增强-融合"框架,通过引入特征增强机制的双路编码架构,有效提升了融合图像中设备的拓扑完整性与特征显著性。具体来说,首先,构建由共享分支与互补分支构成的双路编码架构,实现设备共性结构特征与细节纹理、热特征的有效解耦;其次,针对设备共性结构特征,设计结构增强模块(structure enhancement module, SEM)以强化设备轮廓;针对设备细节纹理与热特征,引入多分支特征增强模块(multibranch feature enhancement module, MFEM)以强化其对设备外观及温度等关键信息的表达;最后,通过共享-互补特征融合模块实现多模态特征有效整合,经由解码器重构出结构清晰、细节丰富的高质量融合图像。**结果** 本文通过在变电站设备红外与可见光图像数据集上的实验表明:在红外与可见光图像融合任务中,本文模型在EN、SF、MI、$Q^{AB/F}$、PSNR以及SSIM六项常规指标上均展现了优异的性能。在变电站设备检测任务中,本文算法在悬式绝缘子、柱式绝缘子、电流互感器、电压互感器以及套管上均表现出较高的准确性,在平均准确率(mAP50)方面,较红外图像提升40.1%,较可见光图像提升1.2%,并优于先进图像融合算法3.9%,显著提升了变电站设备检测的鲁棒性。此外,本文通过消融实验与可视化分析进一步验证了算法有效性及模型合理性。**结论** 本文算法显著提升了在变电站场景下,红外与可见光图像融合生成的图像质量以及设备检测性能,相较于单模态图像与现有融合方法具有明显优势,实现了图像融合技术与变电站应用场景的紧密结合,有助于推动多模态融合技术在电力智能巡检中的实际应用与进一步发展。

关键词： 电力巡检;目标检测;图像融合;特征增强;注意力机制

Dual-Branch Perceptual Enhanced Infrared and Visible Image Fusion Algorithm for Substation Equipment Detection

Guo Yurong[1,3,4], He Yufei[2], Zhang Ke[1,3,4*], Zhang Tiefeng[1,3,4], Yang Hong[2,3,4]

1. Yanzhao Electric Power Laboratory of North China Electric Power University, Baoding 071003, Hebei, China; 2. Department of Elec-

收稿日期:2025-11-17;**修回日期**:2026-05-10

* **通信作者**:张珂 zhangkeit@ncepu.edu.cn

基金项目:国家自然科学基金项目(62506129);中央高校基本科研业务费专项资金项目(2025MS117);河北省自然科学基金项目(F2024502017)

Supported by: National Natural Science Foundation of China (62506129); Fundamental Research Funds for the Central Universities (2025MS117); Natural Science Foundation of Hebei Province(F2024502017)

tronic and Communication Engineering of North China Electric Power University, Balding 071003, Hebei, China; 3. Hebei Key Laboratory of Power Internet of Things Technology, Baoding 071003, Hebei, China; 4. Hebei Engineering Research Center of Intelligent Technology for Power Internet of Things, Baoding 071003, Hebei, China

Abstract: Objective The precise identification of substation equipment is crucial for ensuring the stability and operational security of power systems. Traditional detection methods relying solely on single-modality imaging, whether infrared thermal imaging or visible images, struggle to comprehensively characterize the multidimensional attributes of complex electrical apparatus. While the infrared modality effectively captures thermal information of equipment, it lacks the spatial resolution and textural details required for structural assessment. Conversely, visible images provide excellent detail fidelity but are fundamentally blind to thermal phenomena directly related to equipment health status, and are significantly affected by weather and environmental conditions. This representational dichotomy highlights the core value of infrared-visible image fusion, a technique that synthesizes complementary diagnostic information into a unified visual representation. Despite its theoretical promise, existing fusion methods exhibit significant shortcomings in substation inspection scenarios: fused results often suffer from insufficient target saliency, ambiguous structural definition, and poor foreground-background differentiation. These deficiencies severely compromise the discriminative features required by downstream detection algorithms, ultimately reducing the inspection accuracy and operational robustness of the entire power infrastructure network. To address these long-standing limitations through architectural innovation, this study proposes a dual-branch perceptual enhancement framework specifically designed to optimize the efficacy of substation equipment inspection. **Method** The architecture adopts a dual-branch paradigm. The shared-branch encoder utilizes a Transformer architecture to extract high-level structural commonalities invariant across both infrared and visible spectra, capturing the geometric continuity, topological relationships, and spatial configurations that define substation equipment categories. The complementary-branch encoder employs domain-optimized convolutional blocks to isolate modality-specific features, specifically the detailed texture information from visible images and the thermal information from infrared images. Compared to traditional methods, our algorithm achieves a tight integration between the feature enhancement mechanism and the substation equipment detection scenario. By targeting the enhancement of equipment structure and key details, it establishes a novel structure-detail decoupling fusion paradigm, significantly improving the topological integrity and feature saliency of equipment in the fused image. Specifically, we introduce a self-attention and structure enhancement Module (SEM) operating on the shared features. This module dynamically constructs attention maps through cross-modal feature correlation, utilizing learned spatial weighting to selectively enhance equipment contours while suppressing irrelevant background structures, thereby improving the saliency of equipment structure and its distinguishability from the background. Simultaneously, the multibranch feature enhancement module (MFEM) processes complementary features through parallel convolutional streams with cascaded refinement blocks to enhance the expression of detailed textures and thermal information of the target equipment. The refined feature tensors then undergo modality-specific fusion via feature fusion modules. Finally, the decoder reconstructs the fused features into the fused image, ensuring the preservation of both global structural coherence and local diagnostic details throughout the inverse transformation process. **Result** Experimental validation was conducted using rigorously curated substation-specific infrared-visible image pairs capturing diverse equipment types under various operating conditions. In downstream equipment detection tasks evaluated using industry-standard frameworks, the algorithm demonstrated exceptional performance across key categories including suspension insulators, post insulators, current transformers, voltage transformers, and bushings. Quantitative assessment revealed significant improvements in the authoritative metric for object detection reliability, mean average precision (mAP@0.5): achieving a 40.1% enhancement compared to infrared-only detection, a 1.2% improvement over visible-only baselines, and a 3.9% gain over existing State-Of-The-Art (SOTA) methods, significantly enhancing the robustness of substation equipment detection. Regarding fused image performance, a comprehensive evaluation across six established dimensions consistently demonstrated superiority. The fused outputs excelled in information entropy (EN), spatial frequency (SF), mutual information (MI), $Q^{AB/F}$, peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM), comprehensively outperforming existing fusion methods without exception. Ablation studies systematically isolating architectural contributions confirmed that both the SEM and MFEM modules

are effective in both image fusion and object detection tasks. **Conclusion** This research establishes a transformative paradigm for intelligent substation inspection through perceptually enhanced image fusion. By fundamentally reimagining the synthesis of infrared and visible representations, the framework overcomes the long-standing limitations of conventional methods that compromise equipment detectability. Beyond the direct performance improvements, this study bridges a crucial gap between computer vision theory and power engineering practice, demonstrating that domain-aware fusion architectures are essential for mission-critical infrastructure applications. The methodology provides a foundational advancement toward fully autonomous power grid maintenance, inherently adaptable to the evolving inspection requirements within the energy sector.

Key words: power equipment inspection; object detection; image fusion; feature enhancement; attention mechanism

+中图法分类号:(此号在中国图书馆分类法中查) **文献标识码**:A **文章编号**:1006-8961(年)-

论文引用格式:Guo Yurong, He Yufei, Zhang Ke, Zhang Tiefeng, Yang Hong. Dual-Branch Perceptual Enhanced Infrared and Visible Image Fusion Algorithm for Substation Equipment Detection — SCID[J/OL]. Journal of Image and Graphics. DOI:10.11834/jig.250581.(郭玉荣,何雨非,张珂,张铁峰,杨宏.面向变电站设备检测的双分支感知增强红外与可见光图像融合算法—SCID[J/OL].中国图象图形学报.DOI:10.11834/jig.250581.

0 引 言

随着工业化和城市化的快速发展,电力基础设施规模不断扩大,系统复杂度显著提升。变电站作为电力系统中实现电能传输与分配的关键节点,其设备的正常可靠运行对于保障供电安全至关重要(Wang等,2022)。变电站设备检测能够帮助全站设备空间分布拓扑与类型标识的快速构建,为变电站巡检系统提供设备级的空间感知基础。传统的人工检测方式需要耗费大量的人力物力与时间资源,不仅巡检速度缓慢,同时存在较高漏检的风险,难以满足变电站快速精确检测的需求。

目前,基于深度学习的目标检测方法已取得显著进展,基于单一模态图像(包括可见光图像(Xu等,2021;Wang等,2024)或者红外图像(Zhu等,2021;Li等,2023))的变电站设备检测任务也取得了一定成果(Wu等,2024)。然而,单模态图像难以全面表征设备的多维信息,严重制约了单模态图像在变电站设备检测中的精度与鲁棒性。具体而言,可见光图像包含丰富的纹理细节和色彩信息,但其成像质量易受环境光照条件影响,在弱光、雾霾或设备局部遮挡等场景下,这些信息显著退化;而红外图像主要反映目标的温度分布,具有抗环境干扰和全天候工作优势,但其成像质量受热平衡状态、波长限制等因素影响,存在对比度不足、边缘模糊等问题。

近年来,多模态融合算法利用不同模态交互,融合不同模态多维度信息,为下游任务提供更丰富全面的特征表示,显著提升目标检测、图像分割、图像分类等多个下游任务性能(Zhao等,2024)。其中,红外与可见光图像融合算法能够生成具有可见光高频纹理信息与红外图像高对比度温度信息的融合图像,已被广泛应用于医疗卫生(Xu等,2022)、遥感(Shao等,2018)等领域。而在电力巡检领域,红外与可见光图像融合算法能够有效整合可见光图像中设备的细节纹理信息与红外图像中设备的热分布信息,为模型提供更加丰富的特征表示,从而显著提升变电站设备检测的精度与鲁棒性。现有的红外与可见光图像融合算法通常分为传统方法与基于深度学习的方法两大类。传统方法主要包括基于变换域(Liu等,2015)、空间域(Ma等,2016)及稀疏表示(Liu等,2016)的算法等,这些传统算法高度依赖人工设计的特征与规则,难以有效建模复杂的非线性关系,存在泛化能力弱、自适应性差以及计算开销大等问题;基于深度学习的红外与可见光图像融合算法通过其强大的端到端特征学习能力以及对复杂非线性映射的建模优势,为解决传统方法的瓶颈提供了新途径。这类方法根据框架可进一步分为基于自编码器-解码器(AE)(Li等,2019)、卷积神经网络(CNN)(Li等,2021)、生成对抗网络(GAN)(Ma等,2019)与Vision Transformer(VIT)(Zhao等,2023)的红外与可见光图像融合算法,并已在图像融合任务中展现出优异性能。

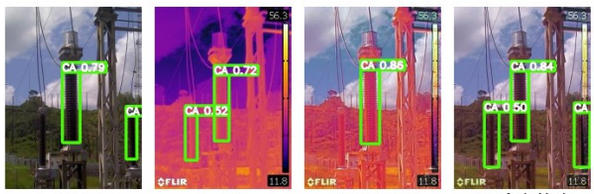

图1　本文算法检测效果与现有算法及源图像的对比

Fig. 1　Comparison of detection performance between our proposed algorithm, existing algorithms, and source images.

然而,现有图像融合算法在变电站设备检测任务中仍存在明显不足,具体表现为融合图像中目标设备不突出,设备结构性信息被弱化,与背景界限模糊,上述问题为检测模型提取判别性特征带来阻碍。如图1(c)所示,设备混叠于背景(植被)中,关键结构辨识度不足,最终引发检测模型错检现象。

究其根源,现有融合方法难以显式区分图像中的低频共性结构与高频独有细节,而是采用统一的融合策略处理所有特征,导致低频结构信息难以被保留与强化,高频细节也缺乏针对性增强。这直接造成融合图像中设备结构性信息不显著,检测模型难以从中提取有效的判别性特征。

然而,红外与可见光图像中的低频成分与高频成分在信息属性上存在本质差异,需要采用差异化的处理策略。Zhao等(2023)通过实证分析表明,红外与可见光图像中的低频成分通常相关性较强,对应模态间共享的结构性信息,如场景中对象的整体轮廓和相对布局;而高频成分则更多体现各模态特异性,如可见光图像中的细微纹理和红外图像中的热辐射分布。在变电站场景下,这一特性尤为明显:低频信息刻画了电力设备共有的空间结构与拓扑关系,是目标检测中位置与类别判别的重要依据;高频信息则分别承载了表观细节与热状态特征,为模型提供更丰富的判别特征。

基于上述分析,本文提出一种面向变电站设备检测的双分支感知增强红外与可见光图像融合算法。本文的核心创新在于设计了一种面向变电站设备结构-细节特征的"解耦-增强-融合"框架,通过引入特征增强机制的双路编码架构,有效提升了融合图像中设备的拓扑完整性与特征显著性。本文首先通过共享分支编码器与互补分支编码器分别提取图像的共享特征与互补特征。共享特征聚焦于低频共性结构信息,用于保持设备的拓扑完整性与结构一致性;互补特征则对应高频独有信息,分别强化设备的可见光纹理与红外热信息表达。进一步,针对共享特征,本文设计结构增强模块(structure enhancement module, SEM),首先利用自注意力机制在单模态内增强设备部件间的结构关联性,构建完整设备拓扑,再通过交叉注意力实现双模态在结构层面的协同增强,提升设备结构显著性及其与背景的区分度;针对互补特征,引入多分支特征增强模块(multi-branch feature enhancement module, MFEM),以突出设备的关键细节纹理与热信息。最后,通过共享特征融合模块与互补特征融合模块整合两类特征,并经解码器重构得到融合图像。本文的主要工作如下:

(1)提出一种面向变电站设备结构-细节特征的"解耦-增强-融合"框架。本框架采用双分支编码器分别提取源图像的共享特征与互补特征,并通过共享特征融合模块与互补特征融合模块整合不同模态的特征信息,最终利用解码器生成设备细节丰富且结构清晰的融合图像。

(2)设计了结构增强模块SEM,通过自注意力机制挖掘单模态共享特征的内部结构关联,并利用交叉注意力机制促进跨模态共享特征间的交互,增强设备结构的显著性,从而提高设备与背景的区分性。

(3)引入了多分支特征增强模块MFEM,强化模型对设备外观及温度等关键信息的表达,强化设备的细节纹理与热特征。

(4)在变电站红外与可见光图像数据集上进行了红外与可见光图像融合与设备检测实验,实验表明,本文算法不仅能取得性能良好的融合图像,同时在变电站设备检测精度上优于现有算法。

1　相关工作

本节对基于深度学习的的目标检测算法以及红外与可见光融合算法进行简要回顾。

1.1　目标检测算法

近年来,得益于其强大的特征自动提取能力、卓越的非线性建模性能以及对大规模数据的高效学习机制,深度学习算法已基本取代了依赖手工特征与浅层分类器的传统检测方法,成为目标检测领域的主导范式。

R-CNN(Girshick等,2014)通过引入卷积神经网

络(CNN)来自动提取区域特征取代手工特征,显著提升了检测精度,然而其采用选择性搜索生成候选区域,同时对每个区域独立进行CNN前传,导致计算冗余严重,严重影响了模型的检测效率;Fast-RCNN(Girshick等,2015)采用了共享卷积计算机制,将整张图像输入至CNN一次性生成特征图,解决了R-CNN因重复卷积而导致的计算效率偏低的问题;Faster R-CNN(Ren等,2016)则在此基础上提出了区域提议网络(RPN),将候选区域生成任务整合到神经网络中,实现了端到端的目标检测,同时RPN与检测网络共享卷积特征图,大幅减少了区域提议的额外计算开销,显著提高了模型的检测速度。

在工业级检测对高精确度以及高实时性的双重需求下,YOLO(You Only Look Once)系列通过端到端的单阶段检测架构持续突破性能边界,并逐渐发展成熟。Ultralytics团队于2020年推出的YOLOv5采用模块化设计,通过改进的CSPDarknet53骨干网络与自适应锚框计算,在COCO数据集上实现了高精度的同时,保持了140FPS的推理速度;在此基础上,YOLOv8采用了新的骨干网络,将YOLOv5中的C3模块替换为C2f,实现了进一步的轻量化,同时抛弃了以往的Anchor-Base,使用了Anchor-Free的思想,在保持实时性能的基础上进一步提升了精度;YOLOv9(Wang等,2024)则通过可编程梯度信息(PGI)和广义高校层聚合网络(GELAN)技术,在参数量减少的情况下提高了模型的精确度,为工业级高精度检测提供了新的基准。

当前,YOLO系列算法在电力设备检测中已取得良好进展,并针对具体任务需求进行了针对性改进。然而,现有方法主要基于单模态可见光图像,检测性能仍存在局限。采用红外与可见光图像融合技术获取高质量融合图像,为提升电力设备检测性能提供了新的解决思路。

1.2 红外与可见光融合算法

目前,深度学习成为红外与可见光图像融合领域的主要方法,按照网络架构主要分为基于自编码器-解码器(AE),基于生成对抗网络(GAN),基于卷积神经网络与基于Vision Transformer(VIT)的红外与可见光图像融合算法四个分支。

基于AE的红外与可见光图像融合算法通常采用编码器提取图像特征,在对图像进行特征级融合后将融合特征输入解码器,得到最终的融合图像;DenseFuse(Li等,2019)受密集连接卷积网络(densely connected convolutional networks)(Huang等,2017)的启发,设计Dense Block作为编码器,充分提取源图像的特征,并对特征分别利用直接相加与加权相加的方式融合,通过解码器得到最终的融合图像;RFN-Nest(Li等,2021)通过巢式连接结构(Nest)作为自编码器提取图像特征,并首次利用神经网络代替传统手工设计的融合方法,设计残差融合网络(RFN)根据特定的融合任务自适应学习融合策略,并在主观与客观指标上取得了效果更加优异的融合图像。

基于GAN的红外与可见光图像融合算法主要通过生成器与判别器的相互对抗使得融合图像在分布上类似于源图像,保留源图像的重要特征并在感知上令人满意。FusionGAN(Ma等,2019)通过生成器生成具有较大红外强度的融合图像,并在判别器中加入可见光图像,强制融合图像逼近可见光图像的分布,使得融合图像能够体现更多可见光图像的纹理特征,实现了红外特征与可见光特征在融合图像中的平衡;DDcGAN(Ma等,2020)则在此基础上,采用了双判别器的结构,分别加入红外图像与可见光图像,使得融合图像能够同时具有红外图像的强度分布以及可见光图像的梯度分布,从而同时保留并增强融合图像中的红外图像热目标显著性以及可见光图像的纹理细节;TarDAL(Liu等,2022)保留了双判别器的结构,分别在判别器中输入红外图像的前景目标与可见光图像的背景细节,使融合图像在学习差异的同时寻求共同点,从红外图像中保留目标的结构信息,从可见光图像中保留纹理细节,不仅能够生成优异的融合图像,同时在在下游检测任务中展现了较好的性能。

基于CNN的红外与可见光图像融合算法通常采用卷积神经网络(convolutional neural networks,CNN)实现图像特征的提取与融合,显著提升了融合质量,同时,研究者们逐渐开始处理图像融合相关的局限性以及实际问题。MetaLearning-Fusion(Li等,2021)提出了一种基于元学习的红外与可见光图像融合深度框架,可根据实际需求通过元升采样模块以任意合适倍数进行上采样,突破了输入-输出图像空间分辨率的约束;DIVFusion(Tang等,2023)充分考虑了微光图像增强与图像融合之间的内在联系,实现了二者的有效耦合与互补,有效解决了弱光场

景下可见光纹理退化问题，生成具有真实色彩与高对比度的融合图像。

随着Transformer（Yang等，2022）技术的发展，基于Vision Transformer（ViT）（Dosovitskiy等，2021）的红外与可见光图像融合算法因其强大的长距离依赖建模能力展现出新的潜力。其中，CddFuse（Zhao等，2023）沿用了双分支结构，将图像全局特征解耦为背景特征与细节特征将图像特征解耦为全局背景特征与高频细节特征，并分别采用轻量化Transformer和可逆神经网络（invertible neural networks，INN）（Gomez等，2017）进行针对性提取，在自然场景融合任务中取得突破性进展。DCEvo（Liu等，2025）首次将进化学习引入图像融合领域，将图像融合与下游双任务的优化统一建模为多目标问题，采用进化算法动态平衡损失函数参数，从而根据任务需求自适应地学习不同模态的互补特征。该方法在提升融合图像质量的同时，有效改善了后续高层任务的性能。TDFusion（Bai等，2025）则将元学习引入图像融合的优化过程中，首次将融合损失设计为可学习的神经网络模块，借助下游任务损失结合元学习反向优化其参数，使损失函数能够动态调整，以适应不同任务的需求。RISFuse（Wang等，2025）实现了融合过程的可控性，通过多模态流形先验，引导文本响应网络从实例分割结果中识别与文本匹配的目标对象，并在融合过程中对目标区域与非目标区域分别施加约束，从而使融合模型能够根据文本增强用户所关注的目标对象。

然而，现有算法在面向变电站设备检测任务时存在显著局限性。如何构建适配变电站复杂环境（如绝缘子串密集排列、设备热斑特性显著等）的图像融合模型，进而有效提升设备检测任务的精度与鲁棒性，已成为电力智能巡检领域的核心研究课题。

2 方 法

本节详细介绍所提出的面向变电站设备检测的双分支感知增强红外与可见光图像融合算法的模型结构以及具体的训练策略，其模型结构如图2所示。

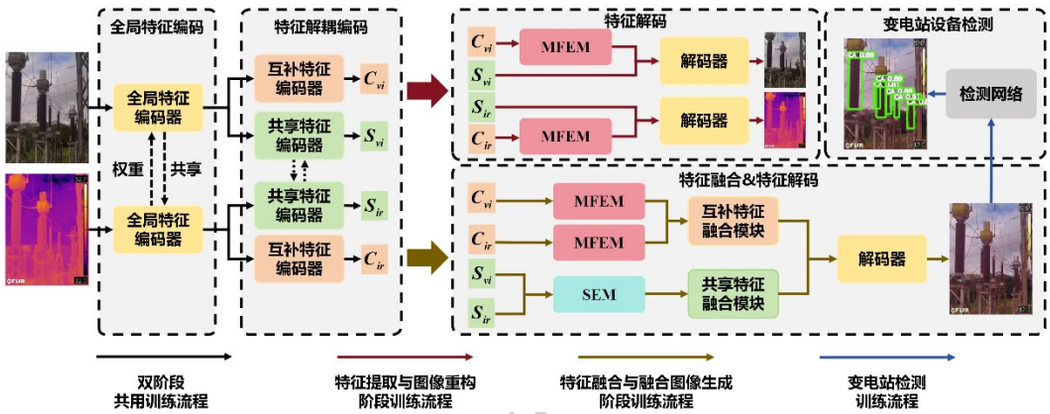

图2 本文方法模型结构示意图

Fig. 2 Architecture diagram of the proposed model

2.1 总体结构

本文提出的面向变电站设备检测的双分支感知增强红外与可见光图像算法采取双阶段模型结构范式。在特征提取与图像重构阶段，通过全局特征编码器、互补特征编码器、共享特征编码器、多分支特征增强模块MFEM以及解码器的联合训练，使模型初步具备特征提取与图像重构能力；在特征融合与融合图像生成阶段，冻结全局特征编码器、互补特征编码器及共享特征编码器，对结构增强模块SEM、共享特征融合模块与互补特征融合模块进行训练，同时微调解码器，使模型具备特征融合与融合图像生成能力。以下将分别通过两个阶段对本文模型进行详细介绍。

2.2 特征提取与图像重构

在特征提取与图像重构阶段，通过全局特征编码器、互补特征编码器、共享特征编码器、多分支特征增强模块MFEM以及解码器联合训练，旨在强化模型的特征提取与图像重构能力。

2.2.1　全局特征编码器

为便于叙述，可见光图像与红外图像分别表示为 $\boldsymbol{I}_{vi} \in \mathbf{R}^{H\times W\times C}$ 与 $\boldsymbol{I}_{ir} \in \mathbf{R}^{H\times W\times C}$。图像全局特征通常表现为多模态图像的大尺度浅层特征，实现图像空间至特征空间的变换。Vision Transformers（Dosovitskiy等，2021）因其全局注意力具备提取图像大尺度特征的能力，但存在计算成本高的问题，为此，本文采用其轻量级变体Restormer，通过共享的全局特征编码器{E_F: $\boldsymbol{I} \to F$}（由Patch Embedding卷积层与Restormer块组成）提取不同模态图像的全局特征，其中Restormer块由多个Restormer单元组成。对于红外图像，其全局特征可表示为：

$$F_{IR} = E_F(\boldsymbol{I}_{IR}) \tag{1}$$

同理，可见光全局特征可表示为：

$$F_{VI} = E_F(\boldsymbol{I}_{VI}) \tag{2}$$

2.2.3　特征解耦编码器

特征解耦旨在将复杂交错的特征分解为多个相互独立、且各自表达特定属性的子特征。本文认为，变电站设备图像中混杂的视觉信息可被有效分解为两个语义明确的组成部分，即共享特征与互补特征。如引言中所述，共享特征主要聚焦于低频共性结构信息，用于维持设备整体的拓扑完整性与结构一致性；而互补特征则对应高频独有信息，分别用于表达可见光图像中的表面纹理与红外图像中的热分布信息。

从信息论的角度出发，该特征解耦过程本质上是对源图像信息在特征层面的有目的重构。具体而言：共享特征的学习，可阐释为对可见光与红外模态在特征空间中互信息的最大化，可表示为：

$$\max I(S_{VI}, S_{IR}) \tag{3}$$

式中，$I(\cdot,\cdot)$为特征互信息算子，S_{VI}与S_{IR}分别为可见光与红外图像的共享特征通过最大化两者之间的互信息，可促使模型强化对共有的低频结构信息的提取，从而在融合过程中保持设备宏观轮廓与拓扑结构的一致性。

互补特征的学习，可阐释为最小化两模态间似有特征的互信息，可表示为：

$$\min I(C_{VI}, C_{IR}) \tag{4}$$

式中，C_{VI}与C_{IR}分别为可见光与红外图像的互补特征，该约束旨在降低两部分互补特征之间的信息冗余，通过拉远其在特征空间中的分布距离，促使模型更纯净地提取并保留各模态特有的高频信息，如可见光中的设备细节纹理与红外图像中的设备热信息。

基于上述特征解耦策略，为有效提取变电站设备红外与可见光图像中的共享特征与互补特征，本文进一步设计了一种双分支编码器结构。

对于共享特征，由于其具有大尺度全局特性，本文设计基于Restormer块的共享特征编码器{E_S: $F \to S$}对其进行提取，利用具有全局注意力机制的Restormer有效捕获这类特征。因此，可见光共享特征可表示为：

$$S_{VI} = E_S(F_{VI}) \tag{5}$$

$$S_{IR} = E_S(F_{IR}) \tag{6}$$

对于互补特征，由于其具有细节以及局部特性，本文设计基于可逆神经网络INN块的互补特征编码器对其进行提取，利用可逆神经网络无损特征信息传输的能力，保证不同模态互补特征中的细节信息能够得到最大程度保留。同时，由于不同模态图像互补特征具有独立性，本文在提取不同模态图像互补特征过程中采用独立权重的两个互补特征编码器，分别为可见光互补编码器{E_{VI}: $F \to C$}与红外互补编码器{E_{IR}: $F \to C$}，则可见光互补特征C_{VI}与红外互补特征C_{IR}可表示为：

$$C_{VI} = E_C^{VI}(F_{VI}) \tag{7}$$

$$C_{IR} = E_C^{IR}(F_{IR}) \tag{8}$$

2.2.4　多分支特征增强模块

变电站红外与可见光图像场景复杂，互补特征编码器中可逆神经网络特征提取能力有限，特征所包含的语义信息较少，且感受野较窄，图像互补特征表达的不同区域的信息容易发生混淆，使检测模型在处理融合图像的过程中容易出现与目标特征相似的干扰特征，为检测模型区分目标与背景带来困难。

因此，本文引入多分支特征增强模块MFEM增强互补特征编码器提取的可见光与红外互补特征，增强图像的细节信息，使得检测模型能够提取更多的判别性特征，从而区分目标与背景。

MFEM主要从两个角度增强图像细节特征：首先，采用多分支卷积结构，以增加特征的丰富度，获取多种语义信息；其次，应用空洞卷积，以获取更加丰富，感受野更大的局部上下文信息。整体结构如图3所示，共分为四个分支，每个分支首先对特征进

行1×1的卷积操作，用于调整特征通道便于后续处理。第一个分支为残差结构，以保留原特征的关键信息；后三个分支执行标准卷积操作，分别运用了3×3，1×3与3×1卷积，以丰富特征语义信息，其中两个分支加入了空洞卷积层，以扩大局部感受野，使得特征图保留更多的上下文信息。具体表示如下：

$$B_1 = f_{conv}^{3\times3}(f_{conv}^{1\times1}(F)) \tag{9}$$

$$B_2 = f_{dconv}^{3\times3}(f_{conv}^{3\times1}(f_{conv}^{1\times3}(f_{conv}^{1\times1}(F)))) \tag{10}$$

$$B_3 = f_{dconv}^{3\times3}(f_{conv}^{1\times3}(f_{conv}^{3\times1}(f_{conv}^{1\times1}(F)))) \tag{11}$$

$$B_4 = f_{conv}^{1\times1}(F) \tag{12}$$

$$F^e = concat(B_1,B_2,B_3)\oplus B_4 \tag{13}$$

式中，B_1，B_2，B_3，B_4分别为MFEM四个分支的输出特征，F^e为MFEM输出的增强特征。$f_{conv}^{1\times1}$，$f_{conv}^{3\times3}$，$f_{conv}^{1\times3}$，$f_{conv}^{3\times1}$分别表示卷积核大小为1×1，3×3，1×3与3×1的标准卷积，$f_{dconv}^{3\times3}$表示卷积核大小为3×3的空洞卷积，且扩张率设置为5，$concat(\cdot)$表示通道拼接操作，$\oplus$表示元素加法。

最后，通过MFEM增强后的可见光互补特征C_{VI}^e与红外互补特征C_{IR}^e可表示为：

$$C_{VI}^e = FEM(C_{VI}) \tag{14}$$

$$C_{IR}^e = FEM(C_{IR}) \tag{15}$$

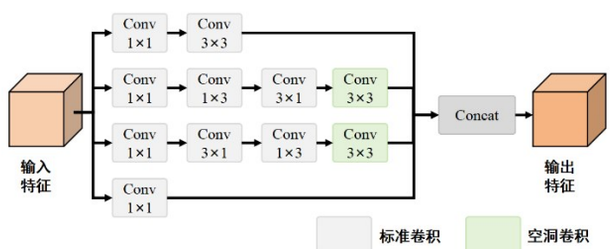

图3　多分支特征增强模块结构示意图

Fig. 3　Architecture diagram of MFEM

为增强模型的图像重构能力，使得解码器能够完成由特征空间至图像空间的转换，模型引入了图像重构损失函数，表示为：

$$L_{recn} = L_{recnIR} + L_{recnVI} \tag{21}$$

式中，L_{recnIR}与L_{recnVI}可分别表示为：

$$L_{recnIR} = \left\| \boldsymbol{I}_{IR} - \widetilde{\boldsymbol{I}_{IR}} \right\|_2^2 + k_1 SSIM(\boldsymbol{I}_{IR},\widetilde{\boldsymbol{I}_{IR}}) \tag{22}$$

2.2.5　解码器

特征解码旨在通过解码器对同一模态下的共享特征与互补特征进行反编码，完成特征空间至图像空间的还原。由于全局特征编码采用基于Restormer块的编码器实现图像空间至特征空间的转换，为保证编码与解码过程的平行，本文采用基于Restormer块的解码器$\{D:(S,C)\rightarrow \boldsymbol{I}\}$，具体可表示为：

$$D(S,C) = \sigma(Conv(RTM(Conv(S,C)))) \tag{16}$$

式中，$\sigma(\cdot)$表示sigmoid函数，$RTM(\cdot)$表示Restormer块，两个卷积层$Conv(\cdot)$分别用于完成共享特征与互补特征的聚合以及特征图降维至图像的过程。

因此，在特征解耦与图像重构阶段，解码器输出的重构可见光图像$\widetilde{\boldsymbol{I}_{VI}}$与重构红外图像$\widetilde{\boldsymbol{I}_{IR}}$可以分别表示为：

$$\widetilde{\boldsymbol{I}_{VI}} = D(S_{VI},C_{VI}^e) \tag{17}$$

$$\widetilde{\boldsymbol{I}_{IR}} = D(S_{IR},C_{IR}^e) \tag{18}$$

2.2.6　损失函数

在特征解耦与图像重构阶段，为增强模型对图像特征的编码及解耦能力，模型引入了特征分解损失，表示为：

$$L_{dis} = \frac{(L_{cc}^C)^2}{L_{cc}^S} = \frac{(CC(C_{VI},C_{IR}))^2}{CC(S_{VI},S_{IR}) + \varepsilon} \tag{19}$$

该损失函数用于增大共享特征的相关性，并减小互补特征的相关性，从而进一步优化互信息目标，促使共享特征最大化其共有信息，同时确保互补特征最小化其信息冗余。其中，ε为防零因子，通常取0.00001。$CC(\cdot,\cdot)$为Pearson相关系数，若给定两个特征向量$X\in R^{N\times C\times H\times W}$与$Y\in R^{N\times C\times H\times W}$，则二者相关系数可表示为：

$$CC(X,Y) = \frac{1}{NC}\sum_{n=1}^{N}\sum_{c=1}^{C}\left(\frac{\sum_{l=1}^{L}(X_{n,c,l}-\overline{X_{n,c}})(Y_{n,c,l}-\overline{Y_{n,c}})}{\sqrt{\sum_{l=1}^{L}(X_{n,c,l}-\overline{X_{n,c}})^2}\cdot\sqrt{\sum_{l=1}^{L}(Y_{n,c,l}-\overline{Y_{n,c}})^2}}\right) \tag{20}$$

$$L_{recnVI} = \left\| \boldsymbol{I}_{VI} - \widetilde{\boldsymbol{I}_{VI}} \right\|_2^2 + k_2 SSIM(\boldsymbol{I}_{VI},\widetilde{\boldsymbol{I}_{VI}}) \tag{23}$$

式中，$\|\cdot\|_2$为L2范数，$SSIM(\cdot,\cdot)$[28]为结构相似性函数，k_1与k_2为损失函数权重超参数。

为增强模型对梯度信息的感知能力，本模型通过引入梯度保留损失，减小重构可见光图像与可见光源图像的差异，使得重构的图像能最大程度保留

高频特征，梯度保留损失可表示为：

$$L_{grad}=\left\|\nabla\widetilde{\boldsymbol{I}}_{VI}-\nabla\boldsymbol{I}_{VI}\right\|_1 \tag{24}$$

式中，∇表示梯度算子，$\|\cdot\|_1$表示L1范数。

因此，在特征提取与图像重构阶段，模型总损失函数可表示为：

$$L_{phase-a}=\alpha_1L_{dis}+\alpha_2L_{recn}+\alpha_3L_{grad} \tag{25}$$

式中，α_1、α_2、α_3为损失函数权重超参数。

2.3 特征融合与融合图像生成

在特征融合与融合图像生成阶段，冻结经特征提取与图像重构阶段训练好的全局特征编码器、互补特征编码器及共享特征编码器，引入结构增强模块SEM、共享特征融合模块与互补特征融合模块进行训练，同时微调解码器，以增强模型的特征融合与融合图像生成能力。

2.3.1 全局特征编码与特征解耦编码

在特征融合与融合图像生成阶段，利用在特征提取与图像重构阶段已训练好的全局特征编码器、共享特征编码器与互补特征编码器提取源图像的特征，同时利用MFEM增强红外与可见光互补特征，得到可见光共享特征S_{VI}与互补特征C_{VI}^e，红外共享特征S_{IR}与互补特征C_{IR}^e。

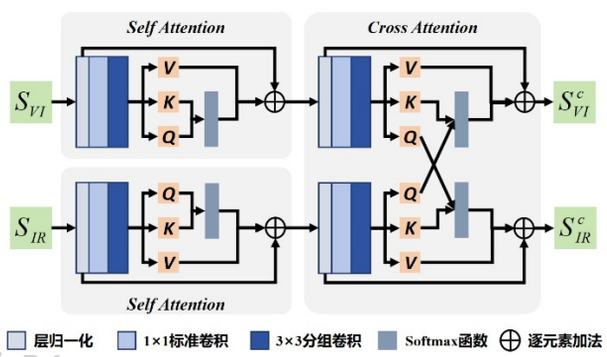

图4 结构增强模块结构示意图

Fig. 4 Architecture diagram of SEM

2.3.2 结构增强模块

针对融合图像中设备目标不显著，结构信息被弱化的问题，本文引入了结构增强模块SEM，其结构示意图如图4所示。

SEM包含两个子模块，首先通过自注意力模块在单模态内强化设备部件的结构关联性，使设备在特征图中呈现完整拓扑结构，计算流程如下：

首先计算源图像互补特征S的查询Q^s，键K^s与值V^s，表示为：

$$\{Q_{IR}^c,K_{VI}^c,V_{VI}^c\}=F_{qkv}^c(S_{VI}^s) \tag{26}$$

式中，$F_{qkv}^s(\cdot)$为自注意力查询键值计算模块，包含一个层归一化层，一个1×1标准卷积层，与一个3×3分组卷积层。

随后，计算自注意力增强后的互补特征，可表示为：

$$S^s=softmax(\frac{Q^sK^s}{d_k})V^s \tag{27}$$

式中，d_k为一个可学习的正标量参数，初始化为1，并随着模型的训练自适应调整，以调节模型输出的特征分布。$softmax(\cdot)$为softmax函数。接着，通过上述计算流程得到自注意力增强后的可见光互补特征S_{VI}^s与红外互补特征S_{IR}^s，并输入至交叉注意力模块实现双模态互补特征在设备结构上的协同增强，从而增强设备结构的显著性，计算流程如下：

首先计算不同模态互补特征的查询、键、值，可表示为：

$$\{Q_{VI}^c,K_{VI}^c,V_{VI}^c\}=F_{qkv}^{c_1}(S_{VI}^s) \tag{28}$$

$$\{Q_{IR}^c,K_{IR}^c,V_{IR}^c\}=F_{qkv}^{c_2}(S_{IR}^s) \tag{29}$$

式中，$F_{qkv}^{c_i}(\cdot)$为交叉注意力查询键值计算模块，与自注意力查询键值计算模块结构相同。

随后，交换两个模态的查询，计算得到交叉注意力增强后的可见光互补特征S_{VI}^c与红外互补特征S_{IR}^c，可表示为：

$$S_{VI}^c=softmax(\frac{Q_{IR}^cK_{VI}^c}{d_k})V_{VI}^c \tag{30}$$

$$S_{IR}^c=softmax(\frac{Q_{VI}^cK_{IR}^c}{d_k})V_{IR}^c \tag{31}$$

2.3.3 特征融合模块

在特征融合与融合图像生成阶段，本文设计了共享特征融合模块与互补特征融合模块，分别融合不同模态的共享特征与互补特征。对于可见光共享特征S_{VI}与红外共享特征S_{IR}，由于其全局大尺度特性，本文采用Restormer块作为共享特征融合模块Γ_S，则融合共享特征S_{FU}可表示为：

$$S_{FU}=\Gamma_S(Conv(Concat(S_{VI}^c,S_{IR}^c)) \tag{32}$$

式中，$Concat(\cdot)$为通道维度拼接，$Conv(\cdot)$为卷积层，用于聚合降维拼接后的可见光共享特征S_{VI}^c与红外共享特征S_{IR}^c。

同理，对于可见光互补特征C_{VI}^e与红外互补特征C_{IR}^e，由于其局部细节特性，本文采用可逆卷积神经网络块作为互补特征融合模块Γ_C，则融合互补特征

C_{FU}可表示为：

$$C_{FU} = \Gamma_C(Conv(Concat(C_{VI}^e, C_{IR}^e)) \tag{33}$$

2.3.4　解码器

在特征融合与融合图像生成阶段，本文运用特征解耦与图像重构阶段已经训练好的解码器，实现融合图像$\boldsymbol{I}_{FU}$的生成，可表示为：

$$\boldsymbol{I}_{FU} = D(S_{FU}, C_{FU}) \tag{34}$$

2.3.5　损失函数

在特征融合与融合图像生成阶段，为尽可能使融合图像保留红外与可见光图像的细节纹理信息，本文分别引入了最大纹理保留损失$L_{texture}$，可表示为：

$$L_{texture} = \left\| \nabla I_{FU} - Max(\nabla I_{VI}, \nabla I_{IR}) \right\|_1 \tag{35}$$

为了更好地融合来自不同区域（设备与场景）以及不同源图像的内容信息，本文引入了内容分布损失。本文首先通过对红外图像灰度图设置阈值，提取设备掩码M_T，将图像分割为两个区域。对于设备区域，内容分布损失定义如下：

$$\begin{aligned} L_{int-target} &= \lambda L_{content}(I_{IR}^T, I_{FU}^T) \\ &+(1-\lambda)L_{content}(I_{VI}^T, I_{FU}^T) \end{aligned} \tag{36}$$

$$I^T = M_T \otimes I \tag{37}$$

$$L_{content}(I_1, I_2) = \left\| I_1 - I_2 \right\|_2 + SSIM(I_1, I_2) \tag{38}$$

式中，ρ是用于控制$L_{content}$中结构相似性损失SSIM Loss分量权重的超参数，$\otimes$为逐元素乘法，$\lambda \in (0,1)$为控制红外图像特征在设备区域内分布的比例因子。类似地，对于场景区域，内容分布损失可表示为：

$$\begin{aligned} L_{int-scene} &= \gamma L_{content}(I_{IR}^{Sc}, I_{FU}^{Sc}) \\ &+(1-\gamma)L_{content}(I_{VI}^{Sc}, I_{FU}^{Sc}) \end{aligned} \tag{39}$$

$$I^{Sc} = (1 - M_T) \otimes I \tag{40}$$

式中，$\gamma \in (0,1)$为控制红外图像特征在场景区域内分布的比例因子。因此，内容分布损失可表示为：

$$L_{int} = L_{int-target} + L_{int-scene} \tag{41}$$

为控制红外图像特征在区域内分布的比例因子综上所述，模型第二阶段训练损失函数可表示为：

$$L_{phase-b} = \beta_1 L_{texture} + \beta_2 L_{int} \tag{42}$$

式中，β_1、β_2为损失函数权重超参数。

3　实验结果

本节详细阐述本文的实验设置，并设计了一系列定量对比实验、消融实验以及可视化分析对本文模型的有效性与合理性进行验证。

3.1　实验设置

3.1.1　变电站红外与可见光图像融合实验设置

本文通过采集变电站数据，构建了包含133对图像的变电站设备红外与可见光图像数据集，并将数据集划分为训练集（103对）与测试集（30对），所有图像分辨率统一为640×480。本文在数据预处理阶段对训练集进行裁剪与筛选，最终生成978组分辨率大小为120×120的红外-可见光匹配图像对。本文模型在训练集上训练，并在测试集上评估性能。

本文采用六项指标定量评估融合图像质量：图像熵（EN），空间频率（SF），互信息（MI），$Q^{AB/F}$（Xydeas等，2000），和结构相似性指数（SSIM）（Wang等，2004），峰值信噪比（PSNR）（Hore等，2010）。其中，信息熵（EN）用于衡量融合图像的信息丰富度，数值越高表明图像包含的细节信息越多；空间频率（SF）反映图像在空间域的整体活跃度，侧重评价融合结果的清晰度与纹理复杂度；互信息（MI）用于度量融合图像与两幅源图像之间的信息共享程度，能够有效反映源图像特征传递的完整性；$Q^{AB/F}$则是一种基于梯度的融合指标，评估融合图像对源图像边缘和细节信息的保持能力；结构相似性（SSIM）从亮度、对比度和结构三个维度衡量融合图像与源图像的相似性，关注融合结果在视觉结构上的保真度；峰值信噪比（PSNR）则从像素级误差角度评价融合图像的噪声水平与重建质量，数值越高表明失真越小。

本文搭载实验平台配置NVIDIA GeForce RTX 4090显卡，数据集图像已预对齐。模型训练分两阶段进行：第一阶段训练全局特征编码器、特征解耦编码器、文本驱动特征增强模块，第二阶段则对融合模块进行训练，并微调第一阶段预训练的模块。两阶段分别训练40与80个epoch，批大小设置为4，采用AdamW优化器，初始学习率为0.0001，每20个epoch衰减50%。

损失函数超参数设置方面，本文设置k_1、k_2为5，α_1、α_2、α_3为2、1、5，ρ为1，β_1、β_2为10、1，λ设置为0.5，γ设置为0.1。

模型架构方面，每个Restormer块包含8个串联的Restormer单元，每个Restormer单元包含8个注意力头，且特征输入输出维度为64；每个INN块包含8个交替连接的INN单元，每个INN单元的输入输出维度均为64。

3.1.2　变电站设备检测实验设置

为验证本文算法在下游任务中的有效性，本文采用可见光图像（VI）、红外图像（IR）以及现有通用图像融合算法和本文算法生成的融合图像作为变电站设备检测数据集。利用YOLOv9（Wang等，2024）对每个数据集进行目标检测实验，并以mAp50作为评价指标评估检测性能。

由于变电站设备数据集的训练样本数量有限，本文首先对原始数据集（103张图像）首先进行划分，其中训练集72张，验证集31张，后利用旋转、裁剪等方法对原始数据集进行数据增强，生成包含412张图像的增强数据集，其中训练集288张，验证集124张。测试集设置为30张图像，与图像融合任务保持一致。表1详细列出了训练集与验证集中目标设备的类型与实例数量。实验参数设置如下：训练轮次（epochs）为250，批量大小（batch size）为8，优化器采用SGD，学习率大小设置为0.01。

表1　训练集与验证集中目标设备的类型与实例数量

Table 1　The types and quantities of target equipment in the training set and validation set

设备类型	标签	训练集实例数量	验证集实例数量
悬式绝缘子	SI	279	170
柱式绝缘子	PI	235	214
电流互感器	CT	319	140
电压互感器	PT	347	124
套管	CA	814	338

3.2　变电站设备检测定量评价

本节定量评估了本文算法在变电站设备检测上的性能，并与源图像（红外IR与可见光VI图像）以及当前通用融合算法（RISFuse（Wang等，2025），FDFuse（Cheng等，2025），CddFuse，CoCoNet，MURF，TarDAL，SDCFusion，RFN-Nest，DenseFuse）进行对比，本文采用AP50（Average Precision at IoU threshold 0.50）作为各类别的检测精度评估指标，并以mAP50（mean Average Precision at IoU threshold 0.50）作为整体检测性能的综合评价依据。如表2所示，最优、次优和第三性能分别用红色粗体、黑色粗体和下划线标出。

由表2可知，本文算法在悬式绝缘子与柱式绝缘子两类小目标设备检测中均取得次优性能；在大目标设备检测中，对电压互感器与套管的检测同样达到次优，在电流互感器上位列第三。尽管基于可见光图像的单模态检测方法及其他融合算法在部分设备类别上略优于本文算法，但其检测鲁棒性普遍不足。例如，可见光图像在电压互感器上的性能远低于本文算法，而基于MURF的融合方法在悬式绝缘子与电压互感器上的表现也明显较差。相比之下，本文算法在平均检测性能上达到最优，表明其在变电站设备检测任务中具有更优的综合鲁棒性。

为深入比较本文算法与单模态及其他融合方法在变电站设备检测中的性能差异，本研究进一步借助混淆矩阵对识别结果进行细粒度分析。如图5所示（图中BG表示背景类别），本文算法在正确检出目标（TP）方面表现优异。同时可见，多数对比算法在悬式绝缘子上存在显著漏检，如红外图像将背景误检为悬式绝缘子的FP值高达0.89，DenseFuse、TarDAL、MURF与CddFuse在该类别上的FP值也均超过0.5，而本文算法将其控制在0.5，一定程度上缓解了漏检问题。此外，在类间混淆方面，DenseFuse、TarDAL和CddFuse在形态相似的电流互感器与电压互感器之间易产生误判，将电流互感器误检为电压互感器的FN值分别为0.10、0.08与0.07，而本文算法仅0.02，显示出更优的类间判别能力。综上所述，本文算法在变电站设备检测任务中表现出更全面的综合性能。

3.3　变电站设备红外与可见光图像融合定量评价

本节定量评估了本文算法在变电站设备红外与可见光数据集上的图像融合性能，并与当前通用算法（RISFuse，FDFuse（Cheng等，2025），CddFuse，CoCoNet（Liu等，2024），MURF（Xu等，2023），TarDAL，SDCFusion（Liu等，2024），RFN-Nest，DenseFuse）进行对比，为保证统计可靠性，本次实验对各算法均重复训练10次，取各指标的平均值与方差。如表3所示，最优、次优和第三优性能分别用红色粗体、黑色粗体和下划线标出。

由表3可知，本文算法在EN上表现最优，表明其融合图像具有最丰富的细节信息；在SF上表现次优，略低于CddFuse算法，表明融合图像的高频特征较为丰富。同时，本文算法在MI和PSNR指标上均表现最优，说明其在保留源图像信息方面优于现有算法。而在$Q^{AB/F}$和SSIM上，本文算法表现次优，表明其能有效保留源图像的显著特征，且生成的融合图像具有较高的视觉保真度。综上所述，本文算法

表2　变电站设备检测性能对比

Table 2　Performance comparison of substation equipment detection

方法	SI	PI	CT	PT	CA	mAP50
IR	0.137	0.439	0.768	0.792	0.796	0.586
VI	**0.588**	0.782	**0.918**	0.827	**0.943**	**0.811**
DenseFuse	0.464	0.741	0.869	**0.892**	0.935	0.780
RFN-Nest	0.201	0.410	0.488	0.887	0.756	0.548
SDCFusion	0.567	0.786	0.821	0.837	0.889	0.780
TarDAL	0.225	0.741	0.888	0.847	0.930	0.726
MURF	0.421	**0.837**	**0.908**	0.843	0.937	0.789
CoCoNet	0.451	0.785	0.900	0.885	0.927	0.790
CddFuse	0.394	0.698	0.896	0.824	0.925	0.747
FDFuse	0.423	0.780	0.877	0.871	**0.941**	0.778
RISFuse	0.473	0.751	0.893	0.865	0.905	0.777
本文	**0.582**	**0.795**	0.901	**0.886**	0.938	**0.821**

有效提升了变电站设备红外与可见光融合图像的综合质量。

3.4　特征解耦性能评价

为验证本文算法在特征解耦方面的有效性，我们对测试集图像解耦后的特征进行了可视化处理，结果如图6所示。由图6可知，可见光共享特征S_{VI}与红外共享特征S_{IR}呈现出高度的一致性，共同捕捉了跨模态共享的大尺度结构信息，如设备与场景之间的位置关系以及设备的拓扑结构。在互补特征方面，可见光互补特征C_{VI}主要包含丰富的细节纹理信息，例如设备内部的复杂结构纹理；而红外互补特征C_{IR}则通过高亮区域凸显了设备的温度分布情况。这一结果与本文对解耦后特征语义的分析相符。同时，可见光共享特征S_{VI}与红外共享特征S_{IR}的相关系数接近1，而可见光互补特征C_{VI}与红外互补特征C_{IR}的相关系数趋近于0，进一步验证了特征解耦方法的有效性。

3.5　消融实验

为验证本文方法的合理性，本节设计了多组消融实验。

3.5.1　多尺度特征增强模块

本节定量评估了多尺度特征融合模块（MFEM）在图像融合与设备检测任务上的有效性，实验结果分别如表4、表5所示，最优性能用红色粗体标出。

由表4可知，引入MFEM后，模型在EN和SF指标上均有提升，表明MFEM能有效增强融合图像的细节及高频特征表现。同时，模型在MI与$Q^{AB/F}$指标上也获得一定提升，说明MFEM增强了模型对源图像信息的感知能力，从而提高了融合图像的质量。

由表5可知，引入多尺度特征增强模块（MFEM）后，模型在悬式绝缘子、柱式绝缘子等小目标设备的

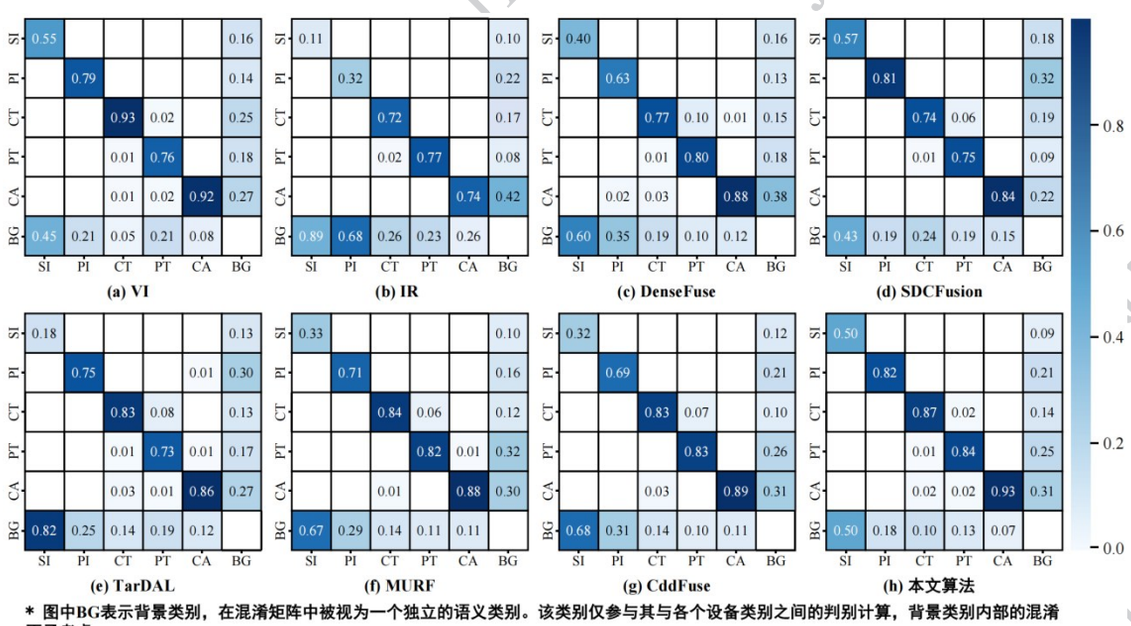

图5　变电站设备检测混淆矩阵对比

Fig. 5　Comparison of confusion matrix for power substation equipment detection

表3　变电站设备红外与可见光图像融合的性能对比

Table 3　Performance comparison of infrared and visible image fusion for substation equipment

方法	EN(bit)	SF	MI(bit)	$Q^{AB/F}$	SSIM	PSNR(dB)
DenseFuse	6.391±0.031	7.393±0,094	1.727±0.115	0.229±0.009	0.914±0.017	2.811±0.049
RFN-Nest	6.271±0.027	6.331±0.119	1.436±0.147	0.157±0.007	0.745±0.010	2.764±0.021
SDCFusion	6.730±0.036	7.208±0.107	2.001±0.129	0.374±0.012	1.029±0.016	2.805j±0.048
TarDAL	7.127±0.017	7.088±0.193	2.132±0.176	0.275±0.011	0.938±0.011	2.828±0.037
MURF	6.803±0.029	7.470±0.097	**2.337　0.185**	0.381±0.011	1.062±0.012	2.794±0.024
CoCoNet	7.079±0.021	7.603±0.074	2.034±0.157	**0.442　0.010**	1.080±0.019	2.810±0.053
CddFuse	7.102±0.016	**7.917　0,072**	2.110±0.106	0.415±0.008	**1.094　0.013**	**2.861　0.044**
FDFuse	7.086±0.014	7.432±0.069	1.906±0.133	0.371±0.009	1.058j±0.012	**2.816　0.061**
RISFuse	7.081±0.019	7.131±0.065	2.315±0.113	0.403±0.008	1.084±0.011	**2.815　0.026**
本文	**7.158　0,019**	**7.903　0.094**	**2.482　0.107**	**0.425　0.007**	**1.087　0.010**	**2.874　0.029**

检测精度上取得了显著提升；在电压互感器与套管类别上检测性能达到最佳，而在电流互感器上虽略有下降，仍保持了良好的检测能力。总体而言，MFEM的引入使得模型在整体检测性能上实现较大幅度提高，有效验证了该模块在电力设备检测任务中的有效性。

同时，本文对MFEM中空洞卷积的扩张率进行了实验分析。如表4所示，当扩张率设置为5时（即本文采用设定），模型在EN、SF和MI三项指标上均表现最优，并在$Q^{AB/F}$上表现次优。这主要是因为扩张率为5能够匹配变电站目标的空间尺度，获得适当的感受野，从而在增强互补特征时既保留了丰富的细节信息与边缘结构，又有效避免了特征丢失与混叠，因此与细节、边缘保留相关的EN、SF、MI、$Q^{AB/F}$等融合指标均得到提升。当扩张率为3时，感受野过小，所能捕获的上下文信息有限；而当扩张率增大至7时，采样间隔过大，导致特征图部分区域信息缺失并产生混叠干扰，进而造成各项指标下降。进一步地，表5所示的变电站设备检测实验结果表明，扩张率为5的设定在悬式绝缘子、柱式绝缘子、电流互感器和电压互感器四类设备上均优于扩张率为3和7的对比模型，在套管检测任务中也表现出具有竞争力的性能。这得益于扩张率为5时，特征增强不仅能获取更丰富的上下文信息，还能避免因上下文区域混叠而造成的特征间相互干扰，从而使检测模型能够更好地区分目标与背景。综合来看，该设定下的模型整体检测性能最佳。上述结果一致验证了将扩张率设置为5的合理性与有效性。3.5.2 结构增强模块

本节定量评估了结构增强模块SEM在图像融合任务与设备检测任务上的有效性，并分别探究了自注意力与交叉注意力对模型性能的影响。实验结果如表6、表7所示，最优、次优性能分别用红色粗体与黑色粗体标出。

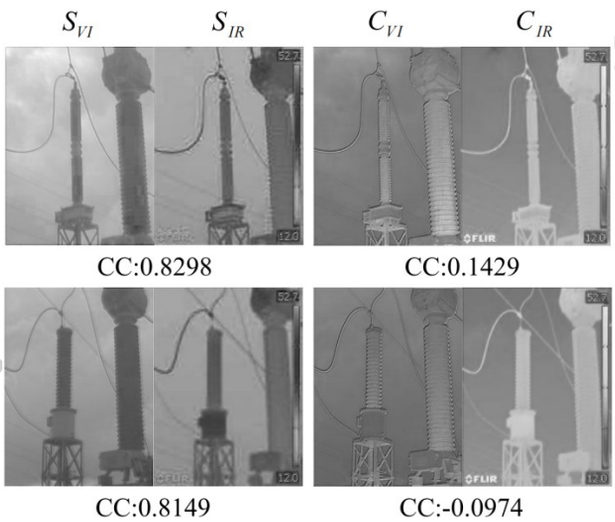

图6　解耦特征可视化效果

Fig. 6　Visualization of the decomposed features

由表6可知，分别引入自注意力SA和交叉注意力CA后，模型在EN、SF、MI与$Q^{AB/F}$四项指标上均有所提升，验证了两种注意力机制的有效性。而在同时引入自注意力与交叉注意力（即SEM）后，模型在上述所有指标上均达到最优性能。这表明SEM不

表4　面向图像融合的多尺度特征融合模块消融实验

Table 4　Ablation study of the MFEM for image fusion

方法	EN(bit)	SF	MI(bit)	$Q^{AB/F}$
w/o MFEM	7.141	7.861	2.389	0.421
扩张率=3	7.153	7.873	**2.414**	0.427
扩张率=5	**7.160**	**7.889**	**2.479**	**0.426**
扩张率=7	**7.157**	**7.883**	2.401	0.425

表5　面向变电站设备检测的多尺度特征融合模块消融实验

Table 5　Ablation study of the MFEM for substation equipment detection

方法	SI	PI	CT	PT	CA	mAP50
w/o MFEM	0.552	0.786	0.925	0.837	0.931	0.806
扩张率=3	0.563	**0.791**	0.900	0.858	0.940	0.810
扩张率=5	0.582	0.795	**0.901**	0.886	**0.938**	0.821
扩张率=7	**0.578**	**0.791**	0.898	**0.863**	0.936	**0.813**

仅能够增强图像自身的细节与高频特征表征，还能促进模态间的有效交互，确保融合图像充分保留来自双源图像的关键信息。最终结果证明了自注意力与交叉注意力具有显著的协同效应，进一步验证了SEM的有效性。

表6　面向图像融合的结构增强模块消融实验

Table 6　Ablation study of the SEM for image fusion

方法	EN(bit)	SF	MI(bit)	$Q^{AB/F}$
w/o SEM	7.151	7.873	2.430	0.420
w/o SA	**7.157**	7.875	**2.467**	**0.424**
w/o CA	7.155	**7.883**	2.455	0.423
本文	**7.160**	**7.889**	**2.479**	**0.426**

表7　面向变电站设备检测的结构增强模块消融实验

Table 7　Ablation study of the SEM for substation equipment detection

方法	SI	PI	CT	PT	CA	mAP50
w/o SEM	0.503	0.841	0.880	0.805	0.919	0.789
w/o SA	**0.554**	0.784	**0.901**	**0.896**	0.923	**0.812**
w/o CA	0.513	**0.803**	**0.882**	**0.907**	**0.924**	0.806
本文	**0.582**	0.795	**0.901**	0.886	**0.938**	**0.821**

由表7可知，未引入结构增强模块SEM的模型在柱式绝缘子类别上检测性能最优，而在分别引入交叉注意力CA与自注意力SA机制后，模型在多数设备类别以及整体检测性能上均获得显著提升。同时，引入SEM的模型在悬式绝缘子、电流互感器和套管类别上均达到最佳检测效果，在柱式绝缘子与电压互感器上也保持了具有竞争力的性能。总体而言，本文模型在综合性能上优于其他对比模型，充分证明了SEM在设备检测任务中的有效性。

为进一步验证结构增强模块SEM的可解释性，本文分别对包含与不包含该模块的模型输出特征进行了可视化处理：通过对输出特征在通道维度上计算平均值，生成了特征注意力热力图，结果如图7所示。

由图7可知，当模型未引入SEM时，红外与可见光图像的特征注意力图虽能大致反映设备的结构信息，但响应区域较为模糊，显著性不足；而在加入SEM后，设备区域及其结构特征在热力图中的激活响应显著增强，表现更为突出。这一对比结果充分说明，SEM能够有效提升模型对图像中设备结构信息的聚焦与凸显能力。

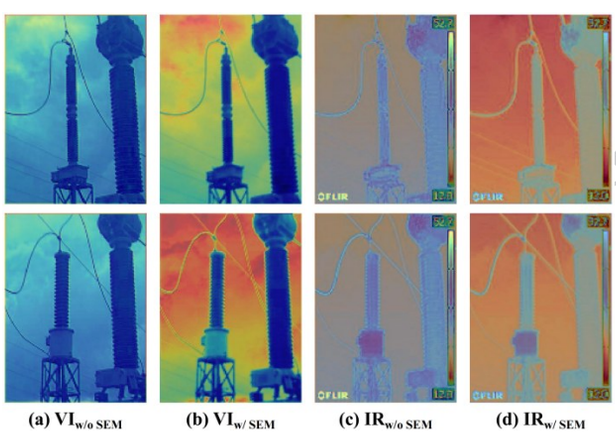

图7　有无结构增强模块的特征注意力热力图对比

Fig. 7　Comparison of feature attention heatmaps with and without the SEM

3.5.3　骨干网络

本节对共享特征编码器与互补特征编码器中的骨干网络进行了定量分析。值得注意的是，本文采用的Swin Transformer基于SwinFuse（Wang等，2022）提出的残差Swin Transformer模块（Residual Swin Transformer Block）。实验结果如表8所示，最优、次优和第三优性能分别用红色粗体、黑色粗体与下划线标出。

由表8可知，在共享特征编码方面，ResNet-18

和INN凭借其局部细节感知能力，在EN指标上表现优异；而Swin Transformer作为ViT架构的一种，在PSNR方面取得了最佳性能。相比之下，采用Restormer作为共享特征编码器的模型则在SF、MI、$Q^{AB/F}$和SSIM等多个指标上均表现最优，同时在EN和PSNR方面也位居第三，展现出全面而竞争力的性能，充分验证了Restormer在共享特征提取任务中的有效性。在互补特征编码方面，使用INN作为编码器的模型在EN与SF指标上显著优于CNN与ResNet-18，表明INN在提取图像细节与高频特征方面具有优势；同时，本文提出的模型在MI、$Q^{AB/F}$与SSIM指标上均达到最优，PSNR指标次优，展现出优异的综合性能。这一结果既验证了INN在促进融合图像充分保留双源关键信息方面的有效性，又体现了Restormer作为共享编码器与INN作为互补编码器的协同作用，共同提升了模型效能，进一步证明了二者组合架构的有效性。

表8　不同骨干网络的消融结果对比

Table 8　Ablation study on different backbone networks

方法		EN(bit)	SF	MI(bit)	$Q^{AB/F}$	SSIM	PSNR(dB)
共享特征编码	CNN	7.143	7.101	2.209	0.401	1.024	2.825
	ResNest-18	7.264	7.219	2.013	0.353	0.913	2.817
	INN	**7.171**	7.307	2.146	0.379	1.037	2.810
	Swin Transformer	7.142	6.884	2.193	0.396	**1.061**	2.889
互补特征编码	CNN	7.146	**7.692**	2.310	**0.415**	1.043	2.829
	ResNet18	7.149	7.437	**2.327**	0.412	1.052	**2.884**
本文		7.160	7.889	2.479	0.426	1.087	2.869

3.5.4　损失函数

本节对本文模型在特征融合与融合图像生成阶段中不同损失函数，即$L_{content}$中的L2范数分量（L2 norm）与SSIM分量以及$L_{texture}$对于模型（融合图像）性能的影响进行了定量分析，实验结果如表7所示，最优与次优性能分别用红色粗体、黑色粗体标出。

由表9可知，本文模型（在特征融合与图像生成阶段引入L2范数分量、SSIM分量及$L_{texture}$损失项）在SSIM指标上达到最优，EN与SF指标表现次优，PSNR指标也接近最优水平，验证了模型损失函数设计的有效性。当损失函数中移除L2范数分量时，模型虽在EN指标上表现最优，但SSIM值显著下降，表明融合图像损失了源图像的结构、亮度及对比度信息；移除SSIM分量时，模型虽在SF指标上最优，但SSIM同样大幅降低；而移除$L_{texture}$时，模型虽获得最优PSNR值，但EN与SF指标急剧下滑，说明生成图像丢失了大量高频细节特征。上述结果共同佐证了损失函数中保留三项分量的必要性，充分体现了当前损失函数设置的合理性。

3.6　方法计算效率评价

本节定量比较了本文算法与现有算法的计算效率。由表10可知，本文方法在推理时间和浮点运算量上高于部分已有方法，参数量处于中等水平。这主要归因于引入了多分支特征增强模块以及由多头自注意力和交叉注意力构成的结构增强模块，在一定程度上增加了模型的计算开销与参数量，延长了推理时间，但也换来了更优的融合质量与下游任务性能。尽管当前模型在计算效率上并非最优，但其参数量相对较小，具备进一步轻量化的潜力。因此，如何在保持模型性能的前提下，实现模型轻量化与计算加速，以满足边缘设备的实时部署需求，将是未来研究的重点方向。

3.7　变电站设备检测定性评价

本节对本文算法、源图像及通用图像融合算法在变电站设备检测任务上的性能进行了定性分析。如图8所示，第一行和第二行的结果表明，由于变电站设备与背景的复杂性，在检测悬式绝缘子、柱式绝缘子等小目标设备时，易出现误检和漏检。例如，第一行展示的悬式绝缘子检测中，现有算法及源图像均忽略了被套管部分遮挡的悬式绝缘子，而CddFuse算法则误将电线杆支架识别为悬式绝缘子；第二行柱式绝缘子的检测结果显示，现有算法及

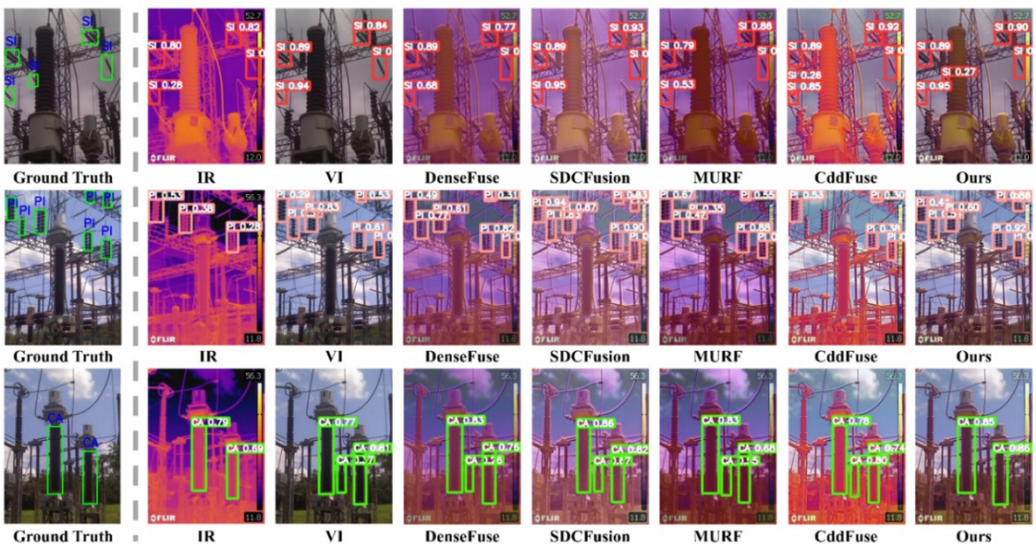

图8　变电站设备检测可视化对比

Fig. 8　Visual comparison of substation equipment detectio

表9　不同损失函数的消融结果评估

Table 9　Ablation study on different loss functions

方法	EN(bit)	SF	PSNR(dB)	SSIM
w/o L2 norm	7.163	7.819	**2.890**	0.982
w/o SSIM	7.154	7.896	2.837	0.835
w/o $L_{texture}$	7.069	7.607	2.892	**1.081**
本文	**7.160**	**7.889**	2.869	1.087

源图像均出现漏检，而本文算法则能准确识别图中所有目标。

表10　不同方法的计算效率对比

Table 1　Comparison of computational efficiency of different methods

方法	推理时间(ms)	浮点计算量(G)	参数量(M)
DenseFuse	251.10	48.92	0.074
RFN-Nest	178.8	1346.34	7.524
SDCFusion	110.23	112.87	0.570
TarDAL	16.57	233.39	0.297
MURF	565.55	186.21	0.117
CoCoNet	29.06	498.58	9.114
CddFuse	463.8	1402.07	1.186
FDFuse	177.35	65.50	0.578
RISFuse	560.53	430.10	0.476
本文	429.43	1764.8	1.715

第三行的结果表明，在检测套管等易混淆设备时，现有算法与可见光源图像均将电线杆体误检为套管，红外源图像虽能正确检测但置信度较低；相比之下，本文算法在保证准确率的同时，对正确检测结果保持了较高的置信度。上述结果验证了本文算法在设备检测任务中的有效性。

3.8　变电站设备红外与可见光图像融合定性评价

本节对本文算法与通用图像融合算法在变电站设备红外与可见光图像融合任务上的性能进行了定性分析。如图9所示，第一行和第二行的结果表明，现有对比算法生成的融合图像中场景部分的红外特征过于突出，导致设备结构在植被背景下难以辨识；相比之下，本文算法能有效区分设备与环境，所得融合图像中设备结构清晰、纹理细节丰富。第三行的结果进一步显示，本文算法融合图像中设备与楼房等背景物体的轮廓界限分明，设备结构清晰可见，显著提升了目标设备在图像中的显著性，满足了变电站设备图像融合的核心需求。此外，本文算法所得融合图像在色彩呈现上也更为自然，视觉效果更佳。

3.9　泛化性能评价

为充分分析本文算法的泛化能力，本节在公共数据集M3FD（训练集3900张，测试集300张）上比较了不同算法的性能，如表11所示，本文算法在M3FD数据集多个指标上表现良好，其中SSIM与PSNR表现最佳，SF与$Q^{AB/F}$上表现次佳，表明该算法

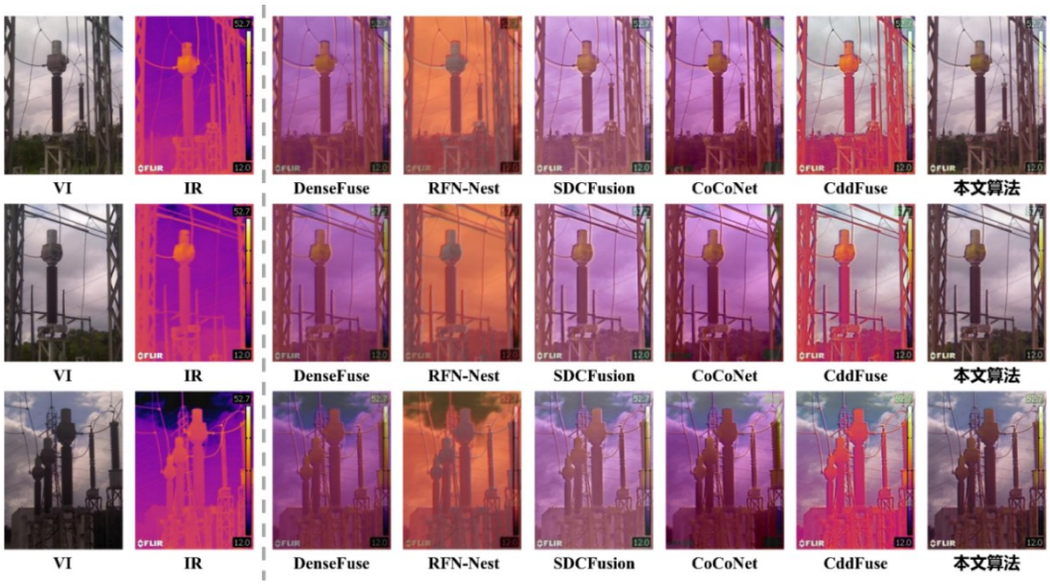

图9　变电站设备图像融合可视化对比

Fig. 9　Visualization Comparison of Image Fusion for Substation Equipment

不仅在变电站场景中具有卓越性能，迁移至通用场景时仍能展现出较强的泛化能力。然而，在EN与MI指标上，本文算法仅位列第三，与现有最优方法相比仍存在一定差距，说明其在有效保留源图像细节信息方面仍有较大提升空间。因此，如何设计一种能够自适应不同领域与场景的融合算法，使其在变电站、通用及其他领域中均取得优异性能，将是后续研究的重点突破方向。

此外，本文进一步对比了不同算法在M3FD数据集上的融合图像可视化结果。如图10所示，本文算法所得融合图像中的目标（如行人、车辆）更为突出，表明其能够有效提升融合图像中目标的显著性。同时，该算法融合图像中的背景（如楼房、树木）在色彩上更为自然，进一步体现出其生成更佳视觉效果融合图像的能力。

图10　M3FD数据集图像融合可视化对比

Fig. 10　Visual comparison of image fusion on M3FD Dataset

4　结　论

本文提出一种面向变电站设备检测的双分支感知增强红外与可见光图像融合算法，有效提升了融合图像中设备的拓扑完整性与特征显著性。具体而言，首先通过由共享分支与互补分支构成的双路编码架构，实现设备共性结构特征与细节纹理、热特征的有效解耦；其次，针对设备共性结构特征，设计结构增强模块（structure enhancement module，SEM）以强化设备轮廓；针对设备细节纹理与热特征，引入多分支特征增强模块（multibranch feature enhancement module，MFEM）以强化其对设备外观及温度等关键信息的表达；最后，通过共享-互补特征融合模块实现多模态特征有效整合，经由解码器重构出结构清晰、细节丰富的高质量融合图像。实验结果表明，本

表11　M³FD数据集红外与可见光图像融合的性能对比

Table 11　Performance comparison of infrared and visible image fusion on M³FD dataset

方法	EN (bit)	SF	MI (bit)	$Q^{AB/F}$	SSIM	PSNR (dB)
DenseFuse	6.431	7.348	2.907	0.397	0.754	61.368
RFN-Nest	6.790	10.774	2.627	0.534	0.801	62.220
SDCFusion	6.682	12.010	3.019	0.517	0.856	61.491
TarDAL	7.225	12.153	3.452	0.430	0.892	62.743
MURF	**7.013**	11.981	3.327	0.558	0.907	62.467
CoCoNet	6.658	13.830	3.042	0.557	0.943	**62.864**
CddFuse	6.955	16.542	4.532	0.647	**1.030**	60.838
FDFuse	6.913	15.236	4.311	0.609	1.001	59.725
RISFuse	6.772	15.777	**4.499**	0.694	0.998	61.317
本文	7.006	**16.326**	4.445	**0.663**	1.031	63.128

文算法在红外与可见光图像融合及变电站设备检测任务中均表现出卓越性能，有效解决了现有融合方法在变电站场景下面临的关键问题，即目标设备显著性不足、结构信息弱化、与背景界限模糊，导致检测模型难以获取判别性特征而制约检测精度与鲁棒性的问题。

值得注意的是，本算法在设备检测任务中虽整体性能优于现有方法，但在特定单体设备的检测精度上尚未达到最优，表明其仍有提升空间。未来研究可探索融入变电站设备领域相关知识，协同优化图像融合与目标检测任务，以进一步提升针对不同电力设备的检测性能。

参考文献（References）

Qiang W, Zhang X D, Yang G, et al. 2022. Substation power equipment monitoring system based on infrared detection. In: Proceedings of 2022 International Conference on Electronics and Devices, Computational Science (ICEDCS). IEEE: 294-298

Wu T, et al. 2024. ISE-YOLO: A Real-Time Infrared Detection Model for Substation Equipment. IEEE Transactions on Power Delivery, vol. 39, no. 4, pp. 2378-2387. [DOI: 10.1109/TPWRD.2024.3404621]

Xu L, Song Y K, Zhang W S, An Y Y, Wang Y and Ning H S. 2021. An efficient foreign objects detection network for power substation. Image and Vision Computing 109: 104159 [DOI: 10.1016/j.imavis.2021.104159]

Wang Y L, Feng T B, Sun N, Yang C, Yu H W and Cui H Y. 2024. Power insulator defect detection method integrating attention and multi-scale features. High Voltage Engineering, 50(5): 1933-1942 (王韵琳, 冯天波, 孙宁, 杨程, 余恒文, 崔昊杨. 2024. 融合注意力与多尺度特征的电力绝缘子缺陷检测方法. 高电压技术, 50(5): 1933-1942)

Zhu H L, Niu Z W, Huang K C and Tang W H. 2021. Target recognition and location of substation equipment in infrared images based on single-stage object detection algorithm. Electric Power Automation Equipment, 41(8): 217 (朱惠玲, 牛哲文, 黄克灿, 唐文虎. 2021. 基于单阶段目标检测算法的变电设备红外图像目标识别及定位. 电力自动化设备, 41(8): 217 [DOI: 10.16081/j.epae.202104015]

Li J, Xu Y, Nie K, Cao B, Zuo S and Zhu J. 2023. PEDNet: a lightweight detection network of power equipment in infrared image based on YOLOv4-Tiny. IEEE Transactions on Instrumentation and Measurement 72: 1-12 [DOI: 10.1109/TIM.2023.3235416]

Wu H, Jia D H, Zhang T T, Bai X J, Sun L and Pu M Y. 2025. Multimodal zero-shot anomaly detection using dual-experts for electri-cal power equipment inspection images. Journal of Image and Graphics, 30(3): 0672-0682 (吴华, 贾栋豪, 张婷婷, 白晓静, 孙笠, 蒲梦杨. 2025. 基于双专家的巡检影像多模态零样本缺陷检测. 中国图象图形学报, 30(3): 0672-0682) [DOI: 10. 11834/jig.240246]

Xu G, Deng X, Zhou X, Pedersen M, Cimmino L and Wang H. 2022. FCFusion: fractal componentwise modeling with group sparsity for medical image fusion. IEEE Transactions on Industrial Informatics 18(12): 9141-9150 [DOI: 10.1109/TII.2022.3185050]

Zhao F, Zhang C, and Geng B. 2024. Deep multimodal data fusion. ACM Computing Surveys 56(9): Article 216, 1-36 [DOI: 10.1145/3649447]

Liu Y, Liu S, and Wang Z. 2015. A general framework for image fusion based on multi-scale transform and sparse representation. Information Fusion 24: 147-164. [DOI: 10.1016/j.inffus.2014.09.004]

Liu Y, Chen X, Ward R K, et al. 2016. Image fusion with convolutional sparse representation. IEEE Signal Processing Letters 23(12): 1882-1886. [DOI: 10.1109/LSP.2016.2618779]

Ma J, Chen C, Li C, and Huang J. 2016. Infrared and visible image fusion via gradient transfer and total variation minimization. Information Fusion 31: 100-109. [DOI: 10.1016/j.inffus.2016.02.001]

Shao Z and Cai J. 2018. Remote sensing image fusion with deep convolutional neural network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11(5): 1656-1669 [DOI: 10.1109/JSTARS.2018.2805923]

Girshick R., Donahue J., Darrell T. and Malik J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 580-587. [DOI: 10.1109/CVPR.2014.81]

Girshick R., 2015. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448. [DOI: 10.1109/ICCV.2015.169]

Ren S., He K., 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), pp.1137-1149. [DOI: 10.1109/TPAMI.2016.2577031]

Wang C. Y., Yeh, I. H. Liao, H. Y., 2025. YOLOv9: Learning what you want to learn using programmable gradient information. In *Computer Vision - ECCV 2024* (Lecture Notes in Computer Science, vol 15089), pp.1-20. Springer. [DOI: 10.1007/978-3-031-72751-1_1]

Bai Y F, Wang L B, Gao W D and Ma Y L. 2024. Multi-modal hierarchical classification for power equipment defect detection. Jour-nal 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]

Ma J Y, Yu W, Liang P W, Li C and Jiang J J. 2019. FusionGAN: a generative adversarial network for infrared and visible image fusion. Information Fusion 48: 11-26 [DOI: 10.1016/j. inffus. 2018. 09.004]

Ma J, Xu H, Jiang J, Mei X and Zhang X P. 2020. DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing 29: 4980-4995 [DOI: 10.1109/TIP.2020.2977573]

Liu J, Zhang Y, Wang J, Li X and Chen Z. 2022. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 5792-5801 [DOI: 10.1109/CVPR52688.2022.00571]

Li H and Wu X J. 2019. DenseFuse: a fusion approach to infrared and visible images. IEEE Transactions on Image Processing 28 (5): 2614-2623 [DOI: 10.1109/TIP.2018.2887342]

Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]

Li H, Wu X J and Kittler J. 2021. RFN-Nest: an end-to-end residual fusion network for infrared and visible images. Information Fusion 73: 72-86 [DOI: 10.1016/j.inffus.2021.02.023]

Li H., Cen Y., Liu Y., Chen X. and Yu, Z., 2021. Different input resolutions and arbitrary output resolution: A meta learning-based deep framework for infrared and visible image fusion. IEEE Transactions on Image Processing, 30, pp. 4070-4083. [DOI: 10.1109/TIP.2021.3069339]

Tang L., Xiang X., Zhang H., Gong M. and Ma J., 2023. DIVFusion: Darkness-free infrared and visible image fusion. Information Fusion, 91, pp.477-493. [DOI: 10.1016/j.inffus.2022.09.012]

Dosovitskiy A., Beyer L., Kolesnikov A., et al., 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations (ICLR), pp.1-21. [DOI: 10.48550/arXiv.2010.11929]

Zhao Z, Zhang Z, Xu S, Zhang C and Wu X J. 2023. CDDFuse: correlation-driven dual-branch feature decomposition for multi-modality image fusion//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 5906-5916 [DOI: 10.1109/CVPR52729.2023.00572]

Yang C, Wang Y, Zhang X, Zhang H, Wei Z, Lin Y and Xie W. 2022. Lite vision transformer with enhanced self-attention//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 11988-11998 [DOI: 10.1109/CVPR52688.2022.01169]

Gomez A N, Ren M, Urtasun R and Grosse R B. 2017. The reversible residual network: backpropagation without storing activations//Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach: NeurIPS: 2214-2224

Liu J, Zhang B, Mei Q, Li X, Zou Y, Jiang Z, Ma L, Liu R, Fan X. 2025. DCEvo: Discriminative Cross-Dimensional Evolutionary Learning for Infrared and Visible Image Fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 1 - 10 [DOI: 10.1109/CVPR52734.2025.00213]

Bai H, Zhang J, Zhao Z, Wu Y, Deng L, Cui Y, Feng T and Xu S. 2025. Task-driven Image Fusion with Learnable Fusion Loss. In * 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA: 7457-7468 [DOI: 10.1109/CVPR52734.2025.00699]

Wang Z, Zhang J, Song H, Ge M, Wang J and Duan H. 2025. Highlight What You Want: Weakly-Supervised Instance-Level Controllable Infrared-Visible Image Fusion. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV): 12637-12647 [DOI: 10.1109/ICCV51070.2025.01234]

Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G and Sutskever I. 2021. Learning transferable visual models from natural language supervision//Proceedings of the 38th International Conference on Machine Learning. PMLR 139: 8748-8763

Kim G, Kwon T and Ye J C. 2022. DiffusionCLIP: text-guided diffusion models for robust image manipulation//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 2416-2425 [DOI: 10.1109/CVPR52688.2022.00246]

Zhang B, Zhang P, Dong X, Zang Y and Wang J. 2024. Long-CLIP: unlocking the long-text capability of CLIP//Proceedings of the 17th European Conference on Computer Vision. Cham: Springer: 1-15 [DOI: 10.1007/978-3-031-72983-6_18]

Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M. 2022.

Restormer: efficient transformer for high-resolution image restoration//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 5718-5729 [DOI: 10.1109/CVPR52688.2022.00564]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16x16 words: transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. [DOI: 10.48550/arXiv.2010.11929]

Xydeas C S and Petrović V. 2000. Objective image fusion performance measure. Electronics Letters 36(4): 308-309 [DOI: 10.1049/el: 20000267]

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]

Horé A and Ziou D. 2010. Image quality metrics: PSNR vs. SSIM//Proceedings of the 20th International Conference on Pattern Recognition. Istanbul: IEEE: 2366-2369 [DOI: 10.1109/ICPR.2010.579]

Wang C Y, Yeh I H and Liao H Y M. 2024. YOLOv9: learning what you want to learn using programmable gradient information//Proceedings of the 18th European Conference on Computer Vision. Cham: Springer: 1-15 [DOI: 10.1007/978-3-031-72751-1_1]

Cheng M, Huang H, Liu X, Mo H, Wu S, Zhao X. 2025. FDFuse: Infrared and Visible Image Fusion Based on Feature Decomposition. IEEE Transactions on Instrumentation and Measurement 74: 1 - 13 [DOI: 10.1109/TIM.2025.3551460]

Liu J, Lin R, Wu G, Zhang Y, Li X and Wang B. 2024. CoCoNet: coupled contrastive learning network with multi-level feature ensemble for multi-modality image fusion. International Journal of Computer Vision 132(5): 1748-1775 [DOI: 10.1007/s11263-023-01952-1]

Xu H, Yuan J and Ma J. 2023. MURF: mutually reinforcing multi-modal image registration and fusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 45(10): 12148-12166 [DOI: 10.1109/TPAMI.2023.3283682]

Liu X W, Huo H T, Li J, Pang S and Zheng B W. 2024. A semantic-driven coupled network for infrared and visible image fusion. Information Fusion 105: 102352 [DOI: 10.1016/j.inffus.2024.102352]

Wang Z, Chen Y, Shao W, Li H and Zhang L. 2022. SwinFuse: A Residual Swin Transformer Fusion Network for Infrared and Visible Images. IEEE Transactions on Instrumentation and Measurement 71: 1-12 [DOI: 10.1109/TIM.2022.3191664]

作者简介

郭玉荣,女,讲师,主要研究方向为计算机视觉、电力计算机视觉和电力人工智能。Email: guoyurong@ncepu.edu.cn

张珂,通信作者,男,教授,主要研究方向为计算机视觉、电力计算机视觉和电力人工智能。Email: zhangkeit@ncepu.edu.cn

何雨非,男,硕士研究生,主要研究方向为计算机视觉和电力计算机视觉。Email: 220232215006@ncepu.edu.cn

张铁峰,男,副教授,主要研究方向为配电网规划与运行分析、新能源智慧化运维及决策支持。Email: zhangtiefeng@ncepu.edu.cn。

杨宏,男,讲师,主要研究方向风光功率预测和调度。Email: yanghong@ncepu.edu.cn