

DOI: 10.19650/j.cnki.cjsi.J2514647

# 基于多模成像融合的弓网电弧检测方法研究 \*

蔡智超<sup>1,2</sup>, 王健芬<sup>1,2</sup>, 程宏波<sup>1,2</sup>, 韦宝泉<sup>1,2</sup>, 孙洪宇<sup>1,3</sup>

(1. 山区土木工程安全与韧性全国重点实验室 南昌 330013; 2. 华东交通大学电气与自动化工程学院  
南昌 330013; 3. 北京交通大学物理科学与工程学院 北京 100044)

**摘要:**弓网电弧故障是制约高速铁路与城市轨道交通安全稳定运行的重要隐患之一,其产生的强光、高温及电磁干扰会加剧接触网部件磨损,缩短运用寿命,并可能诱发供电系统故障,造成重大安全事故。传统基于可见光的弓网电弧识别易受光照变化、遮挡及天气条件等环境因素干扰,导致检测精度与鲁棒性不足,难以满足复杂场景下在线监测需求。为提升检测性能,提出了一种融合可见光、红外、声信号的多模成像弓网电弧检测方法。首先,利用麦克风阵列采集的弓网电弧声信号并构建时频矩阵;随后,引入基于变分推断的噪声抑制策略,抑制环境背景噪声并保留电弧声信息;在此基础上,采用时域波束形成实现声源成像与能量聚焦,得到声学成像图。进一步,将声学图像与可见光、热成像数据进行配准与空间对齐,获得电弧形态的多模态图像表达,并将配准后的图像输入多模态目标检测模型,最终获得弓网电弧位置与置信度信息,完成电弧故障的检测与定位。为论证关键环节的有效性,搭建声学传播模型和实验平台,系统分析电弧声源传播规律并验证噪声抑制策略对信噪比与成像性能的提升作用。实验结果表明,所提多模成像融合方法,相较单一可见光模态与可见光/红外双模态方案,识别精度分别提升15.9%与8.1%,能够在多工况干扰环境下保持稳定检测性能,为弓网电弧的在线监测提供技术支撑。

**关键词:**弓网电弧;多模态;噪声抑制;故障检测

中图分类号: TH701 TP391.4 文献标识码: A 国家标准学科分类代码: 470.40

## Research on pantograph-catenary arc detection via multimodal imaging fusion

Cai Zhichao<sup>1,2</sup>, Wang Jianfen<sup>1,2</sup>, Cheng Hongbo<sup>1,2</sup>, Wei Baoquan<sup>1,2</sup>, Sun Hongyu<sup>1,3</sup>

(1. State Key Laboratory of Safety and Resilience of Civil Engineering in Mountain Area, Nanchang 330013, China;  
2. School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China;  
3. School of Physical Sciences and Engineering, Beijing Jiaotong University, Beijing 100044, China)

**Abstract:** Pantograph-catenary arc faults pose a serious threat to the safe and stable operation of high-speed railway and urban rail transit systems. The intense light, high temperature, and electromagnetic interference generated by these arcs accelerate the wear of catenary components, shorten their service life, and may even trigger power supply system failures, leading to serious safety incidents. Traditional visible-light-based pantograph-catenary arc detection methods are susceptible to environmental interferences such as illumination variations, occlusions, and adverse weather conditions, leading to reduced detection accuracy and robustness and thereby limiting their applicability in complex online monitoring scenarios. This paper proposes a multimodal imaging-based arc detection method that integrates visible, infrared, and acoustic signals to enhance performance in complex scenes. Initially, the acoustic signals of arc are collected by a microphone array and transformed into time-frequency matrices. Subsequently, a variational inference-based noise suppression strategy is introduced to attenuate environmental background noise while preserving arc-related acoustic information. Building on this, time-domain beamforming is employed to achieve acoustic source imaging and energy focusing, yielding acoustic intensity maps. The acoustic images are then registered and spatially aligned with visible and thermal imagery to obtain a multimodal representation of arc morphology. The registered images are then fed into a multimodal object detection model to produce arc locations and confidence scores, thereby completing the detection and localization of the arc fault. To evaluate the proposed method, an acoustic propagation model and an experimental platform have been

established to analyze the propagation characteristics of arc sources and systematically verify the impact of the noise-suppression strategy on signal-to-noise ratio and imaging performance. The experimental findings demonstrate that, in comparison with single visible-light modality and visible/infrared bimodal schemes, the proposed multimodal imaging fusion method enhances recognition accuracy by 15.9% and 8.1%, respectively, thus providing an effective solution for robust online detection of pantograph-catenary arcs.

**Keywords:** pantograph-catenary arc; multimodal imaging fusion; noise suppression; fault detection

## 0 引言

截至 2024 年 12 月底, 我国铁路营业里程将突破 16.2 万公里, 高速铁路总里程将超过 4.8 万公里<sup>[1]</sup>。高速列车运行所需的能量主要通过“接触网-受电弓系统”(pantograph-catenary system, PCS)滑动接触系统实现能量传输, 因而, PCS 在维持列车安全、平稳运行中起着至关重要的作用<sup>[2]</sup>。受流性能是衡量 PCS 系统运行质量的重要指标, 其中, 弓网电弧是评估受流质量的核心参量之一<sup>[3]</sup>, 电弧会加剧受电弓滑板与接触网导线的磨损, 释放强烈的电磁辐射与干扰信号, 严重时甚至可能造成列车控制系统异常或受电弓烧蚀<sup>[4]</sup>。因此, 针对 PCS 系统中电弧的实时精准检测, 对保障电气化铁路的安全与稳定运行具有重要的现实意义。

目前, 针对 PCS 系统的电弧检测技术可归为两类: 接触式检测与非接触式检测。前者主要依托于电信号参数的变化来实现识别, 包括振荡电流法<sup>[5-6]</sup>、电流波形特征分析法<sup>[7-8]</sup>及电阻测量法<sup>[9-10]</sup>等; 后者则利用多源外部信号进行监测, 如可见光图像检测<sup>[11]</sup>、电磁辐射信号检测<sup>[12]</sup>、红外图像检测<sup>[13]</sup>、紫外信号检测<sup>[14]</sup>以及声学信号检测<sup>[15]</sup>等多种方式。

综合对比, 非接触式视觉检测具有实现简便、成本较低等优势, 但视觉检测的好坏依赖于图像质量的高低, 如飞溅等会污染和遮挡视觉传感器<sup>[16]</sup>, 同时弱光照<sup>[17]</sup>、强光、烟雾及雨雾雪<sup>[18]</sup>等在复杂环境中稳定性不足、检测精度不足所造成的漏检问题值得关注。

值得注意的是, 声学检测技术正逐步由传统信号分析拓展至声学成像领域, 并广泛应用于工业噪声监测与声源定位研究, 作为传统视觉检测的补充检测方法适用性强<sup>[19]</sup>。Chu 等<sup>[20]</sup>提出基于模态组合波束成形(modal composition beamforming, MCB)方法的多旋转声源高分辨率定位方法, 成功实现了轴流风扇叶片噪声源的精准识别; 文献[21]则提出紧凑型麦克风阵列结合高级反卷积的波束形成算法, 实现了大型风力涡轮机的精确噪声定位。针对高噪声环境下声信号难以成像的问题, 对角线去除法是声学成像过程中常用的处理手段, 削弱非相关背景噪声的影响<sup>[22]</sup>。

结合上述检测方法的各自特点, 本研究提出了一种多模成像融合的弓网电弧检测方法, 融合了声学检测、可

见光检测以及红外检测的优势, 旨在实现复杂运行环境下弓网电弧的高精度识别与定位, 以降低恶劣环境下弓网电弧漏检情形。主要内容包括检测框架设计、多模态数据采集和处理、检测模型搭建与评估等。

## 1 弓网电弧多模态目标检测框架

多模态传感器融合的弓网电弧检测框架如图 1 所示, 通过光学、热学与声学信息的联合感知与融合识别, 多模态采集装置由可见光相机、热红外相机及麦克风阵列组成, 分别采集光学、热学与声学信号。

弓网电弧多模态目标检测系统安装于列车车顶, 其中, 检测系统最重要的部分为多模态数据采集装置, 其中包括采集可见光图片的相机、红外热成像的传感设备以及电弧声学信号的麦克风阵列。由于各传感器在成像机制与数据格式上存在差异, 随后需进行多模态数据预处理: 可见光与热红外图像经预训练的语义引导所有过程的网络模型(semantics to lead all, SemLA)<sup>[23]</sup>进行像素级配准, 确保空间一致性; 麦克风阵列数据经声学信号处理转化为声成像图像。

在目标识别阶段, 检测系统构建了基于 transformer 端到端的目标检测架构(detection transformer, DETR)<sup>[24]</sup>的多模态目标检测模型, 包括 50 层残差神经网络(50 layer residual neural network, ResNet-50)、高效混合编码器、Top-K 模态选择模块(Top-K Modality Selection module, TKMS)及多模态可变形交叉注意力机制。该模型通过多尺度特征提取与跨模态关联建模, 以增强电弧特征的表征能力。

### 1.1 可见光和红外图像处理

由于可见光与热红外相机在分辨率、视角及成像角度上的差异, 同一目标在两幅图像中的空间映射常存在偏移。本文采用基于深度学习的预训练配准模型 SemLA, 实现可见光与热成像图像的自动化空间对齐。

首先, 利用轻量级网络提取两类图像的注册与语义感知特征; 随后, 通过跨模态语义校准(cross-modal semantic calibration, CSC)模块自适应调整红外图像的语义特征, 并生成重新校准的红外语义图; 再由语义结构表示(semantic structure representation, SSR)模块聚合两类语义图在不同空间位置的概率分布; 最后, 语义引导特征匹配(semantic-guided feature matching, SFM)模块对可见

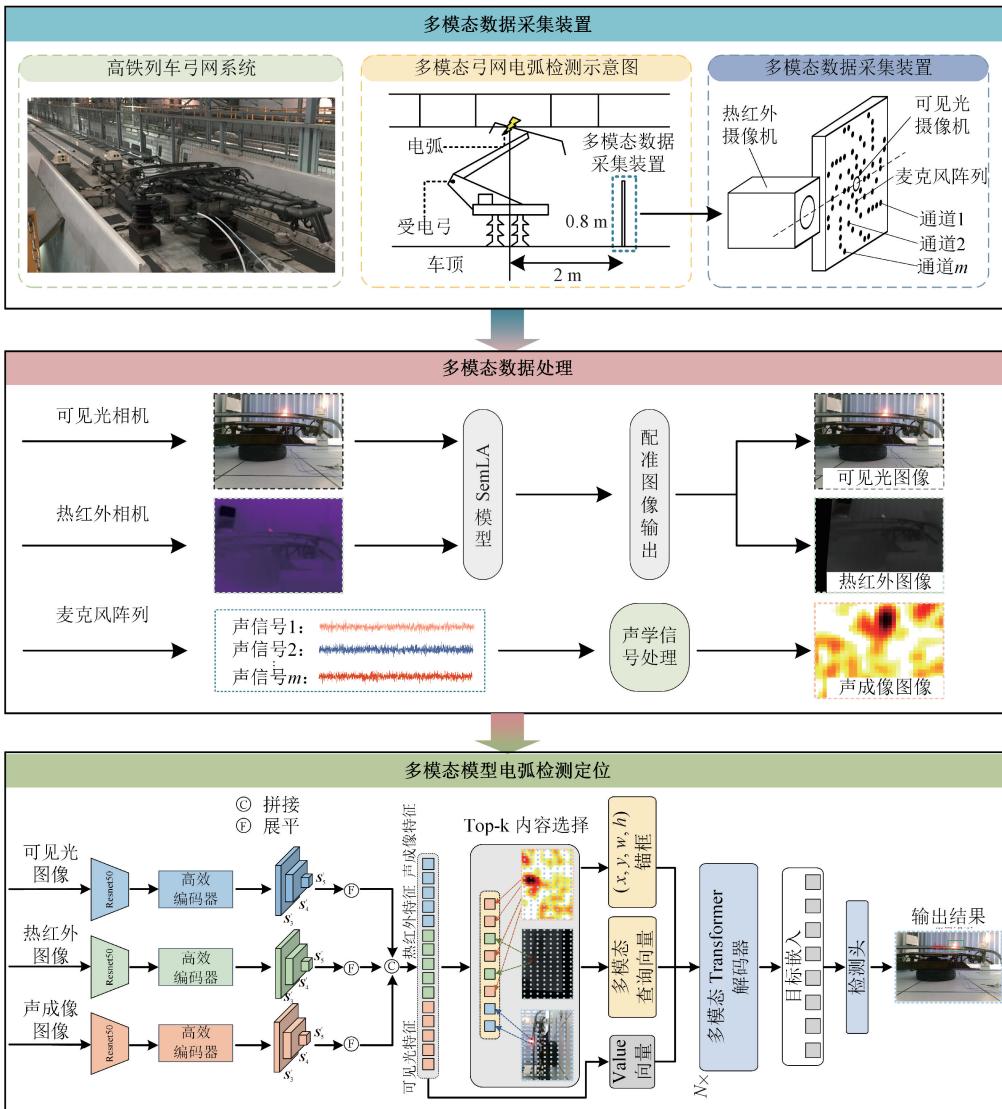


图 1 弓网电弧多模态目标检测系统框架设计

Fig. 1 Framework of multi-modal target detection system using pantograph-catenary arc

光特征、热成像特征及红外语义图进行精确匹配,从而得到空间对齐的红外图像。

## 1.2 声学数据处理

检测系统声学数据处理流程,如图 2 所示。该流程主要包括声学矩阵表达、噪声抑制与信号分离以及声学成像 3 个阶段。

第 1 阶段是声学矩阵表达。声学信号的传播可视为一个或多个辐射源在空间中产生的声场,并通过麦克风阵列在不同位置进行接收与观测。假设声场由一个辐射源在声源位置  $x$  处产生,麦克风阵列在若干离散位置  $y_i (i = 1, 2, \dots, M)$  进行测量,则阵列第  $i$  个通道在频率  $f$  下接收到的声压信号如式(1)所示。

$$\mathbf{p}(y_i, f) = \int_{S_x} s(x, f) G(y_i, f | x) dS_x + \mathbf{e} \quad (1)$$

其中,  $S_x$  表示声源表面,  $y_i$  表示第  $i$  个麦克风位置,  $x$  是声源位置,  $\mathbf{e}$  表示测量噪声,  $s(x, f)$  表示声源  $x$  的频域辐射强度。

利用该声学传播特性进行对采集的多通道声压信号进行处理,如图 2 中的声学矩阵表达流程所示,第  $m$  通道的信号时域声压信号被分成  $N$  个快照 ( $M < N$ )。其中每个快照的长度相等,相邻快照之间有重叠部分,从而获得更多快照。接着添加汉明窗,对每个快照进行傅里叶变换,将需要分析的频率的复数值保留在矩阵中。从而求得如式(1)中的复数值  $\mathbf{p}(y_i, f)$ ,再将  $N$  个快照的复数值拼接为一个行向量,再将  $M$  个麦克风阵列通道的行向量拼接成一列,最后形成时频矩阵  $\mathbf{P}_{if} \in \mathbb{C}^{M \times N}$ ,如式(2)所示。

$$\mathbf{P}_{if} = \mathbf{L}_{if} + \mathbf{E}_{if} \quad (2)$$

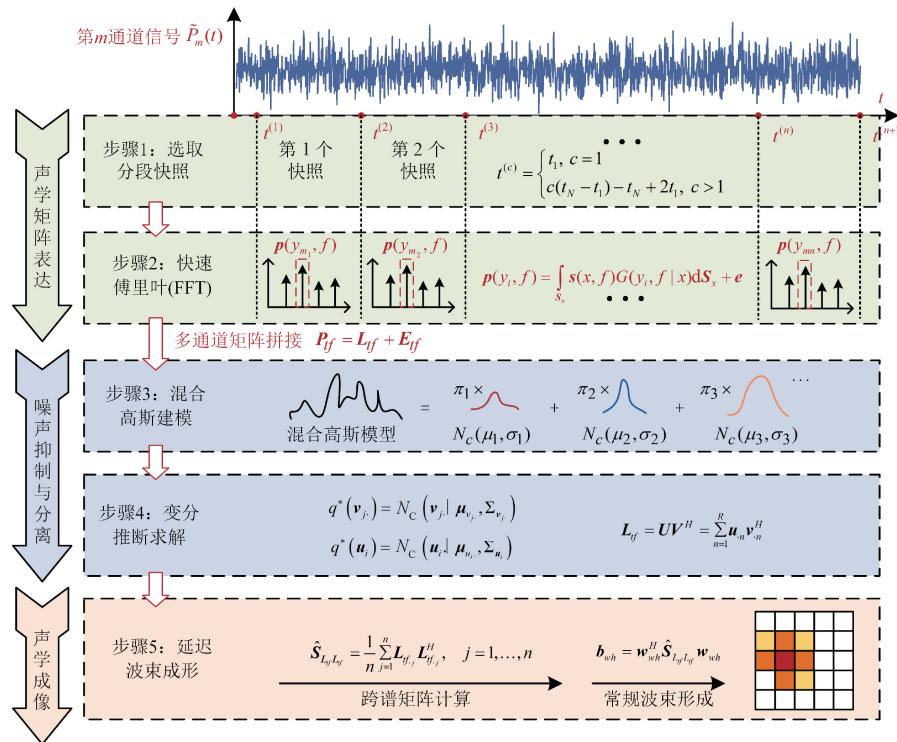


图 2 声学信号处理过程

Fig. 2 Acoustic signal processing procedure

其中,  $\mathbf{P}_f \in \mathbb{C}^{M \times N}$  表示为激励频率  $f$  的测量声压矩阵,  $\mathbf{L}_f \in \mathbb{C}^{M \times N}$  表示为声源信号矩阵,  $\mathbf{E}_f \in \mathbb{C}^{M \times N}$  表示为背景噪声干扰矩阵。

第 2 阶段是噪声抑制与分离。在获取频域观测矩阵  $\mathbf{P}_f$  后, 需要从中有效分离出真实的声源成分。由于声学测量过程中普遍存在复杂的背景噪声和非平稳干扰, 传统的单分布噪声模型往往无法准确拟合观测数据。为此, 本研究采用高斯混合模型对背景噪声矩阵  $\mathbf{E}_f$  进行统计建模, 从而提升信号去噪的精度与鲁棒性。

为了获得参数的最优估计, 本研究采用变分贝叶斯推断方法, 通过最小化 KL 散度实现对后验分布的近似。因此根据平均场理论重写为因式分解形式以促进求解, 如式(3)所示。

$$Q(\boldsymbol{\theta}) = q(\mathbf{U})q(\mathbf{V})q(\boldsymbol{\gamma})q(\mathbf{Z})q(\boldsymbol{\mu})q(\boldsymbol{\tau})q(\boldsymbol{\pi}) \quad (3)$$

$\boldsymbol{\mu}_k, \boldsymbol{\tau}_k$  的变分后验形式如式(4)所示。

$$q^*(\boldsymbol{\mu}_k | \boldsymbol{\tau}_k) = N_c(\boldsymbol{\mu}_k | \mathbf{m}_k, (\boldsymbol{\beta}_k \boldsymbol{\tau}_k)^{-1}) \times$$

$$\text{Gam}(\boldsymbol{\tau}_k | \mathbf{c}_k, \mathbf{d}_k) \quad (4)$$

$\boldsymbol{\pi}$  的变分后验形式如式(5)所示。

$$q^*(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}) \quad (5)$$

由于真实声源信号矩阵  $\mathbf{L}_f \in \mathbb{C}^{M \times N}$  具有显著的低秩特性, 因此  $\mathbf{L}_f$  可以表示为  $\mathbf{U} \in \mathbb{C}^{M \times N}$  和  $\mathbf{V} \in \mathbb{C}^{M \times N}$  的乘积, 表达公式如式(6)所示。

$$\mathbf{L}_f = \mathbf{U}\mathbf{V}^H = \sum_{n=1}^R \mathbf{u}_{..n} \mathbf{v}_{..n}^H, \quad R \leq M \quad (6)$$

其中,  $\mathbf{U} \in \mathbb{C}^{M \times N}$  和  $\mathbf{V} \in \mathbb{C}^{M \times N}$  分别为左右奇异向量矩

阵,  $R > r$ ,  $\mathbf{u}_{..n}$  为  $\mathbf{U}$  的第  $n$  列,  $\mathbf{v}_{..n}$  是  $\mathbf{V}$  的第  $n$  列。进一步地,  $\mathbf{u}_{..i}$  与  $\mathbf{v}_{..i}$  的后验分布可以近似如式(7)、(8)所示。

$$q^*(\mathbf{u}_{..i}) = N_c(\mathbf{u}_{..i} | \boldsymbol{\mu}_{u_{..i}}, \Sigma_{u_{..i}}) \quad (7)$$

$$q^*(\mathbf{v}_{..j}) = N_c(\mathbf{v}_{..j} | \boldsymbol{\mu}_{v_{..j}}, \Sigma_{v_{..j}}) \quad (8)$$

$\boldsymbol{\gamma}_r$  的变分后验形式如式(9)所示。

$$q^*(\boldsymbol{\gamma}_r) = \text{Gam}(\boldsymbol{\gamma}_r | \mathbf{a}_r, \mathbf{b}_r) \quad (9)$$

模型通过迭代优化对  $\mathbf{U}$ 、 $\mathbf{V}$  的后验参数, 当满足如式(10)所示的收敛条件时停止迭代, 即:

$$\| \mathbf{L}_{tf-new} - \mathbf{L}_{tf-old} \|_F < \sigma \quad (10)$$

第 3 阶段是声学成像。在获得去噪后的声源信号矩阵后, 将探测的平面区域划分成网格, 指定某一频率通过波束形成算法扫描每个网格点并计算出网格点上的声功率。以阵列的中心点为原点建立三维笛卡尔坐标系, 麦克风分布在原点的  $x$ - $y$  平面上。假设扫描平面与阵列的距离为  $z$ , 然后将扫描区域划分为  $w \times h$  个网格, 扫描平面一共有  $(w+1) \times (h+1)$  个网格点。通过对每个通道麦克风进行延时补偿接收过程中产生的时间差, 使得各个通道的声信号同步, 然后再经过加权求和输出最大值。跨谱矩阵 (cross spectral matrix, CSM)<sup>[25]</sup> 的引入使得波束形成求解更加方便灵活。基于 CSM 进行波束形成, 去噪后的时频矩阵  $\mathbf{L}_f$  的由傅里叶系数的协方差矩阵定义, 并通过对  $n$  快照进行平均估算, 具体计算如式(11)所示。

$$\hat{\mathbf{S}}_{L_y L_f} = \frac{1}{n} \sum_{j=1}^n \mathbf{L}_{tf,j} \mathbf{L}_{tf,j}^H, \quad j = 1, \dots, n \quad (11)$$

其中,  $H$  是 Hermitian 转置算子,  $\hat{\mathbf{S}}_{L_{if} L_{if}}$  表示为估计量,  $\mathbf{L}_{if,j}$  表示  $\mathbf{L}_{if}$  矩阵的第  $j$  列。波束形成计算如式(12)所示。

$$\mathbf{b}_{wh} = \mathbf{w}_{wh}^H \hat{\mathbf{S}}_{L_{if} L_{if}} \mathbf{w}_{wh} \quad (12)$$

其中,  $\mathbf{b}_{wh}$  表示扫描的第  $w$  行  $h$  列点网格输出的声源功率;  $\mathbf{w}_{wh}$  表示为权重矢量。

## 2 电弧声信号特性分析

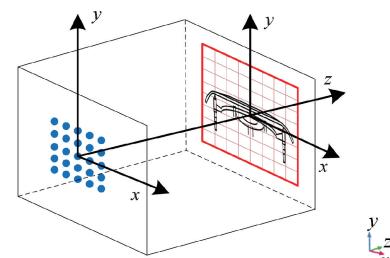
### 2.1 声信号模型搭建

采用有限元方法分析弓网系统中声学信号的传播规律, 如图 3 所示。图 3(a) 中麦克风阵列距离受电弓电弧声源点为 2 m, 均匀阵列中心与坐标原点重合, 扫描平面是一个 2 m 的正方形平面, 分为 900 个 0.06 m 大小的网格点。每个网格点代表一个等效源, 声平面与麦克风阵列平面平行。有限元模型如图 3(b) 所示, 左边点为模拟麦克风阵列的信号采集点, 麦克风阵列形状为均匀阵列, 直径大小为 0.8 m。用单极子源模拟弓网处的电弧声信号  $y(t)$ , 如式(13)所示。

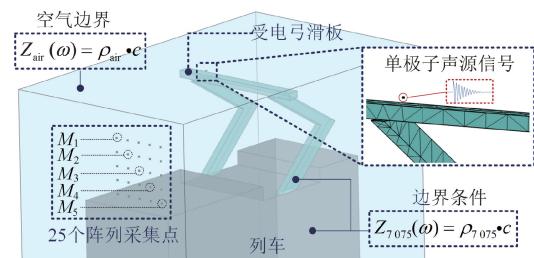
$$y(t) = 4 \times \exp(-5 \times (t - 2T_0)^2 / (T_0/2)) \times \sin(2 \times \pi \times f_0 \times (t - 2 \times T_0)) \quad (13)$$

其中, 声源中心频率  $f_0$  为 700 Hz,  $T_0$  表示中心频率的信号周期,  $t$  为时间。

随着声信号从模拟弓网电弧声源点向外传播, 由麦克风阵列的接收测量声压。其中, 利用简化的局部反应近似模型来模拟受电弓滑板、受电弓和车顶的边界声反射。



(a) 模型简易示意图  
(a) Diagram of the model



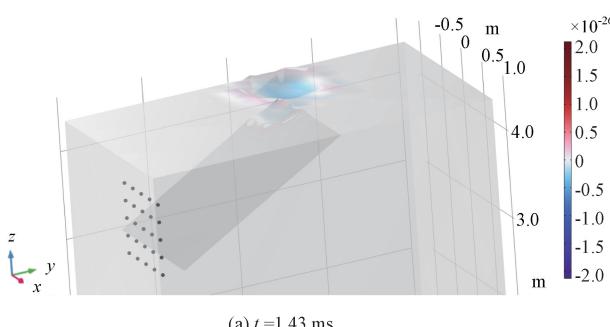
(b) 声信号传播模型  
(b) Acoustic signal propagation model

图 3 声信号模型示意图

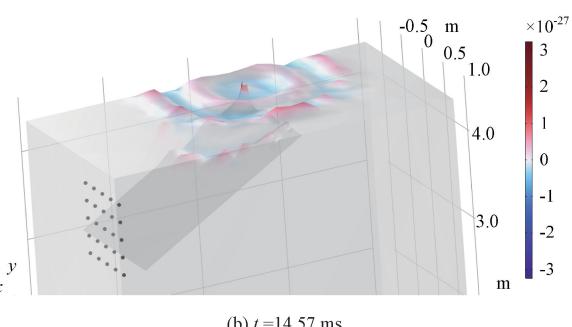
Fig. 3 The illustration of the acoustic signal model

### 2.2 模型结果分析

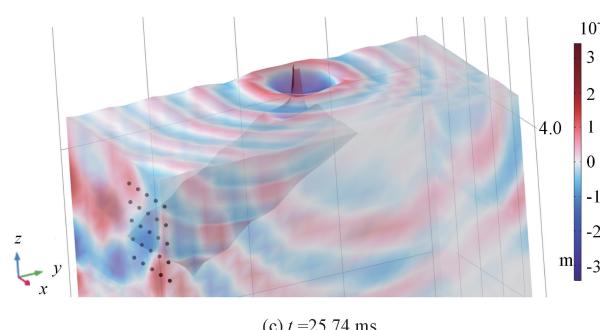
图 4 为模拟弓网电弧声源点到阵列采集点之间的三维平面声场能量图。图 4(a)、(b) 和(c) 分别对应时间为 1.43、14.57 和 25.74 ms。



(a)  $t = 1.43$  ms



(b)  $t = 14.57$  ms



(c)  $t = 25.74$  ms

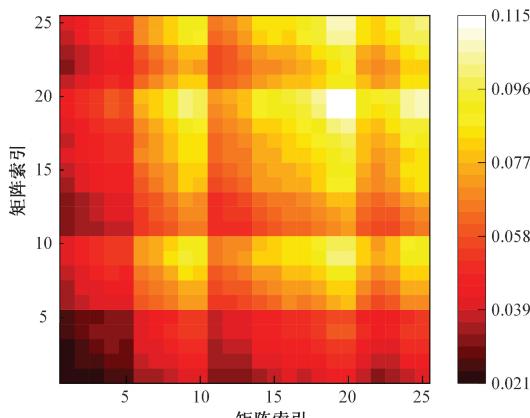
图 4 声源点到阵列采集点之间的三维声场能量图

Fig. 4 Three-dimensional sound field energy map from the sound source point to the array measurement point

通过图 3(b)的声信号模型,所设计的麦克风阵列能够得到 25 个通道的声学时域信号,由于电弧声在传播过程中产生的时间差,而波束成形理论也是根据这一特性进行声源定位。以图 3(b)中的  $M_1$ 、 $M_2$ 、 $M_3$ 、 $M_4$  和  $M_5$  这 5 个采集点时域信号为例,各采集点存在一定的时间差以及峰值差,如图 5 所示。

随后,基于采集的声信号,本文基于去噪模型以及声学成像的算法验证。如图 6 所示。

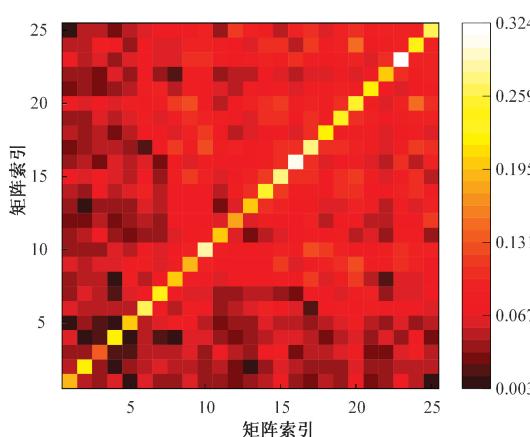
图 6(a)为模型采集信号(未施加噪声)计算的 CSM 能量图,图 6(b)为使用该信号进行波束成形的图像。通过施加随机噪声、高斯噪声和拉普拉斯噪声,使得加噪后信号的信噪比为 -5 dB,具体的结果 CSM 能量图如图 7(a)所示。可以看出噪声污染后的 CSM 矩阵的能量进行提升,同时矩阵内的特征点也逐渐被覆盖,声源信息被污染。通过第 2 阶段中的噪声抑制与分离方法处理后,如图 7(b)所示,未受污染 CSM 的特征信息得以保留,重建 CSM 的振幅复现度高。



(a) 原始信号CSM能量图  
(a) Energy map of the raw signal CSM

图 6 未加噪信号的 CSM 能量图像与波束成形图像

Fig. 6 CSM and beamforming image without adding noise signal



(b) 原始信号波束成形  
(b) Original signal beamforming

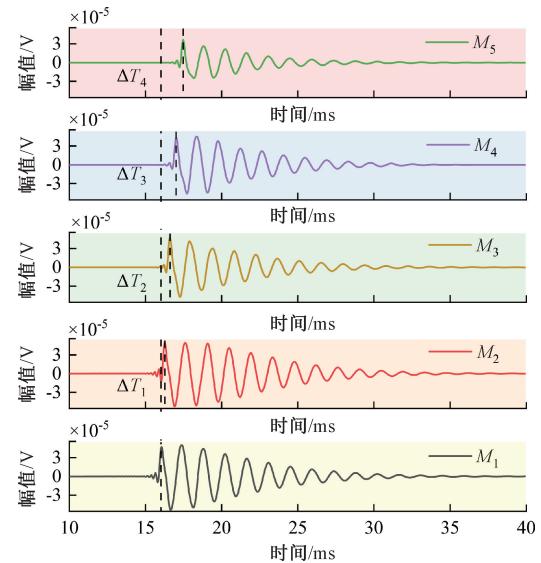
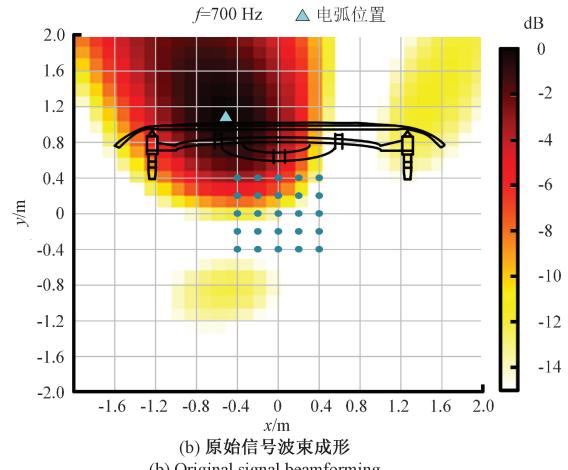
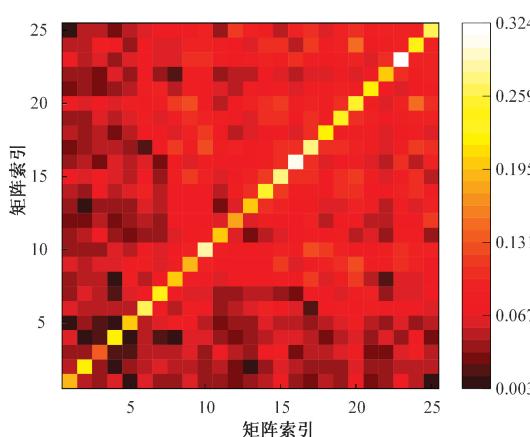


图 5 模型产生的部分声学信号

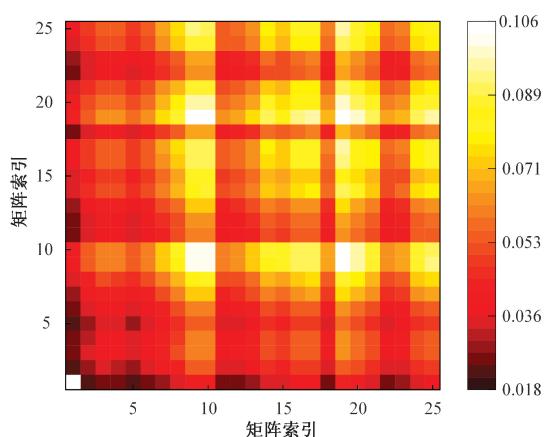
Fig. 5 Acoustic signals generated by the model



(b) 原始信号波束成形  
(b) Original signal beamforming



(a) 噪声污染后的CSM能量图  
(a) CSM energy map with noise contamination

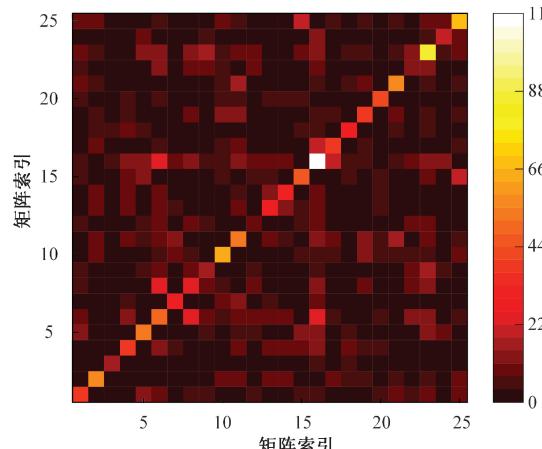


(b) 经过去噪处理后的CSM能量图  
(b) CSM energy map after denoising processing

图 7 -5 dB 背景噪声环境下 CSM 能量图对比

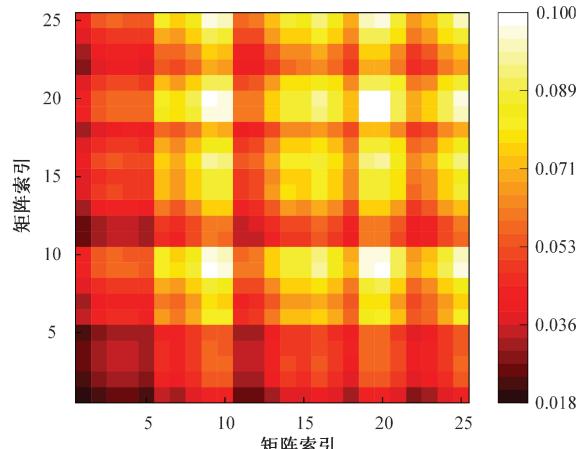
Fig. 7 CSM energy map under a -5 dB background-noise condition

随着加噪程度不断加强,施加噪声后信号信噪比达到了-27 dB,此时噪声污染后的CSM的能量出现大幅度提升,如图8(a)所示。通过噪声抑制与分离方法处理后,如图8(b)所示,特征信息被重建,重建后幅值复原度高。



(a) 噪声污染后的CSM能量图  
(a) CSM energy map with noise contamination

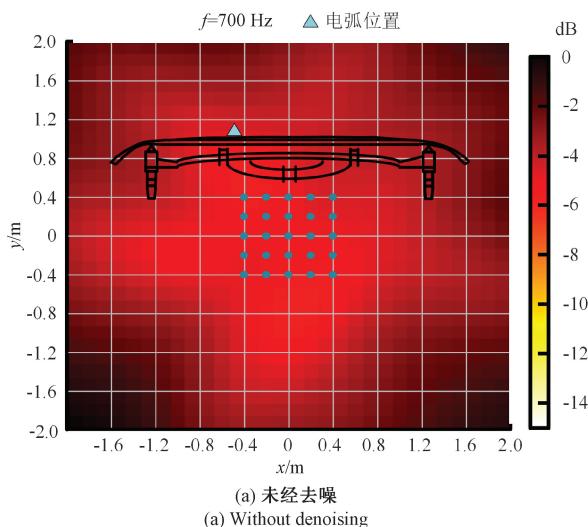
为了更加直观地展示噪声抑制与分离方法的效果,将原始信号、加噪信号(-27 dB)和去噪信号的CSM进行了波束成形,如图9所示。可看出图9(b)和6(b)中的波束成形图像保持一致,如对加噪后的波束成形进行成像,如图9(a)所示,则无法分辨出声源的区域。



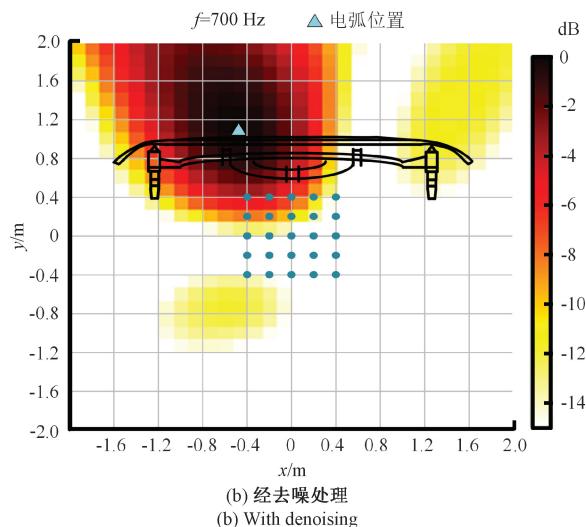
(b) 经过去噪处理后的CSM能量图  
(b) CSM energy map after denoising processing

图8 -27 dB 背景噪声环境下 CSM 能量图对比

Fig. 8 CSM energy map under a -27 dB background-noise condition



(a) 未经去噪  
(a) Without denoising



(b) 经去噪处理  
(b) With denoising

图9 -27 dB 背景噪声环境下的波束成形图

Fig. 9 Beamforming map under a -27 dB background-noise condition

通过仿真模型验证了,采取了CSM能量图和波束成形的方式能够进行算法的可视化,噪声抑制与分离算法能够在极低信噪比的条件下进行稳定去噪,而且去噪前后效果基本保持一致。

### 3 实验验证与分析

#### 3.1 实验平台搭建及数据集构建

实验装置与参数设置如表1所示。为实现高精度多

模态数据采集,实验将扫描平面划分为900个网格,网格尺寸随阵列与声源距离自适应调整,以便实现声学与视觉信号的空间对齐。受电弓碳滑板上安装尺寸为65 mm×40 mm×25 mm的电弧激发装置,用于模拟电弧放电行为,设置5个采样点以获取不同空间位置下的电弧信号特征。每次实验连续采集5 s数据,在1、1.5、2 m这3种不同距离获得15组多模态数据。

经时间同步与几何校准后,使用 Albumentations 工具进行数据增强,每组扩充200张,最终形成约3 000组

表 1 多模态弓网电弧数据采集的实验装置

Table 1 Experimental setup for multimodal pantograph catenary arc data acquisition

麦克风阵列排布	实验设备布局	实验参数设置	
		可见光分辨率	1 920 pixels×1 080 pixels
		热红外分辨率	160 pixels×120 pixels
		声源-阵列距离	1、1.5、2 m
		信号采样频率	192 kHz
		采集电弧位置	5

多模态样本。数据集按 8:2 比例划分为训练集与验证集, 以保证模型的泛化能力。为训练检测模型, 对 RGB 图像中的电弧区域采用 LabelImg 进行人工标注, 生成 COCO 格式标注文件。

### 3.2 声学去噪结果对比

图 10 展示了实验室条件下获取的电弧阵列信号声学成像结果。成像频率设为 20 kHz。声源位于受电弓滑板右侧, 由电弧激发装置产生, 对实验信号加噪声后进行

去噪, 可看出 3 组不同距离下声成像图。

进一步分析, 当成像距离为 1 m 时, 成像效果最佳, 声源聚焦清晰; 当距离增加至 1.5 和 2 m 时, 图像中出现伪声源区域。这种现象主要原因是麦克风阵列与受电弓距离较远, 产生了声波反射, 因此产生了伪声源干扰。

为验证声源定位的准确性, 本文将波束成形图与可见光图像进行融合, 实现声源位置的可视化对比。如图 11 所示。

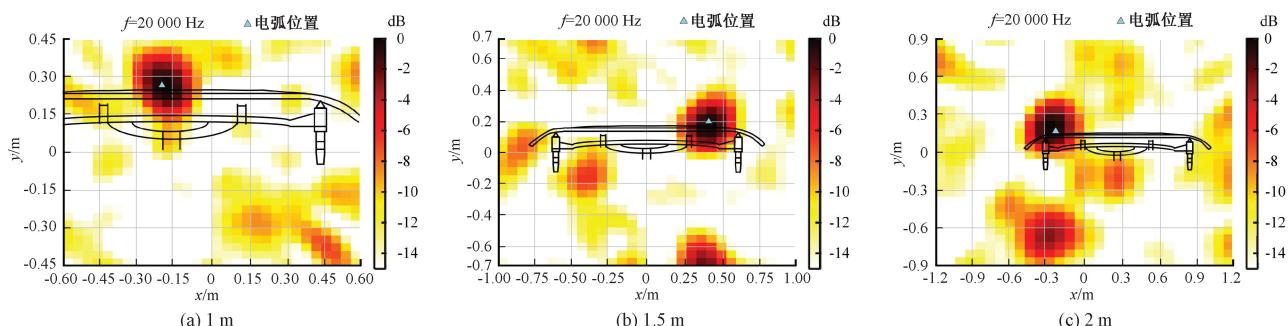


图 10 不同距离下实验信号的波束成形可视化

Fig. 10 Beamforming visualization of experimental signals at different distances.

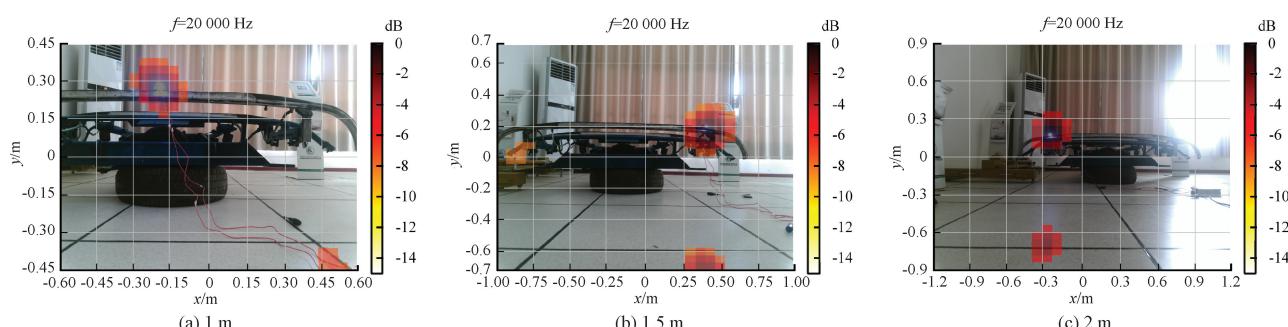


图 11 不同距离下实验信号的波束成形可视化

Fig. 11 Beamforming visualization of experimental signals at different distances.

展示了成像距离为 1、1.5 和 2 m 时的融合结果。通过提取波束成形声源区域特征,以透明化方式叠加至可见光图像,两者在空间位置上一致,声源定位点与实际电弧位置精确重合,验证了所提定位方法的可靠性与精度。

### 3.3 多模态模型性能对比

以 mAP (mean average precision) 作为主要精度指标。基于自建的单类目标(电弧)多模态图像数据集,在服务器端分别训练和测试单模态、双模态及三模态模型,比较不同模态融合方式的检测性能。训练配置如表 2 所示。

表 2 计算机资源配置

Table 2 The computer resource configuration

计算机资源	型号
CPU	Inter Core i7-14700K
GPU	NVIDIA RTX 4090D
GPU 数量	1
Memory	128 G
API	PaddleDetection

优化器采用解耦权重衰减改进的 Adam 优化器(adam with decoupled weight decay, AdamW),权重衰减系数与初始学习率均设为  $10^{-4}$ ,并使用线性学习率热身策略以避免初期不稳定。训练共 50 个轮次,批量大小为 4。

基于该数据集,本文对所提方法进行评估,并与单模态(VIS、TIR)、双模态(RGT、RGBT、GBT)及三模态(RTC、GTC、BTC)模型进行对比。所有模型均在相同的数据集与超参数下,基于 DETR 检测框架训练双模态 RGT 模型输入为可见光 R、G 通道与热成像 T 通道;三模态 RTC 模型输入为可见光 R 通道、热成像 T 通道及声成像 C 通道;RGBT 模型输入为 3 通道可见光与 3 通道热成像。

表 3 结果所示,多模态模型在各阈值下的检测精度均显著优于单模态与双模态模型。所提三模态方法的 mAP@0.5、mAP@0.75 及 mAP@0.5:0.95 分别达到 97.8%、93.5% 和 84%。相较于单模态可见光与热红外模型,mAP@0.5:0.95 分别提升 15.9% 与 20.8%,较双模态 RGBT 模型提升 8.1%。这表明多模态融合能够充分发挥不同信息源的互补优势,显著提高电弧检测的精度与鲁棒性。可见光模型易受光照与环境干扰,热红外模型则受遮挡和噪声影响,而多模态方法通过融合光学、热成像与声学特征,有效克服单一模态的局限。

表 3 模型的准确度评估对比

Table 3 The comparison of accuracy

检测模态	模型	模态					骨干网络	mAP		
		R	G	B	T	CBF		0.5:0.95/%	0.75/%	0.5/%
单模态	VIS	✓	✓	✓			Resnet50	68.1	74.6	91.4
	TIR				✓		Resnet50	63.2	65.1	95.2
双模态	RGT	✓	✓		✓		Resnet50	70.2	79.3	96.1
	RBT	✓		✓	✓		Resnet50	70.7	79.6	96.1
三模态	GBT		✓	✓	✓		Resnet50	69.4	78.5	95.7
	RGBT	✓	✓	✓	✓		Resnet50	75.9	87.6	96.9
	RTC	✓			✓	✓	Resnet50	72.1	82.4	96.9
	GTC		✓		✓	✓	Resnet50	73.2	84.5	96.9
	BTC			✓	✓	✓	Resnet50	73.6	83.4	96.6
	本文	✓	✓	✓	✓	✓	Resnet50	<b>84.0</b>	<b>93.5</b>	<b>97.8</b>

在双模态模型的比较中,RGBT 模型相较于 RGT、RBT 和 GBT 等传统通道融合模型表现更优,其 mAP@0.5:0.95、mAP@0.75 和 mAP@0.5 分别提升 5.2%、8.0% 和 0.8%。结果表明,RGBT 模型通过融合可见光 R、G、B 通道与热成像 T 通道,能更有效提取和整合多模态特征,从而提升电弧检测精度。相比之下,传统通道融合模型未能充分挖掘各模态特征潜力,在复杂背景噪声、反射光和遮挡条件下表现较弱。

在三模态模型中,由于引入声成像模态,检测性能进一步提升。与 RTC、GTC 和 BTC 模型相比,所提模型在 mAP@0.5:0.95、mAP@0.75 和 mAP@0.5 上分别提高 10.4%、9.0% 和 0.9%。这一结果表明,声学信息的加入显著增强了模型对环境干扰的抵抗力,改善了特征融合效果。

综上,三模态模型在检测精度与鲁棒性方面均优于单模态和双模态方法。多模态融合策略有效整合可见

光、热红外及声学特征,显著提升了电弧检测的准确性与稳定性。

### 3.4 实验结果对比

如图 12 所示,为验证模型在不同环境条件下的检测

性能,对验证集样本进行了可视化分析。可视化时置信度阈值设为 0.85,红框实线表示模型预测结果,绿框虚线为人工标注。图 12 中展示了单模态、双模态与三模态模型在多种典型工况下的检测表现。

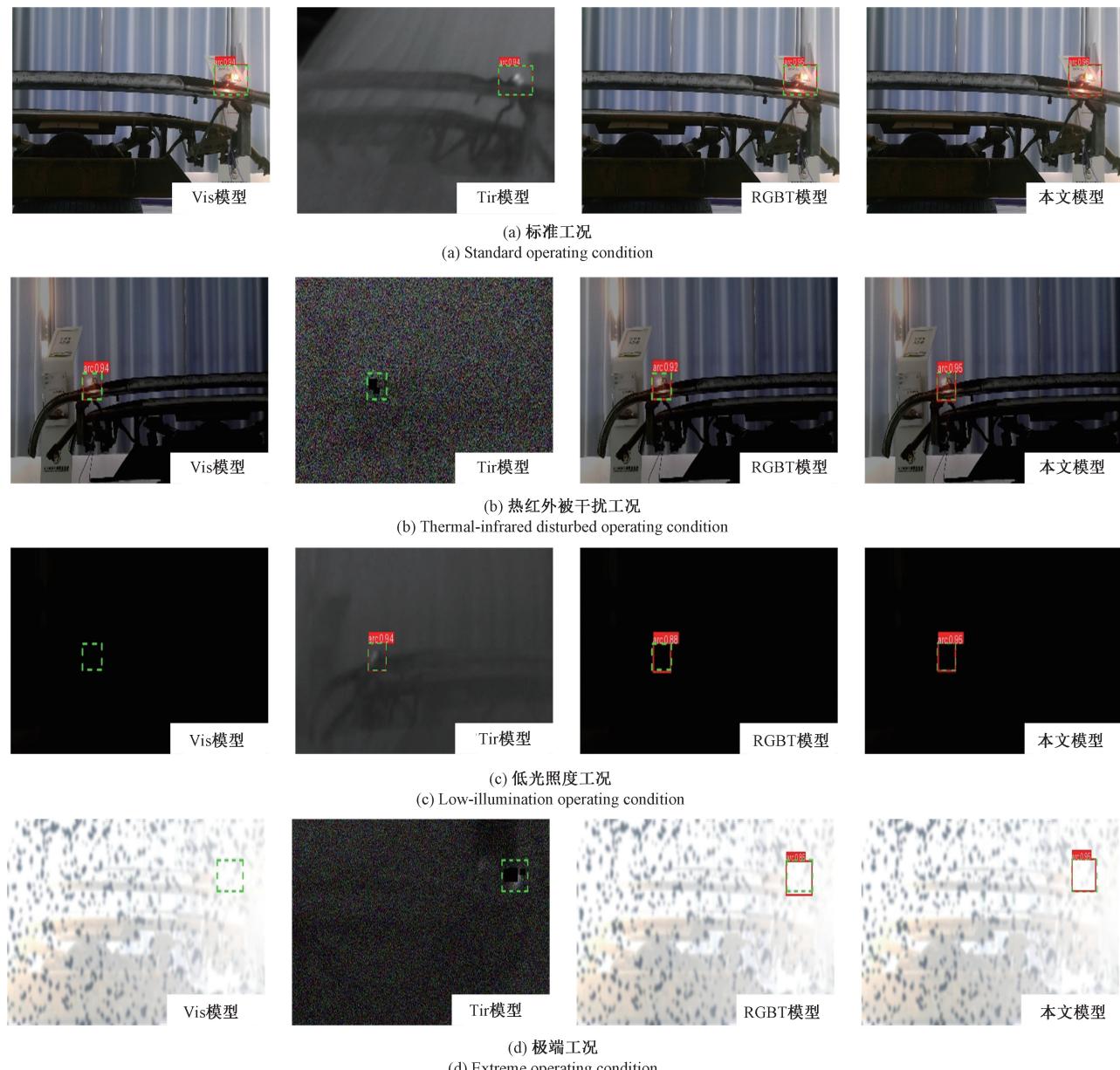


图 12 不同环境下目标检测实验结果

Fig. 12 Experimental results of object detection in various environments.

在标准工况中,如图 12(a)所示,可见光与热红外摄像机均正常工作,三模态模型在声成像信息辅助下的检测置信度较双模态模型提升约 1%,说明声学特征可增强电弧特征稳定性与辨识度。在热红外受强噪声干扰的场景中,如图 12(b)所示,单一热红外模型未能识别电弧,而三模态模型的检测置信度较双模态模型提升约 3%,表

明跨模态融合能显著提升抗干扰能力。在低光照度条件下,如图 12(c)所示,可见光模态失效,三模态模型的检测置信度达 95%,略高于热红外单模态,验证了声学模态在弱信号场景中的增强作用。在极端环境中,如图 12(d)所示,可见光与热红外均受污染或遮挡,单模态检测失败,双模态置信度仅 85%,而三模态模型仍能准确识别

别电弧,置信度达95%。结果表明,三模态模型在多源信息互补下具备显著的鲁棒性与稳定性。

综上所述,三模态模型在多种复杂环境与传感器故障情境下均表现出优异的检测性能,电弧检测置信度始终保持在90%以上,定位精度高且误检率低。

## 4 结 论

针对单一模态数据在列车运行环境复杂多变条件下难以实现高精度弓网电弧检测的问题,本研究首次提出基于多模态融合的弓网电弧检测方法。研究的主要创新点在于:利用声学成像算法实现对弓网电弧声源的空间定位与成像,并设计了基于变分推断的声学阵列去噪流程,有效解决了列车运行环境中高噪声干扰对检测精度的影响。在声学传播特性方面,本文构建了弓网电弧声场仿真模型,对声波传播规律进行了系统分析,并通过仿真信号验证了声学成像算法在高噪声加载条件下的稳定去噪与成像能力。实验结果表明,所提出的多模态检测算法在实验室环境下能够实现对模拟弓网电弧的高精度定位与有效识别,具有良好的可扩展性与工程应用前景,可为弓网系统运行状态监测及故障诊断提供技术强力支撑。

## 参考文献

- [1] 张天龙,魏少伟,魏培勇,等. 基于无人机的铁路外部环境智能巡检技术研究[J]. 铁道运输与经济, 2025, 47(11): 94-102.  
ZHANG T L, WEI SH W, WEI P Y, et al. Intelligent inspection technology for railway external environment based on UAVs [J]. Railway Transport and Economy, 2025, 47(11): 94-102.
- [2] 陈仁祥,王帅,杨黎霞,等. 弓网接触力长短时记忆网络预测的受电弓主动控制与仿真[J]. 仪器仪表学报, 2021, 42(5): 192-198.  
CHEN R X, WANG SH, YANG L X, et al. Active control and simulation for pantograph based on contact force prediction of long short-term memory network [J]. Chinese Journal of Scientific Instrument, 2021, 42(5): 192-198.
- [3] 邢立成,张晓冬,刘彪,等. 弓网电弧温度及导电特性的光谱分析[J]. 光谱学与光谱分析, 2018, 38(3): 890-894.  
XING L CH, ZHANG X D, LIU B, et al. Spectroscopy method used in temperature and electron density of pantograph-catenary arc [J]. Spectroscopy and Spectral Analysis, 2018, 38(3): 890-894.
- [4] YAN Y, LIU H, GAN L F, et al. A novel arc detection and identification method in pantograph-catenary system based on deep learning [J]. Scientific Reports, 2025, 15: 3511.
- [5] SALEH S A, VALDES M E, MARDEGAN C S, et al. The state-of-the-art methods for digital detection and identification of arcing current faults[J]. IEEE Transactions on Industry Applications, 2019, 55(5): 4536-4550.
- [6] JIANG D Z, ZOU H Q, GUO Y K, et al. Simulation on operating overvoltage of dropping pantograph based on pantograph-catenary arc and variable capacitance model[J]. Applied Sciences, 2024, 14(16): 6861.
- [7] 李斌,舒嘉辉,严灵潇,等. 改进黑翅鸢算法的1D-2D-GAF-PCNN-GRU-MSA弓网电弧检测应用[J]. 电子测量与仪器学报, 2024, 38(10): 201-211.  
LI B, SHU J H, YAN L X, et al. 1D-2D-GAF-PCNN-GRU-MSA pantograph arc detection application based on improved black-winged kite algorithm [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(10): 201-211.
- [8] 李斌,娄璟,杜典松. 基于SOA-SVM的弓网电弧识别方法[J]. 电子测量与仪器学报, 2022, 36(10): 83-91.  
LI B, LOU J, DU D S. Pantograph arc recognition method based on SOA-SVM [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(10): 83-91.
- [9] LIU ZH G, ZHOU H Y, HUANG K, et al. Extended black-box model of pantograph-catenary detachment arc considering pantograph-catenary dynamics in electrified railway[J]. IEEE Transactions on Industry Applications, 2019, 55(1): 776-785.
- [10] PAN L K, WANG H CH, YU ZH Q, et al. Electrical characterization of pantograph and catenary offline under train operating conditions based on improved habedank model and at power supply network [J]. IEEE Access, 2024, 12: 174343-174353.
- [11] 张书朝,彭立强,郭阿康,等. 基于改进YOLOv8的受电弓燃弧检测算法[J]. 电子测量技术, 2025, 48(19): 95-105.  
ZHANG SH ZH, PENG L Q, GUO A K, et al. Pantograph arc detection algorithm based on improved YOLOv8 [J]. Electronic Measurement Technology, 2025, 48(19): 95-105.
- [12] 刘冬,闻映红. 现场环境下弓网离线放电辐射测量方法[J]. 电子测量与仪器学报, 2021, 35(1): 88-95.  
LIU D, WEN Y H. In-situ measurement method for radiation of pantograph arcing [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(1): 88-95.
- [13] GAO SH B, WANG Y, LIU ZH G, et al. Thermal distribution modeling and experimental verification of

- contact wire considering the lifting or dropping pantograph in electrified railway [J]. IEEE Transactions on Transportation Electrification, 2016, 2(2): 256-265.
- [14] YU X Y, SU H SH. Pantograph arc detection of urban rail based on photoelectric conversion mechanism [J]. IEEE Access, 2020, 8: 14489-14499.
- [15] WEI W F, LIANG CH L, YANG Z F, et al. A novel method for detecting the pantograph-catenary arc based on the arc sound characteristics [J]. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2019, 233(5): 506-515.
- [16] 周治国, 邸顺帆, 冯新. 语义信息增强的 3D 激光 SLAM 技术进展 [J]. 仪器仪表学报, 2023, 44(3): 209-220.
- ZHOU ZH G, DI SH F, FENG X. Advances in SIE 3D LiDAR SLAM technology [J]. Chinese Journal of Scientific Instrument, 2023, 44(3): 209-220.
- [17] 郭润泽, 孙备, 孙晓永, 等. 无人机弱光条件下多模态融合目标检测方法 [J]. 仪器仪表学报, 2025, 46(1): 338-350.
- GUO R Z, SUN B, SUN X Y, et al. Multimodal fusion object detection method for UAVs under low light conditions [J]. Chinese Journal of Scientific Instrument, 2025, 46(1): 338-350.
- [18] 邹亮, 郭泽沛, 李颖娜, 等. 基于掩码孪生网络与幂调节损失的恶劣场景车道线检测方法 [J]. 仪器仪表学报, 2025, 46(9): 134-145.
- ZOU L, GUO Z P, LI Y N, et al. Robust lane detection in challenging scenarios using a masked siamese network with power-modulated loss [J]. Chinese Journal of Scientific Instrument, 2025, 46(9): 134-145.
- [19] 刘立君, 柳学杨, 兰虎, 等. 声信号辅助同轴视觉激光模具修复飞溅控制 [J]. 焊接学报, 2015, 36(3): 9-13, 1.
- LIU L J, LIU X Y, LAN H, et al. Splash control of laser mould strengthen based on coaxial visual supplemented by acoustical signal [J]. Transactions of the China Welding Institution, 2015, 36(3): 9-13, 1.
- [20] CHU N, HU K Y, HAN H M, et al. Variational bayesian modal composition beamforming for fast-rotating axial-fan blade-noise localization and its application condition [J]. Mechanical Systems and Signal Processing, 2024, 208: 110991.
- [21] RAMACHANDRAN R C, RAMAN G, DOUGHERTY R P. Wind turbine noise measurement using a compact microphone array with advanced deconvolution algorithms [J]. Journal of Sound and Vibration, 2014, 333(14): 3058-3080.
- [22] LI J N, WANG X. Super-resolution localization and orientation estimation of multiple dipole sound sources: From a maximum likelihood framework to wind tunnel validation [J]. Journal of Sound and Vibration, 2025, 595: 118764.
- [23] XIE H SH, ZHANG Y K, QIU J H, et al. Semantics lead all: Towards unified image registration and fusion from a semantic perspective [J]. Information Fusion, 2023, 98: 101835.
- [24] 秦伦明, 张云起, 崔昊杨, 等. 基于改进 RT-DETR 的极端天气下交通标志检测方法 [J]. 电子测量技术, 2025, 48(9): 56-64.
- QIN L M, ZHANG Y Q, CUI H Y, et al. Improved RT-DETR based method for traffic sign recognition in extreme weather [J]. Electronic Measurement Technology, 2025, 48(9): 56-64.
- [25] CHU N, NING Y, YU L, et al. A fast and robust localization method for low-frequency acoustic source: Variational Bayesian inference based on nonsynchronous array measurements [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1-18.

## 作者简介



蔡智超, 2010 年于华东交通大学获得学士学位, 2016 年于河北工业大学获得博士学位, 现为华东交通大学副教授, 主要研究方向为电磁传感器和电磁超声检测。

E-mail: zccai@ecjtu.edu.cn

**Cai Zhichao** received his B. Sc. degree from East China Jiaotong University in 2010, received his Ph. D. degree from Hebei University of Technology in 2016. Now he is an associate professor at East China Jiaotong University. His main research interests include electromagnetic sensors and electromagnetic ultrasonic testing.



孙洪宇(通信作者), 2015 年于华北电力大学获得学士学位, 2018 年于华北电力大学获得硕士学位, 2022 年于清华大学获得博士学位, 现为北京交通大学讲师, 主要研究方向为无损评估、声学超材料、等离子体物理和人工智能。

E-mail: hysun@bjtu.edu.cn

**Sun Hongyu** (Corresponding author) received his B. Sc. and M. Sc. degrees both from North China Electric Power University in 2015 and 2018, and his Ph. D. degree from Tsinghua University in 2022. Now he is a lecturer at Beijing Jiaotong University. His main research interests include nondestructive evaluation, acoustic metamaterials, plasma physics, and artificial intelligence.