智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

面向复杂电力环境场景理解的可见光和红外图像特征级融合方法

黄志鸿, 杜瑞, 张辉

引用本文:
黄志鸿, 杜瑞, 张辉. 面向复杂电力环境场景理解的可见光和红外图像特征级融合方法[J]. 智能系统学报, 2025, 20(3): 631-640.
HUANG Zhihong, DU Rui, ZHANG Hui. Feature-level fusion method of visible and infrared images for scene understanding in complex power environments[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(3): 631-640.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202404014

DOI: 10.11992/tis.202404014
网络出版地址: https://link.cnki.net/urlid/23.1538.tp.20250418.0955.002

面向复杂电力环境场景理解的可见光和红外图像特征级融合方法

黄志鸿[1,2], 杜瑞[3], 张辉[3]

（1. 国网湖南省电力有限公司 电力科学研究院, 湖南 长沙 410017; 2. 湖南省湘电试验研究院有限公司, 湖南 长沙 410017; 3. 湖南大学 机器人视觉感知与控制技术国家工程研究中心, 湖南 长沙 410082）

摘　要: 随着电力系统自动化和智能化程度的不断提高, 变电站和配电网设备的有效监测与故障诊断成为保证电网稳定运行的重要手段。针对传统单模态图像处理方法在复杂电力环境中面临的挑战, 本文提出了一种基于可见光和红外图像特征级融合的场景理解方法。通过深入分析可见光图像和红外图像的互补特性, 设计了一个双分支的对称融合网络框架, 有效结合了可见光图像的高分辨率纹理信息和红外图像的温度信息。此外, 引入多尺度特征融合层和多尺度注意力解码器, 以提高模型的分割精度和细节恢复能力。实验结果表明, 该方法在变电站设备监测中取得了优异的性能, 尤其是在处理光照不足和遮挡情况下的图像时, 展现出了较好的鲁棒性。该研究不仅为复杂电力环境的监测提供了一种有效的技术手段, 而且对于推动电力系统智能化管理具有重要的理论和实践意义。

关键词: 特征级融合; 场景理解; 电力系统监测; 变电站设备; 智能化电网; 多模态融合; 图像语义分割; 红外可见光图像
中图分类号: TP391.4　　**文献标志码:** A　　**文章编号:** 1673−4785(2025)03−0631−10

中文引用格式: 黄志鸿, 杜瑞, 张辉. 面向复杂电力环境场景理解的可见光和红外图像特征级融合方法 [J]. 智能系统学报, 2025, 20(3): 631–640.
英文引用格式: HUANG Zhihong, DU Rui, ZHANG Hui. Feature-level fusion method of visible and infrared images for scene understanding in complex power environments[J]. CAAI transactions on intelligent systems, 2025, 20(3): 631–640.

Feature-level fusion method of visible and infrared images for scene understanding in complex power environments

HUANG Zhihong[1,2], DU Rui[3], ZHANG Hui[3]

(1. Electric Power Research Institute, State Grid Hu'nan Electric Power Company Limited, Changsha 410017, China; 2. Hu'nan Xiangdian Test and Research Institute Co., Ltd., Changsha 410017, China; 3. Engineering Research Center for Robot Visual Perception and Control Technology, Hu'nan University, Changsha 410082, China)

Abstract: With the continuous increase in the automation and intelligence levels of power systems, the effective monitoring and fault diagnosis of substation and distribution network equipment have become crucial to ensuring stable grid operation. To address the challenges faced by traditional single-modal image processing methods in complex power environments, a scene understanding method based on the feature-level fusion of visible and infrared images is proposed here. By deeply analyzing the complementary characteristics of visible and infrared images, a dual-branch symmetric fusion network framework is designed, and it effectively integrates the high-resolution texture information of visible images with the temperature information of infrared images. Furthermore, multi-scale feature fusion layers and multi-scale attention decoders are introduced to enhance the segmentation precision and detail recovery capabilities of the model. The experimental results reveal that this method performs excellently in substation equipment monitoring, particularly demonstrating good robustness in processing images under insufficient lighting and occlusion conditions. This research presents an effective technical approach for monitoring complex power environments and offers significant theoretical and practical implications for advancing intelligent management in power systems.

Keywords: feature-level fusion; scenario understanding; power system monitoring; substation equipment; intelligent grid; multimodal fusion; image semantic segmentation; infrared-visible image

收稿日期: 2024−04−16.　**网络出版日期:** 2025−04−18.
基金项目: 国网湖南省电力有限公司科技项目 (5216A522001Y).
通信作者: 张辉. E-mail: zhanghuihby@126.com.

电力人工智能领域的研究主要聚焦于常见电力场景缺陷类型或特定问题的检测和识别[1]。作

为电力传输与分配的关键节点，变电站设备的健康状况直接关系到电力系统的稳定运行和供电可靠性，因此其有效监测和故障诊断至关重要。变电站存在金属锈蚀、设备渗漏油、悬挂物等缺陷，基于可见光图像的缺陷检测算法能及时发现并排除这些安全隐患，对保障变电站的安全可靠运行有着重大意义[2]。例如，冯晗等[3]提出了一种基于改进 YOLOv5 的绝缘子检测方法，解决了变电站绝缘子串水冲洗机器人识别绝缘子的问题。然而，这类方法往往受限于光照条件和设备表面特性：一方面在光照不足的情况下难以正常工作，另一方面难以全面捕捉设备与温度变化密切相关的潜在故障。

近年来，红外成像技术因其非接触式监测物体的热辐射而成为监测设备温度的有效手段。采用红外热成像技术对变电站设备进行热异常诊断，能够有效监测电力设备表面的温度波动，及时识别潜在的故障[4-5]。然而，目前大多数的热异常检测手段主要依赖于单一的红外图像来分析故障特点，忽略了故障区域的外貌和纹理细节。这种方法在设备温度相近时分辨故障区域变得困难，而可见光图像却能清楚地展示物体的轮廓和细节，从而简化了与背景的区分过程。因此，通过结合使用红外和可见光两种传感器获得的图像，并对这些图像进行融合处理，不仅可以综合利用两种图像的互补特性来更精确地定位电力设备，还有助于对设备缺陷进行更有效的判断。

因此，将可见光图像与红外图像结合起来，利用多模态双光信息对变电站设备进行分析和诊断，成为一项具有重要研究价值和应用前景的技术。通过图像融合技术，可以综合利用可见光图像的高分辨率纹理信息和红外图像的温度信息，实现对变电站设备的全面监测。这种多模态融合不仅有助于提高故障检测的准确性和效率，还能在一定程度上克服单一模态图像在设备监测中的局限性，如可见光图像无法有效表征设备的温度信息，红外图像则缺乏足够的细节信息以支持复杂场景下的准确分割。但由于红外图像和可见光图像的模态差异，如何实现高效的融合是开展基于双光信息检测的变电站分析和诊断的基础。在可见光和红外双光融合方面，陶岩等[6]提出了一种结合自适应配准方法来解决弱光照条件下红外与可见光图像融合质量差的问题，但这种决策级融合难以针对具体的变电站场景任务融合双光的互补信息，难以显著改善具体任务的效果。在电力场景下，可见光和红外图像的融合往往基于特征级，Choi 等[7]提出了一种多模态图像特征融合模块，利用可见光和红外图像，以提高输电线路检测性能，但这种方法仅针对特定形态的电力线，难以应用到复杂的变电站场景。而 Xu 等[8]基于双空间图的交互网络 (dual-space graph-based interaction network, DSGBINet)，通过双光图像的协同工作，实现高压输电线路和变电站场景中电力设备的全天时间语义分割，但是图卷积带来了较大的计算量。

本文针对变电站与配电网等复杂电力场景中多模态图像融合效率低、语义分割精度不高的问题，提出了一种基于可见光与红外图像特征级融合的场景理解方法。所构建的对称双分支网络融合框架，结合多尺度特征融合层与注意力解码模块，能够充分挖掘多模态图像的互补信息，增强目标边缘细节表达与鲁棒性。该方法面向电力设备监测与故障诊断等典型任务，旨在为复杂环境下的智能巡检与视觉感知提供更加精准、稳定的技术支撑。

1　可见光和红外图像特征融合框架

为了充分利用可见光图像和红外图像的互补性，特征级融合在特征提取之后执行，以更有效地整合这两种模态的信息。早期特征级融合方法依赖于机器学习，例如主成分分析 (principal component analysis, PCA) 融合通过线性变换将数据转移到新的坐标系中，并选取主要成分进行数据融合。随机森林融合则通过建立多个决策树来对数据进行分类或回归分析，并在特征层面上融合来自不同源的数据。深度学习因其鲁棒性和免除手动特征提取的优点，逐渐成为特征融合领域的主流方法。深度学习特征融合通过深度神经网络自动提取和融合原始图像的特征，能够在网络的不同层次提取深层特征，并利用网络结构的设计实现有效的特征融合。

常见的两种基于深度学习的可见光和红外特征级融合方式如图 1 所示，在多尺度特征提取的基础上在每一层中实现跨模态特征融合。如图 1(a)，对称的融合方式[9-15]旨在设置两个对称的分支分别提取可见光和红外特征，然后将每一尺度下融合后的特征一起输入到解码器中解码。而非对称的特征融合模型[16-25]往往是利用某一模态的特征去补充另一模态的信息，实现特征增强。在图 1(b) 中，红外图像类似于先验信息或者显著图加入可见光图像的特征提取过程中。然后融合

的特征一方面会作为可见光的特征继续提取到深层, 另一方面也输入到解码器中进行解码。

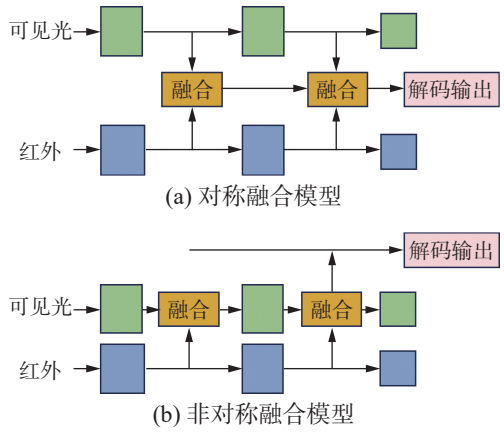


图 1　可见光和红外图像两种典型的特征级融合方式

Fig. 1　Two typical feature-level fusion methods of visible light and infrared images

文献 [26] 在分析夜间场景下自然图像中红外图像对可见光图像的增强作用时指出, 红外图像往往具有稳定的质量, 但在光照充足场景下可见光图像具有更清晰的结构。因此该文献采用非对称的多模态融合结构, 首先利用红外图像稳定的语义信息, 将其加入可见光编码器中, 以消除可见光中的特征干扰。受其启发, 本文针对复杂电力环境下的场景理解 (无人机配电网巡检), 指出在能见度高的情况下, 高分辨可见光图像具备更加清晰的轮廓, 可以为红外图像补充纹理和颜色信息, 这对准确识别红外图像中的电力部件, 并根据部件的热成像信息识别三相温度和相间温差至关重要。同时, 可见光图像也需要红外图像辅助进行光照不佳情况下的部件识别, 以便于进行表面缺陷检测 (例如配电网绝缘子裂痕、变电站油污等)。对称结构可以同等对待红外模态和可见光模态的重要性, 对可见光和红外特征进行相同的融合操作, 更加符合电力场景下的多模态巡检需求。因此, 本文选择以对称的融合模型作为基础框架, 在多尺度融合中构建高效的跨模态融合方式, 充分挖掘红外和可见光图像的互补信息, 构建出更全面、更具判别性的特征表达, 实现更精准的场景理解和更高精度的分割指标。

2　算法设计

本文提出一种基于可见光和红外图像特征级融合的复杂电力环境场景理解方法。针对变电站和配电网电力设备的双光图像 (如图 2 所示), 本文首先构建了一个双分支的对称融合网络, 采用视觉 Transformer[27] 作为基本块来提取两个分支的特征, 以更好地捕捉变电站和配电网设备的全局依赖关系, 从而更好地表征电力设备的独有特征。然后, 基于现有的特征提取双分支架构, 设计了高效的多尺度特征融合层, 在不同模态的特征建模关系中, 深入挖掘来自另一模态的互补信息, 以补全本模态的自身不足, 例如红外模态缺乏严格的纹理和轮廓信息和可见光能见度下降导致的特征表达能力弱。相比以往多模态融合模型的简单拼接和互注意力计算, 本文提出的融合层从多个层面完善跨模态融合学习。此外, 本文还设计了一个高效的特征解码模块, 可以很好地从融合特征中恢复红外和可见光的细节信息, 并用于输出最终的分割结果。

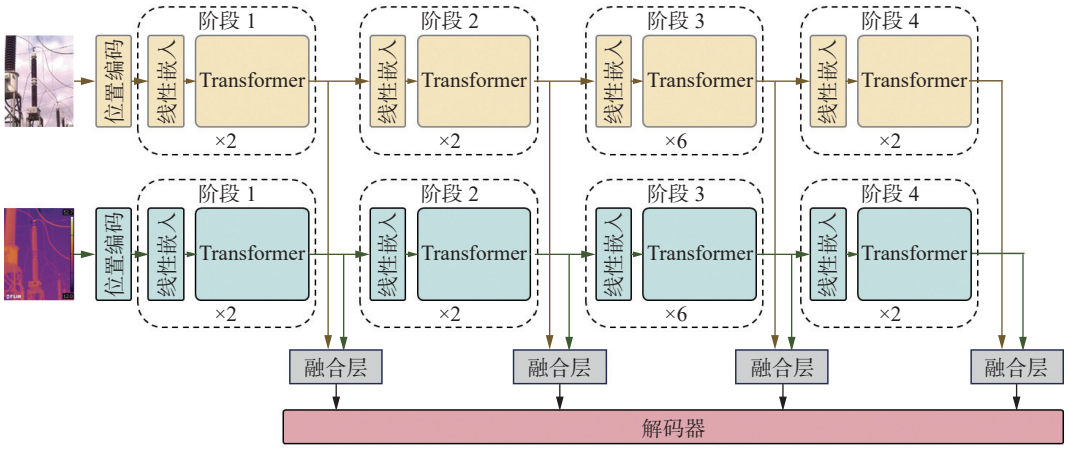


图 2　双分支对称融合网络

Fig. 2　Dual-branch symmetric fusion network

2.1　多尺度特征融合层

为了在不同尺度实现精准高效的融合效果, 本文提出了一种多尺度特征融合层, 其中每一个尺度的可见光和红外融合层如图 3 所示。考虑到实现准确的分割需要清晰的轮廓和像素的准确性, 因此本文旨在充分利用可见光模态和红外模

态特征的相关性完成特征融合。在融合层中，每一个可见光特征和红外特征表示为$\boldsymbol{V}_n$和$\boldsymbol{I}_n$，其中n表示特征提取的第n个阶段，n={1,2,3,4}。

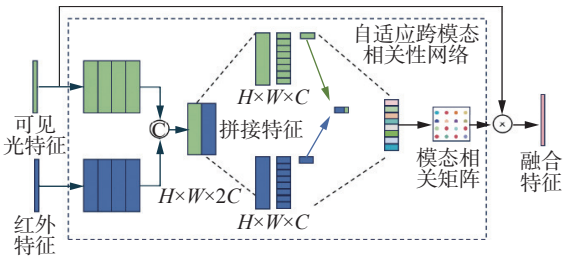


图 3　可见光红外特征融合层

Fig. 3　Visible and infrared feature fusion layer

本文设计了一个自适应跨模态相关网络，用于计算模态相关矩阵。自适应跨模态相关根据可见光和红外图像的不同特征进行优化，以实现在多个尺度上精准而高效的融合效果。该网络能够基于输入数据的具体特性自动调整权重，有效利用可见光的高分辨率纹理信息和红外图像的温度信息，从而增强模型的场景解析能力。具体而言，首先使用四层卷积层分别动态学习可见光和红外模态的代表性特征：

$$\boldsymbol{V}'_n = f(\boldsymbol{V}_n) \tag{1}$$

$$\boldsymbol{I}'_n = g(\boldsymbol{I}_n) \tag{2}$$

然后将特征图进行拼接，得到拼接后的代表性特征：

$$\boldsymbol{F}_n = \left[\boldsymbol{V}'_n \| \boldsymbol{I}'_n\right] \in \mathbf{R}^{h\times w\times 2n} \tag{3}$$

式中："‖"表示矩阵按列方向拼接。此时，$\boldsymbol{F}_n$中的每个像素点的每一个通道对应原图中每一个位置的红外和可见光的代表特征，接下来，通过多个多层感知机 (multilayer perceptron, MLP) 层筛选出重要的互补性特征$\boldsymbol{F}'_n$，对于变电设备而言，此时互补特征凸显了可见光图像较为清晰的边界特征，也强化了电力设备在红外图像上的发热区域。具体地，对于拼接特征$\boldsymbol{F}_n$，网络逐通道对红外模态和可见光模态进行特征响应比对，计算该特征通道内每个区域的跨模态权重，并对该特征通道中同一区域加权，组合成新的跨模态特征。

$$\boldsymbol{F}'_n = \mathrm{MLP}(\boldsymbol{F}_n) \tag{4}$$

跨模态特征获取的具体实现步骤为，首先通过变换函数$\phi: \mathbf{R}^{h\times w\times 2n} \to \mathbf{R}^{h\times w\times p}$将拼接特征$\boldsymbol{F}_n$映射到一个新特征空间$\boldsymbol{R}$：

$$\boldsymbol{R} = \phi(\boldsymbol{F}) \tag{5}$$

这一变换可以由参数化的线性变换 (本文选用 MLP) 实现。然后，使用归一化函数将每个特征向量的响应映射为跨模态权重：

$$\boldsymbol{W}^{\mathrm{CM}} = \psi(\boldsymbol{R}) \in \mathbf{R}^{h\times\omega\times 2p} \tag{6}$$

接下来，跨模态权重分成两个部分，分别对红外模态特征和可见光模态特征进行加权：

$$\boldsymbol{I}'_n = \boldsymbol{W}_I^{\mathrm{CM}} \odot \boldsymbol{I}'_n + \boldsymbol{W}_V^{\mathrm{CM}} \odot \boldsymbol{V}'_n \tag{7}$$

通过互补特征对于每一个位置最具有判别性的特征进行选择，获取可见光和红外图像的模态相关矩阵：

$$\boldsymbol{M} = \max_{1<d<D} \boldsymbol{F}'_n(i,j,d) \tag{8}$$

式中：D表示$\boldsymbol{F}'_n$的通道数，因此$\boldsymbol{F}'_n$在高度i、宽度j处的通道d上的特征最大值，表示红外图像和可见光图像的相关性的高低。通过这种方式，获得的模态相关矩阵可以动态获取不同模态特征的统计特性，对于每一个高度i和宽度j处的通道d，自适应选择其参数以最大化融合特征的表达能力。这种机制使得网络不再仅根据静态的预设规则操作，而是依据每个输入样本的具体内容自适应地调整处理策略。例如，在特征较为丰富的区域，网络可能会提高可见光图像特征的权重；而在热信息显著的区域，则加强红外图像的特征表达。模态相关矩阵不仅优化了融合过程的信息利用率，也显著提升了模型在处理多样化场景下的鲁棒性和准确性。最终得到融合后的特征：

$$\boldsymbol{F}_n^{\mathrm{fuse}} = \boldsymbol{M} \times \boldsymbol{V}_n \tag{9}$$

式中：$\boldsymbol{F}_n^{\mathrm{fuse}}$通过学习可见光和红外图像的相关性，将红外图像中具有判别性的特征表达融合到了可见光特征上，因此具备可见光的轮廓和纹理信息，又可以在受到光照等影响时，通过红外图像的温度分布特性加以补充。

2.2　多尺度注意力解码器

在获得多尺度融合特征$\boldsymbol{F}_n^{\mathrm{fuse}}, n=\{1,2,3,4\}$后，本文设计了一个多尺度注意力解码器，对不同尺度的融合特征进行尺度恢复和细节还原。解码器的具体结构如图 4 所示。

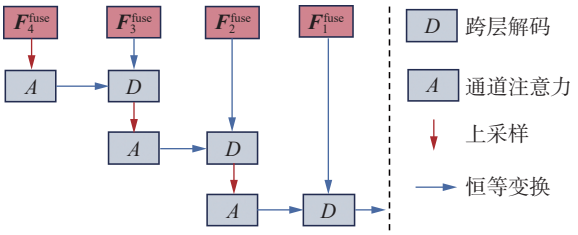


图 4　多尺度注意力解码

Fig. 4　Multi-scale attention decoding

由于高层级特征往往具备更多的全局信息，因此本文对于每一层级的融合特征$\boldsymbol{F}_n^{\mathrm{fuse}}$采用逐级解码的方式，分别从高层级到低层级进行恢复，并与上一层级更具细节信息的特征拼接。在多尺度解码框架中，本文采用一个跨层解码层D来恢

复相邻尺度的融合特征, 并引入通道注意力 A 加强更具判别性的特征表达。为体现本文所设计多尺度解码的有效性, 我们避免直接引用特定的网络层, 而是通过多层解码层和通道注意力来抽象整个解码过程, 从而突显本方法的数学本质, 以及解码层的灵活性和可拓展性。基于此, 对于连续层级的融合特征 $\boldsymbol{F}_n^{\text{fuse}}$ 和 $\boldsymbol{F}_{n-1}^{\text{fuse}}$, 跨层解码特征被表示为

$$\boldsymbol{F}_n^{\text{decode}} = \mathcal{D}\left(A(\text{Upsample}(\boldsymbol{F}_{n+1})), \boldsymbol{F}_n\right) \quad n<4 \tag{10}$$

式中: $\mathcal{D}$ 表示从高层特征到低层特征的转换过程, 包括但不限于线性映射、非线性映射或者其他复合函数, 最常用的方式是使用 MLP 和二维卷积。A 表示注意力机制, 其中一种通道注意力计算过程为, 首先计算全局平均池化, 获得对融合特征的全局描述, 即哪些特征在分割中更具判别性。对于每一个通道计算全局判别性权重 $\boldsymbol{G}_{\text{c}}$:

$$\boldsymbol{G}_{\text{c}} = \text{Sigmoid}\left(\text{FC}\left(\frac{1}{H\times W}\sum_{h=1}^{H}\sum_{w=1}^{W}\boldsymbol{F}_{i,h,w}\right)\right) \tag{11}$$

式中: FC(·) 表示全连接层; $\boldsymbol{F}_{i,h,w}$ 表示待计算注意力的特征图, 通常为 $\boldsymbol{F}_4^{\text{fuse}}$ 和 $\boldsymbol{F}_n^{\text{decode}}$, n={1,2,3}, 得到计算注意力的特征为

$$\boldsymbol{F}_{\text{ca}} = \boldsymbol{F}_{i,h,w}\cdot\boldsymbol{G} \tag{12}$$

式中 $\boldsymbol{G}$ 表示所有通道的全局判别性权重, 最终, F_1^{decode} 经过后处理, 得到最终的分割结果。在本文设计的多尺度解码器中, $\mathcal{D}$ 后续可替换为更多映射方式, 例如 Transformer 和 mamba 模块。A 可以根据后续电力场景的其他任务 (例如小目标检测) 中设计更加尺度敏感的注意力机制。

2.3　损失函数

本研究选择一种组合损失函数来优化多尺度特征融合网络, 以应对复杂电力环境下的图像分割任务。损失函数的设计关键在于能够有效处理类别不平衡问题, 并且促进模型在各种尺度上精确预测细节。具体地, 损失函数结合交叉熵损失 (cross-entropy loss)[28-29] 和 Dice 损失 (Dice loss)[30-31], 这两种损失各有优势, 它们的组合被证明在多种图像分割任务中非常有效。交叉熵的计算方式为

$$L_{\text{ce}} = -\sum_{c=1}^{C} y_{o,c}\log\left(p_{o,c}\right) \tag{13}$$

式中: C 表示类别总数; $y_{o,c}$ 是指示函数, 表示类别 c 是否是像素集 o 的正确分类; $p_{o,c}$ 表示 o 预测成类别 c 的概率。

Dice 损失基于 Dice 系数, Dice 系数是常用于医学图像分割的相似度度量指标。本文选用此损失的目的是为了消除电力样本中存在的不平衡性, 其计算方式为

$$L_{\text{Dice}} = 1 - \frac{2\times\sum_{i=1}^{N} y_i\cdot p_i + \epsilon}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} p_i^2 + \epsilon} \tag{14}$$

式中: y_i 是二进制标签; p_i 是预测概率; ϵ 是一个小常数, 用于避免分母为零的情况。本文模型采用交叉熵损失与 Dice 损失的线性组合, 这有助于同时优化类别间的区分性和预测区域的几何一致性。组合损失函数表达为

$$L = \alpha\times L_{\text{ce}} + (1-\alpha)\times L_{\text{Dice}} \tag{15}$$

式中: 权重因子 α 设置为 0.7, 该值是基于实验验证结果选择的, 旨在联合两种损失对模型进行训练。

3　实验分析

3.1　数据集和实验设置

本文选取湖南省某变电站红外和可见光数据集进行实验, 采集设备为 FLIR E8 XT 多成像手持热像仪。该手持设备同时搭载了红外热像仪和数码相机。红外分辨率为 320 像素×240 像素, 可见光分辨率为 640 像素×480 像素。对每一组可见光图像进行裁剪, 并和红外图像特征点匹配, 得到可见光和红外图像的分辨率均为 320×240 像素。对配准后的数据扩增, 最终得到 500 组训练集和 100 组测试集图片, 每组图片包括一对红外和可见光图像。对所有图像对进行像素级标注, 标注的类别包括: 电压互感器、电流互感器、套管和隔离开关。评价指标包括 (对以上的变电站设备即每一个类别的) 交并比 I_{iou} 和像素分类准确率 A_{acc}, 来评价本文分割性能的好坏。I_{iou} 和 A_{acc} 的计算方式分别为

$$I_{\text{iou}} = \frac{T_{\text{p}}}{T_{\text{p}} + F_{\text{n}} + F_{\text{p}}} \tag{16}$$

$$A_{\text{acc}} = \frac{T_{\text{p}} + T_{\text{n}}}{T_{\text{p}} + T_{\text{n}} + F_{\text{p}} + F_{\text{n}}} \tag{17}$$

式中: T_{p} 表示正确标记为该类别的像素数量; T_{n} 表示没有正确标记为该类别的像素数量 (即正确标记为其他类别的像素); F_{p} 表示错误标记为该类别的像素数量; F_{n} 表示没有错误地标记为该类别的像素数量。I_{iou} 是预测和真实像素集的交集与并集的比例, 衡量了分割结果中预测区域和真实区域的重叠程度, 特别是对于单一类别。A_{acc} 是在整个图像中, 正确分类像素的比例, 包括目标类别和非目标类别的正确预测。因此实验给出了每一类目标的 I_{iou} 和平均 I_{iou} (M_{iou}), 以及平均

$A_{acc}(M_{acc})$。

本文实验的模型参数设置如表 1 所示。本文模型具有 4 个多尺度融合层，因此对于每一个融合层，给出 4 个相同的卷积核参数。在多尺度注意力解码器中，根据待解码的融合特征中 n 的数值来命名 MLP 层的顺序，并依次给出每一层中神经元的个数。此外，还给出了该网络层中输入和输出张量的大小。

表 1　实验模型参数设置

Table 1　Model parameter settings in experiments

模块	网络层	参数 (卷积: 卷积核大小， 输入输出通道数) (MLP: 神经元个数)	输入维度 (B,H,W,C)	输出维度 (B,H,W,C)
多尺度特征融合层	融合层1	卷积核3×3,128,128	(8,160,120,128)	(8,160,120,128)
	融合层2	卷积核3×3,256,256	(8,80,60,256)	(8,80,60,256)
	融合层3	卷积核3×3,512,512	(8,40,30,512)	(8,40,30,512)
	融合层4	卷积核3×3,1 024,1 024	(8,20,15,1 024)	(8,20,15,1 024)
多尺度注意力解码器	MLP层4	1 024,512	(8,40,30,1 024)	(8,40,30,512)
	MLP层3	1 024,256	(8,80,60,1 024)	(8,80,60,256)
	MLP层2	512,128	(8,160,120,512)	(8,160,120,128)
	MLP层1	256,128,64,4	(8,320,240,256)	(8,320,240,4)

本文在包含 NVIDIA GTX 3090 GPU (24 GB RAM) 的服务器上使用 PyTorch 进行训练。在训练过程中，批量大小设置为 2，AdamW 优化器用于优化训练过程。初始学习率和权重衰减分别设置为 6×10^{-5} 和 0.01，在构建的数据集上训练了 500 Epoch，模型的损失变化曲线如图 5 所示。

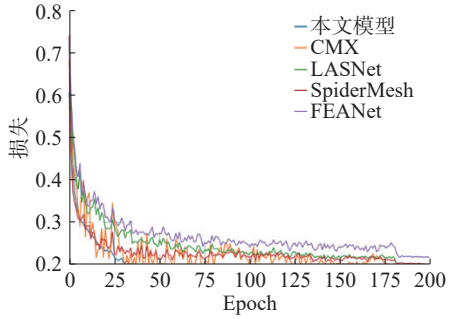


图 5　模型损失函数曲线

Fig. 5　Loss function curves of the model

3.2　与不同网络结构的对比实验

为了验证本文方法对复杂电力环境场景理解的性能，本文选取常用的多模态分割网络作为对比，并且与一些常见的可见光分割网络也进行了对比，以验证多模态对于场景理解的有效性，实验结果如表 2 所示。在与其他几种双光方法的对比中，本文方法取得了最高的 M_{iou}，并且在电压互感器、电流互感器、套管和隔离开关等类别上也取得了最好的结果。实验证明，本文提出的复杂电力环境场景理解方法在变电站数据集上是有效的，可以通过跨模态融合方式优化变电站的分割结果。此外，相对于单独使用可见光的方法，双光融合的方法可以利用来自另一个模态的互补性信息，从而提升信息的判别性，以实现最终的结果提升。

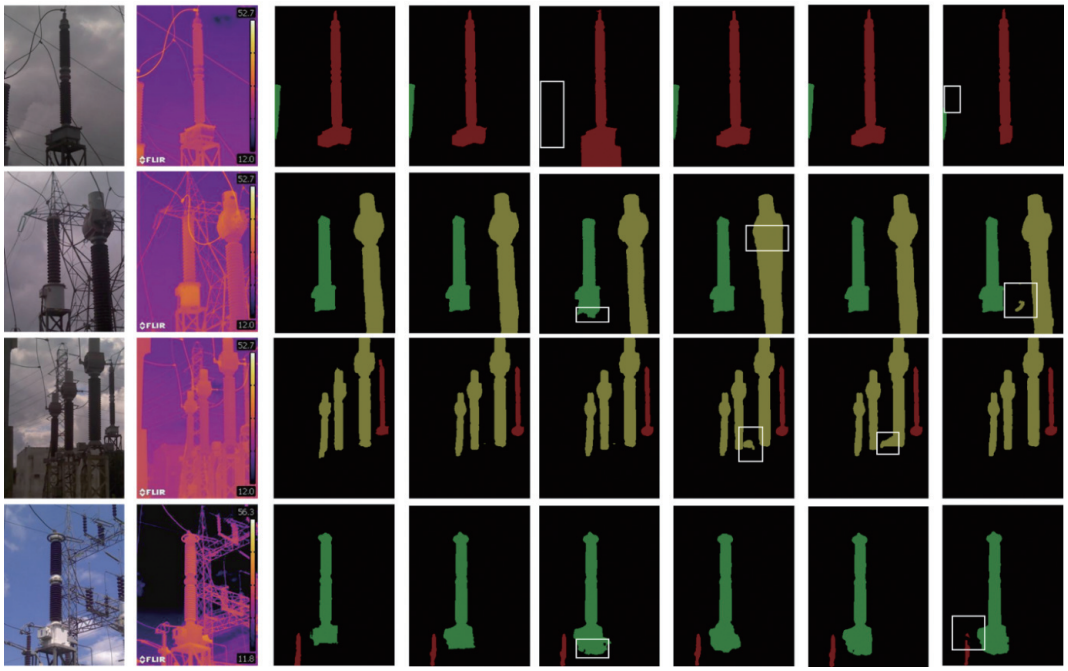
表 2　与不同方法的对比结果

Table 2　Comparison with the results of different methods　%

方法	模态	背景I_{iou}	电压互感器I_{iou}	电流互感器I_{iou}	套管I_{iou}	隔离开关I_{iou}	M_{iou}	M_{acc}
FEANet[18]	双光	94.32	69.02	79.75	79.28	70.91	78.25	83.67
SpiderMesh[19]	双光	96.67	70.03	79.20	82.50	70.90	79.86	84.74
LASNet[20]	双光	95.64	69.70	80.76	83.56	72.45	80.42	84.59
CMX[21]	双光	96.74	70.82	79.57	84.52	73.07	80.94	85.65
本文方法	双光	**96.66**	**72.99**	**81.83**	**85.82**	**72.13**	**81.89**	**86.89**

实验的可视化案例如图 6 所示。本文在测试集中随机选取了部分案例，展示了可见光、红外图像、真实标签以及对应方法的分割结果。分割结果与真实标签的偏差在图中用矩形框标出。实验证明，本文方法在细节上发生偏差的次数均小于其他多模态方法，这说明其对红外和可见光图

像的利用更加充分, 且更全面地挖掘互补信息。这一现象表明, 本研究在融合可见光与红外图像时更为高效, 能更全面挖掘地这两种模态之间的互补信息, 进而提高了图像分割的准确性和可靠性。



(a) 可见光 (b) 红外图像 (c) 真实标签 (d) 本文方法 (e) CMX (f) LASNet (g) SpiderMesh (h) FEANet

图 6　模型的视觉结果对比

Fig. 6　Comparison of visual results of the model

3.3　消融实验

为验证本文所提出模块的有效性, 对多尺度特征融合层和多尺度注意力解码器进行消融实验。对照实验中, 去掉多尺度特征融合, 只是把每一个层级的可见光和红外特征简单拼接和进行 MLP 处理。而对于多尺度注意力层, 去掉逐级解码方式, 直接对不同尺度的融合特征进行上采样和拼接。此外, 为验证本文多尺度特征融合的偏向性, 将模态相关矩阵相乘的特征设置为红外特征进行对照。实验结果如表 3 所示。由实验结果可知, 本文提出的特征融合和注意力解码对于网络最终的分割结果均是有效的。相较于直接融合和直接解码, 本文方法可以更加充分地挖掘来自不同模态信息的互补性, 并且在解码过程中能够重点加权网络提取的判别性特征, 从而更有效地恢复图像中的细节。此外, 模态相关矩阵与可见光模态相乘的结果优于与红外模态相乘的结果, 这说明对于场景理解任务, 相比红外温度分布而言, 可见光中的纹理和边缘信息平均占据更大的比重。

表 3　消融实验结果

Table 3　Ablation study results　　%

特征融合层	注意力解码	电压互感器I_{iou}	电流互感器I_{iou}	套管I_{iou}	隔离开关I_{iou}	M_{iou}	M_{acc}
红外		71.56	74.57	67.80	70.78	75.94	81.24
可见光		71.29	75.98	67.93	70.70	76.12	81.67
不使用融合层	√	70.15	72.87	65.58	69.87	74.73	80.45
红外	√	**73.24**	80.14	66.35	72.28	77.56	83.36
可见光	√	71.89	**80.73**	**85.52**	**73.13**	**81.89**	**86.63**

为验证本文采用的对称融合结构的有效性, 设置了两种非对称的结构与本文结构对照: 1) 将多尺度编码器中红外图像每一层级的特征融合到同一层级的可见光分支; 2) 将每一层级可见光分支融合到红外分支中。为保证公平比较, 特征融合和解码同样采用本文提出的多尺度特征融合层和多尺度注意力解码器。实验结果如表 4 所示, 可以看出本文所采用的对称结构性能优于两种非对称结构, 而非对称结构之间的性能较为接近。这说明电力场景解析中, 两种模态的重要性较为等同, 本文采用的对称融合结构是有效的。

表 4　不同融合结构实验结果
Table 4　Experimental results of different fusion structures　%

结构	融合层位置	电压互感器I_{iou}	电流互感器I_{iou}	套管I_{iou}	隔离开关I_{iou}	M_{iou}	M_{acc}
非对称结构一	可见光分支	70.84	77.98	80.48	72.24	80.11	85.21
非对称结构二	红外分支	71.45	78.54	81.72	72.64	80.47	85.45
本文对称结构	独立分支	**71.89**	**80.73**	**85.52**	**73.13**	**81.89**	**86.63**

3.4　一般性实验

为进一步验证可见光和红外图像融合的作用，本文在可见光图像中引入了天气和光照变化。如图 7 所示，对测试集进行了模拟天气，天气情况包括下雨、雾、阳光等情况。在相同实验条件下进行测试，实验结果如表 5 所示。可以看出，即使在极端条件下，可见光的能见度降低，红外图像可以为其提供基本补充，保证指标的下降在可控的范围内。可见光图像在分割中占据更多的权重，但是热成像可以在能见度下降时保持基本的感知能力，因此本文对于复杂电力场景下可见光和红外图像的处理设计是有效的。这是因为通过红外和可见光图像融合技术，有效克服了光照不足和视线遮挡情况下的监控挑战。红外图像捕捉的是物体表面的热辐射，与可见光图像 (依赖光照的反射信息) 截然不同。红外成像技术不受光照条件的限制，因此在低光或无光环境下仍能提供有效信息。此外，红外图像对天气条件如雾和雨的干扰也具有较强的抵抗力。由于这些特性，红外图像在可见光图像质量受损时 (如光照不足或视线遮挡) 提供了宝贵的补充信息，尤其是在物体的热特性显著时。本文方法能够在各种环境条件下保持优异的图像质量和监控性能。

(a) 原始图像　(b) 晴天　(c) 雾天　(d) 雨天

图 7　模拟天气情况
Fig. 7　Simulate weather conditions

表 5　模拟天气实验结果
Table 5　Results of the simulated weather experiments　%

模拟天气情况	电压互感器I_{iou}	电流互感器I_{iou}	套管I_{iou}	隔离开关I_{iou}	M_{iou}
原始图像	71.89	80.73	85.52	73.13	81.89
雨天	70.32	78.67	80.34	70.47	78.54
晴天	70.85	80.30	84.50	73.07	80.67
雾天	71.11	78.27	80.12	70.82	78.09

4　结束语

本研究提出了一种基于可见光和红外图像特征级融合的复杂电力环境场景理解方法，旨在提高变电站和配电网设备监测的准确性和效率。通过对比传统的单模态图像分析方法，证明了双模态融合方法在提高故障检测准确率、提升场景理解能力方面的优越性。实验结果显示，该方法在多个变电站设备分类任务中取得了优异的性能，尤其是在光照不足或有遮挡的情况下，展现了明显的优势。通过深入分析可见光与红外图像的互补信息，本文成功设计并实现了一种有效的特征级融合框架和算法，使得融合后的图像能够同时反映设备的表面细节与温度变化信息，从而为电力系统的安全运行提供了更全面的监测数据。此外，通过在多尺度下实施特征融合和注意力机制，本方法进一步增强了模型的分割准确性和细节恢复能力，为复杂电力环境下的设备监测与故障诊断提供了新的技术手段。

本文的局限性和潜在问题主要体现在三个方面。首先，尽管本文研究成功实现了可见光和红外图像的特征级融合，但由于两种模态在物理属性上的根本差异，融合效果和精确度仍然存在一定局限。例如，红外图像在低温差环境下可能无法提供足够的细节信息，而可见光图像在光照不足或过强的情况下也可能失真。其次，当前的融合算法虽然在实验室条件下表现良好，但其计算

复杂度较高, 在需要快速响应的电力系统监测和故障诊断场景中, 处理速度成为限制其应用的一个关键因素。最后, 该方法主要针对特定类型的电力设备和环境进行优化, 可能在不同设备或更复杂或变化的环境条件下的泛化能力有限。

针对上述局限性, 未来研究将聚焦于降低特征融合的计算复杂度, 例如通过简化模型结构或采用更高效的算法来提升处理速度。同时, 研究如何减少模态差异对融合效果的负面影响, 可能通过引入更先进的模态对齐技术来实现。其次, 通过在更广泛的数据集上训练模型, 包括不同类型的电力设备和更多样化的环境条件, 可以提高模型的泛化能力。此外, 未来将深入探索不同模态之间的互补信息, 研究如何更有效地利用这些信息来提高故障检测的准确性。

参考文献:

[1] 傅博, 姜勇, 王洪光, 等. 输电线路巡检图像智能诊断系统[J]. 智能系统学报, 2016, 11(1): 70−77.
FU Bo, JIANG Yong, WANG Hongguang, et al. Intelligent diagnosis system for patrol check images of power transmission lines[J]. CAAI transactions on intelligent systems, 2016, 11(1): 70−77.

[2] 张铭泉, 邢福德, 刘冬. 基于改进 Faster R-CNN 的变电站设备外部缺陷检测[J]. 智能系统学报, 2024, 19(2): 290−298.
ZHANG Mingquan, XING Fude, LIU Dong. External defect detection of transformer substation equipment based on improved Faster R-CNN[J]. CAAI transactions on intelligent systems, 2024, 19(2): 290−298.

[3] 冯晗, 姜勇. 使用改进 Yolov5 的变电站绝缘子串检测方法[J]. 智能系统学报, 2023, 18(2): 325−332.
FENG Han, JIANG Yong. A substation insulator string detection method based on an improved Yolov5[J]. CAAI transactions on intelligent systems, 2023, 18(2): 325−332.

[4] 黄志鸿, 颜星雨, 陶岩, 等. 基于多模态图像信息的配电网部件定位方法[J]. 湖南电力, 2024, 44(6): 83−89.
HUANG Zhihong, YAN Xingyu, TAO Yan, et al. Distribution network component positioning methods based on multi-modal image information[J]. Hunan electric power, 2024, 44(6): 83−89.

[5] 张辉, 杜瑞, 钟杭, 等. 电力设施多模态精细化机器人巡检关键技术及应用[J]. 自动化学报, 2025, 51(1): 20−42.
ZHANG Hui, DU Rui, ZHONG Hang, et al. The key technology and application of multi-modal fine robot inspection for power facilities[J]. Acta automatica sinica, 2025, 51(1): 20−42.

[6] 陶岩, 张辉, 黄志鸿, 等. 面向配电网典型部件的热故障精准判别方法[J]. 智能系统学报, 2025, 20(2): 506−515.
TAO Yan, ZHANG Hui, HUANG Zhihong, et al. Accurate identification of thermal faults for typical components of distribution networks[J]. CAAI transactions on intelligent systems, 2025, 20(2): 506−515.

[7] CHOI H, YUN J P, KIM B J, et al. Attention-based multimodal image feature fusion module for transmission line detection[J]. IEEE transactions on industrial informatics, 2022, 18(11): 7686−7695.

[8] XU Chang, LI Qingwu, JIANG Xiongbiao, et al. Dual-space graph-based interaction network for RGB-thermal semantic segmentation in electric power scene[J]. IEEE transactions on circuits and systems for video technology, 2023, 33(4): 1577−1592.

[9] ZHOU Wujie, LIU Jinfu, LEI Jingsheng, et al. GMNet: graded-feature multilabel-learning network for RGB-thermal urban scene semantic segmentation[J]. IEEE transactions on image processing, 2021, 30: 7790−7802.

[10] ZHOU Wujie, DONG Shaohua, XU Caie, et al. Edge-aware guidance fusion network for RGB-thermal scene parsing [EB/OL]. (2021−12−09)[2024−01−01]. https://arxiv.org/abs/2112.05144.

[11] LIU Jinfu, ZHOU Wujie, CUI Yueli, et al. GCNet: Grid-like context-aware network for RGB-thermal semantic segmentation[J]. Neurocomputing, 2022, 506: 60−67.

[12] LIN Baihong, LIN Zengrong, GUO Yulan, et al. Variational probabilistic fusion network for RGB-T semantic segmentation[EB/OL]. (2023−06−17)[2024−01−01]. https://arxiv.org/abs/2307.08536.

[13] LI Ping, CHEN Junjie, LIN Binbin, et al. Residual spatial fusion network for RGB-thermal semantic segmentation [EB/OL]. (2023−06−17)[2024−01−01]. https://arxiv.org/abs/2306.10364.

[14] ZHANG Jiyou, ZHANG Rongfen, LIU Yuhong, et al. RGB-T semantic segmentation based on cross-operational fusion attention in autonomous driving scenario[J]. Evolving systems, 2024, 15: 1429−1440.

[15] LIN Zengrong, LIN Baihong, GUO Yulan. Label-guided real-time fusion network for RGB-T semantic segmentation[C]//Proceedings of the British Machine Vision Conference. Aberdeen: BMVC, 2023: 767−770.

[16] SUN Yuxiang, ZUO Weixun, YUN Peng, et al. FuseSeg: semantic segmentation of urban scenes based on RGB and thermal data fusion[J]. IEEE transactions on automation science and engineering, 2021, 18(3): 1000−1011.

[17] ZHOU Wujie, LIN Xinyang, LEI Jingsheng, et al. MFFENet: multiscale feature fusion and enhancement network for RGB-thermal urban road scene parsing[J]. IEEE transactions on multimedia, 2021, 24: 2526−2538.

[18] DENG Fuqin, FENG Hua, LIANG Mingjian, et al. FEANet: feature-enhanced attention network for RGB-thermal real-time semantic segmentation[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems. Prague: IEEE, 2021: 4467–4473.

[19] FAN Siqi, WANG Zhe, WANG Yan, et al. SpiderMesh: spatial-aware demand-guided recursive meshing for RGB-T semantic segmentation[EB/OL]. (2023−09−27)[2024−01−01]. https://arxiv.org/abs/2303.08692v2.

[20] LIN Baihong, LIN Zengrong, GUO Yulan, et al. Asymmetric multimodal guidance fusion network for real-time visible–thermal semantic segmentation[J]. Robotics and computer-integrated manufacturing, 2024, 86: 103822.

[21] ZHANG Jiaming, LIU Huayao, YANG Kailun, et al. CMX: cross-modal fusion for RGB-X semantic segmentation with transformers[J]. IEEE transactions on intelligent transportation systems, 24(12): 14679−14694.

[22] LI Gongyang, WANG Yike, LIU Zhi, et al. RGB-T semantic segmentation with location, activation, and sharpening[J]. IEEE transactions on circuits and systems for video technology, 2022, 33(3): 1223–1235.

[23] SHIN U, LEE K, KWEON I S, et al. Complementary random masking for RGB-thermal semantic segmentation[C]//2024 IEEE International Conference on Robotics and Automation. Yokohama: IEEE, 2014: 11110−11117.

[24] WANG Yuxin, LI Gongyang, LIU Zhi. SGFNet: semantic-guided fusion network for RGB-thermal semantic segmentation[J]. IEEE transactions on circuits and systems for video technology, 2023, 33(12): 7737−7748.

[25] LI Gongyang, WANG Yike, LIU Zhi, et al. RGB-T semantic segmentation with location, activation, and sharpening[J]. IEEE transactions on circuits and systems for video technology, 2023, 33(3): 1223−1235.

[26] ZHOU Zikun, WU Shukun, ZHU Guoqing, et al. Channel and spatial relation-propagation network for RGB-thermal semantic segmentation[EB/OL]. (2023−08−24)[2024−01−01]. https://arxiv.org/abs/2308.12534.

[27] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. (2021−06−03)[2024−01−01]. https://arxiv.org/abs/2010.11929.

[28] MAO Anqi, MOHRI Mehryar, ZHONG Yutao. Cross-entropy loss functions: theoretical analysis and applications[EB/OL]. (2023−06−15)[2024−01−01]. https://arxiv.org/pdf/2304.07288v1.

[29] ZHANG Zhi, SABUNCU M R. Generalized cross entropy loss for training deep neural networks with noisy labels[J]. Advances in neural information processing systems, 2018, 31: 8792−8802.

[30] SUDRE C H, LI Wenqi, VERCAUTEREN T, et al. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations[EB/OL]. (2017−07−14)[2024−01−01]. https://arxiv.org/abs/1707.03237.

[31] LI Xiaoya, SUN Xiaofei, MENG Yuxian, et al. Dice loss for data-imbalanced NLP tasks[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020: 465–476.

作者简介：

黄志鸿，高级工程师，博士研究生，主要研究方向为电力人工智能。E-mail：zhihong_huang111@163.com。

杜瑞，博士研究生，主要研究方向为电力人工智能、多模态感知。E-mail：durui@hnu.edu.cn。

张辉，教授，博士生导师，博士，主要研究方向为机器人视觉检测、深度学习、图像识别、机器人智能控制、嵌入式系统应用。近年来，主持科技创新 2030—“新一代人工智能”重大项目课题、国家自然科学基金共融机器人重大研究计划重点项目，国家重点研发计划子课题、国家科技支撑计划项目子课题等 20 余项。技术成果获 2018 年国家技术发明奖二等奖，以主要完成人获得省部级科学技术奖励一等奖 8 项。发表学术论文 50 余篇，获国家发明专利授权 38 项、计算机软件著作权 5 项。E-mail：zhanghuihby@126.com。