arXiv:2602.01681v1 [eess.IV] 2 Feb 2026

高光谱图像 I　mage 与光谱波段及融合尺度无关性的图像融合

Yu-Jie Liang *1 Zihan Cao *2 Liang-Jian Deng [2 3] Yang Yang [1] Malu Zhang [1]

摘要

当前用于多光谱与高光谱图像融合（多光谱／高光谱融合）的深度学习模型通常针对固定的光谱波段和空间尺度进行设计，因此难以迁移至各种不同的传感器。为此，我们提出 SSA，这是一个具备光谱波段和融合尺度无关性的通用多光谱／高光谱融合框架。具体而言，我们引入了套娃核（MK），这是一种新型算子，能够使单个模型适应任意数量的光谱通道。同时，我们以隐式神经表示（INR）骨干网络构建 SSA，将高光谱信号建模为连续函数，从而能够在任意空间分辨率下进行重建。这两种无关性结合起来，使单个多光谱/高光谱融合模型能够有效泛化到未见过的传感器和空间尺度。大量实验表明，我们的单一模型在取得最先进性能的同时，还能良好地泛化到未见过的传感器和尺度，为未来的高光谱基础模型铺平道路。

1. 引言

高光谱成像（HSI）（Amigo，2019年）在数百个窄波段上提供丰富的光谱信息，使其在医学诊断（Fauvel 等人，2013年）、遥感和目标检测（Uzair 等人，2013年；Cheng 等人，2025）等领域具有不可替代的价值。然而，光学系统中的一个基本权衡意味着，这种丰富的光谱细节往往以较低的空间分辨率为代价，从而阻碍分类（He 等人，2017）和目标跟踪（Chen 等人，2021a年）等下游任务。为了解决这一问题，

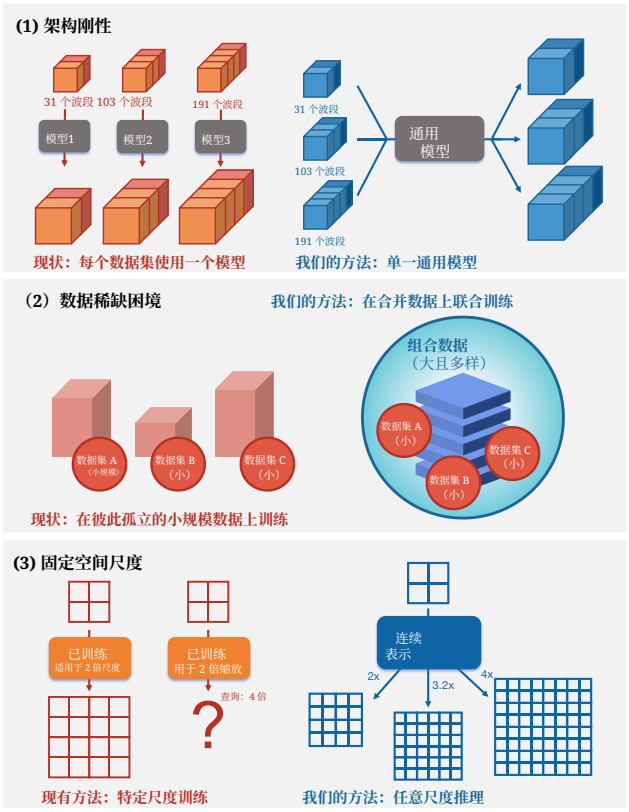

图1. 高光谱图像融合中现有范式与我们通用方法的对比。我们的方法能够适应不同光谱波段的联合训练，并可在任意尺度下进行融合推理。

多光谱与高光谱图像融合（MS/HS融合）已成为一种关键解决方案，旨在通过融合同一场景的低分辨率高光谱图像与高分辨率多光谱图像，重建高分辨率高光谱图像。近年来，随着深度学习的快速发展，基于深度学习的方法（Masi等人，2016；Yang等人，2017；Dian等人，2018）逐渐成为MS/HS融合的主流。这些方法能够自动学习复杂的非线性关系，在高光谱图像融合任务中展现出巨大的潜力。

尽管取得了显著进展，基于深度学习的方法在现实世界中的适用性仍受到这一根深蒂固的挑战所制约：传感器多样性（Hong 等人，2025）。不同传感器采集的数据集（如 AVIRIS（224 个波段）或帕维亚大学（103 个波段））在光谱波段数量上差异显著，这暴露了一个根本性问题：

*共同贡献　[1]中国电子科技大学计算机科学与工程学院（网络空间安全学院），中国成都　[2]中国电子科技大学数学科学学院，中国成都　[3]中国电子科技大学四川省多灾种早期预警重点实验室，中国成都。通信作者：Malu Zhang <maluzhang@uestc.edu.cn>。

预印本。*2026*年*2*月*3*日。

当前模型缺乏通用性。如图1所示，这一局限主要体现在三个方面：（1）架构僵化：大多数深度神经网络的通道维度都是硬编码的，迫使研究人员要么针对每种传感器分别训练独立模型（Deng 等，2023b；Wu 等，2025），要么采用代价高昂的微调（Gon-zalez 等，2025）。有些研究甚至会手动选择波段子集（Suryanarayana 等，2025），这会损失宝贵的光谱信息。（2）数据稀缺与模型规模的两难：高光谱图像数据集通常规模较小。在这类孤立的小规模数据上训练会限制模型容量，阻碍可泛化特征的学习，从而导致较高的过拟合风险（Hoffmann 等，2022；Yan 等，2025）。（3）固定的空间缩放因子：当前模型（Deng 等，2023a；Hu 等，2022）均针对固定的整数上采样因子（例如，$4\times, 8\times$）进行训练。它们学习的是离散的网格到网格映射，无法处理许多应用所需的任意或非整数缩放比例。

为打破这些壁垒，我们提出了 SSA 这一统一框架，通过两项技术创新应对上述挑战。为解决光谱刚性（1）及其对数据稀缺问题的影响（2），鉴于硬编码架构阻碍了跨传感器汇集数据进行大规模训练，我们提出了套娃核（MK）。受套娃表示学习（MRL）（Kusupati 等人，2022）的启发，MK 使单个模型能够处理数量可变的光谱波段，从而支持在异构数据集上联合训练。同时，为解决空间刚性（3），我们基于隐式神经表示（INR）骨干网络构建框架，将信号建模为连续函数，并支持以任意（包括非整数）缩放因子进行重建。在本研究中，我们将这些概念整合为一个统一的多光谱/高光谱融合模型。我们的贡献有三点：

- 我们提出了俄罗斯套娃核（MK），这是一种通过自适应卷积层实现的新型架构原理。它使模型能够原生处理任意数量的光谱波段，从而实现光谱波段无关性。
- 我们以隐式神经表示（INR）为基础构建框架，以连续函数取代离散缩放，从而能够在任意空间尺度上进行重建。
- 大量实验表明，SSA 达到了最先进的性能，同时能够很好地泛化到未见过的传感器和缩放因子，显著降低了针对每个传感器重新训练的需求。

2. 相关工作

2.1. MS/HS图像深度学习融合

继 PanNet（Yang 等人，2017）将深度学习引入高光谱与多光谱（MS/HS）图像融合之后，研究范式已从矩阵分解（Kawakami 等人，2011；Huang 等人，2013）和稀疏编码（Nezhad 等人，2016）等传统方法转向数据驱动的深度学习。当前研究主要聚焦于设计新颖的网络架构，例如引入注意力机制和 Transformer（Fang 等人，2024；Wu 等人，2025；Cao 等人，2025），或采用生成对抗网络（Shang 等人，2024；Zhou 等人，2022）和扩散模型（Cao 等人，2024）等流行范式。然而，这些高度专业化的架构往往与特定数据规格（例如光谱波段数量）深度耦合，导致其通用性和可迁移性严重不足。

近年来，一些研究（Zhang 等人，2022；Chen 等人，2023）开始将隐式神经表示（INR）应用于高光谱图像数据，将图像建模为坐标的连续函数。这种方法能够在任意尺度上进行重建，从而巧妙地解决了传统模型在空间上的僵化问题。后续研究探索了多种扩展方法，例如融入低秩属性（Wang 等人，2025b）、在频域中进行融合（Liang 等人，2024）或采用自适应采样（Deng 等人，2025）。然而，尽管解决了空间僵化问题，这些方法仍面临两项关键局限：首先，它们缺乏光谱灵活性，因为面对波段数量不同的数据集时，必须重新设计模型。其次，作为数据驱动方法，在可用高光谱图像数据有限的情况下，很难实现强泛化能力。

2.2. 套娃表示学习与灵活模型设计

我们的方法在理念上受到机器学习中构建更具自适应能力的通用模型这一更广泛趋势的启发。从表示粒度自适应的角度来看，套娃表示学习（MRL）（Kusupati等人，2022；Hojjat等人，2025；Cappellazzo等人，2025）被提出，用以探索模型的灵活性。MRL的核心思想是训练一个单一的高维特征向量 $z \in \mathbb{R}^d$，使其以嵌套方式编码不同粒度的信息，从而其较短的前缀同样可作为有效表示。该范式为计算需求各异的下游任务提供灵活的表示。

形式上，通过深度神经网络 z 获得的嵌套表示 $z := F(\cdot, \theta_F) : \mathcal{X} \to \mathbb{R}^d$ 包含不同粒度的表示：$z_{1:m} \in \mathbb{R}^m$，其中 m 是所需维度，而 $(r : s)$ 表示一个子表示

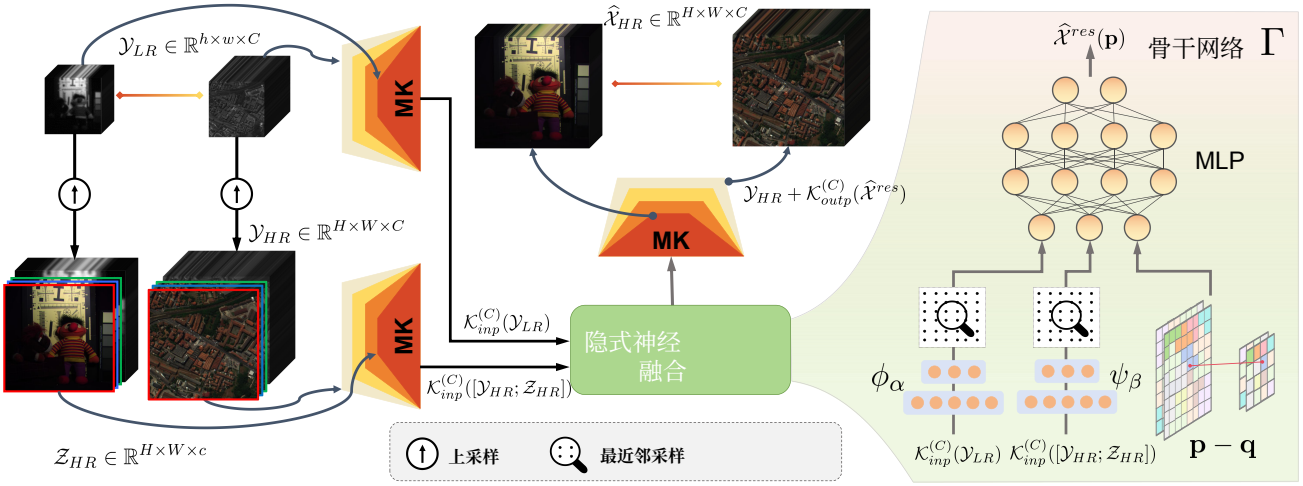

图2. 我们提出的 SSA 框架的整体架构。该端到端模型实现了通用映射函数 $\mathcal{F}$，以来自各种传感器的 LR-HSI 和 HR-MSI 为输入，并以任意尺度重建高保真高分辨率高光谱图像。

选择算法。通过最小化经验风险（即获得有效表示）得到套娃表示。在分类场景下，将嵌套 z 参数化为分类器权重$\mathbf{W}^{(m)} \in \mathbb{R}^{N,\ m}$，即可使用交叉熵损失对经验风险进行监督：

$$\min_{\substack{\theta_F,\\ \{\mathbf{W}^{(m)}\}_{m\in\mathcal{M},\\ m\le d}}} \frac{1}{N}\sum_{m\in\mathcal{M}} \mathcal{L}\left(\mathbf{W}^{(m)}\cdot F(x,\theta_F)_{1:m}; y\right), \quad (1)$$

其中，$\mathcal{M}$ 是一个嵌套的选择列表（例如，$\mathcal{M}:=\{m_1,\ldots,m_i,d\}, m_1\le m_i\le d$），$N$是总分类数，($x$，$y$)是数据/标签对。在 MRL 论文中，子表示选择算法被设计为维度切片，其中 r 固定为1，且 $s=m$。

与旨在学习灵活有效的表示并使用不同尺寸模型维度进行推理的 MRL 不同（Lin 等人，2024；Wang 等人，2024b），我们将嵌套原则引入高光谱数据融合内核，以处理各种输入图像通道，并将其称为套娃内核（MK，见 §3.2）。这一转换消除了此前的两个关键缺陷：架构刚性、数据稀缺以及模型规模问题。

3. 方法

3.1. 问题表述与模型概述

给定低分辨率高光谱图像（LR-HSI） $\mathcal{Y}_{LR}\in\mathbb{R}^{h\times w\times C}$和高分辨率多光谱图像（HR-MSI）$\mathcal{Z}_{HR}\in\mathbb{R}^{H\times W\times c}$（$c\ll C$），多光谱／高光谱融合^的目标是学习一个映射函数 $\mathcal{F}_\theta$：（$\mathcal{Y}_{LR},\mathcal{Z}_{HR}$）$\to\mathcal{X}_{HR}$，以重建高分辨率高光谱图像（HR-HSI）$\mathcal{X}_{HR}\in\mathbb{R}^{H\times W\times C}$，其中 θ 表示模型的可学习参数。

构建通用融合模型的一个关键挑战是传感器异构性。对于不同的传感器和光学相机，光谱波段数量 C 和空间缩放因子 s 可能有所不同。这意味着我们的训练数据来自一组数据集 $\mathcal{D}=\{D_1,\cdots,D_N\}$，并且对于从第 i个数据集中采样的任意图像对（$\mathcal{Y}^{(i)}_{LR},\mathcal{Z}^{(i)}_{HR}$），它具有相应的光谱波段数量 $C^{(i)}$ 和缩放因子 $r^{(i)}$。这就要求通用融合模型应具备光谱与空间无关性。

为实现这样一种通用的 $\mathcal{F}$，我们提出了 SSA 框架，其整体架构如图2所示。我们的模型被设计为一个基于连续表示的端到端融合流程。首先，模型针对给定的数据对进行光谱统一。如图所示，任意输入 $\mathcal{Y}_{LR}$，无论其原始波段数 C如何，都会通过套娃核层（MKL）$\mathcal{K}^{(C)}_{inp}(\cdot)$ 映射到固定的预定义特征维度（见 §3.2）。与此同时，将高分辨率多光谱图像 $\mathcal{Z}_{HR}$ 与空间上采样后的低分辨率高光谱图像 $\mathcal{Y}_{HR}\in\mathbb{R}^{H\times W\times c}$拼接起来，并由一个类似的层处理这个组合张量。

随后，这两个特征图被输入连续表示融合骨干网络（见 §3.3）。该模块的核心是对一个由多层感知机参数化的隐式函数进行建模，负责深度融合光谱与空间信息。融合骨干网络输出的连续特征随后被输入输出 MKL $\mathcal{K}^{(C)}_{outp}(\cdot)$，映射到具有原始波段数的像素空间$C$。最后，受（He 等人，2016）中残差学习的启发，^将输出 $\mathcal{X}^{res}$ 与上采样后的低分辨率高光谱图像相加，生成最终的高保真高分辨率高光谱图像。

3.2. 俄罗斯套娃核：光谱无关性 cism

为了解决§ 1中提到的缺点，尤其是处理通道数各异的输入时，我们引入了如下定义的套娃核层（MKL）$\mathcal{K}_{\star}^{(C)}$，$\star \in \{inp, outp\}$，用作输入层和输出层。每个MKL 都包含一个套娃式嵌套核，见图 3。这种方法避免了为每种可能的通道配置分别设置独立层。对于输入 MKL，给定一个包含 $\mathbf{X}$ 个 C 波段的高光谱立方体，其目标是将输入嵌入为具有 $\mathbf{Y}$ 个 D嵌入维度的特征。对于输出 MKL，过程则相反。

具体而言，每个 MKL 内部都将一个 MK 作为可学习权重。以 inputMKLweight 为例，MK 的权重为 $\mathbf{W}_{\text{nested}} \in \mathbb{R}^{D\times C_{\max}\times k\times k}$，其中 D 是固定的输出通道数，$C_{\max}$ 是该层可支持的预定义最大输入通道数（$C_{\text{in}} \le C_{\max}$），$k\times k$ 是核大小。此外，还会维护相应的偏置向量 $\mathbf{b} \in \mathbb{R}^{D}$。在前向传播过程中，使用基于切片的方法从嵌套核 $\mathbf{W}_{\text{valid}} \in \mathbb{R}^{D\times C_{\text{in}}\times k\times k}$ 中生成有效核：

$$\mathbf{W}_{\text{valid}} = \mathbf{W}_{\text{nested}}[:, : C_{\text{in}}, :, :], \tag{2}$$

其中，运算 $[:]$ 遵循 NumPy 数组切片规则，沿 $\mathbf{W}_{\text{nested}}$ 的输入通道维度选取前 C_{in} 个元素，以构成当前计算所用的核。利用当前有效核，通过标准二维卷积即可生成最终输出特征图：

$$\mathbf{Y} = \texttt{Conv2D}(\mathbf{X}, \mathbf{W}_{\text{valid}}, \mathbf{b}, \text{stride}, \text{padding}). \tag{3}$$

对于第 d个输出通道中的每个元素，其计算可形式化表示为：

$$\mathbf{Y}_{d,i,j} = \mathbf{b}_d + \sum_{c=0}^{C_{\text{in}}-1}\sum_{m=0}^{k-1}\sum_{n=0}^{k-1} \mathbf{X}_{c,i\cdot s+m,j\cdot s+n}\cdot(\mathbf{W}_{\text{valid}})_{d,c,m,n},$$

其中 s 表示步幅，索引 d、 c、 m和n 分别指输出通道、输入通道以及卷积核的空间维度。

类似的嵌套理念被应用于模型的输出MKL，使其能够动态生成与输入样本原始波段数 C 相匹配的输出。在前向传播过程中，会指定所需的输出通道数 C_{out}（$C_{\text{out}} \le C_{\max}$）。有效卷积核$\mathbf{W}_{\text{valid}} \in \mathbb{R}^{C_{\text{out}}\times D\times k\times k}$ 和偏置$\mathbf{b}_{\text{comp}} \in \mathbb{R}^{C_{\text{out}}}$ 通过对超集参数沿输出通道维度进行切片而生成：

$$\mathbf{W}_{\text{valid}} = \mathbf{W}_{\text{nested}}[: C_{\text{out}}, :, :, :], \tag{4}$$

$$\mathbf{b}_{\text{valid}} = \mathbf{b}_{\text{nested}}[: C_{\text{out}}]. \tag{5}$$

通过嵌套卷积核与动态子集切片这一机制，我们的SSA 模型实现了真正的光谱不可知-
不可知性，使其能够无缝处理异构

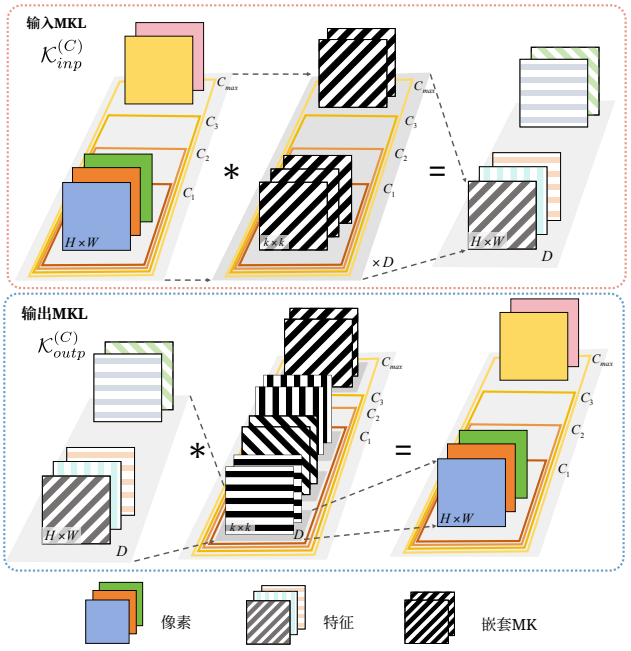

图3. MKL 编码与解码示意图。

来自通用架构中任意传感器的数据。详细的输入和输出 MKL 算法见附录§ A。

3.3. 作为连续函数的任意尺度图像融合

通过隐式融合主干网络 Γ 将目标HR-HSI实现为连续函数，使我们的模型具备真正的空间灵活性。该模型可以在任意坐标网格上进行查询，从而生成任意所需分辨率的输出。

如图2(b)所示，我们的模型首先通过两个并行编码器 ϕ_α，ψ_β，将MK之后的嵌入特征映射到潜在空间。该过程可表示为：

$$\begin{cases} \mathcal{E}_{pe} = \phi_\alpha(\mathcal{K}_{inp}^{(C)}(\mathcal{Y}_{LR})), \\ \mathcal{E}_{pa} = \psi_\beta\left(\mathcal{K}_{inp}^{(C)}([\mathcal{Y}_{HR}; \mathcal{Z}_{HR}])\right), \end{cases}$$

其中 $[;\cdot]$ 表示通道拼接。这些潜在编码 $\mathcal{E}_{pe}$ 和 $\mathcal{E}_{pa}$为后续的连续函数建模提供了所需的光谱与空间语义信息。

为了预测目标高分辨率高光谱图像空间中任意连续坐标 $\mathbf{p}$ 处的光谱向量，我们的融合模块采用一种局部的、基于查询的过程。对于给定像素，我们以其中心表示其坐标，并将高分辨率坐标映射到归一化方形网格 $\Omega = [-1, 1]^2$。随后，将 $H \times W$ 源网格中像素的整数坐标 (i, j) 映射为查询坐标$\mathbf{p}_{(i,j)} \in \Omega \subset \mathbb{R}^2$，映射方式如下：

表*1*。 在 10 个 CAVE 示例和 2 个 PaviaC 数据集示例上，对所有比较方法计算的平均值和标准差，模拟 $\times 4$（分布内）和 $\times 8$，$\times 16$，$\times 32$（分布外）。结果以 最佳和 次佳进行突出显示。

	指标	CAVE 数据集							PaviaC 数据集						
		DSPNet	MIMO-SST	DCINN	DCFormer	PSTUN	LRTN	Ours	DSPNet	MIMO-SST	DCINN	DCFormer	PSTUN	LRTN	Ours
In-Dist.	峰值信噪比(↑)	40.37 ± 3.25	45.35 ± 3.35	43.75 ± 3.77	45.81 ± 5.56	40.57 ± 3.72	41.13 ± 3.24	45.96 ± 4.69	43.89 ± 1.78	42.04 ± 0.78	44.79 ± 1.99	42.61 ± 1.19	44.94 ± 1.50	45.13 ± 1.01	45.92 ± 1.75
$\times 4$	光谱角映射(↓)	2.72 ± 0.89	2.74 ± 0.86	2.72 ± 0.81	2.70 ± 1.03	2.94 ± 0.89	2.92 ± 1.38	2.55 ± 1.00	3.15 ± 0.43	3.29 ± 0.45	2.59 ± 0.41	3.55 ± 0.43	2.95 ± 0.46	3.18 ± 0.45	2.02 ± 0.30
	相对全局综合误		2.08 ± 0.68	2.37 ± 0.95	2.01 ± 1.82	2.70 ± 0.89	2.38 ± 0.82	1.72 ± 1.13	1.58 ± 0.10	1.73 ± 0.02	1.34 ± 0.12	1.82 ± 0.05	1.47 ± 0.07	1.48 ± 0.01	1.07 ± 0.11
	结构相似性指数(↑)	0.993 ± 0.005	0.994 ± 0.005	0.992 ± 0.004	0.993 ± 0.004	0.992 ± 0.005	0.992 ± 0.006	0.994 ± 0.005	0.988 ± 0.000	0.986 ± 0.000	0.992 ± 0.000	0.983 ± 0.000	0.990 ± 0.000	0.988 ± 0.000	0.996 ± 0.000
	峰值信噪比(↑)	36.92 ± 3.47	29.73 ± 3.21	33.72 ± 3.39	41.46 ± 4.98	34.69 ± 3.17	33.40 ± 3.13	38.80 ± 3.79	38.03 ± 0.80	34.45 ± 0.55	32.82 ± 0.58	36.57 ± 0.88	37.27 ± 0.93	33.57 ± 0.40	38.59 ± 0.67
	光谱角映射(↓)	3.82 ± 1.59	5.55 ± 2.66	4.33 ± 1.94	3.65 ± 1.65	4.30 ± 1.59	4.55 ± 2.03	3.70 ± 1.45	5.66 ± 0.25	6.70 ± 0.14	6.09 ± 0.19	5.49 ± 0.26	5.48 ± 0.20	6.67 ± 0.15	4.74 ± 0.06
$\times 8$	相对全局综合误		9.91 ± 4.03	6.24 ± 2.62	3.08 ± 2.21	5.13 ± 1.93	6.00 ± 2.03	3.46 ± 1.57	4.05 ± 0.05	4.53 ± 0.07	4.53 ± 0.09	3.79 ± 0.06	3.60 ± 0.06	4.51 ± 0.05	3.47 ± 0.05
	结构相似性指数(↑)	0.980 ± 0.015	0.904 ± 0.099	0.958 ± 0.029	0.985 ± 0.014	0.964 ± 0.022	0.956 ± 0.030	0.982 ± 0.013	0.947 ± 0.004	0.923 ± 0.006	0.929 ± 0.005	0.948 ± 0.004	0.953 ± 0.003	0.931 ± 0.006	0.960 ± 0.004
Out-of-Dist.	峰值信噪比(↑)	30.87 ± 3.46	24.76 ± 2.93	27.65 ± 3.43	-	29.47 ± 3.16	27.53 ± 2.97	31.84 ± 3.47	32.92 ± 0.43	31.49 ± 0.20	28.74 ± 0.05	-	34.28 ± 0.39	28.65 ± 0.11	35.00 ± 0.02
$\times 16$	光谱角映射(↓)	6.51 ± 2.83	8.36 ± 4.48	6.93 ± 3.48	-	6.68 ± 2.44	6.73 ± 3.08	5.93 ± 2.42	8.43 ± 0.39	8.94 ± 0.13	8.72 ± 0.18	-	7.48 ± 0.30	9.64 ± 0.10	6.89 ± 0.08
	相对全局综合误		16.69 ± 6.56	12.19 ± 4.99	-	8.85 ± 2.80	11.44 ± 4.40	7.30 ± 2.88	6.79 ± 0.05	6.25 ± 0.07	6.78 ± 0.17	-	5.04 ± 0.04	7.39 ± 0.30	5.16 ± 0.07
	结构相似性指数(↑)	0.947 ± 0.032	0.834 ± 0.102	0.901 ± 0.061	-	0.926 ± 0.037	0.905 ± 0.057	0.953 ± 0.029	0.910 ± 0.001	0.898 ± 0.003	0.880 ± 0.002	-	0.932 ± 0.001	0.883 ± 0.005	0.935 ± 0.002
	峰值信噪比(↑)	27.91 ± 2.95	21.85 ± 2.56	23.96 ± 2.77	-	27.55 ± 2.75	24.48 ± 2.69	28.06 ± 2.64	30.51 ± 0.21	29.91 ± 0.02	26.51 ± 0.17	-	32.63 ± 0.46	26.53 ± 0.41	32.79 ± 0.26
$\times 32$	光谱角映射(↓)	10.03 ± 3.94	12.26 ± 5.90	10.34 ± 4.73	-	9.43 ± 3.28	9.64 ± 3.96	8.64 ± 3.17	10.26 ± 0.51	10.69 ± 0.33	10.93 ± 0.47	-	9.16 ± 0.66	11.89 ± 0.43	8.49 ± 0.19
	相对全局综合误		23.54 ± 8.22	18.98 ± 4.18	-	12.35 ± 3.30	16.80 ± 5.97	11.40 ± 6.69	8.62 ± 0.57	7.39 ± 0.33	8.44 ± 0.50	-	6.00 ± 0.32	9.26 ± 0.85	6.24 ± 0.30
	结构相似性指数(↑)	0.912 ± 0.036	0.774 ± 0.099	0.843 ± 0.066	-	0.896 ± 0.034	0.860 ± 0.061	0.918 ± 0.032	0.893 ± 0.002	0.887 ± 0.000	0.853 ± 0.003	-	0.919 ± 0.004	0.856 ± 0.001	0.922 ± 0.001

通过以下变换:

$$\mathbf{p}_{(i,j)} = \left(\frac{2i}{H-1} - 1, \frac{2j}{W-1} - 1 \right). \tag{6}$$

以查询坐标$\mathbf{p}$ 为枢轴，我们从空间潜码 $\mathcal{E}_{pa}$中采样精确的局部空间特征。同时，我们在低分辨率光谱潜码 $\mathcal{E}_{pe}$ 中找到与$\mathbf{p}^{(i,j)}$ 对应的最近邻$\mathbf{q}_{(i,j)} \in \Omega$，并采样该邻居的特征。

对于采样得到的局部光谱特征，我们将先前的光谱潜码、空间潜码以及描述细粒度位置的相对坐标$\mathbf{p} - \mathbf{q}$输入共享的MLP解码器:

$$\widehat{\mathcal{X}}^{res}(\mathbf{p}) = \texttt{MLP}(\mathcal{E}_{pe}(\mathbf{p}), \mathcal{E}_{pa}(\mathbf{q}), \mathbf{p} - \mathbf{q}), \tag{7}$$

其中，$\mathbf{p}$ 是高分辨率域中每个查询像素的归一化坐标。随后，将MLP的输出输入输出MKL层。借助MKL，该层可将MLP特征动态映射回当前样本的原始光谱波段数 C，生成最终预测的残差分量 $\mathcal{K}^{(C)}_{outp}(\widehat{\mathcal{X}}^{res}(\mathbf{p}))$。最后，将预测残差加到双三次插值基准图像上，即可得到完整的融合图像:

$$\widehat{\mathcal{X}}_{HR}(\mathbf{p}) = \mathcal{Y}_{HR}(\mathbf{p}) + \mathcal{K}^{(C)}_{outp}(\widehat{\mathcal{X}}^{res}(\mathbf{p})). \tag{8}$$

此外，为确保跨越网格边界时的空间连续性，我们遵循 LIIF（Chen 等人，2021b）的做法。在低分辨率特征图中，我们确定四个最近邻网格点$\mathbf{q}_i$，$i \in \{1, \dots 4\}$。在解码过程中，我们为每个邻点预测一个候选光谱残差 $\mathcal{X}^{res}_i(\mathbf{p})$以及一个标量权重 w_i。随后，通过对所有候选残差进行 Softmax加权融合，得到最终的残差向量:

$$\widehat{\mathcal{X}}^{res}(\mathbf{p}) = \sum_{i=1}^{4} \frac{\exp(w_i)}{\sum_{j=1}^{4} \exp(w_j)} \widehat{\mathcal{X}}^{res}_i(\mathbf{p}). \tag{9}$$

这种基于坐标的方法实现了真正的尺度无关性。通过将高分辨率高光谱图像建模为连续函数，我们的架构无需固定上采样层，因此能够生成任意分辨率的输出。

3.4. 跨传感器联合训练策略

传统的“一数据集一模型”范式不仅效率低下，而且还容易过拟合（Yan 等，2025），尤其是在高光谱图像数据集规模较小的情况下。在各种高光谱数据上训练通用模型通常面临一个挑战：不同的波段和空间尺寸无法组成批次，从而无法进行联合训练。

我们的训练流程旨在处理 N 异构数据集，记为 $\mathcal{D} = \{D_1, \dots, D_N\}$。首先，我们将所有训练样本划分到不同的桶中，其中每个桶D_i 仅包含来自单一源数据集的数据，因此具有相同的空间尺寸和光谱波段数量。在每个训练步骤中，我们通过两阶段采样过程构建一个小批次。首先，根据均匀随机分布从 D_i 中选择一个桶$\mathcal{D}$。随后，仅从选定的桶中抽取一批样本。这一策略确保单个小批次内的所有样本在结构上保持同质，而不同小批次之间的样本可以是异构的。正是这种批次间的异构性，使我们的模型能够适应这种训练方式。重要的是，所有样本均由通用模型参数 θ进行处理，并通过统一的混合重建损失函数进行优化。根据公式（1），重建损失可以用 MRL形式表示:

$$\min_{\substack{\theta, C \in \text{Dim}(D_i), \\ (\mathcal{Y}_{LR}, \mathcal{Z}_{HR}) \in D_i}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{total} \left[\mathcal{F}_\theta(\mathcal{Y}_{LR}, \mathcal{Z}_{HR}); \mathcal{X}_{HR} \right],$$

其中，$\mathcal{F}_\theta$ 是模型（即，$\mathcal{F}_\theta := \mathcal{K}^{(C)}_{outp} \circ \Gamma \circ \mathcal{K}^{(C)}_{inp}$ 和 $\circ$ 表示网络组合），Dim(·)表示数据桶中的波段数量。我们的总损失 $\mathcal{L}_{total}$ 是 L1损失与结构相似性指数的加权和

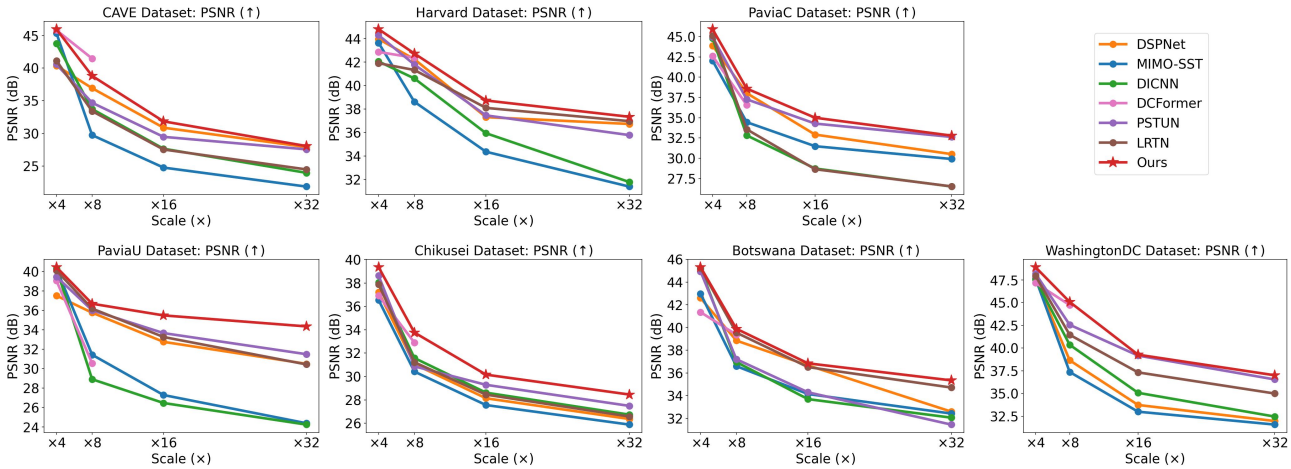

图 4. 七个数据集在多个缩放因子下的峰值信噪比（dB）定量比较。

（结构相似性指数）损失，定义如下：

$$\begin{aligned}\mathcal{L}_{total} &= \mathbb{E}_{(\mathcal{Y}_{LR},\mathcal{Z}_{HR},\mathcal{X}_{HR})\sim D_i}\left[\mathcal{L}_1 + \lambda\mathcal{L}_{SSIM}\right],\\ \mathcal{L}_1 &= \|\mathcal{F}_\theta(\mathcal{Y}_{LR},\mathcal{Z}_{HR}) - \mathcal{X}_{HR}\|_1,\\ \mathcal{L}_{SSIM} &= 1 - \mathrm{SSIM}(\mathcal{F}_\theta(\mathcal{Y}_{LR},\mathcal{Z}_{HR}),\mathcal{X}_{HR}),\end{aligned} \tag{10}$$

其中，λ 是加权系数，$\|\cdot\|$ 是 ℓ_1 范数，而 SSIM(·, ·) 是结构相似性指数。

通过这种联合训练策略，我们的 SSA 模型被迫学习通用的融合映射，而非记忆特定数据集的细节。这种方法赋予模型出色的泛化能力，使其能够适应未见过的传感器（见 §4.4.2）。

4. 实验

4.1. 数据集

为评估我们模型的性能和泛化能力，我们选取了七个多样化的公开高光谱数据集进行实验，涵盖室内、室外和遥感场景。数据集信息汇总于表B.1中。

我们的设置有意涵盖多种异质性：光谱波段数从 31（CAVE）到 191（华盛顿特区）不等，且来自 ROSIS 和 Hyperion 等传感器的数据在空间分辨率和光谱范围方面也存在显著差异。这种多样性为验证 SSA 框架的光谱无关性提供了理想的测试平台。我们遵循各数据集的标准训练/测试划分，并采用表中详述的适当图像块尺寸。

4.2. 实现细节

我们的 SSA 模型及所有实验均使用 PyTorch 框架实现（Paszke 等人，2019年）。所有训练和推理均在单台 NVIDIA RTX 4090 GPU 上进行。训练过程中，我们使用 AdamW 优化器（Loshchilov 和 Hutter，2017），并以批量大小为 4 训练模型 2000 个训练轮次。学习率由余弦退火学习率调度器管理，从 2e-4 开始退火至 1e-5。为实现 MK 的光谱无关性，最大通道数这一关键超参数 $C_{\max}$ 被设为 194。两个编码器均基于 EDSR 架构（Lim 等人，2017）。

4.3. 主要结果

为凸显我们提出的模型与现有方法之间的根本差异，对于每个数据集（在4× 比例下），我们分别为所有待比较的最先进方法独立训练专用模型，并在比例为 4×、8×、16×和 32×的测试集上进行评估。对于我们的方法，我们仅训练一个统一模型，其训练集由来自这七个数据集的训练样本混合构成。测试集评估采用相同方式。此外，我们采用了四种广泛使用的质量指标，即峰值信噪比（PSNR）、光谱角映射（SAM）、相对全局综合误差（ERGAS）和结构相似性指数（SSIM）。上述指标的计算公式见附录 § B.3。

4.3.1. 定量比较

表 1 展示了所有方法在 CAVE 和 PaviaC 数据集上的定量比较结果，涵盖分布内（即 ×4 训练尺度）和分布外（×8、×16和 ×32 尺度）两种设置。表中数据表明，我们的方法在所有分布内指标上均达到最先进水平。在分布外评估中，我们的模型在几乎所有分布外指标上都取得了最佳或次佳结果。尽管 DC-

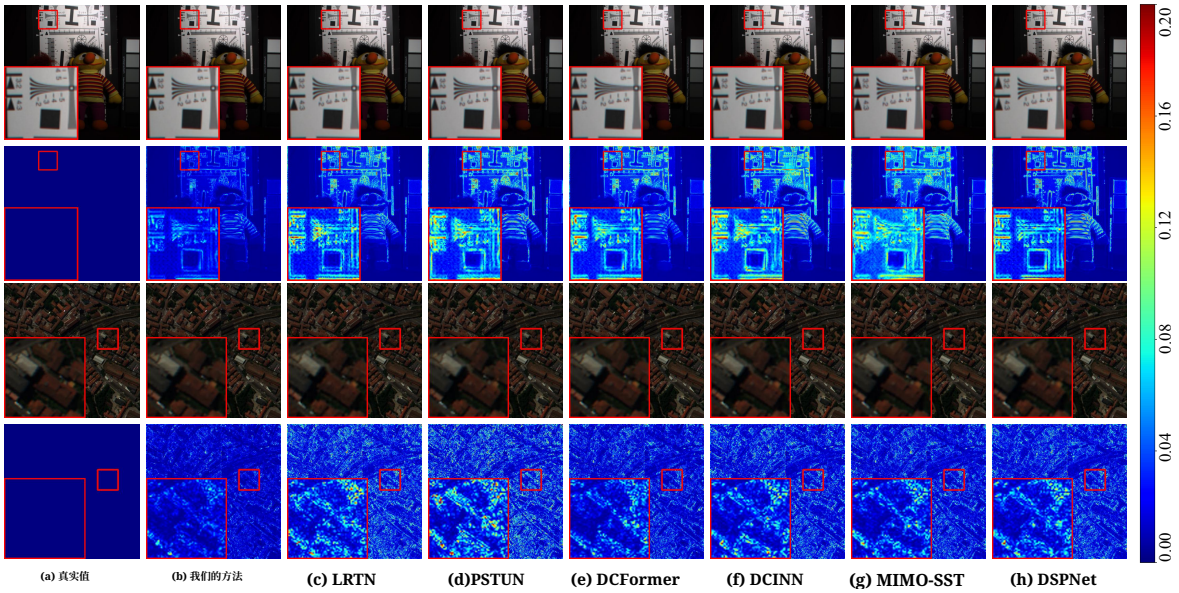

图5. 上半部分和下半部分分别展示了 CAVE 和 PaviaC 数据集的融合图像和误差图。红色矩形框标示了一些局部放大图。

表2. PaviaC 数据集上 MK 的消融研究（参见 §4.4.1）。

模型		峰值信噪比(↑)	光谱角映射(↓)	相对全局综合误差(↓)	结构相似性指数(↑)
×4	不含 MK	45.85 ± 1.80	**2.01 ± 0.29**	**1.06 ± 0.11**	0.995 ± 0.001
	Ours	**45.92 ± 1.75**	2.02 ± 0.30	1.07 ± 0.11	**0.996 ± 0.000**
×8	不含 MK	**38.62 ± 0.65**	4.75 ± 0.07	3.49 ± 0.06	0.959 ± 0.004
	Ours	38.59 ± 0.67	**4.74 ± 0.06**	**3.47 ± 0.05**	**0.960 ± 0.004**
×16	不含 MK	34.95 ± 0.03	6.92 ± 0.08	5.18 ± 0.07	**0.936 ± 0.002**
	Ours	**35.00 ± 0.02**	**6.89 ± 0.08**	**5.16 ± 0.07**	0.935 ± 0.002
×32	不含 MK	**32.81 ± 0.25**	8.51 ± 0.20	6.25 ± 0.29	0.921 ± 0.001
	Ours	32.79 ± 0.26	**8.49 ± 0.19**	**6.24 ± 0.30**	**0.922 ± 0.001**

Former（Wu 等人，2025）、DSPNet（Sun 等人，2023）和 PSTUN（Wang 等人，2025a）在某些特定情况下也取得了优异的次佳结果，但它们无法处理所有尺度，且性能表现不稳定。其余五个数据集的定量比较结果见附录 §B.6。为全面展示各方法在所有数据集上的表现，我们绘制了所有方法的峰值信噪比指标折线图（见图 4）。可以看出，我们的模型并未牺牲重建质量；其性能甚至超过了针对每个特定数据集分别训练的专用模型。我们的方法在分布内准确率方面表现最佳，并对未见尺度展现出稳健的泛化能力。

4.3.2. 定性比较

为了说明我们的方法所具备的优势，我们在图5中提供了与最先进方法的直观对比，其中包括局部放大图和误差图，以突出具体细节。从视觉效果来看，我们的 SSA 模型生成的结果明显更出色，也最忠实于真实值。从视觉效果来看，我们的 SSA 模型生成的结果最忠实于真实值。局部放大图显示，我们的方法能够有效还原细节，例如毛绒玩具的纹理和图表上的清晰线条，而其他方法往往存在模糊和伪影。同样，该方法也能更清晰地重建遥感场景中的建筑结构。误差图进一步印证了这一优势：我们的方法所得误差图整体呈深蓝色（表示残差极小），而最先进的方法在物体边界处则出现明显的高能量区域（黄色/红色），这证明了我们统一方法的稳健准确性。

4.4. 消融实验

4.4.1. 俄罗斯套娃核的必要性

为验证 MK 的作用，我们通过将 MKL 替换为标准固定通道卷积，构建了一个不含 MK 的变体，并在每个数据集上分别独立训练。

表 2 比较了这一单独训练的变体与我们的单个联合训练模型。结果表明，这两个模型在所有尺度上的性能都非常接近。虽然不含 MK 的变体偶尔表现略好，但我们的含 MK 统一模型通常能达到或超过其性能。这表明，我们的模型取得了与专用模型相当的性能。

表 3. **对未见波段与分数尺度的泛化能力**。上：在 ×4 尺度下对未见数据集的泛化能力，包括 Houston 和 Loukia。下：在 CAVE 上对未见尺度的泛化能力（以 ×4 比例训练）。DCFormer/PSTUN/LRTN 的结果是基于 ×4 比例，通过双三次插值缩放得到的。

模型	峰值信噪比(↑)	光谱角映射(↓)	相对全局综合误差(↓)	结构相似性指数(↑)
未见数据集 — *Houston*（*48* 个波段， ×4尺度）				
DCFormer	38.55	3.21	4.09	0.972
PSTUN	38.52	3.15	3.98	0.975
LRTN	38.65	3.12	3.91	**0.980**
Ours	**38.71**	**3.09**	**3.85**	0.978
未见数据集 — *Loukia*（*176* 波段， ×4尺度）				
DCFormer	36.25	**3.88**	3.65	0.958
PSTUN	35.80	4.12	3.82	0.952
LRTN	35.92	4.05	3.78	0.955
Ours	**36.42**	3.95	**3.55**	**0.962**
分数尺度 — *CAVE*（仅在 ×4 上训练） ×3.2尺度				
双三次插值	38.50	5.12	4.85	0.945
DCFormer	44.65	**2.45**	1.95	0.990
PSTUN	43.90	2.62	2.25	0.985
LRTN	44.35	2.58	2.08	0.988
我们的方法（直接）	**45.02**	2.48	**1.81**	**0.993**
分数尺度 — *CAVE*（仅在 ×4 ×5.7尺度上训练）				
双三次插值	32.20	6.85	6.10	0.895
DCFormer	42.95	3.02	2.45	0.984
PSTUN	41.25	3.25	2.72	0.980
LRTN	42.80	**2.88**	2.55	0.982
我们的方法（直接）	**43.49**	2.91	**2.17**	**0.991**

4.4.2. 未见波段与分数尺度

我们进一步评估了预训练模型在两种具有挑战性的情形下的泛化能力：具有不同光谱波段的未见数据集，以及未见分数尺度。

未见数据集（未见波段）。我们首先在两个光谱配置未在训练期间出现过的数据集上进行测试：包含 48 个波段的 Houston 和包含 176 个波段的 Loukia，两者均在 ×4 尺度下进行评估。如表 3所示，我们的方法仅经过 500 次迭代微调，便在各项指标上始终达到顶尖水平。在 Houston 数据集上，我们的模型取得了最佳的峰值信噪比/光谱角映射/相对全局综合误差（38.71 dB / 3.09 / 3.85），且结构相似性指数（0.978）也具有竞争力，展现出对不同波段配置的稳健泛化能力。在 Loukia 数据集上，我们的方法再次实现最佳重建质量，取得最高的峰值信噪比（36.42 dB）、最低的相对全局综合误差（3.55）和最高的结构相似性指数（0.962），同时光谱角映射也具有竞争力。这些结果表明，我们学习到的光谱表示能够很好地迁移至波段数量和分布各异的数据集。

未见过的分数尺度。我们进一步通过在 CAVE 上以分数尺度 ×3.2和 ×5.7进行测试来评估尺度泛化能力，而训练仅在整数尺度 ×4下进行。对于不支持原生连续缩放的方法，我们采用强基线方案：生成 ×4 输出，然后通过双三次插值将其调整至目标分数尺度（×4 →双三次插值）。表 3 显示，我们的方法可直接预测分数尺度，并在两个尺度下均取得最佳峰值信噪比（在 ×3.2 下为 45.02 dB，在×5.7下为 43.49 dB），同时获得最低的相对全局综合误差（1.81 和 2.17）以及最高的结构相似性指数（0.993 和 0.991）。总体而言，这些结果证实，我们的连续表示具备真正的尺度无关能力，能够有效泛化至训练期间未见过的离散尺度之外，并优于强劲的 ×4→双三次插值基线。

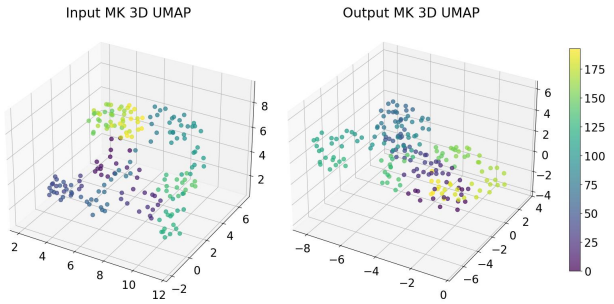

图 6. 统一流形逼近与投影（McInnes 等人，2018）对学习到的 MK 权重进行的可视化。这些权重形成了一个平滑流形，揭示了模型对光谱连续性的编码，而这正是模型具备泛化能力的关键。

4.4.3. 学习到的核结构分析

为了探究 MK 的内部工作机制，我们使用统一流形逼近与投影（McInnes 等人，2018）可视化了输入核和输出核的权重结构，如图 6 所示。可视化图中的每个点对应一个光谱通道，并根据其索引着色。如图所示，投影后的点形成了平滑且连续的流形，而非随机散布的点云。这表明核的学习过程形成了有意义的分层流形结构，网络能够主动捕捉相邻光谱波段之间的高度相关性。这种学习到的内部结构是模型能够在不同传感器之间有效泛化的关键。

4.5. 效率分析

我们对所有对比模型进行了效率分析，并针对模型缩放开展了消融实验，详情见附录 §B.4 和附录 §B.5。

5. 结论

在这项工作中，我们针对多光谱/高光谱融合模型普遍存在的刚性问题，提出了 SSA 这一通用框架。该框架包含俄罗斯套娃核（MK）和隐式神经表示（INR）骨干网络，克服了架构刚性、数据匮乏和空间尺度固定这三重困境。我们的实验表明，单个联合

训练的 SSA 模型即可达到最先进的性能，能够以零样本方式有效泛化至未见过的传感器和尺度，并学习有意义的内部表征。尽管我们承认 INR 的计算成本可能构成限制，但我们相信，该模型为开发高光谱分析领域的通用基础模型奠定了坚实基础。

影响声明

本文旨在推动机器学习领域的发展。我们的工作可能对社会产生诸多影响，但我们认为没有哪项影响需要在此特别强调。

参考文献

Amigo, J. M. 高光谱成像，第 32 卷。爱思唯尔，阿姆斯特丹，2019年。

Cao, Z., Cao, S., Deng, L.-J., Wu, X., Hou, J., and Vivone, G. 采用解耦调制的扩散模型，用于多光谱和高光谱图像锐化。*Inf.Fusion*，104:102–158，2024年。

Cao, Z.-H., Liang, Y.-J., Deng, L.-J., 和 Vivone, G. 一种利用统一语言和掩码引导的高效图像融合网络。*IEEETPAMI*，2025。Cappellazzo, U., Kim, M., Ma, P., Chen, H., Liu, X., Petridis, S., 和 Pantic, M. MoME：用于视听语音识别的套娃专家混合模型。*arXiv*预印本 *arXiv:2510.04136*，2025。

Chen, H., Zhao, W., Xu, T., Shi, G., Zhou, S., Liu, P., 和 Li, J. 用于高光谱图像重建的逐光谱隐式神经表示。*IEEE*电路与系统视频技术汇刊，34(5):3714–3727，2023。

Chen, L., Zhao, Y., Yao, J., Chen, J., Li, N., Chan, J. C.-W., 和 Kong, S. G. 利用快速空间-光谱特征对高光谱视频进行目标跟踪。*Remote Sens.(Basel)*，13(10):1922，2021a年。

Chen, Y.、Liu, S.与Wang, X. 基于局部隐式图像函数学习连续图像表示。载于计算机视觉与模式识别会议，第8628–8638页，2021b。

Cheng, M.-F.、Mukundan, A.、Karmakar, R.、Valappil, M. A. E.、Jouhar, J.与Wang, H.-C. 高光谱成像的现代趋势与最新应用：综述。《技术》，13(5):170，2025。ISSN 2227-7080。URL https://www.mdpi.com/2227-7080/13/5/170。

Deng, S., Deng, L.-J., Wu, X., Ran, R. 和 Wen, R. 用于多光谱与高光谱图像融合的双向膨胀率 Transformer。载于 *IJCAI*，2023a。

Deng, S., Ma, J., Deng, L.-J. 和 Wei, P. OTIAS：用于多光谱与高光谱图像融合的八叉树隐式自适应采样。载于*AAAI*，第 39 卷，第 2708–2716 页，2025。

Deng, S.-Q., Deng, L.-J., Wu, X., Ran, R., Hong, D. 和 Vivone, G. PSRT：用于多光谱与高光谱图像融合的金字塔式打乱与重排 Transformer。*IEEE*地球科学与遥感汇刊，61：1–15，2023b。Dian, R., Li, S., Guo, A. 和 Fang, L. 深度高光谱图像锐化。*IEEE*神经网络与学习系统汇刊，29(11)：5345–5355，2018。

Fang, J.、Yang, J.、Khader, A.与Xiao, L. MIMO-SST：面向高光谱与多光谱图像融合的多输入多输出空谱Transformer。*IEEE*地球科学与遥感汇刊，62:1–20，2024。

Fauvel, M.、Tarabalka, Y.、Benediktsson, J. A.、Chanussot, J.与Tilton, J. C. 高光谱图像空谱分类的进展。*IEEE*会刊，101(3):652–675，2013年。

Gonzalez, R., Albrecht, C. M., Braham, N.A.A., Lambhate, D., Almeida, J. L. d. S., Fraccaro, P., Blumenstiel, B., Brunschwiler, T.和Bangalore, R. 使用预训练基础模型从多光谱到高光谱。*arXiv*预印本*arXiv:2502.19451*，2025。

He, K., Zhang, X., Ren, S.和Sun, J. 面向图像识别的深度残差学习。载于计算机视觉与模式识别会议，第770–778页，2016。

He, L., Li, J., Liu, C.和Li, S. 光谱–空间高光谱图像分类的最新进展：综述与新指南。*IEEE*地球科学与遥感汇刊，56(3):1579–1597，2017。

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A.等. 训练计算最优大语言模型。*arXiv*预印本*arXiv:2203.15556*，2022。

Hojjat, A., Haberer, J., Pirk, S., 和 Landsiedel, O. ThinkingViT：用于弹性推理的嵌套思维视觉Transformer。载于 *ES-FoMo III*：第三届基础模型高效系统研讨会，2025。网址：https://openreview.net/forum?id=mFQRyJkFys。

Hong, D., Li, C., Yokoya, N., Zhang, B., Jia, X., Plaza, A., Gamba, P., Benediktsson, J. A., 和 Chanussot, J. 高光谱成像。*arXiv* 预印本 *arXiv:2508.08107*，2025。

Hu, J., Huang, T., Deng, L., Dou, H., Hong, D. 和 Vivone, G. Fusformer：一种用于

高光谱图像超分辨率的基于Transformer的融合网络。*IEEE Geosci. Re-moteSens. Lett.*，19：1–5，2022。

Huang, B.、Song, H.、Cui, H.、Peng, J. 和 Xu, Z. 使用稀疏矩阵分解进行空间与光谱图像融合。*IEEE*地球科学与遥感汇刊，52(3)：1693–1704，2013年。

Kawakami, R.、Matsushita, Y.、Wright, J.、Ben-Ezra, M.、Tai, Y.-W. 和 Ikeuchi, K. 通过矩阵分解实现高分辨率高光谱成像。载于计算机视觉与模式识别会议，第2329–2336页。I EEE，2011年。

Kusupati, A.、Bhatt, G.、Rege, A.、Wallingford, M.、Sinha, A.、Ramanujan, V.、Howard-Snyder, W.、Chen, K.、Kakade, S.、Jain, P. 等。套娃表示学习。*NeurIPS*，35：30233–30249，2022年。

Liang, Y., Cao, Z., Deng, S., Dou, H.-X., and Deng, L.- J. 面向多光谱与高光谱图像融合的傅里叶增强隐式神经融合网络。*NeurIPS*, 37:63441–63465, 2024.

Lim, B., Son, S., Kim, H., Nah, S., andLee, K.M. 用于单幅图像超分辨率的增强型深度残差网络。发表于计算机视觉与模式识别会议研讨会, pp. 136–144, 2017.

Lin, B., Zeng, Z., Xiao, Z., Kou, S., Hou, T., Gao, X., Zhang, H., and Deng, Z. Matryoshkakv：通过可训练正交投影实现自适应KV压缩。*arXiv*预印本 *arXiv:2410.14731*, 2024.

Liu, Y., Dian, R., and Li, S. 用于高分辨率高光谱计算成像的低秩Transformer。*IJCV*, 133(2):809–824, 2025.

Loshchilov, I. 和 Hutter, F. 解耦权重衰减正则化。*arXiv* 预印本 *arXiv:1711.05101*，2017。

Masi, G.、Cozzolino, D.、Verdoliva, L. 和 Scarpa, G. 基于卷积神经网络的全色锐化。*RemoteSens.(Basel)*，8(7):594，2016。

McInnes, L.、Healy, J. 和 Melville, J. UMAP：用于降维的统一流形逼近与投影。*arXiv* 预印本 *arXiv:1802.03426*，2018。

Nezhad, Z. H., Karami, A., Heylen, R., and Scheunders, P. 使用光谱解混和稀疏编码融合高光谱图像与多光谱图像。*I EEE J. Sel. Top.Appl. Earth Obs. Remote Sens.*, 9(6):2377–2389, 2016. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. PyTorch：一种命令式、高性能深度学习库。载于*NeurIPS*，第 8024–8035 页。Curran Associates, Inc., 2019年。

Shang, Y., Liu, J., Zhang, J., and Wu, Z. MFT-GAN：一种用于无监督高光谱全色锐化的多尺度特征引导 Transformer 网络。*IEEETrans.Geosci. Remote Sens.*, 62:1–16, 2024. Sun, Y., Xu, H., Ma, Y., Wu, M., Mei, X., Huang, J., and Ma, J. 一种用于高光谱图像融合的 Transformer 双空间–光谱金字塔网络。*IEEE Trans.Geosci. Remote Sens.*, 61:1–16, 2023。

Suryanarayana, G.、Ramtej, K. S.、Reddy, D. S.、Prasad, P. K.、Prasad, A. 和 Srikanth, K. 基于深度插值的高光谱-多光谱图像融合：通过各向异性依赖主成分分析实现。*Mul-timedia Tools Appl.*，84(4):1649–1669，2025。

Uzair, M., Mahmood, A., and Mian, A. S. 使用三维离散余弦变换和偏最小二乘法进行高光谱人脸识别。载于*BMVC*，第1卷，第10页，2013年。

Wald, L. 高分辨率合成图像的质量：是否存在简单的判据? 载于*Proc. Int. Conf. Fusion EarthData*，第99–103页，2000年。

Wang, B., Xiong, X., Lian, Y., Cao, X., Zhou, H., Yu, K., and Liu, Z. 采用多尺度混合训练的感知光谱变换器展开网络，用于任意尺度高光谱与多光谱图像融合。*Inf.Fusion*，122:103166，2025a。

Wang, T., Yan, Z., Li, J., Zhao, X., Wang, C., and Ng, M. 通过自监督表征实现任意分辨率的高光谱与多光谱图像融合。*IJCV*，第1–21页，2025b。

Wang, W.、Deng, L.-J.、Ran, R. 与 Vivone, G. 一种面向图像融合的细节保持条件可逆网络通用范式。*IJCV*, 132(4):1029–1054, 2024a.

Wang, Y.、Yue, Z.、Zeng, H.、Wang, D. 与 McAuley, J. 一次训练，随处部署：面向多模态推荐的套娃表示学习。*arXiv*预印本 *arXiv:2409.16627*, 2024b.

Wu, X.、Cao, Z.-H.、Huang, T.-Z.、Deng, L.-J.、Chanussot, J. 与 Vivone, G. 面向多源图像融合的全连接 Transformer。*IEEE TPAMI*, 47(3):2071–2088, 2025.

Yan, H.-F.、Zhao, Y.-Q.、Chan, J. C.-W.、Kong, S. G.、EI-Bendary, N. 与 Reda, M. 高光谱与多光谱图像融合：当模型驱动遇见数据驱动策略。*Inf. Fusion*, 116:102803, 2025.

Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., 和 Paisley, J. PanNet：一种用于全色锐化的深度网络架构。发表于*I CCV*，第 5449–5457 页，2017。

Zhang, K., Zhu, D., Min, X., 和 Zhai, G. 用于高光谱图像超分辨率的隐式神经表示学习。*IEEE*地球科学与遥感汇刊，61：1–12，2022。

Zhou, H.、Liu, Q.、Weng, D.与Wang, Y. 无监督循环一致生成对抗网络用于全色锐化。*IEEE*地球科学与遥感汇刊，60:1–14，2022。

A. 套娃核层算法

我们提出的套娃核层（MKL）的伪代码详见§3.2，并在算法1 和算法2中给出。在我们的SSA 模型中，套娃核输入层和输出层均使用一个 3×3 卷积核。步幅、膨胀率和分组数均设为1。因此，输入与输出特征图的空间维度在MKL运算过程中保持不变。

算法 1 套娃核输入层前向传播

要求：
输入 $\mathbf{X} \in \mathbb{R}^{N \times C_{\text{in}} \times H \times W}$
输入 嵌套卷积权重$\mathbf{W}_{\text{nested}} \in \mathbb{R}^{D \times C_{\max} \times k \times k}$
输入 卷积偏置$\mathbf{b} \in \mathbb{R}^{D}$（可选）
输入 卷积参数：步幅 $\mathbf{s}$，填充 $\mathbf{p}$，膨胀率 $\mathbf{d}$，分组数 $\mathbf{g}$
确保：
输出 $\mathbf{Y} \in \mathbb{R}^{N \times D \times H' \times W'}$

函数 Forward($\mathbf{X}$,$\mathbf{W}_{\text{nested}}$,$\mathbf{b}$)
  if $C_{\text{in}} = C_{\max}$ **then**
    $\mathbf{Y} \leftarrow$ 二维卷积($\mathbf{X}$, $\mathbf{W}_{\text{nested}}$, $\mathbf{b}$,$\mathbf{s}$, $\mathbf{p}$, $\mathbf{d}$,$\mathbf{g}$) {使用完整权重}
  else
    $\mathbf{W}_{\text{valid}} \leftarrow \mathbf{W}_{\text{nested}}[:, : C_{\text{in}}, :, :]$ {针对输入通道 C_{in}} 切片权重
    $\mathbf{Y} \leftarrow$ 二维卷积($\mathbf{X}$, $\mathbf{W}_{\text{valid}}$, $\mathbf{b}$,$\mathbf{s}$, $\mathbf{p}$, $\mathbf{d}$,$\mathbf{g}$)
  结束如果
  返回 Y
结束函数

算法 2 套娃核输出层前向传播

要求：输入 Y $\in \mathbb{R}^{N \times D \times H' \times W'}$**输入目标输出通道** C_{out}**输入预定义卷积权重 W**$_{\text{nested}} \in \mathbb{R}^{C_{\max} \times D \times k \times k}$**输入预定义卷积偏置 b**$_{\text{nested}} \in \mathbb{R}^{C_{\max}}$ **（可选）输入卷积参数：步幅 s、填充 p、膨胀率 d、分组数 g 确保：输出 X** $\in \mathbb{R}^{N \times C_{\text{out}} \times H \times W}$**1：函数 Forward(Y** C_{out}**, W**$_{\text{nested}}$**, b**$_{\text{nested}}$**) 2：if** $C_{\text{out}} = C_{\max}$ **then 3：X** $\leftarrow$ **二维卷积(Y, W**$_{\text{nested}}$**, b**$_{\text{nested}}$**, s, p, d, g) 4：else 5：W**$_{\text{valid}} \leftarrow$ **W**$_{\text{nested}}$**[：** C_{out}**, :,:, :]** {**为输出通道切片权重** C_{out}}**6：b**$_{\text{valid}} \leftarrow$**None 7：if b**$_{\text{nested}} \neq$ **None then 8：b**$_{\text{valid}} \leftarrow$ **b**$_{\text{nested}}$**[：** C_{out}] {**切片偏置**}**9：end if 10：X** $\leftarrow$ **二维卷积(Y, W**$_{\text{valid}}$**, b**$_{\text{valid}}$**, s, p, d, g) 11：end if 12：返回 X 13：end function**

B. 附加实验详情

B.1. 数据集信息

本节介绍所用数据集的信息，详见表 B.1。

表 *B.1.* 实验数据集的详细元信息和训练/测试划分。

属性	CAVE	Harvard	PaviaC	PaviaU	Chikusei	博茨瓦纳	华盛顿特区
波段	31	31	102	103	128	145	191
像素分辨率 (m)	0.001	0.001	1.3	1.3	2.5	30	2.5
光谱范围 (nm)	400–700	420–720	430–860	430–860	363–1018	400–2500	400–2500
传感器	制冷型电荷耦合器件	Nuance FX	ROSIS	ROSIS	Headwall Hyperspec-VNIR-C	Hyperion	Hydice
Size	512×512	1040×1392	1096×715	610×340	2517×2335	1476×256	1208×307
训练/测试图像块尺寸	128×128/ 512×512	128×128/ 1024×1024	128×128/ 256×256	128×128/ 512×512	128×128/ 1024×1024	128×128/ 256×256	128×128/ 256×256
训练/测试划分	3718/10	5360/10	2257/2	434/2	5175/4	765/5	876/5

B.2. 数据模拟

为全面评估所提方法的有效性，我们在七个公开高光谱数据集上进行了实验，其中包括两个室内/室外数据集：CAVE[1] 和 Harvard[2], ，以及五个遥感数据集：PaviaC[3], PaviaU[4], Chikusei[5], 博茨瓦纳[6], 和华盛顿特区[7]。所有数据模拟流程均严格遵循 Wald协议（Wald，2000）。各数据集的详细参数汇总于表B.1。相应的测试集由互不重叠的图像块构成，具体而言，Chikusei 使用四个 512×512 图像块，其他四个遥感数据集则各使用 256×256 个图像块。对于所有训练集，我们以固定步幅生成 128×128 图像块。低分辨率高光谱图像（LR-HSIs）通过使用抗混叠滤波器将原始高光谱图像下采样 s 倍合成，而相应的高分辨率多光谱图像（HR-MSIs）则利用 ROSIS 和 Landsat-8 等传感器的光谱响应函数进行模拟。

B.3. 质量指标

为了定量评估高光谱图像重建的性能，我们采用了四种广泛使用的质量指标：峰值信噪比、光谱角映射、结构相似性指数和相对全局综合误差。这些指标共同评估空间保真度、光谱准确度、结构相似性和全局重建一致性。

峰值信噪比（峰值信噪比）

峰值信噪比是一种空间指标，用于衡量信号的最大可能功率与干扰其表示的噪声功率之比。它为评估重建高光谱图像的空间质量提供了有效方法。对于$\widehat{}$重建图像 $\mathcal{X}_{HR}$ 和真实值 $\mathcal{X}_{HR}$，峰值信噪比定义如下：

$$PSNR(\widehat{\mathcal{X}}_{HR}, \mathcal{X}_{HR}) = 20 \log \left(\frac{255}{RMSE(\widehat{\mathcal{X}}_{HR}, \mathcal{X}_{HR})} \right),$$

其中，RMSE 表示均方根误差：

$$RMSE(\widehat{\mathcal{X}}_{HR}, \mathcal{X}_{HR}) = \frac{\|\widehat{\mathcal{X}}_{HR} - \mathcal{X}_{HR}\|_F}{\sqrt{b \cdot m_l}}, \tag{11}$$

其中，b 是光谱波段数，m_l 是空间像素数。PSNR 值越高，表示重建越准确。

[1]https://cave.cs.columbia.edu/repository/Multispectral[2]
http://vision.seas.harvard.edu/hyperspec/index.html[3]
https://www.ehu.eus/ccwintco/index.php/Hyperspectral_遥感_场景#Pavia_Centre 场景__
https://www.ehu.eus/ccwintco/index.php/Hyperspectral[4]遥感_场景#Pavia_University_场景_
https://www.sal.t.u-tokyo.ac.jp/hyperdata/_https://www.ehu.eus/ccwintco/index.php/Hyperspectral[5]遥感[6]场景
#Botswana_https://engineering.purdue.edu/_biehl/MultiSpec/hyperspectral.html

光谱角映射器（SAM）

光谱角映射是一种光谱度量，用于计算两个光谱向量之间的夹角。它常用于评估光谱保真度，确保重建过程中光谱形状得到良好保留。光谱角映射的公式如下：

$$SAM(\widehat{\mathcal{X}}_{HR}, \mathcal{X}_{HR}) = \frac{1}{b}\sum_{j=1}^{b} \arccos\left(\frac{\langle \widehat{x}_j, x_j \rangle}{\|\widehat{x}_j\|\|x_j\|}\right),$$

其中 $\widehat{x}_j$ 和 x_j 分别表示第 j个波段的重建光谱向量与真实值光谱向量。

结构相似性指数（SSIM）

SSIM用于评估重建图像与真实值之间的结构相似性。它综合考虑亮度、对比度和结构信息，并被广泛用于感知图像质量评估。对于每个波段，SSIM定义为：

$$SSIM(\widehat{x}, x) = \frac{(2\mu_{\widehat{x}}\mu_x + C_1)(2\sigma_{\widehat{x}x} + C_2)}{(\mu_{\widehat{x}}^2 + \mu_x^2 + C_1)(\sigma_{\widehat{x}}^2 + \sigma_x^2 + C_2)},$$

其中 μ 和 σ 分别表示均值和方差， $\sigma_{\widehat{x}x}$ 为协方差， C_1、 C_2 为防止不稳定的常数。SSIM值越高（最大为1），表示结构相似性越好。

合成相对全局无量纲误差（ERGAS）

ERGAS 是衡量所有光谱波段相对重建误差的全局指标。ERGAS 值越低，整体重建性能越好。ERGAS 定义如下：

$$ERGAS(\widehat{\mathcal{X}}_{HR}, \mathcal{X}_{HR}) = 100 \cdot \frac{1}{s}\sqrt{\frac{1}{b}\sum_{j=1}^{b}\left(\frac{RMSE_j}{\mu_j}\right)^2},$$

^其中 s 是尺度因子， $RMSE_j$ 是 11 式中 与 $\mathcal{X}_{HR}$之间各波段均方根误差（RMSE）的均值，而 μ_j 是真实值波段的均值。

B.4. 效率分析

表 B.2 总结了代表性融合方法的模型复杂度和实际推理效率。对于数据驱动的深度学习高光谱融合方法，为了有效利用大规模异构训练数据并提升泛化能力，增加模型容量往往是自然而然的，有时也是必要的。因此，我们有意扩大网络规模，以增强其拟合能力。即便如此，我们的模型参数量仍少于 PSTUN（Wang 等人，2025a），同时实现了更高的重建精度。

从复杂度角度来看，表B.2 与表1共同揭示了一种一致趋势：计算预算较大的方法通常能在分布内和分布外实现更高的重建质量，而极为轻量级的模型（例如 DCFormer（Wu 等人，2025））则往往因容量有限而出现明显的质量下降。

值得注意的是，容量的提升并未导致运行时间高到难以接受。尽管我们的方法理论浮点运算次数较高，但在现代图形处理器上实际运行仍然高效：在 $\times 4$ 和 512×512 输入下，我们的模型每张图像的运行时间为 83.4 ms（3.1 MPix/s），峰值内存占用为 2.46 GB。其运行时间与若干容量相近或更小的强基线方法（例如 PSTUN 和 LRTN）相当（甚至更快），同时比基于查询的基线方法（例如 DCINN（Wang 等人，2024a））快数个数量级。总体而言，这些结果表明，扩大模型规模是提升准确性和鲁棒性的有效方式，同时不会牺牲实际推理效率。

B.5. 模型缩放消融实验

为系统地研究模型容量如何影响最终重建质量，我们通过逐步增加空间/光谱编码器的深度进行消融分析。这些结构调整使参数总量从 2.735M 增至 18.65M，浮点运算次数从 58.873G 增至 0.202T。

表 *B.2.* RTX 4090 上的模型复杂度和推理效率。在 $\times 4$ 下，使用批量大小 1 和输入尺寸 512×512 ($C = 191$) 测量时间/显存/百万像素每秒。每种方法使用其验证前向入口，预热运行 20 次后，报告 50 次迭代的平均运行时间。支持时启用自动混合精度（半精度浮点数）；否则使用单精度浮点数。峰值显存是推理期间分配的最大 cuda.max 内存。吞吐量按 $(HW/10^6)/$（时间，单位为秒）计算。

方法	参数量（M）↓	浮点运算次数（G）↓	时间（毫秒）↓	峰值显存（GB）↓	吞吐量（百万像素/秒）↑
DSPNet（Sun 等人，2023）	6.056	27.279	38.7	1.52	6.8
MIMO-SST（Fang 等人，2024）	4.983	6.164	14.4	1.22	18.2
DCINN (Wang 等人，2024a)	2.417	51.531	860.9	2.25	0.3
DCFormer(Wu 等人，2025)	0.082	0.579	40.2	1.00	6.5
PSTUN (Wang 等人，2025a)	29.948	102.400	102.4	1.58	2.6
LRTN (Liu 等人，2025)	3.691	8.396	153.7	3.32	1.7
Ours	18.654	206.848	83.4	2.46	3.1

表 *B.3.* 模型缩放消融实验。 E_{spe} 和 E_{spa} 分别表示空间编码器和光谱编码器的深度。

(E_{spe}, E_{spa})	参数量（M）	浮点运算次数	峰值信噪比(↑)	光谱角映射(↓)	相对全局综合误差(↓)	结构相似性指数(↑)
(1, 1)	2.735	58.873G	$42.45_{\pm 3.24}$	$5.32_{\pm 1.20}$	$3.92_{\pm 2.05}$	$0.985_{\pm 0.004}$
(2, 2)	9.213	0.120T	$44.15_{\pm 3.12}$	$3.94_{\pm 1.05}$	$2.88_{\pm 1.50}$	$0.988_{\pm 0.003}$
(4, 4)	13.934	0.161T	$44.72_{\pm 3.04}$	$3.05_{\pm 0.95}$	$2.11_{\pm 1.28}$	$0.991_{\pm 0.002}$
(6, 6)	18.654	0.202T	$45.96_{\pm 4.69}$	$2.55_{\pm 1.00}$	$1.72_{\pm 1.13}$	$0.994_{\pm 0.005}$

如表 B.3明确所示，采用更深的编码器能够持续提升所有指标上的重建精度；例如，从最小配置到最大配置，PSNR 提高了 3.5 dB 以上。这一明显趋势表明，更大的模型具有显著更强的空谱表征能力，能够更好地捕捉细粒度特征，并拟合高光谱数据复杂且高维的分布。尽管模型尺寸和计算成本也会随之增加，但这种扩展在意料之中；正如我们的结果所表明的，在处理大规模异构高光谱信号时，数据驱动融合框架往往需要进行此类扩展，才能达到最先进的性能。

B.6. 更多主要结果

为全面评估我们所提出的 SSA 框架的通用性与鲁棒性，我们将实验扩展至另外五个基准：Harvard、PaviaU、Chikusei、博茨瓦纳和华盛顿特区。这些数据集涵盖了多种多样的传感器特性，光谱波段数量从 31 到 191 不等。如表 B.4 至 B.8所示，我们的统一模型在这些多样化场景中始终取得最先进的性能。值得注意的是，尽管我们的通用模型是在异构数据集上联合训练的，其性能仍优于 LRTN（Liu 等人，2025）和 PSTUN（Wang 等人，2025a）等针对每个特定数据集单独训练的专用方法。例如，在同分布设置（$\times 4$ 尺度）下的 WashingtonDC 数据集（表 B.8）上，我们的方法在峰值信噪比上比次优方法高出约 1.0 dB。这一结果有力地验证了我们的套娃核设计能够有效缓解通常与多数据集训练相关的干扰，使单一网络无需修改架构即可适应不同的光谱分布。

我们进一步通过图B.1 至B.6中的全面视觉对比，佐证这些定量结果。为明确展示我们方法的尺度泛化能力，我们将另外五个数据集上的可视化结果（图B.3–B.6）并列呈现训练尺度（$\times 4$，上方两行）和极端未见尺度（$\times 32$，下方两行）的结果。对于 CAVE 和 PaviaC 数据集，我们特别展示了图 $\times 32$ B.1中的高难度结果。可以看到，尽管大多数具有竞争力的方法在 $\times 4$ 尺度下仍能保持合理的质量，但在 $\times 32$尺度下性能显著下降。采用固定上采样模块的方法（例如 DCFormer（Wu 等人，2025））在这一未见分辨率下完全无法生成有效结果（以打叉的方框标示），而其他方法则会出现严重模糊或网格状伪影。相比之下，得益于 INR 骨干网络的连续建模能力，SSA 在所有尺度下都能保持较高的重建保真度，误差图也大多呈深蓝色。值得注意的是，在 Harvard 数据集（图B.2）中，即使在 $\times 32$ 尺度下，我们的模型也能成功恢复清晰可辨的文本（例如“30th EDITION”），而其他方法则受到严重失真影响。

表 *B.4.* 哈佛数据集在四种尺度下的定量比较。单个单元格中的值为均值 ± 标准差。

	指标		哈佛数据集						
			DSPNet	MIMO-SST	DCINN	DCFormer	PSTUN	LRTN	Ours
In-Dist.	×4	峰值信噪比(↑)	43.97 ± 5.99	43.65 ± 6.85	42.06 ± 4.55	42.87 ± 5.30	44.30 ± 4.59	41.91 ± 5.33	44.81 ± 5.21
		光谱角映射(↓)	2.70 ± 0.83	2.98 ± 0.86	2.96 ± 0.85	3.19 ± 0.91	3.59 ± 0.98	4.58 ± 1.54	2.73 ± 0.89
		相对全局综合误差(↓)	4.30 ± 2.29	5.11 ± 2.08	5.46 ± 0.90	5.12 ± 2.46	4.54 ± 1.80	5.78 ± 1.38	4.41 ± 2.63
		结构相似性指数(↑)	0.983 ± 0.001	0.983 ± 0.007	0.983 ± 0.003	0.982 ± 0.009	0.983 ± 0.008	0.975 ± 0.008	0.984 ± 0.010
Out-of-Dist.	×8	峰值信噪比(↑)	42.25 ± 6.20	38.62 ± 6.74	40.61 ± 6.23	42.37 ± 5.25	41.79 ± 4.04	41.33 ± 5.63	42.73 ± 5.09
		光谱角映射(↓)	2.99 ± 0.90	3.26 ± 0.98	3.33 ± 0.90	3.51 ± 0.93	4.00 ± 0.99	4.82 ± 1.63	3.02 ± 0.92
		相对全局综合误差(↓)	5.19 ± 2.69	6.68 ± 2.43	5.69 ± 2.51	5.30 ± 2.15	5.62 ± 1.84	5.98 ± 1.23	4.92 ± 2.08
		结构相似性指数(↑)	0.976 ± 0.012	0.944 ± 0.047	0.974 ± 0.012	0.980 ± 0.012	0.977 ± 0.008	0.972 ± 0.013	0.978 ± 0.013
	×16	峰值信噪比(↑)	37.31 ± 6.99	34.36 ± 7.28	35.94 ± 6.79	_	37.46 ± 5.23	38.10 ± 6.48	38.73 ± 6.00
		光谱角映射(↓)	3.60 ± 1.15	4.01 ± 1.25	3.95 ± 1.11	_	4.89 ± 1.24	5.05 ± 1.56	3.85 ± 1.00
		相对全局综合误差(↓)	7.53 ± 4.48	10.74 ± 4.12	8.90 ± 2.98	_	9.40 ± 5.05	8.56 ± 4.39	7.68 ± 5.21
		结构相似性指数(↑)	0.956 ± 0.031	0.903 ± 0.069	0.950 ± 0.024	_	0.951 ± 0.036	0.952 ± 0.036	0.960 ± 0.037
	×32	峰值信噪比(↑)	36.73 ± 6.72	31.39 ± 7.65	31.78 ± 7.14	_	35.78 ± 4.84	36.97 ± 6.37	37.34 ± 6.49
		光谱角映射(↓)	4.28 ± 1.38	5.63 ± 1.56	4.99 ± 1.27	_	5.94 ± 1.81	5.68 ± 1.91	4.66 ± 1.57
		相对全局综合误差(↓)	9.21 ± 5.44	13.41 ± 5.37	12.46 ± 4.68	_	11.57 ± 5.72	9.50 ± 4.14	8.70 ± 4.79
		结构相似性指数(↑)	0.942 ± 0.044	0.863 ± 0.090	0.902 ± 0.053	_	0.941 ± 0.033	0.945 ± 0.038	0.949 ± 0.045

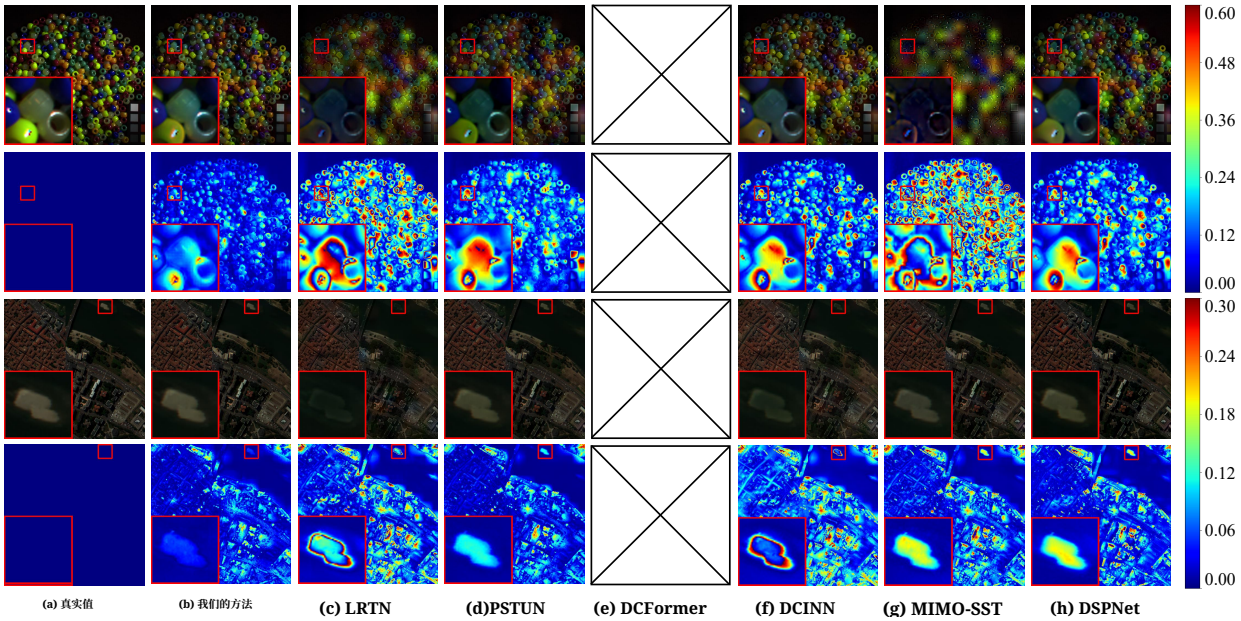

图 *B.1*。 在未见过的 ×32 **尺度**下，对 **CAVE**（上方各行）和 **PaviaC**（下方各行）数据集进行定性比较。红色方框突出显示了细节放大图，偶数行展示了相应的误差图。与在这种极端尺度下出现明显模糊的最先进方法相比，我们的 SSA 模型具有更出色的细节恢复能力和更小的重建误差。

表 *B.5.* 四种尺度下 **PaviaU** 数据集上的定量比较。数值为均值 ± 标准差。

		指标	PaviaU 数据集						
			DSPNet	MIMO-SST	DCINN	DCFormer	PSTUN	LRTN	Ours
In-Dist.	×4	峰值信噪比(↑)	37.51 ± 0.15	40.28 ± 0.15	40.31 ± 0.42	39.05 ± 0.25	39.46 ± 0.17	40.09 ± 0.63	40.44 ± 0.15
		光谱角映射(↓)	2.27 ± 0.12	3.46 ± 0.11	1.96 ± 0.07	2.60 ± 0.17	2.54 ± 0.02	2.05 ± 0.14	1.95 ± 0.07
		相对全局综合误差(↓)	1.99 ± 0.07	1.85 ± 0.11	1.53 ± 0.02	1.77 ± 0.07	1.58 ± 0.02	1.55 ± 0.00	1.56 ± 0.06
		结构相似性指数(↑)	0.981 ± 0.001	0.977 ± 0.002	0.988 ± 0.001	0.977 ± 0.000	0.981 ± 0.000	0.987 ± 0.001	0.988 ± 0.001
Out-of-Dist.	×8	峰值信噪比(↑)	35.76 ± 0.50	31.43 ± 0.55	28.91 ± 0.44	30.56 ± 1.74	36.01 ± 0.26	36.18 ± 0.03	36.68 ± 0.26
		光谱角映射(↓)	5.57 ± 0.06	5.83 ± 0.17	5.36 ± 0.26	5.39 ± 0.24	4.79 ± 0.27	5.16 ± 0.20	4.97 ± 0.30
		相对全局综合误差(↓)	4.64 ± 0.01	4.47 ± 0.12	4.89 ± 0.19	4.06 ± 0.56	3.50 ± 0.15	4.01 ± 0.14	3.03 ± 0.06
		结构相似性指数(↑)	0.917 ± 0.001	0.895 ± 0.001	0.884 ± 0.003	0.927 ± 0.004	0.959 ± 0.002	0.940 ± 0.001	0.960 ± 0.002
	×16	峰值信噪比(↑)	32.76 ± 0.79	27.29 ± 1.12	26.47 ± 0.56	–	33.66 ± 0.49	33.26 ± 0.00	35.48 ± 0.16
		光谱角映射(↓)	7.63 ± 0.31	9.18 ± 0.16	8.37 ± 0.34	–	6.35 ± 0.14	6.38 ± 0.23	6.26 ± 0.34
		相对全局综合误差(↓)	5.72 ± 0.47	7.01 ± 0.78	6.51 ± 0.25	–	5.62 ± 0.01	5.03 ± 0.15	4.40 ± 0.06
		结构相似性指数(↑)	0.873 ± 0.004	0.844 ± 0.005	0.833 ± 0.005	–	0.840 ± 0.003	0.879 ± 0.003	0.912 ± 0.000
	×32	峰值信噪比(↑)	30.50 ± 1.66	24.39 ± 1.25	24.25 ± 0.21	–	31.49 ± 0.64	30.44 ± 0.18	34.35 ± 0.14
		光谱角映射(↓)	10.50 ± 0.16	12.63 ± 0.87	11.47 ± 0.49	–	9.17 ± 0.08	8.81 ± 0.06	8.43 ± 0.04
		相对全局综合误差(↓)	7.90 ± 1.16	9.98 ± 1.69	8.38 ± 0.10	–	6.60 ± 0.21	6.02 ± 0.47	5.74 ± 0.13
		结构相似性指数(↑)	0.849 ± 0.014	0.802 ± 0.018	0.809 ± 0.003	–	0.829 ± 0.005	0.872 ± 0.006	0.898 ± 0.000

表 *B.6.* 在**Chikusei** 数据集上对四种尺度进行定量比较。数值为均值 ± 标准差。

		指标	Chikusei 数据集						
			DSPNet	MIMO-SST	DCINN	DCFormer	PSTUN	LRTN	Ours
In-Dist.	×4	峰值信噪比(↑)	37.21 ± 3.89	36.54 ± 4.12	38.05 ± 3.76	36.92 ± 3.55	38.62 ± 4.26	37.88 ± 3.91	39.35 ± 4.59
		光谱角映射(↓)	2.87 ± 0.62	3.05 ± 0.58	2.53 ± 0.47	2.91 ± 0.54	2.22 ± 0.50	2.18 ± 0.49	2.12 ± 0.46
		相对全局综合误差(↓)	5.28 ± 1.47	5.63 ± 1.52	4.89 ± 1.39	5.35 ± 1.41	4.54 ± 1.35	4.97 ± 1.43	4.50 ± 1.42
		结构相似性指数(↑)	0.951 ± 0.018	0.947 ± 0.021	0.958 ± 0.016	0.949 ± 0.019	0.965 ± 0.014	0.976 ± 0.015	0.972 ± 0.016
Out-of-Dist.	×8	峰值信噪比(↑)	31.05 ± 2.67	30.42 ± 2.79	31.58 ± 2.53	32.90 ± 2.81	30.87 ± 2.48	31.22 ± 2.64	33.76 ± 1.92
		光谱角映射(↓)	4.68 ± 0.91	4.92 ± 0.87	3.85 ± 0.83	3.87 ± 0.84	4.75 ± 0.89	3.21 ± 0.86	3.25 ± 0.78
		相对全局综合误差(↓)	7.65 ± 1.02	7.98 ± 0.97	7.32 ± 0.99	6.89 ± 0.94	7.72 ± 1.01	7.41 ± 0.96	6.23 ± 0.51
		结构相似性指数(↑)	0.862 ± 0.024	0.855 ± 0.027	0.871 ± 0.023	0.892 ± 0.022	0.888 ± 0.021	0.867 ± 0.024	0.905 ± 0.015
	×16	峰值信噪比(↑)	28.14 ± 2.19	27.56 ± 2.33	28.62 ± 2.08	–	29.29 ± 2.31	28.45 ± 2.15	30.16 ± 2.94
		光谱角映射(↓)	7.24 ± 1.38	7.59 ± 1.42	6.93 ± 1.31	–	6.56 ± 1.25	7.08 ± 1.34	5.42 ± 1.09
		相对全局综合误差(↓)	10.87 ± 0.89	11.23 ± 0.95	10.54 ± 0.92	–	10.09 ± 0.94	10.62 ± 0.91	8.70 ± 1.07
		结构相似性指数(↑)	0.795 ± 0.034	0.788 ± 0.037	0.802 ± 0.032	–	0.818 ± 0.031	0.836 ± 0.031	0.832 ± 0.033
	×32	峰值信噪比(↑)	26.37 ± 1.05	25.89 ± 1.18	26.74 ± 1.02	–	27.49 ± 0.99	26.58 ± 1.07	28.46 ± 1.61
		光谱角映射(↓)	9.15 ± 1.62	9.53 ± 1.71	8.87 ± 1.58	–	8.49 ± 1.55	8.96 ± 1.60	6.64 ± 1.23
		相对全局综合误差(↓)	12.68 ± 0.81	13.05 ± 0.87	12.34 ± 0.79	–	11.90 ± 0.75	12.47 ± 0.82	9.92 ± 0.25
		结构相似性指数(↑)	0.761 ± 0.011	0.753 ± 0.014	0.768 ± 0.010	–	0.783 ± 0.007	0.765 ± 0.012	0.803 ± 0.018

表 *B.7*。 在四种尺度下对<stayle id='3'>博茨瓦纳 数据集进行定量比较。数值为均值 ± 标准差。

		指标	博茨瓦纳数据集						
			DSPNet	MIMO-SST	DCINN	DCFormer	PSTUN	LRTN	Ours
In-Dist.	×4	峰值信噪比(↑)	42.60 ± 2.32	42.97 ± 3.98	45.16 ± 3.42	41.35 ± 2.81	44.94 ± 3.98	45.23 ± 4.47	45.31 ± 4.30
		光谱角映射(↓)	1.42 ± 0.21	1.28 ± 0.19	0.99 ± 0.11	1.47 ± 0.26	1.01 ± 0.11	0.97 ± 0.10	0.80 ± 0.04
		相对全局综合误差(↓)	2.01 ± 0.51	1.91 ± 0.62	1.61 ± 0.49	2.14 ± 0.54	1.75 ± 0.58	1.72 ± 0.60	1.55 ± 0.66
		结构相似性指数(↑)	0.975 ± 0.005	0.980 ± 0.004	0.990 ± 0.003	0.972 ± 0.005	0.989 ± 0.004	0.989 ± 0.004	0.992 ± 0.004
Out-of-Dist.	×8	峰值信噪比(↑)	38.85 ± 2.38	36.60 ± 2.49	36.96 ± 2.44	39.37 ± 1.72	37.21 ± 2.59	39.55 ± 2.51	39.89 ± 2.36
		光谱角映射(↓)	2.15 ± 0.61	2.67 ± 1.10	2.45 ± 0.93	1.96 ± 0.50	2.52 ± 0.97	2.09 ± 0.82	2.13 ± 0.83
		相对全局综合误差(↓)	2.87 ± 0.52	3.37 ± 0.87	2.95 ± 0.67	2.60 ± 0.43	3.12 ± 0.58	2.47 ± 0.52	2.49 ± 0.54
		结构相似性指数(↑)	0.956 ± 0.004	0.926 ± 0.020	0.940 ± 0.012	0.960 ± 0.003	0.946 ± 0.013	0.964 ± 0.009	0.965 ± 0.008
	×16	峰值信噪比(↑)	36.52 ± 2.39	34.15 ± 2.76	33.70 ± 2.28	_	34.30 ± 2.47	36.67 ± 2.26	36.85 ± 2.24
		光谱角映射(↓)	2.90 ± 1.39	3.51 ± 1.67	3.26 ± 1.52	_	3.40 ± 1.55	2.89 ± 1.07	2.78 ± 1.27
		相对全局综合误差(↓)	3.71 ± 0.81	4.34 ± 1.41	4.02 ± 1.00	_	4.25 ± 0.93	3.21 ± 0.73	3.28 ± 0.83
		结构相似性指数(↑)	0.945 ± 0.009	0.906 ± 0.035	0.922 ± 0.024	_	0.928 ± 0.023	0.950 ± 0.017	0.951 ± 0.017
	×32	峰值信噪比(↑)	32.60 ± 2.09	32.43 ± 2.76	32.07 ± 2.27	_	31.46 ± 2.44	34.72 ± 2.15	35.35 ± 2.22
		光谱角映射(↓)	4.23 ± 2.08	4.30 ± 2.15	4.09 ± 2.07	_	4.34 ± 1.68	3.79 ± 1.67	3.64 ± 1.89
		相对全局综合误差(↓)	4.89 ± 1.30	5.07 ± 1.76	4.75 ± 1.42	_	4.91 ± 1.16	4.23 ± 0.94	3.97 ± 1.22
		结构相似性指数(↑)	0.918 ± 0.031	0.896 ± 0.044	0.909 ± 0.034	_	0.938 ± 0.025	0.939 ± 0.014	0.942 ± 0.024

Table B.8. 在四种尺度下对**WashingtonDC** 数据集进行定量比较。数值为均值 ± st d.

		指标	WashingtonDC 数据集						
			DSPNet	MIMO-SST	DCINN	DCFormer	PSTUN	LRTN	Ours
In-Dist.	×4	峰值信噪比(↑)	47.64 ± 2.23	47.68 ± 2.39	47.55 ± 1.78	47.18 ± 1.66	48.20 ± 1.61	47.95 ± 3.07	48.91 ± 3.46
		光谱角映射(↓)	1.01 ± 0.12	1.04 ± 0.16	1.04 ± 0.08	1.25 ± 0.20	1.06 ± 0.16	0.86 ± 0.13	0.76 ± 0.09
		相对全局综合误差(↓)	2.15 ± 0.37	2.15 ± 0.44	2.24 ± 0.21	2.41 ± 0.31	1.99 ± 0.28	2.48 ± 0.50	1.92 ± 0.58
		结构相似性指数(↑)	0.995 ± 0.001	0.994 ± 0.002	0.994 ± 0.000	0.994 ± 0.000	0.996 ± 0.001	0.994 ± 0.001	0.996 ± 0.002
Out-of-Dist.	×8	峰值信噪比(↑)	38.65 ± 1.59	37.36 ± 1.48	40.36 ± 2.18	44.74 ± 1.36	42.56 ± 2.60	41.46 ± 2.60	45.08 ± 1.42
		光谱角映射(↓)	4.46 ± 0.60	5.29 ± 0.82	3.24 ± 0.51	1.99 ± 0.23	2.03 ± 0.37	1.86 ± 0.25	1.80 ± 0.23
		相对全局综合误差(↓)	5.49 ± 0.75	6.37 ± 0.72	4.86 ± 1.01	3.28 ± 0.36	3.78 ± 0.73	4.44 ± 0.96	3.32 ± 0.45
		结构相似性指数(↑)	0.957 ± 0.007	0.938 ± 0.009	0.974 ± 0.008	0.989 ± 0.001	0.983 ± 0.004	0.981 ± 0.005	0.990 ± 0.000
	×16	峰值信噪比(↑)	33.73 ± 1.60	32.99 ± 1.70	35.07 ± 2.12	_	39.17 ± 4.13	37.32 ± 2.24	39.27 ± 2.49
		光谱角映射(↓)	7.96 ± 0.78	8.69 ± 0.98	6.22 ± 0.74	_	3.02 ± 0.57	3.04 ± 0.42	2.97 ± 0.74
		相对全局综合误差(↓)	8.94 ± 0.85	9.43 ± 1.06	8.12 ± 1.09	_	5.68 ± 1.68	6.44 ± 1.03	5.35 ± 1.00
		结构相似性指数(↑)	0.912 ± 0.012	0.896 ± 0.018	0.939 ± 0.014	_	0.968 ± 0.025	0.961 ± 0.010	0.969 ± 0.007
	×32	峰值信噪比(↑)	31.96 ± 1.17	31.57 ± 1.42	32.45 ± 1.70	_	36.54 ± 4.39	34.99 ± 1.95	37.00 ± 2.02
		光谱角映射(↓)	10.35 ± 1.07	11.13 ± 1.28	8.74 ± 0.96	_	3.99 ± 1.17	3.94 ± 0.50	3.54 ± 0.54
		相对全局综合误差(↓)	11.66 ± 0.97	11.61 ± 1.05	10.99 ± 1.11	_	7.13 ± 1.89	8.31 ± 1.18	6.95 ± 1.31
		结构相似性指数(↑)	0.887 ± 0.009	0.880 ± 0.010	0.910 ± 0.012	_	0.951 ± 0.033	0.945 ± 0.011	0.958 ± 0.009

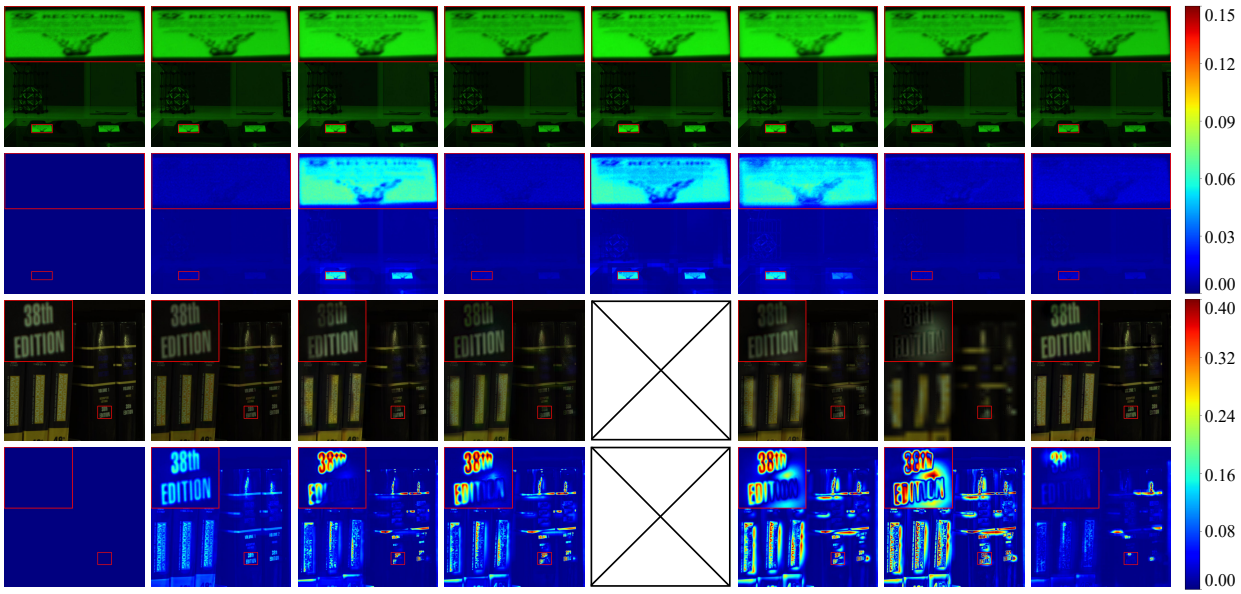

图 *B.2.* **Harvard** 数据集上的视觉对比。前两行对应分布内的 $\times 4$ 尺度，后两行对应未见过的 $\times 32$ 尺度。即使在较大的放大倍数下，我们的方法仍能保留清晰的文字细节（见放大图）。

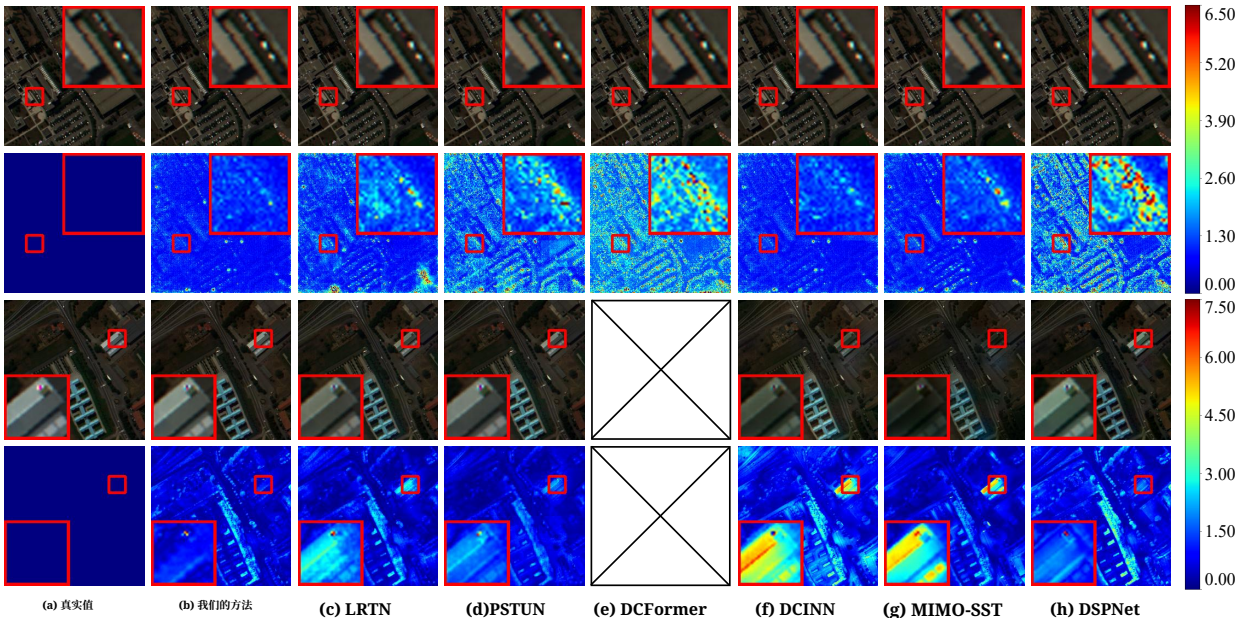

图 *B.3*。 在不同尺度下对 **PaviaU** 数据集进行的视觉比较。前两行展示训练尺度（$\times 4$）下的结果，而后两行展示分布外尺度（$\times 32$）下的结果。对于每个尺度，我们展示融合图像及其对应的误差图。请注意，DCFormer（Wu 等人，2025）无法支持任意 $\times 32$ 尺度（以带叉方框标示）。

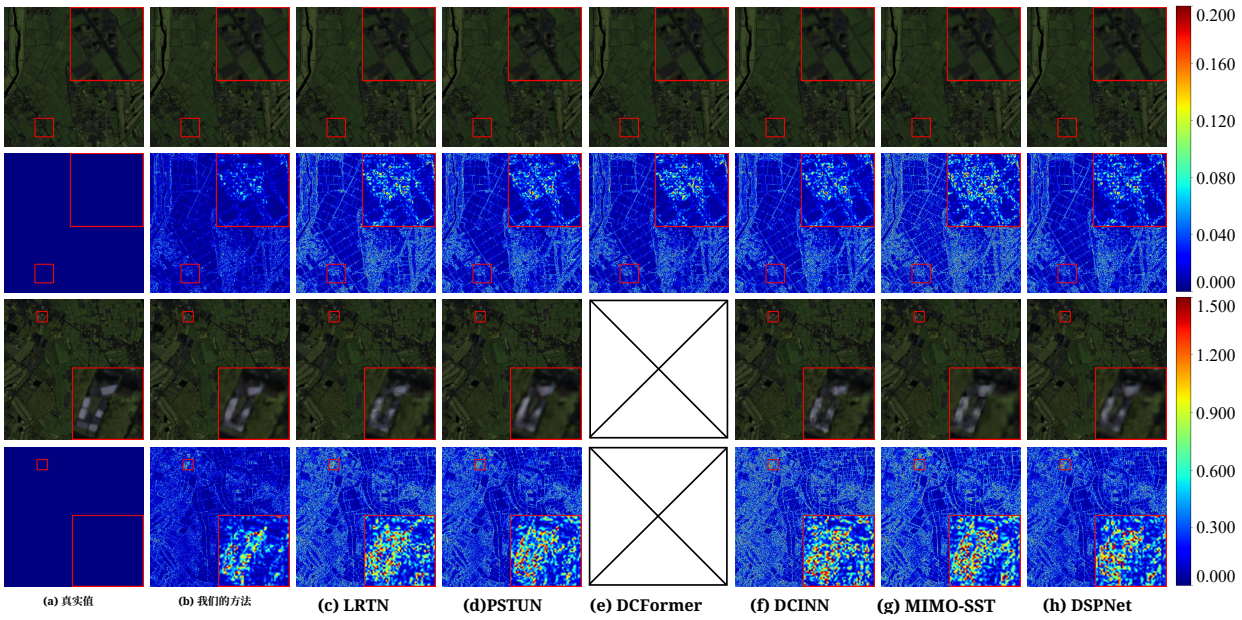

图 *B.4*。 在 **Chikusei** 数据集上的视觉对比。前两行： ×4 尺度；后两行： ×32 尺度。误差图（蓝色表示误差较低）表明，我们的模型能够很好地泛化到具有不同空间分辨率的大规模遥感场景。

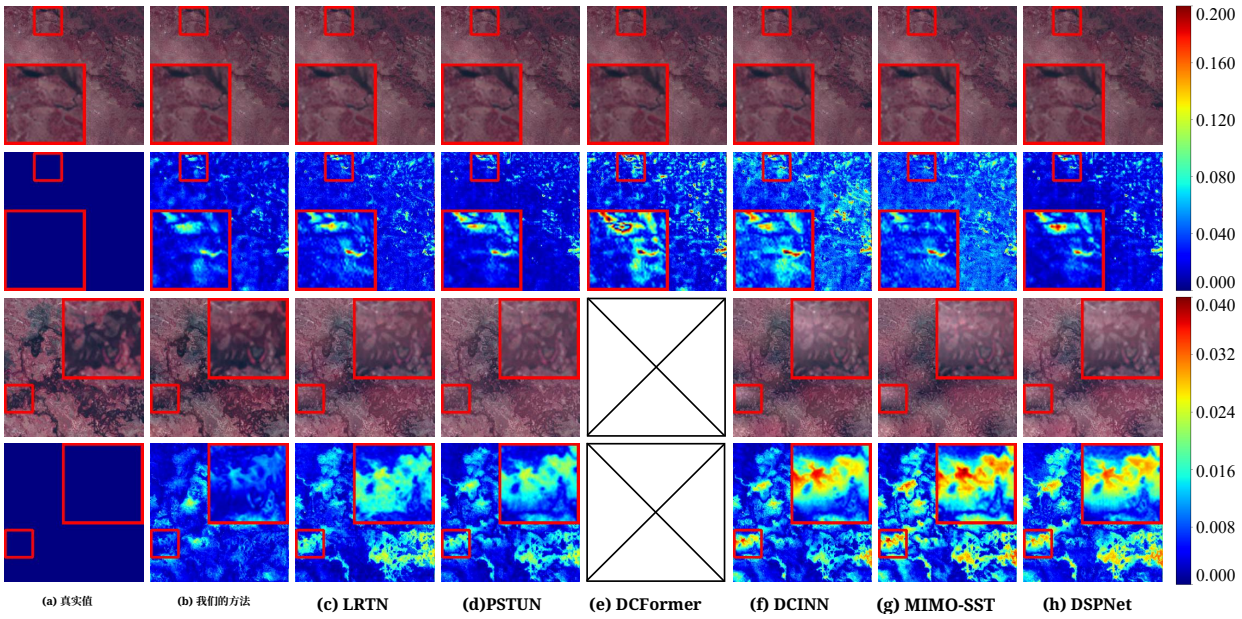

图 *B.5*。 在**博茨瓦纳** 数据集上的视觉对比。前两行： ×4 缩放；后两行： ×32 缩放。与对比方法相比，我们的方法能够更好地保持光谱和空间一致性；对比方法在较大的放大因子下通常会显示出高能量残差（红色/黄色）。

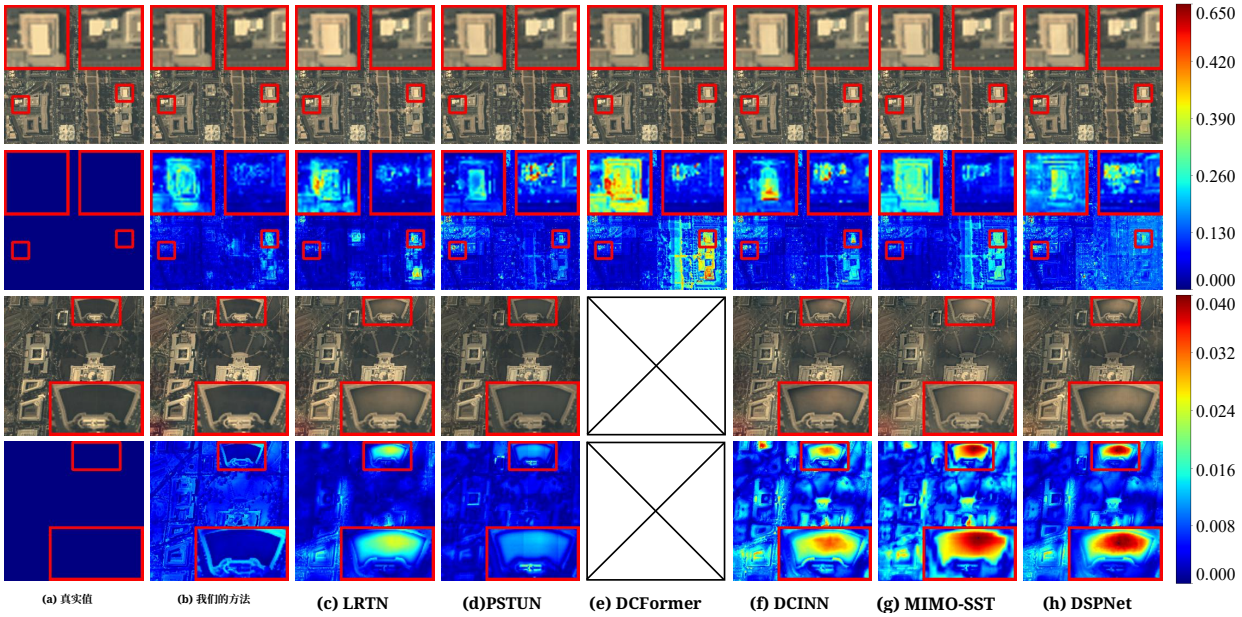

图 *B.6.* **华盛顿特区**数据集上的视觉对比。上两行： $\times 4$ 尺度；下两行： $\times 32$ 尺度。尽管波段数量较多（191 个波段），我们的统一模型无需重新训练模型，即可稳健地重建不同缩放因子下的空间细节。