

特征恢复的抗模糊检测：一种先验引导的红外无人机目标检测端到端框架

王啸林¹, 方厚章^{1*}, 李青山¹, 王璐¹, 昌毅², 颜露新²

¹西安电子科技大学 计算机科学与技术学院, 中国

²华中科技大学 人工智能与自动化学院, 中国

wxl@stu.xidian.edu.cn, {houzhangfang,wanglu}@xidian.edu.cn, qshli@mail.xidian.edu.cn, {yichang,yanluxin}@hust.edu.cn

摘要

红外无人机 (UAV) 目标图像通常遭受由于传感器快速移动引起的运动模糊退化，显著降低了目标与背景之间的对比度。通常，检测性能严重依赖于目标与背景之间的判别性特征表示。现有方法通常将去模糊视为专注于视觉质量的预处理步骤，而忽略了增强对检测至关重要的任务相关特征。在模糊条件下提升用于检测的特征表示仍然具有挑战性。在本文中，我们提出了一种新颖的联合特征域去模糊和检测的端到端框架 (Joint Feature-Domain Deblurring and Detection end-to-end framework)，称为 JFD³。我们设计了一个具有共享权重的双分支架构，其中清晰分支引导模糊分支以增强判别性特征表示。具体而言，我们首先引入了一个轻量级特征恢复网络，其中清晰分支的特征充当特征级监督以引导模糊分支，从而增强其用于检测的判别能力。然后，我们提出了一个频率结构引导模块，该模块细化来自恢复网络的结构先验，并将其集成到检测骨干网络浅层中，以丰富目标结构信息。最后，在双分支检测骨干之间施加特征一致性自监督损失，驱动模糊分支逼近清晰分支的特征表示。我们还构建了一个名为 IRBlurUAV 的基准测试，包含 30,000 张合成的和 4,118 张真实的具有不同运动模糊的红外无人机目标图像。在 IRBlurUAV 上的大量实验表明，JFD³ 在保持实时效率的同时实现了卓越的检测性能。

代码和数据集 — <https://github.com/IVPILabX/JFD3>

1 引言

红外无人机 (UAV) 目标 (IRUT) 检测由于其全天候可操作性和对不同光照条件的鲁棒性，在许多应用中发挥着至关重要的作用，例如无人机监视和侦察任务^[1]。然而，由于为了及时跟踪快速移动无人机的平台突变运动或突然的机械振动，运动模糊经常出现在红外图像中（见图 1 左侧）。这种运动模糊在长期无人机监视中频繁且通常不可避免，给精确的目标检测带来了巨大挑战。此外，IRUT 通常表现出微弱的特征，并嵌入在复杂的杂波背景中。运动模糊进一步降低了目标与周围环境之间的对比度，使得判别性特征提取更加困难。近年来，在图像去模糊^[2-6] 和 IRUT 检测^[7,8] 方面都取得了显著进展。然而，大多数现有方法将去模糊和检测视为两个独立的任务，并以不同的目标独立解决。据我们所知，目前还没有专门解决运动模糊条件下 IRUT 检测的工作。

为了解决上述问题，一种直接的方法是直接在模糊图像上应用检测器^[9]。然而，模糊退化显著降低

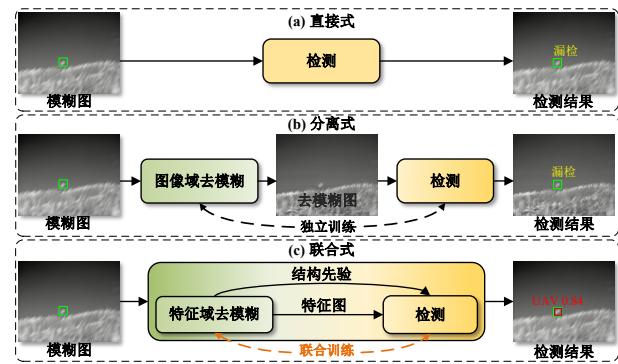


图 1: 运动模糊下无人机目标检测的三种策略。(a) 直接法：检测器直接处理模糊图像。(b) 分离法：图像域去模糊作为检测前的预处理步骤。(c) 联合法：特征域去模糊和检测在端到端框架中同时解决。我们的 JFD³ 联合处理这两项任务，并利用来自去模糊网络的结构先验来增强检测网络的特征表示。

了无人机目标与背景之间的对比度，导致频繁的漏检（见图 1(a)）。另一种方法是将去模糊^[10] 作为检测前的预处理步骤^[9]（见图 1(b)）。然而，该流程存在几个局限性。首先，基于深度学习的去模糊方法计算复杂，增加大量延迟时间，限制了其在无人机监视这一时敏任务中的适用性。其次，这些方法通常针对视觉增强进行优化，而不是针对特定任务的特征恢复，这可能会引入难以察觉的噪声，从而对检测性能产生负面影响^[11,12]。最近，一些研究^[11-13] 探索了低级和高级视觉任务的联合优化。然而，这些努力大多集中在雾、霾等恶劣天气条件下，并且主要是在可见光中针对一般物体类别进行的。相比之下，IRUT 检测中运动模糊的独特挑战在很大程度上仍未得到充分探索。

为了弥合红外底层去模糊和高层 IRUT 检测之间的差距，我们提出了联合特征域去模糊和检测网络 (Joint Feature-Domain Deblurring and Detection end-to-end framework, JFD³)。它在训练期间采用双分支架构，其中清晰图像分支监督模糊图像分支，以联合优化特征恢复和检测。在推理时，仅保留模糊分支以实现高效推理。具体而言，为了解决图像域去模糊方法延迟高且缺乏面向检测意识的局限性，我们设计了一个由清晰特征域去模糊分支引导并与检测网络联合训练的轻量级特征恢复网络。该网络有效地增强了对检测至关重要的退化表示。此外，为了提高运动模糊下的结构感知能力，我们设计了一个频率结构引导模块。它首先使用自适应高通滤波模块提取高频细节特征，然后通过细节保

*通讯作者。

持注意力机制细化结构信息。细化后的结构先验随后被注入到检测骨干网络的 stem 和 stage 1 之间，补偿模糊图像中缺失的结构线索并提高 IRUT 的可判别性。最后，为了增强骨干网络从退化输入中提取有意义的目标特征的能力，我们在模糊分支和清晰分支之间引入了特征一致性自监督损失。这种约束鼓励模糊分支在特征空间中逼近清晰分支，从而在模糊下实现更准确的目标判别。

为了评估 JFD³ 的有效性，我们构建了一个新的基准测试，名为 IRBlurUAV。它包含 30,000 对合成模糊-清晰的 IRUT 图像 (IRBlurUAV-syn) 和 4,118 张真实世界的模糊图像 (IRBlurUAV-real)，涵盖了不同的运动方向、模糊强度、多尺度无人机目标和复杂的背景。大量实验表明，我们的 JFD³ 在运动模糊下检测 IRUT 方面显著优于最先进的方法。

我们的主要贡献总结如下：

- 我们提出了一个联合框架 JFD³，它以端到端的方式统一了特征域恢复和 IRUT 检测。优化恢复组件以增强有利于检测的特征，这由检测目标引导而非常见的视觉质量评估。这种任务驱动的设计增强了运动模糊条件下的检测性能。据我们所知，这是第一项在统一框架中解决运动模糊条件下 IRUT 检测的工作。
- 我们首次引入了一种专门为红外模糊图像设计的特征域恢复策略，重点关注目标检测任务的需求。它专注于恢复与检测相关的特征。该方法增强了退化图像中目标特征的表示，并提高了检测性能。
- 我们设计了一种新颖的频率结构引导模块，将去模糊网络中的目标频率结构先验集成到检测骨干中。这种集成通过补充高频结构细节增强了模糊条件下 IRUT 的结构表示。该模块显著提高了局部判别能力和目标定位能力。

2 相关工作

2.1 图像去模糊方法

图像去模糊长期以来一直是一项核心的低级视觉任务，对于提高退化视觉输入的质量至关重要。早期的基于卷积神经网络的方法^[10]直接将模糊图像映射到清晰图像，构成了许多编码器-解码器架构的基础。基于 Transformer 的模型^[14]通过长程依赖建模提高了性能，但也引入了更高的复杂性。基于扩散的方法^[15]和最近的 Mamba 风格^[16,17]架构已经展示了令人印象深刻的感知质量和通用性，但计算量仍然很大。然而，大多数现有的去模糊方法很少与检测等高级任务集成。相比之下，我们的方法在特征域中恢复与检测相关的特征，以更好地支持下游任务。

2.2 红外无人机目标检测方法

近年来，已经提出了几种 IRUT 检测方法^[18-20]，其中大多数侧重于清晰图像的检测。大多数公开可用的 IRUT 数据集^[21-23]也由清晰的高质量图像组成。据我们所知，UniCD^[24]是唯一一项明确考虑 IRUT 检测中低频非均匀性噪声退化的工作。然而，IRUT 检测中的另一种常见退化，即运动模糊，在该领域仍然是一个很大程度上未探索的空白。在这项工作中，我们旨在通过提出第一个联合特征域去模糊和 IRUT 检测的端到端框架来填补这一空白。

2.3 联合去模糊和检测方法

最近的研究^[11,25]已经开始探索底层视觉复原与高级视觉任务的集成，但大多数都集中在雾霾和可见光

场景。Sayed 等人^[26]使用五类补救措施增强了运动模糊检测。Aakanksha 等人^[27]引入了以类别为中心的运动模糊增强方法用于分割。DREB-Net^[13]提出了一种用于运动模糊下可见光汽车目标检测的双流融合架构。然而，这些方法忽略了运动模糊下红外小目标检测的独特挑战，例如纹理稀缺和精细结构退化。为了解决这个问题，我们提出了一个频率引导模块，该模块增强了小目标的结构线索并提高了模糊退化下的鲁棒性。

3 本文提出的方法

3.1 双分支联合学习框架

为了解决红外无人机图像中运动模糊导致的判别性特征退化挑战，我们提出了一种双分支联合学习框架，使特征域去模糊和检测能够在端到端的方式中协同优化。核心设计理念是在训练期间利用清晰图像监督来指导鲁棒表示的学习，从而增强模型在模糊退化下的检测性能。

如图 2 所示，该框架由两个并行分支组成：清晰图像分支和模糊图像分支，它们共享权重以确保特征空间对齐。在训练期间，两个分支都被激活。清晰分支在清晰输入上运行并提供高质量的特征引导，而模糊分支暴露于运动退化的输入，并学习通过监督恢复和对齐其表示。在推理期间，仅保留模糊图像分支。

每个分支都始于一个轻量级特征恢复网络，旨在补偿低级退化。恢复后的特征图随后通过检测骨干网络，该骨干网络由 stem 和多个 stage 组成。我们采用 DEIM^[28]作为我们的基础检测架构。在 stem 和骨干网络 stage 1 之间，我们集成了一个频率结构引导模块，该模块注入细化的结构先验以丰富目标定位线索。这种设计使得特征去模糊和检测任务的联合优化成为可能。

为了增强网络的特征提取能力，我们在模糊分支和清晰分支的检测骨干之间引入了特征一致性自监督损失。该损失鼓励模糊分支产生与清晰分支对齐的中间表示，从而提高其在运动模糊下的鲁棒性。

与其他直接从退化输入检测或将去模糊视为孤立预处理步骤的工作相比，我们的联合框架通过协同监督和结构引导的信息流实现了任务感知的特征增强，最终在模糊条件下产生了卓越的检测性能。

3.2 特征域去模糊网络

传统的图像域去模糊方法通常会产生巨大的计算成本，并引入与检测无关的冗余视觉细节。相比之下，我们采用特征域去模糊 (Feature-Domain Deblurring, FDD) 策略，直接在潜在空间中恢复语义上有意义的表示，从而实现高效且与任务对齐的增强。

如图 2 所示，特征恢复网络在模糊图像上运行，并通过紧凑的编码器-解码器结构对其进行细化。我们将该模块建立在 MIMO-UNet^[29]之上，这是一个通用的去模糊网络，并通过将基础通道数减少到 2，且每个阶段的残差块数量减少到 2 来使其适应我们的场景。这种设计侧重于调节特征分布，促进与清晰分支的语义一致性，并减少表示空间中的域偏移，同时保持低计算代价以应用于实际场景。

为了指导特征域恢复过程，我们设计了一个包含两部分的损失函数，用于监督恢复网络的编码器和解码器阶段。具体而言，我们在编码器中使用 L_1 损失强制模糊分支和清晰分支之间的特征级对齐，并在解码器中使用基于结构相似性的损失强调结构一致性。

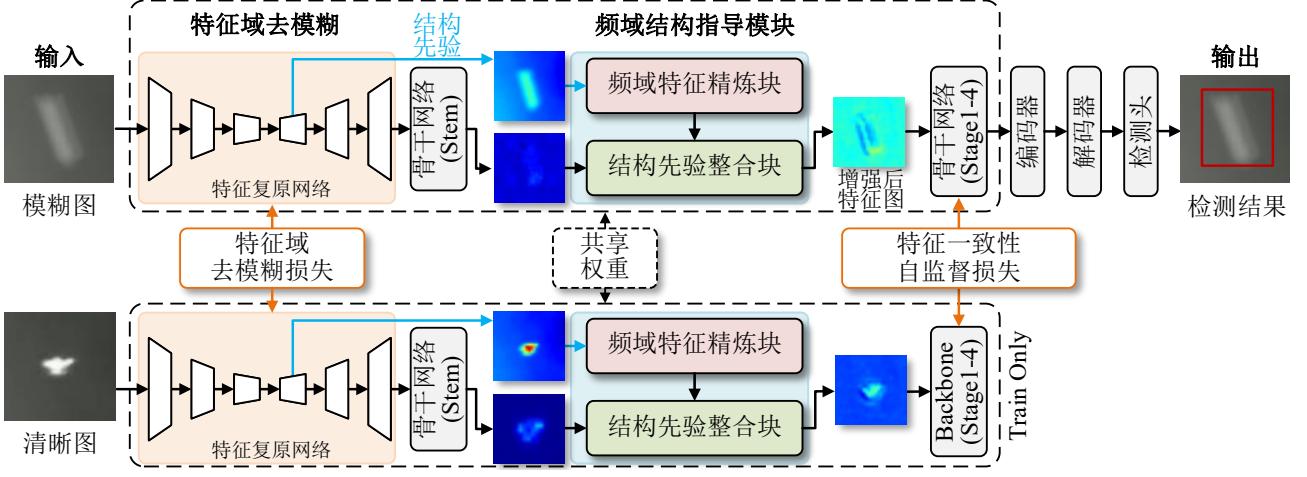


图 2: 所提出的 JFD³ 概述, 它首先通过特征域恢复增强退化特征, 然后使用频率结构引导模块细化结构信息。清晰图像分支使用特征恢复损失和特征一致性自监督损失来监督模糊图像分支。

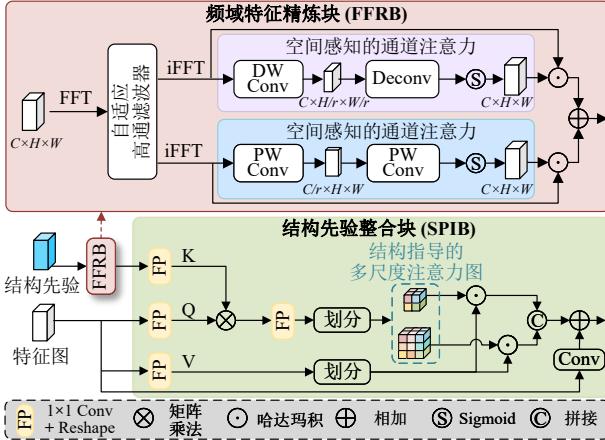


图 3: FSGM 概述。FFRB 通过高通滤波和注意力机制处理先验以细化特征表示。SPIB 将细化后的结构先验集成到特征图中。

设 E_b^i 和 E_c^i 分别表示模糊分支和清晰分支的第 i 个编码器阶段的中间特征图。同样, 设 D_b^j 和 D_c^j 表示第 j 个解码器阶段的输出。为了在解码过程中保留结构先验, 我们计算相应解码器特征之间的结构相似性指数测量 (structural similarity index measure, SSIM), 这鼓励解码器保留结构模式而不是像素相似性, 从而促进下游检测的更鲁棒的特征表示。最终的总特征域去模糊损失 \mathcal{L}_{deb} 可以表示为以下公式:

$$\mathcal{L}_{deb} = \sum_{i=1}^3 \|E_b^i - E_c^i\|_1 + \sum_{j=1}^3 (1 - SSIM(D_b^j, D_c^j)). \quad (1)$$

3.3 频率结构引导模块

在模糊的红外无人机图像中, 小目标区域通常遭受边界结构退化和细节被抑制的问题。为了解决这个问题, 我们在框架中集成了一个频率结构引导模块 (frequency structure guidance module, FSGM), 它细化高频结构先验并将其注入检测骨干网络。具体而言, 我们将特征域去模糊网络中第一个解码器的

输出特征图作为结构先验 P 。该特征图包含相对丰富的结构和语义信息, 并且与其原始图像相比下采样了 4 倍, 这与检测骨干的 stem 和 stage 1 输出的特征图的分辨率相匹配。如图 3 所示, FSGM 由两个子组件组成: 频率特征细化块 (frequency feature refine block, FFRB) 和结构先验集成块 (structure prior integration block, SPIB)。它们共同提取、增强并将频域结构线索集成到特征表示中, 以改善目标判别。

频率特征细化块。 FFRB 旨在通过从结构先验 P 中提取和细化高频结构分量 P_{high} 来增强判别性细节。我们首先通过快速傅里叶变换 (fast Fourier transform, FFT) 将 P 变换到频域, 然后应用可学习的高通滤波器 $\mathcal{H}_{high}(\cdot)$ 来抑制低频分量。该滤波器以 0.5 的频率阈值初始化, 随后在训练期间自适应调整。该滤波器保留了对于区分小目标至关重要的关键高频细节。最后, 使用快速傅里叶逆变换 (inverse fast Fourier transform, iFFT) 将滤波结果变换回空间域。该过程定义为:

$$P_{high} = \text{iFFT}(\mathcal{H}_{high}(\text{FFT}(P))), \quad (2)$$

接下来, 我们采用两种类型的注意力机制来细化特征图: 空间感知通道注意力 (spatial-aware channel attention, SCA) 和通道感知空间注意力 (channel-aware spatial attention, CSA)。这些机制旨在以防止在空间或通道维度上过度压缩的方式增强特征图, 这对于小目标检测尤其重要 [30]。与将空间或通道维度压缩为单个标量的传统全局注意力不同 (传统方法可能会消除关键的细粒度线索), 我们采用部分压缩来保留与目标相关的细节, 这对小物体尤为重要的。细化后的先验 $P_{refined}$ 计算如下:

$$F_{SCA} = P_{high} \odot \sigma(\text{DeConv}(\text{DWConv}(P_{high}))), \quad (3)$$

$$F_{CSA} = P_{high} \odot \sigma(\text{PWConv}(\text{PWConv}(P_{high}))), \quad (4)$$

$$P_{refined} = F_{SCA} + F_{CSA}, \quad (5)$$

其中 DWConv、DeConv 和 PWConv 分别指深度卷积、反卷积和逐点卷积操作; $\sigma(\cdot)$ 是 sigmoid 激活函数; \odot 表示逐元素乘法。

结构先验集成块。 SPIB 接收结构先验 $P_{refined}$ 和中间特征图 f 作为输入。两个流首先通过特征投影 (表示为 FP), 由 1×1 卷积后接 reshape 组成。

这产生注意力分量: $Q = \phi_q(f)$, $K = \phi_k(P_{\text{refined}})$, 以及 $V = \phi_v(f)$, 其中 ϕ 表示投影算子。

然后计算交叉注意力矩阵 $A = Q^\top K$ 。它产生查询和结构引导键之间的成对对应关系。为了注入多尺度空间线索, 注意力图 A 被分成两个分支, 表示为 A_1 和 A_2 , 分别作为对应于 5×5 和 7×7 动态核的结构引导多尺度注意力图。并行地, 投影的值特征 V 被分成两个匹配的分支 V_1 和 V_2 , 确保与注意力图对齐。然后, 每个注意力图通过逐元素 Hadamard 积调制其对应的值特征, 融合结果通过通道级联进行聚合。最后, 添加来自原始特征图 f 的具有 3×3 卷积的残差连接以增强梯度流和表示保真度:

$$F_{\text{PG}} = \text{Concat}(A_1 \odot V_1, A_2 \odot V_2) + \text{Conv}(f). \quad (6)$$

结果特征 F_{PG} 随后被转发到后续的检测网络。这种设计能够将结构感知引导分层集成到模糊特征表示中。

3.4 联合去模糊和检测损失

为了提高运动模糊下的检测性能, 我们采用多损失优化策略。具体而言, 我们的框架使用三个损失分量的组合进行训练: 检测损失 \mathcal{L}_{det} 、特征去模糊损失 \mathcal{L}_{deb} 和特征一致性自监督 (feature consistency self-supervised, FCSS) 损失 $\mathcal{L}_{\text{FCSS}}$ 。这些分量协同监督双分支网络, 确保从清晰图像到退化图像的有效知识转移, 并促进特征去模糊和检测的联合优化。

检测损失 \mathcal{L}_{det} 基于模糊分支的预测计算, 与原始 DEIM [28] 保持一致。它作为特定于任务的目标来指导端到端优化。

为了进一步弥合分支之间的表示差距, 我们引入了 FCSS 损失, 它约束模糊分支的中间特征与它们的清晰图像对应物对齐。对于共享检测骨干中的每个阶段 i , 我们分别从清晰和模糊分支提取中间特征图 $F_C^{(i)}$ 和 $F_B^{(i)}$ 。所有阶段的一致性损失平均为:

$$\mathcal{L}_{\text{FCSS}} = \frac{1}{4} \sum_{i=1}^4 \left(1 - \frac{\mathbf{F}_C^{(i)} \cdot \mathbf{F}_B^{(i)}}{\|\mathbf{F}_C^{(i)}\| \|\mathbf{F}_B^{(i)}\|} \right). \quad (7)$$

这种自监督约束促进了分支之间的结构对齐和语义一致性。正如我们的网络设计所支持的那样, 这鼓励模糊分支更好地逼近清晰域表示, 并促进模糊下的更准确检测。

总体训练目标定义为: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}} + \lambda_1 \mathcal{L}_{\text{deb}} + \lambda_2 \mathcal{L}_{\text{FCSS}}$ 。根据经验, 我们将初始权重设置为 $\lambda_1 = 0.4$, $\lambda_2 = 0.2$, 并且 λ_2 在 20 个 epoch 后降低至 0.01, 允许网络在随后的收敛阶段专注于检测精度。

4 实验

4.1 数据集和评估指标

数据集。 我们构建了一个新的基准数据集 IRBlurUAV, 以促进对运动模糊条件下 IRUT 检测的评估。它包含 30,000 对合成模糊和清晰的 IRUT 图像 (IRBlurUAV-syn) 和 4,118 张真实世界模糊 IRUT 图像 (IRBlurUAV-real)。所有图像大小均为 640×512 。合成图像是使用类似于 Sayed 等人 [26] 的方法生成的, 但采用了改进的运动轨迹建模, 该建模采用具有随机方向和长度的线性路径, 以更好地逼近现实的无人机运动模式。该数据集涵盖了不同的背景、多种无人机尺度和各种无人机类型。所有图像都标注了用于检测任务的边界框。IRBlurUAV-syn 集合按 8:1:1 的比例分为训练、验证和测试子

集, 而 IRBlurUAV-real 仅用作测试集, 以评估真实场景中的泛化性能。更多细节在补充材料中提供。

指标。 我们使用标准 COCO 指标评估模型的检测性能: AP , AR , AP_{50} , 和 AR_{50} 。 AP 和 AP_{50} 分别表示交并比 (Intersection over Union, IoU) 范围 0.50:0.95 和 IoU=0.50 时的检测精度。 AR_{50} 测量 maxDets=100 时 IoU=0.50 的平均召回率, AR 表示 maxDets=1 时 IoU 范围 0.50 到 0.95 的平均召回率。为了评估模型复杂性, 我们考虑参数量 (Params)、浮点运算数 (FLOPs) 和每秒帧数 (frames per second, FPS) 以评估实时性能。此外, 我们采用信杂比 (Signal to-Clutter Ratio, SCR) 来评估特征域中目标信号的增强。

4.2 实验细节

实验在配备 CUDA 12.1 和 PyTorch 2.7 的 NVIDIA RTX 3090 GPU 上进行。模型使用 AdamW 优化器训练了 150 个 epoch, 所有其他设置与 DEIM 一致。由于缺乏广泛采用的红外去模糊方法, 我们利用了几种通用的方法: DeepRFT (基于 CNN) [10], MDT (基于 Transformer) [14], MaIR [17] 和 EVSSM (基于 Mamba) [16]。对于目标检测, 我们应用基于 CNN 的方法, 包括 YOLOv1-N, YOLOv1-L [31], MSHNet [32], 和 PCConv (YOLOv8-N 版本) [33], 以及基于 Transformer 的方法, 如 RT-DETR (ResNet18 版本) [9], D-FINE (N 版本) [34], 和 DEIM (D-FINE-N 版本) [28]。MSHNet 和 PCConv 专为红外小目标检测而设计。最后, DREB-Net [13] 作为联合去模糊和检测方法用于比较。为了公平起见, 我们在 IRBlurUAV-syn 上重新训练了所有方法, 并在 IRBlurUAV-syn 和 IRBlurUAV-real 上进行了评估, 以评估性能和泛化能力。

4.3 定量结果

如表 1 所示, 当直接检测模糊图像时, 特别是对于像 MSHNet 和 PCConv 这样的红外目标检测方法, 由于未考虑运动模糊导致的目标与背景之间判别性特征的丢失, 导致准确率和召回率较低。当使用分离检测方法时, 虽然去模糊模块恢复了图像的视觉质量, 但去模糊过程并未充分关注有利于检测的特征。这导致某些方法的恢复性能较差, 例如 MDT, 甚至导致检测性能显著下降。此外, 分离方法中的去模糊模块通常涉及高计算复杂度, 限制了流程的整体实时性能, 从而降低了效率。

与 DREB-Net 等非实时的联合方法相比, 我们的方法在 RTX 3090 上以 25.7 FPS 的速度实现了最佳检测性能, 并具有高部署效率, 仅需 120W 功耗和 606 MB GPU 内存。这得益于我们的轻量级模块 (FDD 和 FSGM), 仅引入了 0.02M 参数。

在表 2 中, 我们进一步在 IRBlurUAV-real 上验证了在表 1 中表现良好的方法, 以测试它们在真实世界模糊条件下的性能。结果表明, 我们的方法在面对真实模糊图像时仍然实现了最佳检测性能, 进一步证明了我们框架的鲁棒性和实用价值。

4.4 定性结果

如图 4 所示, 我们展示了几种有效的方法, 包括直接检测方法 (RT-DETR, D-FINE)、检测前去模糊的方法 (DeepRFT+RT-DETR, DeepRFT+D-FINE), 以及联合方法 (DREB-Net) 和我们的方法。该图的第一行和第二行分别显示了这些方法在 IRBlurUAV-syn 和 IRBlurUAV-real 上的结果。具体而言, 第三

策略	模块			指标						
	去模糊	检测	发表'年份	$AP_{50} \uparrow$	$AR_{50} \uparrow$	$AP \uparrow$	$AR \uparrow$	Params (M)↓	FLOPs (G)↓	FPS↑
直接式	/	YOLO11-N	2024	0.510	0.530	0.213	0.258	<u>10.2</u>	2.7	69.2
		YOLO11-L	2024	0.551	0.565	0.232	0.282	86.6	25.3	39.3
		RT-DETR	CVPR'24	0.716	<u>0.811</u>	<u>0.369</u>	<u>0.400</u>	19.0	30.2	50.2
		D-FINE	ICLR'25	<u>0.722</u>	0.795	0.347	0.382	3.5	3.5	45.8
		DEIM	CVPR'25	0.654	0.734	0.290	0.330	3.5	3.5	45.8
		MSHNet	CVPR'24	0.358	0.428	0.099	0.148	15.5	38.2	53.1
分离式	DeepRFT	RT-DETR	AAAI'23	0.673	0.749	0.284	0.329	36.1	110.1	9.6
		D-FINE	-	0.660	0.743	0.255	0.303	20.6	83.4	9.7
	MDT	RT-DETR	CVPR'25	0.195	0.297	0.062	0.082	31.6	591.8	1.5
		D-FINE	-	0.181	0.268	0.052	0.069	16.1	565.1	1.5
	EVSSM	RT-DETR	CVPR'25	0.636	0.713	0.244	0.285	35.3	822.6	0.6
		D-FINE	-	0.589	0.662	0.194	0.234	19.8	795.9	0.6
	MaIR	RT-DETR	CVPR'25	0.342	0.471	0.118	0.150	39.7	582.4	0.1
		D-FINE	-	0.292	0.403	0.084	0.114	24.2	555.7	0.1
联合式	DREB-Net	TGRS'25	-	0.710	0.754	0.300	0.357	34.6	684.2	10.9
	我们的 JFD ³	-	-	0.767	0.850	0.428	0.458	3.5	4.7	25.7

表 1: 各种方法在 IRBlurUAV-syn 数据集上的性能比较。加粗和下划线分别表示最好和次好的结果。

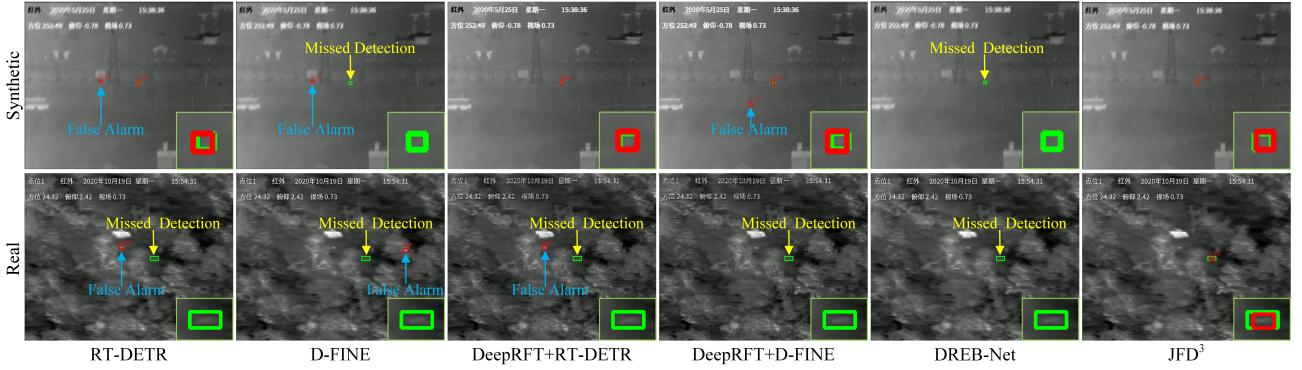


图 4: 在 IRBlurUAV-syn 和 IRBlurUAV-real 上的检测结果比较，包括直接式、分离式和联合式的方法。绿色和红色框分别代表真值和检测到的目标。右下角显示了局部放大图。

方法	$AP_{50} \uparrow$	$AR_{50} \uparrow$	$AP \uparrow$	$AR \uparrow$
RT-DETR	0.480	0.633	0.135	0.170
D-FINE	0.514	<u>0.693</u>	<u>0.151</u>	0.190
DeepRFT + RT-DETR	0.419	0.600	<u>0.124</u>	0.170
DeepRFT + D-FINE	0.437	0.638	0.129	0.181
DREB-Net	<u>0.520</u>	0.619	0.143	<u>0.196</u>
我们的 JFD ³	0.623	0.730	0.251	0.291

表 2: 各种方法在 IRBlurUAV-real 数据集上的性能比较。

列和第四列显示了使用 DeepRFT 去模糊后的图像，而其他列显示了模糊图像。

从结果可以看出，由于运动模糊，目标特征显著退化，使得检测更加困难。直接检测方法通常会导致虚警或漏检，因为它们难以区分模糊图像中的判别性特征。当在检测前应用去模糊时，现有的去模糊算法在严重模糊下仍然难以恢复退化的特征，进一步影响检测性能。相比之下，我们的方法可以在运动模糊条件下准确检测无人机目标。

FDD	FSGM	$AP_{50} \uparrow$	$AR_{50} \uparrow$	$AP \uparrow$	$AR \uparrow$	SCR↑
×	×	0.654	<u>0.734</u>	0.290	0.330	0.463
✓	✗	<u>0.763</u>	0.852	0.390	0.426	0.473
✓	✓	0.765	0.852	<u>0.420</u>	<u>0.451</u>	0.477

表 3: FDD 和 FSGM 的消融研究。

4.5 消融实验

本节介绍了消融研究以验证 JFD³ 的创新点。所有实验均在 IRBlurUAV-syn 上进行。更多实验在补充材料中提供。

FDD 和 FSGM 的影响。 如表 3 所示，引入 FDD 模块显着改善了所有评估指标，表明其在减轻运动模糊方面的强大有效性。这表明特征域去模糊在增强模糊条件下的检测中起着至关重要的作用。当进一步添加 FSGM 时，性能甚至提高得更多。这表明 FSGM 增强了目标结构的可判别性，补充了 FDD 并进一步提高了检测精度。此外，骨干网络不同阶段的 SCR 值反映了目标特征 SCR 的改善，表明特征提取得到了增强。可以看出，我们提出的两个模块都有效地在特征域中增强了目标特征。

图像域去模糊 (IDD) 和特征域去模糊 (FDD) 的影响。 表 4 探讨了图像域去模糊和特征域去模糊之间

IDD	FDD	$AP_{50} \uparrow$	$AR_{50} \uparrow$	$AP \uparrow$	$AR \uparrow$
\times	\times	0.007	0.091	0.001	0.011
\checkmark	\times	0.660	0.743	0.255	0.303
\times	\checkmark	0.763	0.852	0.390	0.426
\checkmark	\checkmark	0.703	0.809	0.307	0.356

表 4: IDD 和 FDD 的消融研究。

模糊程度 $\in (\text{PSNR} \text{ 范围})$	方法	$AP_{50} \uparrow$	$AR_{50} \uparrow$
严重 $\in [10, 20)$	DeepRFT + RT-DETR	0.542	0.638
	JFD ³	0.562	0.723
中等 $\in [20, 22.5)$	DeepRFT + RT-DETR	0.664	0.745
	JFD ³	0.672	0.788
轻微 $\in [22.5, 32)$	DeepRFT + RT-DETR	0.772	0.831
	JFD ³	0.767	0.853

表 5: 不同模糊程度下的性能比较。

层级	w/o FDD		w/ FDD	
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
Stem	14.66	0.5977	15.64	0.6201
Stage1	15.29	0.5189	18.18	0.6464

表 6: FDD 在骨干网络浅层的影响。

的关系，并在 IRBlur-syn 上进行了测试。第一行代表我们的基线架构，在没有 FDD 和 FSGM 的情况下在清晰图像上训练，并直接在模糊图像上测试以模拟真实世界场景。性能几乎为零，凸显了运动模糊对仅考虑清晰图像的检测的重大影响。第二行显示了在检测前对模糊图像应用 DeepRFT 的结果，这导致分离去模糊方法的性能有所提高。第三行展示了我们最终的 JFD³，其中特征域去模糊达到了最佳性能。最后一行展示了使用去模糊图像作为 JFD³ 的输入，显示出比第二行全面的改进，表明特征域和图像域去模糊相互补充。这种组合增强了仅靠图像域去模糊无法有效恢复的目标特征。

不同模糊程度的影响。 在表 5 中，我们将测试集按模糊程度划分为三个等级。针对每个等级，我们研究了分离方法和我们提出的 JFD³ 的性能。随着模糊严重程度的增加，所有方法的检测性能都会下降，这突显了由更严重模糊引起的干扰的不断增大。然而，单独分析每个模糊程度时，JFD³ 在模糊严重程度增加的情况下展现出更大的性能提升。对于更严重的模糊，JFD³ 的表现优于分离方法，这表明我们的方法在处理严重模糊条件方面相比于分离式方法更有效。

FDD 在骨干网络浅层的影响。 如表 6 所示，我们通过计算从清晰和模糊输入中提取的浅层特征图之间的峰值信噪比 (PSNR) 和结构相似性指数测量 (SSIM) 来评估 FDD 的效果 (有和没有 FDD)。结果表明，FDD 显著改善了这两个指标，表明早期阶段特征的判别质量和整体表示更好。

更多关于 IRBlurUAV 的可视化结果和其他数据集的实验结果可以在补充材料中找到。

5 结论

在本文中，我们提出了 JFD³，一种用于运动模糊下 IRUT 检测的端到端双分支框架。首先，我们引入了一个专注于特征域去模糊的轻量级特征恢复网络。

接下来，我们提出了一个频率结构引导模块，以增强有利于检测的目标结构信息。此外，我们构建了一个名为 IRBlurUAV 的数据集，其中包含各种运动模糊红外无人机图像。实验表明，JFD³ 在模拟和真实场景中均优于现有方法，同时保持实时性能。

致谢

本工作得到了多谱信息智能处理技术全国重点实验室开放研究基金（批准号 61421132301）、江苏省自然科学基金（BK20232028）以及中国国家自然科学基金项目（批准号 62472341, 62372351, U21B2015, 62371203 和 62301228）的部分资助。

参考文献

- [1] Houzhang Fang, Zikai Liao, Xuhua Wang, Yi Chang, and Luxin Yan. Differentiated attention guided network over hierarchical and aggregated features for intelligent uav surveillance. *IEEE Transactions on Industrial Informatics*, 19(9):9909–9920, 2023.
- [2] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3883–3891, July 2017.
- [3] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8174–8182, June 2018.
- [4] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8183–8192, June 2018.
- [5] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8878–8887, October 2019.
- [6] Songnan Lin, Jiawei Zhang, Jinshan Pan, Yicun Liu, Yongtian Wang, Jing Chen, and Jimmy Ren. Learning to deblur face images via sketch synthesis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11523–11530, Apr. 2020.
- [7] Houzhang Fang, Lan Ding, Liming Wang, Yi Chang, Luxin Yan, and Jinhui Han. Infrared small uav target detection based on depthwise separable residual dense network and multiscale feature fusion. *IEEE Transactions on Instrumentation and Measurement*, 71:1–20, 2022.
- [8] Houzhang Fang, Lan Ding, Xiaolin Wang, Yi Chang, Luxin Yan, Li Liu, and Jinrui Fang. Scinet: Spatial and contrast interactive super-resolution assisted infrared uav target detection.

- IEEE Transactions on Geoscience and Remote Sensing, 62:1–22, 2024.
- [9] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. Detrs beat yolos on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 16965–16974, June 2024.
- [10] Xintian Mao, Yiming Liu, Fengze Liu, Qingli Li, Wei Shen, and Yan Wang. Intriguing findings of frequency selection for image deblurring. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, pages 1905–1913, Jun. 2023.
- [11] Chengyang Li, Heng Zhou, Yang Liu, Caidong Yang, Yongqiang Xie, Zhongbo Li, and Liping Zhu. Detection-friendly dehazing: Object detection in real-world hazy scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(7):8284–8295, 2023.
- [12] Tianyang Xu, Yifan Pan, Zhenhua Feng, Xuefeng Zhu, Chunyang Cheng, Xiao-Jun Wu, and Josef Kittler. Learning feature restoration transformer for robust dehazing visual object tracking. International Journal of Computer Vision, 132(12):6021–6038, 2024.
- [13] Qingpeng Li, Yuxin Zhang, Leyuan Fang, Yuhan Kang, Shutao Li, and Xiao Xiang Zhu. Dreb-net: Dual-stream restoration embedding blur-feature fusion network for high-mobility uav object detection. IEEE Transactions on Geoscience and Remote Sensing, 63:1–18, 2025.
- [14] Duosheng Chen, Shihao Zhou, Jinshan Pan, Jinglei Shi, Lishen Qu, and Jufeng Yang. A polarization-aided transformer for image deblurring via motion vector decomposition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 28061–28070, June 2025.
- [15] Mengwei Ren, Mauricio Delbracio, Hossein Talebi, Guido Gerig, and Peyman Milanfar. Multiscale structure guided diffusion for image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 10721–10733, October 2023.
- [16] Lingshun Kong, Jiangxin Dong, Jinhui Tang, Ming-Hsuan Yang, and Jinshan Pan. Efficient visual state space model for image deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 12710–12719, June 2025.
- [17] Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, and Xi Peng. Mair: A locality- and continuity-preserving mamba for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 7491–7501, June 2025.
- [18] Artem Rozantsev, Vincent Lepetit, and Pascal Fua. Detecting flying objects using a single moving camera. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(5):879–892, 2017.
- [19] Mingjing Zhao, Wei Li, Lu Li, Ao Wang, Jin Hu, and Ran Tao. Infrared small uav target detection via isolation forest. IEEE Transactions on Geoscience and Remote Sensing, 61:1–16, 2023.
- [20] Houzhang Fang, Zikai Liao, Lu Wang, Qingshan Li, Yi Chang, Luxin Yan, and Xuhua Wang. Danet: Multi-scale uav target detection with dynamic feature perception and scale-aware knowledge distillation. In Proceedings of the 31st ACM International Conference on Multimedia (ACMMM), page 2121–2130, 2023.
- [21] Bo Huang, Jianan Li, Junjie Chen, Gang Wang, Jian Zhao, and Tingfa Xu. Anti-uav410: A thermal infrared benchmark and customized scheme for tracking drones in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(5):2852–2865, 2024.
- [22] Nan Jiang, Kuiran Wang, Xiaoke Peng, Xuehui Yu, Qiang Wang, Junliang Xing, Guorong Li, Guodong Guo, Qixiang Ye, Jianbin Jiao, Jian Zhao, and Zhenjun Han. Anti-uav: A large-scale benchmark for vision-based uav tracking. IEEE Transactions on Multimedia, 25:486–500, 2023.
- [23] Jie Zhao, Jingshu Zhang, Dongdong Li, and Dong Wang. Vision-based anti-uav detection and tracking. IEEE Transactions on Intelligent Transportation Systems, 23(12):25323–25334, 2022.
- [24] Houzhang Fang, Xiaolin Wang, Zengyang Li, Lu Wang, Qingshan Li, Yi Chang, and Luxin Yan. Detection-friendly nonuniformity correction: A union framework for infrared uav target detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 11898–11907, June 2025.
- [25] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-adaptive yolo for object detection in adverse weather conditions. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 1792–1800, Jun. 2022.
- [26] Mohamed Sayed and Gabriel Brostow. Improved handling of motion blur in online object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1706–1716, June 2021.
- [27] Aakanksha and A. N. Rajagopalan. Improving robustness of semantic segmentation to motion-blur using class-centric augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 10470–10479, June 2023.
- [28] Shihua Huang, Zhichao Lu, Xiaodong Cun, Yongjun Yu, Xiao Zhou, and Xi Shen. Deim: Detr with improved matching for fast convergence. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 15162–15171, June 2025.

- [29] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 4641–4650, October 2021.
- [30] Yimian Dai, Yiquan Wu, Fei Zhou, and Kobus Barnard. Asymmetric contextual modulation for infrared small target detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pages 950–959, January 2021.
- [31] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024.
- [32] Qiankun Liu, Rui Liu, Bolun Zheng, Hongkui Wang, and Ying Fu. Infrared small target detection with scale and location sensitivity. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 17490–17499, June 2024.
- [33] Jiangnan Yang, Shuangli Liu, Jingjun Wu, Xinyu Su, Nan Hai, and Xueli Huang. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 39, pages 9202–9210, Apr. 2025.
- [34] Yansong Peng, Hebei Li, Peixi Wu, Yueyi Zhang, Xiaoyan Sun, and Feng Wu. D-fine: Re-define regression task of detrs as fine-grained distribution refinement. In Proceedings of the International Conference on Representation Learning (ICLR), volume 2025, pages 44015–44031, 2025.