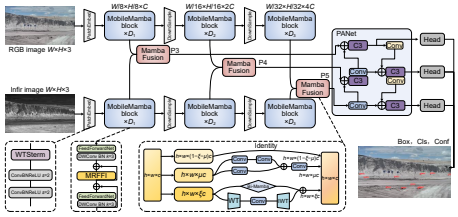
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 2 月, 第 53 卷, 第 2 期

# 红外与可见光渐进式融合的矿区道路落石检测



马文宁*, 孟　彬

西安翻译学院, 陕西 西安 710105

**摘要：**矿区道路落石实时精准检测对矿区安全生产意义重大。矿区落石存在目标微小、光照剧烈变化、恶劣天气干扰等挑战，导致传统单模态检测在低能见度下失效，现有双模态融合方法面临网络结构复杂、计算开销大、难以兼顾实时性与小目标检测精度等问题。为此，本文提出新型红外与可见光渐进式融合检测算法，首先设计基于小波变换的 WTSterm 下采样模块并优化 MobileMamba 骨干网络结构，以减少微小落石细节特征损失，实现红外与可见光特征高效提取；其次设计 MambaFusion 跨模态融合模块，结合 Mamba 全局建模能力与卷积局部细节保留特性，通过动态状态转移和门控机制实现红外与可见光特征高效互补融合；最后结合尺度与位置敏感损失函数 (SLS Loss) 以增强模型对微小目标的敏感性并优化定位精度。在自建露天矿落石数据集上的实验显示，该算法将精度、召回率和 $mAP_{0.5}$ 分别提升至 0.856、0.715 和 0.721，推理速度为 16 ms，有效提升复杂场景下小目标落石的检测性能与鲁棒性，为矿区道路安全实时检测提供有效方法。

**关键词：**落石检测；红外与可见光融合；Mamba；小目标检测

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250226 | CSTR: 32245.14.oee.2026.250226 | 中图分类号：TP391.4 | 文献标识码：A

马文宁, 孟彬. 红外与可见光渐进式融合的矿区道路落石检测 [J]. 光电工程, 2026, **53**(2): 250226

Ma W N, Meng B. Progressive fusion of infrared and visible light for road rockfall detection in mining areas[J]. *Opto-Electron Eng*, 2026, **53**(2): 250226

## 1　引　言

矿区道路落石检测是煤矿安全生产的重要保障，对防范地质灾害、确保运输通道畅通具有关键作用[1]。针对复杂地形中边坡失稳、人工巡检滞后等问题，利用机器视觉方法构建实时监测系统，准确识别道路落石能够有效减少车辆损坏，保障矿区道路畅通。但矿区落石与道路材质相近、受光照和恶劣天气影响，进行准确、全时域的道路落石检测仍是一项挑战。

现阶段落石检测可分为基于激光雷达方法与基于机器视觉方法两种。基于激光雷达的传统落石检测方法大多采用点云聚类算法[2]，如密度聚类[3]、区域增长算法[4] 等，此类算法存在参数敏感、计算效率低、动态抗噪差及小目标漏检等问题[5]，制约复杂矿区场景下的实时精准检测。而深度学习的激光雷达落石检测方法[6]，受限于点云特征稀疏，落石不规则等难以达到良好检测效果，同时激光雷达易受灰尘、雨雾干扰，限制了矿区应用。基于机器视觉的落石检测方法可分为图像特征分析与深度学习目标检测算法两种，图像特征分析算法[7]易受现场环境中光影变化、移动车辆和行人的影响而出现漏报、误报等问题，基于深度学习的目标检测方法具有较好的检测性能，在矿区障碍物检测方面得到了广泛

收稿日期：2025-07-29；修回日期：2025-10-18；录用日期：2025-10-20；网络出版日期：2026-02-13

*通信作者：马文宁，cctn2015@163.com

版权所有©2026 中国科学院光电技术研究所

应用。文献 [8] 基于 YOLOv3 算法开发了铁路边坡落石检测模型，文献 [9] 通过改进 YOLOv5 算法并融合双向加权特征提升矿区检测精度，文献 [10] 则结合改进 YOLOv8 算法在提升检测精度的同时，优化模型计算效率，便于部署于边缘计算设备。但是此类方法均采用可见光相机，在夜间或雾天等低能见度环境下，仅依靠可见光图像进行检测效果较差。

为解决这一问题，已有研究人员提出将可见光图像与红外图像相结合的方法，通过融合两种图像有效提升模型的检测性能。在跨模态融合检测框架的设计中，融合位置的选择是决定模型性能的关键因素。现有融合策略可归纳为 3 类[11]：1) 早期融合在输入层直接拼接红外与可见光原始数据并输入单一网络处理，该策略因忽略模态间成像机制差异，易受噪声干扰且难以有效学习互补信息。2) 晚期融合则采用双模态分支独立提取特征，仅在预测层通过结果加权实现融合，其局限性在于模态间特征交互发生在决策末端，无法充分利用细粒度层级的互补信息，导致目标表征能力受限。3) 作为中间优化方案的半程融合，选择在骨干网络中间层进行双模态特征融合，该策略通过平衡模态交互强度与特征独立性展现出显著优势[12-13]，后续研究[14-17]表明中间层特征既保留了底层纹理和热辐射等基础表征，又具备中层语义聚合能力，为跨模态信息的深度交互提供了理想载体。中期融合方法已有众多研究，文献 [18] 开发了基于双分支特征分解框架的图像融合网络，结合域自适应策略生成高清晰度融合图像，以改善识别效果；文献 [19] 基于 Mask R-CNN 框架，融合多模态图像数据并优化深度卷积网络，提升复杂场景下的行人识别能力，但检测速度较慢；文献 [20] 在 YOLOv5s 架构中集成通道注意力机制，优化网络结构以增强红外与可见光融合数据下的行人检测精度，有效提升了检测速度；文献 [21] 针对车载红外图像特性改进 YOLOv5s 模型，在保持轻量化的同时实现检测精度与计算效率的平衡；文献 [22] 设计自适应注意力特征融合机制，有效整合红外与可见光模态优势特征，同时抑制无效特征干扰，增强目标判别能力，但对于小目标物体检测效果较差。

以上多模态融合检测方法，虽然实现了可见光与红外双模态图像融合检测，但网络结构过深、推理速度有限，并且无法满足露天矿区运输道路中常见的直径为 10~50 cm 的落石检测。因此本文提出一种基于 Mamba[23] 的轻量型红外与可见光融合矿区落石检测算法。为复杂环境下露天矿区非结构化路面落石检测提供了更高效的方法，进而降低安全事故发生率，提升运输效率。本文的主要贡献有以下三点：

1) 提出小波变换下采样模块并优化了 MobileMamba[24] 骨干网络结构，以减少微小落石细节特征损失；

2) 提出了 MambaFusion 特征融合模块，有着良好的全局建模能力，可关联跨模态的语义对应区域，实现红外图像与可见光图像的高效融合；

3) 在落石检测任务中引入了尺度与位置敏感损失函数，改善落石小目标检测与定位，实验表明本文方法有效提升了复杂场景下小目标落石的检测性能与鲁棒性，为矿区道路安全实时检测提供可靠保障。

## 2 基于 Mamba 的红外与可见光渐进式融合模型

### 2.1 模型架构

受复杂矿山环境的影响，落石目标在红外与可见光成像中通常表现出尺寸小、形态不规则、特征微弱且易与背景混淆等显著特性。这些特性使得传统目标检测模型难以有效地区分目标与背景噪声，导致检测精度大幅下降。虽然卷积神经网络 (CNN) 和 Transformer[25] 架构在各类视觉任务中取得了显著成功，但在处理矿山监控所需的大尺寸高分辨率图像输入时，普遍面临计算复杂度呈平方级增长或感受野受限与局部建模导致的效率低下问题，难以满足实时性要求，同时也限制了模型捕捉长距离上下文依赖以识别小目标和复杂背景中微弱特征的能力。针对上述问题，本研究提出了一种基于状态空间模型 (state space model, SSM) Mamba[26] 的轻量型红外与可见光渐进式融合矿区落石检测算法。模型结构如图 1 所示，充分利用了 Mamba 架构在建模长序列依赖方面的独特优势，以及硬件感知并行架构和选择性扫描机制带来的高计算效率。其中：$W$, $H$, $C$ 分别表示骨干网络中特征图的宽、高、通道数；$w$, $h$, $c$ 分别表示 MobileMamba block 中特征图的宽、高、通道数；$D$ 为堆叠层数/块的数量；$\mu$和$\xi$分别为 Mamba 和小波变化分支的通道数比例；$s$ 为步长。

模型首先使用 MobileMamba 特征提取骨干网络进行优化构建双流架构，用于高效挖掘红外图像中显著的热辐射特征，该特征对昼夜光照变化和灰尘雨雾具有固有鲁棒性，同时提取可见光图像中丰富的空间细节纹理特征，这些特征提供了目标形状和边缘的关键信息。利用 MobileMamba 内置的选择性状态空间块，能够动态、选择性地捕获输入图像中的关键信息流，在保留深层高

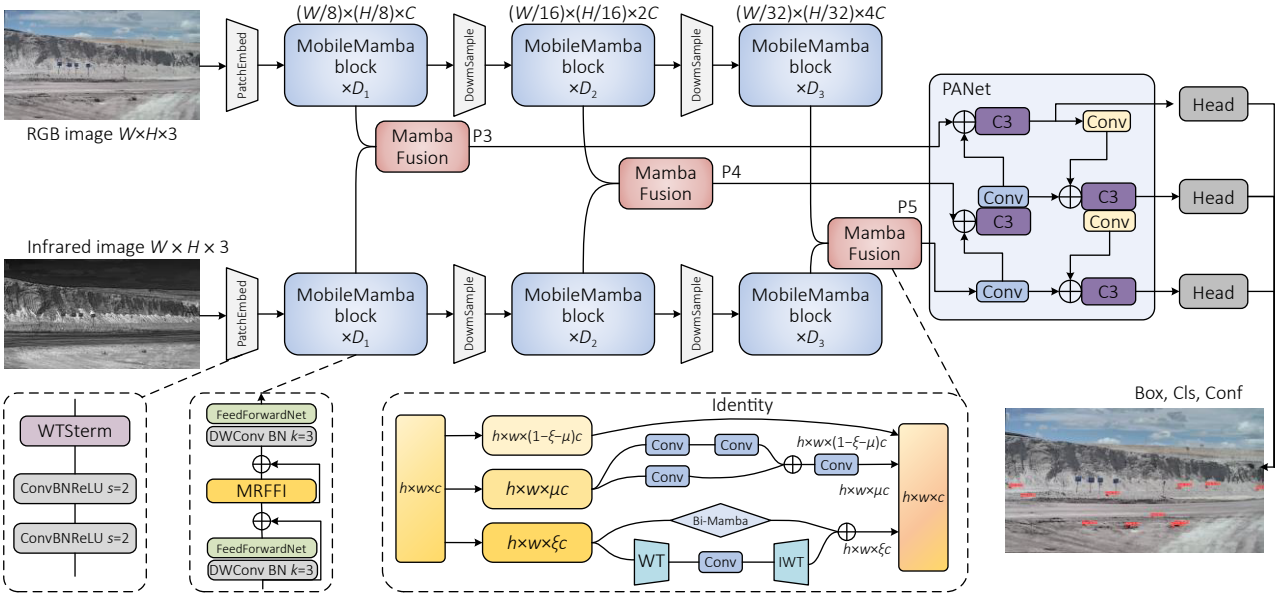



图 1 基于 Mamba 的红外与可见光渐进式融合模型

Fig. 1 Mamba-based infrared and visible light progressive fusion model

维语义信息的同时，精准维护浅层的细粒度视觉细节。其次设计了 MambaFusion 跨模态特征融合机制，渐进式融合大、中、小三个尺寸的特征图，通过方向性扫描过程同步处理红外与可见光分支的特征图序列，结合 Mamba 固有的动态参数化能力，即其线性变换权重由输入数据动态生成，以及选择性状态转移特性，自动筛选并保留来自双模态最具判别性的互补信息，有效过滤掉各模态特有的噪声和冗余数据。通过学习得到的调制权重对融合过程中的多尺度特征实施自适应加权融合，确保最终融合特征能依据场景上下文动态整合多模态及多尺度的显著信息，从而增强模型在矿区常见的恶劣场景，如强烈逆光、昏暗照明、眩光、烟尘弥漫及目标部分遮挡等条件下的目标表征能力，实现红外热辐射与可见光纹理特征在全局信息依赖指导下的深度高效交互，生成融合特征图。随后，将这些具有强互补性与鲁棒性的融合特征图 P3、P4、P5 输入到双向特征融合金字塔结构，即路径聚合网络 PANet 中，通过其自底向上和自顶向下的双向信息融合路径，深入挖掘不同分辨率特征图中的多尺度上下文信息，有效融合深层特征的语义理解与浅层特征的空间细节，以提升模型在多种尺度上、特别是对小型落石目标的精准捕捉能力和尺寸适应性。最终，经过上述强化的多尺度融合特征被送入一个标准的分类预测与边界框回归检测头，同时输出检测到的落石目标精确的边界框位置坐标、具体的类别归属置信度以及目标真实存在的置信度分数。

## 2.2 基于 MobileMamba 的图像特征提取优化

针对露天矿区道路落石检测中红外与可见光图像特征提取难题，本文对 MobileMamba[24] 进行了针对性优化。MobileMamba 是一种新型轻量级视觉状态空间模型，其通过多感受野特征交互模块 (MRFFI) 与三阶段网络架构，在复杂场景的小目标检测任务中展现出显著优势。该模型采用长程小波变换增强的 Mamba 模块 (WTE-Mamba)，通过 Haar 小波分解提取高频边缘特征，结合全局状态空间建模，显著增强了细粒度特征表征能力。同时，其多核深度卷积 (MK-DeConv) 模块构建多尺度局部感受野，有效改善露天矿区道路场景中落石目标尺度微小的检测问题。在红外图像处理分支，WTE-Mamba 模块强化热辐射特征的长程依赖建模，有效捕捉温度异常区域；可见光分支则通过 MK-DeConv 的 3/5/7 多核卷积并行结构，提取不同尺度的纹理细节。双模态特征在 MRFFI 模块中经通道交互与非线性融合后，可生成具有空间-光谱联合表征能力的融合特征，克服传统方法在低信噪比环境下特征混淆的缺陷。

但是将 MobileMamba 应用于目标检测任务时，PatchEmbed 层使用了四个卷积块连续进行下采样，极大地损失了细节特征，影响落石这类小尺寸物体的检测召回率。因此本文提出了 WTSterm 下采样模块，并对骨干网络结构进行了重新设计，如图 2 所示。WTSterm 模

块利用小波变换将输入的 640×640×3 尺寸图像分解为近似图像 (低频信息)、水平细节 (垂直边缘)、垂直细节 (水平边缘) 和对角细节 (斜向纹理)，这 4 个图像的尺寸为 320×320×3，在通道维度进行拼接得到 320×320×12 的特征图，而后使用步长为 1 的 1×1 卷积整合通道实现图像的下采样。相较于使用步长为 2 的 3×3 卷积进行下采样，WTSterm 能够尽可能地保留落石的细节特征信息，并减少对于大尺寸图像的计算量，加快检测速度。在网络结构方面，本文进行 3 次下采样后利用 MobileMamba block 进行特征提取，使最终输出的基础特征图尺寸为 80×80。该尺寸特征图在空间分辨率上保留了更丰富的原始图像细节，能够有效捕捉落石目标的边缘轮廓、热辐射梯度等细粒度特征，尤其适合远处小尺寸落石的表征。同时通过 MobileMamba 的硬件感知并行计算单元对特征图的通道维度进行优化，结合扫描窗口状态缓存机制，在保持特征提取能力的同时显著降低计算量。针对原架构中小尺寸高维特征图对于落石检测存在特征冗余问题，由于 10×10 尺寸特征图对小目标的表征能力较弱，且带来较大计算负担，因此移除该尺寸特征图分支。通过减少冗余特征处理流程，进一步降低模型计算量，优化网络结构对小目标检测的适应性，在保持特征提取效率的同时提升模型对远处落石目标的表征能力。

## 2.3 红外图像与可见光图像特征融合模块 MambaFusion

传统的卷积模块进行红外与可见光图像特征融合时，存在较多的信息丢失问题，本文针对红外图像对热辐射敏感但缺乏纹理细节、可见光图像包含丰富空间信息但受光照影响大等问题，提出一种 MambaFusion 的融合模块。如图 2 所示，结合了 mamba、卷积和小波变换的优势，有着良好的全局建模能力，可关联跨模态的语义对应区域，而局部卷积分支能有效保留各模态的细节特征，实现流程如下：

给定输入特征图 $x^{\mathrm{I}} \in \mathbf{R}^{h\times w\times c}$，MRFFI 将其沿通道维度划分为三个子部分 $(\xi c, \mu c, (1-\xi-\mu)c)$，分别通过以下三个组件处理：

1) 长程小波变换增强 Mamba 模块通过双向扫描 Mamba 模块捕获全局依赖关系，并利用小波变换增强高频细节。首先通道比例为 $\xi$ 的特征子集 $x_{\mathrm{G}}^{\mathrm{I}} \in \mathbf{R}^{h\times w\times \xi c}$ 通过双向扫描 Mamba 模块学习全局信息。

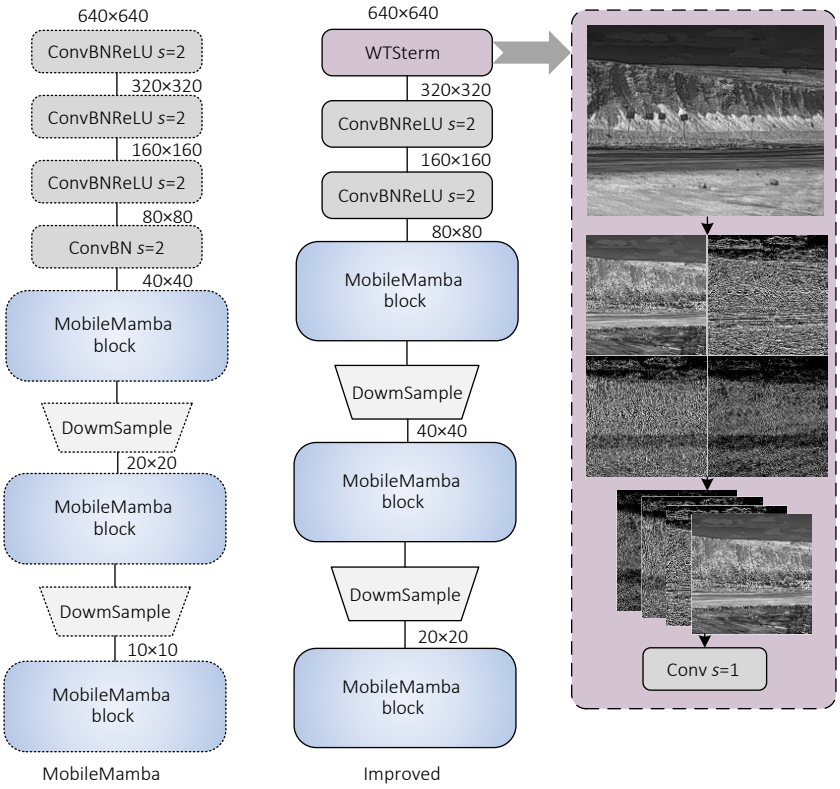



图 2 改进的 MobileMamba 特征提取网络

Fig. 2 Improved MobileMamba feature extraction network

$$x_{m1}^{\mathrm{I}} = SSM\left(\sigma\left(Conv\left(Linear(x_{\mathrm{G}}^{\mathrm{I}})\right)\right)\right),$$
$$x_{m2}^{\mathrm{I}} = \sigma\left(Linear(x_{\mathrm{G}}^{\mathrm{I}})\right),$$
$$x_{m}^{\mathrm{O}} = Linear\left(x_{m1}^{\mathrm{I}} \otimes x_{m2}^{\mathrm{I}}\right), \tag{1}$$

式中：*Linear* 表示线性变换；*SSM* 表示 Mamba 变换模块；*Conv* 表示卷积；$\sigma$表示超参数。

其次对同一特征图进行小波变换，获得不同频率尺度的特征表达 $x_{\mathrm{w}}^{\mathrm{I}} \in \mathbf{R}^{\frac{h}{2}\times\frac{w}{2}\times 4\xi c}$：

$$x_{\mathrm{wt}}^{\mathrm{I}} = WT\left(x_{\mathrm{w}}^{\mathrm{I}}, [\boldsymbol{f}_{\mathrm{LL}}, \boldsymbol{f}_{\mathrm{LH}}, \boldsymbol{f}_{\mathrm{HL}}, \boldsymbol{f}_{\mathrm{HH}}]\right), \tag{2}$$

$$x_{\mathrm{w}}^{\mathrm{O}} = IWT\left(Conv(x_{\mathrm{wt}}^{\mathrm{I}}), [\boldsymbol{f}_{\mathrm{LL}}, \boldsymbol{f}_{\mathrm{LH}}, \boldsymbol{f}_{\mathrm{HL}}, \boldsymbol{f}_{\mathrm{HH}}]\right), \tag{3}$$

式中：*WT* 为小波变换；*IWT* 为逆小波变换。滤波器组定义为

$$\begin{cases} \boldsymbol{f}_{\mathrm{LL}} = \dfrac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\ \boldsymbol{f}_{\mathrm{LH}} = \dfrac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \\ \boldsymbol{f}_{\mathrm{HL}} = \dfrac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \\ \boldsymbol{f}_{\mathrm{HH}} = \dfrac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{cases}. \tag{4}$$

最终输出通过残差连接融合全局信息与高频细节：$x_{\mathrm{G}}^{\mathrm{O}} = x_{m}^{\mathrm{O}} + x_{\mathrm{w}}^{\mathrm{O}}$，其中$x_{\mathrm{G}}^{\mathrm{O}} \in \mathbf{R}^{h\times w\times \xi c}$。

2) 多梯度流卷积模块，通过并行 3×3 和 1×1 多尺度卷积捕获局部特征，实现多尺度感受野信息融合。

$$\begin{cases} x_{\mathrm{L1}}^{\mathrm{O}} = Conv(x_{\mathrm{L1}}^{\mathrm{I}}, k=1) \\ x_{\mathrm{L2}}^{\mathrm{O}} = Conv(x_{\mathrm{L2}}^{\mathrm{I}}, k=3) \\ x_{\mathrm{L}}^{\mathrm{O}} = Concat([x_{\mathrm{L1}}^{\mathrm{O}}, x_{\mathrm{L2}}^{\mathrm{O}}], dim=1) \end{cases}, \tag{5}$$

式中：$k$ 表示卷积核尺寸；$x_{\mathrm{L}i}^{\mathrm{I}}$为卷积处理的特征；Concat 表示在特征拼接；*dim*=1 表示在通道维度拼接。

3) 冗余特征消除，剩余通道通过恒等映射直接传递，降低计算冗余。

$$x^{\mathrm{O}} = \mathrm{Concat}(x_{\mathrm{G}}^{\mathrm{O}}, x_{\mathrm{L}}^{\mathrm{O}}, x^{\mathrm{I}}[(1-\xi-\mu)c:]). \tag{6}$$

4) 设红外特征为 $F_{\mathrm{IR}} \in \mathbf{R}^{h\times w\times c}$，可见光特征为 $F_{\mathrm{VIS}} \in \mathbf{R}^{h\times w\times c}$，表示融合过程可表示为

$$F_{\mathrm{fusion}} = MambaFusion(Concat(F_{\mathrm{IR}}, F_{\mathrm{VIS}}))$$
$$= \sum_{k=1}^{3} W_k \times F_k(F_{\mathrm{IR}}, F_{\mathrm{VIS}}). \tag{7}$$

式中：$F_k$表示特征提取算子，$W_k$ 为融合权重。

### 2.4 小目标检测定位优化

在矿区道路落石检测任务中，针对目标尺度微小且定位精度要求高的特点，本文引入了一种尺度与位置敏感损失函数 (scale and location sensitive loss, SLS Loss)，以提升模型对小目标的鉴别能力。该损失函数通过显式建模目标尺度差异与中心点偏移的敏感性，有效解决了传统分割损失 (如 IoU、Dice) 对目标尺寸和位置变化不敏感的问题。其数学形式定义为：$L_{\mathrm{SLS}} = L_{\mathrm{S}} + L_{\mathrm{L}}$，其中，$L_{\mathrm{S}}$为尺度敏感损失项，$L_{\mathrm{L}}$为位置敏感损失项。

尺度敏感损失基于目标像素集合 $A_{\mathrm{p}}$ (预测值) 与 $A_{\mathrm{gt}}$ (真实值)，引入尺度自适应权重 $w$ 重构 IoU 损失。

$$L_{\mathrm{S}} = 1 - w \cdot \frac{|A_{\mathrm{p}} \cap A_{\mathrm{gt}}|}{|A_{\mathrm{p}} \cup A_{\mathrm{gt}}|},$$
$$w = \frac{*\min(|A_{\mathrm{p}}|, |A_{\mathrm{gt}}|) + \mathrm{Var}(|A_{\mathrm{p}}|, |A_{\mathrm{gt}}|)}{*\max(|A_{\mathrm{p}}|, |A_{\mathrm{gt}}|) + \mathrm{Var}(|A_{\mathrm{p}}|, |A_{\mathrm{gt}}|)}, \tag{8}$$

式中：$*\min(\cdot)$和$*\max(\cdot)$为针对目标检测中预测框面积与真实框面积的比较运算，代表两个数值中的最小值和最大值；$\mathrm{Var}(\cdot)$为方差函数。该设计使尺度差异 ($|A_{\mathrm{p}}|$与$|A_{\mathrm{gt}}|$的偏差) 与损失值正相关，迫使模型更关注尺度预测偏差较大的目标，尤其强化微小落石的监督信号。

位置敏感损失通过极坐标转换量化目标中心点偏移误差。首先计算预测目标中心$c_{\mathrm{p}} = (x_{\mathrm{p}}, y_{\mathrm{p}})$与真实中心$c_{\mathrm{gt}} = (x_{\mathrm{gt}}, y_{\mathrm{gt}})$的极坐标参数：$d_{\mathrm{p}} = \sqrt{x_{\mathrm{p}}^2 + y_{\mathrm{p}}^2}$, $\theta_{\mathrm{p}} = \arctan(y_{\mathrm{p}}/x_{\mathrm{p}})$。

损失函数融合径向距离与角度偏差：

$$L_{\mathrm{L}} = \left(1 - \frac{\min(d_{\mathrm{p}}, d_{\mathrm{gt}})}{\max(d_{\mathrm{p}}, d_{\mathrm{gt}})}\right) + \frac{4}{\pi^2}(\theta_{\mathrm{p}} - \theta_{\mathrm{gt}})^2. \tag{9}$$

该组合惩罚项对中心点偏移具有强区分性，适用于落石定位精度的优化，在矿区道路场景中，SLS 损失通过尺度权重 $w$ 增强模型对微小落石的敏感性，同时利用极坐标约束提升落石中心点定位精度，为复杂背景下落石检测的可靠性提供保障。

## 3 实验结果与分析

### 3.1 实验数据集构建

鉴于当前缺乏公开的矿区道路落石数据集，本文从内蒙古和新疆的露天矿区采集了 5634 组红外与可见光图像。可见光相机采用 1920×1080 分辨率，热红外相机采用 640×480 分辨率，8~14 μm 长波红外工作波段，

6.8 mm 焦距，视场角为 39.3°×29.4°。为明确算法的工程适用范围，本系统在部署时建议将成像单元的俯仰角设置在−10°至 15°之间，以优化对路面的覆盖；其可见光模态的工作照度范围为 1 lx 至 100000 lx，红外模态则不受环境光照制约，从而共同确保了算法在矿区昼夜、晴雨等各种复杂光照场景下的全时域鲁棒检测能力。

数据包含了金属矿与非金属矿，涵盖白天、夜间、傍晚、灰尘等天气。由于热红外图像与可见光图像相机参数不同，需要对两个相机进行标定，使用棋盘格标定板，通过拍摄多个角度的标定板图像，计算并求解各自相机的内部参数 (包括焦距、主点坐标、径向与切向畸变系数)，从而矫正镜头引起的图像畸变并获得每个相机的精确成像模型，如图 3 所示。然后在两幅已矫正畸变的图像上人工选择寻找一组数量足够且分布合理的相同物理位置的特征点对作为公共特征点。利用这些精确匹配的公共特征点对，可以采用诸如最小二乘法的优化技术计算一个描述两个成像平面之间几何变换关系的单应性矩阵 $\boldsymbol{H}$。最终，利用求解出的这个投影变换矩阵，对红外图像的每一个像素坐标应用变换计算其在可见光图像坐标系中的对应位置。

由于红外图像可以投影到 RGB 图像，因此在数据标注时，以红外图像作为辅助，在 RGB 图像上进行标注。在露天矿区环境中，落石广泛存在于路面、挡墙、爆堆、排土场等环境中，仅标注在车辆可行驶区域内部和边缘的落石，使用 LabelImg 标注工具，构建了露天矿区热红外与可见光落石数据集，将数据集依据 8:2 的比例，随机划分生成训练集和验证集。标注结果如图 4 所示，横坐标表示落石在图像中占据的横坐标像素，最大的落石目标尺寸仅 150 pixels，在目标检测领域中均属于小目标，其中大部分尺寸在 0~40 pixels 之间的落石属于微小目标，这对落石精确检测提出了更高的要求。

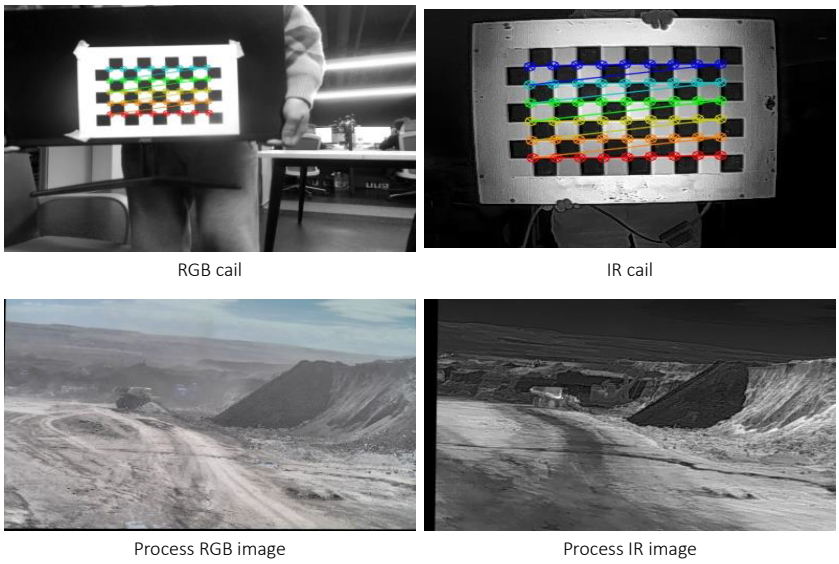



图 3 红外与可见光相机标定与投影

Fig. 3 Calibration and projection of infrared and visible light cameras

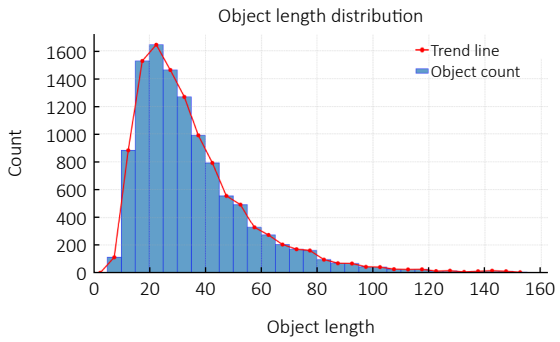



图 4 落石目标信息统计

Fig. 4 Statistical information of rockfall targets

在后续实验中，本文将对比前融合和中期融合的优劣，因此本文采用不同权重的 Blend 融合、CDDFuse[27] 和 DDFM[28] 红外与可见光融合方法，同时构建了融合图像的数据集，如图 5 所示。

## 3.2 实验设置与评价指标

本研究采用 Ubuntu 22.04 操作系统，使用 Intel Core

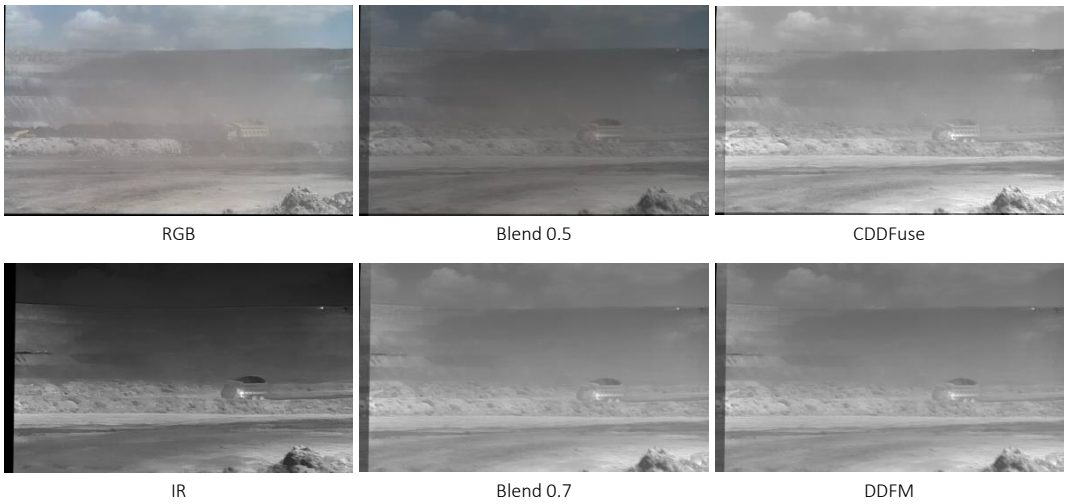



图 5 RGB 与 IR 图像融合结果

Fig. 5 Fusion results of RGB and IR images

(TM) i9-14900K CPU，Nvidia RTX 3090 24 G GPU，系统内存 64 G，采用 Cuda 12.1，Pytorch 2.1.0 深度学习框架，模型的训练参数设置如表 1 所示。

表 1 模型训练参数设置

Table 1 Model training parameter settings

| Params | Set |
|---|---|
| Epoch | 100 |
| BatchSize | 8 |
| Image size | 640×320 |
| Optimizer | SGD |
| Learning rate | 0.01 |
| Moment | 0.947 |

本文使用性能指标与复杂度指标两类进行模型评估：精确率 (Precision) 反映检测结果中正确正样本的比例；召回率 (Recall) 衡量模型找出真实正样本的能力；IoU 阈值为 0.5 的平均精度 ($mAP_{0.5}$) 评估宽松定位精度；IoU 阈值为 0.5~0.95 的平均精度 ($mAP_{0.5:0.95}$) 综合评价严格定位能力；参数量 (Params) 表征模型存储复杂度；浮点运算量 (FLOPs) 量化计算资源需求。

$$Precision = \frac{TP}{TP+FP}, \tag{10}$$

$$Recall = \frac{TP}{TP+FN}, \tag{11}$$

$$mAP_{0.5} = \frac{1}{N}\sum_{i=1}^{N}\int_0^1 p_i(r)\mathrm{d}r|_{IoU=0.5}, \tag{12}$$

$$mAP_{0.5:0.95} = \frac{1}{10}\sum_{k=0}^{9} mAP_{0.5+0.05k}, \tag{13}$$

式中：TP (true positive) 表示 IoU≥阈值且分类正确的检测框数量；FP (false positive) 表示 IoU<阈值或分类错误的检测框数量；FN (false negative) 表示未被检出的真实目标数量；$N$ 表示目标类别总数；$p(r)$表示召回率值为 $r$ 时对应的最大精确率；d$r$表示召回率的微分变量。

## 3.3 对比实验分析

### 3.3.1 不同融合方法对比实验分析

为验证前融合、中间融合及单模态图像在矿区落石检测中的有效性，本研究基于 YOLOv5L 模型分别训练了可见光图像、红外图像、Blend 加权融合图像、CDDFuse 融合图像及 DDFM 融合图像，并与提出的渐进式融合方法进行对比。从表 2 可以分析出：单一模态模型中，可见光模型虽以 0.804 的精度为最高，但其召回率仅为 0.547，导致综合检测性能 $mAP_{0.5}$ 仅为 0.589；红外模型因自建数据集中相机模组未实现硬件级时空同步，且标注仅基于可见光图像引入标签偏差，导致性能表现欠佳。Blend 融合方法在图像层面以固定权重融合模态时，因特征干扰导致性能甚至低于单可见光模型，且 57 ms 的推理速度远劣于单模态的12 ms。基于先进融合模型 CDDFuse 和 DDFM 的前融合方法虽在检测性能上有所提升，CDDFuse 的 $mAP_{0.5}$ 为 0.591，DDFM 精度达 0.832 且 $mAP_{0.5}$ 为 0.629，但 DDFM 因依赖扩散模型需 100 次迭代，造成严重速度瓶颈，无法满足实时需求。

相比之下，本文渐进式融合方法以 0.856 的精度、0.715 的召回率、0.721 的 $mAP_{0.5}$ 及 0.389 的 $mAP_{0.5:0.95}$ 全面超越对比方法，同时 16 ms 的推理速度接近单模态模型，证明传统图像级前融合效果有限，而复杂模型前融合虽提升性能却牺牲实用性。本文采用的渐进式融合方法有效解决了性能与速度的矛盾，通过在模型内部高效协同 RGB 的纹理细节与 IR 的环境感知优势，能够有效提升检测性能并维持实时推理速度，为矿区落石检测的实时可靠监控提供了有效方法。

图 6 展示了单模态 RGB 方法，前融合方法 (CDDFuse) 和本文的中间融合方法的检测结果，在白天光照良好的情况下，三者都有着较好的检测结果，能够在 50 m 的距离检测到 20 cm×20 cm 的碎石，100 m 的距离下检测到 50 cm×50 cm 的碎石。当存在逆光、灰尘等场景时单模态 RGB 方法存在一定的精度衰减，前融合方法由于图像融合模型不适用于露天矿落石，在部分场景存在一

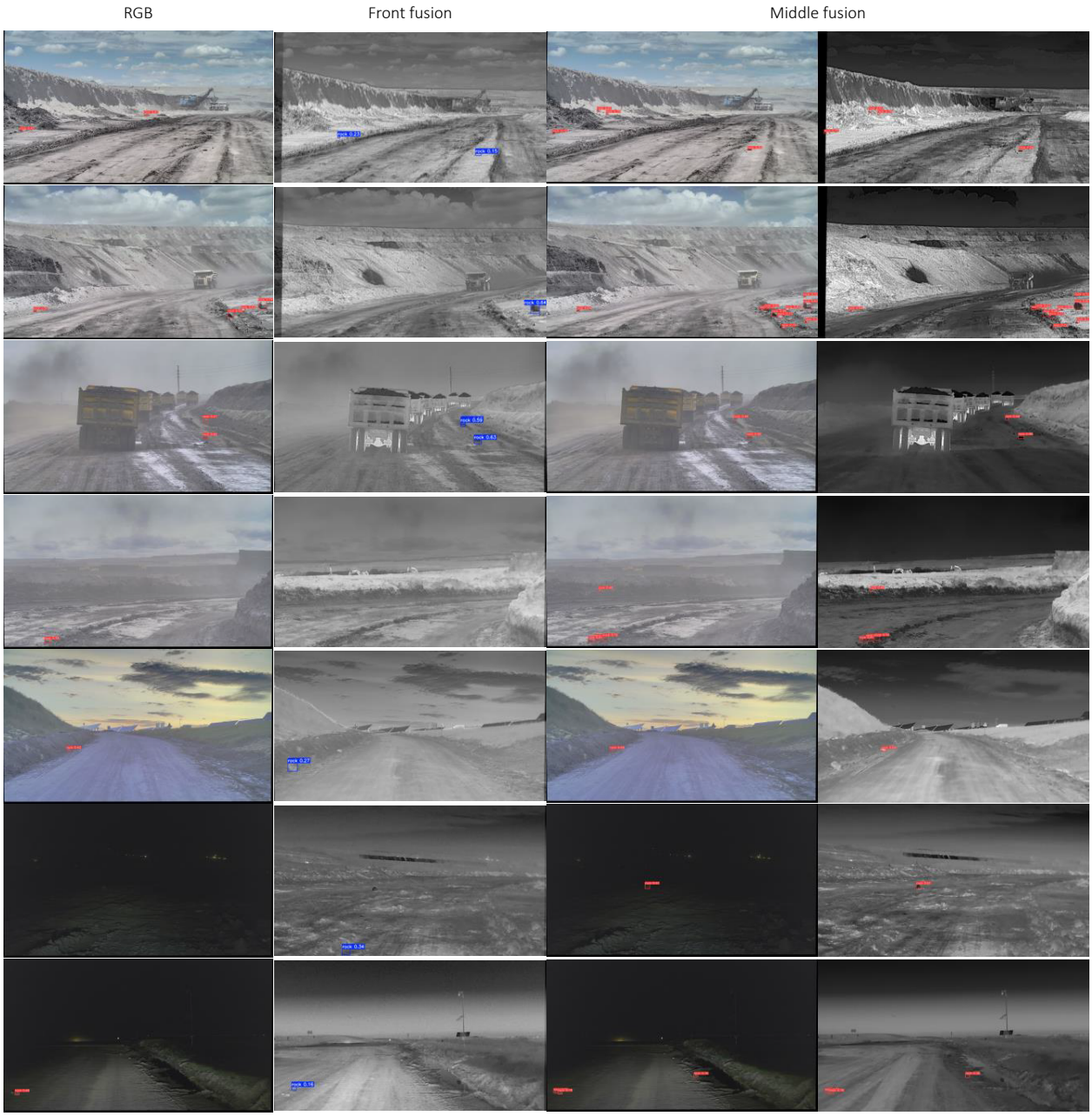



图 6 不同融合方法检测结果

Fig. 6 Detection results of different fusion methods

定的误检问题，而本文方法不仅能够准确识别路面落石，同时由于 IR 和 RGB 图像的渐进式有机融合，使得 IR 增强了 RGB 图像在灰尘和逆光场景下的检测性能，同时借助 RGB 图像丰富的色彩纹理信息，可以有效区分路面的落石和土堆，减少土堆引起的误报。在夜间光照不理想场景下，单模态 RGB 方法因缺乏有效视觉信息，检出率显著降低甚至失效；前融合方法虽能检测到部分落石，但对非落石热源 (如地面余热、机械设备) 的误检增多。相比之下，本文方法凭借红外图像对温度差异的敏感特性，能清晰捕捉到落石目标，并通过渐进式融合有效结合 RGB 模态的纹理信息，消除了干扰热源，实现夜间复杂场景下的精准、低误检识别。

### 3.3.2 主流融合模型对比实验分析

本文模型与主流的多模态融合目标检测网络 GAFF[15]、CFT[11]、ICAFusion[29] 进行了对比分析，实验结果如表 3 所示。在红外与可见光融合的露天矿落石检测实验中，各对比模型呈现出不同的性能特征。GAFF 作为轻量化模型，有着 15 ms 的快速推理速度和 22.6 M 的较小参数规模，但检测精度不足，$mAP_{0.5}$为 0.672、$mAP_{0.50\text{-}0.95}$ 仅为 0.304，在多尺度目标检测和复杂场景下的鲁棒性较弱。CFT 参数规模达到了 196.7 M，尽管 $mAP_{0.5}$ 略高，$mAP_{0.5}$ 达到 0.674，但 $mAP_{0.5\text{-}0.95}$ 仅 0.295，整体性能难以满足露天矿实时检测的效率与精度需求。ICAF-NiF 达到了 0.862 的精确率和 0.735 的 $mAP_{0.5}$ 指标，使用不同骨干网络的 ICAF-Transform、ICAF-ResNet50、ICAF-VGG16 均有着较好的性能，但其模型参数量大，检测速度较慢。

相比之下，本文模型在综合性能表现良好，有着 0.856 的精确率与 0.715 的召回率，实现了漏检与误检风险的有效平衡，42.6 M 的参数规模不足 ICAF 系列模型的四分之一，16 ms 的推理速度接近轻量化模型 GAFF 却在 $mAP_{0.5}$ 和 $mAP_{0.5\text{-}0.95}$ 指标中有着较大提升，体现了对红外与可见光跨模态特征的高效融合能力。为了进一步验证本文所提出的特征融合模块有效性，使用 GradCam 对融合模块的输出特征进行了可视化，如图 7 所示，GAFF 模型由于结构简单，能够捕捉到大量特征但是也增加了部分背景信息，CFT 和 ICAFusion 有着较好的特征融合性能，但是存在部分特征丢失现象。本文模型通过捕捉红外模态的温度差异与可见光模态的纹理细节，增强了落石目标在复杂背景中的区分度，尤其提

表 2　不同融合方法对比实验结果

Table 2　Comparative experiment results of different fusion methods

| Model | Precision | Recall | $mAP_{0.5}$ | $mAP_{0.5\text{-}0.95}$ | Inference speed |
|---|---|---|---|---|---|
| Yolov5L+RGB | 0.804 | 0.547 | 0.589 | 0.255 | 12 ms |
| Yolov5L+IR | 0.659 | 0.524 | 0.492 | 0.234 | 12 ms |
| Yolov5L+ Blend 0.5 | 0.735 | 0.502 | 0.538 | 0.231 | 57 ms |
| Yolov5L+ Blend 0.7 | 0.772 | 0.525 | 0.567 | 0.244 | 57 ms |
| Yolov5L+ CDDFuse | 0.807 | 0.549 | 0.591 | 0.257 | 3.5 s |
| Yolov5L+ DDFM | 0.832 | 0.571 | 0.629 | 0.311 | >2 min |
| Our model+RGB+IR | 0.856 | 0.715 | 0.721 | 0.389 | 16 ms |

表 3　主流融合模型对比实验结果

Table 3　Comparative experiment results of mainstream fusion models

| Model | Precision | Recall | $mAP_{0.5}$ | $mAP_{0.5\text{-}0.95}$ | Params/M | Inference speed/ms |
|---|---|---|---|---|---|---|
| GAFF | 0.765 | 0.694 | 0.672 | 0.304 | 22.6 | 15 |
| CFT[26] | 0.796 | 0.702 | 0.674 | 0.295 | 196.7 | 48 |
| ICAF-NiF[27] | 0.862 | 0.697 | 0.735 | 0.371 | 180.6 | 25 |
| ICAF-Transform | 0.885 | 0.722 | 0.726 | 0.384 | 181.5 | 26 |
| ICAF-ResNet50 | 0.876 | 0.723 | 0.731 | 0.377 | 299.3 | 32 |
| ICAF-VGG16 | 0.832 | 0.711 | 0.694 | 0.357 | 188.1 | 18 |
| Our model | 0.856 | 0.715 | 0.721 | 0.389 | 42.6 | 16 |

升了小目标和隐蔽目标的检测效果。这种轻量化与高精度的结合，使其能够满足实时监测的预警时效性要求，在露天矿落石检测中展现良好的应用价值。

### 3.4 消融实验

在红外与可见光融合的露天矿落石检测消融实验中，基准模型实现了 0.815 的精确率、0.652 的召回率、0.668 的 $mAP_{0.5}$ 及 0.328 的 $mAP_{0.5\text{-}0.95}$。如表 4 所示，结合优化后的 MobileMamba、MambaFusion 和 SLS 损失函数后模型精确率达到 0.856、召回率为 0.715、$mAP_{0.5}$ 和 $mAP_{0.5\text{-}0.95}$ 分别为 0.721 和 0.389，各项指标均为消融实验中的最优表现。采用重构的 MobileMamba 替换基础特征提取器，凭借其选择性状态空间机制，有效增强矿区复杂背景下小目标特征的捕获能力，初步提升特征表征质量；嵌入 MambaFusion 模块，通过方向性扫描与自适应加权，实现红外热辐射特征与可见光纹理特征的深度互补交互，改善遮挡与光照突变场景的目标判别能力；引入 SLS Loss 通过约束边界框几何一致性与语义对齐，缓解因落石形态不规则导致的定位漂移问题。实验结果表明，优化的 MobileMamba 提供的轻量化特征提取框架、MambaFusion 实现的跨模态信息交互机制以及 SLS Loss 对训练过程的优化具有互补效应，三者协同在保持模型推理效率的同时提升了检测性能，验证了骨干网络、融合模块与损失函数的协同优化对红外与可见光融合检测模型在露天矿复杂场景下检测效果的有效性。

### 3.5 模型环境适应性分析

矿区道路环境受自然光照、季节气候等因素影响显著，导致红外与可见光成像特性在不同时间和季节下存在较大差异。白天可见光图像纹理丰富但易受逆光、眩光干扰；夜间可见光信息匮乏，而红外图像依赖热辐射，易受地面余热、设备散热等干扰；黄昏时段光照剧烈变化，可见光与红外模态均面临特征不稳定挑战。为验证

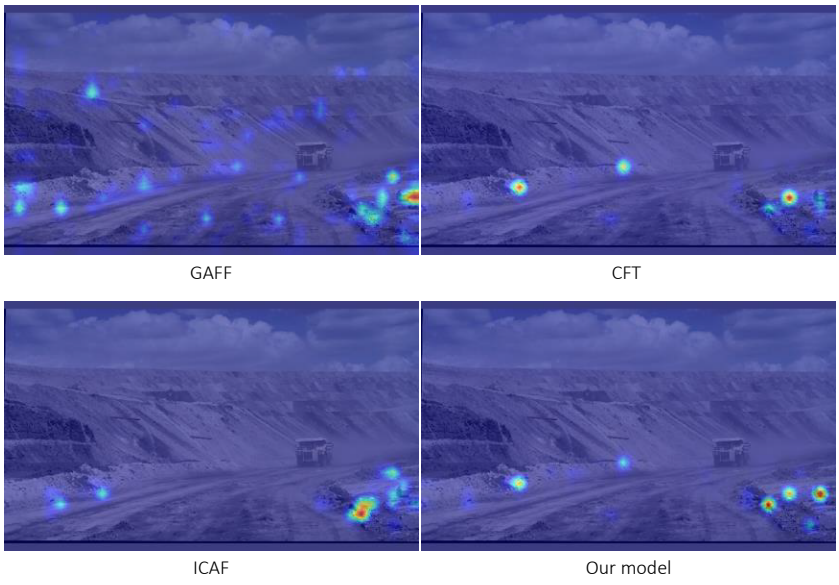



图 7 融合模块可视化结果

Fig. 7 Visualization results of the fusion module

表 4 消融实验结果

Table 4 Ablation experiment results

| MobileMamba | Improve MobileMamba | MambaFusion | SLS loss | Precision | Recall | $mAP_{0.5}$ | $mAP_{0.5\text{-}0.95}$ |
|---|---|---|---|---|---|---|---|
| √ | | | | 0.815 | 0.652 | 0.668 | 0.328 |
| | √ | | | 0.828 | 0.685 | 0.692 | 0.352 |
| | √ | √ | | 0.840 | 0.698 | 0.704 | 0.373 |
| | √ | √ | √ | 0.856 | 0.715 | 0.721 | 0.389 |

模型在多变环境下的鲁棒性，本文从自建数据集中提取了不同时间段 (白天、夜间、黄昏) 和子集进行测试。

如表 5 所示，模型在不同环境下性能波动较小，主要得益于 MambaFusion 模块的动态门控机制与选择性状态转移能力，能够自适应调整双模态特征权重，抑制环境噪声，增强目标判别特征。此外，SLS 损失函数通过强化小目标的位置与尺度敏感性，进一步提升了模型在低对比度、弱光照等极端场景下的定位精度。综上，本文模型具备良好的时间与季节适应性，可为矿区全时域、全季节道路安全监控提供可靠技术支持。

表 5 模型环境适应性分析结果

Table 5 Analysis results of model environment adaptability

| Time | Precession | Recall | $mAP_{0.5}$ | $mAP_{0.5\text{-}0.95}$ |
|---|---|---|---|---|
| Day | 0.850 | 0.715 | 0.725 | 0.395 |
| Dusk | 0.838 | 0.701 | 0.715 | 0.372 |
| Night | 0.855 | 0.688 | 0.704 | 0.381 |

## 4 结 论

针对露天矿复杂场景下道路落石检测存在的目标微小、背景干扰复杂、光照剧烈变化及单模态信息局限等问题，本文提出基于 Mamba 的轻量型红外与可见光渐进式融合检测算法。通过优化 MobileMamba 骨干网络，设计 MambaFusion 融合模块，并结合尺度与位置敏感损失函数 SLS Loss，增强模型对微小目标的敏感性并优化定位精度。实验表明，本文方法在保持轻量和实时推理的同时，有着良好的精度，有效解决了传统单模态方法低能见度失效及现有融合方法难以兼顾性能与速度的问题，为矿区道路安全监控提供有效方案。然而受设备限制，数据集的红外与可见光图像未实现硬件级时空同步，单应性矩阵对齐存在微小偏差。对于图像中占比极小的落石目标，模型检测性能有待提升。因此提升多模态传感器时空同步方案精度，改善极端场景下检测精度是下一步工作方向。

### 作者贡献

马文宁：数据采集，软件设计，创新点设计，论文撰写；孟彬：论文格式检查。

### 利益冲突

所有作者声明无利益冲突

## 参考文献


1. Chen Z F, Yu G Z, Zhang C Y, et al. Driving risk assessment and prevention strategies for autonomous vehicle in open-pits[J]. *J China Coal Soc*, 2023, **48**(4): 1782−1797.
陈志发, 余贵珍, 张传莹, 等. 露天矿区无人驾驶行车风险评估及防控策略仿真研究[J]. 煤炭学报, 2023, **48**(4): 1782−1797.
2. Gu Q H, Li J W, Chen L, et al. Small-scale rockfall detection method based on solid-state lidar for unstructured transportation roads in open-pit mines[J]. *Laser Opto Prog*, 2024, **61**(8): 0812006.
顾清华, 李佳威, 陈露, 等. 基于固态激光雷达的露天矿非结构化运输道路小尺寸落石检测方法[J]. 激光与光电子学进展, 2024, **61**(8): 0812006.
3. Chen Y W, Zhou L D, Bouguila N, et al. BLOCK-DBSCAN: fast clustering for large scale data[J]. *Pattern Recogn*, 2021, **109**: 107624.
4. Shi T D, Zhong D Y, Bi L. A new challenge: detection of small-scale falling rocks on transportation roads in open-pit mines[J]. *Sensors*, 2021, **21**(10): 3548.
5. Chen X D, Han L L, Xiao Y H, et al. 3-D object detection for point clouds based on Voxel-Keypoint Feature Aggregation Network[J]. *Laser Technol*, 2026, **50**(2): 291−299
陈旭东, 韩丽丽, 肖英豪, 等. 基于体素-关键点特征聚合网络的点云 3 维目标检测[J]. 激光技术, 2026, **50**(2): 291−299
6. Zheng W, Tang W L, Chen S J, et al. CIA-SSD: confident IoU-aware single-stage object detector from point cloud[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021: 3555–3562.
https://doi.org/10.1609/aaai.v35i4.16470.
7. Jian Y R, Xiao S. Research on rockfall recognition algorithm based on digital image processing for slopes[J]. *Sci Technol Inn Her*, 2017, **14**(8): 88−89.
简云瑞, 肖硕. 基于数字图像处理的边坡落石识别算法研究[J]. 科技创新导报, 2017, **14**(8): 88−89.
8. Liu L Y, Wu S Y, Zuo Z Y, et al. Research on rockfall detection method of mountain railway slope based on YOLOv3 algorithm[J]. *Comput Sci*, 2021, **48**(S2): 290−294.
刘林芽, 吴送英, 左志远, 等. 基于 YOLOv3 算法的山区铁路边坡落石检测方法研究[J]. 计算机科学, 2021, **48**(S2): 290−294.
9. Gu Q H, Du Y F, Li P F, et al. Road rockfall detection in mining area based on weighted bidirectional feature fusion[J]. *Gold Sci Technol*, 2023, **31**(6): 953−963.
顾清华, 杜艺凡, 李萍丰, 等. 基于加权双向特征融合的矿区道路落石检测[J]. 黄金科学技术, 2023, **31**(6): 953−963.
10. Peng P, Gao L C, Li J C. Road rockfall detection method based on enhanced Yolov8-GCB[J]. *J Chang'an Univ (Nat Sci Ed)*, 2025, **45**(2): 24−35.
彭朋, 高浪超, 李家春. 基于改进 Yolov8-GCB 的公路落石检测方法[J]. 长安大学学报 (自然科学版), 2025, **45**(2): 24−35.
11. Fang Q Y, Han D P, Wang Z K. Cross-modality fusion transformer for multispectral object detection[Z]. arXiv: 2111.00273, 2021.
https://doi.org/10.48550/arXiv.2111.00273.
12. Liu J J, Zhang S T, Wang S, et al. Multispectral deep neural networks for pedestrian detection[C]//*Proceedings of the British Machine Vision Conference 2016*, 2016.
https://doi.org/10.5244/C.30.73.
13. Zheng H J, Ge B, Xia C X, et al. Infrared-visible person re-

identification based on multi feature aggregation[J]. *Opto-Electron Eng*, 2023, **50**(7): 230136.
郑海君, 葛斌, 夏晨星, 等. 多特征聚合的红外-可见光行人重识别[J]. 光电工程, 2023, **50**(7): 230136.

14. Li C Y, Song D, Tong R F, et al. Multispectral pedestrian detection via simultaneous detection and segmentation[C]//*British Machine Vision Conference 2018*, 2018.
15. Zhang H, Fromont E, Lefèvre 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.
https://doi.org/10.1109/WACV48630.2021.00012.
16. Zhang L, Zhu X Y, Chen X Y, et al. Weakly aligned cross-modal learning for multispectral pedestrian detection[C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*, 2019: 5126–5136.
https://doi.org/10.1109/ICCV.2019.00523.
17. Ge B, Xu N, Xia C X, et al. Quadrupl-stream input-guided feature complementary visible-infrared person re-identification[J]. *Opto-Electron Eng*, 2024, **51**(9): 240119.
葛斌, 许诺, 夏晨星, 等. 四流输入引导的特征互补可见光-红外行人重识别[J]. 光电工程, 2024, **51**(9): 240119.
18. Xu J, He X. DAF-Net: a dual-branch feature decomposition fusion network with domain adaptive for infrared and visible image fusion[Z]. arXiv: 2409.11642, 2024.
https://doi.org/10.48550/arXiv.2409.11642.
19. Lai K C, Zhao J, Wang C, et al. Improved mask R-CNN algorithm for pedestrian detection in low-light road conditions[J]. *J Chongqing Univ Technol (Nat Sci)*, 2021, **35**(7): 154−160.
赖坤城, 赵津, 王超, 等. 改进 Mask R-CNN 算法在低光道路环境下行人检测研究[J]. 重庆理工大学学报 (自然科学), 2021, **35**(7): 154−160.
20. Chen S Q, Wang C Q, Zhou Y J. A pedestrian detection method based on YOLOv5s and image fusion[J]. *Electron Opt Control*, 2022, **29**(7): 96−101,131.
陈世权, 王从庆, 周勇军. 一种基于 YOLOv5s 和图像融合的行人检测方法[J]. 电光与控制, 2022, **29**(7): 96−101,131.
21. Jiang Y, Liu R R, Li H Y, et al. Target detection algorithm for vehicle-mounted infrared images based on improved YOLOv5s[J]. *Infrared Technol*, 2024.
姜宇, 刘冉冉, 李杭宇, 等. 基于改进 YOLOv5s 的车载红外图像目标检测算法[J]. 红外技术, 2024.
22. Zhao S P, Yang L P, Zhao X, et al. Object detection in visible light and infrared images based on adaptive attention mechanism[J]. *Infrared Technol*, 2024, **46**(4): 443−451.
赵松璞, 杨利萍, 赵昕, 等. 基于自适应注意力机制的红外与可见光图像目标检测[J]. 红外技术, 2024, **46**(4): 443−451.
23. Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces[Z]. arXiv: 2312.00752, 2023.
https://doi.org/10.48550/arXiv.2312.00752.
24. He H Y, Zhang J N, Cai Y X, et al. MobileMamba: lightweight multi-receptive visual mamba network[C]//*Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025: 4497–4507.
https://doi.org/10.1109/CVPR52734.2025.00424.
25. Han K, Wang Y H, Chen H T, et al. A survey on vision transformer[J]. *IEEE Trans Pattern Anal Mach Intell*, 2023, **45**(1): 87−110.
26. Zhang H W, Zhu Y, Wang D, et al. A survey on visual mamba[J]. *Appl Sci*, 2024, **14**(13): 5683.
27. Wang A, Chen H, Lin Z J, et al. Rep ViT: revisiting mobile CNN from ViT perspective[C]//*Proceedings of the2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024: 15909–15920.
https://doi.org/10.1109/CVPR52733.2024.01506.
28. Zhao Z X, Bai H W, Zhu Y Z, et al. DDFM: denoising diffusion model for multi-modality image fusion[C]//*Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*, 2023: 8048–8059.
https://doi.org/10.1109/ICCV51070.2023.00742.
29. Shen J F, Chen Y F, Liu Y, et al. ICAFusion: iterative cross-attention guided feature fusion for multispectral object detection[J]. *Pattern Recogn*, 2024, **145**: 109913.

## 作者简介



【通信作者】马文宁 (1983-)，女，硕士，讲师，主要从事图像处理、目标检测信号处理、计算机应用技术研究。

E-mail：cctn2015@163.com

# Progressive fusion of infrared and visible light for road rockfall detection in mining areas

Ma Wenning*, Meng Bin

Xi'an Fanyi University, Xi'an, Shaanxi 710105, China

## Abstract

**Objective** Real-time and accurate detection of rockfalls on mining roads is critical for ensuring transportation safety, preventing geological hazards, and maintaining operational continuity in open-pit mining environments. Detecting rockfalls in such scenarios presents substantial challenges due to the small size of targets, irregular shapes, low contrast against the background, drastic illumination variations, and interference from adverse weather conditions such as fog, dust, and rain. Conventional single-modality visual detection methods, particularly those based solely on visible light imagery, often fail under low-visibility conditions, while existing dual-modality fusion approaches typically face limitations including complex network structures, high computational costs, and difficulties in simultaneously achieving real-time performance and high detection accuracy for small targets.

**Methods** To address these challenges, this study proposes a novel progressive infrared–visible light fusion detection framework based on the Mamba state space model (SSM), which is specifically designed to efficiently integrate complementary information from infrared and visible light images while preserving fine-grained details of small targets, even under challenging environmental conditions such as low illumination, strong backlighting, partial occlusion, and adverse weather. The framework consists of three key components: 1) optimized feature extraction using MobileMamba and Wavelet Transform-based downsampling, which reduces feature loss while maintaining high-resolution details, 2) cross-modal feature fusion via the MambaFusion module, enabling adaptive integration of thermal and texture cues across multiple scales, and 3) small-object detection and localization enhancement through a Scale and Location Sensitive Loss (SLS Loss), which explicitly models scale variations and center-point offsets to improve both classification confidence and spatial precision for micro-scale rockfall targets. This multi-component design ensures robust detection in complex, unstructured mining environments and provides a balanced trade-off between computational efficiency and detection accuracy.

First, a Wavelet Transform-based WTSterm downsampling module is introduced to mitigate the loss of fine-grained features during feature extraction. Unlike conventional stride-based convolutional downsampling, the WTSterm module decomposes input images into multiple frequency components, including low-frequency approximations and directional high-frequency details. These components are concatenated along the channel dimension and projected through a 1×1 convolution, effectively preserving edge and texture information of small rock targets while reducing computational overhead. This module is integrated into an optimized MobileMamba backbone network, a lightweight visual SSM architecture that combines multi-kernel convolutions, multi-receptive field feature interaction, and wavelet-enhanced Mamba blocks. The MobileMamba backbone enables efficient extraction of thermal radiation cues from infrared images and rich spatial-texture features from visible light images, while selectively maintaining both shallow fine-grained details and deep high-level semantic features. To reduce computational redundancy, the network removes high-dimensional, small-scale feature branches that contribute little to small-object representation, thereby balancing feature richness and efficiency.

Second, a MambaFusion module is designed to progressively fuse multi-scale features from infrared and visible light modalities. The fusion process leverages Mamba's dynamic state transition and gating mechanisms, which enable adaptive selection of discriminative information while suppressing modality-specific noise. MambaFusion combines the global dependency modeling ability of the SSM with local feature retention through convolution, facilitating cross-modal correspondence at semantic and spatial levels. Multi-scale fusion is performed across large, medium, and small feature maps, with adaptive weighting applied to emphasize salient information according to scene context. This mechanism allows the network to maintain robustness under challenging conditions such as strong backlighting, low illumination, partial occlusion, and airborne dust or smoke, enhancing both target discrimination and representation for small rockfall instances.

Third, to improve small-object detection and localization, the framework incorporates a Scale and Location Sensitive Loss (SLS Loss). This loss function explicitly models variations in target scale and center-point offsets, addressing the insensitivity of conventional losses (e.g., IoU, Dice) to size and position discrepancies. The scale-sensitive component reweights loss contributions according to the predicted

*Correspondence: Ma W N, E-mail: cctn2015@163.com

and true target sizes, focusing training on underrepresented or mispredicted small targets. The location-sensitive component employs polar-coordinate quantification of center-point offsets, providing a strong penalty for positional errors and refining bounding box localization. Together, SLS Loss ensures that the network maintains high sensitivity to micro-scale rockfalls and achieves precise positioning in cluttered or unstructured road environments.

**Results and Discussions** Extensive experiments were conducted on a self-collected open-pit mine rockfall dataset, containing diverse scenes with varying illumination, weather conditions, and target sizes (typically 10–50 cm in diameter). Results demonstrate that the proposed framework achieves precision of 0.856, recall of 0.715, and mAP@0.5 of 0.721, with an inference speed of 16 ms per frame. Compared with conventional single-modality detection and existing dual-modality fusion approaches, the proposed method significantly improves small-target detection performance and robustness while maintaining real-time computational efficiency. In conclusion, the progressive infrared–visible light fusion framework based on MobileMamba and MambaFusion modules effectively integrates complementary thermal and textural cues, preserves fine-grained details of micro-scale rockfall targets, and enhances small-object localization and discrimination in complex visual scenarios. By combining efficient feature extraction, adaptive cross-modal fusion, and a scale- and location-aware loss function, the proposed approach achieves high robustness under diverse environmental conditions, including low visibility, dynamic lighting, dust, and partial occlusion.

**Conclusions** The method offers a reliable, high-performance solution for real-time rockfall detection and monitoring in open-pit mining environments, significantly contributing to safer road operation, reduced operational risk, and improved transportation efficiency. These results demonstrate the potential of lightweight state-space models combined with progressive multi-modal fusion and targeted loss functions for deployment in practical industrial safety and autonomous monitoring systems.

**Keywords:** rockfall detection; infrared and visible light fusion; mamba; small target detection

Ma W N, Meng B. Progressive fusion of infrared and visible light for road rockfall detection in mining areas[J]. *Opto-Electron Eng*, 2026, **53**(2): 250226; DOI: 10.12086/oee.2026.250226

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文