arXiv:2611.33032v1 [cs.CV] 24 Nov 2026

SHARPnP：基于真实噪声未对齐相机测量的自监督与即插即用高光谱超分辨率

Mohamad Jouni[1] Brayan Monroy[1,2] Jorge Bacca[1,2] Mauro DallaMura[1,3]法国阿尔卑斯大学、法国国家科学研究中心、格勒诺布尔国立高等工程学院、GIPSA实验室，法国格勒诺布尔38000[2]桑坦德工业大学，哥伦比亚布卡拉曼加[3]法国大学研究院（IUF），法国巴黎

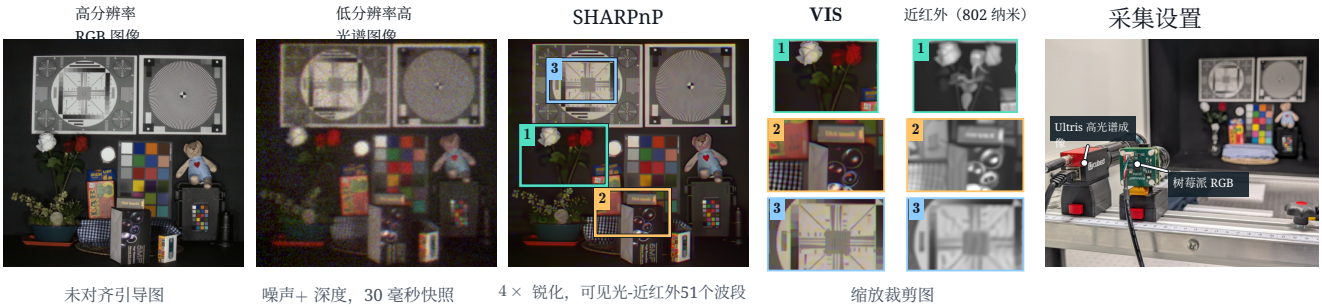


图1。**未对齐相机下的高光谱超分辨率**。SHARPnP将带噪的30毫秒Ultris采集结果与未对齐的高分辨率RGB视图融合，重建覆盖450–850纳米范围的51个波段。重建结果实现了超过 4× 的有效细节放大。配对的可见光（VIS）和近红外（NIR，802纳米）裁剪图展示了相同的三个区域：（1）花朵，（2）前景深度边界，以及（3）分辨率图表。右侧面板展示了Ultris–RPi采集装置。

摘要

经济实惠的快照式高光谱相机可通过单次曝光捕获光谱图像，但空间分辨率有限且存在传感器噪声。低成本 *RGB* 相机可提供精细的可见光细节，以引导近红外重建，但会引入空间变化的跨相机错位。有监督融合方法需要相对较大的无噪、高分辨率高光谱图像数据集，对传感器噪声敏感，且难以推广到新传感器。我们提出 *SHARPnP*，这是一种可解释、灵活的框架，将配准与重建相结合，能够基于含噪、未对齐的测量数据实现 *RGB* 引导的高光谱超分辨率，且无需无噪参考图像。我们通过跨视图损失调整预训练配准模型，并将该损失反向传播至固定的即插即用融合模块。时间配对数据和重新加噪配对数据使含噪输入能够进行无参考自监督。测试时，我们校准跨相机辐射测量并细化对齐。即插即用融合采用顺序三波段去噪。在 *REAL*基准上，*SHARPnP* 在无噪输入上以 4×、 8× 和 16× 超越有监督融合基线，并在 4× 高斯噪声下（$\sigma = 5/255$、$10/255$）也表现更优。在原生 *RGB-HSI* 快照测量数据上，该方法无需高分辨率参考图像，即可重建全部 *51* 个可见光-近红外波段，超越传统的 *31* 波段可见光设置。

1. 引言

紧凑型快照式高光谱成像（HSI）能够在单次曝光中捕获密集的光谱测量数据，从而支持对动态场景进行材料表征 [18]。在 RGB 图像中外观相似的材料，其光谱特征可能截然不同；这些特征在机器人检测、精准农业、废物分选、无损检测、文化遗产成像、气体检测和生物医学等应用中都很有用 [4, 9, 20–22, 26]。低成本快照传感器以牺牲空间分辨率和信噪比为代价，换取丰富的光谱信息和更快的采集速度，因而会使精细结构变得模糊。

RGB引导的高光谱超分辨率（HSR）利用一台独立的相对高分辨率相机来恢复空间细节 [10, 12]。在地面层，随深度变化的视差和遮挡使跨相机对齐更加复杂，

而传感器响应差异和高光谱图像噪声则使像素对应关系更难匹配。未对齐的引导图可能引入伪边缘，而不加区分地抑制引导会使重建结果模糊 [10]。

监督方法（如 SSC [33]和 UAFL [34]）通过从合成输入和无噪高分辨率参考图像中学习空间-光谱特征，改善非配准高光谱超分辨率；然而，这些数据在现实中几乎无法获得。其训练条件规定了固定退化、光谱采样和尺度因子，因此难以迁移到存在噪声且传感器响应不同的原生 RGB-高光谱成像相机对。这促使我们直接从采集到的图像中估计跨相机对齐和辐射差异。CMF-GSA-wg [10] 通过依据配准置信度传递 RGB 细节来应对这一挑战。然而，其单遍融合在不可靠区域依赖低分辨率数据，并且抑制噪声的能力有限。

基于置信度感知的细节传递，我们提出了**SHARPnP**，这是一种可解释且灵活的模型，将配准与重建相结合，用于从带噪且未对齐的相机测量中进行 RGB引导的高光谱超分辨率，无需无噪参考图像。我们通过将跨视图损失反向传播通过固定融合模块，对预训练配准模型 Cross-RAFT [37], 进行适配。这使融合估计的质量能够指导光流更新。时序配对数据和重新加噪配对数据在没有无噪高分辨率参考图像的情况下，为含噪输入提供自监督，以限制噪声导致的光流变化。在测试时，SHARPnP会校准跨相机辐射测量并细化残差对齐。即插即用（PnP）融合通过逐像素配准置信度对 RGB引导进行加权，并复用冻结的预训练去噪器。在变形映射不确定的区域，估计结果更多地依赖高光谱图像观测和去噪先验。高效的冻结预训练 RGB 去噪器 [23] 将该先验应用于光谱组，使其能够在不同光谱网格之间运行。我们的贡献如下:

- **自监督融合感知配准**。 我们利用含噪观测而非Clean高分辨率参考图像，针对高光谱重建调整配准方法。
- **传感器感知的即插即用融合**。 我们结合显式的空间退化模型、拟合的跨相机辐射度、置信度加权的RGB引导以及可复用的去噪先验。
- **受控与原生传感器评估**。SHARPnP在受控的REAL基准上超越了有监督基线，在干净的 $4\times$ 至 $16\times$ 以及含噪的 $4\times$上均表现出色，其中还包括使用相同去噪器的顺序“先去噪后融合”控制实验。在我们采集的原生RGB-高光谱成像快照设置中，该方法还重建了450至850纳米范围内原生可见光-近红外采集的51个波段。

2. 相关工作

非配准高光谱超分辨率。ESSAformer等单图像方法 [31] 仅从高光谱图像中恢复结构。RGB引导提供了额外的空间细节，同时也带来了建立视图间对应关系的需求。HSIFN [12]引入了REAL，用于评估来自非配准地面视图的引导式高光谱超分辨率。在此设置下，SSC [33]改进了光流估计和空间-光谱聚合，而UAFL [34] 则利用可变形聚合和交叉注意力重建丰度图。SA-RFM [32]通过黎曼流匹配将配准与融合相结合，并利用多光谱梯度保留结构边界。

测量驱动和基于模型的融合。 无监督融合方法可以从观测中推断跨传感器关系。例如，u^2-MDN [25]通过共享表示和互信息关联不同模态，而BUSIFusion [14]估计退化参数，并从单个图像对重建高光谱图像。另一个相关问题是所需的高分辨率光谱信息量。SparseHSR [35]研究了稀疏高分辨率可见光-近红外引导的光谱覆盖范围如何影响基于光谱完整的低分辨率高光谱图像进行的重建。

预训练先验与噪声视图自监督。当测量使重建问题欠定时，去噪先验有助于对估计进行正则化。即插即用复原将去噪器融入迭代式逆问题求解器中，例如DPIR [29], ；Picone等人 [23]探索了光谱分组，以便在高光谱数据上复用RGB去噪器。当Clean参考图像稀缺时，含噪观测也可以提供监督。Noise2Noise [13]从重复的含噪观测中学习，而GR2R [19]则在给定噪声模型的情况下，从单次测量构造训练视图。

3. 图像形成模型

3.1. 采集模型

我们希望根据低分辨率高光谱观测 $Y \in \mathbb{R}^{n_\ell \times B}$ 以及从不同视角拍摄的同一场景的高分辨率RGB图像 $R \in \mathbb{R}^{n_R \times 3}$，重建高分辨率高光谱图像 $X \in \mathbb{R}^{n \times B}$。每一列都是一个向量化波段或颜色通道。我们采用REAL 使用的模糊并下采样模型 [34],

$$Y = A_s X + N, \qquad A_s = S_s B_h, \tag{1}$$

其中， $B_h \in \mathbb{R}^{n \times n}$ 应用空间模糊， $S_s \in \mathbb{R}^{n_\ell \times n}$按因子 s进行下采样。加性噪声 $N \in \mathbb{R}^{n_\ell \times B}$不可观测。

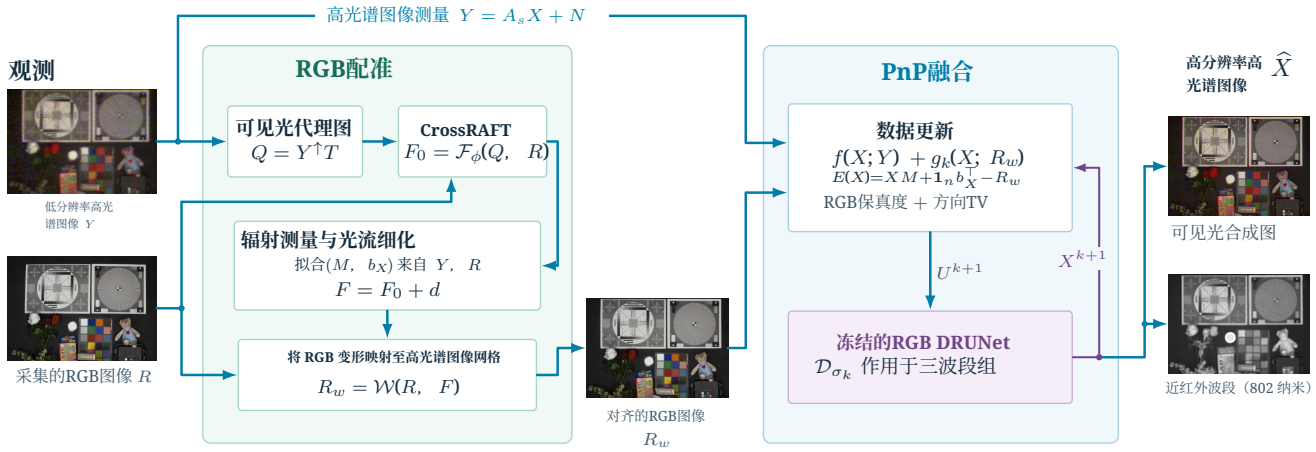


图2。**SHARPnP 的前向过程**。CrossRAFT 根据采集的RGB图像 R 和低分辨率高光谱图像的可见光代理图 Y 估计初始光流。我们拟合辐射测量并优化光流，使其对齐到高光谱图像网格 R，生成 R_w。即插即用融合交替执行数据更新（结合高光谱图像保真度、RGB保真度和方向TV）以及冻结的RGB DRUNet；后者一次处理三个波段。可见光合成图和 802 纳米波段是原生双相机采集得到的同一重建数据立方体 $\widehat{X}$ 的两种视图。

3.2. 配准模型

配准函数 $\mathcal{W}$ 使用稠密对应关系 F，在高光谱图像输出网格上对 RGB R 进行采样。在有效重叠区域内，我们将对齐的RGB图像建模为潜在图像 [10, 27],的仿射光谱投影

$$R_w = \mathcal{W}(R, F) \simeq XM + \mathbf{1}_n b_X^\top, \tag{2}$$

其中，$M \in \mathbb{R}^{B\times 3}$ 是有效光谱响应，$b_X \in \mathbb{R}^3$ 是特定场景的RGB偏移量，$\mathbf{1}_m$ 是由m 个1组成的列向量。对于校准对 j，$M_j = P_R$=diag(g_j)，其中$P_R \in \mathbb{R}^{B\times 3}$ 的列和为1，$g_j \in \mathbb{R}^3_{>0}$ 包含该校准对的正RGB增益。

4. 方法

SHARPnP包含两个主要模块（图2）。融合感知RGB配准将采集的图像 R 配准到高光谱图像网格，以获得引导图 R_w。传感器感知即插即用（PnP）融合将 R_w 与高光谱图像观测 Y 结合，并使用其感知算子 A_s 生成最终的高光谱估计 $\widehat{X}$。

4.1. 融合感知RGB配准

配准分三个阶段进行。CrossRAFT适应利用观测视图对的重建反馈。在测试时，辐射校准使用初始光流 F_0拟合跨相机响应仿射变换，而残差对齐则通过有界位移d_Θ优化该光流。优化后的光流 F 在高光谱图像输出网格上对采集的RGB图像 R进行采样，从而生成用于融合的引导图 R_w，

$$F = F_0 + d_\Theta, \qquad R_w = \mathcal{W}(R, F_0 + d_\Theta).$$

CrossRAFT适应。 为在共同的三通道表示中比较RGB和高光谱图像，我们在输出网格 [10]上构建可见光代理图像 $Q = Y^{\uparrow} T$。其中，$Y^{\uparrow} \in \mathbb{R}^{n\times B}$ 是对 Y进行的双三次上采样，$T \in \mathbb{R}^{B\times 3}$ 选择三个可见光波段。我们利用高光谱重建的反馈对预训练的CrossRAFT [37] 进行微调。适应后的模型预测用于残差对齐的初始光流，

$$F_0 = \mathcal{F}_\phi(Q, R),$$

其中，$F_0 \in \mathbb{R}^{n\times 2}$ 将输出网格上的位置映射到采集的RGB图像 R，而 ϕ 表示可训练参数。

在微调过程中，同一场景的两个视图分别根据其可见光代理图 $Q_i = Y_i^{\uparrow} T$和RGB图像 R_i生成光流 F_i。按照第 4.2 节中的即插即用方案，固定的重建模块基于双三次高光谱图像升采样$Y_i^{\uparrow}$ 和变形后的RGB引导图$R_{w,\,i} = \mathcal{W}(R_i, F_i)$生成初步高分辨率估计 Z_i。该目标函数结合了每个视图内的对齐以及跨视图监督，

$$\begin{aligned}\mathcal{L}_{\text{ft}}(\phi) = \mathcal{L}_{\text{proxy}} + \mathcal{L}_{\text{frame}} \\ + w_f \mathcal{L}_{\text{fusion}} + w_s \mathcal{L}_{\text{smooth}}.\end{aligned} \tag{3}$$

我们设置 $w_f = 0.25$ 用于跨视图融合，并设置 $w_s = 10^{-3}$用于光流平滑度。表1 定义了四个项。代理图一致性在每个视图内对齐RGB与高光谱图像，帧一致性则促使两个重建结果保持一致。跨视图融合利用另一视图的两种模态来引导光谱内容和空间细节。光流平滑度限制估计光流中的突变。

噪声适应。 为了提高模型对含噪高光谱图像输入的鲁棒性，我们在 GR2R [19]视图 $Y_1 = Y + \tau N'$上进行微调，并

表 1. **配准损失**。 Z_i 表示初始高光谱图像估计值。$\sim R_{w,\,i}$ 和 R_i 分别表示变形后的RGB引导图和代理图像。 $\bar{V}_i$ 是跨颜色通道共享的单位均值掩膜。MAE表示平均绝对误差，sg表示停止梯度， $d \in \{x, y\}$ 表示空间方向。

损失名称	公式
$\mathcal{L}_{\text{proxy}}$	$\sum_i$ 平均绝对误差$(\bar{V}_i \odot (\tilde{R}_i - Q_i))$
$\mathcal{L}_{\text{frame}}$	$\frac{1}{2}\sum_{i\neq j}$ 平均绝对误差$(Z_i - \text{sg}(Z_j))$
$\mathcal{L}_{\text{fusion}}$	$\frac{1}{2}\sum_{i\neq j}\Big[$平均绝对误差$(Z_i - \text{sg}(Y_j^{\uparrow}))$ + 平均绝对误差$(\text{sg}(\bar{V}_j) \odot (Z_i T - \text{sg}(R_{w,\,j})))\Big]$
$\mathcal{L}_{\text{smooth}}$	$\frac{1}{2}\sum_{i,d}$ 平均绝对误差$(\partial_d F_i)$

$Y_2 = Y - \tau^{-1}N'$由一次测量构成 Y。对于标准差为 σ_n 的高斯测量噪声，我们从 $N' \sim \mathcal{N}(0,\ \sigma_n^2 I)$ 中独立于测量噪声进行采样，并选择 $\tau \in (0,\ 1)$。在此模型下，各视图共享相同的信号，且噪声相互独立，因此每个视图都能监督另一个视图。对于原生时序数据，我们使用同一静态场景中相邻的两帧作为 Noise2Noise [13] 配对。

辐射校准。 在共享的响应曲线形状 P_R 固定后，我们拟合 g_s 的增益 M 以及 RGB 偏置b_X，具体方法是根据第 3.2 节，在低分辨率网格上使用初始变形映射和传感器响应拟合的加权变体进行拟合 [27]。RGB 预测 R_ℓ 及其残差 E_ℓ 为:

$$R_\ell = YM + \mathbf{1}_{n_\ell} b_X^\top, \qquad E_\ell = R_\ell - A_s\mathcal{W}(R, F_0) \in \mathbb{R}^{n_\ell \times 3}.$$

空间退化与光谱混合可交换，并在拟合支撑域上保留偏置。定标损失为

$$\mathcal{L}_{\text{cal}}(g_s, b_X) = \sum_p w_p \|E_\ell(p, :)\|_2^2 + \|g_s\|_2^2. \tag{4}$$

H其中 $w_p \geq 0$是低分辨率像素 p的固定拟合权重。

残差对齐。 对于每个测试对，我们将有界残差位移参数化为

$$d_\Theta = v \odot \Delta \tanh(\mathcal{I}(\Theta)), \tag{5}$$

此处， $\mathcal{I}$ 对粗位移网格 Θ进行双线性插值， Δ将每个分量限制在输出像素范围内，而 v则将校正限制在有效位置。

令 $c(\Theta)$ 表示 R_ℓ 与 $A_s\mathcal{W}(R,\ F)$ 之间的局部归一化互相关（LNCC），采用固定权重 V^{align}，并在 RGB通道和像素上取平均。

$$\mathcal{L}_{\text{align}}(\Theta) = 1 - c(\Theta) + \frac{\beta_d}{n_e}\|\nabla_s d_\Theta\|_F^2 + \frac{\beta_m}{n}\|d_\Theta\|_F^2. \tag{6}$$

正则项惩罚 n_e 空间边缘上的变化以及跨越 n像素的位移。在已配置的情况下，我们会在对齐后重新拟合增益和偏置。

4.2. 传感器感知即插即用融合

重建目标。 固定对齐和辐射度后，我们估计 X ，由 Y 和 R_w 在其原始数值单位下得到。数据更新目标为 $f(X;\ Y) + g_k(X;\ R_w)$，其中包含高光谱图像保真度 f 和RGB引导 g_k。利用式（2）中的仿射预测，可写为

$$\begin{aligned} f(X;Y) &= \frac{\|Y - A_s X\|_F^2}{2Bn_\ell}, \\ g_k(X;R_w) &= \underbrace{\ell_{\text{RGB},k}(E(X))}_{\text{RGB fidelity}} + \underbrace{\mu\mathcal{V}(X;R_w,C_P)}_{\text{edge preservation}}, \\ E(X) &= XM + \mathbf{1}_n b_X^\top - R_w. \end{aligned} \tag{7}$$

RGB保真度$\ell_{\text{RGB},\ k}$ 结合了像素值和空间梯度的置信度加权平方残差。映射 C_P 以强度 μ控制RGB边缘对方向总变差 [3], 的影响。对于原生数据，RGB引导同时采用数值和梯度保真度，方向总变差强度为 $\mu = 10^{-5}$。

迭代更新。每次迭代都会交替进行数据更新和去噪更新。采用半二次分裂时，数据更新计算近似极小值

$$U^{k+1} \approx \arg\min_{U\in\mathbb{R}^{n\times B}} \left\{ f(U;Y) + g_k(U;R_w) + \frac{\rho_k}{2}\|U - X^k\|_F^2 \right\}. \tag{8}$$

二次项以权重 ρ_k 将数据更新与 X^k 关联起来。我们使用方向TV的二次主化函数，并通过共轭梯度法求解由此得到的子问题。去噪器 $\mathcal{D}_{\sigma_k}$ 提供隐式的即插即用先验 [29], ，其条件水平为 σ_k ，校正强度为 a_k,

$$X^{k+1} = U^{k+1} + a_k\left[\mathcal{D}_{\sigma_k}(U^{k+1}) - U^{k+1}\right]. \tag{9}$$

光谱去噪。我们使用预训练RGB DRUNet [29]。它按波长顺序处理原始光谱波段，每次处理三个 [23]。

5. 实验

5.1. REAL上的受控重建

数据与指标。REAL由真正未对齐的高光谱图像视图组成，基准通过光谱投影构建RGB引导，并通过模糊/下采样生成低分辨率高光谱图像 [12, 34]。其31个波段覆盖400至700纳米。我们使用官方的10个测试场景，以及保留的47个场景拟合划分。基准模糊采用宽度为3的 8×8高斯函数，并使用左上角采样相位。

表 2. **REAL 数据集，干净重建**。 所有复现结果均使用十个测试场景。文献行（†）和（‡）分别来自 [34] 和 [33]，遵循 [34]的训练/测试协议。粗体/下划线分别表示所展示均值中的最佳/次佳结果。

Type	方法	场地	4×				8×				16×			
			峰值信噪比↑ (dB)	结构相似性指数↑	光谱角映射↓ (°)	均方根误差↓ (10^{-2})	峰值信噪比↑ (dB)	结构相似性指数↑	光谱角映射↓ (°)	均方根误差↓ (10^{-2})	峰值信噪比↑ (dB)	结构相似性指数↑	光谱角映射↓ (°)	均方根误差↓ (10^{-2})
SISR	双三次插值	–	35.40	0.951	2.137	2.255	31.90	0.908	2.735	3.389	27.49	0.850	4.015	5.599
	SSPSR [8]†	TCI'20	39.04	0.976	2.292	–	31.13	0.906	3.953	–	27.20	0.851	4.412	–
	MCNet [15]†	RS'20	39.07	0.974	2.177	–	31.31	0.904	3.610	–	27.04	0.845	5.042	–
	BiQRNN [7]†	JSTAR'21	37.80	0.969	2.521	–	31.66	0.908	3.438	–	26.53	0.832	6.360	–
	ESSA [31]†	ICCV'23	39.94	0.982	2.292	–	31.85	0.913	3.667	–	26.70	0.918	6.360	–
融合	优化 [6]†	CVPR 2019	27.26	0.916	10.829	–	26.99	0.909	11.173	–	26.93	0.912	12.605	–
	u^2-MDN [25]‡	TGRS 2022	30.58	0.936	4.985	–	30.18	0.936	5.500	–	29.84	0.932	6.245	–
	NonReg [36]†	TGRS 2022	31.60	0.951	4.641	–	30.00	0.923	9.568	–	29.09	0.904	33.461	–
	MoE-PNP [24]†	TNNLS'25	32.57	0.951	5.787	–	31.12	0.946	6.818	–	29.91	0.934	6.933	–
	CMF-GSA-wg [10]	BMVC'26	37.66	0.976	2.117	1.612	33.90	0.947	2.687	2.605	28.34	0.872	3.836	5.080
	HSIFN [12]	TNNLS'25	40.28	0.983	2.146	1.129	34.76	0.956	3.016	2.219	27.98	0.864	4.149	5.193
	LRTN [16]†	IJCV'25	38.62	0.978	3.094	–	34.79	0.957	3.782	–	31.41	0.937	4.641	–
	SRLF [17]†	CVPR'25	39.06	0.978	2.464	–	34.63	0.953	3.667	–	31.12	0.929	5.214	–
	SSC-S [33]	IJCV'25	40.70	0.985	2.151	1.070	35.72	0.962	2.963	1.986	31.51	0.939	4.512	3.220
	UAFL [34]	CVPR'26	41.91	0.988	1.908	0.910	37.14	0.971	2.669	1.645	31.89	0.940	3.784	3.096
SSL/PnP SHARPnP（我们的方法）		–	**43.37**	**0.991**	**1.595**	**0.770**	**37.38**	**0.973**	**2.174**	**1.640**	**32.36**	**0.946**	**3.125**	**2.970**

我们报告逐波段峰值信噪比和结构相似性指数的均值、平均光谱角（光谱角映射，单位为度）以及全局均方根误差。复现指标遵循 [34]的训练/测试协议。对于含噪情形，在低分辨率高光谱图像中加入标准差为 $5/255$ 或 $10/255$ 的独立高斯噪声，并采用固定的按数据集索引的实现。几何和辐射拟合使用含噪观测。

基线方法。 我们按照 [34],的训练/测试协议，重新评估了双三次插值、CMF-GSA、HSIFN、SSC 和 UAFL，并纳入了从同一来源移植而来的已发表 SISRand 融合评估结果。在含噪情形下，没有可直接比较的最先进水平方法，因为上述方法会将噪声与信号一同传播。因此，我们采用相同的 DRUNet 机制，先去噪再融合，构建顺序式基准；随后分别使用 SSC 或 UAFL，称为 DRU-SSC 和 DRU-UAFL。这些对照用于检验先去噪再融合这一实用替代方案。

干净重建。 表 2显示，在每种尺度下，保留的 SHARPnP 配置在四项指标上均优于有监督基线。在 4×，基于 CrossRAFT 的重建达到 43.37 dB，而 UAFL 为 41.91 dB。在 8×，重建达到 37.38 dB，高于 UAFL 的 37.14 dB，且四项指标的均值均优于 UAFL。在 16×，基于 CrossRAFT 的重建达到 32.36 dB 和 3。125° SAM，高于 UAFL 的 31.89 dB 和 3。784° 。因此，保留的配置在三种空间采样条件下均实现了重建增益。

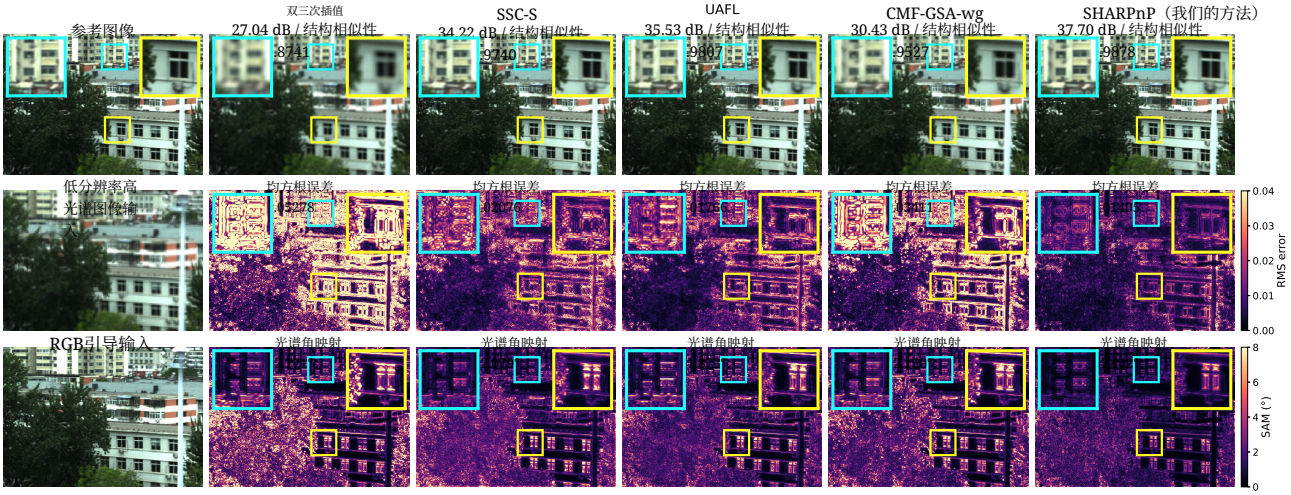
噪声污染下的重建。 在加入高光谱图像噪声后，直接监督基线的性能显著下降

表 3。**噪声 REAL，位于** 4×。 仅向 LR- HSI 添加高斯噪声。DRU 表示在冻结的有监督融合模型之前应用冻结的 DRUNet。评估使用十个测试场景。

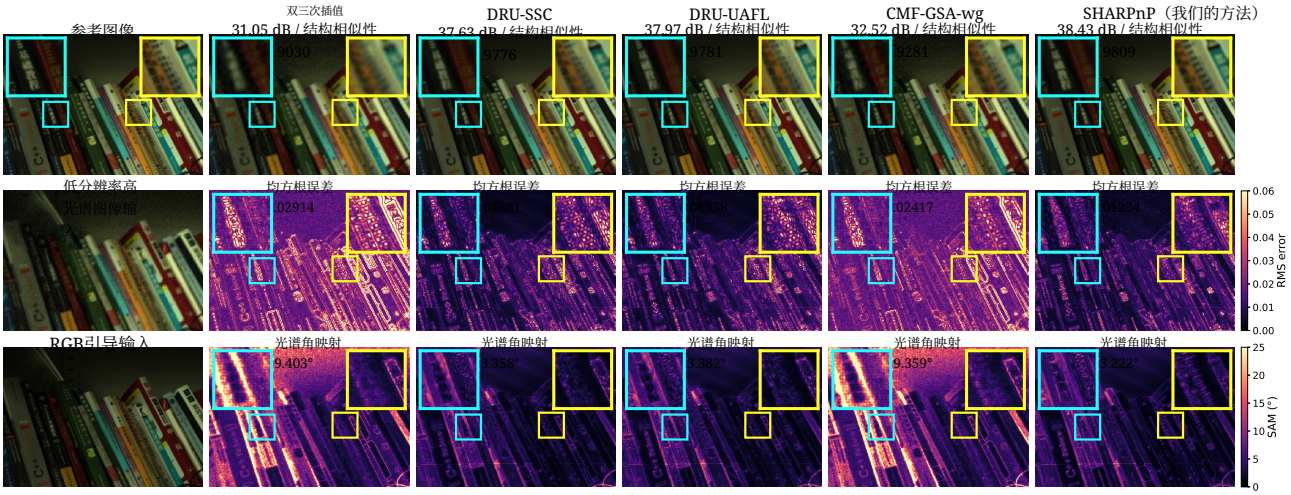
方法	$\sigma_n = 5/255$				$\sigma_n = 10/255$			
	峰值信噪比↑ (dB)	结构相似性指数↑	光谱角映射↓ (°)	均方根误差↓ (10^{-2})	峰值信噪比↑ (dB)	结构相似性指数↑	光谱角映射↓ (°)	均方根误差↓ (10^{-2})
双三次插值	31.60	0.901	8.170	2.870	27.90	0.785	14.458	4.135
CMF-GSA [10]	32.85	0.927	8.052	2.382	28.66	0.822	13.630	3.745
HSIFN [12]	32.34	0.905	9.349	2.481	27.18	0.751	16.079	4.463
SSC-S [33]	31.93	0.891	9.642	2.611	26.78	0.725	16.625	4.693
UAFL [34]	32.21	0.896	9.813	2.496	26.53	0.706	17.677	4.790
DRU-SSC	39.01	0.978	2.837	1.298	37.08	0.969	3.546	1.604
DRU-UAFL	39.21	0.979	2.868	1.231	37.17	0.969	3.557	1.567
SHARPnP	**39.33**	**0.981**	**2.736**	**1.185**	**37.25**	**0.971**	**3.375**	**1.512**

（表 3）。对其输入进行去噪，比直接进行带噪推理更具挑战性：DRU-UAFL 在两个噪声水平下分别达到 39.21/37.17 dB。SHARPnP 达到 39.33/37.25 dB，并在展示的四项指标上均优于两个 DRU 基线。其均方根误差为 0.01185/0.01512，光谱角映射为 2。736°/3。375°这些结果表明，其优势不仅仅来自在现有融合网络之前加入去噪器。

图 3(a) 展示了干净的 4× 重建效果，与表 2中的定量比较相辅相成。与评估的融合基线相比，SHARPnP 能恢复更精细的结构。图3(b) 展示了一个带噪的 4× 示例，其 $\sigma_n = 5/255$对应表3。在此，先去噪后融合基线先对低分辨率高光谱图像应用冻结的 DRUNet，然后进行有监督融合。而 SHARPnP 则通过其即插即用



(a) 真实场景 0； Clean； 4×



(b) REAL 场景 7； 5/255 低分辨率高光谱图像噪声； 4×；先去噪后融合对比

图3. REAL 与双三次插值、SSC-S 的对比 [33], UAFL [34], CMF-GSA-wg [10], 和 SHARPnP（本文方法）。含噪面板使用 DRU-SSC 和 DRU-UAFL。各行依次展示重建的 RGB 预览图、均方根误差图和光谱角映射图；参考图像和输入视图位于左侧。标注为全帧指标。

优化进行联合去噪与重建，以更少的噪声伪影恢复更精细的细节。这一视觉比较印证了相较于顺序处理流程的定量改进。

5.2. 消融研究

先验的选择。 在匹配的无噪声 4× 对照实验中，将 DRUNet 替换为恒等映射后，峰值信噪比从 43.37 降至 41.83dB（表 4）。其他预训练去噪器和 Haar收缩仍可使用，但在测试设置下精度较低。这表明应复用性能强大的预训练先验，而不应将全部结果归因于校准或 RGB保真度。

几何与引导相互作用。 表5分别考察了残余配准和 RGB 引导强度在 final Clean 4× SHARPnP 配置。残差校正使 0.1× 引导下的峰值信噪比提高了 0.54 dB，并使保留强度下的峰值信噪比提高了 1.7 dB。没有残差校正时，更强的引导会使峰值信噪比降低 0.94 dB，并使全部四项平均指标变差。采用校正后的几何后，峰值信噪比提高了 0.245 dB，结构相似性指数和均方根误差均有所改善，但光谱角映射变差。随后加入方向 VTV，可使每个场景中的全部四项指标均得到改善，达到 43.37 dB，且光谱角映射为1.595°。这些结果表明，对应关系校正、细节传递和空间正则化具有互补作用。有关实现的更多细节，请参阅 SM。

表 4. **去噪器先验消融**。十个场景的 Clean $4\times$ 消融。Haar 条目为小波收缩对照。其阈值是在四个基准场景上选取的。恒等映射不进行去噪。

去噪器	峰值信噪比↑ (dB)	结构相似性指数↑	光谱角映射↓ (°)	均方根误差↓ ($\times 10^{-2}$)
DRUNet [29]	**43.37**	**0.991**	**1.595**	**0.770**
FFDNet [28]	42.88	0.990	1.624	0.823
SCUNet [30]	42.79	0.989	1.632	0.838
GRUNet [11]	42.39	0.989	1.686	0.860
三维 Haar [5]	41.88	0.988	1.716	0.905
二维 Haar [5]	41.88	0.988	1.718	0.906
恒等映射	41.83	0.988	1.722	0.910

表5。**几何、RGB 引导和方向 VTV**。在迭代 50 次时，对十个场景取平均的 Clean $4\times$。RGB 强度相对于保留值（10^{-4}）缩放值系数和梯度系数。校准、置信度、冻结的 DRUNet 和求解器设置均固定。残差校正限制在 ± 2 个 HR 像素以内；dVTV 的权重为 10^{-5}。

残差	RGB 强度	dVTV	峰值信噪比↑ (dB)	光谱角映射↓ (°)	结构相似性指数↑	均方根误差↓ (10^{-2})
Off	$\times 0.1$	Off	41.61	1.694	0.985	1.062
On	$\times 0.1$	Off	42.15	1.658	0.987	0.953
Off	$\times 1$	Off	40.67	2.040	0.981	1.259
On	$\times 1$	Off	42.40	1.833	0.989	0.909
On	$\times 1$	On	**43.37**	**1.596**	**0.991**	**0.771**

5.3. 原生真实 RGB-高光谱成像快照数据 a

采集与适配。 我们使用 Ultris SR5 紧凑型快照式高光谱相机和 Raspberry Pi HQ RGB 相机，采集包含不同深度物体的静态场景序列。该高光谱图像包含 51 个可见光-近红外波段，范围为 450 至 850 纳米，间隔为 8 纳米；其中前 31个波段覆盖 450 至 690 纳米。我们采用 480 毫秒曝光以获取更高信噪比的观测，并采用 30 毫秒曝光以拍摄 10 fps 的高光谱视频，此时传感器噪声更强。测量数据经过暗场/平场校正，未进行合成退化或添加噪声。时间 Noise2Noise 对对用于适配 Cross-RAFT，而辐射测量和残差对齐则根据测量数据进行拟合。每次曝光使用 70 帧适配帧、10 帧验证帧和 10 帧诊断帧，不使用 Clean 高分辨率参考图像。

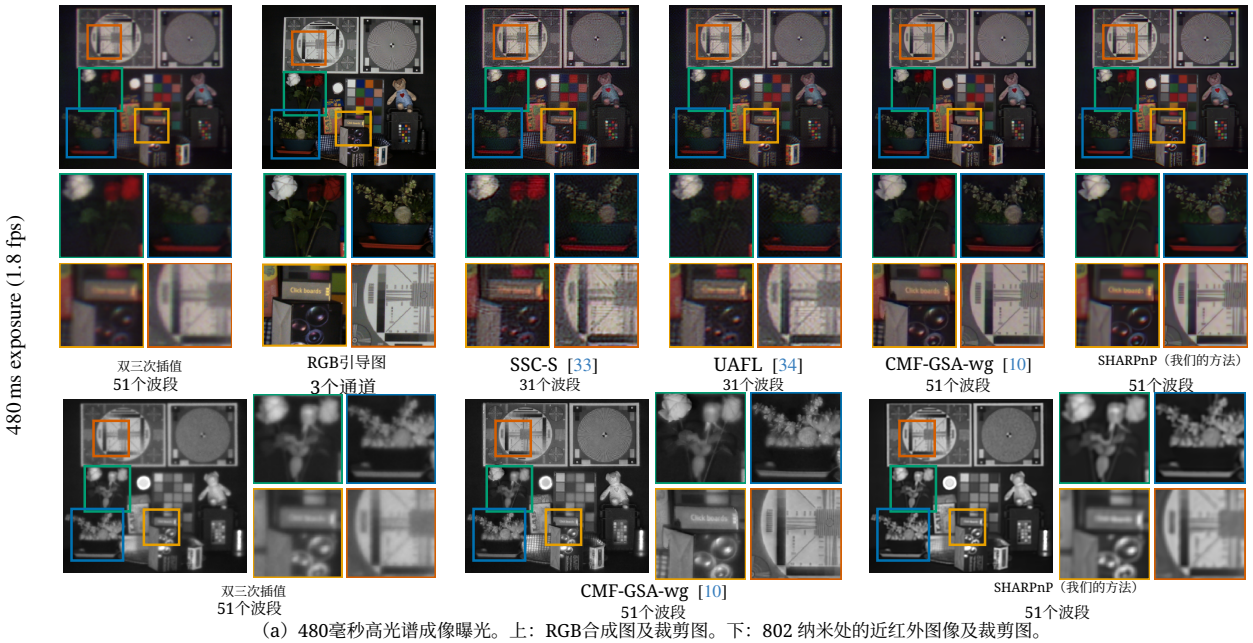
原生配置。我们使用 20 次 HQS迭代和冻结的 DRUNet 应用 $4\times$ 超分辨率，并采用 $P_D = I_{51}$。方向 VTV 使用 $\mu = 10^{-5}$；共轭梯度的容差为 10^{-5}。RGB 值/梯度系数在 30 毫秒时为 $(0.304, 0.0909)$，在 480 毫秒时为 $(0.0297, 0.0888)$。

图 4对比了原生重建结果。480 毫秒的采集提供了更强的信号，但仍然存在噪声，而 30 毫秒视频则展示了低信噪比快照成像造成的劣化。SSC 和 UAFL 会传播测量噪声，并丢失细微的目标结构。CMF-GSA-wg 能够迁移 RGB 细节，但仍残留可见噪声，且局部区域较为平滑。SHARPnP 能够联合抑制噪声，并恢复目标标记和物体边界；即使曝光时间较短，也能做到这一点，但存在一定局限。它使用同一个冻结的去噪器重建全部 51 个可见光-近红外波段，而迁移后的神经网络基线则仅处理 31 个可见光波段子集。材料光谱呈现出独特的可见光-近红外响应，且未进行单独归一化。由于没有可用的干净高分辨率高光谱图像参考，这些原生结果仅作定性评估。

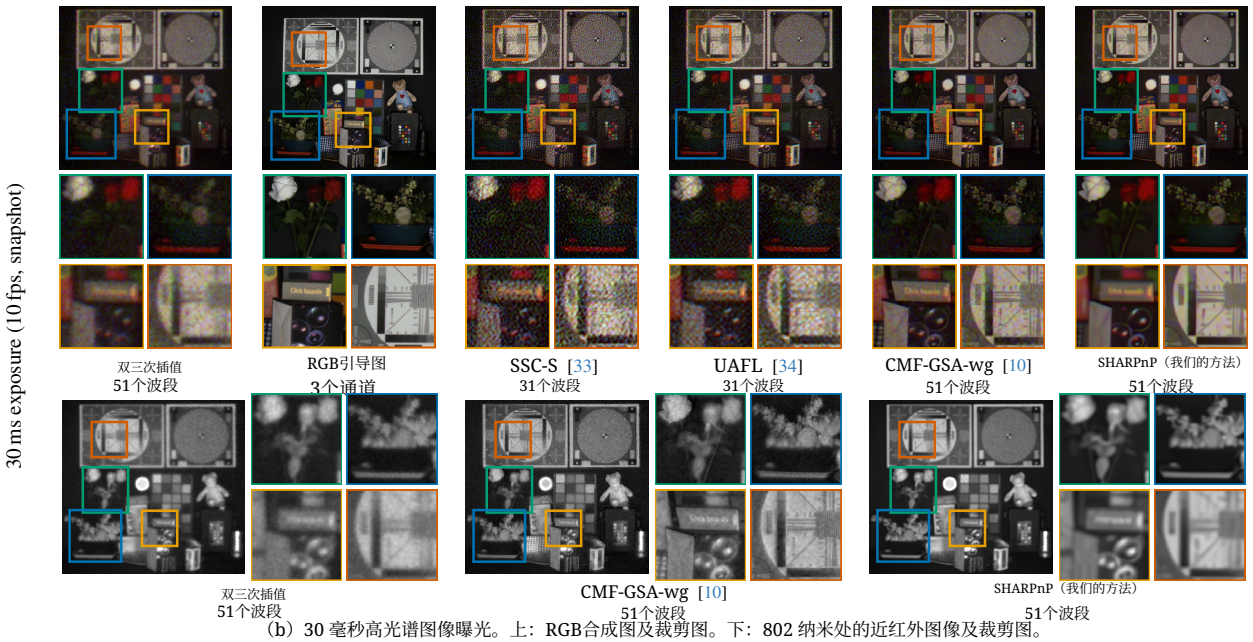
在近红外领域，SHARPnP仅在有限程度上应用超分辨率，同时也谨慎避免从错误的通道注入细节。例如，左上方的裁剪图显示了三朵玫瑰，其中两朵是仿制花，而中间的一朵是真花；其花茎和叶片在近红外领域表现出更高的反射率。在近红外图像的裁剪区域中可以观察到这一点：SHARPnP重建出了真花的花茎，而CMF-GSA-wg则将三朵花的花茎全部投影出来。场景右侧的黑色物镜也是一个有趣的物体，它在RGB预览中呈黑色，但在近红外预览中具有反射性，这再次表明，SHARPnP能够在一定的超分辨率能力范围内保留近红外特征。

6. 结论

我们提出了SHARPnP，这是一种从带噪且未对齐的相机测量中进行RGB引导的高光谱超分辨率的自监督即插即用框架。SHARPnP不需要高分辨率高光谱参考数据。它通过重建反馈对预训练配准模型进行适配，并将校准后的RGB引导信息与冻结的去噪先验进行融合。在REAL 数据集上，它达到了当前最佳水平，在所有尺度上均超过了有监督融合方法，并在 $4\times$ 达到 43.37 dB，而最强基线为41.91 dB。在带噪图像对上进行自监督微调后，该方法能够处理含噪输入，在 $\sigma = 5/255$达到39.33 dB，几乎比直接有监督融合高出7 dB。在原生快照成像装置上，它能够重建全部 51个VIS–NIR波段。本研究的一个重要发现是，即使在RGB引导信息有限的区域，NIR信息也有助于重建空间细节。这一观察结果促使我们在未来改进NIR范围内的光谱一致性，确保重建出的细节忠实于底层光谱信息。此外，我们计划通过加速其测试时优化，将SHARPnP扩展到实时高光谱视频超分辨率，从而更高效地重建动态场景和运动物体。



（a）480毫秒高光谱成像曝光。上：RGB合成图及裁剪图。下：802 纳米处的近红外图像及裁剪图。



（b）30 毫秒高光谱图像曝光。上：RGB合成图及裁剪图。下：802 纳米处的近红外图像及裁剪图。

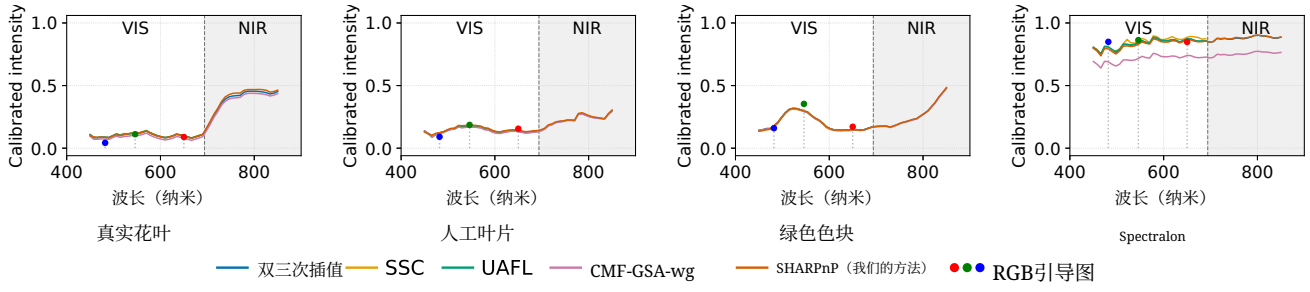


图4. Ultris-RPi 在 480 和 30 毫秒曝光下的原生重建。RGB 列：双三次插值、RGB 引导图、SSC [33], UAFL [34], CMF- GSA-wg [10], 和 SHARPnP（我们的方法）。近红外部分展示了双三次插值、CMF-GSA-wg 和 SHARPnP，仅适用于 51个波段重建。裁剪图展示了真实/仿真玫瑰（花瓣和叶片）、真实仙人掌、视差区域和空间靶标。没有可用的干净高分辨率参考图像。

补充材料

本补充材料分为四个主要部分：(1) 额外的实现细节，包括配准过程和重建设置；(2) 额外的消融研究；(3) 额外的 REAL 对比；最后是 (4) 原生 RGB-高光谱成像数据上的定性和定量结果，其中我们在 480 毫秒和 30 毫秒下重新考察主论文第 5.3 节中的场景，并评估 Jouni等人此前使用的第二个场景。 [10]。代码实现将以开源形式提供。

A. 额外的实现细节

A.1. 配准微调设置

微调会更新选定的 CrossRAFT 参数 ϕ，而去噪器保持冻结。我们使用 AdamW，学习率为 10^{-5}，权重衰减为 10^{-6}，梯度裁剪值为 1，随机种子为 2022年。REAL 微调使用 470 次更新，并匹配 256×256 裁剪区域。批归一化统计量保持不变。适配器微调会冻结主干网络（RAFT）；主干网络微调会冻结适配器。

带噪图像对。 (1) *REAL*基准数据： 对于含噪输入，正文中的 GR2R 使用 $\tau \simeq 0.42$。对于无噪输入，微调时两个视图使用相同的观测。(2)我们的原生*RGB-HSI*数据： Noise2Noise 图像对取自视频采集的相邻帧；该视频是在给定噪声配置下对静态场景采集的。在正文第 5.3 节中，分别针对 480 毫秒和 30 毫秒曝光进行微调。对于动态场景，可以先使用在相同采集设置下采集的静态场景数据进行微调，然后将其用于动态序列的逐帧重建。

A.2. 重建设置

置信度加权方向 VTV（dVTV）。 我们采用置信度加权的方向TV [3],，通过向量TV [2]耦合不同光谱波段的空间梯度。该方法在配准的RGB引导图像的可靠边缘处减少平滑，同时对不同光谱波段的空间变化进行联合正则化。

记 $X \in \mathbb{R}^{n\times B}$ 为重建的高光谱图像，其中 n和 B 分别表示输出像素和光谱波段， $L \in \mathbb{R}^{n\times 1}$ 为配准后的RGB引导图 $R_w \in \mathbb{R}^{n\times 3}$的灰度图像， C_P 为置信度图， $\nabla X\,(p) \in \mathbb{R}^{2\times B}$ 为像素 $p \in \{1,\dots,n\}$处所有波段的水平与垂直梯度。我们将dVTV正则项 $\mathcal{V}\,(X; R_w, C_P)$ 定义为：

$$\xi_p = \frac{\eta\sqrt{C_P(p)}\ \ \nabla L(p)}{\sqrt{\|\nabla L(p)\|_2^2 + \delta^2}},\ H_p = I_2 - \xi_p\xi_p^\top. \qquad \text{(S1)}$$

$$\mathcal{V}(X; R_w, C_P) = \frac{1}{n\sqrt{B}}\sum_p \|H_p \nabla X(p)\|_F, \qquad \text{(S2)}$$

其中 η 为方向强度， δ 为防止除以零的梯度稳定项 [3]， $\xi_p \in \mathbb{R}^{2\times 1}$ 为置信度加权的方向向量， $I_2 \in \mathbb{R}^{2\times 2}$ 为2×2 单位矩阵， $H_p \in \mathbb{R}^{2\times 2}$ 为作用于水平与垂直梯度分量的方向矩阵， $\|\cdot\|_F$ 表示Frobenius范数。我们使用$\eta = 0.99$ 和 $\delta = 0.01$。为了数值稳定性，我们使用 $\epsilon = 0.001$ 平滑TV惩罚项，以避免在均匀区域 [1]中出现除以零。

去噪配置。 对于去噪，我们在HQS中使用冻结的RGB DRUNet配置，并采用DPIR [29]参数化 [23]。对于Clean REAL重建，我们在后续迭代中减弱去噪校正的强度。对于含噪输入，我们保留完整的校正。对于我们的原生RGB-HSI数据，去噪器噪声水平根据每次曝光时的时间噪声估计固定。

B. 附加消融实验

RGB配准消融实验。 表 S1 评估了SHARPnP在十个干净REAL场景上的三个组件，使用第50次迭代且未重新调参$4\times$，$8\times$和 $16\times$，在 $4\times$ 和 $16\times$处使用第50次迭代且未重新调参。移除残差对齐后，峰值信噪比分别下降2.09、6.59和1.50 dB。将拟合的辐射度映射替换为固定的三波段代理，同时保留几何结构和有效RGB权重，会导致峰值信噪比分别下降0.64、2.36和3.41 dB。因此，这两个组件在所有测试尺度上都带来了显著收益。CrossRAFT微调的影响较小且取决于尺度：在 $4\times$ 处提升峰值信噪比0.05 dB，在 $16\times$处提升0.40 dB，但在 $8\times$处降低0.14 dB，同时略微改善光谱角映射。因此，当发布的权重已经提供有效的初始对应关系时，任务特定微调并非始终必要。这些发现支持优先进行辐射度校准和残差对齐，同时将前端微调视为取决于设置的改进。

表S1. 在三个尺度下对Clean REAL进行SHARPnP组件消融实验。每个区块报告十个场景在第50次迭代时、未经重新调参的平均结果。粗体表示每个区块内未四舍五入的最佳得分。

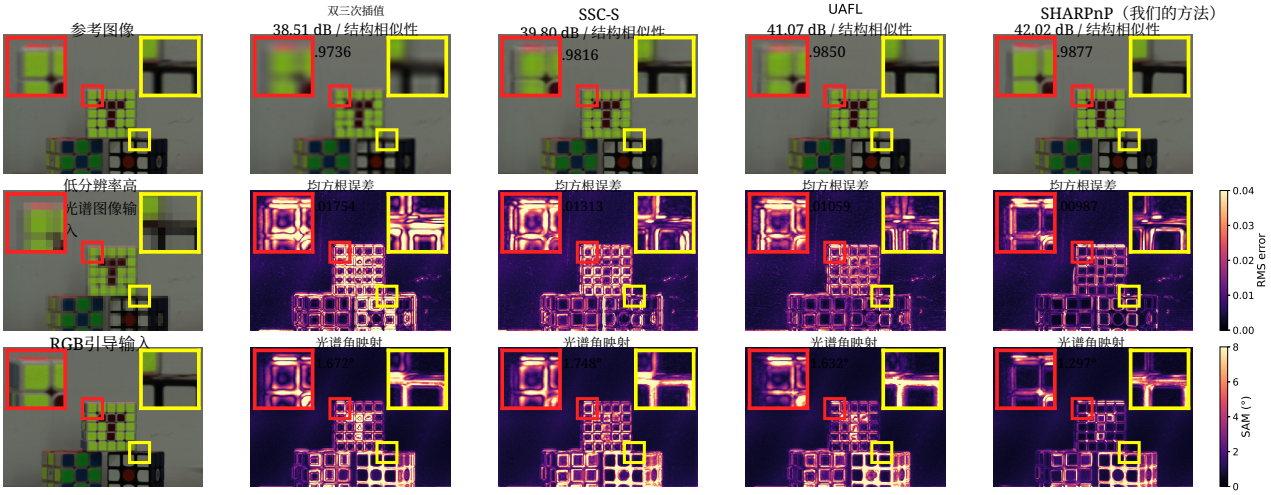
尺度	变体	峰值信噪比↑ (dB)	结构相似性指数↑	SAM↓ (°)	均方根误差↓ (10^{-2})
REAL $4\times$	SHARPnP	**43.35**	**0.991**	**1.596**	**0.7736**
	无 CrossRAFT适应	43.30	0.991	1.596	0.7780
	无辐射校准	42.71	0.989	1.689	1.0806
	无残差对齐	41.26	0.983	1.726	1.1240
REAL $8\times$	SHARPnP	37.37	0.974	**2.278**	1.6437
	无CrossRAFT适应	**37.51**	**0.974**	2.283	**1.6178**
	无辐射校准	35.01	0.950	2.728	2.3267
	无残差对齐	30.78	0.884	3.443	4.0326
REAL $16\times$	SHARPnP	**32.37**	**0.947**	**3.126**	**2.9777**
	不使用CrossRAFT适应	31.97	0.942	3.198	3.1175
	不使用辐射校准	28.96	0.887	5.084	4.9438
	不使用残差对齐	30.87	0.920	3.354	3.6917

其在 16× 处的四项指标均有所提升，证明了其在这一具有挑战性的场景中的价值。

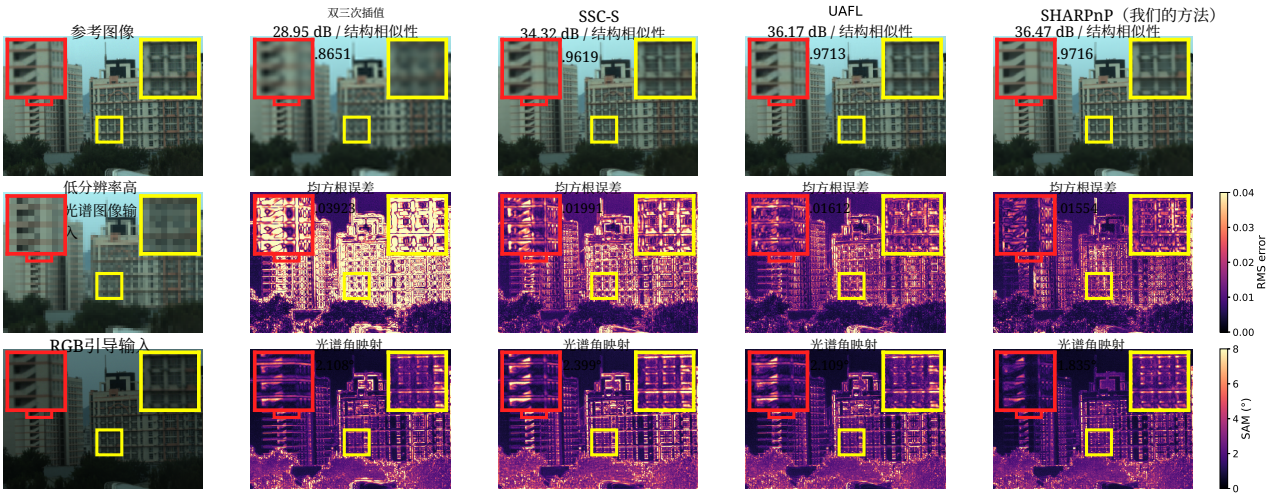
C. REAL 数据集中的更多结果

Clean x8. 图 S1通过两个 Clean 8×示例补充了正文，其中包含一个比 4×更具挑战性的案例。场景 59 展示了从两个视角拍摄的近距离魔方，其中一些面仅在一个相机视图中可见。SHARPnP 在减少边缘误差的同时，更好地保留了平滑区域。场景 35 展示了阳台和窗户等精细的户外细节，这些细节在低分辨率高光谱图像输入中模糊不清。SHARPnP 在四项指标上均取得最佳分数。

10/255 噪声 x4。 图 S2展示了两个 4× 示例，其中 10/255 低分辨率高光谱图像中加入了噪声。场景 0 在更强 10/255 噪声下重新展示了主论文中的 Clean 示例，而场景 31 则提供了另一个示例。我们比较了双三次插值、直接 UAFL [34]，以展示将有监督融合直接应用于含噪输入的效应；还比较了使用与 SHARPnP 中相同的 DRUNet 去噪器的先去噪后融合基线，以及 SHARPnP 本身。场景 0 展示了建筑窗户和深色植被。直接 UAFL 保留了明显的噪声，而在融合前进行去噪则显著改善了重建结果。SHARPnP 获得了最佳的峰值信噪比、结构相似性指数和均方根误差，但其光谱角映射略高于 DRU 基线。场景 31 展示了平滑的墙面瓷砖以及细长的自行车结构。SHARPnP 在降低较暗自行车区域噪声的同时，保留了瓷砖边界。在四项指标上，它均取得了最佳分数。

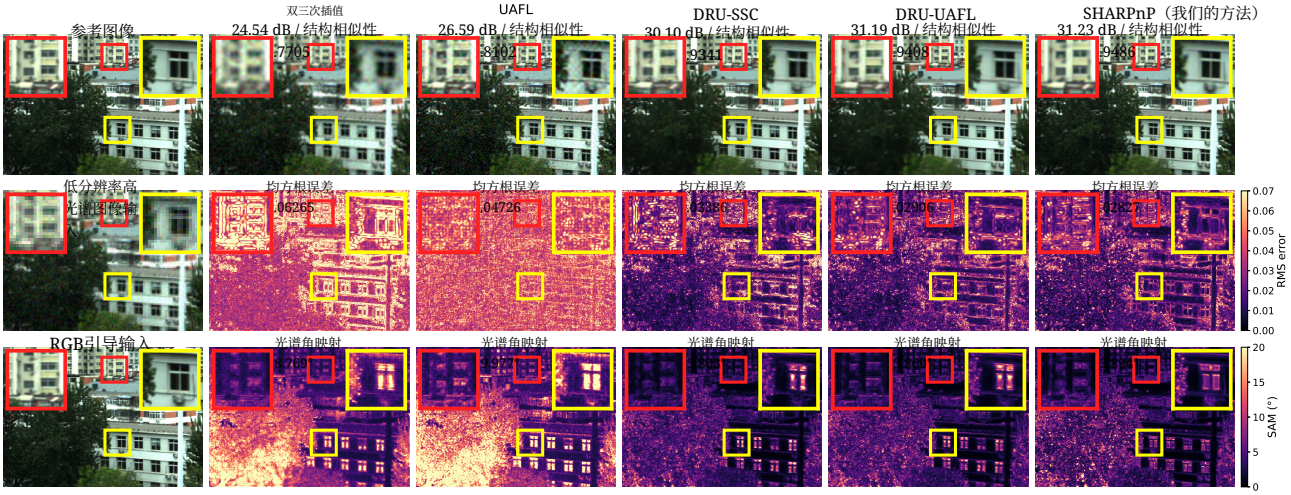


(a) 真实场景 59；Clean； 8×

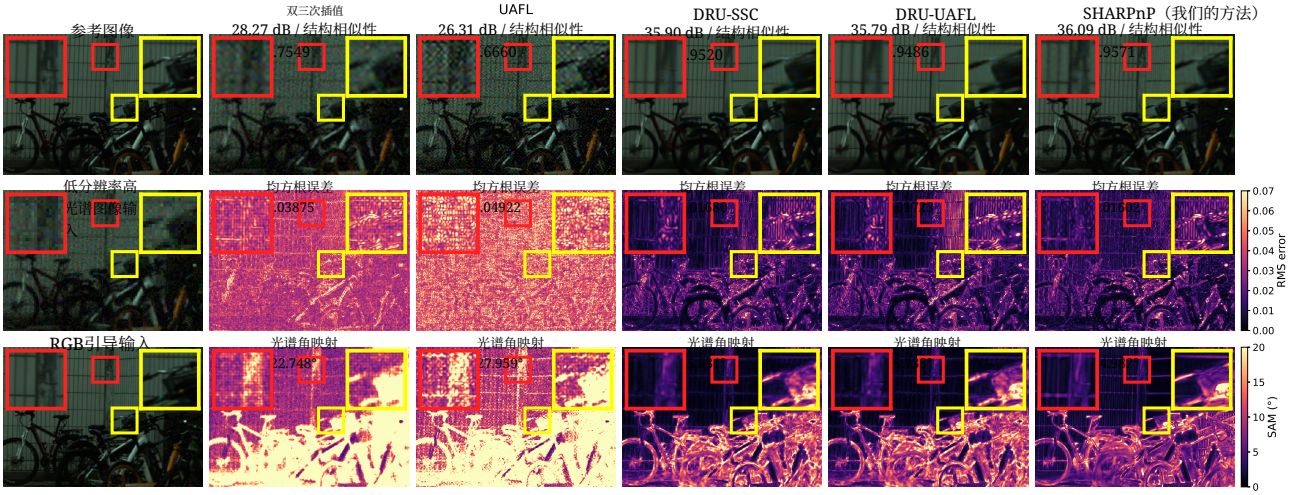


(b) 真实场景 35；Clean； 8×

图 S1. 使用双三次插值、SSC-S [33], UAFL [34], 和 SHARPnP（我们的方法）进行的 Clean REAL 对比，展示于 8× 。各行分别显示 RGB 预览图、均方根误差以及与 Clean 高分辨率参考图像相比的光谱角映射图。所有方法的局部放大图均采用相同的区域和比例尺。标注为基于所显示预测结果计算的全帧指标。



（a）REAL 场景 0； 10/255 低分辨率高光谱图像噪声； 4×



（b）REAL 场景 31； 10/255 低分辨率高光谱图像噪声； 4×

图 S2。在 4× 和 10/255 采用双三次插值、UAFL [34], DRU-SSC、DRU-UAFL 和 SHARPnP（本文方法）的含噪 REAL 对比。UAFL 直接接收含噪的低分辨率高光谱图像。DRU 表示先去噪后融合的串行处理，即先使用 DRUNet 进行去噪，再使用冻结的有监督融合模型。SHARPnP 在场景 31 中提升了全部四项全帧指标；而场景 0 虽然峰值信噪比、结构相似性指数和均方根误差均有所改善，但光谱角映射略有权衡。各行分别展示 RGB 预览图、相对于 Clean 高分辨率参考图像的均方根误差和光谱角映射图。

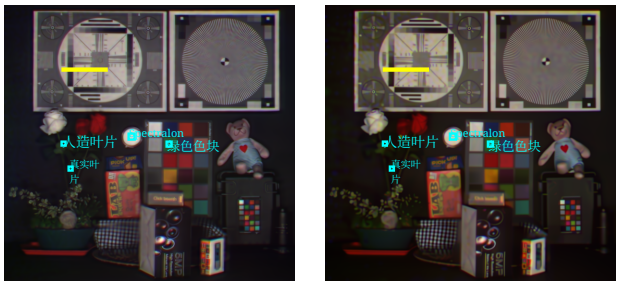
D. 原生RGB-高光谱成像数据中的其他结果

本节通过定量分析和一个额外的真实原生 RGB-高光谱成像场景，进一步扩展真实原生 RGB-高光谱成像实验部分。在第 D.1 节中，我们重新考察了主论文第 5.3 节中在两种曝光时间（480 毫秒和 30 毫秒）下的场景，并在没有 Clean 高分辨率参考图像的情况下，对其结果进行了定量分析。在第D.2 节中，我们补充了另一个真实原生场景的结果，该场景用于 [10], , 同时采用与主论文第 5.3 节数据相同的定性和定量分析方式。

两种采集方式都无法提供干净的高分辨率高光谱图像参考。在高分辨率网格上，我们依赖 con-

控的空间和光谱目标，来检验各方法在最终超分辨率输出中恢复空间和光谱细节的效果。对于报告的全图指标，分析是在低分辨率高光谱图像网格上进行的。具体而言，我们对每个超分辨率输出进行模糊和下采样，使其达到观测到的低分辨率高光谱图像网格，然后与同一场景中时间平均后的高光谱图像观测进行比较。这样做是为了减弱单帧观测高光谱图像中的噪声影响，并更多地关注空间和光谱结构。

我们分别在 31 个可见光波段和 20 个近红外波段上报告均方根误差和光谱角映射。我们还报告所谓的邻域光谱角映射（归一化光谱角映射）和 Spectralon 变异（光谱-



（a）480 毫秒　（b）30 毫秒

图 S3。黄色标记用于目标剖面的条带。青色标记用于获取光谱的四个区域。

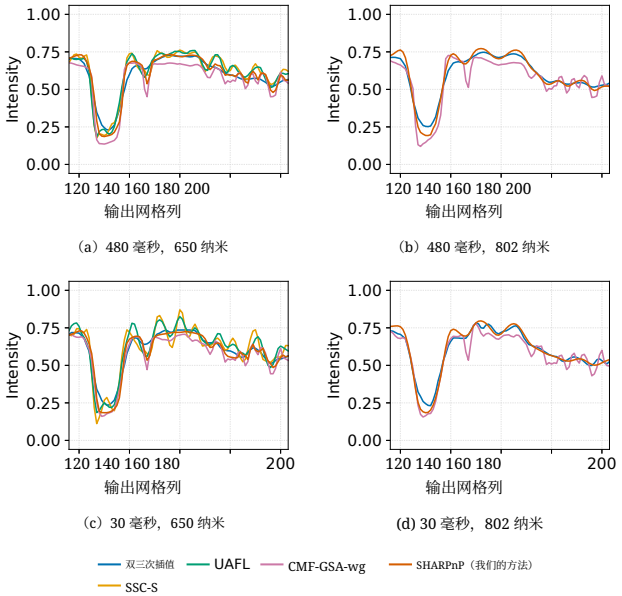


图 S4. 相同目标在两种曝光下的强度。每条曲线为四个图像行的平均值。

方差）。对于归一化光谱角映射，我们在实验中使用的每个 ColorChecker 色块内选择一个小邻域（色块），并在白色 Spectralon 表面上选择一个邻域。然后，我们对每个邻域内的光谱取平均值，并计算测量可见光谱与重建可见光谱之间的夹角，从而为每个邻域得到一个角度。我们报告这些角度的中位数。对于光谱-方差，我们在白色 Spectralon 表面上选择一个均匀邻域，该表面应当平坦且平滑。我们计算该邻域内像素强度的标准差，以衡量重建白色表面在不同像素之间的变化程度。

D.1. 不同曝光时间

我们重新考察主论文第 5.3 节中的静态场景，分别在 480 毫秒和 30 毫秒曝光下进行。在每种曝光下，我们对另外十帧高光谱成像帧取平均。此处归一化光谱角映射使用 20 个可见的 ColorChecker色块以及 Spectralon。图S3 标出

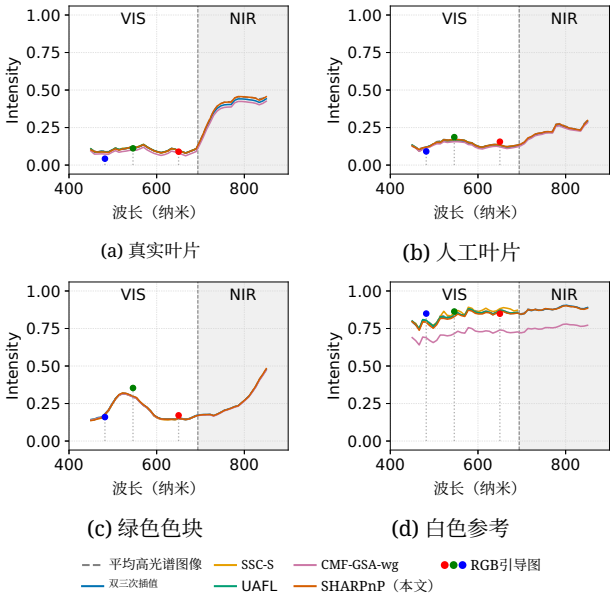


图 S5。480 毫秒曝光下的色块光谱。虚线表示十个高光谱成像帧的平均值。

表 S2。静态场景，曝光时间为 480 毫秒。

方法	VIS [450, 690] 纳米		NIR [698, 850] 纳米		可见光图像块	
	均方根误差↓ (10^{-2})	光谱角映射↓ (°)	均方根误差↓ (10^{-2})	光谱角映射↓ (°)	归一化光谱角映射↓ (°)	光谱方差↓ (10^{-2})
双三次插值	1.891	1.879	2.087	1.182	0.711	1.431
SSC-S [33]	2.240	2.298	–	–	1.330	2.424
UAFL [34]	2.131	1.972	–	–	0.882	2.137
CMF-GSA [10]	2.316	2.106	2.544	1.287	1.069	**1.209**
SHARPnP	**0.815**	**1.580**	**0.787**	**0.861**	**0.702**	1.388

目标条带以及用于光谱的四个区域。

空间细节。 总体而言，SHARPnP 在空间剖面附近的锐度与过冲之间实现了最干净的折衷。例如，在 650 纳米处，双三次插值在两种曝光下都会将从亮到暗的过渡扩散到约六个像素。SHARPnP 则将其缩窄至约四个像素（图 S4）。SSC-S 能产生更锐利的边缘，但在 30 毫秒时，其强度在边缘处的过冲约为亮暗强度差的 24%。SHARPnP 对应的过冲则低于 1%。

光谱一致性。在 480 毫秒时，SHARPnP 在可见光和近红外波段都与平均高光谱成像测量最为接近，均方根误差和光谱角映射值最低（表S2）。CMF-GSA 得到的 Spectralon 色块最均匀，但低估了其亮度。在 802 纳米处，测得的强度为 0.90，而 CMF-GSA 和 SHARPnP 的强度分别为 0.78 和 0.90（图S5）。

表 S3. 静态场景，30 毫秒曝光。

方法	VIS [450, 690] 纳米		NIR [698, 850] 纳米		可见光图像块	
	均方根误差↓	光谱角映射↓	均方根误差↓	光谱角映射↓	归一化光谱角映射↓	光谱方差↓
	(10^{-2})	(°) (10^{-2})		(°)	(°) (10^{-2})	
双三次插值	2.759	7.983	2.254	2.397	4.396	5.077
SSC-S [33]	3.056	8.506	–	–	4.544	8.099
UAFL [34]	3.087	8.599	–	–	4.903	7.743
CMF-GSA [10]	3.073	8.385	2.728	2.471	5.150	4.916
SHARPnP	**2.201**	**6.957**	**1.389**	**2.152**	**3.622**	**1.646**

在 30 毫秒时，SHARPnP 在全部六项指标上的数值均最低（表S3）。差异最大的是 Spectralon 色块：该白色斑块内的变化幅度约为次优方法 CMF-GSA 的三分之一。SHARPnP 还保留了两片叶片之间的差异（图S6）。在 802 纳米处，它重建出的真实叶片和人工叶片强度分别为 0.45 和 0.25，与测量值 0.46 和 0.26 接近。

D.2. 第二个原生场景。

实验设置。我们从 [10] 中描述的静态 Ultris SR5 和 Raspberry Pi HQ 序列中选取一帧。SHARPnP 使用 20 次 HQS 迭代和冻结的 DRUNet，从 450 纳米到 850 纳米范围内重建全部 51个波段，其中$P_D = I_{51}$ 为 51×51 恒等映射（单位矩阵）。RGB值和梯度的归一化权重分别为 0.009 和 0.200。

SSC-S 和 UAFL 仅处理前 31个波段，覆盖 450 至 690 纳米的可见光范围。另一种做法是在全部 51个波段上重新训练它们，但这需要干净的高分辨率高光谱图像参考，而这里并没有这些参考。

由于我们没有干净的高分辨率高光谱图像参考，因此基于场景中受控的光谱和空间材料进行定性和定量分析。

图S7、S8和S9 分别展示了可见光重建图像、802 纳米波段和色块光谱。表S4 将重建结果与同一场景的时间平均高光谱图像观测进行了比较，以减少传感器噪声，并聚焦于结构误差和物体边缘。由于平均后的观测仍可能含有噪声，因此仍可能出现一些残差变化。

此示例补充了主论文第 5.3 节中的原生采集。

定性分析。图S7 展示了结果的定性分析。在可见光域中，我们可以看到 RGB 细节在哪些地方有所帮助，以及在哪些地方可能引入伪影。通过在 REAL 数据上进行监督训练后迁移而来的 SSC-S 和 UAFL，无法很好地泛化到真实的原生 RGB-高光谱图像传感器设置。例如，整体图像的属性

表 S4。来自 Ultris-RPi 序列的额外原生帧如 [10]所述。

方法	VIS [450, 690] 纳米		NIR [698, 850] 纳米		可见光图像块	
	均方根误差↓	光谱角映射↓	均方根误差↓	光谱角映射↓	归一化光谱角映射↓	光谱方差↓
	(10^{-2})	(°)	(10^{-2})	(°)	(°) (10^{-2})	
双三次插值	1.378	0.845	1.118	0.625	**0.497**	1.478
HSIFN [12]	2.050	1.405	–	–	1.164	1.734
SSC-S [33]	2.180	1.384	–	–	1.606	1.708
UAFL [34]	1.737	0.899	–	–	0.665	1.728
CMF-GSA [10]	1.720	0.858	1.907	0.631	0.927	1.367
SHARPnP	**0.860**	**0.819**	**0.691**	**0.567**	0.810	**1.275**

会引入一些噪声，呈现模糊且重建效果较差的区域，传播未配准RGB引导图中看似伪影的内容（在显示Spectralon边缘的右上裁剪图中部分可见），在穿过Spectralon边缘的位置呈现微弱的水平线，并在白板下边缘形成额外的轮廓，同时在RMSE图中显示出更大的误差。CMF-GSA-wg能够恢复更干净且细节更丰富的图像，但其降噪能力有限，并采用置信度门控策略，在不可靠区域回退到原始低分辨率高光谱图像（在显示Spectralon边缘的右上裁剪图中部分可见）。SHARPnP呈现出更干净的Spectralon边缘，并恢复出更多清晰的细节和笔画，包括小目标上的灰色阶梯和内环，以及圆形目标的线条。

图 S8 展示了802纳米处的重建。我们

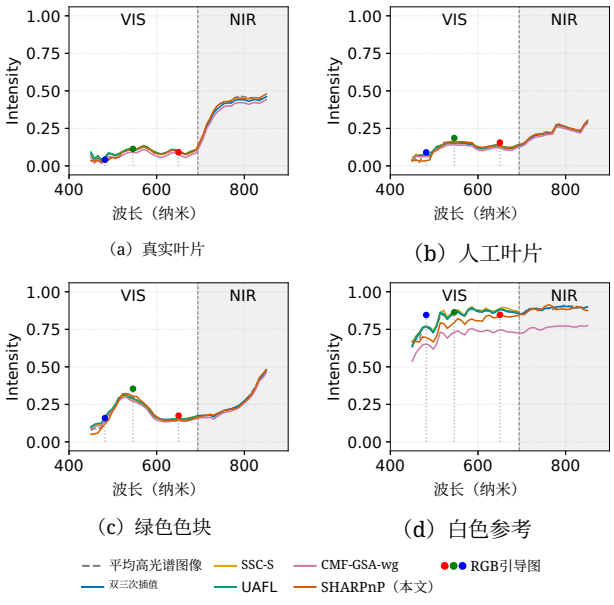


图 S6。30 毫秒时的色块光谱。虚线表示平均高光谱图像帧。

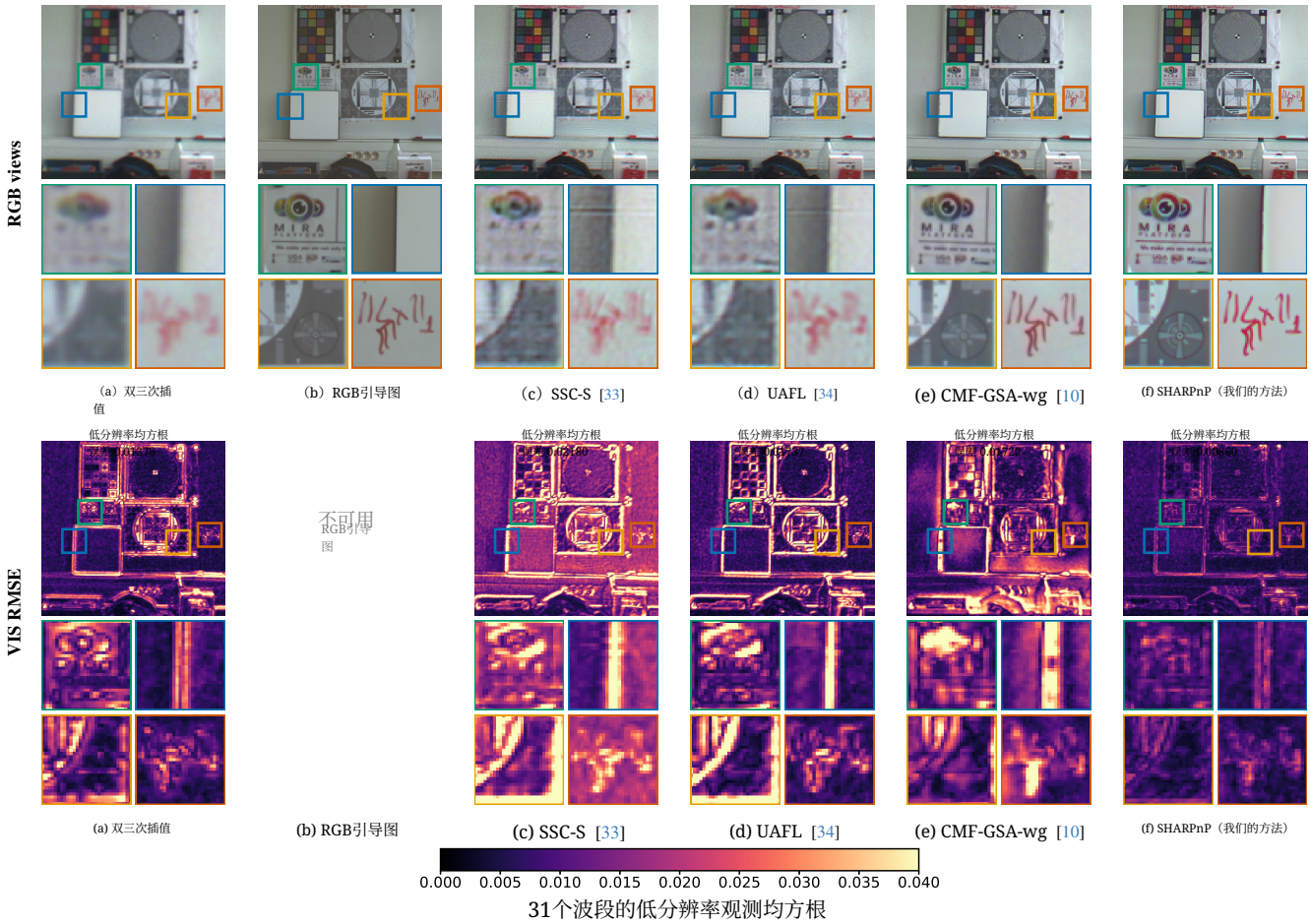


图 S7。来自 CMF-GSA-wg [10]所用装置的另一个 Ultris-RPi 原生示例。各行展示了 RGB 视图和相对于平均高光谱图像观测计算的可见光均方根误差图。裁剪图展示了细小纹理、Spectralon边缘、目标角和墙上文字。没有可用的 Clean 参考图像。

比较双三次插值、CMF-GSA-wg和SHARPnP这三种能够重建全部51个波段的方法。RGB引导图中可见的文字也出现在CMF-GSA-wg的重建结果中，尽管在双三次插值的近红外裁剪图中这些文字较为 faint 或缺失。另一方面，SHARPnP细化了近红外细节，并相较于双三次插值恢复出一些细节；但与CMF-GSA-wg不同，它对可见光域细节投影的抑制程度要高得多。

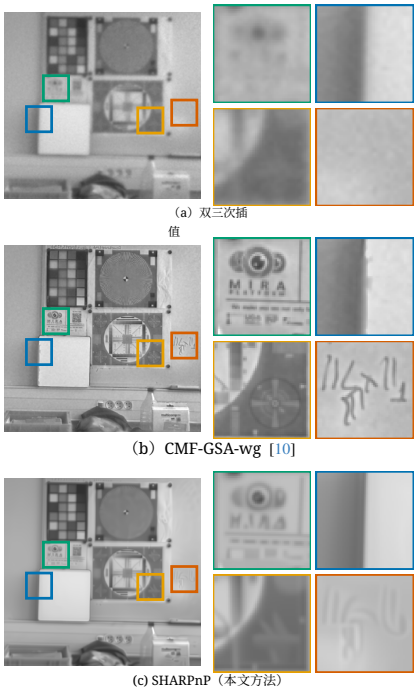
图 S9 中的光谱取自蓝色、绿色和红色ColorChecker色块以及 Spectralon；通过计算像素邻域的平均光谱，减轻单个像素光谱中的噪声影响。这些探测对象之所以被选中，是因为它们是光谱特性已知的受控元素，且覆盖可见光-近红外范围。SHARPnP处理全部51个波段，结果与参考图像接近。

定量分析。 对于表S4，我们对静态场景的四帧高光谱成像帧取平均。图S7 显示了 可见光均方根误差 图。我们报告 ColorChecker 的 24 个色块以及 Spectralon 色块上的中位角。

在两个光谱范围内，SHARPnP的均方根误差和光谱角映射均最低，光谱方差也最低。双三次插值和 UAFL 的归一化光谱角映射低于 SHARPnP；值得注意的是，双三次插值和 UAFL 更好地保留了色块的平均光谱，而 SHARPnP 则能更清晰地恢复字迹和目标环，见图 S7。

参考文献

[1] Peter Blomgren 和 Tony F Chan。《彩色电视：用于向量值图像复原的全变分方法》。*IEEE* 图像处理汇刊，7(3):304–309，1998年。[2] Xavier Bresson、Tony F Chan 等。《向量全变分范数的快速对偶最小化及其在彩色图像处理中的应用》。逆问题与成像，2(4):455–484，2008年。[3] Leon Bungert、David A. Coomes、Matthias J. Ehrhardt、Jennifer Rasch、Rafael Reisenhofer 和 Carola-Bibiane



图S8. 802 纳米处的重建结果。每行显示完整图像和与图S7相同的四个裁剪区域。

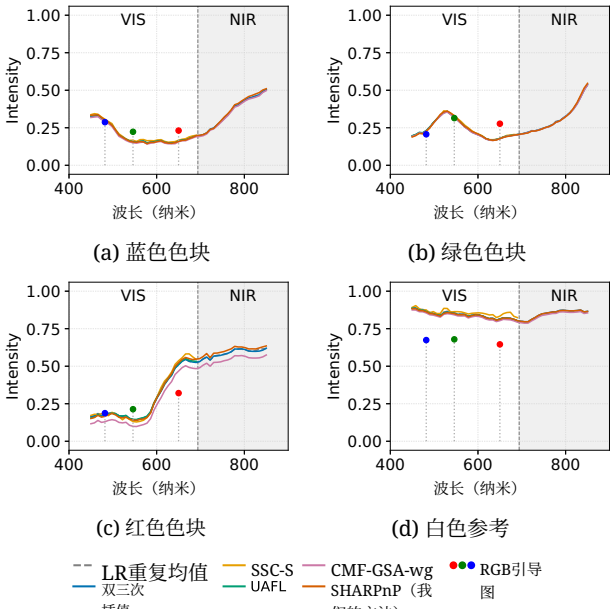


图 S9. 三个 ColorChecker色块和 Spectralon 的平均光谱。虚线表示四个高光谱成像帧的平均值。694 纳米处的分界线将可见光和近红外分开。

Sch¨onlieb。《基于方向全变分的高光谱成像盲图像融合》。*Inverse Problems*，34(4): 044003，2018年。[4] Kebin Contreras、Mohamad Jouni、Mauro Dalla Mura 和 Jorge Bacca. 一种带有理化标注的可可豆多模态高光谱数据集。*Scientific Data*，2025年。[5] David L. Donoho and Iain M. Johnstone. 通过小波收缩实现理想的空间自适应。*Biometrika*，81 455，1994 3 425( ): – 。[6] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang. 采用优化的RGB 引导进行高光谱图像超分辨率。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第11661–11670页，2019年。[7] Y. Fu, Z. Liang, and S. You. 用于高光谱图像超分辨率的双向三维准循环神经网络。*IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*，14:2674–2688，2021年。[8] J. Jiang, H. Sun, X. Liu, and J. Ma. 学习空间-光谱先验，用于高光谱图像超分辨率。*IEEE Transactions on Computational Imaging*，6:1082–1096，2020年。[9] Mohamad Jouni, Daniele Picone, and Mauro Dalla Mura. 多光束干涉光谱学：仪器分析与光谱重建。*IEEE Transactions on Instrumentation and Measurement*，2025年。[10] Mohamad Jouni, Aur´elie n Godet, and Mauro Dalla Mura. 真实跨分辨率错位条件下轻量且可解释的RGB引导高光谱超分辨率。载于*BMVC*，2026年。[11] Zeqiang Lai, Kaixuan Wei, and Ying Fu. 用于高光谱图像复原的深度即插即用先验。*Neurocomputing*，481:281–293，2022年。[12] Zeqiang Lai, Ying Fu, and Jun Zhang. 采用真实未对齐 RGB 引导的高光谱图像超分辨率。*IEEETransactions on Neural Networks and Learning Systems*，36 (2):2999–3011，2025年。[13] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise：无需干净数据的图像复原学习。载于国际机器学习会议论文集，第2965–2974页，2018年。[14] Jiabao Li, Yuqi Li, Chong Wang, Xulun Ye, and Wolfgang Heidrich. BUSIFusion：高光谱图像与 RGB 图像的盲无监督单图像融合。*IEEE Transactions on Computational Imaging*，9:94–105，2023年。[15] Q. Li, Q. Wang, and X. Li. 用于高光谱图像超分辨率的混合2D/3D卷积网络。*Remote Sensing*，12(10):1660，2020年。[16] Y. Liu, R. Dian, and S. Li. 用于高分辨率高光谱计算成像的低秩变换器。*International Journal of Computer Vision*，133(2):809–824，2025年。[17] Y. Liu, J. Liu, R. Dian, and S. Li. 高光谱融合成像的选择性再学习机制。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第7437–7446页，2025年。[18] Dimitris G Manolakis, Ronald B Lockwood, and Thomas W Cooley. 高光谱成像遥感：物理、传感器与算法。剑桥大学出版社，2016年。[19] Brayan Monroy, Jorge Bacca, and Juli´an Tachella. 广义再污染到再污染：超越高斯噪声的自监督学习。载于*CVPR*，第28155–28164页，2025年。

[20] Pai Chet Ng, Zhixiang Chi, Yannick Verdie, Juwei Lu, and Konstantinos N Plataniotis. Hyper-skin：一种用于从 RGB 图像重建面部皮肤光谱的高光谱数据集。神经信息处理系统进展，36:24158–24170，2023年。[21] Subir Paul, Vinayaraj Poliyapram, Nevrez ˙Imamo˘glu, Kuniaki Uto, Ryosuke Nakamura, and D. Nagesh Kumar. 利用高光谱数据上的卷积自编码器预测梨树冠层平均叶绿素含量。*IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*，13:1426–1437，2020年。[22] Daniele Picone, Mohamad Jouni, and Mauro Dalla Mura. 基于干涉采集的光谱-空间高光谱图像重建。载于*ICASSP 2024 - 2024 IEEE* 国际声学、语音与信号处理会议 *(ICASSP)*，第2590–2594页，2024年。[23] Daniele Picone, Mohamad Jouni, and Mauro Dalla Mura. 利用预训练RGB去噪器进行高光谱图像复原。载于*IEEE*国际图像处理会议，2026年。[24] Jiahui Qu, Jizhou Cui, Wenqian Dong, Qian Du, Xiaoyang Wu, Song Xiao, and Yunsong Li. 配准-融合一致性的原理性设计：迈向可解释深度非配准高光谱图像融合。*IEEETransactionson Neural Networks and Learning Systems*，36(5):9648–9662，2025年。[25] Ying Qu, Hairong Qi, Chiman Kwan, Naoto Yokoya, and Jocelyn Chanussot. 基于互协Dirichlet-Net的无监督、非配准高光谱图像超分辨率。*IEEE Transactions on Geoscience and Remote Sensing*，60，2022年。[26] Bernard Schmitt, Zahira Souidi, Fr´ed´erique Duquesnoy, and Fr´ed´eric-Victor Donz´e. 从RGB相机到高光谱成像：新石器时代岩画分析的突破。遗产科学，11(1):91，2023年。[27] Miguel Simoes, Jos´e Bioucas-Dias, Luis B Almeida, and Jocelyn Chanussot. 通过基于子空间的正则化实现高光谱图像超分辨率的凸优化形式。*IEEE Transactions on Geoscience and Remote Sensing*，53 (6):3373–3388，2014年。[28] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDNet：迈向一种快速灵活的基于CNN的图像去噪方案。*IEEE* 图像处理汇刊，2018年。[29] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. 基于深度去噪器先验的即插即用图像复原。*IEEE Transactions on Pattern Analysis and Machine Intelligence*，44 6376，2021年 10 6360( ): – 。[30] Kai Zhang, Yawei Li, Jingyun Liang, Jiezhang Cao, Yulun Zhang, Hao Tang, Deng-Ping Fan, Radu Timofte, and Luc Van Gool. 通过Swin-Conv-UNet和数据合成实现实用的盲图像去噪。机器智能研究，2023年。[31] M. Zhang, C. Zhang, Q. Zhang, J. Guo, X. Gao, and J. Zhang. ESSAformer：用于高光谱图像超分辨率的高效Transformer。载于*IEEE/CVF*国际计算机视觉会议论文集，第23073–23084页，2023年。

[32] Quan Zhang、Jun Li、Weilong Zhu、MINGYANG LI、Qinmu Shen 和 Yuanxi Peng。用于高光谱与多光谱图像配准和融合的结构感知黎曼流匹配。发表于第四十三届国际机器学习会议，2026年。[33] Yingkai Zhang、Zeqiang Lai、Tao Zhang、Ying Fu 和 Chenghu Zhou。利用空间光谱一致性的未对齐RGB引导高光谱图像超分辨率。国际计算机视觉期刊，133:6590–6610，2025年。[34] Yingkai Zhang、Tao Zhang、Jing Nie 和 Ying Fu。通过基于解混的丰度融合学习，提升未配准高光谱图像超分辨率。发表于*CVPR*，第41573–41583页，2026年。[35] Jiangsan Zhao、Jakob Geipel 和 Krzysztof Kusnierek。SparseHSR：从RGB引导到稀疏可见光–近红外传感，实现全范围高光谱超分辨率。*Remote Sensing*，18(21):3823，2026年。[36] K. Zheng、L. Gao、D. Hong、B. Zhang 和 J. Chanussot。NonRegSRNet：一种非刚性配准高光谱超分辨率网络。*IEEE Transactionson Geoscience andRemote Sensing*，60:1–16，2022年。[37] Shili Zhou、Weimin Tan 和 Bo Yan。面向多样化跨模态流估计的单模态光流网络增强方法。发表于*AAAI*人工智能会议，第3562–3570页，2022年。