OEE | 光电工程 Opto-Electronic Engineering

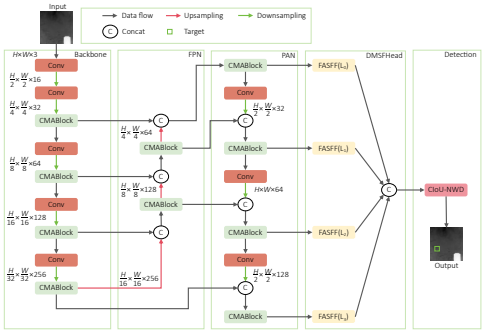
Article

2026 年 6 月, 第 53 卷, 第 6 期

用于红外小目标检测的多尺度特征学习网络

张　策[1], 武丹凤[2,3*], 孙　涛[4*], 匡敏驰[5]

[1]辽宁工程技术大学软件学院，辽宁 葫芦岛 125105；
[2]北京联合大学，北京市信息服务工程重点实验室，北京 100101；
[3]北京联合大学机器人学院，北京 100027；
[4]中国人民解放军 63963 部队，北京 100072；
[5]清华大学精密仪器系，北京 100084



摘要： 红外小目标检测 (IRSTD) 在预警系统、精确制导等领域具有重要应用价值，但现有方法存在计算开销大、小目标易受背景干扰以及 IoU 度量方法对小目标定位偏差敏感等问题。针对上述问题，提出一种多尺度特征学习网络 MFLNet。首先设计 CMABlock 特征提取模块，通过线性注意力机制在轻量化前提下强化微弱目标感知，且提升特征判别性；其次设计 DMSFHead 检测头，通过动态加权融合机制自适应聚合多尺度特征，平衡细节保留与语义增强，提升特征利用率；此外，采用归一化高斯 Wasserstein 距离 (NWD) 替代传统 IoU 度量方法，并设计 CIoU-NWD 混合损失，降低训练过程中模型对小目标定位偏差的敏感性。选取 IRSTD-1k、NUDT-SIRST、NUAA-SIRST 三个数据集进行实验，结果表明 MFLNet 在三个数据集上的精度、召回率与 F1 分数均优于当前 IRSTD 的主流 SOTA 方法，且消融实验验证了本文网络各组件的有效性。本文提出的多尺度特征学习网络 MFLNet 能有效解决红外小目标检测中的低信杂比、特征利用率低与 IoU 敏感等问题。

关键词： 红外小目标；目标检测；线性注意力；多尺度特征；二维高斯分布

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

DOI: 10.12086/oee.2026.250358 | CSTR: 32245.14.oee.2026.250358 | 中图分类号：TP391.4 | 文献标识码：A

张策，武丹凤，孙涛，等. 用于红外小目标检测的多尺度特征学习网络 [J]. 光电工程，2026，**53**(6): 250358

Zhang C, Wu D F, Sun T, et al. Multi-scale feature learning network for infrared small object detection[J]. *Opto-Electron Eng*, 2026, **53**(6): 250358

1　引　言

目标检测是计算机视觉领域的核心技术之一，其中，红外小目标检测 (infrared small target detection, IRSTD) 融合计算机视觉与红外遥感技术，凭借红外成像可在低能见度或夜间捕捉目标热辐射[1] 的独特优势，在交通管理[2]、海上救援[3]、预警系统[4] 和精确制导[5] 等关键领域发挥着至关重要的作用。

红外小目标具有显著的固有特性。一是尺寸极小，通常在图像中以数个至数十个像素的斑点形式存在，导致其特征表达有限。二是红外图像背景噪声与杂波较多，目标易被淹没，导致图像对比度低、信杂比 (signal-to-clutter ratio, SCR) 低。三是目标数量少且在图像中占比极低，目标与背景样本极度不平衡，导致目标易被背景

收稿日期：2025-11-21；修回日期：2026-03-06；录用日期：2026-03-10；网络出版日期：2026-06-25
基金项目：时空信息精密感知技术全国重点实验室开放基金资助项目 (STL2023-B-06-01(K))；北京联合大学科研项目 (ZK20202201)
*通信作者：武丹凤，jqrdanfeng@buu.edu.cn；孙涛，43176757@qq.com
版权所有©2026 中国科学院光电技术研究所

噪声掩盖。

针对这些特点，早期红外小目标检测方法以手工设计的特征提取技术为主。这些方法主要分为三类，一是基于滤波器的方法，如最大中值[6]、Tophat[7]、二维最小均方滤波[8]、二维变分模态分解[9]和应用于无人机探测系统的双边滤波与改进 Top-hat 算法[10]，这类方法在简单背景下表现尚可，但在复杂场景下的检测性能与鲁棒性仍有待提升；二是基于局部对比度的方法，包括加权增强局部对比算法[4]、三层局部对比度算法[11]、改进的局部对比度算法[12]和相对局部对比度算法[13]，这类方法依赖目标亮度高于邻域的假设，在检测昏暗目标时存在明显局限性；三是基于低秩分解的方法，如 IPI 算法[14]、非凸秩近似最小化关节 $L_{2,1}$ 范数[15]、重加权红外贴片张量[16]、基于张量核范数[17]的部分和，以及近期采用随机化技术与时空张量以提升计算效率的改进方法[18]等，这类方法基于低秩背景和稀疏目标假设实现目标背景分离，但对背景杂波适应性较弱，易受干扰。

深度学习技术的兴起为红外小目标检测带来突破。相关方法可分为基于分割的方法和基于检测的方法两类。

基于分割的方法通过逐像素阈值分割生成包含目标位置和大小的分割掩码。例如，Wang 等[19]提出了生成式对抗网络 (GAN)，实现训练中漏检与虚警的均衡；Dai 等[20]设计了不对称上下文调制 (ACM) 模块，结合自顶向下和自底向上的点状注意机制增强语义和空间细节编码；Dai 等[21]提出模型驱动的注意局部对比网络 (ALCNet)，融合标记数据和领域知识解决建模不准确等问题；Zhang 等[22]提出了注意力引导的金字塔上下文网络 (AGPCNet)，显著提升了复杂背景下红外小目标检测的精度与鲁棒性；Zhang 等[23]构建红外形状网络 (ISNet)，专注红外小目标形状信息检测；Li 等[24]提出了密集嵌套注意网络 (DNA-Net)，缓解池化层导致的深度信息损失；Hou 等[25]设计了一种将手工特征方法与 CNN 相结合的鲁棒红外小目标检测网络 (RISTDNet)；Chen 等[26]开发了分层重叠小贴片变压器 (HOSPT)，替代 CNN 卷积核实现多尺度特征编码，解决图像中远程依赖关系建模的挑战；Zhang 等[27]改进 SAM 得到 IRSAM 模型，增强了红外小目标的特征表示。

基于检测的方法直接输出目标的位置与尺度信息，计算效率更高。例如，Li 等[28]提出了结合输入图像超分辨率增强与 YOLOv5 结构改进的方法；Dai 等[29]提出了一级级联细化网络 (OSCAR)，解决目标固有特征不足与边界框回归不准确问题；Yao 等[30]开发了轻量级网络，融合传统滤波方法与卷积一级目标检测 (FCOS) 提升红外小目标响应；Du 等[31]采用帧间能量积累 (IFEA) 增强了机制放大运动目标的能量，并搭配小交联策略解决了样本误识别；Ju 等[32]通过图像滤波模块实现了类似目标增强。Yang等[33]提出 EFLNet，使用动态头进行红外小目标检测。

尽管深度学习方法取得显著进展，但现有多数方法仍将 IRSTD 视为分割任务[34-37]，这存在资源占用大、实时性差的缺陷。部分方法已将 IRSTD 建模为检测任务[29,31,38-39]，但仍存在一些问题，一是小目标信号弱、背景杂波多，导致模型难以精准感知目标；二是即使提取到微弱特征，模型对多尺度特征的孤立使用也会造成有效信息浪费，特征利用率低；三是模型在训练过程中对小目标定位偏差敏感，导致特征提取阶段的优势无法通过训练转化为精准检测结果。

针对上述挑战，本文提出了一种多尺度特征学习网络 MFLNet (multi-scale feature learning network)，以检测范式解决 IRSTD 问题，既避免分割方法资源占用大、实时性差的缺陷，又提高了检测准确性。首先，设计了包含 MILA 线性注意力机制[40]的特征提取模块 CMABlock，强化模型对红外小目标的关注，实现目标和背景的有效区分；其次，针对红外小目标易被背景淹没的问题，设计了 DMSFHead 提升小目标特征的利用率，显著提高检测精度；最后，为缓解边界框偏移带来的显著影响，本文采用二维高斯分布重构边界框，并结合归一化高斯 Wasserstein 距离 (normalized Wasserstein distance，NWD) 解决红外小目标对 IoU 高度敏感的问题。本文工作的核心在于通过有洞察力的架构设计与协同策略，构建一个深度适配红外物理特性的检测框架。

具体而言，本文主要贡献如下：

1) 针对红外小目标在低信杂比下的弱特征感知问题，提出 CMABlock 模块，引导模型在噪声中聚焦微弱目标区域，为后续特征处理提供高判别性的多尺度特征基础。

2) 针对红外小目标尺度微小且多尺度特征利用效率低的问题，提出基于多层级特征融合机制的 DMSFHead，通过动态平衡不同层级特征的贡献，解决红外小目标在浅层特征与深层特征间的表征冲突。

3) 针对红外小目标因像素占比低导致 IoU 度量敏感、训练不稳定的问题，重新评估 IoU 在红外小目标上的失效机理，引入 NWD 度量并设计 CIoU-NWD 混合损失函数，有效缓解模型对红外小目标定位偏差的敏感性，稳定训练过程。

2 检测方法设计

2.1 整体架构

本文提出的 MFLNet 整体架构如图 1 所示，由主干网络、特征金字塔网络和检测头三部分组成，以实现复杂背景下红外小目标的高精度实时检测。CMABlock 作为主干网络核心，负责从复杂背景中挖掘小目标的有效特征；特征金字塔网络对其输出的多尺度特征进行初步整合后，送入 DMSFHead 实现高效利用；最后通过改良的损失函数优化模型参数，使得前两模块的特征优势能转化为精准的检测结果，这三个部分协同工作，共同提升了模型的检测性能。具体而言，主干网络部分首先以 CMABlock 为核心构建多级特征提取结构，在接收到输入图像后将对其进行逐步下采样生成基础特征图；之后特征金字塔部分采用上采样与横向连接，将深层语义信息与浅层空间细节融合，各层级再经 CMABlock 强化特征表达生成四个尺度的特征图；之后将其送入 DMSFHead 进行多尺度融合与预测，通过对不同层级的特征分别进行处理，提升模型对红外小目标特征的利用率；最终由 CIoU-NWD 评估检测结果，计算损失并指导模型对检测过程进行优化。

MFLNet 的整体架构设计体现了针对红外小目标的系统性优化思想。CMABlock、DMSFHead 与 CIoU-NWD 损失构成了一个协同增强的闭环，CMABlock 提供高判别性的基础特征，DMSFHead 在此基础上实现跨尺度特征的高效利用，而 CIoU-NWD 损失则确保了前两者的优势能在训练中被稳定地转化为检测性能。这种组件间的协同设计，是 MFLNet 有效性的关键。整个架构兼顾轻量化，以模块化设计实现了高效的特征提取、多尺度特征融合与稳定的训练过程，为红外小目标检测提供了有力模型支撑。

2.2 CMABlock

红外小目标具有低分辨率、弱边缘信息以及与背景相似度高的特点，这对传统卷积神经网络的特征提取能力提出更高要求。然而，用于增强模型对局部细节关注的通用视觉注意力机制在红外场景下面临挑战，这类机制通常通过重新加权来强调特征图中的显著区域，但在

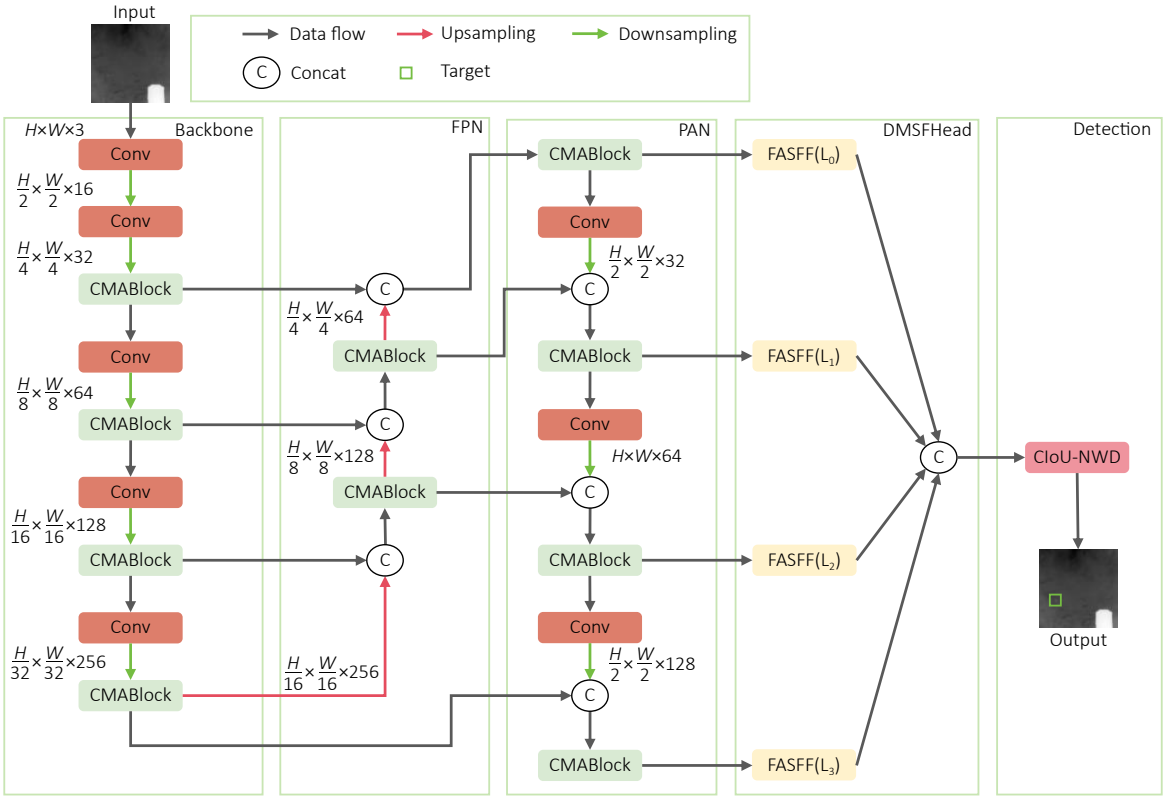


图 1 MFLNet 网络结构图

Fig. 1 MFLNet structure diagram

低信杂比的红外图像中，微弱的小目标往往并非视觉显著区域，直接应用此类机制反而可能抑制本就微弱的目标信号，放大背景噪声。

为克服这一局限，本文设计着重于利用位置信息来增强对微弱信号的感知，本文提出了一种深度融合 CSP (cross stage partial) 架构[41] 与线性注意力机制的特征提取模块 CMABlock (CSP with namba-inspired linear attention block)。该模块的核心在于，在 CSP 架构的一个分支中创新性地引入集成旋转位置编码 (RoPE) 的 MILA 子模块，从而实现对输入特征图中潜在目标区域的、位置感知的多层次建模，有效提升模型在复杂红外背景下的特征感知与判别能力。

如图 2 的 CMABlock 结构图所示，该模块使用分流后聚合的设计，对于输入的特征图 $\boldsymbol{x}\in\mathbb{R}^{H\times W\times C_1}$，其中 H 和 W 分别表示特征图的高和宽，C_1 为输入特征图的通道数，它通过模块后最终会输出特征图 $\boldsymbol{x}'\in\mathbb{R}^{H\times W\times C_2}$，其中 C_2 为输出特征图的通道数。输入特征图 $\boldsymbol{x}$ 首先经 1×1 卷积将通道数映射为 $2C$，其中 $C=C_2\times e$，e 为控制中间通道数的扩展因子，$e\in(0,1]$；随后调整后的特征沿通道维度分裂为转换分支 a 和残差分支 b。

残差分支 b 保留原始特征分布，转换分支 a 用于精细化特征增强。转换分支送入集成线性注意力与位置编码的特征转换单元 MILABlock 进行特征增强，特征图经其处理后与残差分支沿通道维度拼接，融合后的特征图通过 1×1 卷积压缩通道后，再经过特征分布归一化层与非线性特征激活层，得到最终输出特征 $\boldsymbol{x}'$。

本文使用的 MILA (mamba-inspired linear attention) 模块整合了线性注意力的并行计算优势与 Mamba[42] 的高效设计，实现了视觉任务中的高效序列建模。该模块通过线性注意力捕捉全局特征依赖，并结合 RoPE 旋转位置编码为特征注入明确的空间位置信息，以及通过深度卷积强化局部特征关联，从而引导模型聚焦于红外图像中缺乏纹理但位置关键的微弱目标区域。该设计在减少冗余计算的同时，借助位置增强的注意力机制显著提升了模型在噪声背景中对局部目标的感知能力。

MILABlock 的结构如图 3 所示，输入特征图首先被重塑为序列形式，即将形状由 (C,H,W) 转换为 $(H\times W,C)$，以适配后续序列建模。随后，特征经过 3×3 深度卷积增强局部位置信息，并通过残差连接与原始输入特征融合，以强化特征流动。

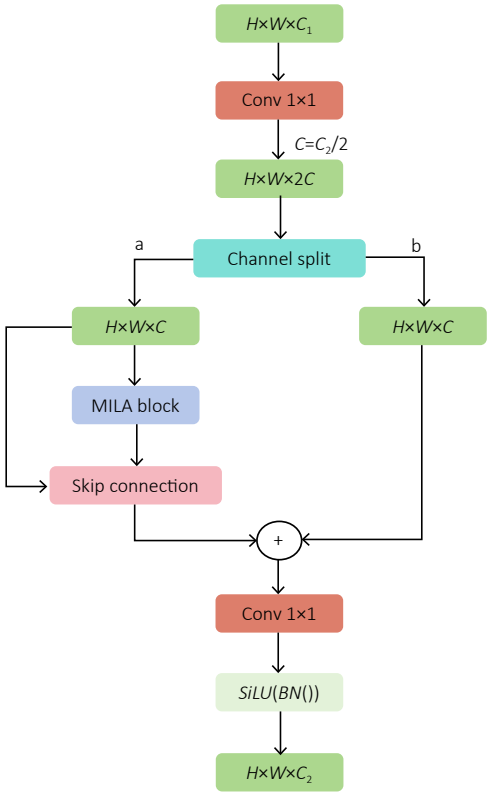


图 2 CMABlock 结构图

Fig. 2 CMABlock structure diagram

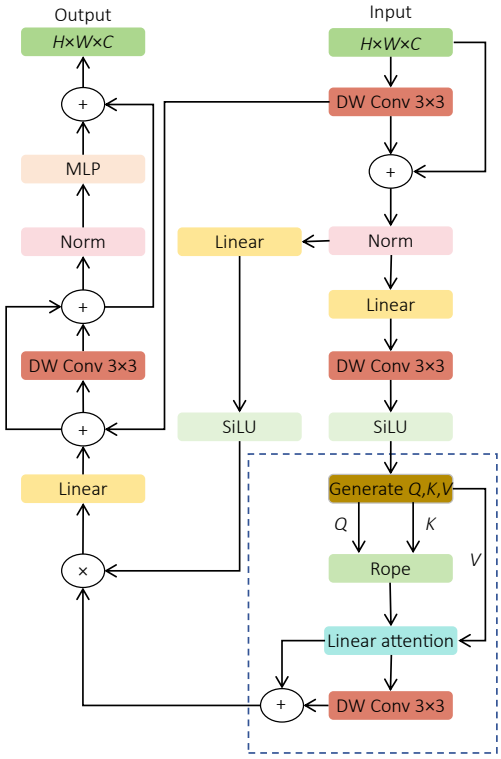


图 3 MILABlock 结构图

Fig. 3 MILABlock structure diagram

之后特征进入规范化与转换阶段，首先经过 LayerNorm 稳定特征分布，再通过线性投影调整特征维度，并重塑回空间维度(H,W,C)，随后通过 3×3 深度卷积与 SiLU 激活函数完成非线性特征转换。转换后的特征输入至 LinearAttention 部分进行全局聚合，首先通过线性层生成查询 ($\boldsymbol{Q}$) 和键 ($\boldsymbol{K}$)，值 ($\boldsymbol{V}$) 直接使用当前特征；随后 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 被重塑为空间维度并施加 RoPE 旋转位置编码，以注入位置信息；同时，$\boldsymbol{V}$ 经深度卷积增强局部关联，补充局部偏置。

旋转位置编码 (RoPE) 作为模块的位置感知核心，它通过三角函数生成与位置相关的旋转矩阵，将特征转换为复数形式并施加旋转后再转回实数，无需可学习参数即可感知序列中令牌的位置关系，在保持并行计算的同时增强对输入顺序的敏感性。

在聚合阶段，LinearAttention 按线性注意力机制计算，先计算 $\boldsymbol{K}$ 与 $\boldsymbol{V}$ 的矩阵乘得到全局上下文特征 ($\boldsymbol{KV}$)，再通过 $\boldsymbol{Q}$ 与 $\boldsymbol{KV}$ 的乘积分步聚合，并引入基于 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 均值计算的归一化因子 z 以稳定训练，全过程保持了 O(N) 的线性复杂度。

随后聚合结果与由线性投影和 SiLU 生成的门控特征逐元素相乘，之后特征通过 3×3 深度卷积进一步增强位置信息，最后经过 LayerNorm 规范化后，由 MLP 完成特征维度扩展与转换，并经过残差连接，最终重塑为原始特征图形态(C,H,W)输出。

CMABlock 通过 CSP 架构与集成旋转位置编码的线性注意力机制的融合，既实现了轻量化，又针对红外小目标低信杂比的特点强化了特征的判别性，为模型后续的多尺度特征融合提供了高质量的特征输入。

2.3 DMSFHead

基于模型提取的小目标特征方法，传统的检测头存在权重固定与尺度适配差的问题，导致有效特征被浪费。为解决此问题，本文设计了具有多层级特征融合机制的 DMSFHead。特别地，针对红外小目标尺寸微小、特征信息有限的特点，其在特征金字塔中的表征面临一个实际问题，在高分辨率特征图中尚存的微弱像素级信息，经过几次下采样后，在较低层特征图中极易丢失。这意味着，通用的三层动态融合架构在融合时已经难以获取到足够丰富的原始细节信息来进行有效的加权与平衡。

因此，我们对融合架构进行了针对性的扩展，将核心融合机制从通用的三层，重新设计为适配红外小目标的四层特征金字塔的 DMSFHead。这一设计使动态加权网络能够在更早、更高分辨率的层级介入，确保那些关键的微弱细节在丢失前就被有效地捕获并纳入融合决策，从而实现对红外小目标多尺度特征的更充分利用。

该检测头通过动态加权融合机制充分利用各层级特征的互补信息，生成高信噪比的融合特征，以提升模型对多尺度特征的利用率，强化其处理多尺度特征时的性能，进一步提高对红外小目标的检测精度。

红外小目标在不同分辨率特征图中呈现极强的尺度敏感性，在高分辨率特征图中目标易受噪声干扰导致特征淹没，在低分辨率特征图中目标背景区分度提升但空间定位精度下降。为此，本文设计出可接收四个不同尺度特征图的 DMSFHead，这四个特征图分别为$\boldsymbol{x}_0$、$\boldsymbol{x}_1$、$\boldsymbol{x}_2$和$\boldsymbol{x}_3$，其中$\boldsymbol{x}_i \in \mathbb{R}^{C_i \times H_i \times W_i}$，$H_i$和$W_i$分别为特征图的高和宽，$C_i$为通道数，它们的分辨率逐级降低，但语义信息逐级增强，DMSFHead 通过设计 FASFF (feature aggregation and selective feature fusion) 核心模块实现跨尺度特征的自适应加权特征融合机制[43]。如图 4 所示，各输入特征通过其对应层级的 FASFF 模块，与相邻两层级的特征图完成自适应加权融合，融合后的特征用于输出原始检测结果。

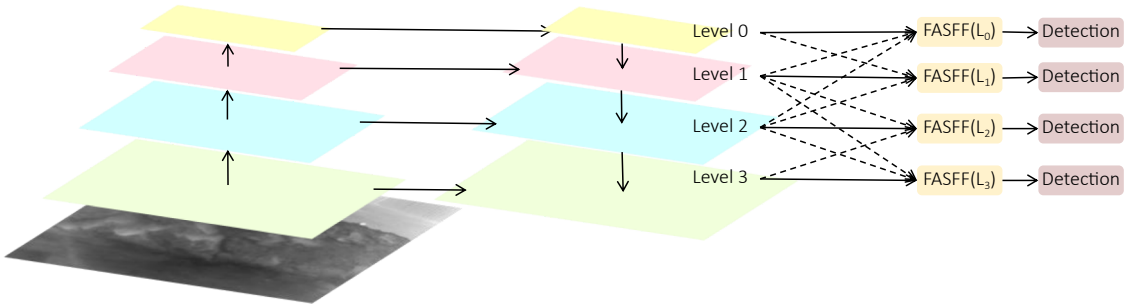


图 4 DMSFHead 检测过程图

Fig. 4 DMSFHead detection process diagram

FASFF 作为 DMSFHead 的核心模块，承担多尺度特征动态融合功能。如图 5 以第二级融合模块为例，首先将输入的 3 个不同尺度特征$\boldsymbol{x}_1$、$\boldsymbol{x}_2$和$\boldsymbol{x}_3$通过上采样或下采样转换为同维度、同分辨率的适配特征$\boldsymbol{x}_1^*$、$\boldsymbol{x}_2^*$、$\boldsymbol{x}_3^* \in \mathbb{R}^{C_l \times H \times W}$，随后轻量级权重学习网络通过 1×1 卷积压缩各适配特征，生成中间权重特征$\boldsymbol{w}_1$、$\boldsymbol{w}_2$、$\boldsymbol{w}_3 \in \mathbb{R}^{8 \times H \times W}$，拼接后经 1×1 卷积映射为 3 通道原始权重图$\boldsymbol{w}_{\text{raw}} \in \mathbb{R}^{3 \times H \times W}$，再通过 softmax 归一化得到位置敏感的动态权重，具体如下：

$$\boldsymbol{\omega}_k(x,y) = \frac{\exp(\boldsymbol{w}_{\text{raw}}[k,x,y])}{\sum_{k'=1}^{3}\exp(\boldsymbol{w}_{\text{raw}}[k',x,y])},\ (k=1,2,3), \tag{1}$$

式中：$\boldsymbol{\omega}_k(x,y)$与对应位置的目标响应强度正相关，可在红外小目标区域自动提升高分辨率特征的权重以保留细节，在背景区域通过权重平均抑制杂波。

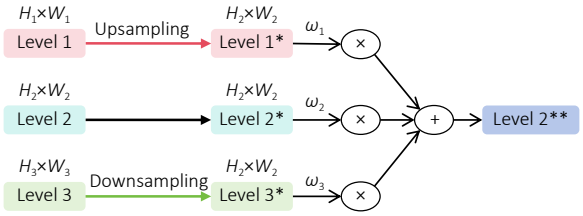


图 5 FASFF 处理过程图

Fig. 5 FASFF processing diagram

最终，融合特征$\boldsymbol{x}_2^{**}$通过加权求和生成，具体如下：

$$\boldsymbol{x}_2^{**}(x,y) = \sum_{k=1}^{3}\boldsymbol{\omega}_k(x,y)\cdot\boldsymbol{x}_2^*(x,y). \tag{2}$$

最后输出的高信噪比特征有效解决了红外小目标在单一尺度下易被噪声掩盖、尺度波动大的问题，显著增强了红外小目标与背景的区分度。

DMSFHead 使用动态权重学习平衡不同尺度特征的贡献，结合专为红外小目标多尺度矛盾定制的四层融合架构，实现了对不同层次特征的有效整合，进而增强模型对红外小目标的定位和分类能力。

2.4 归一化高斯 Wasserstein 距离

CMABlock 的特征感知与 DMSFHead 的特征融合为高精度的红外小目标检测提供了基础，但小目标在图像中像素占比过小，会导致模型训练收敛困难，无法充分发挥特征处理阶段带来的优势。因此，本文将改进损失函数，通过优化训练过程，提高模型检测小目标的能力。

传统的 IoU 通过计算两边界框交集与并集的比例衡量重叠度，其计算公式为

$$IoU = \frac{A \cap B}{A \cup B}. \tag{3}$$

其衍生的 GIoU、DIoU 和 CIoU 应用广泛，其中 CIoU 性能较均衡。对于预测框$R=(c_x,c_y,w,h)$，其中$b=(c_x,c_y)$为预测框的中心坐标，w 和 h 分别为预测框的宽度和高度；Ground truth (GT) 框 $R^{\text{gt}}=(c_x^{\text{gt}},c_y^{\text{gt}},w^{\text{gt}},h^{\text{gt}})$，其中$b^{\text{gt}}=(c_x^{\text{gt}},c_y^{\text{gt}})$为 GT 的中心坐标，$w^{\text{gt}}$和$h^{\text{gt}}$分别为 GT 的宽度和高度。CIoU 定义如下：

$$CIoU = IoU - \frac{\rho^2(b,b^{\text{gt}})}{c^2} - \alpha v, \tag{4}$$

式中：$\rho(\,)$为欧几里得距离；c为能框住两个边界框的最小矩形的对角线长度；

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan\frac{w}{h}\right)^2, \tag{5}$$

$$\alpha = \frac{v}{(1-IoU)+v}. \tag{6}$$

然而在常规目标检测中常用的这些 IoU 度量，应用于红外小目标检测时表现出极强敏感性。尽管预测框和 GT 的位置仅轻微偏移，IoU 值就可能出现显著波动。如图 6 所示，红外小目标的微小位置偏移可使 IoU 从

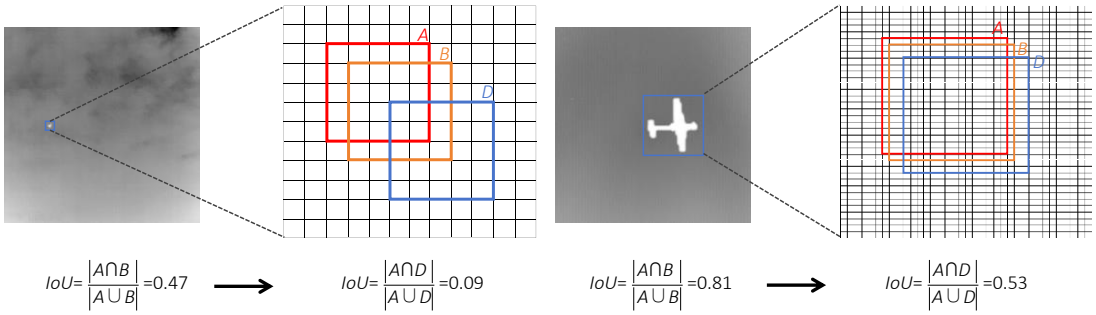


图 6 IoU 敏感性分析图

Fig. 6 IoU sensitivity analysis diagram

0.47 骤降至 0.09；而正常尺寸的目标在相同偏移下，IoU 仅从 0.81 降至 0.53。

这种高敏感性为红外小目标检测的训练带来挑战，正负样本间相似度偏高导致网络难以有效收敛，更极端的是，红外小目标有时只占图像中的一两个像素，这会使真实框和所有预测框的 IoU 均低于阈值，导致图像中无正样本。因此，需要更合适的评价指标来准确评估检测效果。

IoU 本质是计算样本间的相似度，只是它对目标尺度变化敏感。为此，本文引入对微小目标更友好的 NWD 作为新度量。Wasserstein 距离可度量低重叠目标框的相似性，且对尺度不敏感。这正能解决红外小目标训练中正负样本相似度高、正样本稀疏的问题。具体做法为，将红外小目标的边界框建模成二维高斯分布来进行计算，具体定义如下：

对于边界框$R=(c_x,c_y,w,h)$，其中(c_x,c_y)为中心坐标，w 和 h 分别为宽度和高度。该边界框将被建模为均值向量 $\boldsymbol{\mu}$ 和协方差矩阵 $\boldsymbol{\Sigma}$ 的高斯分布$\mathbb{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$，其中：

$$\boldsymbol{\mu}=\begin{bmatrix} c_x \\ c_y \end{bmatrix},\ \boldsymbol{\Sigma}=\begin{bmatrix} \left(\frac{w}{2}\right)^2 & 0 \\ 0 & \left(\frac{h}{2}\right)^2 \end{bmatrix}.$$

值得注意的是，协方差矩阵的对角线元素与边界框的宽高相关，用来反映目标尺度信息。由以上定义可得出二维高斯分布的概率密度函数为

$$f(\boldsymbol{x}\mid\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)}{2\pi\sqrt{|\boldsymbol{\Sigma}|}},\tag{7}$$

式中：$\boldsymbol{x}=(x,y)$为空间坐标；$|\boldsymbol{\Sigma}|$为协方差矩阵的行列式。

对于两个边界框 A 和 B，其对应的高斯分布分别为$\mathbb{N}_1(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$和$\mathbb{N}_2(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$，二者之间的 Wasserstein 距离 (二阶) 定义为

$$W^2(\mathbb{N}_1,\mathbb{N}_2)=\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|^2+\left\|\boldsymbol{\Sigma}_1^{1/2}-\boldsymbol{\Sigma}_2^{1/2}\right\|_{\mathrm{F}}^2,\tag{8}$$

式中：$\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|^2$为两个均值向量的欧氏距离平方，反映边界框中心位置的偏差；$\left\|\boldsymbol{\Sigma}_1^{1/2}-\boldsymbol{\Sigma}_2^{1/2}\right\|_{\mathrm{F}}^2$为两个协方差矩阵平方根的 Frobenius 范数平方，反映边界框尺度 (宽高) 的偏差；$\|\cdot\|_{\mathrm{F}}$为 Frobenius 范数，用于度量矩阵的“大小”。

为了便于将此距离值作为评价指标，需要将其限制在 [0,1] 范围内，所以对 Wasserstein 距离进行指数归一化[44]，得到 NWD：

$$NWD(A,B)=\exp\left(-\frac{W(A,B)}{C_n}\right),\tag{9}$$

式中：$W(A,B)$ 为边界框 A 和 B 对应高斯分布的 Wasserstein 距离 (取平方根)；C_n 是根据数据集调整的常数，用于控制归一化的尺度。

NWD 通过分布建模有效缓解了 IoU 的敏感性，为训练提供了平滑的相似度信号。然而，若完全依赖 NWD，其全局分布匹配的特性可能会弱化对边界框的宽高比与中心点等精确几何属性的约束。为此，本文将 NWD 与 CIoU 结合，构建 CIoU-NWD 混合损失函数。该设计能同时利用 NWD 对微小偏移的平滑性以稳定训练，以及 CIoU 对几何属性的精确约束以保证定位精度，从而为红外小目标检测提供了一个更鲁棒的优化目标。

本文提出的 CIoU 与 NWD 混合的损失函数通过权重参数λ来动态平衡二者的贡献，具体如下：

$$loss=\lambda\times l_{\mathrm{CIoU}}+(1-\lambda)\times l_{\mathrm{NWD}},\tag{10}$$

式中：$\lambda\in[0,1]$，λ为与数据集相关的权重参数，l_{CIoU}为 CIoU 几何回归损失，l_{NWD}为 NWD 分布相似度损失。

3 实验设计

3.1 实验设置

为验证本文所提网络的有效性和先进性，实验采用 IRSTD-1k[23]、NUDT-SIRST[24] 和 NUAA-SIRST[20] 三个公开红外小目标数据集对网络进行训练与测试。这些数据集涵盖了从极端微小到典型尺寸的红外小目标，统计表明，IRSTD-1k 与 NUDT-SIRST 数据集中大部分目标的面积在数十至百余像素之间，且包含宽度或高度仅为 1 pixel 的极端微小目标，NUAA-SIRST 则补充了平均宽度与高度在 10~20 pixels 之间的典型小目标样本，共同确保了模型训练与测试的全面性。具体而言，IRSTD-1k 数据集包含 1000 幅真实红外图像，背景覆盖 12 类真实场景；NUDT-SIRST 数据集包含 1327 幅图像，其中包含合成目标叠加真实背景模拟的复杂图像；NUAA-SIRST 数据集包含 427 幅真实红外图像，标注了 480 个小目标。为确保模型评估结果的客观性与准确性，所有数据集均按 6:2:2 的比例随机划分为训练集、验证集与测试集。

模型的训练和测试均在 NVIDIA GeForce RTX 3080 GPU 上完成，显存大小为 10 GB，使用 PyTorch 1.13 深度学习框架，采用 CUDA 11.7 对 GPU 进行加速。我们

在预实验中发现模型在约 250 个 epoch 后验证集损失趋于平稳，因此将模型训练周期设定为 300 个 epoch，同时避免过拟合。基于 GPU 显存容量与训练稳定性的权衡，将批量大小设为 8，这样设定可以在硬件允许范围内使模型在小目标样本稀疏、背景多变的情况下进行更稳健的梯度更新和学习。超参数 C_n 设置为 12.8，该值参考了 NWD[44] 为小目标检测提出的典型取值，为适配红外小目标的特性，基于样本充足且由真实场景构成的 IRSTD-1k 数据集，在其验证集上确认了该值的有效性，最终采用此数值。

与基于模型的 IRSTD 方法相比，深度学习方法表现出明显更优越的性能，因此本文的方法将与基于深度学习的当前主流 SOTA 方法进行比较，包括 MDvsFA[19]、AGPCNet[22]、ACM[20]、ISNet[23]、ALCNet[21]、DNANet[24]，并且所有模型都使用相同的划分好的数据集进行训练，并将其余参数保持在默认值。

3.2 评价指标

为实现对网络的客观评价，本文选取 3 种广泛使用的评价指标进行定量分析，包括精度、召回率和 F1。每个指标的定义如下：

精度 (Precision)：衡量模型预测为正样本的结果中真实真样本的比例，反映模型抗假阳性能力，精度越高则表明模型在复杂背景下的抗干扰性能力越强，其计算公式为

$$Precision = TP/(TP+FP), \tag{11}$$

式中：TP (true positive) 表示正确检测到的目标数量；FP (false positive) 表示误检为目标的背景或噪声数量。

召回率 (Recall)：衡量真实正样本中被成功检测出的比例，反映模型捕捉真实小目标的能力，召回率越高则表明模型对低信噪比目标漏检率越低，其计算公式为

$$Recall = TP/(TP+FN), \tag{12}$$

式中：FN (false negative) 表示未被检测出的真实目标数。

F1 分数 (F1-score)：精度与召回率的调和平均，用于平衡二者矛盾，更全面反映在目标与背景极度不平衡场景下模型的综合性能，其计算公式为

$$F1 = 2\cdot Precision\cdot Recall/(Precision+Recall). \tag{13}$$

4 数据分析

4.1 对比实验

表 1 呈现了不同检测方法的定量对比结果，其中，加粗数字代表所属指标的最高值，下划线则为次高值，下文中的其他表格表达方式相同。本文所提方法在 IRSTD-1k、NUDT-SIRST 及 NUAA-SIRST 三个数据集上的各项评估指标均处于领先水平，充分验证了其有效性。

从表 1 可见，在三个数据集上，本方法的 F1 分数均显著优于第二名 DNANet。这主要得益于 DMSFHead 对多尺度特征的高效利用，有效减少了复杂背景下的虚警并提升了弱小目标的检出率。

当前多数深度学习方法将红外小目标检测归为像素级分割任务，这类方法需输出像素级分割结果，但其对细节误差较为敏感，即使微小偏差也易引发虚警或漏检，且这类方法对红外小目标检测中存在的目标与背景不平衡问题，以及训练过程稳定性欠佳的问题关注不足，因此在目标级检测任务中表现欠佳，对应的精度、召回率及 F1 分数均相对较低。

相比之下，本文设计的 CMABlock 通过对特征图的

表 1 本文方法与其他 SOTA 方法在不同数据集上的定量结果比较

Table 1 Quantitative comparison of results between the proposed method and other state-of-the-art (SOTA) methods on different datasets

Method	IRSTD-1k			NUDT-SIRST			NUAA-SIRST		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
MDvsFA	0.550	0.483	0.475	0.608	0.192	0.262	0.845	0.507	0.597
ACM	0.679	0.605	0.640	0.732	0.745	0.738	0.765	0.762	0.763
ALCNet	0.843	0.656	0.738	0.868	0.772	0.817	0.848	0.780	0.813
ISNet	0.718	0.741	0.729	0.742	0.834	0.785	0.820	**0.847**	0.834
AGPCNet	0.415	0.470	0.441	0.368	0.684	0.479	0.390	0.810	0.527
DNANet	0.768	0.721	0.744	0.914	0.889	0.901	0.847	0.836	0.841
Ours	**0.902**	**0.794**	**0.845**	**0.947**	**0.944**	**0.945**	**0.941**	0.843	**0.889**

多层次建模，有效提升了模型对红外小目标的关注能力，再结合 DMSFHead 以及 NWD 度量，模型可更充分地学习红外小目标的特征，从而在红外小目标检测任务中展现出更优性能。

4.2 泛化实验

为验证 MFLNet 的泛化能力，本文进行一对一交叉数据集泛化实验，具体方法是使用一个数据集的训练集完成模型训练，随后将训练后的模型在另一个未参与训练的数据集的测试集上进行评估，以此模拟模型面对全新数据分布时的性能。所有训练与评估设置均与 3.1 节保持一致，确保测试公平性。评估指标包括精度、召回率、F1 分数与 AP50，结果如表 2 所示。

整体而言，MFLNet 在全部六组测试中均表现稳定，未出现极端的性能退化。这证明其并未过拟合于单一数据集的特定分布，而是学习到了更具通用性的红外小目标表征，具备了良好的跨数据集泛化能力。

深入分析发现，以 IRSTD-1k 为训练源时，模型泛化性能最为突出，这得益于 IRSTD-1k 数据涵盖场景丰富、目标尺度分布均衡，而 MFLNet 中的 CMABlock 模块通过线性注意力机制有效抑制了多样背景的干扰，从中提炼出的目标特征本身就具有更强的场景无关性与鲁棒性，从而在跨域测试中受益。此外，模型展现了优秀的尺度泛化能力，即使使用目标尺寸普遍较大的 NUAA-SIRST 进行训练，模型在目标尺寸更小的 IRSTD-1k 和 NUDT-SIRST 上测试，F1 分数仍能保持在 0.697 和 0.770。这验证了 DMSFHead 模块所实现的多尺度动态融合机制的有效性，它使模型能够学习并协调不同层级的特征表示，而非机械记忆特定尺度，从而实现了向未知尺度的良好迁移。

值得讨论的是，当使用合成特性明显的 NUDT-SIRST 训练并迁移至真实场景 IRSTD-1k 时，模型性能出现一定波动。这恰好反映了合成数据与真实数据之间存在固有的域差异，是评估泛化性的一个典型挑战场景。尽管如此，MFLNet 在该极具挑战性的跨域测试中依然保持了有效的检测能力，进一步证明了其核心架构在面对域偏移时具有的固有鲁棒性。这一案例恰恰反证了，在数据分布更具一般性的真实场景上进行训练，是构建高性能、高泛化能力红外小目标检测模型更可靠的路径。

综上所述，交叉数据集泛化实验从跨场景与跨尺度两个维度证明，MFLNet 凭借其鲁棒的特征提取与高效的多尺度融合设计，能够学习到泛化性强的特征表示。这标志着 MFLNet 不仅是一个高效的检测器，更是一个具备强适应能力的特征学习框架，为算法在实际复杂多变红外场景中的部署应用提供了关键依据。

4.3 消融实验

为验证本文方法的有效性，本节对 MFLNet 中的 CMABlock 模块、DMSFHead 模块以及 CIoU-NWD 混合损失函数进行了消融实验。其中对 CMABlock 模块和 DMSFHead 模块的消融实验选取 IRSTD-1k、NUDT-SIRST 和 NUAA-SIRST 三个数据集，所有实验均使用相同的划分好的数据集和参数设置；对 CIoU-NWD 混合损失函数的消融实验采用 IRSTD-1k 数据集，因为它由真实图像组成且数据量充足，最能验证其性能。

为验证 CMABlock 模块的有效性，设计了不含其线性注意力分支的消融实验，移除该注意力部分并将其替换为常规卷积及归一化，其余网络结构保持不变。实验结果如表 3 所示，相较于移除注意力分支的结构，本文方法的精度与召回率更高，究其原因，是改进后的 CMABlock 能更精准地聚焦于红外图像中的微弱目标区域，显著增强模型对红外小目标的感知能力，提高检测精度。

如表 4 所示，将本文架构中的 DMSFHead 替换为标准检测头，其余部分不变，虽模型参数量有所下降，但精度与召回率均显著降低，这是因为 DMSFHead 对于特征图的利用率远高于标准检测头，其实现了对不同层次

表 2 交叉数据集泛化实验结果

Table 2 Results of cross-dataset generalization experiments

Training dataset	Testing dataset	Precision	Recall	F1-Score	AP50
IRSTD-1k	NUDT-SIRST	0.911	0.702	0.793	0.826
IRSTD-1k	NUAA-SIRST	0.871	0.858	0.864	0.889
NUDT-SIRST	IRSTD-1k	0.636	0.614	0.625	0.605
NUDT-SIRST	NUAA-SIRST	0.837	0.817	0.827	0.836
NUAA-SIRST	IRSTD-1k	0.751	0.650	0.697	0.695
NUAA-SIRST	NUDT-SIRST	0.860	0.697	0.770	0.783

表 3 本文方法与去除注意力后的 CMABlock 模块的消融实验定量结果比较

Table 3 Quantitative result comparison of the ablation experiment between the proposed method and the CMABlock module without attention

Dataset	With attention				Without attention			
	Precision	Recall	F1-Score	AP50	Precision	Recall	F1-Score	AP50
IRSTD-1k	**0.902**	0.794	**0.845**	**0.849**	0.856	**0.806**	0.830	0.842
NUDT-SIRST	**0.947**	**0.944**	**0.945**	**0.967**	0.927	0.941	0.934	0.965
NUAA-SIRST	**0.941**	**0.843**	**0.889**	**0.941**	0.895	0.830	0.861	0.912

表 4 本文方法与替换本文头部后的模型的消融实验定量结果比较

Table 4 Quantitative result comparison of the ablation experiment between the proposed method and the model with the proposed head replaced

Dataset	DMSFHead				DetectHead			
	Precision	Recall	F1-Score	AP50	Precision	Recall	F1-Score	AP50
IRSTD-1k	**0.902**	**0.794**	**0.845**	**0.849**	0.835	0.711	0.768	0.806
NUDT-SIRST	0.947	**0.944**	**0.945**	**0.967**	**0.966**	0.891	0.927	0.953
NUAA-SIRST	**0.941**	**0.843**	**0.889**	**0.941**	0.823	0.832	0.827	0.873

特征的有效整合，并平衡了不同尺度特征的贡献，使模型能够牺牲较低的参数量换来精度以及召回率的大幅提升，且推理速度的损失也微弱到可以忽略不计。

尽管在 NUDT-SIRST 上替换头部的模型取得了更高的精度，但其召回率大幅下降，导致 F1 分数显著低于本文方法。这证明 DMSFHead 通过多尺度特征的动态融合，在精度和召回率之间取得了更好的平衡。

如前所述，CIoU-NWD 损失函数中参数 λ 的取值将影响训练精度，本文通过梯度法动态调整 λ 的取值开展对比实验，如表 5 所示，NWD 通过重构正负样本的特征判别边界，显著强化了两类样本的特征可区分性，推动模型性能提升，实验发现，在 λ 值为 0.1 时，模型训练的精度与召回率可达到最佳。

表 5 不同 λ 值对实验定量结果的影响

Table 5 Effects of different λ values on experimental quantitative results

λ	Precision	Recall	F1-Score
0.0	0.874	**0.799**	0.835
0.1	**0.902**	0.794	**0.845**
0.2	0.891	0.781	0.832
0.3	0.893	0.749	0.815
0.4	0.825	0.759	0.791
0.5	0.887	0.78	0.830
0.6	0.878	0.775	0.823
0.7	0.866	0.797	0.830

4.4 计算效率分析

为定量评估 MFLNet 的计算效率，本节在统一的测试环境下，通过模型参数量 (Params)、浮点运算数 (GFLOPs) 与推理速度 (FPS) 三项指标，将其与当前红外小目标检测领域的主流 SOTA 方法进行对比。所有模型的计算效率均在 NVIDIA GeForce RTX 3080 GPU、CUDA 11.7、PyTorch 1.13 环境下测试，采用批量大小为 1 的 512×512 随机张量输入，以排除数据 I/O 的影响，纯粹衡量模型核心前向推理的计算吞吐能力。结果如表 6 所示，本文方法在计算效率与推理速度的权衡上取得了显著优势。

表 6 不同方法的计算效率对比

Table 6 Computational efficiency comparison of different methods

Method	Params/M	GFLOPs	FPS
MDvsFA	**3.14**	164.45	15.75
ACM	4.62	83.40	26.79
ALCNet	5.92	157.90	12.65
ISNet	8.78	122.52	17.78
AGPCNet	12.36	172.73	8.22
DNANet	4.70	89.26	14.58
Ours	4.12	**13.80**	**47.40**

MFLNet 的浮点运算数在所有对比方法中最低，同时实现了最高的推理速度。该优势首先源于采用检测范式从根本上规避了分割任务中逐像素计算的冗余开销，在此范式的基础上，CMABlock 的线性注意力机制与 DMSFHead 的轻量化特征融合结构进一步压缩了计算量，从而使 MFLNet 能以轻量化参数量实现较高效的计算吞吐。

横向对比显示，MFLNet 在计算量上仅为 ACM 的约六分之一，同时推理速度也大幅领先；与参数量相近的 ALCNet 相比，MFLNet 以其约十一分之一的计算量，实现了约 3.7 倍的推理加速。此外，相较于 AGPCNet 等高计算负荷模型，MFLNet 的计算量仅为其约十二分之一，同时带来近六倍的帧率提升。上述数据充分印证

了本文采用的检测范式及其核心模块，在根本上显著降低了传统分割方法在红外小目标检测中存在的巨大计算开销。

综合来说，MFLNet 不仅通过多尺度特征学习优化了检测精度，更在计算效率上实现了显著提升，为红外小目标检测技术在边缘计算场景的部署提供了可行性。

4.5 可视化结果

图 7 展示了不同方法在 IRSTD-1k、NUDT-SIRST 数据集上的部分可视化结果。其中正确检测目标、未检

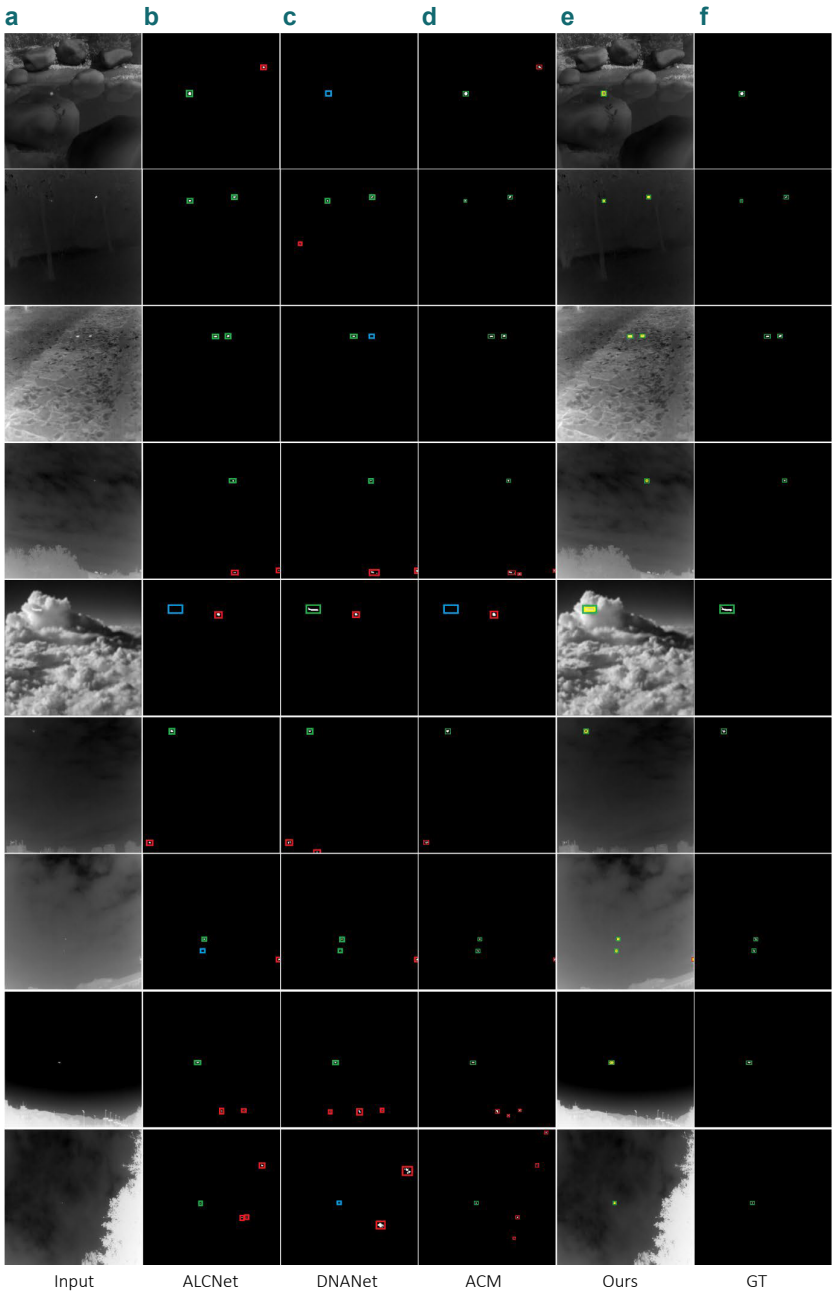


图 7 不同方法在 IRSTD-1k、NUDT-SIRST 数据集上获得的部分可视化结果。(a) 输入图像；(b) ALCNet；(c) DNANet；(d) ACM；(e) MFLNet (本文方法)；(f) 真值标注。目标用绿色、蓝色和红色的方框表示，分别表示正确检测到的目标、未检测到的目标和错误检测到的目标

Fig. 7 Partial visualization results of different methods on the IRSTD-1k and NUDT-SIRST datasets. (a) Input image; (b) ALCNet; (c) DNANet; (d) ACM; (e) MFLNet (Ours); (f) Ground truth. Targets are represented by green, blue, and red boxes, which respectively indicate correctly detected targets, undetected targets, and incorrectly detected targets

测到的目标及错误检测的目标分别以绿色、蓝色和红色方框标注。性能优越模型表现为绿色方框更多而红、蓝色方框更少。

主流深度学习的方法虽凭借自适应特征学习具备一定鲁棒性，能有效检测多数目标，但在局部高光干扰下易产生假警报，面对暗淡目标时存在漏检问题。本文方法通过有效学习红外小目标特性，在局部高光干扰与弱小目标场景中仍能实现精准检测与定位。

5 结　论

针对红外小目标检测中因目标尺度微小、信杂比低导致的特征提取不充分及特征利用不足，以及传统 IoU 对小目标位置偏移敏感引发的正样本稀疏问题，本文提出多层级特征融合学习网络 MFLNet。该网络通过 CMABlock 强化模型对微弱目标的局部特征感知；借助 DMSFHead 的动态加权融合机制平衡了多尺度特征贡献，提升了小目标多尺度特征利用率；采用归一化高斯 Wasserstein 距离替代传统 IoU，并设计 CIoU-NWD 混合损失函数缓解了模型在训练过程中对红外小目标定位偏差的高敏感性。综合理论分析与实验验证，MFLNet 被证明能够有效应对前述挑战。

本文研究为进一步提升红外小目标检测性能指明了若干有价值的探索方向。未来工作可致力于优化序列建模中的维度转换机制，通过减少冗余计算，在保持精度的同时进一步提升推理效率；另一方面，可探索构建多帧时空融合框架，结合帧间运动估计与目标时空连续性建模，以增强在信杂比极低场景下对超微小红外目标的判别能力，从而提升复杂环境中的整体检测性能。

作者贡献

张策：研究资料的收集、分析，论文模型性能分析，论文写作；武丹凤：指导论文选题、论文修改；孙涛：指导论文研究；匡敏驰：论文修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Chen Z Y, Liu H T, Wu X H, et al. Progress in the research of directed thermal radiation[J]. *Opto-Electron Eng*, 2024, **51**(9): 240128.
   陈志营, 刘皓佗, 吴小虎, 等. 定向热辐射研究进展[J]. 光电工程, 2024, **51**(9): 240128.
2. Deng H, Sun X P, Liu M L, et al. Small infrared target detection based on weighted local difference measure[J]. *IEEE Trans Geosci Remote Sens*, 2016, **54**(7): 4204−4214.
3. Teutsch M, Krüger W. Classification of small boats in infrared images for maritime surveillance[C]//*2010 International Waterside Security Conference*, 2010: 1–7. https://doi.org/10.1109/WSSC.2010.5730289.
4. Han J H, Moradi S, Faramarzi I, et al. Infrared small target detection based on the weighted strengthened local contrast measure[J]. *IEEE Geosci Remote Sens Lett*, 2021, **18**(9): 1670−1674.
5. Sun Y, Yang J G, An W. Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model[J]. *IEEE Trans Geosci Remote Sens*, 2021, **59**(5): 3737−3752.
6. Deshpande S D, Er M H, Venkateswarlu R, et al. Max-mean and max-median filters for detection of small targets[C]//*Proceedings of SPIE 3809, Signal and Data Processing of Small Targets 1999*, 1999: 74–83. https://doi.org/10.1117/12.364049.
7. Rivest J F, Fortin R. Detection of dim targets in digital infrared imagery by morphological image processing[J]. *Opt Eng*, 1996, **35**(7): 1886−1893.
8. Hadhoud M M, Thomas D W. The two-dimensional adaptive LMS (TDLMS) algorithm[J]. *IEEE Trans Circuits Syst*, 1988, **35**(5): 485−494.
9. Dragomiretskiy K, Zosso D. Two-dimensional variational mode decomposition[C]//*10th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2015: 197–208. https://doi.org/10.1007/978-3-319-14612-6_15.
10. Zhang M C, Niu C H, Liu L S, et al. Infrared small target detection algorithm for UAV detection system[J]. *Laser Technol*, 2024, **48**(1): 114−120.
    张明淳, 牛春晖, 刘力双, 等. 用于无人机探测系统的红外小目标检测算法[J]. 激光技术, 2024, **48**(1): 114−120.
11. Han J H, Moradi S, Faramarzi I, et al. A local contrast method for infrared small-target detection utilizing a tri-layer window[J]. *IEEE Geosci Remote Sens Lett*, 2020, **17**(10): 1822−1826.
12. Han J H, Ma Y, Zhou B, et al. A robust infrared small target detection algorithm based on human visual system[J]. *IEEE Geosci Remote Sens Lett*, 2014, **11**(12): 2168−2172.
13. Han J H, Liang K, Zhou B, et al. Infrared small target detection utilizing the multiscale relative local contrast measure[J]. *IEEE Geosci Remote Sens Lett*, 2018, **15**(4): 612−616.
14. Gao C Q, Meng D Y, Yang Y, et al. Infrared patch-image model for small target detection in a single image[J]. *IEEE Trans Image Process*, 2013, **22**(12): 4996−5009.
15. Zhang L D, Peng L B, Zhang T F, et al. Infrared small target detection via non-convex rank approximation minimization joint $l_{2,1}$ norm[J]. *Remote Sens*, 2018, **10**(11): 1821.
16. Dai Y M, Wu Y Q. Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection[J]. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 2017, **10**(8): 3752−3767.
17. Zhang L D, Peng Z M. Infrared small target detection based on partial sum of the tensor nuclear norm[J]. *Remote Sens*, 2019, **11**(4): 382.
18. Jian Y, Huang Z L, Wang X. Infrared small target detection based on randomized tensor algorithm[J]. *Laser Technol*, 2024, **48**(1): 127−134.
    蹇渊, 黄自力, 王询. 基于随机化张量算法的红外弱小目标检测[J]. 激光技术, 2024, **48**(1): 127−134.

19. Wang H, Zhou L P, Wang L. Miss detection vs. false alarm: adversarial learning for small object segmentation in infrared images[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 8508–8517. https://doi.org/10.1109/ICCV.2019.00860.
20. Dai Y M, Wu Y Q, Zhou F, et al. Asymmetric contextual modulation for infrared small target detection[C]//*2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021: 949–958. https://doi.org/10.1109/WACV48630.2021.00099.
21. Dai Y M, Wu Y Q, Zhou F, et al. Attentional local contrast networks for infrared small target detection[J]. *IEEE Trans Geosci Remote Sens*, 2021, **59**(11): 9813−9824.
22. Zhang T F, Li L, Cao S Y, et al. Attention-guided pyramid context networks for detecting infrared small target under complex background[J]. *IEEE Trans Aerosp Electron Syst*, 2023, **59**(4): 4250−4261.
23. Zhang M J, Zhang R, Yang Y X, et al. ISNet: shape matters for infrared small target detection[C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022: 867–876. https://doi.org/10.1109/CVPR52688.2022.00095.
24. Li B Y, Xiao C, Wang L G, et al. Dense nested attention network for infrared small target detection[J]. *IEEE Trans Image Process*, 2023, **32**: 1745−1758.
25. Hou Q Y, Wang Z P, Tan F J, et al. RISTDnet: robust infrared small target detection network[J]. *IEEE Geosci Remote Sens Lett*, 2022, **19**: 7000805.
26. Chen G, Wang W H, Tan S R. IRSTFormer: a hierarchical vision transformer for infrared small target detection[J]. *Remote Sens*, 2022, **14**(14): 3258.
27. Zhang M J, Wang Y C, Guo J, et al. IRSAM: advancing segment anything model for infrared small target detection[C]//*18th European Conference on Computer Vision*, 2025: 233–249. https://doi.org/10.1007/978-3-031-72855-6_14.
28. Li R H, Shen Y. YOLOSR-IST: a deep learning method for small target detection in infrared remote sensing images based on super-resolution and YOLO[J]. *Signal Process*, 2023, **208**: 108962.
29. Dai Y M, Li X, Zhou F, et al. One-stage cascade refinement networks for infrared small target detection[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5000917.
30. Yao S B, Zhu Q Y, Zhang T, et al. Infrared image small-target detection based on improved FCOS and spatio-temporal features[J]. *Electronics*, 2022, **11**(6): 933.
31. Du J M, Lu H Z, Zhang L P, et al. A spatial-temporal feature-based detection framework for infrared dim small target[J]. *IEEE Trans Geosci Remote Sens*, 2022, **60**: 3000412.
32. Ju M R, Luo J N, Liu G Q, et al. ISTDet: an efficient end-to-end neural network for infrared small target detection[J]. *Infrared Phys Technol*, 2021, **114**: 103659.
33. Yang B, Zhang X Y, Zhang J, et al. EFLNet: enhancing feature learning network for infrared small target detection[J]. *IEEE Trans Geosci Remote Sens*, 2024, **62**: 5906511.
34. Ying X Y, Liu L, Wang Y Q, et al. Mapping degeneration meets label evolution: learning infrared small target detection with single point supervision[C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 15528–15538. https://doi.org/10.1109/CVPR52729.2023.01490.
35. Nian B K, Jiang B, Shi H J, et al. Local contrast attention guide network for detecting infrared small targets[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5607513.
36. Li R J, An W, Xiao C, et al. Direction-coded temporal U-shape module for multiframe infrared small target detection[J]. *IEEE Trans Neural Netw Learn Syst*, 2025, **36**(1): 555−568.
37. Sun H, Bai J X, Yang F, et al. Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5000513.
38. Wan M J, Ye X B, Zhang X J, et al. Infrared small target tracking via gaussian curvature-based compressive convolution feature extraction[J]. *IEEE Geosci Remote Sens Lett*, 2022, **19**: 7000905.
39. Wang Z, Zang T, Fu Z L, et al. RLPGB-Net: reinforcement learning of feature fusion and global context boundary attention for infrared dim small target detection[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5003615.
40. Han D C, Wang Z Y, Xia Z F, et al. Demystify mamba in vision: a linear attention perspective[C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2024: 4039.
41. Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020: 1571–1580. https://doi.org/10.1109/CVPRW50498.2020.00203.
42. Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces[EB/OL]. arXiv: 2312.00752, 2023. https://doi.org/10.48550/arXiv.2312.00752.
43. Liu S T, Huang D, Wang Y H. Learning spatial fusion for single-shot object detection[EB/OL]. arXiv: 1911.09516, 2019. https://doi.org/10.48550/arXiv.1911.09516.
44. Wang J W, Xu C, Yang W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[EB/OL]. arXiv: 2110.13389, 2021. https://doi.org/10.48550/arXiv.2110.13389.

作者简介



张策 (2002-)，男，硕士研究生，主要研究方向为图像处理、红外小目标检测。

E-mail：fwktzxc@163.com



【通信作者】孙涛 (1981-)，男，硕士，工程师，主要研究方向为地面无人装备论证。

E-mail：43176757@qq.com



【通信作者】武丹凤 (1982-)，女，博士，副教授，硕士生导师，主要研究方向为智能无人系统感知与规划决策。

E-mail：jqrdanfeng@buu.edu.cn

Multi-scale feature learning network for infrared small object detection

Zhang Ce[1], Wu Danfeng[2,3*], Sun Tao[4*], Kuang Minchi[5]

[1] School of Software, Liaoning Technical University, Huludao, Liaoning 125105, China;
[2] Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China;
[3] College of Robotics, Beijing Union University, Beijing 100027, China;
[4] Unit 63963, Chinese People's Liberation Army, Beijing 100072, China;
[5] Department of Precision Instrument, Tsinghua University, Beijing 100084, China

Abstract

Objective Infrared small target detection (IRSTD) serves as a foundational enabling technology in computer vision, with extensive and critical deployment across military and civilian domains including infrared early warning, precision guidance, maritime search and rescue, and intelligent traffic monitoring. The task presents inherent and unresolved challenges stemming from the unique physical properties of infrared small targets, which typically manifest as spot-like structures spanning only several to dozens of pixels, yielding extremely sparse texture and structural information. Compounding these difficulties, infrared imagery is frequently degraded by intense background clutter, low contrast, and low signal-to-clutter ratio (SCR), all of which obscure the faint signals of dim targets. Existing mainstream methods exhibit notable limitations in addressing these interconnected challenges: segmentation-based deep learning frameworks, while proficient in fine-grained local detail capture, incur prohibitive computational overhead and limited real-time performance; detection-based architectures, despite superior inference efficiency, demonstrate insufficient sensitivity to dim targets, suffer from target feature suppression under clutter interference, underutilize multi-scale feature information, and remain highly sensitive to minor localization deviations during training due to inherent constraints of standard intersection over union (IoU)-based loss functions. To mitigate these critical bottlenecks, this paper proposes a novel multi-scale feature learning network, hereafter referred to as MFLNet, for high-precision and efficient IRSTD.

Methods To address the aforementioned challenges, MFLNet is built on a synergistic architecture with three core technical innovations. First, to enhance weak target feature perception in low-SCR scenarios, we develop a CSP with mamba-inspired linear attention block, termed CMABlock, for feature extraction. This module integrates cross stage partial architecture with a tailored linear attention mechanism, and incorporates rotary position encodings within a mamba-inspired linear attention submodule to realize position-aware hierarchical modeling of potential target regions. This design directs the model to focus on faint, location-critical target cues in intense background noise, improving feature discriminability while preserving a lightweight computational profile. Second, to resolve insufficient multi-scale feature utilization, we design a dynamic multi-scale feature fusion head, termed DMSFHead, as the detection decoder. Unlike conventional detection heads with fixed fusion weights, DMSFHead adopts a dynamic weighted fusion strategy, extending standard three-layer feature fusion to a four-layer feature pyramid optimized for IRSTD to retain critical high-resolution details before they are degraded in deep network layers. Its core feature aggregation and selective feature fusion modules adaptively balance contributions from multi-scale features, preserving spatial details from high-resolution layers while integrating semantic information from low-resolution layers, thus enhancing feature utilization and target-clutter discrimination. Third, to alleviate the extreme sensitivity of standard IoU metrics to minor positional shifts in tiny targets, we introduce normalized Gaussian Wasserstein distance (NWD), which models bounding boxes as two-dimensional Gaussian distributions to generate a smoother, scale-robust similarity metric. To balance this with precise geometric regression constraints, we construct a hybrid CIoU-NWD loss function, which combines the training stability of NWD for tiny targets with the accurate bounding box regression capability of complete IoU (CIoU), mitigating issues of sparse positive samples and unstable model training.

Results and Discussions Extensive validation experiments are conducted on three widely used challenging public IRSTD datasets: IRSTD-1k, NUDT-SIRST, and NUAA-SIRST, with MFLNet benchmarked against six state-of-the-art (SOTA) methods including MDvsFA, ACM, ALCNet, ISNet, AGPCNet, and DNANet. Quantitative results show that MFLNet consistently outperforms all competing methods across all benchmarks. Specifically, MFLNet achieves an F1-score of 0.845 on IRSTD-1k, outperforming the second-best method by 10.1% relative; an F1-score of 0.945 on NUDT-SIRST, corresponding to a 4.4% relative improvement; and an F1-score of 0.889 on NUAA-SIRST, exceeding the runner-up by 4.8% relative. Ablation studies systematically validate the effectiveness of each proposed component: replacing the attention branch of CMABlock with standard convolutions incurs consistent performance degradation across all

Foundation item: Open Foundation of the State Key Laboratory of Precision Space-time Information Sensing Technology (STL2023-B-06-01(K)), and Academic Research Projects of Beijing Union University (ZK20202201)

*Correspondence: Wu D F, E-mail: jqrdanfeng@buu.edu.cn; Sun T, E-mail: 43176757@qq.com

metrics, while substituting DMSFHead with a conventional detection head leads to significant drops in both precision and recall, verifying the critical contribution of dynamic multi-scale fusion. Experiments on the hybrid loss function confirm that a balanced weighting factor of λ=0.1 achieves optimal performance by reconciling training stability and localization accuracy. Cross-dataset generalization experiments further demonstrate MFLNet's robustness to domain shifts, with stable performance across all six transfer tasks, indicating its ability to learn generalizable target representations rather than dataset-specific biases. Computational efficiency analysis shows that MFLNet maintains an exceptionally lightweight profile, with 4.12 M parameters, 13.80 GFLOPs of computational complexity, and 47.40 f/s inference speed, outperforming all compared methods in efficiency. This performance stems from its detection-based paradigm and lightweight module design, making it well-suited for deployment on resource-constrained edge platforms.

Conclusions This paper proposes MFLNet, a novel high-performance multi-scale feature learning network for IRSTD. Through the synergistic design of three core components—CMABlock for enhanced weak target feature perception, DMSFHead for efficient multi-scale feature utilization, and the CIoU-NWD hybrid loss for stable model training—MFLNet systematically addresses the core technical challenges of IRSTD. Extensive experimental results confirm that MFLNet not only establishes new SOTA detection accuracy across multiple mainstream benchmarks but also exhibits superior computational efficiency. This work provides a robust and practical framework for real-world high-performance infrared detection systems, with important application value for military surveillance, autonomous navigation, and related fields. Future work will focus on optimizing sequence modeling mechanisms and exploring multi-frame spatiotemporal fusion to further improve detection performance in extremely low-SCR scenarios.

Keywords: infrared small target; object detection; linear attention; multi-scale feature; two-dimensional gaussian distribution

Zhang C, Wu D F, Sun T, et al. Multi-scale feature learning network for infrared small object detection[J]. *Opto-Electron Eng*, 2026, **53**(6): 250358; DOI: 10.12086/oee.2026.250358

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文