

DISTA-Net: 动态空间邻近红外小目标解混

韩圣东^{1,*}, 杨尚东^{1,*}, 张鑫², 李宇轩², 李翔^{2,3}, 杨健², 程明明^{2,3}, 戴一冕^{2,3†}

¹ 计算机学院, 南京邮电大学, ²PCA Lab, VCIP, 计算机学院, 南开大学, ³NKIARI, 福田, 深圳

* 相同贡献, † 通讯作者

在红外成像中, 密集成团的空间邻近小目标解混存在严峻的技术挑战, 因为信号混叠会严重影响目标数量统计、亚像素级定位以及辐射强度测定的准确性。尽管深度学习在红外小目标检测领域取得了一定的进展, 但其在空间邻近红外小目标上的应用尚未得到探索。这一空白主要源于混叠特征分离的复杂性以及开源基础设施的缺失。在本研究中, 我们提出了动态迭代收缩阈值网络 (Dynamic Iterative Shrinkage Thresholding Network, DISTA-Net), 该网络将传统稀疏重建方法重构为动态框架。DISTA-Net 能自适应地生成卷积权重和阈值参数, 以实时调整重建过程。据我们所知, DISTA-Net 是首个专门针对空间邻近红外小目标解混设计的深度学习模型, 实现了卓越的亚像素检测精度。此外, 我们建立了该领域首个开源生态系统以促进后续研究。该生态系统包含三大核心组件: (1) CSIST-100K, 一个公开的基准数据集; (2) CSO-mAP, 一个用于亚像素检测的自定义专用评估指标; (3) GrokCSO, 一个开源工具包, 其中包含 DISTA-Net 及其他模型。代码和数据集详见 <https://github.com/GrokCV/GrokCSO>。

日期: 5.20.2025

项目主页: <https://github.com/GrokCV/GrokCSO>

GrokCV

1 引言

得益于红外成像对热辐射的卓越敏感性以及对照明条件的独立性, 红外成像在各种长距离检测和监视任务中发挥着关键作用 [Sobrino et al. \(2016\)](#)。然而, 由于远程目标与成像系统距离较大, 从远程目标捕获的辐射强度往往较弱 [Kou et al. \(2023\)](#)。特别是当目标密集聚集时, 这一挑战将进一步加剧。这是因为空间邻近目标 (Closely-Spaced Objects, CSO) 成像相互交叠 [Pruett et al. \(2023\)](#) 会形成混叠的团簇状像斑。

如图 1, 这种混叠使得人眼无法通过视觉直接分辨出独立目标, 从而影响对目标数量、精确位置及辐射强度的感知 [Zhao et al. \(2022\)](#)。这一现象对红外搜索与跟踪 (Infrared Search and Tracking,IRST) 系统后续的检测、跟踪和识别阶段带来了巨大挑战。因此, 探索解混并重建此类空间邻近红外小目标 (Closely-Spaced Infrared Small Targets, CSIST), 准确辨别其精确位置和辐射强度的高效技术, 具有重要意义。

尽管在各种应用中, 对 CSIST 进行解混具有至关重要的意义, 但针对这一特定任务的研究仍极为匮乏。这些方法通常将问题表述为参数估计任务, 并采用优化算法进行求解 [Meyer and Williams \(2021\)](#)。例如 [Zhang et al. \(2013\)](#) 利用成像平面上的目标稀疏性, 提出了一种基于离散采样的稀疏重建方法, 该方法采用过完备字典, 并在 ℓ_1 范数正则化下求解二阶锥规划问题。

然而, 现有基于优化的模型性能高度依赖于精细的超参数调优 [Chen et al. \(2021\)](#), 制约了实际应用。此外, 目标数量或位置的变化进一步增加了最优超参数选择的难度, 限制了这些方法的泛化能力和实用性 [Wan et al. \(2018\)](#)。因此, 亟需开发对超参数变化具有更强鲁棒性、并能有效应用于多样化现实场景的解混算法。

深度学习方法已彻底改变了图像超分辨率技术 [Li et al. \(2023\)](#), 然而, 其在 CSIST 解混中的应用尚未被探索, 这主要归因于任务目标的根本差异以及生态系统限制。与通过恢复高频细节来提升清晰度的通用超分辨率技术不同 [Wu et al. \(2024\)](#), CSIST 解混需要精确估算重叠目标的数量、位置及辐射强度——本质上是将图像映射到特定目标的属性, 而非传统的像素空间超分辨率。这一挑战由于缺乏标准化基准数据集、专用评估指标和开源实现而变得更加严峻, 极大地阻碍了该领域深度学习方法的研发与性能比较。

为了解决上述挑战, 本文提出了一种新型深度展开网络, 用于解混 CSIST, 称为动态迭代收缩阈值网络 (Dynamic Iterative Shrinkage Thresholding Network, DISTA-Net)。我们将传统稀疏重建方法重构为动态深度学习框架, 该框架能自适应生成卷积权重和阈值参数, 以实时调整重建过程。与先前的方法不同, 近端映射相关参数 (包括非线性变换和收缩阈值) 会根据输入数据动态调整, 而非手动设定或训练后固定。据我们所知, 这是首次针对该任务的基于深度学习的尝试。

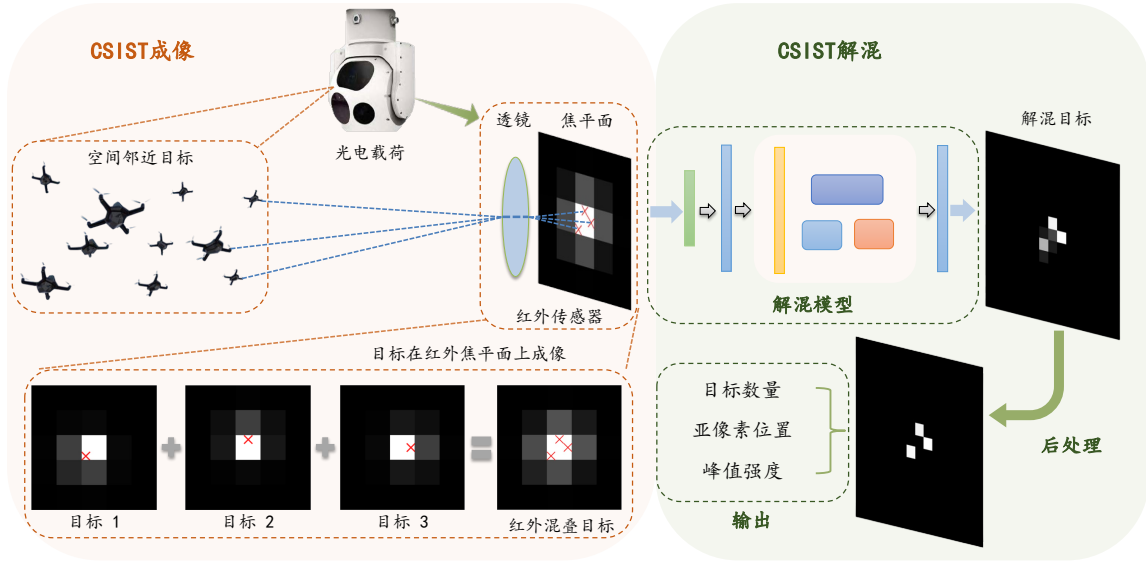


图 1 空间邻近红外小目标成像与解混过程的概念示意图。解混技术旨在分离并精确估算重叠目标的数量、位置及强度。

此外，我们构建了一个完整的开源生态系统以推动该领域研究。该生态系统包括 CSIST-100K，这是一个开放的基准数据集，包含 10 万对低分辨率混叠目标图像及其精确的位置和辐射强度标注。我们还引入了 CSO-mAP 评估指标，该指标受目标检测中的平均精度（mean average precision, mAP）指标启发，专门用于评估解混后红外小目标的数量、空间定位及辐射强度。与这些资源配套的是 GrokCSO，这是一个基于 PyTorch 的开源工具箱，封装了我们的 DISTA-Net 以及其他最先进模型，使研究人员能够便捷使用 CSIST-100K 数据集。

综上所述，我们的贡献可归纳为 4 个主要方面：

1. 我们将 CSIST 的解混问题重新表述为一个可解释的深度展开问题。据我们所知，这是首个基于深度学习的此类任务尝试。
2. 我们提出的 DISTA-Net 是一种动态深度展开网络，能够根据输入数据自适应地生成卷积权重和阈值参数，以优化重建过程。
3. 我们建立了首个针对此任务的开源生态系统，其中包括 CSIST-100K 数据集、CSO-mAP 评估指标以及 GrokCSO 工具箱。
4. 我们对我们的方法进行了全面分析，验证了动态深度展开的重要性，以及 DISTA-Net 在 CSIST 解混任务中的有效性。

2 相关工作

2.1 红外小目标检测

受一系列开源数据集的驱动 Dai et al. (2021a, 2023); Li et al. (2024)，红外小目标检测任务在近年来吸引了大量研究关注。当前研究主要集中于开发多尺度

特征融合模型以应对目标固有特征稀缺的问题 Xu et al. (2023)。Dai et al. (2021b) 提出了一种非对称上下文调制模块，通过逐点通道注意力机制构建自底向上的路径的非对称上下文调制模块，实现高层语义与底层细节的融合。Wang et al. (2023) 将强化学习与金字塔特征融合相结合，并提出了一个全局上下文边界注意力模块，以缓解局部亮噪声。Mei et al. (2023) 提出一种差异感知注意力模块，包含两个关键组件：双时相聚合模块，用于学习全局特征并激活目标敏感通道；差异注意力模块，通过局部相关性建模实现多尺度小目标检测。Tong et al. (2023) 采用编码器-解码器结构，通过空洞空间金字塔池化和双注意力模块增强特征提取，并利用多尺度标签聚焦于目标边缘和内部特征。Zhang et al. (2025) 提出一种基于图像-文本融合的红外小目标检测框架。通过 CLIP 语言提示引导 SAM 分割，并设计去噪适配器使其适配红外域，充分发挥其泛化能力。

本研究聚焦于红外小目标检测，但与先前研究存在两点主要差异。首先，尽管红外小目标检测是本研究的前期工作，但我们的核心关注点在于对 CSIST 进行解混。因为检测潜在重叠目标对于后续任务（如亚像素定位和辐射强度预测）至关重要。其次，我们的任务超越了简单的检测，通过在亚像素级别精确定位目标并估计其辐射强度，提供了比先前二值检测方法更详细的目标特征理解。

2.2 深度展开

深度展开，如 Monga et al. (2021) 中所描述的，可追溯至 2010 年的学习迭代缩放阈值算法（Learned Iterative Shrinkage-Thresholding Algorithm, LISTA）Gregor and LeCun (2010)，该算法将 ISTA Daubechies et al. (2004) 重构为全连接前

馈神经网络，对新样本展现出优异的泛化能力，能以更少的迭代次数达到类似 ISTA 的精度。后续研究如 ADMM-Net [yang et al. \(2016\)](#) 将交替方向乘法器方法 (Alternating Direction Method of Multipliers, ADMM) 的步骤展开为深度学习框架，通过基于端到端判别式训练优化的压缩感知模型，提升了 MRI 重建的精度与效率。同样，ISTA-Net [Zhang and Ghanem \(2018\)](#) 采用端到端学习策略实现近端映射，增强了自然图像压缩感知的重建性能。

受此类进展启发，深度展开技术已在多种计算机视觉任务中找到应用。[Li et al. \(2020\)](#) 将一个广义梯度域总变差算法转化为深度可解释网络，用于盲图像去模糊，通过学习参数实现了优异性能。在图像超分辨率领域，[Guo et al. \(2019\)](#) 将可训练卷积层融入离散余弦变换框架，有效抑制伪影并实现有限数据下的学习能力。[Solomon et al. \(2020\)](#) 将鲁棒主成分分析展开为深度网络，提升了超声成像中微气泡与组织信号的区分度。

与之前采用静态参数的方法 [Zhang and Ghanem \(2018\)](#); [You et al. \(2021\)](#) 不同，本研究的 DISTA-Net 根据输入动态调整近端映射权重，从而实现适应性重建过程，以应对不同场景。

3 CSIST 基准测试、指标与工具包

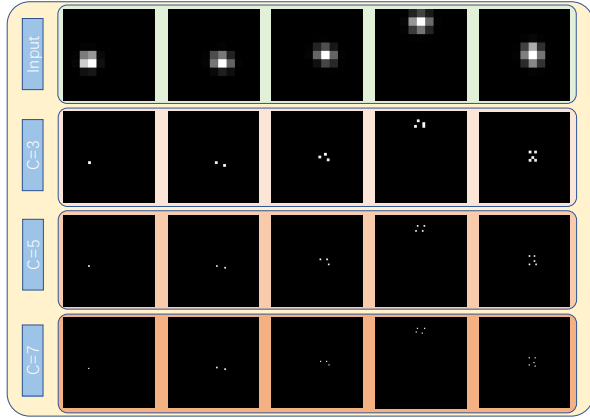


图 2 CSIST-100K 可视化：顶部一行显示 1 至 5 个重叠目标的成像结果，随后各行展示了针对 3×、5× 和 7× 亚像素分割的解混结果。

CSIST-100K 数据集。 当点光源通过透镜成像时，在焦点处会出现一种称为“艾里斑”的衍射图案，其特征为中央呈现明亮圆斑，周围环绕着多个明暗相间的同心环状条纹。这种衍射效应通常用点扩散函数 (Point Spread Function, PSF) 近似表达，其标准差取决于传感器焦比 (f 值) 和检测带宽，用于量化能量扩散程度。在多目标成像中，每个像素捕获的辐射强度是多点源的累积响应。特别是对于遥远物体，每个物体都被视为点光源，其产生的艾里斑半径为 $1.22\lambda/D$ (λ 为探测波长， D 为透镜直径)，也就是

二维高斯 PSF 的 1.9σ ，是传感器对两个点源目标的物理分辨率极限，称为瑞利判据 (Rayleigh Unit, 1R)。

在本研究中，设定 σ_{PSF} 为 0.5 像素。我们的仿真实验涵盖了包含 1 至 5 个重叠目标的图像，每个目标具有二维坐标和辐射强度信息 (220 至 250 单位)。目标在 11×11 像素网格的单个像素内随机分布，同时保持至少 0.52 瑞利单位的间距。在此配置下，生成了一个包含 100,000 个样本的数据集 **CSIST-100K**，其中 80,000 个样本构成训练集，剩余的 20,000 个样本平均分配给验证集和测试集。如图 2 所示，CSIST 会在像平面发生扩散，能量主要集中在 3×3 像素区域内。这导致显著的混叠现象，使得确定目标的数量及其精确空间坐标变得极为困难。

CSO-mAP 评价指标。 传统边界框度量在目标分离度低于瑞利准则时失效 (由于艾里斑干扰)，我们的 CSO-mAP 通过严格的亚像素位置/强度匹配进行 CSIST 评估，重新定义真阳性 (True Positive, TP) 和假阳性 (False Positive, FP) 如下：

$$\mathbb{1}_k(\hat{t}_j, t_i) = \begin{cases} 1, & \text{if } d(\hat{t}_j, t_i) < \delta_k \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

其中， \hat{t}_j 是第 j 个预测目标， t_i 是第 i 个真实目标，而 $\delta_k \in \{0.05, 0.1, 0.15, 0.2, 0.25\}$ 是一系列距离阈值，其中 $k = 1, 2, 3, 4, 5$ ，用于控制所需的定位精度。通过强度优先匹配生成精确率-召回率曲线，并在每个 δ_k 处计算平均精度 (Average Precision, AP)。最终的 CSO-mAP 指标通过对所有阈值下的 AP 值进行平均计算得出，明确量化了在不同空间分辨率需求下的性能表现，详见[补充材料](#)。

GrokCSO 工具箱。 为了解决该领域缺乏专用工具箱的问题，我们推出了 GrokCSO 开源工具包，旨在提升对 CSIST 的重建效果。基于 PyTorch 构建，GrokCSO 提供了预训练模型、可重复性脚本以及专门针对 CSO 挑战设计的专用评估指标。该工具包的详细架构和功能在[补充材料](#)中有所说明。

4 方法

在本节中，我们将介绍 CSIST 成像模型、传统的稀疏重建方法以及 DISTA-Net 架构。

4.1 成像与解混框架

CSIST 高斯成像模型。 鉴于目标和红外探测器之间的显著距离，目标近似点光源。光学系统的衍射效应会将能量扩散到相邻像素，这一过程可由二维高斯点扩散函数描述，详见 [Liaudat et al. \(2021\)](#)：

$$p(x, y) = \frac{1}{2\pi\sigma_{\text{PSF}}^2} \exp \left[-\frac{(x - x_t)^2 + (y - y_t)^2}{2\sigma_{\text{PSF}}^2} \right], \quad (2)$$

其中 σ_{PSF}^2 表示扩散方差, (x_t, y_t) 表示目标的坐标。在 $U \times V$ 像素的红外焦平面上, 像元 (i, j) 对目标 (x_t, y_t) 的响应被定义为 PSF 在其边界内的积分值:

$$g_{i,j}(x_t, y_t) = \int_{x_{i,j}-1/2D}^{x_{i,j}+1/2D} \int_{y_{i,j}-1/2D}^{y_{i,j}+1/2D} p(x, y) dx dy, \quad (3)$$

其中 $(x_{i,j}, y_{i,j})$ 表示像素的中心位置, $i = 0, \dots, U, j = 0, \dots, V - 1, D$ 为像素宽度。测量模型被向量化为:

$$\mathbf{z} = \mathbf{G}(\mathbf{x}, \mathbf{y})\mathbf{s} + \mathbf{n}, \quad (4)$$

其中 $\mathbf{G}(\mathbf{x}, \mathbf{y})$ 表示导向矩阵, \mathbf{s} 表示目标辐射强度向量, 而 \mathbf{n} 表示高斯白噪声。

纯高斯框架。对于 N 个空间邻近目标, 若每个目标用 $t_i = (x_i, y_i, g_i, \sigma_{\text{PSF}})$ 表示, 其中 $i \in 1, \dots, N$, (x_i, y_i) 表示精确坐标, g_i 表示辐射强度。所有目标的坐标和峰值强度集合分别表示为 $\mathcal{P} = \{p_i | i \in \{1, \dots, N\}\}$ 和 $\mathcal{G} = \{g_i | i \in \{1, \dots, N\}\}$ 。CSIST 解混需要预测的中心点和强度的集合分别表示为 $\hat{\mathcal{P}} = \{\hat{p}_j | j \in \{1, \dots, M\}\}$ 和 $\hat{\mathcal{G}} = \{\hat{g}_j | j \in \{1, \dots, M\}\}$, 其中 M 是预测的目标的数量, \hat{g}_j 是预测目标 \hat{p}_j 的峰值强度。假设 \hat{p}_j 是对真实点 p_i 的预测。我们的目标是最小化 \hat{p}_j 与 p_i 之间的距离以及 \hat{g}_j 与 g_i 之间的差距。因此, 预测目标的数量 M 也应与真实目标的数量 N 密切匹配。本质上, 纯高斯框架会同时预测目标的数量、亚像素位置和辐射强度。

基于稀疏重构的 CSIST 解混。在允许一定量化误差的前提下, CSIST 的位置可离散化为有限个亚像素位置的集合 $\Omega = \{(x_l, y_l)\}_{l=1, \dots, L}$, 其中实际位置构成该集合的稀疏子集。

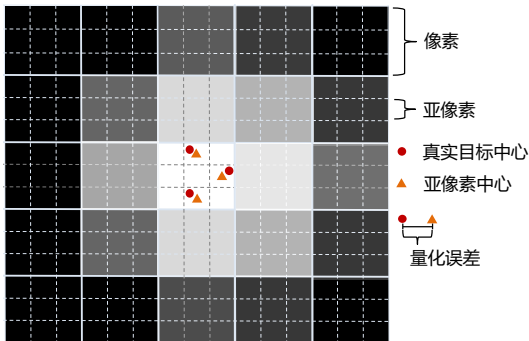


图 3 将每个像素划分为一个 $n \times n$ 的亚像素网格, 代表潜在的目标位置。

如图 3 所示, 每个像素被划分为一个 $n \times n$ 的网格, 从而产生 $L = UVn^2$ 个亚像素。在足够多的网格划分下, 每个亚像素最多包含一个目标, 且最大位置偏差为 $\frac{\sqrt{2}D}{2n}$ 。测量模型可扩展至 Ω :

$$\mathbf{z} = \mathbf{G}(\Omega)\tilde{\mathbf{s}} + \mathbf{w}, \quad (5)$$

其中 $\mathbf{G}(\Omega)$ 是一个以位置集 Ω 中的导向矢量为列的矩阵, \mathbf{w} 表示高斯白噪声, 而 $\tilde{\mathbf{s}} \in \mathbb{R}^L$ 是稀疏的, 且

$L \gg UV$ 。CSIST 解混问题可被表述为带 ℓ_1 正则化的稀疏重建问题:

$$\min_{\tilde{\mathbf{s}}} \|\mathbf{z} - \mathbf{G}(\Omega)\tilde{\mathbf{s}}\|_2^2 + \lambda \|\tilde{\mathbf{s}}\|_1, \quad (6)$$

其中 λ 是正则化参数。解 $\tilde{\mathbf{s}}$ 直接给出目标属性: 其非零元素表示目标数量和强度, 而它们在 Ω 中的对应位置则表示亚像素坐标。

优化方法。ISTA Daubechies et al. (2004) 通过两个交替步骤解决稀疏重建问题:

$$\mathbf{r}^{(k)} = \tilde{\mathbf{s}}^{(k-1)} - \rho \mathbf{G}^\top (\mathbf{G}\tilde{\mathbf{s}}^{(k-1)} - \mathbf{z}), \quad (7)$$

$$\tilde{\mathbf{s}}^{(k)} = \arg \min_{\tilde{\mathbf{s}}} \frac{1}{2} \|\tilde{\mathbf{s}} - \mathbf{r}^{(k)}\|_2^2 + \lambda \|\Psi \tilde{\mathbf{s}}\|_1. \quad (8)$$

其中 Ψ 表示变换矩阵, k 表示迭代索引, ρ 表示步长。第二步表示一个近端映射:

$$\text{prox}_{\lambda\phi}(\mathbf{r}) = \arg \min_{\tilde{\mathbf{s}}} \frac{1}{2} \|\tilde{\mathbf{s}} - \mathbf{r}\|_2^2 + \lambda\phi(\tilde{\mathbf{s}}). \quad (9)$$

虽然 ISTA 使用正交变换 (如小波变换) 时具有高效的解决方案, 但它在处理复杂变换时面临挑战, 并且需要大量迭代。为了克服这些限制, ISTA-Net 用可学习的非线性变换 $\mathcal{F}(\cdot)$ 替换:

$$\tilde{\mathbf{s}}^{(k)} = \arg \min_{\tilde{\mathbf{s}}} \frac{1}{2} \|\mathcal{F}(\tilde{\mathbf{s}}) - \mathcal{F}(\mathbf{r}^{(k)})\|_2^2 + \theta \|\mathcal{F}(\tilde{\mathbf{s}})\|_1, \quad (10)$$

其中 θ 是一个可学习的参数。然而, ISTA-Net 在训练后静态的网络权重限制了其对输入数据的适应性, 这一问题在对输入的敏感性至关重要的 CSIST 解混场景中表现得尤为明显。

4.2 DISTA-Net: 一个动态框架

Algorithm 1 动态迭代收缩阈值网络 (DISTA-Net)

Input: CSIST 图像 \mathbf{z} , 导向矩阵 $\mathbf{G}(\mathbf{x}, \mathbf{y})$, 初始化矩阵 Q_{init} , 阶段数 N , 步长 $\{\rho^{(k)}\}_{k=1}^N$

Output: 重建结果 $\tilde{\mathbf{s}}^{(N)}$

可学习参数:

$\{\rho^{(k)}\}_{k=1}^N, \{\text{DTG}^{(k)}\}_{k=1}^N, \{\mathcal{F}_d^{(k)}\}_{k=1}^N, \{\tilde{\mathcal{F}}^{(k)}\}_{k=1}^N$
(满足 $\tilde{\mathcal{F}}^{(k)} \circ \mathcal{F}_d^{(k)} = \mathbf{I}$)

初始化:

1: $\tilde{\mathbf{s}}^{(0)} \leftarrow Q_{\text{init}}\mathbf{z}$

迭代步骤:

2: **for** $k = 1$ **to** N **do**
3: $\mathbf{r}^{(k)} \leftarrow \tilde{\mathbf{s}}^{(k-1)} - \rho^{(k)} \mathbf{G}^\top (\mathbf{G}\tilde{\mathbf{s}}^{(k-1)} - \mathbf{z})$
4: $\hat{\mathbf{s}}^{(k)} \leftarrow \mathcal{F}_d(\tilde{\mathbf{s}}^{(k-1)}, \mathbf{r}^{(k)})$
5: $\theta_d^{(k)} \leftarrow \text{DTG}^{(k)}(\hat{\mathbf{s}}^{(k)})$
6: $\tilde{\mathbf{s}}^{(k)} \leftarrow \tilde{\mathcal{F}}(\text{Soft}(\mathcal{F}_d(\mathbf{r}^{(k)}), \theta_d^{(k)}))$
7: **end for**

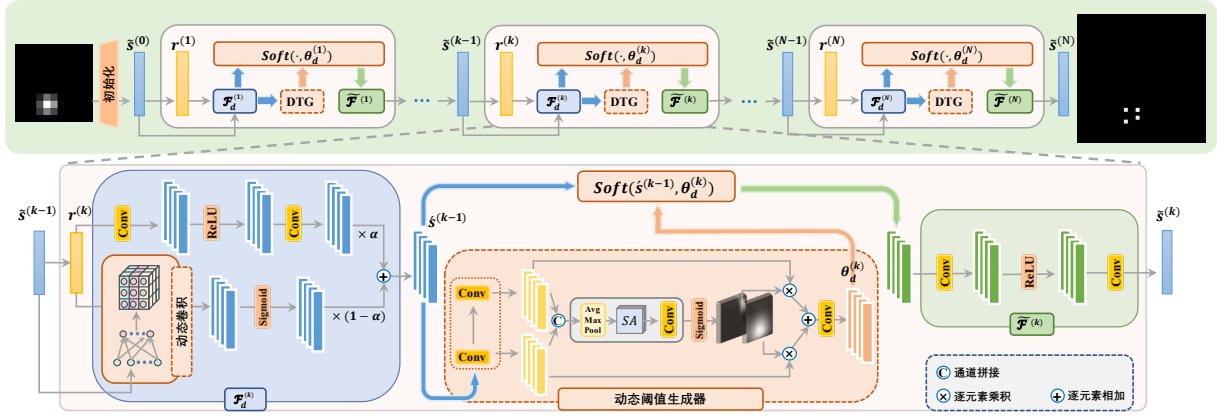


图 4 提出的 DISTA-Net 架构。整体框架由多个级联阶段组成。每个阶段包含三个主要组件：用于特征提取的双分支动态变换模块 ($\mathcal{F}_d^{(k)}$)、用于特征精化的动态阈值模块 ($\Theta_d^{(k)}$) 以及用于重建的逆变换模块 ($\tilde{\mathcal{F}}^{(k)}$)。

为了将 ISTA 与基于网络的稀疏表示相结合, DISTA-Net 将每个 ISTA 迭代映射到网络层。通过构建这些网络层并进行端到端训练, DISTA-Net 能够自动优化参数, 相较于传统 ISTA 中手动设计的参数, 既提升了性能又保持了可解释性。基于 ISTA-Net 的压缩感知框架, 我们引入了两项关键改进。首先, 我们设计了一个数据自适应的非线性变换函数 $\mathcal{F}_d(\cdot)$, 该函数将图像映射到更丰富的维度表示, 同时强调图像中的重要区域。该变换如下:

$$\mathcal{F}_d(\tilde{s}^{(k)}) = \text{Soft}(\mathcal{F}_d(\mathbf{r}^{(k)}), \theta_d), \quad (11)$$

其中 $\text{Soft}(\cdot, \theta_d)$ 表示可学习的软阈值参数 θ_d , k 为阶段索引。左逆 $\tilde{\mathcal{F}}$ 满足 $\tilde{\mathcal{F}}(\cdot) \circ \mathcal{F}_d(\cdot) = \mathbf{I}$, 且与 $\mathcal{F}_d(\cdot)$ 不要求结构对称性。则有:

$$\tilde{s}^{(k)} = \tilde{\mathcal{F}}(\text{Soft}(\mathcal{F}_d(\mathbf{r}^{(k)}), \theta_d)). \quad (12)$$

其次, 我们引入一个动态阈值模块, 以适应输入图像的多样性, 解决图像生成中稀疏向量扰动的敏感性问题。这种灵活的阈值机制可以简单有效地避免固定参数导致的过严格或过宽松问题。如图 4 所示, DISTA-Net 包含 N 个阶段。每个阶段 k 包含三个组件: $\mathcal{F}_d^{(k)}$, $\text{Soft}(\cdot, \theta_d^{(k)})$ 和 $\tilde{\mathcal{F}}^{(k)}$ 。输入 $\mathbf{r}^{(k)}$ 通过式 $\tilde{s}^{(k-1)}$ 从 (7) 计算获得, 依次经过这些组件处理, 以生成下一迭代的 $\tilde{s}^{(k)}$ 。整体而言, 本文提出的 DISTA-Net 方法可参照算法 1 实现。

动态变换。虽然可训练的非线性变换克服了手动设计方法的局限性, 但其训练后固定的参数导致了静态的变换模式。我们通过在第 k 阶段引入双分支动态变换模块 $\mathcal{F}_d^{(k)}$ 解决这一问题 (图 4)。

为处理稀疏图像 $\tilde{s}^{(k-1)}$ 的扰动敏感性, 我们设计了一个辅助分支, 通过动态卷积核引导 $\mathbf{r}^{(k)}$ 。此方法可自适应地提升特征表示能力。

该模块首先通过全连接网络处理 $\tilde{s}^{(k-1)}$ 以生成权重向量:

$$W = f(\tilde{s}^{(k-1)}). \quad (13)$$

动态卷积模块 (图 4) 将该权重向量作为自适应卷积核应用于 $\mathbf{r}^{(k)}$:

$$w_r = C(W, \mathbf{r}^{(k)}). \quad (14)$$

最终输出将 Conv-ReLU-Conv 分支与 sigmoid 激活的辅助分支相结合:

$$\mathcal{F}_d^{(k)} = \alpha \cdot A(\text{ReLU}(B(\mathbf{r}^{(k)}))) + (1 - \alpha) \cdot \text{sigmoid}(w_r), \quad (15)$$

其中 $A(\cdot)$ 和 $B(\cdot)$ 表示卷积运算, 而 $\alpha \in [0, 1]$ 控制两条分支之间的贡献。

动态软阈值。与 ISTA-Net 的固定阈值 θ 不同, 我们提出了一个动态软阈值模块, 该模块根据输入图像信息动态调整 θ_d 。这种方法更好地处理了密集重叠的目标和空间上下文变化。

如图 4 所示, 该模块采用双卷积层捕获多尺度特征, 与 Li et al. (2025, 2019a) 中的方法类似。 $\mathcal{F}_d^{(k)}$ 的输出经过两个 3×3 卷积层处理, 生成特征图 \tilde{U}_1 和 \tilde{U}_2 。这些特征图被拼接形成 $\tilde{U} = [\tilde{U}_1, \tilde{U}_2]$ 。空间关系则通过并行池化操作捕获:

$$SA_{\text{avg}} = P_{\text{avg}}(\tilde{U}), \quad SA_{\text{max}} = P_{\text{max}}(\tilde{U}). \quad (16)$$

池化后的特征通过卷积处理:

$$(\hat{S}A) = \text{Conv}^{2 \rightarrow N}([SA_{\text{avg}}; SA_{\text{max}}]), \quad (17)$$

再通过 sigmoid 激活函数生成空间选择性掩码:

$$(\tilde{S}A)_i = \text{sigmoid}((\hat{S}A)_i). \quad (18)$$

动态阈值 θ_d 随后通过将这些掩码与多尺度特征图结合, 并进行最终卷积运算来计算。

$$\theta_d = C\left(\sum_{i=1}^N (\tilde{S}A)_i \cdot \tilde{U}_i\right). \quad (19)$$

其中 $C(\cdot)$ 表示卷积运算。

初始化与学习目标。 DISTA-Net 采用与迭代稀疏编码算法类似的线性初始化方法，其中初始估计 $\tilde{s}_i^{(0)}$ 通过从训练对 $\{(z_i, s_i)\}$ 中学习到最优线性投影 Q_{init} 获得（详情见[补充材料](#)）。

训练目标结合了重建保真度与结构保存：

$$\mathcal{L} = \mathcal{L}_{\text{discrepancy}} + \gamma \mathcal{L}_{\text{constraint}}, \quad (20)$$

其中：

$$\mathcal{L}_{\text{discrepancy}} = \frac{1}{MN_s} \sum_{i=1}^M \|\tilde{s}_i^{(N)} - s_i\|_2^2, \quad (21)$$

$$\mathcal{L}_{\text{constraint}} = \frac{1}{MN_s} \sum_{i=1}^M \sum_{k=1}^N \|\tilde{\mathcal{F}}^{(k)}(\mathcal{F}_d^{(k)}(s_i)) - s_i\|_2^2. \quad (22)$$

其中， $\mathcal{L}_{\text{discrepancy}}$ 计算重建的 $\tilde{s}_i^{(N)}$ 与真实值 s_i 之间的均方误差，而 $\mathcal{L}_{\text{constraint}}$ 通过 $\tilde{\mathcal{F}}(\cdot) \circ \mathcal{F}_d(\cdot) = \mathbf{I}$ 强制执行多阶段约束， $\gamma = 0.01$ 用于平衡这些目标。

5 实验

5.1 实验设置

训练：采用 CSIST-100K 数据集中的图像作为输入，我们运用 4.1 中提到的方法进行亚像素分割，采样网格比例为 c ，生成一个非混叠的高分辨率图像作为真值。具体而言，对于第 i 个目标，其参数为 (x_i, y_i, s_i) ，强度值 s_i 被分配给位置为 $(c \cdot x_i + \frac{c-1}{2}, c \cdot y_i + \frac{c-1}{2})$ 的像素，其他像素均设置为零。通过适当选择 c 值，每个目标可在亚像素网格中独立存在，从而实现精确的空间分离。

测试：以 50 为阈值保留解混图像中高于该值的像素作为预测目标，亮度值作为预测目标的预测峰值强度，其在图像中的位置需要重新投影回原始的 11×11 坐标空间，以计算 CSO-mAP。具体地，对于第 i 个预测目标，其坐标计算为 $\left(\frac{x_i - \lfloor \frac{c-1}{2} \rfloor}{c}, \frac{y_i - \lfloor \frac{c-1}{2} \rfloor}{c}\right)$ 。为了更全面地评估模型的性能，我们还通过直接比较预测的高分辨率图像与亚像素划分的真实图像，计算 PSNR 和 SSIM 指标。

超参数。我们设定 $c = 3$ 作为基线。随机选取的 64 张图像为一个批次训练，阶段数 $N = 6$ ，动态变换分支系数 $(1 - \alpha)$ 设置为 0.3。

5.2 与最先进方法的对比实验

实验结果。表 1 比较了 DISTA-Net 与传统优化方法 (ISTA)、图像超分辨率网络以及深度展开方法在 CSIST-100K 数据集上的性能，评估指标包括计算效率 (#P/FLOPs)、定位精度 (CSO-mAP) 和图像质量 (PSNR/SSIM)。

在定位精度评估方面，我们采用 CSO-mAP 指标并设置不同距离阈值（从 AP-05 到 AP-25），其中较小的阈值表示更严格的定位精度要求。例如，AP-05 评估 0.05 像元宽度内的检测精度，代表极高的精度需求。考虑到亚像素划分 $c = 3$ 时存在 0.236 像元宽度的固有系统误差，本研究以 0.20 像元和 0.25 像元精度的 AP-20 和 AP-25 作为主要性能指标。实验结果表明 DISTA-Net 可实现 86.18% 和 97.14% 的准确率，显著优于现有方法。在反映空间邻近小目标的平均精度指标 mAP 中，我们的方法以 46.74% 的准确率保持领先优势，展现出其不同精度要求下的稳健定位能力。

在模型效率方面，DISTA-Net 以适中的计算成本 (2.179M 参数和 35.103G Flops) 实现了上述结果，与 ACTNet (46.212M 参数, 62.798G Flops) 和 HAN (64.342M 参数, 495G Flops) 等方法相比，显著优于现有方法。此外，DISTA-Net 在所有对比方法中展现出最优的图像质量，其 PSNR (37.8747) 和 SSIM (99.79) 得分最高，展现了其卓越的图像细节保留能力。

这些综合结果验证了 DISTA-Net 在计算效率、定位精度和重建质量之间实现了有效的平衡。

可视化比较。图 5 比较了不同方法在包含 3-5 个相邻目标的场景中，采用 $3 \times$ 亚像素划分进行重建的结果，评估了在空间邻近红外多目标场景中的性能。现有方法在重建空间邻近目标时表现不佳，常出现模糊边界或目标合并现象，且随着目标数量增加（如五目标场景），性能退化更为明显。相比之下，DISTA-Net 在多目标高重叠度条件下仍能同时保留目标数量与亚像素位置，同时保持清晰边界和准确的空间分布。

5.3 消融实验

不同组件的影响。我们通过消融实验评估 DISTA-Net 中各组件的贡献，结果如表 2 所示。第二行 (DISTA-Net w/o DT) 对应不包含动态变换的模型变体，第三行 (DISTA-Net w/o Thres.) 则代表不包含动态软阈值的模型。与基线模型 ISTA-Net 相比，我们的完整模型在 CSO-mAP（从 45.16% 提升至 46.74%）和 AP-20（从 82.58% 提升至 86.18%）指标上均取得了显著提升。

移除动态变换导致性能轻微下降（CSO-mAP 降至 46.32%），凸显其在提升模型性能中的作用。移除动态软阈值导致性能最显著下降（CSO-mAP 降至 46.17%），强调其在确保准确性中的关键作用。

模型性能与采样网格。为了研究采样网格比率对模型性能和计算效率的影响，我们在不同采样网格比率 ($c = 5$ 和 7) 下对多种方法进行了实验。表 3 展示了在不同采样配置下模型性能的全面比较。

随着采样网格比的增加，所有方法的检测性能均呈现出整体提升的趋势，这归因于目标位置表示精度

Method	#P ↓	FLOPs ↓	CSO-mAP						PSNR ↑	SSIM ↑
			mAP	AP-05	AP-10	AP-15	AP-20	AP-25		
传统优化方法										
ISTA Daubechies et al. (2004)	-	-	7.46	0.01	0.31	2.39	9.46	25.14	-	-
图像超分辨率方法										
ACTNet Zhang et al. (2023)	46.212M	62.80G	45.61	0.38	7.46	41.13	83.12	95.95	35.4526	99.70
CTNet Wang et al. (2021)	0.400M	2.756G	45.11	0.38	7.53	40.39	82.11	95.14	35.1499	99.70
DCTLSA Zeng et al. (2023)	0.865M	13.69G	44.51	0.39	7.35	39.35	81.15	94.34	34.6314	99.65
EDSR Lim et al. (2017)	1.552M	12.04G	45.32	0.33	7.07	40.58	83.24	95.41	35.3724	99.71
EGASR Qiu et al. (2023)	2.897M	17.73G	45.51	0.42	8.03	41.32	85.71	95.08	34.5681	99.66
FeNet Wang et al. (2022)	0.348M	2.578G	45.77	0.42	8.19	42.13	83.30	94.80	34.1531	99.66
RCAN Zhang et al. (2018a)	1.079M	8.243G	45.87	0.42	7.96	41.81	83.61	95.57	35.2119	99.69
RDN Zhang et al. (2018b)	22.306M	173.0G	45.81	0.35	7.11	41.07	84.07	96.43	36.4686	99.74
SAN Dai et al. (2019)	4.442M	34.05G	45.95	0.36	7.35	41.17	84.32	96.57	36.5037	99.74
SRCNN Dong et al. (2015)	0.019M	1.345G	29.06	0.23	4.10	21.65	49.95	69.39	28.7608	98.44
SRFBN Li et al. (2019b)	0.373M	3.217G	46.05	0.43	8.31	42.83	83.72	94.95	34.0174	99.68
HAN Niu et al. (2020)	64.342M	495.0G	45.70	0.39	7.46	40.90	83.61	96.17	35.2703	99.71
HiT-SIR Zhang et al. (2024)	0.769M	7.357G	44.51	0.40	7.16	39.55	80.94	94.50	34.7400	99.70
HiT-SNG Zhang et al. (2024)	0.952M	13.324G	45.01	0.39	7.34	40.19	81.98	95.17	35.1390	99.71
HiT-SRF Zhang et al. (2024)	0.844M	7.945G	44.71	0.40	7.71	40.03	80.94	94.49	34.5449	99.67
深度展开方法										
ISTA-Net Zhang and Ghanem (2018)	0.171M	12.77G	45.16	0.41	7.71	40.57	82.58	94.53	33.9215	99.68
ISTA-Net+ Zhang and Ghanem (2018)	0.337M	24.33G	46.06	0.42	7.66	41.58	84.46	96.17	36.0892	99.72
LAMP Daubechies et al. (2004)	2.126M	0.278G	14.22	0.05	1.11	7.31	21.56	41.06	27.8299	96.89
LIHT Daubechies et al. (2004)	21.10M	1.358G	10.35	0.06	0.92	4.99	14.74	30.5	27.5107	96.42
LISTA Daubechies et al. (2004)	21.10M	1.358G	30.13	0.25	4.13	22.29	51.18	72.82	29.8936	99.12
FISTA-Net Xiang et al. (2021)	0.074M	18.96G	44.66	0.45	7.68	39.74	81.24	94.19	35.7519	99.67
TiLISTA Daubechies et al. (2004)	2.126M	0.278G	14.95	0.06	1.23	7.72	22.50	46.23	27.7038	97.40
★ DISTA-Net (Ours)	2.179M	35.10G	46.74	0.38	7.54	42.44	86.18	97.14	37.8747	99.79

表 1 与当前最先进方法在 CSIST-100K 数据集上的比较。



图 5 不同数量相邻目标场景中 3× 亚像素划分重建的视觉比较。红色方框标出了目标表现出显著亚像素特征的区域。

Method	#P ↓	FLOPs ↓	CSO-mAP
ISTA-Net	0.171M	12.766G	45.16
DISTA-Net w/o DT	0.171M	39.544G	46.32
DISTA-Net w/o Thres.	5.111M	92.62G	46.17
★ DISTA-Net (Ours)	2.179M	35.103G	46.74

具包，为后续研究提供支持。

表 2 不同组件的作用。

Method	#P ↓	FLOPs ↓	CSO-mAP		
			mAP	AP-10	AP-15
<i>c=5</i>					
ISTA-Net	0.171M	39.544G	66.90	56.73	87.26
ISTA-Net+	0.225M	48.158G	<u>68.50</u>	57.96	<u>89.52</u>
CFGN	0.538M	4.122G	67.95	<u>58.08</u>	88.35
★ DISTA-Net (Ours)	5.153M	102.4G	69.58	60.95	90.85
<i>c=7</i>					
ISTA-Net	0.171M	89.51G	<u>71.19</u>	<u>76.45</u>	84.16
ISTA-Net+	0.225M	103.0G	71.09	74.90	<u>84.90</u>
CFGN	0.548M	4.202G	70.38	73.88	83.97
★ DISTA-Net (Ours)	6.409M	142.3G	72.84	78.47	86.09

表 3 在 CSIST-100K 数据集上，不同采样网格下方法的比较。

的巨大提升。DISTA-Net 在定位精度方面始终保持领先优势，在所有配置下均优于其他方法。然而，这种性能提升是以计算复杂度大幅增加为代价的。值得注意的是，DISTA-Net 在更高分辨率任务中表现出更强的性能。与其他方法相比，尽管它们显示出类似的趋势但性能较低，我们的方法在增加相同的采样比率时实现了更高的精度提升。基于这些观察，我们建议根据实际需求选择适当的采样网格比率，以在检测精度和计算效率之间实现良好平衡。

本文方法 vs "超分辨率 + 检测器"。解混阶段通常作为检测后的精细化步骤。为严谨验证，我们专门实现了“超分辨率 + 检测器”的基准流程：采用 YOLOv11 作为检测器，配合主流超分辨率方法（使用我们的红外数据及解混真值重新训练，专门优化点源超分辨率任务）。实验结果验证了 DISTA-Net 的持续优势：DISTA-Net + YOLOv11 组合取得 47.82 的 CSO-mAP，优于 SRFBN + YOLOv11 (45.74) 和 CFGN + YOLOv11 (46.71)。通过可视化分析发现，传统 SR 方法和我们的解混方法均能生成具有清晰独立峰的高分辨率图像，此时简单阈值分割即可有效分离目标，这表明解混阶段在该任务中起主导作用。

超参数分析。我们在**补充材料**中分析了模型性能与阶段数及动态分支系数的关系，展示了稳健的架构和参数设计。

6 总结

在本研究中，我们针对空间邻近红外小目标解混问题，提出动态迭代收缩阈值网络 (DISTA-Net)。通过将传统稀疏重构转化为动态深度展开框架，DISTA-Net 自适应生成卷积、阈值参数，显著提升亚像素检测精度。为推动该领域研究，我们引入了 CSIST-100K 数据集、CSO-mAP 评估指标和 GrokCSO 工

References

- Xiaohan Chen, Jialin Liu, Zhangyang Wang, and Wotao Yin. Hyperparameter tuning is all you need for lista. In *Neural Information Processing Systems (NeurIPS)*, pages 11678–11689, 2021. 1
- Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11065–11074, 2019. 7
- Yimian Dai, Yiquan Wu, Fei Zhou, and Kobus Barnard. Attentional local contrast networks for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 59(11):9813–9824, 2021a. 2
- Yimian Dai, Yiquan Wu, Fei Zhou, and Kobus Barnard. Asymmetric contextual modulation for infrared small target detection. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 949–958, 2021b. 2
- Yimian Dai, Xiang Li, Fei Zhou, Yulei Qian, Yaohong Chen, and Jian Yang. One-stage cascade refinement networks for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–17, 2023. 2
- Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004. 2, 4, 7
- Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2015. 7
- Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *International Conference on Machine Learning (ICML)*, pages 399–406, 2010. 2
- Tiantong Guo, Hojjat Seyed Mousavi, and Vishal Monga. Adaptive transform domain image super-resolution via orthogonally regularized deep networks. *IEEE Transactions on Image Processing*, 28(9):4685–4700, 2019. 3
- Renke Kou, Chunping Wang, Zhenming Peng, Zhihe Zhao, Yaohong Chen, Jinhui Han, Fuyu Huang, Ying Yu, and Qiang Fu. Infrared small target segmentation networks: A survey. *Pattern Recognition*, 143:109788, 2023. 1
- Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. Selective kernel networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 510–519, 2019a. 5
- Xin Li, Weisheng Dong, Jinjian Wu, Leida Li, and Guangming Shi. Superresolution image reconstruction: Selective milestones and open problems. *IEEE Signal Processing Magazine*, 40(5):54–66, 2023. 1
- Yuelong Li, Mohammad Tofighi, Junyi Geng, Vishal Monga, and Yonina C. Eldar. Efficient and interpretable deep blind image deblurring via algorithm unrolling. *IEEE Transactions on Computational Imaging*, 6:666–681, 2020. 3
- Yuxuan Li, Xiang Li, Yunheng Li, Yicheng Zhang, Yimian Dai, Qibin Hou, Ming-Ming Cheng, and Jian Yang. Sm3det: A unified model for multi-modal remote sensing object detection. *arXiv preprint arXiv:2412.20665*, 2024. 2
- Yuxuan Li, Xiang Li, Yimian Dai, Qibin Hou, Li Liu, Yongxiang Liu, Ming-Ming Cheng, and Jian Yang. Lsknet: A foundation lightweight backbone for remote sensing. *International Journal of Computer Vision*, 133(3):1410–1431, 2025. 5
- Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3867–3876, 2019b. 7
- T Liaudat, J Bonnin, J-L Starck, MA Schmitz, A Guinot, M Kilbinger, and SDJ Gwyn. Multi-ccd modelling of the point spread function. *Astronomy & Astrophysics*, 646:A27, 2021. 3
- Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, pages 136–144, 2017. 7
- Jie Mei, Yi-Bo Zheng, and Ming-Ming Cheng. D2anet: Difference-aware attention network for multi-level change detection from satellite imagery. *Computational Visual Media*, 9(3):563–579, 2023. 2
- Florian Meyer and Jason L. Williams. Scalable detection and tracking of geometric extended objects. *IEEE Transactions on Signal Processing*, 69:6283–6298, 2021. 1
- Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021. 2
- Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *European Conference on Computer Vision (ECCV)*, pages 191–207, 2020. 7
- K. Pruett, N. McNaughton, and M. Schneider. Closely Spaced Object Classification Using MuyGPys. In *Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*, page 158, 2023. 1

- Zhonghang Qiu, Huanfeng Shen, Linwei Yue, and Guizhou Zheng. Cross-sensor remote sensing imagery super-resolution via an edge-guided attention-based network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 199:226–241, 2023. 7
- José A. Sobrino, Fabio Del Frate, Matthias Drusch, Juan C. Jiménez-Muñoz, Paolo Manunta, and Amanda Regan. Review of thermal infrared applications and requirements for future high-resolution sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):2963–2972, 2016. 1
- Oren Solomon, Regev Cohen, Yi Zhang, Yi Yang, Qiong He, Jianwen Luo, Ruud J. G. van Sloun, and Yonina C. Eldar. Deep unfolded robust PCA with application to clutter suppression in ultrasound. *IEEE Transactions on Medical Imaging*, 39(4):1051–1063, 2020. 3
- Xiaozhong Tong, Shaojing Su, Peng Wu, Runze Guo, Junyu Wei, Zhen Zuo, and Bei Sun. Msaffnet: A multi-scale label-supervised attention feature fusion network for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023. 2
- Zhiqiang Wan, Haibo He, and Bo Tang. A generative model for sparse hyperparameter determination. *IEEE Transactions on Big Data*, 4(1):2–10, 2018. 1
- Shunzhou Wang, Tianfei Zhou, Yao Lu, and Huijun Di. Contextual transformation network for lightweight remote-sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2021. 7
- Zheyuan Wang, Liangliang Li, Yuan Xue, Chenchen Jiang, Jiawen Wang, Kaipeng Sun, and Hongbing Ma. Fenet: Feature enhancement network for lightweight remote-sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12, 2022. 7
- Zhe Wang, Tao Zang, Zhiling Fu, Hai Yang, and Wenli Du. Rlpgb-net: Reinforcement learning of feature fusion and global context boundary attention for infrared dim small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023. 2
- Gang Wu, Junjun Jiang, Kui Jiang, and Xianming Liu. Fully 1×1 convolutional network for lightweight image super-resolution. *Machine Intelligence Research*, 21(6):1062–1076, 2024. 1
- Jinxi Xiang, Yonggui Dong, and Yunjie Yang. Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging*, 40(5):1329–1339, 2021. 7
- Hai Xu, Sheng Zhong, Tianxu Zhang, and Xu Zou. Multiscale multilevel residual feature fusion for real-time infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023. 2
- yan yang, Jian Sun, Huibin Li, and Zongben Xu. Deep admm-net for compressive sensing mri. In *Neural Information Processing Systems (NeurIPS)*, 2016. 3
- Di You, Jingfen Xie, and Jian Zhang. Ista-net++: Flexible deep unfolding network for compressive sensing. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2021. 3
- Kun Zeng, Hanjiang Lin, Zhiqiang Yan, and Jinsheng Fang. Densely connected transformer with linear self-attention for lightweight image super-resolution. *IEEE Transactions on Instrumentation and Measurement*, 72:1–12, 2023. 7
- Hui Zhang, Hui Xu, and Liangkui Lin. Super-resolution method of closely spaced objects based on sparse reconstruction using single frame infrared data. *Acta Optica Sinica*, 33(4):0411001–8, 2013. 1
- Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1828–1837, 2018. 3, 7
- Mingjin Zhang, Qian Xu, Yuchun Wang, Xi Li, and Haojuan Yuan. Mirsam: multimodal vision-language segment anything model for infrared small target detection. *Visual Intelligence*, 3(1):1–13, 2025. 2
- Xiang Zhang, Yulun Zhang, and Fisher Yu. Hit-sr: Hierarchical transformer for efficient image super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 483–500, 2024. 7
- Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision (ECCV)*, pages 286–301, 2018a. 7
- Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2472–2481, 2018b. 7
- Zheng Zhang, Fanchen Liu, Changan Liu, Qing Tian, and Hongquan Qu. Actnet: A dual-attention adapter with a cnn-transformer network for the semantic segmentation of remote sensing imagery. *Remote Sensing*, 15(9):2363, 2023. 7
- Mingjing Zhao, Wei Li, Lu Li, Jin Hu, Pengge Ma, and Ran Tao. Single-frame infrared small-target detection: A survey. *IEEE Geoscience and Remote Sensing Magazine*, 10(2):87–119, 2022. 1

补充材料

1 多目标成像

当目标位于较远距离时,可将其近似为点光源,其在传感器焦平面上的能量分布由 PSF 描述,该函数近似于二维高斯函数。当多个目标在近距离存在时,探测器接收到的能量响应是各 PSF 的线性叠加。这一叠加原理是空间邻近目标解混过程的基础,其中叠加响应必须被分解以识别和表征单个目标,如图 6 所示。

2 CSO-mAP 指标

一旦预测结果被分类为真正例 (True Positive, TP) 或假正例 (False Positive, FP), 将按照 COCO 数据集标准的常见做法构建一个二进制列表, 其中真正例预测用 1 表示, 假正例预测用 0 表示。该二进制列表是生成精确率-召回率 (Precision Recall, PR) 曲线的基础。通过动态调整检测亮度阈值, 可以获得一系列精确度和召回率值, 从而形成 PR 曲线。平均精确度 (Average Precision, AP) 作为 PR 曲线下的面积, 能够全面评估模型在不同灰度阈值下的性能, 准确总结精确度与召回率之间的权衡关系。最后, 我们引入 CSO-mAP, 其通过对不同距离阈值 δ_k 下的 AP 取平均值, 为 CSIST 解混任务中模型性能的比较提供了一个标准化指标。

3 GrokCSO 工具箱

鉴于该领域缺乏专业且易于使用的工具, 我们推出了 GrokCSO——一个专为提升空间邻近红外小目标重建效果而设计的综合性开源工具包。尽管通用计算机视觉领域已拥有 MMDetection 和 GlueCV 等丰富的目标检测工具包, 但由于红外小目标解混任务的特殊性, 迫切需要一个专属平台。缺乏此类平台导致的研究工作碎片化, 阻碍了实验的可重复性和算法的比较分析。

基于强大的 PyTorch 框架开发, GrokCSO 经过精心设计, 旨在解决 CSIST 解混过程中固有的独特挑战。其独特之处在于:

- 预训练模型与可复现性:** GrokCSO 为研究者提供了一套完整的预训练模型库, 包含前沿算法的训练脚本和日志记录。这些资源不仅增强了研究结果的可复现性, 还支持对不同算法方案进行细致入微的对比分析。
- 定制化灵活性与评估严谨性:** GrokCSO 通过大量可适配的主干网络和颈部结构, 支持更广泛的计算策略。该工具包整合了专用数据集加载器、尖端注意力机制以及多功能数据增强流程。更重要的是, GrokCSO 针对空间邻近红外小目

标解混任务的特殊性, 专门设计了精细化评估指标, 这些指标充分尊重了 CSO 问题独有的挑战特性。

4 超参数分析

阶段数量。 我们使用阶段数 $K = 2, 4, 6, 8,$ 和 10 对网络进行了训练。性能曲线如图 7 所示。随着阶段数从 2 增加到 6, CSO-mAP 分数有所提升, 并在 6 个阶段时达到峰值 (46.74%)。此后, 增加阶段数带来的收益递减, 这可能由于模型复杂度的提升。因此, 选择 6 个阶段可在准确性和效率之间实现最优平衡。

动态变换权重。 我们调整了动态变换分支相对于两个分支总贡献的比例, 并观察性能变化。从 0% 系数开始, 随着动态分支影响力的增加, 性能逐渐提升, 并在 30% 系数时达到峰值。模型在 50% 至 90% 的系数设置下保持稳定性能, 展现出鲁棒性。然而, 当系数达到 100% 时, 性能急剧下降至 30.62%, 这证实了合适的动态变换分支系数对实现最优特征表示至关重要。

5 初始化以及学习目标

初始化。 给定数据集 $(z_i, s_i)_{i=1}^M$, 其中 z_i 表示混叠目标的图像, s_i 是对应的超分辨率图像。我们利用目标信息 (x_i, y_i, g_i) 和缩放因子 c 来计算 s_i 中的高分辨率目标坐标。例如, 高分辨率图像中位置 $(c \cdot x_i + \frac{c-1}{2}, c \cdot y_i + \frac{c-1}{2})$ 处的灰度值 g_i 被用于生成 s_i 。

令 $Z = [z_1, \dots, z_M]$ 、 $S = [s_1, \dots, s_M]$ 。用于初始化 $\tilde{s}^{(0)}$ 的矩阵 Q_{init} 计算如下:

$$Q_{\text{init}} = \arg \min_Q \|QZ - S\|_F^2 = SZ^T(ZZ^T)^{-1}.$$

因此, 初始解混图像 $\tilde{s}^{(0)}$ 的计算表达式为:

$$\tilde{s}^{(0)} = Q_{\text{init}} Z.$$

损失函数。 为确保解混后的图像 \tilde{s} 在保持结构约束 $\tilde{\mathcal{F}}(\cdot) \circ \mathcal{F}_d(\cdot) = \mathbf{I}$ 的同时尽可能接近真实图像 s , 我们为 DISTA-Net 设计了如下端到端训练损失函数, 其中训练数据集规模为 M , 网络包含 N 个阶段, 图像尺寸为 N_s :

$$\mathcal{L} = \mathcal{L}_{\text{discrepancy}} + \gamma \mathcal{L}_{\text{constraint}},$$

其中:

$$\mathcal{L}_{\text{discrepancy}} = \frac{1}{MN_s} \sum_{i=1}^M \|\tilde{s}_i^{(N)} - s_i\|_2^2,$$

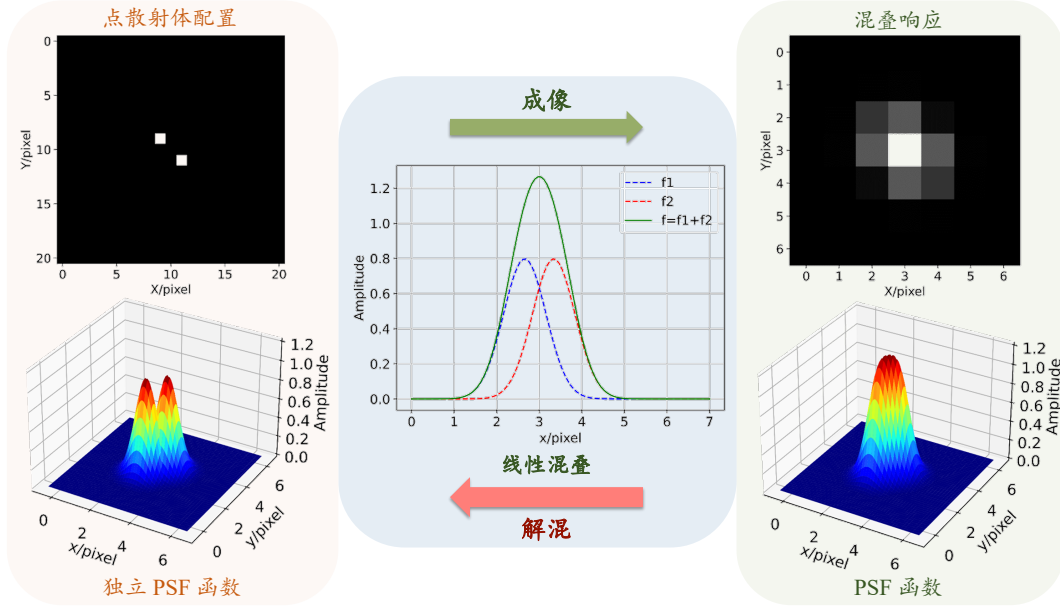


图 6 多目标成像中，远距离目标在成像平面上的成像可视为点源目标通过点扩散函数（Point Spread Function，PSF）进行的能量扩散过程。多目标成像本质上是多个叠加点源累积响应的结果。

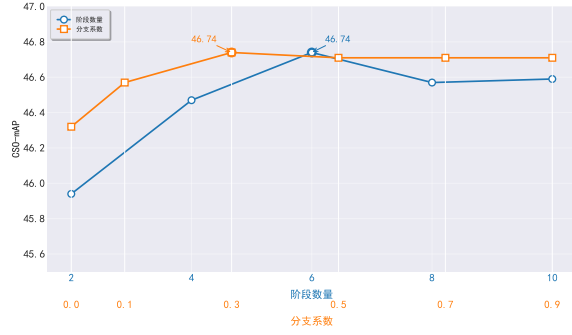


图 7 不同阶段数和分支系数下的 CSO-mAP 性能表现

$$\mathcal{L}_{\text{constraint}} = \frac{1}{MN_s} \sum_{i=1}^M \sum_{k=1}^N \|\tilde{\mathcal{F}}^{(k)}(\mathcal{F}_d^{(k)}(s_i)) - s_i\|_2^2.$$

其中， $L_{\text{discrepancy}}$ 用于衡量超分辨率图像 $\tilde{s}_i^{(N)}$ 与真实图像 s_i 之间的均方误差。 $L_{\text{constraint}}$ 通过确保每阶段 k 中 $\tilde{\mathcal{F}}^{(k)}$ 与 $\mathcal{F}_d^{(k)}$ 的复合变换近似恒等变换，来强制执行结构约束。 γ 是用于平衡差异项与约束项的参数。该损失函数旨在平衡解混图像的精度与变换函数的结构完整性。