

中文引用格式:杨奉展,顾清华,李少博,等. 露天矿低能见度下多模态融合障碍物检测[J]. 中国安全科学学报,2025,35(5):195-203.
英文引用格式:YANG Fengzhan, GU Qinghua, LI Shaobo, et al. Multimodal fusion-based obstacle detection in low-visibility open-pit mines [J]. China Safety Science Journal,2025,35(5):195-203.

露天矿低能见度下多模态融合障碍物检测*

杨奉展^{1,2},顾清华^{1,2}教授,李少博^{1,2},杨建春³ 工程师

(1 西安建筑科技大学 资源工程学院,陕西 西安 710055;2 西安建筑科技大学

西安市智慧工业感知计算与决策重点实验室,陕西 西安 710055;

3 哈密市和翔工贸有限责任公司,新疆 哈密 839200)

中图分类号:X924.3

文献标志码:A

DOI: 10.16265/j.cnki.issn1003-3033.2025.05.1654

基金项目:国家自然科学基金资助(52374135,52074205);陕西省金属矿智能开采理论及技术创新团队(2023-CX-TD-12)。

【摘要】 为解决露天矿区低能见度、低照度环境中无人矿卡对行进障碍物的感知误差问题,减少碰撞风险,提出基于多模态融合的障碍物检测方法;首先,采用轻量粘合(LightGlue)图像配准算法,对齐热红外与可见光不同模态图像的空间,避免融合前的空间错位与几何失真;其次,模态特征提取与融合阶段,在改进的双分支主干网络引入双模态特征融合(DMFF)模块,通过特征压缩、跨模态特征增强,提高提取双模态特征的能力并完成融合;然后,引入迭代学习方法,深入匹配模态间的互补信息,获取双模态特征映射图,提高多模态检测性能;最后,将各尺度融合的特征映射图输入到检测头部,结合边界框回归与分类预测进行精确检测。结果表明:该方法在低能见度等复杂场景下对障碍物的检测效果较好,阈值为0.5时的平均精度均值(mAP@0.5)达到90.8%; F_1 平衡分数达到0.887,较现有方法表现出更高精度和速度,并且误报率和漏检率也较低,能有效提升无人矿卡在复杂环境下的检测精度和安全性。

【关键词】 露天矿; 低能见度; 无人矿卡; 多模态融合; 障碍物检测; 感知预警

Multimodal fusion-based obstacle detection in low-visibility open-pit mines

YANG Fengzhan^{1,2}, GU Qinghua^{1,2}, LI Shaobo^{1,2}, YANG Jianchun³

(1 School of Resources Engineering, Xi'an University of Architecture and Technology, Xi'an Shaanxi

710055, China; 2 Xi'an Key Laboratory of Intelligent Industry Perception Computing and Decision

Making, Xi'an University of Architecture and Technology, Xi'an Shaanxi 710055, China; 3 Hami

City and Xiangkong Trading and Industry Co., Ltd., Hami Xinjiang 839200, China)

Abstract: To address the perception inaccuracies of autonomous mining trucks in open-pit mines under low-visibility and low-illumination conditions—issues that may lead to obstacle collisions. This paper was proposed a multimodal fusion-based obstacle detection method to enhance detection accuracy and operational safety in complex environments. Firstly, local Feature matching at light speed (LightGlue), was employed to achieve spatial alignment between thermal infrared and visible light images, thereby avoiding spatial misalignment and geometric distortion prior to fusion. Secondly, in the modality feature

extraction and fusion stage, a Dual-Modality Feature Fusion (DMFF) module was incorporated into the improved dual-branch backbone network. The extraction capability of dual-modality features was enhanced and fusion was performed through feature compression and cross-modal feature enhancement. An iterative learning method was then introduced to effectively match the complementary information between modalities, generating a fused dual-modality feature map and improving multimodal detection performance. Finally, the fused feature maps at multiple scales were input into the detection head. They were combined with bounding box regression and classification prediction for precise detection. Experimental results demonstrate that the proposed method achieves excellent obstacle detection performance in challenging scenarios with low visibility. Specifically, it achieves a mean Average Precision (mAP@0.5) of 90.8%, and an F_1 -score of 0.887, outperforming existing methods in both accuracy and speed. Moreover, the proposed approach exhibits lower false positive and miss detection rates, effectively ensuring the safe navigation of autonomous mining trucks in complex operational environments.

Keywords: open-pit mine; low visibility; unmanned driving truck; multi-modal fusion; obstacle detection; perceptual warning system

0 引言

露天矿区无人驾驶卡车在复杂环境下的障碍物检测是保障作业安全的关键技术。矿山场景中普遍存在低照度、雨雾天气及粉尘干扰等问题,导致传统检测方法面临挑战:可见光图像在弱光条件下纹理信息模糊,小目标识别能力不足;红外图像虽能在低照度环境保持特征显著性,但分辨率较低且细节表征有限。单一传感器模态难以适应动态环境变化,存在障碍物漏检误检的安全隐患。因此,开展面向矿区低能见度环境的多模态融合检测方法研究具有重要的工程价值。

近年来,多传感器融合技术^[1]已成为解决露天矿区障碍物检测的关键手段,尤其是可见光与热红外相机的融合方法。该方法结合了可见光图像的高分辨率细节与热红外图像的抗干扰特性,提升了检测精度。目前,图像融合技术主要分为传统方法和基于深度学习的融合方法2大类。传统方法通常依赖于人工设计的融合规则,例如:多尺度变换^[2]、稀疏表示^[3]、各向异性扩散融合^[4]、视觉显著性映射结合加权最小二乘法^[5]、多尺度引导的图像与视频融合^[6]等。这些方法虽然具有一定效果,但由于传统的数学变换方法难以有效提取目标特征,未能充分反映源图像的模态差异,容易导致边缘模糊和对比度不足等问题^[7]。

随着深度学习技术的发展,多种方法被应用于图像融合检测任务。SUN Yiming^[8]、YUAN Maoxun^[9]等提出了基于可见光与热红外图像跨模态融合网络的障碍物检测方法,但此类方法在图像

数据匹配、特征不平衡处理及计算效率方面存在局限。为此,WANG Haozhe等^[10]提出了融合多尺度网络,该网络可通过子网络增强弱光条件下的可见光图像多尺度特征,但其增加了模型复杂度和计算开销,难以满足实时性要求。YE Yingchun等^[11]提出了基于双域增强的融合方法,虽提升了热红外与可见光2种模态目标的可识别性,但多针对特定场景,泛化性较差,限制了实际应用。周李兵^[12]、贾桂敏^[13]等提出了基于卷积神经网络的多特征增强融合算法,虽扩展了感受野,提高了局部特征提取,但缺乏全局特征提取能力。

鉴于此,笔者将针对露天矿低能见度环境下无人矿卡障碍物检测难题,提出一种基于双交叉注意力机制的多模态图像融合检测模型,以期提高无人矿卡的安全感知预警能力。

1 多模态融合障碍物检测模型

在露天矿区的夜间低照度、阴雨大雾天气及运输作业产生的粉尘环境下,道路的能见度显著降低。这些不利的环境因素对无人车的安全驾驶构成严峻挑战,尤其是在障碍物检测方面,显著增加系统的复杂度。恶劣环境(如天气/粉尘)导致传统传感器(摄像头、雷达等)感知退化,威胁无人车安全,为克服单模态图像检测的局限性,提高感知预警能力。文中结合目标检测算法^[14]、深度学习模型^[15]及注意力机制等计算机视觉理论,设计基于热红外与可见光融合的多模态障碍检测模型。模型网络结构如图1所示,主要由3部分组成:双分支特征提取网络、双模态特征融合模块(Dual-Modal Feature

Fusion, DMFF) 模块^[16] 和不同尺度的目标检测头模块。双分支网络之间则是 DMFF 模块, 双模态特征融合用数学表达式描述:

$$M_u = \mathcal{F}(M_R, M_T) = \mathcal{F}(\psi_R(I_R), \psi_T(I_T)) \quad (1)$$

式中: M_u 为融合后的特征图; I_R 和 I_T 分别为输入的可见光图像和热红外图像; M_R 为可见光特征图; M_T 为热红外特征图; ψ_R 和 $\psi_T(\cdot)$ 为特征提取函数; $\mathcal{F}(\cdot)$ 为融合函数。

1.1 DMFF 方法

针对露天矿低能见度、低照度等复杂环境下无人矿卡对障碍物检测的影响, 在模型中加入 DMFF 模块, 通过建模全局特征交互, 同时捕获跨模态的互补信息, 引入交叉注意力机制增强了不同对象特征的可辨别性, 进而提高检测精度, 保证无人矿卡的安全行驶与避障。

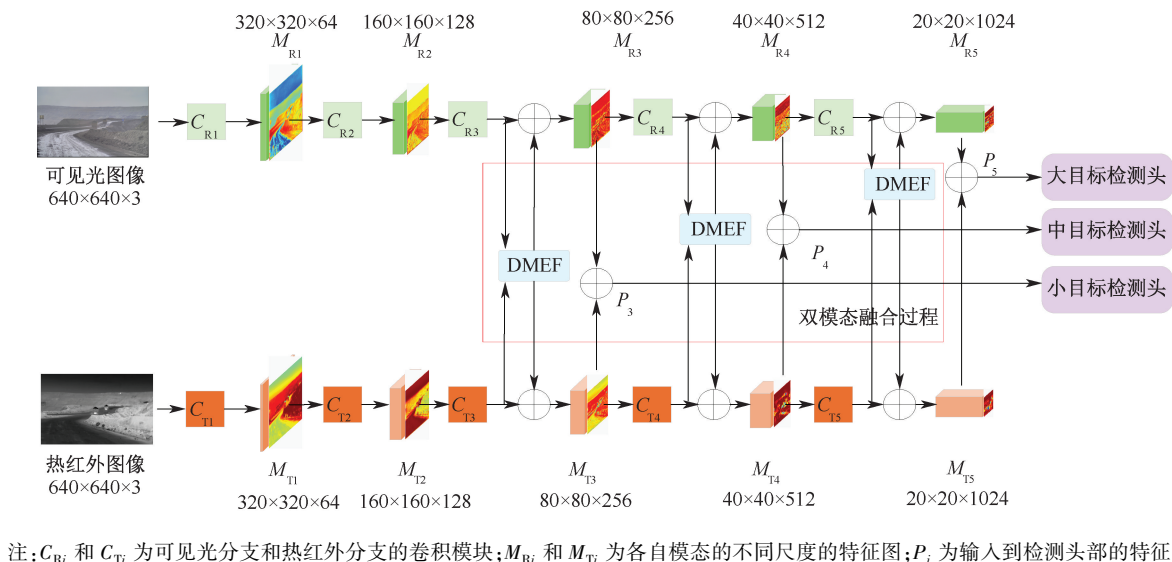
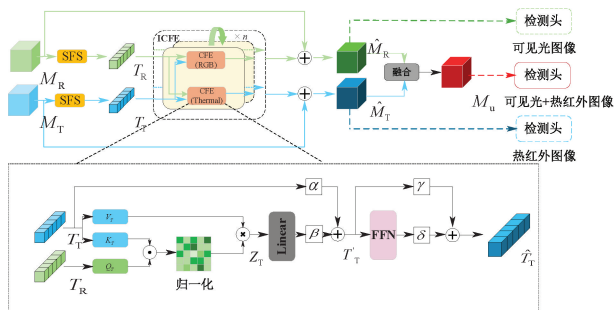


图1 双交叉注意力多模态融合障碍检测模型

Fig. 1 Double-cross attention multimodal fusion disorder detection model

该模块由空间特征收缩 (Spatial Feature Shrinking, SFS) 模块和迭代跨模态特征增强 (Iterative Cross-modal Feature Enhancement, ICFE) 模块 2 部分组成。SFS 模块压缩特征图大小以优化后续处理, ICFE 模块通过双跨模态特征增强 (Cross-modal Feature Enhancement, CFE) 模块细化跨模态特征, 从全局角度在辅助模态中学习更多互补信息, 在加强可见光与热红外模态间互补性的同时, 克服了跨模态特征远程依赖关系建模的不足, 最终通过局部特征融合输出, 具体设计如图 2 所示。

先将每个特征图展平为 1 组标记并添加可学习的位置嵌入, 这是 1 个维度为 $HW \times C$ 的可训练参数, 编码不同标记之间的空间信息。之后可以得到 1 组带有位置嵌入的令牌 $T_R, T_T \in \mathbb{R}^{HW \times C}$ 作为 CFE 模块的输入。由于可见光-热红外图像对通常不完全对齐, 于是采用双 CFE 模块分别获取互补信息, 增强 2 种模态的特征。参数在 2 个 CFE 模块之间共享, 如图 3(底部) 所示, 这里为热红外分支的 CFE 模块:



注: $\hat{M}_R$ 和 $\hat{M}_T$ 为迭代过后的可见光与热红外特征图。

图2 DMFF 模块

Fig. 2 DMFF module

$$\hat{T}_T = \omega - T(\{T_R, T_T\}) \quad (2)$$

$$V_T = T_T W^V, K_T = T_T W^K, Q_R = T_R W^Q \quad (3)$$

式中: T_R 和 T_T 为输入到 CFE 模块的可见光特征和热红外特征标记; $\hat{T}_T$ 为使用 CFE 模块增强的热红外特征; $\omega - T(\cdot)$ 为热红外分支的 CFE 模块; Q_R, V_T, K_T 为特征投影矩阵; W^V, W^K 和 W^Q 为权重矩阵。

首先, 将热红外模态 T_T 的标记投影到 2 个独立

的矩阵 $V_T, K_T \in \mathbb{R}^{H \times W \times C}$ 以计算 1 组键值对,如下式。然后,将可见光模态 T_R 的标记投影到另 1 个单独的矩阵 $Q_R \in \mathbb{R}^{H \times W \times C}$ 以计算 1 组查询。其次,相关矩阵是通过点乘运算构建的,然后使用激活函数 (softmax) 函数归一化相关性分数,表示可见光图像和热红外图像不同特征之间的相似性。之后,通过将相关矩阵与 V_T 相乘得到向量 Z_T ,如下式:

$$Z_T = \text{softmax} \left(\frac{Q_R K_T^T}{\sqrt{D_K}} \right) \cdot V_T \quad (4)$$

$$T'_T = \alpha \cdot Z_T W^0 + \beta \cdot T_T \quad (5)$$

$$\hat{T}_T = \gamma \cdot T'_T + \delta \cdot \text{FFN}(T'_T) \quad (6)$$

式中: D_K 为投影矩阵; FFN (Feed-Forward Network) 为前馈神经网络; W^0 为 FFN 层之前的输出权重矩阵; $\alpha, \beta, \gamma, \delta$ 为不同残差分支的应用可学习系数。

该向量通过利用模态之间的相似性来细化可见光特征,随后借助 FFN 进一步优化全局信息,最终输出增强后的特征。为增强模型对可见光和热红外图像特征相关性的理解,采用具有 8 个平行头的多头交叉注意机制。通过非线性变换将向量 Z_T 投影回原始空间,并通过残差连接^[17] 添加到输入序列中,输出增强后的特征。受到 SReT^[18] 模型的启发,在式(5)和式(6)的残差连接每个分支上应用可学习系数,自适应地从不同分支学习数据,从而提升性能。

1.2 迭代学习模块改进

矿区作业环境复杂,传感器可能受到天气、光照、遮挡等因素的影响,及时检测提高感知预警能力可以避免潜在的安全隐患,确保无人矿卡能迅速检测到障碍物并作出应对决策。尽管用于融合的初始特征图是通过骨干网络下采样得到的,但模型的内存消耗仍然庞大。为了降低后续计算成本,增加网络深度并逐步优化跨模态互补信息,减少参数数量,提升无人矿卡检测的速度,满足实时性。引入了 1 种基于 CFE 的迭代学习模块,称为 ICFE 模块。

如图 3a 所示的传统迭代方法,通过串联堆叠多个块,每个块中没有共享参数,这种显著扩展模型深度的策略不仅增加参数,而且会导致过拟合。而如图 3b 所示的基于 CFE 模块的迭代方法在每一轮迭代中,模型会根据当前融合后的特征进行反向传播,并通过共享参数来改进特征的代表。以 n 次迭代为例,如下式:

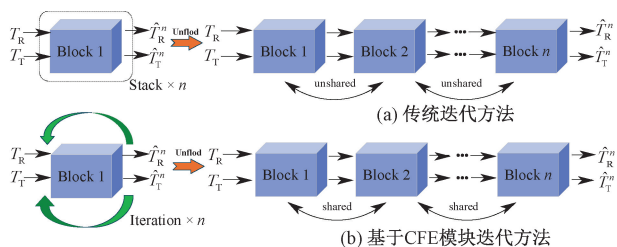


图 3 双 CFE 模块

Fig. 3 Dual CFE module

$$\{\hat{T}_R^n, \hat{T}_T^n\} = \zeta(\{T_R, T_T\}, n) = \underbrace{\omega(\cdots \omega(\{T_R, T_T\}))}_n \quad (7)$$

式中: $\{\hat{T}_R^n, \hat{T}_T^n\}$ 为经过 n 次迭代操作后得到的输出序列; $\{T_R, T_T\}$ 为 ICFE 模块的输入序列; $\zeta(\cdot)$ 为提出的 ICFE 模块。

该模块分别为可见光分支和热红外分支集成了 2 个 CFE 模块。每次迭代操作的输出被用作下 1 次迭代操作的输入,并且每次迭代操作之间的参数是共享的。ICFE 模块的输出序列 $\hat{T}_R^n$ 和 $\hat{T}_T^n$ 首先被转换为特征图,然后通过双线性插值重新校准到特征图的原始大小,形成特征映射最后输入到检测头部如下式所示:

$$[G_l, G_b] = \vartheta_h(\{M'_{R+T}\}_{i=1}^t) \quad (8)$$

式中: G_l, G_b 为分类和回归的损失函数; ϑ_h 为检测头函数; M'_{R+T} 为第 i 层融合特征; t 为卷积层层数。

这些双模态特征图输入到检测头后,显著提高分类精度和边界框定位准确性,提高了不同尺度模态特征的检测能力,增强对各模态大小目标的判别能力,加快模型收敛,提升训练效率。

2 露天矿双模态数据集构建

2.1 双模态图像数据集采集及预处理

文中数据集采集于 2023 年 7 月到 2024 年 6 月的某大型露天矿开采现场,检测目标包括大型运载卡车、其他汽车、行人、挖掘机和装载机 5 类。在不同环境条件下共采集 4 698 张图像数据,其中可见光图像和热红外图像各 2 349 张。

2.2 双模态图像配准算法

由于 2 类图像分别来自不同的传感器,存在几何畸变、角度、分辨率和透视差异,导致图像空间参考不一致,需要提取与匹配 2 类图像特征。为解决双模态图像配准问题,采用 1 种基于深度神经网络的局部特征匹配方法 (LightGlue)^[19]。其步骤如下:

1) 特征点提取与匹配,如图4所示,给定从图像a和b中提取的2组特征点,每个特征点 i 由归一化的2D坐标 $\theta_i:=(x,y)_i \in [0,1]^2$ 和视觉描述子 $d_i \in \mathbb{R}^d$ 组成。图a和b分别有 M 和 N 个特征点,用 $\mathcal{A}:=\{1,2,\dots,M\}$ 和 $\mathcal{B}:=\{1,2,\dots,N\}$ 表示。特征匹配的目标是预测软部分分配矩阵,软分配矩阵的元素二值化,理论上矩阵的每行或每列之和为1,但受视角变化、遮挡、光照等影响,可能存在图a的特征点在图b中无匹配点,导致行或列和不为1。计算点对相似度矩阵 $S \in \mathbb{R}^{M \times N}$ 如下:

$$S_{ij} = L(\mathbf{x}_i^{\mathcal{A}})^T L(\mathbf{x}_j^{\mathcal{B}}) \quad \forall (i,j) \in \mathcal{A} \times \mathcal{B} \quad (9)$$

式中 $L(\cdot)$ 为带偏置的线形变换。相似度可视作该点对源自同一3D点投影的概率,综合考虑特征点的相似度与匹配分数,当2点具备良好匹配关系且相似度显著高于其他点对时,可视为有效关联点。若满足大于阈值且大于所在行列的其他值,则确认为匹配点。最终的分配矩阵 E 可计算为:

$$E_{ij} = \sigma_i^{\mathcal{A}} \sigma_j^{\mathcal{B}} \text{softmax}_{k \in \mathcal{A}}(S_{kj})_i \text{softmax}_{k \in \mathcal{B}}(S_{ik})_j \quad (10)$$

式中: E_{ij} 为图a中特征点 i 与图b中特征点 j 之间的匹配概率; $\sigma_i^{\mathcal{A}}、\sigma_j^{\mathcal{B}}$ 为各点的可匹配性分数。

2) 单应性估计,从相似度矩阵提取匹配点,构建2组点的变换关系,完成图像间映射。单应性描述同一场景下2图像的对应关系,匹配特征点并通过投影优化重合度。以2张图像的匹配点分别是 $X=[x,y]^T$ 和 $X'=[x',y']^T$ 为例,则两者的关系必须满足: $X'=HX$ 。上述关系可以表示为:

$$\begin{pmatrix} H' \\ Y' \\ 1 \end{pmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \quad (11)$$

单应性矩阵 H 由8个参数组成,需4对匹配点即可确定。通过 H 可将1幅图像重投影到另一幅图像的平面进行配准。如图5所示,配准后的图像统一为分辨率1920×1080,确保2类图像空间对齐。

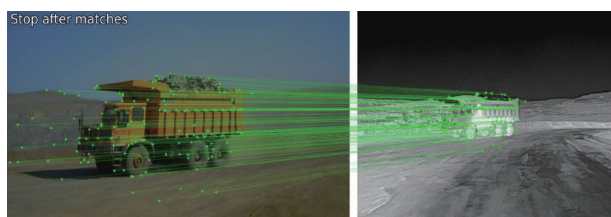


图4 特征点匹配情况

Fig. 4 Matching situation of feature points



图5 双模态融合图像情况

Fig. 5 Situation of dual-modal fusion images

3 多模态融合障碍物检测试验

3.1 试验评估指标

试验采用的评价性能指标有准确率(Precision, P)、召回率(Recall, R)、平均精度均值(mean Average Precision, mAP),参数量(Parameters, Params);其中,mAP@0.5表示在置信度阈值为0.5时的模型对所有类别的mAP均值;mAP@0.5:0.95表示在置信度阈值从0.5到0.95的范围内,以0.05为步长变化时的mAP值,结合 F_1 分数,将其作为无人矿卡安全预警机制的量化评估指标。 F_1 分数是精确率 P 和召回率 R 的调和平均数,较低通常意味着误报率或漏报率较高。高误报率易导致无人矿卡作出错误决策,增加碰撞风险;高漏报则导致障碍物未被及时识别,避障失效引发事故。各指标的计算公式如下:

$$AP = \int_0^1 P(R) dr, mAP = \frac{1}{N} \int_1^e AP_i \quad (12)$$

$$R = \frac{TP}{TP + FN}, P = \frac{TP}{TP + FP} \quad (13)$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (14)$$

式中:TP(True Positive)为预测正确的正样本数量;FP(False Positive)为预测错误正样本数;FN(False Negative)为预测错误负样本数量;AP(Average Precision)为不同召回率水平下平均精确度值; N 为不同目标类别总数; e 为类别序号; i 为类别索引; AP_i 为第 i 个类别平均精度。

3.2 多模态网络障碍物检测效果对比与分析

文中提出的多模态融合障碍物检测模型在复杂环境中表现优异;在低能见度情况下,图像融合帮助模型准确检测障碍物,与主流多模态融合目标检测网络的对比结果见表1。改进模型在统一安全评估和检测指标下,有着90.8%的mAP@0.5和64.4%

表 1 不同多模态网络模型对比

Table 1 Comparison of different multimodal network models

模型	模态	mAP@ 0. 5/ %	mAP@ 0. 5: 0. 95/ %	R/ %	P/ %	Params	F_1
GAFF ^[20]	可见光+热红外	72. 8	37. 4	—	—	23 765 806	—
CFR_3 ^[21]	可见光+热红外	72. 4	36. 1	—	—	64 893 649	—
CFT ^[22]	可见光+热红外	89. 4	62. 7	86. 4	88. 7	206 268 762	0. 875
ICAF-NIN	可见光+热红外	90. 6	60. 9	87. 8	87. 2	189 448 282	0. 874
ICAF-TF	可见光+热红外	90. 2	63. 8	85. 0	89. 0	190 273 534	0. 870
ICAF-ResNet50	可见光+热红外	88. 0	60. 4	82. 3	88. 5	313 810 686	0. 853
文中	可见光+热红外	90. 8	64. 4	86. 9	89. 3	173 866 640	0. 881

注:引导注意力特征融合(Guided Attentive Feature Fusion,GAFF);循环融合的多光谱(Cyclic Fuse-and-Refine, CFR_3);跨模态融合变压器(Cross-Modality Fusion Transformer,CFT);迭代交叉注意引导特征融合(Iterative Cross-Attention Guided Feature Fusion,ICAFusion);ICAF-NIN、ICAF-TF表示ICAFusion网络采用网络嵌套融合(Network In Network Fusion,NIN)和变压融合(Transformer Fusion,TF)2种方式;ICAF-ResNet50表示ICAFusion将骨干网络替换成残差网络(Residual Network 50 layers, ResNet50)后的模型结构。

的 mAP @ 0. 5: 0. 95, 综合召回率 86. 9%、精度 89. 3%, 存储参数量为 173 866 640, 计算资源较少, 响应速度更快, 且 F_1 平衡分数最高, 有着较低的误报率和漏检率。

而 ICAF-ResNet50 和 CFT 模型在矿区复杂环境下虽取得了较高的检测精度, 但较大的参数量导致检测速度较慢, 无法满足无人矿卡在矿区实时的感知预警需求。相比之下, 轻量级多模态网络 GAFF 和 CFR_3 虽然参数量较少, 检测速度较快, 但其检测精度仅为 72. 8% 和 72. 4% 的 mAP@ 0. 5, 导致在复杂环境中目标识别能力较弱, 难以适应矿区的检测需求。ICAF-NIN 与 ICAF-TF, 这 2 种更换了融合方式的 ICAFusion 模型分别取得 90. 6% 和 90. 2% 的 mAP@ 0. 5, 但两者的 F_1 平衡分数都相对较低, 有着较高的误报率和漏检率, 导致模型在实际应用中的可靠性降低。

针对复杂场景下的试验效果如图 6 所示, 文中提出的多模态障碍检测模型分别在这几种复杂环境

下有着良好的综合性能, 其中单模态下可见光图像受粉尘和夜间环境影响, 在不同的低能见度和低照度环境下无法清晰识别障碍物, 出现了漏检情况; 而单模态下的热红外图像因自身分辨率低的特性出现误检和重复识别的情况, 实测结果表明: 改进后的多模态融合模型能够精准识别各类环境下的障碍物, 有效提升无人矿卡在复杂矿区环境中的检测与避障能力, 提供了可靠的安全保障。

3.3 多模态-单模态障碍物检测结果对比

为验证多模态网络的有效性, 借助于主流的单模态网络模型与本模型在露天矿区数据集上进行试验对比。其结果见表 2, 由表 2 可以看出, 单模态网络检测的可见光图像因受矿区复杂多变的环境影响, 整体的检测效果都低于热红外图像, 而文中的多模态网络模型其检测结果有着较高的检测精度, 但因双分支主干网络的原因导致参数量庞大, 相比于轻量化的单模态网络, 有待进一步的提高。

表 2 对比单模态网络

Table 2 Comparison of monomodal networks

模型	模态	mAP@ 0. 5/ %	mAP@ 0. 5: 0. 95/ %	R/ %	P/ %	Params
Faster-RCNN	可见光	64. 5	40. 4	53. 7	52. 3	23 765 806
Faster-RCNN	热红外	73. 8	44. 7	55. 9	61. 8	20 268 762
RT-DETR	可见光	60. 7	40. 2	51. 6	69. 2	32 816 351
RT-DETR	热红外	64. 7	39. 2	62. 0	64. 8	32 816 351
YOLOv8	可见光	77. 0	49. 6	70. 5	83. 2	43 610 463
YOLOv8	热红外	84. 4	61. 4	75. 8	86. 2	43 610 463
YOLOv11	可见光	60. 9	38. 9	59. 0	67. 3	2 590 815
YOLOv11	热红外	57. 9	34. 3	53. 2	68. 7	2 590 815
文中	可见光+热红外	90. 8	64. 4	86. 9	89. 3	173 866 640

注:快速区域卷积神经网络(Faster Region-based Convolutional Neural Network,Faster R-CNN);实时监测变换器(Real-Time Detection Transformer,RT-DETR)。

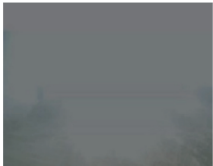
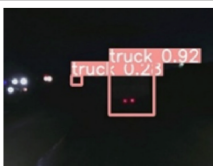
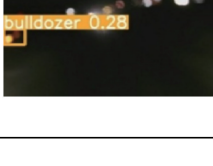
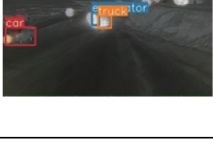
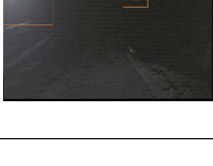
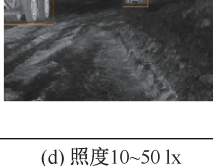
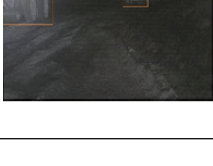
场景	单模态-可见光	单模态-热红外	双模态 R+T
低能见度环境			
			
			
	(a) 能见度50~200 m		
低照度环境			
			
			
	(b) 能见度≤50 m		
正常光照环境			
			
			
	(c) 照度1~10 lx		
正常光照环境			
			
			
	(d) 照度10~50 lx		
正常光照环境			
			
			
	(e) 正常照度		

图 6 复杂环境下不同模态检测效果对比

Fig. 6 Comparison of detection performance across different modalities in complex environments

4 结 论

1) 提出的基于改进 YOLOv8 的双分支架构,通过多模态特征互补融合机制(可见光纹理+热红外辐射特性),在低能见度场景下对障碍物检测 mAP @ 0.5 达 90.8%,较单模态基线提升 21.5%,误报率

降低至 3.2%,满足无人矿卡的实时响应要求。
2) 提出双交叉注意力融合模块,通过通道-空间跨模态交互,有效缓解异源相机分辨率差异与地形起伏导致的空间配准偏差,特征对齐精度提升 28.7%;设计轻量化动态参数共享策略,在参数量仅 17.3 M 条件下,推理速度达 48 帧/s,较主流的多模

态模型提速 2.3 倍,支撑边缘端高效部署。

3) 文中数据集是由实地露天矿区场景采集的,为可见光-热红外图像融合模型提供训练基准,但受不同视觉传感器分辨率差异以及设备时间延迟及

矿区地形起伏干扰影响,部分区域出现空间配准偏差,导致跨模态特征融合产生误差。未来如何进一步优化融合模块、细化多模态特征,以提升融合精准性,仍需深入研究。

参 考 文 献

[1] 王勇,胡斌,康渴楠,等. 智能传感器在消防装备中的应用现状分析[J]. 中国安全科学学报, 2022, 32(增1): 184-188.
WANG Yong, HU Bin, KANG Ke'nan, et al. Analysis of application status of intelligent sensors in firefighting equipment[J]. China Safety Science Journal, 2022,32(S1):184-188.

[2] WANG Zhishe, XU Jiawei, JIANG Xiaolin, et al. Infrared and visible image fusion via hybrid decomposition of NSCT and morphological sequential toggle operator[J]. Optik, 2020,201:DOI:10.1016/j.ijleo.2019.163497.

[3] 娄熙承,冯鑫. 潜在低秩表示框架下基于卷积神经网络结合引导滤波的红外与可见光图像融合[J]. 光子学报, 2021,50(3):188-201.
LOU Xicheng, FENG Xin. Infrared and visible image fusion in latent low rank representation framework based on convolution neural network and guided filtering[J]. Acta Photonica Sinica, 2021,50(3):188-201.

[4] VASU G T, PALANISAMY P. Multi-focus image fusion using anisotropic diffusion filter[J]. Soft Computing, 2022, 26(24):14 029-14 040.

[5] MA Jinlei, ZHOU Zhiqiang, WANG Bo, et al. Infrared and visible image fusion based on visual saliency map and weighted least square optimization[J]. Infrared Physics and Technology, 2017,82:8-17.

[6] BAVIRISETTI P D, XIAO Gang, ZHAO Junhao, et al. Multi-scale guided image and video fusion: a fast and efficient approach[J]. Circuits, Systems, and Signal Processing, 2019,38(12):5 576-5 605.

[7] SUN Changqi, ZHANG Cong, XIONG Naixue. Infrared and visible image fusion techniques based on deep learning: a review[J]. Electronics, 2020,9(12):DOI:10.3390/electronics9122162.

[8] SUN Yiming, CAO Bing, ZHU Pengfei, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6 700-6 713.

[9] YUAN Maoxun, WEI Xingxing. C²Former: calibrated and complementary transformer for RGB-infrared object detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024,62:1-12.

[10] WANG Haozhe, SHU Chang, LI Xiaofeng. Enlighten fusion multiscale network for infrared and visible image fusion in dark environments [J]. IEEE Signal Processing Letters, 2023,30:1 167-1 171.

[11] YE Yingchun, ZHANG Junxuan, LI Zeyi. Infrared and visible image fusion method based on dual domain enhancement in low illumination environment[C]. 2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing(DASC), Intl Conf on Pervasive Intelligence and Computing(PiCom), Intl Conf on Cloud and Big Data Computing(CBDCom), Intl Conf on Cyber Science and Technology Congress(CyberSciTech), 2022:1-7.

[12] 周李兵,陈晓晶,贾文琪,等. 用于井下行人检测的可见光和红外图像融合算法[J]. 工矿自动化, 2023,49(9): 73-83.
ZHOU Libing, CHEN Xiaojing, JIA Wenqi, et al. Visible and infrared image fusion algorithm for underground personneldetection[J]. Journal of Mine Automation, 2023,49(9):73-83.

[13] 贾桂敏,程羽,齐孟飞. 多尺度注意力特征增强融合的红外小目标检测新网络[J]. 中国安全科学学报, 2024, 34(6):90-98.
JIA Guimin, CHENG Yu, QI Mengfei. Multi-scale attention feature-enhanced fusion of a new network for infrared small object detection[J]. China Safety Science Journal, 2024,34(6):90-98.

[14] HUSSAIN M. YOLOv1 to v8: unveiling each variant; a comprehensive review of YOLO[J]. IEEE Access, 2024,12: 42 816-42 833.

[15] VANSWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Proceedings of the Advances in Neural

Information Processing Systems(NeurIPS), 2017:5 998–6 008.

[16] SHEN Jifeng, CHEN Yifei, LIU Yue, et al. Icafusion: iterative cross-attention guided feature fusion for multispectral object detection[J]. Pattern Recognition, 2024,145:DOI:10. 1016/j. patcog. 2023. 109913.

[17] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:770–778.

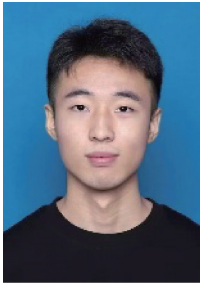
[18] SHEN Zhiqiang, LIU Zechun, XING Eric. Sliced recursive transformer[C]. Proceedings of the Computer Vision-ECCV 2022: 17th European Conference, 2022:727–744.

[19] LINDENBERGER P, SARLIN P E, POLLEFEYS M. LightGlue: local feature matching at light speed[C]. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023:17581–17592.

[20] ZHANG Heng, FROMONT E, LEFEVRE S, et al. Guided attentive feature fusion for multispectral pedestrian detection[C]. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021: 72–80.

[21] ZHANG Heng, FROMONT E, LEFEVRE S, et al. Multispectral fusion for object detection with cyclic fuse-and-refine blocks[C]. Proceedings of the IEEE International Conference on Image Processing(ICIP), 2020:276–280.

[22] FANG Qingyun, HAN Dapeng, WANG Zhaokui. Cross-modality fusion transformer for multispectral object detection [J]. SSRN Electronic Journal, 2022:DOI:10. 2139/ssrn. 4227745.



作者简介： 杨奉展（2001—）,男,山东淄博人,硕士研究生,研究方向为矿山无人驾驶感知预警。E-mail:19854179563@163.com。

《中国安全科学学报》被 Ei 数据库收录

2025 年 4 月,《工程索引》(Ei Compendex)发布最新收录期刊目录,《中国安全科学学报》(简称《学报》)成功被收录。

COMPENDEX SOURCE LIST: UPDATED APRIL 2, 2025

PRINT ISSN	ONLINE ISSN	CHINESE TITLE (中文刊名)	TRANSLITERATED TITLE (刊名翻译)	ENGLISH/TRANSLATED TITLE (英文刊名)	LANGUAGE (语言)	Ei 2024 INDEXING
10014632	-	中国铁道科学	Zhongguo Tiedao Kexue	China Railway Science	Chinese	Renewed (保持收录)
10033033	-	中国安全科学学报	Zhongguo Anquan Kexue Xuebao	China Safety Science Journal	Chinese	New (新收录)
10079289	-	中国表面工程	Zhongguo Biaomian Gongcheng	China Surface Engineering	Chinese	Renewed (保持收录)

自 1991 年创刊以来,《学报》始终秉承“立足安全科技前沿,服务国家重大需求”的办刊宗旨,坚持“学术性、权威性、应用性、信息性”的办刊原则,在安全生产、应急管理、公共安全、防灾减灾、职业健康等领域搭建高水平学术交流平台,发表众多具有创新性和实践价值的研究成果。多年来,《学报》的学术影响力持续攀升,先后入选“中国科技核心期刊”“中文核心期刊”和“中国科学引文数据库核心期刊(CSCD 核心期刊)”,入选中国科技期刊卓越行动计划梯队期刊(一期、二期),并被美国《化学文摘社》(CAS)、Scopus 等国际知名数据库收录。此次成功被 Ei Compendex 数据库收录,更是对《学报》学术地位的有力肯定。