Lightweight Interpretable RGB-Guided Hyperspectral Super-Resolution under Real Cross-resolution Misalignment

Mohamad Jouni[1]
https://mira-imaging.github.io/mj
Aurélien Godet[1]
aurelien.godet@grenoble-inp.fr
Mauro Dalla Mura[1,2]
mauro.dalla-mura@grenoble-inp.fr

[1] Univ. Grenoble Alpes, CNRS
Grenoble INP, GIPSA-Lab
38000 Grenoble, France
[2] Institut Universitaire de France (IUF)
Paris, France

Abstract

Compact snapshot hyperspectral cameras provide rich instantaneous spectral measurements for ground-level machine vision, but at lower spatial resolution than standard RGB cameras. RGB-guided hyperspectral super-resolution (HSR) addresses this limitation by transferring spatial detail from a high-resolution RGB guide to a low-resolution hyperspectral image (HSI). These dual-camera systems are typically in a horizontal rig geometry, requiring cross-camera image alignment due to different fields of view. However, residual misregistration can inject spurious high-frequency details. Existing learned unaligned-fusion methods are usually trained for a fixed spectral support and spatial scale factors and can be computationally demanding, limiting their flexibility across sensors.

We propose a lightweight and interpretable RGB-guided HSR framework combining cross-modal flow alignment with model-based Gram-Schmidt orthogonalization fusion. The method first warps the RGB guide onto the HSI grid, then estimates an energy-based confidence weight map by measuring local alignment reliability. This map is then used both in a weighted least-squares spectral regression and in a gated fusion between the RGB-guided SR estimate and an HSI-preserving estimate. Unlike existing learned methods, the proposed framework has a low computational footprint and supports different VIS-NIR spectral supports and scale factors without retraining.

Experiments on the Real benchmark show that the proposed method improves reconstruction accuracy over learned fusion baselines while remaining substantially faster. On a 34-frame sequence acquired with our real RGB-HSI dual-camera setup, a reduced-resolution quantitative evaluation validates the method under genuine cross-sensor radiometric, noise, and geometric differences, while native-resolution qualitative results demonstrate deployment on the full 51-band visible-near-infrared acquisition beyond the 31-band setting of existing learned baselines.

1 Introduction

Hyperspectral imaging acquires dense spectral measurements across the electromagnetic spectrum, enabling material-level analysis beyond standard RGB imaging [13]. Compact

© 2026. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

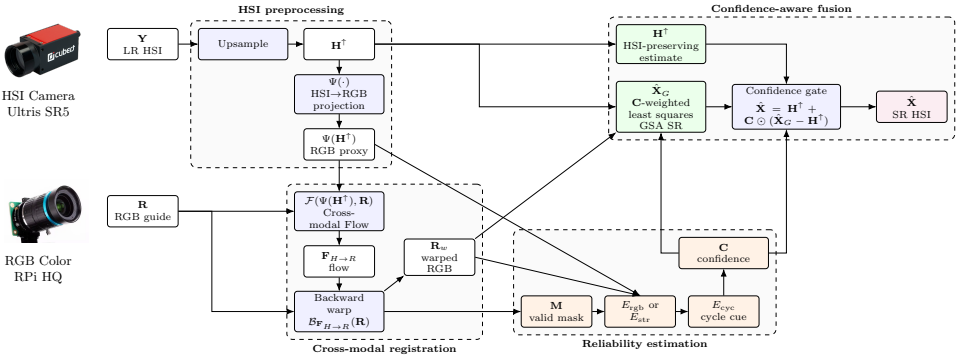


Figure 1: Overview of the proposed confidence-aware RGB-guided HSI super-resolution framework, showing images of the Ultris SR5 HSI and the RPi HQ RGB cameras.

snapshot hyperspectral cameras are particularly attractive for ground-level machine vision because, unlike scan-based counterparts, they can instantly acquire spectral cubes, making them relevant for dynamic applications such as robotic inspection, precision agriculture, food analysis, non-destructive testing, heritage imaging, gas detection, and biomedical imaging [6, 11, 14, 15, 16, 19]. However, for miniaturized and low-cost systems, this additional spectral resolution comes at the cost of spatial resolution, especially for compact or snapshot hyperspectral cameras [10]. A practical solution is therefore to combine a low-resolution (LR) hyperspectral image (HSI) with a high-resolution (HR) RGB image of the same scene, using the RGB image as a spatial guide for HSI super-resolution (HSR). This is a standard configuration in remote sensing [4], but in that case there is basically no parallax.

Most RGB-guided HSR methods assume that the RGB and HSI observations are accurately aligned. This assumption is difficult to satisfy in close-range dual-camera systems. Unlike far-field satellite settings, where viewpoint differences are often negligible at the scene scale, ground-level RGB-HSI rigs use separate sensors with distinct optics, a nonzero baseline, and different fields of view and spectral responses. Even after geometric calibration, the two cameras can observe different scene content near depth discontinuities, producing parallax and occlusions. The problem is further complicated by the cross-resolution nature of real RGB-HSI data. The HSI-derived RGB proxy is smoother and noisier than the RGB observation, which weakens dense correspondences around fine structures.

Recent unaligned RGB-guided HSR methods address this problem with learned alignment and fusion modules [12, 26]. Such methods demonstrate the importance of registration-aware fusion, but they are computationally heavy, lack interpretability, and are tied to the spectral and spatial settings used during training. In particular, standard natural HSI benchmarks commonly use a 31-band visible-domain (VIS) protocol in the range $[400, 700]$ nm [2, 3, 12, 24], but in many applications it would be good to acquire information beyond this range [6, 14, 19]. Applying these models to another number of bands, another spectral range, or another scale factor generally requires retraining or spectral-domain adaptation [17].

In contrast, lightweight model-based fusion methods offer a complementary trade-off. Gram-Schmidt adaptive (GSA) fusion [1] is time-efficient, interpretable, and uses multivariate spectral regression to adapt to different sensing configurations at a low computational footprint. However, such methods typically assume aligned inputs and do not handle locally

unreliable registration. If a warped RGB guide is used without accounting for local registration failures, RGB details may be injected into the wrong HSI pixels, producing ghosting, edge duplication, and spectral distortion, especially with occluded and parallax areas.

This paper investigates a complementary direction. Instead of replacing lightweight model-based fusion with a fully learned reconstruction network, we explicitly model when RGB guidance should be trusted. We build on a cross-modal flow (CMF) estimator [27] to align the RGB image to the HSI grid, and then compute a pixel-wise confidence map from registration validity and local inconsistencies. This confidence controls both the spectral regression used inside GSA and the final amount of detail injected into the HSI estimate. The resulting method is a reliability-aware version of CMF-guided GSA, designed for real RGB-HSI pairs with imperfect cross-resolution alignment. The novelty is not CMF or GSA individually, but their coupling through an explicit reliability map that controls both spectral regression and local detail injection.

Our contributions are as follows. First, we formulate RGB-guided HSI SR under close-range cross-resolution misalignment as a confidence-aware registration-and-fusion problem. Second, we introduce an interpretable energy-based confidence map for warped RGB guidance, combining geometric validity with RGB-HSI consistency cues and bidirectional-flow consistency. Third, we extend GSA with confidence-weighted least-squares (WLS) spectral regression and Tikhonov regularization, reducing the influence of unreliable regions on spectral calibration. Finally, we formulate the reconstruction as confidence-gated detail injection between an HSI-preserving upsampled estimate and an RGB-guided GSA estimate.

2 Related Work

Hyperspectral image super-resolution. Single-image super-resolution (SISR) [17] reconstructs a HR HSI from a single LR HSI. ESSAformer [25], for example, uses an efficient transformer to exploit spectral-spatial correlations. However, SISR remains limited by the lack of observed high-frequency spatial information. Fusion-based approaches instead use an auxiliary HR image, typically RGB, multispectral, or panchromatic data [4]. RGB-guided HSI SR has also been addressed through model-based optimization, such as TV^3 minimization [23]. Model-based HSI fusion methods provide efficient detail injection, while modern deep networks learn spatial-spectral fusion from data. These are mainly developed for remote sensing as this is a largely spread technological configuration.

Ground-level unaligned RGB-guided HSI fusion. Aligned RGB-guided HSR assumes that RGB and HSI pixels correspond to the same scene points, which is difficult to satisfy in real ground-level dual-camera systems. Fewer methods address this configuration compared to that of remote sensing. HSIFN [12] explicitly considers real unaligned RGB guidance by aligning reference features before fusion. SSC-HSR [26] further combines two-stage alignment and feature aggregation with modulation, similarity, and attention-based fusion. These methods demonstrate the importance of registration-aware fusion, but rely on learned black-box pipelines that can be computationally demanding, less interpretable, and tied to training scale factors and spectral supports. Our work instead targets lightweight, interpretable model-based fusion robust to local registration failure and flexible across sensor supports.

Cross-modal dense alignment. Dense optical flow provides pixel-wise correspondences suitable for close-range RGB-HSI systems, where nonrigid misalignments such as parallax

and depth discontinuities cannot be captured by a global affine transform. CMF [27] extends the RAFT [21] dense flow estimator to heterogeneous image pairs through modality-aware processing, making it suitable for images with different sensing characteristics. In this work, we use CMF to estimate correspondences between the HSI-derived RGB proxy and the RGB guide, which exhibit radiometric differences and cross-resolution bandwidth mismatch. Residual flow errors can nevertheless make these correspondences locally unreliable, motivating their explicit reliability assessment before RGB detail is injected.

Registration-aware fusion and confidence. Misalignment-aware guidance also appears in other multimodal problems, including RGB-guided thermal SR with edge-aware or unaligned guidance [8, 9], RGB-guided IR resolution enhancement [22], and recent thermal SR benchmarks [18]. SuperFusion [20], for example, jointly considers registration and fusion using bidirectional deformation fields and symmetric constraints. First, we target HSR from real RGB-HSI VIS-NIR pairs. Second, rather than learning an end-to-end reconstruction, we use registration confidence explicitly as a pixel-wise weight map in spectral regression and in a gated detail injection fusion.

Real RGB-HSI acquisition gap. The closest standard evaluation protocol for unaligned RGB-guided HSR is the "Real" dataset [17]. It uses HSI-HSI pairs from identical scan-based hyperspectral cameras; one HSI is spectrally projected to form the RGB guide and the other is spatially degraded, with evaluation restricted to 31 VIS bands. The pairs are manually cropped to their common support. This provides controlled quantitative evaluation but differs from a native RGB-HSI dual-camera acquisition. We therefore complement it with a real native RGB-HSI acquisition, acquired by our dual-camera setup, which is composed of a Raspberry Pi HQ RGB camera and the Ultris SR5 compact snapshot HSI camera. This setup exposes practical sensor effects absent from the Real protocol. We provide a 34-frame reduced-resolution quantitative evaluation and a native-resolution qualitative example over the full 51-band visible-near-infrared (VIS-NIR) support.

3 Problem Formulation and Notation

We aim to reconstruct a latent HR HSI $\mathbf{X} \in \mathcal{X} \subseteq \mathbb{R}^{H\times W\times C}$, defined on the HSI grid (where H, W, and C are the height, width, and number of spectral bands), from an observed LR HSI $\mathbf{Y} \in \mathcal{Y} \subseteq \mathbb{R}^{h\times w\times C}$ (where $H > h$ and $W > w$) and an observed HR RGB image $\mathbf{R} \in \mathcal{R} \subseteq \mathbb{R}^{H_R\times W_R\times 3}$. We denote the reconstruction by $\hat{\mathbf{X}}$, and the scale factor by $\sigma = H/h = W/w$.

The LR HSI observation is modeled as a spatially degraded version of the latent HR HSI:

$$\mathbf{Y} \approx \mathbf{D}_\sigma \mathcal{L}_\sigma \mathbf{X}, \tag{1}$$

where $\mathcal{L}_\sigma : \mathcal{X} \to \mathcal{X}$ is a spatial blur operator, representing the difference in spatial resolution between LR and HR, and $\mathbf{D}_\sigma : \mathcal{X} \to \mathcal{Y}$ is a downsampling operator.

The RGB observation is related to the latent HSI through spectral projection and geometric correspondence. Let $\phi_{H\to R} : \Omega_H \subset \mathbb{R}^2 \to \Omega_R \subset \mathbb{R}^2$ denote the unknown coordinate map from the HSI grid to the RGB grid. Ideally, for each HSI-grid location $x \in \Omega_H$,

$$\mathbf{R}(\phi_{H\to R}(x)) \approx \Psi(\mathbf{X})(x), \tag{2}$$

where $\Psi : \mathcal{X} \to \mathbb{R}^{H\times W\times 3}$ maps an HSI cube to an RGB proxy. In practice, Ψ is the fixed HSI-to-RGB spectral projection used by the corresponding experimental pipeline.

In practice, $\phi_{H\to R}$ is unknown and is parametrized by a dense flow field $\mathbf{F}_{H\to R} \in \mathbb{R}^{H\times W\times 2}$ between the HSI-derived RGB proxy $\Psi(\mathbf{H}^{\uparrow})$ and the observed RGB image $\mathbf{R}$, where $\mathbf{H}^{\uparrow} \in \mathbb{R}^{H\times W\times C}$ denotes the LR HSI upsampled to the target HR HSI grid, for example by bicubic interpolation, but without the HR spatial details. The flow parametrizes the coordinate map, in the chosen sampling coordinates, as $\phi_{H\to R}(x) = x + \mathbf{F}_{H\to R}(x)$, with $x \in \Omega_H$. Accordingly, the RGB guide warped onto the HSI grid is $\mathbf{R}_w(x) = \mathbf{R}(\phi_{H\to R}(x)) = \mathbf{R}(x + \mathbf{F}_{H\to R}(x))$.

Given $\mathbf{Y}$ and the registered RGB guide $\mathbf{R}_w$, the RGB-guided HSR problem is to reconstruct:

$$\hat{\mathbf{X}} = \mathcal{S}_\sigma(\mathbf{Y}, \mathbf{R}_w), \tag{3}$$

Ideally, it should satisfy the HSI degradation model $\mathbf{Y} \approx \mathbf{D}_\sigma \mathcal{L}_\sigma \hat{\mathbf{X}}$ and, where RGB correspondence is reliable, the registered RGB observation $\Psi(\hat{\mathbf{X}})(x) \approx \mathbf{R}_w(x)$, with $x \in \Omega_H$.

The remaining issue is that $\mathbf{R}_w$ is not uniformly reliable. Even when the sampling coordinate lies inside the RGB image, the correspondence may be incorrect near occlusions, depth discontinuities, non-overlapping regions, or weak cross-modal texture. The proposed method therefore estimates both a warped RGB guide and a lightweight confidence map $\mathbf{C} \in [0,1]^{H\times W}$, used as a reliability map by the reconstruction operator $\mathcal{S}_\sigma$.

4 Methodology

Figure 1 summarizes our instantiation of the reconstruction operator $\mathcal{S}_\sigma$. The method first estimates a dense HSI-to-RGB correspondence and back-warps the RGB image to the HSI grid. It then estimates the reliability of the warped RGB guide, uses this confidence inside the GSA spectral regression, and finally gates the RGB-guided detail residual.

4.1 Cross-modal flow alignment

We estimate the coordinate map $\phi_{H\to R}$ through its flow representation, using a dense flow estimator $\mathcal{F}$ applied to the HSI-derived RGB proxy and the observed RGB guide:

$$\mathbf{F}_{H\to R} = \mathcal{F}(\Psi(\mathbf{H}^{\uparrow}), \mathbf{R}), \qquad \mathbf{F}_{H\to R} \in \mathbb{R}^{H\times W\times 2}, \tag{4}$$

where $\mathcal{F}$ is practically instantiated by CMF/CrossRAFT [27], a recent cross-modal optical-flow estimator designed to promote single-modal optical-flow networks to heterogeneous image pairs. Other dense registration modules could be used in the same formulation.

The RGB image is then resampled on the HSI grid by backward warping:

$$\mathbf{R}_w = \mathcal{B}_{\mathbf{F}_{H\to R}}(\mathbf{R}) \quad \Longleftrightarrow \quad \mathbf{R}_w(x) = \mathbf{R}(x + \mathbf{F}_{H\to R}(x)), \quad x \in \Omega_H, \tag{5}$$

where $\mathcal{B}_{\mathbf{F}_{H\to R}}$ samples $\mathbf{R}$ at the RGB-grid coordinates $x + \mathbf{F}_{H\to R}(x)$ for each $x \in \Omega_H$, using bilinear interpolation in practice. A direct registration-and-fusion baseline is therefore

$$\hat{\mathbf{X}}_{\text{CMF-GSA}} = \text{GSA}_\sigma(\mathbf{H}^{\uparrow}, \mathbf{R}_w). \tag{6}$$

This baseline assumes that every valid warped RGB pixel is equally trustworthy. The proposed method replaces this uniform-trust assumption with a spatial confidence map. Dense flow handles spatial parallax, while true occlusions and non-overlapping fields of view cannot be registered and are instead handled by the validity mask and reliability model below.

4.2 Confidence map and reliability cues

We estimate a confidence map $\mathbf{C} \in [0,1]^{H\times W}$ that measures whether the warped RGB guide should be trusted locally during fusion. The principle is to assign high confidence only to pixels that are geometrically valid and locally consistent with the HSI-derived RGB proxy. As shown in the following, the confidence is not a learned uncertainty model; it is an energy-based reliability score used as a spatial weight.

Validity mask. A binary geometric validity mask $\mathbf{M} \in \{0,1\}^{H\times W}$ is obtained from the warping grid. For each $x \in \Omega_H$, we set $\mathbf{M}(x) = 1$ if the sampled coordinate $x + \mathbf{F}_{H\to R}(x)$ lies inside the RGB image domain Ω_R, and $\mathbf{M}(x) = 0$ otherwise. This mask removes pixels outside the intersection of the fields of view of the two cameras. Equivalently, the validity mask $\mathbf{M}$ is obtained as follows:

$$\mathbf{M}(x) = \begin{cases} 1, & x + \mathbf{F}_{H\to R}(x) \in \Omega_R, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

RGB proxy residual. The first cue compares the HSI-derived RGB proxy with a low-pass version of the warped RGB guide, so that both terms have comparable spatial bandwidth:

$$E_{\text{rgb}}(x) = \frac{1}{3}\sum_{c=1}^{3} \left| \Psi(\mathbf{H}^{\uparrow})_c(x) - \mathcal{L}_R(\mathbf{R}_w)_c(x) \right|. \tag{8}$$

Here, $\mathcal{L}_R$ denotes the RGB low-pass operator used to match the bandwidth of the HSI proxy. We use an ℓ_1 residual because it is less sensitive than a squared residual to local radiometric mismatch and outliers. This cue is appropriate when the proxy and guide are radiometrically comparable, as on Real where both are derived from HSI-domain data.

In real snapshot dual-camera systems, independent sensor responses, exposure, and high HSI noise can keep E_{rgb} high even under correct alignment, as for Ultris-RPi. We therefore use a structure-consistency residual based on local zero-normalized cross-correlation (ZNCC) [7]. We denote two scalar images $U = \mathcal{G}_s\big(\ell(\Psi(\mathbf{H}^{\uparrow}))\big)$, and $V = \mathcal{G}_s(\ell(\mathbf{R}_w))$, where $\ell(\cdot)$ is a fixed RGB-to-luminance conversion and $\mathcal{G}_s$ is a Gaussian smoothing operator. For two scalar images U and V, the local ZNCC over a window Ω_x is:

$$\text{ZNCC}_{\Omega_x}(U,V) = \frac{\sum_{u\in\Omega_x}(U(u)-\mu_{U,x})(V(u)-\mu_{V,x})}{\sqrt{\sum_{u\in\Omega_x}(U(u)-\mu_{U,x})^2}\sqrt{\sum_{u\in\Omega_x}(V(u)-\mu_{V,x})^2+\varepsilon}}, \tag{9}$$

where $\mu_{U,x}$ and $\mu_{V,x}$ are local means over Ω_x, and $\varepsilon > 0$ avoids numerical instability. The structure residual is then:

$$E_{\text{str}}(x) = \mathbf{M}_{\text{str}}(x)\left[1 - \max\left(0, \text{ZNCC}_{\Omega_x}(U,V)\right)\right], \tag{10}$$

where $\mathbf{M}_{\text{str}}(x) = 1$ only when both local standard deviations are above a minimum threshold, and $\mathbf{M}_{\text{str}}(x) = 0$ otherwise. Thus, flat or low-texture regions are treated as uninformative rather than unreliable.

Cycle consistency. If bidirectional flows are available, a cycle-consistency error can complement the above cues:

$$E_{\text{cyc}}(x) = \left\| \mathbf{F}_{H\to R}(x) + \mathcal{B}_{\mathbf{F}_{H\to R}}\left(\mathbf{F}_{R\to H}\right)(x) \right\|_2. \tag{11}$$

This cue penalizes forward-backward disagreement and is expected to identify occlusions and inconsistent correspondences.

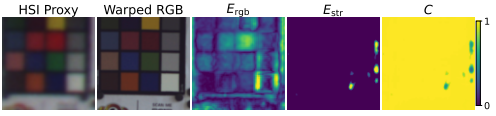


Figure 2: Confidence cues on a native Ultris-RPi crop. From left to right: HSI proxy, warped RGB, E_{rgb}, E_{str}, and confidence C. The residual maps are percentile-normalized for visualization, while C is shown on its native $[0,1]$ scale.

Confidence map. We define the local reliability energy as:

$$\mathcal{E}(x) = \alpha E_{\text{rgb}}(x) + \beta E_{\text{str}}(x) + \gamma E_{\text{cyc}}(x), \tag{12}$$

where $\alpha, \beta, \gamma \geq 0$ control the rejection strengths of raw RGB-proxy inconsistency, local structure inconsistency, and cycle-flow disagreement, respectively. These parameters are selected on the validation set when a reference is available, or fixed for real-device evaluation. The confidence is:

$$\mathbf{C}(x) = \mathbf{M}(x) \exp[-\mathcal{E}(x)]. \tag{13}$$

Although the gate is pixel-wise, E_{rgb} uses bandwidth-matched images and E_{str} is computed over a smoothed local window, so confidence is spatially supported; sharp drops are expected mainly near local misregistration or parallax, where suppressing RGB detail prevents inconsistent structures from being transferred. A native Ultris-RPi example of these cues and the resulting confidence is shown in Figure 2. The exponential maps additive reliability penalties to a multiplicative weight in $[0,1]$; each cue can only decrease the trust assigned to the warped RGB guide. This confidence is not a calibrated posterior probability; it is used as a spatial precision weight for regression and as a gate for detail injection.

4.3 Confidence-weighted Gram-Schmidt adaptive fusion

We propose a method based on GSA, which is a well known technique for the fusion of LR multispectral (e.g., RGB) and HR panchromatic (i.e., single band) satellite remote sensing images [1]. Specifically, we extend the GSA method to address RGB-HSI fusion. The proposed GSA method estimates the relation between the HSI spectral bands and the low-pass RGB guide, then injects the high-frequency component of the RGB guide into the HSI reconstruction. When the RGB image is replaced by a panchromatic one, this corresponds to the standard GSA algorithm [1].

Conversely to standard GSA, here we consider a penalized multivariate regression between the three RGB channels and the HSI bands. Let $\mathbf{A} \in \mathbb{R}^{N \times C}$ denote centered HSI samples extracted from $\mathbf{H}^{\uparrow}$, and let $\mathbf{Q} \in \mathbb{R}^{N \times 3}$ denote centered samples from the low-pass component of the warped RGB guide. Let $w_i = \mathbf{C}(x_i)$ be the confidence value at sample location x_i. We estimate the regression matrix $\mathbf{\Theta} \in \mathbb{R}^{C \times 3}$ by solving

$$\mathbf{\Theta}^{\star} = \arg\min_{\mathbf{\Theta}} \sum_{i=1}^{N} w_i \left\| \mathbf{Q}_i - \mathbf{A}_i \mathbf{\Theta} \right\|_2^2 + \lambda \left\| \mathbf{\Theta} \right\|_F^2 . \tag{14}$$

In matrix form, with $\mathbf{\Lambda} = \operatorname{diag}(w_1, \ldots, w_N)$, this gives

$$\mathbf{\Theta}^{\star} = \left(\mathbf{A}^{\top} \mathbf{\Lambda} \mathbf{A} + \lambda \mathbf{I} \right)^{-1} \mathbf{A}^{\top} \mathbf{\Lambda} \mathbf{Q}. \tag{15}$$

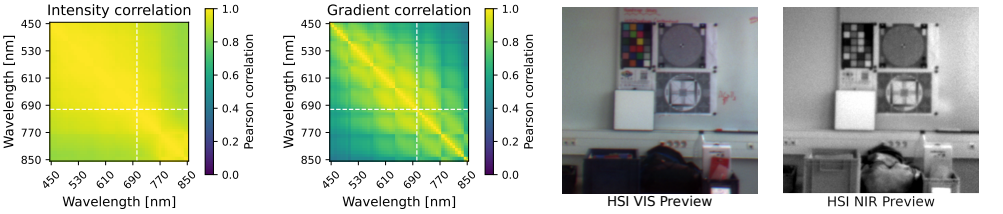


Figure 3: VIS-NIR band correlation on an Ultris HSI. From left to right: intensity correlation, gradient correlation, and VIS and NIR previews of the same HSI. Dashed lines mark the 700 nm VIS-NIR boundary.

The Tikhonov parameter λ stabilizes the regression, especially in the presence of high-frequency noise, which is typical in snapshot HSI acquisitions. The regression predicts the HSI-explained low-frequency RGB component $\mathbf{I} = \mathbf{A}\boldsymbol{\Theta}^{\star}$. The residual RGB detail is then $\mathbf{D} = \mathbf{R}_w - \mathbf{I} \in \mathbb{R}^{N\times 3}$, and is injected through band-wise GSA gains $\mathbf{S} \in \mathbb{R}^{3\times C}$ estimated from $\mathbf{A}$ and $\mathbf{I}$, yielding:

$$\hat{\mathbf{X}}_G = \mathbf{H}^{\uparrow} + \mathbf{DS}. \tag{16}$$

Thus, RGB guidance supplies spatial detail rather than predicting NIR radiance, as the bands remain anchored to $\mathbf{H}^{\uparrow}$, while the RGB residual is transferred through the HSI-derived gains $\mathbf{S}$. This transfer is meaningful when spatial structures remain correlated across the VIS-NIR boundary, as observed in our Ultris data in Figure 3.

This estimate contains spatial details from the RGB guide, but those details may still be unreliable in low-confidence regions. Therefore, we apply a final gating step.

4.4 Confidence-gated detail injection

The confidence map is finally used to control how much RGB-induced detail is retained in the reconstruction. We combine two candidate estimates with complementary behavior. The first one is the upsampled HSI $\mathbf{H}^{\uparrow}$, which preserves the spectral content of the observed HSI but remains spatially smooth. The second one is the confidence-weighted GSA estimate $\hat{\mathbf{X}}_G$, which contains high-frequency detail transferred from the registered RGB guide but may still be locally unreliable when the RGB-HSI correspondence is incorrect.

We use the confidence map as a spatially varying convex gate between these two estimates:

$$\hat{\mathbf{X}}(x) = (1 - \mathbf{C}(x))\mathbf{H}^{\uparrow}(x) + \mathbf{C}(x)\hat{\mathbf{X}}_G(x), \qquad \mathbf{C}(x) \in [0,1]. \tag{17}$$

The confidence is broadcast along the spectral dimension. Thus, when $\mathbf{C}(x) = 0$, the reconstruction falls back to the HSI-preserving estimate; when $\mathbf{C}(x) = 1$, it uses the RGB-guided GSA estimate; intermediate values attenuate the RGB-induced detail.

Equivalently, Eq. (17) can be written as gated detail injection:

$$\hat{\mathbf{X}} = \mathbf{H}^{\uparrow} + \mathbf{C} \odot \left(\hat{\mathbf{X}}_G - \mathbf{H}^{\uparrow}\right). \tag{18}$$

This form makes explicit that the confidence gate does not replace the hyperspectral estimate. It only controls the amount of RGB-guided high-frequency detail added to it.

The gate also admits a local least-squares interpretation. For each spatial location x, Eq. (17) is the minimizer of:

$$E_x(\mathbf{Z}) = \frac{1-\mathbf{C}(x)}{2}\left\|\mathbf{Z}-\mathbf{H}^{\uparrow}(x)\right\|_2^2 + \frac{\mathbf{C}(x)}{2}\left\|\mathbf{Z}-\hat{\mathbf{X}}_G(x)\right\|_2^2. \tag{19}$$

Therefore, the final reconstruction is a reliability-weighted least-squares compromise between a spectrally conservative HSI estimate and an RGB-guided detail-enhanced estimate.

5 Experiments

5.1 Experimental setup

5.1.1 Datasets

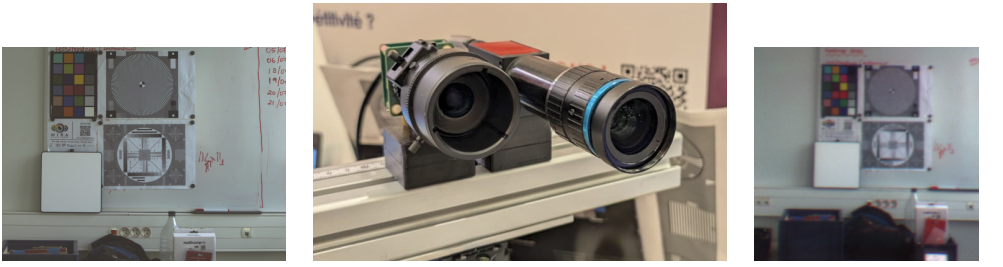
Real dataset. We first evaluate on the Real unaligned RGB-guided HSR benchmark [12]. The dataset contains 57 paired hyperspectral acquisitions with real non-rigid misalignment. One HSI is spectrally projected to simulate the HR RGB guide, and the other is spatially degraded according to Eq. (1), using a Gaussian blur with standard deviation $\sigma/2$ followed by area-based downsampling, with the original HSI used as the HR reference. All methods are re-evaluated using these common inputs and a valid-mask evaluation described in Sec. 5.1.3. Consequently, baseline values may differ slightly from those originally reported. We use the 31-band visible setting and evaluate scale factors $\times 4$ and $\times 8$ on 10 test frames. Hyperparameters are selected on the 47-frame validation split and then fixed for test evaluation.

Ultris-RPi real-device acquisition. To assess deployment beyond the Real benchmark, we additionally evaluate an Ultris-RPi sequence acquired with the Ultris SR5 compact snapshot hyperspectral camera and a Raspberry Pi HQ RGB camera, shown in Figure 4. Unlike Real, whose RGB guidance is synthesized from HSI-HSI pairs, this is a true dual-camera RGB-HSI acquisition with different fields of view, sensor responses, spatial resolutions, and noise characteristics. Since simultaneous HR HSI ground truth is unavailable in this real setup, we use a reduced-resolution quantitative evaluation [4] on a 34-frame static sequence, using the observed HSI as reference, and synthetically degrading it by a factor $\times 4$. We additionally report a qualitative analysis on a single frame in its native resolution. SSC-HSR is evaluated on the 31-band VIS subset, while CMF-GSA-wg is evaluated on both the same 31-band subset and the full 51-band VIS-NIR cube.

5.1.2 Compared methods

We compare against a learning-based SISR baseline, ESSAformer [25], and two learning-based unaligned RGB-guided HSR baselines, HSIFN [12] and SSC-HSR [26]. We also compare several variants of our lightweight CMF-GSA hybrid pipeline. CMF-GSA denotes CMF alignment followed by valid-mask GSA. CMF-GSA-w uses the confidence map only as weighted least-squares weights in the GSA spectral regression. CMF-GSA-wg is the full method, using confidence both in the regression and in the final two-expert detail gate.

On Ultris-RPi, a reduced-resolution quantitative evaluation and a native-resolution qualitative evaluation are provided. SSC-HSR is evaluated in its 31-band VIS subset, whereas CMF-GSA-wg is evaluated both on the same 31-band subset, providing a direct comparison, and on the full 51-band VIS-NIR cube. We do not evaluate 31-band learned baselines in the 51-band setting without retraining or spectral-domain adaptation [17].



(a) RGB camera view **(b)** Acquisition setup **(c)** HSI camera view

Figure 4: Ultris-RPi acquisition system used for the real-device evaluation, composed of a Raspberry Pi HQ RGB camera (left) and an Ultris SR5 compact snapshot hyperspectral camera (right), mounted on a rail. The corresponding observed RGB and HSI images are shown after flat-field correction with the white reference. The two cameras have different fields of view, spatial resolutions, and sensor and noise characteristics.

5.1.3 Metrics

Let $\mathbf{X}^\star, \hat{\mathbf{X}} \in \mathbb{R}^{H\times W\times C}$ denote the reference and reconstructed HSIs. Metrics are computed on the valid spatial support Ω induced by the registration mask, and all images are evaluated in the data range $[0,1]$. We report RMSE, mean PSNR (mPSNR), mean SSIM (mSSIM), and spectral angle mapper (SAM). RMSE is computed globally over all valid pixels and spectral bands, while mPSNR and mSSIM are averaged over bands. SAM is reported as the average spectral angle over valid pixels. In addition to scalar metrics, we visualize pixel-wise RMSE and SAM error maps for selected frames. All methods are evaluated on the HSI grid. Quantitative metrics are computed on the same valid spatial support for a given frame, induced by the CMF mask, so that no method is favored by a different border region. We report runtime on a Linux workstation with an NVIDIA RTX 3500 Ada Generation Laptop GPU, an x86-64 CPU, and PyTorch 2.11.0 with CUDA 13.0.

5.1.4 Hyperparameter selection

For the Real dataset, hyperparameters are selected on the full 47-frame validation split and then fixed for test evaluation. We perform a grid search over the confidence parameters and the GSA Tikhonov weight, using a joint RMSE-SAM validation criterion. The selected values are $(\alpha,\beta,\gamma,\lambda) = (25,0,0.1,10^{-4})$ for $\times 4$, and $(\alpha,\beta,\gamma,\lambda) = (15,0,0.1,10^{-4})$ for $\times 8$. Thus, on Real, the raw RGB proxy residual is the main confidence cue, while cycle consistency acts as a moderate auxiliary cue and the ZNCC structure cue is disabled.

For Ultris-RPi, no reference-based hyperparameter search is performed. The raw RGB proxy residual E_{rgb} is disabled because, in a real dual-camera acquisition, it responds not only to misregistration but also to cross-sensor radiometry, HSI noise, and imperfect spatial-bandwidth matching. For Ultris-RPi, the structural reliability term is instantiated as a local ZNCC cue computed after Gaussian smoothing. For both Ultris-RPi settings we use $(\alpha,\beta,\gamma,\lambda) = (0,10,0.01,0.1)$. The ZNCC structural cue uses a 5×5 window, minimum local standard deviation 0.01, and threshold 0.05; the Gaussian smoothing standard deviation is 0.7 for reduced-resolution evaluation and 1.0 for the native-resolution evaluation.

5.2 Real benchmark evaluation

Quantitative comparison. Table 1 reports the main comparison on the Real test set. The full confidence-gated CMF-GSA-wg outperforms the learned RGB-guided baselines at both scale factors and the HSI-only ESSAformer SISR baseline at $\times 4$, while remaining substantially faster. At $\times 4$, CMF-GSA-wg obtains the best RMSE, mPSNR, mSSIM, and SAM, with values of 0.0168, 37.34 dB, 0.94, and 2.17°, respectively. At $\times 8$, CMF-GSA-wg again achieves the best four reconstruction metrics, including RMSE 0.0257 and SAM 2.94°.

The comparison also shows that weighted regression alone does not explain the main gain. CMF-GSA-w improves SAM and mSSIM over valid-mask CMF-GSA, but gives only marginal changes in RMSE and mPSNR. The major improvement comes from the final two-expert gate, which suppresses unreliable RGB detail injection. At $\times 4$, CMF-GSA-wg reduces RMSE by about 33% relative to CMF-GSA-w.

Table 2 further reports learned parameter counts and peak GPU memory. All 39.76M parameters of CMF-GSA-wg belong to the frozen CMF registration backbone, while the confidence estimation and SR/fusion stages add no learned parameters. Despite the larger registration backbone, CMF-GSA-wg has the lowest peak memory at 0.69 GiB.

Confidence-cue ablation Table 3 isolates the confidence cues in the full gated model. Removing the RGB residual cue, i.e. setting $\alpha = 0$, consistently degrades RMSE, mPSNR, mSSIM, and SAM at both scale factors, confirming that RGB-HSI proxy consistency is the dominant reliability cue on the Real benchmark. Removing cycle consistency, i.e. setting $\gamma = 0$, gives metrics nearly tied to the full model, while reducing runtime because the backward flow no longer needs to be estimated. This supports our interpretation of cycle consistency as an auxiliary diagnostic cue in this cross-resolution setting.

Qualitative analysis. Figure 5 shows representative Real test examples at $\times 4$ and $\times 8$, respectively. The selected examples support the quantitative results. At $\times 4$, the visual differences are subtle but CMF-GSA-wg reduces local RMSE in the highlighted regions. At $\times 8$, SSC-HSR shows more apparent over-enhanced or displaced structures, while the proposed gate suppresses unreliable RGB detail injection. In both frames, CMF-GSA-wg obtains the lowest RMSE and comparable or lower SAM.

5.3 Ultris-RPi real-device evaluation

Reduced-resolution evaluation. Table 4 reports the 34-frame reduced-resolution evaluation. On the common 31-band VIS support, CMF-GSA-wg outperforms SSC-HSR across all four reconstruction metrics, while the 51-band result demonstrates direct deployment over the full VIS-NIR support without retraining.

Native-resolution evaluation. At native resolution, Figure 6 provides the complementary qualitative evaluation. CMF-GSA-wg operates directly on the full 51-band VIS-NIR cube, whereas SSC-HSR remains restricted to its 31-band VIS setting. As shown in the zoomed crops in Figure 6(c) and (d), SSC-HSR produces clear local spatial artifacts under this real cross-sensor setting, while CMF-GSA-wg transfers the corresponding RGB details without retraining. Since no clean HR HSI reference is available at this resolution, these results are interpreted qualitatively rather than as reference-based reconstruction metrics.

	SF	Method	mPSNR ↑	RMSE ↓	mSSIM ↑	SAM [deg] ↓	Time [s] ↓
Real [D]	×4	HSIFN [D]	31.9786±2.7543	0.0275±0.0109	0.8436±0.0369	3.7786±1.0926	1.2328±0.0533
		ESSAformer [D]	32.7880±3.2412	0.0294±0.0131	0.8968±0.0479	4.8850±2.1632	0.4776±0.0109
		SSC-HSR [D]	34.8129±3.8667	0.0237±0.0121	0.9224±0.0393	2.2074±0.5635	0.4764±0.0203
		CMF-GSA	34.2240±3.7267	0.0248±0.0113	0.9234±0.0324	2.3894±0.5173	**0.0461±0.0056**
		CMF-GSA-w	34.1732±3.6950	0.0250±0.0114	0.9281±0.0294	2.3077±0.5429	0.0922±0.0063
		CMF-GSA-wg	**37.3442±3.4779**	**0.0168±0.0074**	**0.9434±0.0211**	**2.1697±0.5515**	0.0956±0.0075
	×8	HSIFN [D]	28.6940±2.9490	0.0423±0.0167	0.7609±0.0652	4.1891±1.2229	1.2325±0.0579
		SSC-HSR [D]	28.7510±3.8835	0.0480±0.0221	0.7829±0.1079	3.3277±1.0593	0.4753±0.0194
		CMF-GSA	32.8469±3.6326	0.0283±0.0116	0.8978±0.0465	3.2291±0.8154	**0.0486±0.0011**
		CMF-GSA-w	32.7655±3.6501	0.0287±0.0119	0.9001±0.0462	3.1147±0.8751	0.0911±0.0034
		CMF-GSA-wg	**33.7938±3.8149**	**0.0257±0.0115**	**0.9039±0.0435**	**2.9398±0.8248**	0.0939±0.0025

Table 1: Quantitative comparison on the Real test set. Values are mean ± standard deviation over 10 test frames. CMF-GSA-w uses confidence only in the weighted regression, while CMF-GSA-wg additionally gates the final RGB-guided detail residual.

Method	Total [M]	Reg. [M]	SR/fusion [M]	Peak mem. [GiB] ↓
HSIFN [D]	21.01	18.75	2.26	7.80
ESSAformer [D]	11.64	–	11.64	3.97
SSC-HSR [D]	27.45	5.26	22.19	2.63
CMF-GSA-wg	39.76	39.76	**0**	**0.69**

Table 2: Learned parameter counts and peak allocated GPU memory. Parameter counts are decomposed into registration and SR/fusion components.

	SF	Energy variant	mPSNR ↑	RMSE ↓	mSSIM ↑	SAM [deg] ↓	Time [s] ↓
Real [D]	×4	**Full gate**	**37.3442±3.4779**	**0.0168±0.0074**	**0.9434±0.0211**	**2.1697±0.5515**	0.0956±0.0075
		w/o cycle cue ($\gamma = 0$)	37.3099±3.4378	**0.0168±0.0073**	0.9428±0.0212	2.1746±0.5510	0.0504±0.0009
		w/o RGB res. ($\alpha = 0$)	35.3151±3.7098	0.0215±0.0099	0.9312±0.0310	2.3445±0.5246	0.0932±0.0038
		w/o energy ($\alpha = \gamma = 0$)	34.1760±3.7427	0.0250±0.0115	0.9257±0.0316	2.3808±0.5071	**0.0502±0.0010**
	×8	**Full gate**	33.7938±3.8149	0.0257±0.0115	**0.9039±0.0435**	**2.9398±0.8248**	0.0939±0.0025
		w/o cycle cue ($\gamma = 0$)	**33.8194±3.7854**	**0.0256±0.0114**	0.9036±0.0438	2.9572±0.8343	0.0509±0.0009
		w/o RGB res. ($\alpha = 0$)	33.2308±3.6130	0.0270±0.0108	0.9032±0.0439	3.1343±0.8089	0.0931±0.0026
		w/o energy ($\alpha = \gamma = 0$)	32.7810±3.6606	0.0286±0.0118	0.8988±0.0465	3.2156±0.7978	**0.0506±0.0012**

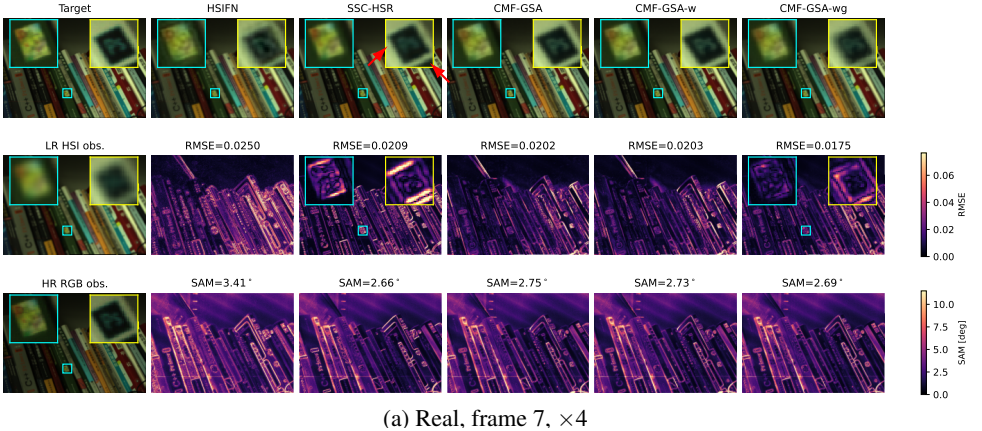
Table 3: Confidence-cue ablation on the Real test set. Values are mean ± standard deviation over 10 test frames. The RGB residual is the dominant confidence cue. Removing cycle consistency gives reconstruction metrics nearly tied to the full model while roughly halving runtime by avoiding the backward flow.

	SF	Bands	Method	mPSNR ↑	RMSE ↓	mSSIM ↑	SAM [deg] ↓	Time [s] ↓
Ultris-RPi	×4	31	SSC-HSR [D]	29.3323±0.0925	0.0342±0.0004	0.8637±0.0017	1.8125±0.0100	0.2306±0.0179
		31	**CMF-GSA-wg**	**29.8677±0.0901**	**0.0322±0.0003**	**0.9087±0.0010**	**1.7713±0.0132**	**0.0873±0.0064**
		51	CMF-GSA-wg	29.7565±0.1034	0.0329±0.0004	0.8909±0.0012	1.7122±0.0081	0.0957±0.0086

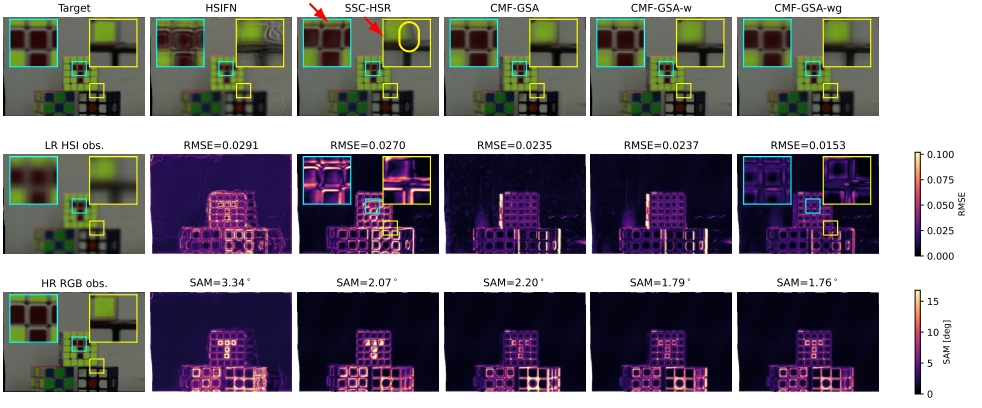
Table 4: Reduced-resolution quantitative evaluation on the 34-frame static Ultris-RPi sequence at ×4. Values are mean ± standard deviation over 34 frames. The 31-band VIS rows provide a direct comparison between SSC-HSR and CMF-GSA-wg, while the 51-band row evaluates CMF-GSA-wg on the full VIS-NIR spectral support.

6 Discussion

The experiments support the proposed lightweight model-based alternative to learned unaligned RGB-guided HSR. On Real, the main reconstruction gain comes from the final confidence gate rather than confidence-weighted regression alone. WLS modifies the global



(a) Real, frame 7, $\times 4$



(b) Real, frame 59, $\times 8$

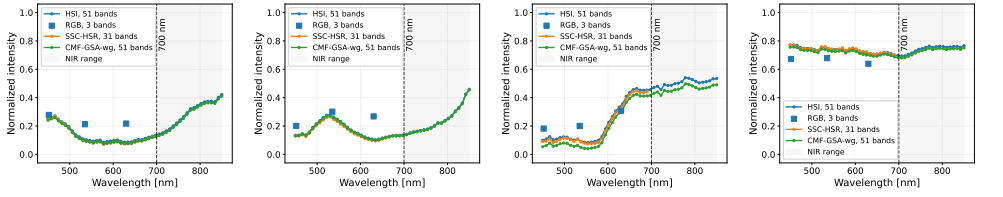
Figure 5: Qualitative comparison on the Real test set at (a) $\times 4$ and (b) $\times 8$. The first column shows the target, LR HSI observation, and HR RGB observation; method columns show RGB previews, RMSE maps, and SAM maps. Two regions are highlighted for local inspection, with corresponding RMSE enlargements for SSC-HSR and CMF-GSA-wg. At $\times 4$, the visual differences are relatively subtle but CMF-GSA-wg reduces local RMSE in the highlighted regions. At $\times 8$, over-enhanced or spatially shifted structures are more apparent in SSC-HSR, whereas CMF-GSA-wg reduces the corresponding local error. In both examples, CMF-GSA-wg obtains the lowest RMSE and comparable or lower SAM. RMSE insets use a shared local scale for SSC-HSR and CMF-GSA-wg, separate from the full-map colorbar.

RGB-HSI spectral calibration, whereas the gate locally suppresses unreliable high-frequency RGB residuals. The RGB-proxy residual is the dominant confidence cue on Real, while cycle consistency is mainly auxiliary, since forward-backward disagreement can also arise from the cross-resolution asymmetry between a smooth HSI proxy and a sharper RGB image.

The Ultris-RPi experiments highlight the deployment flexibility of the model-based formulation beyond the controlled Real protocol. The 34-frame reduced-resolution evaluation validates the method on genuine cross-sensor data, while the native-resolution experiment demonstrates direct operation on the full 51-band VIS-NIR support without retraining, whereas the learned baseline remains restricted to its 31-band VIS setting.



(a) LR HSI (b) RGB-guide (c) SSC-HSR [26] (d) Ours, 51-bands



(e) Blue square (f) Green square (g) Red square (h) White reference

Figure 6: Qualitative real-device evaluation on an Ultris-RPi frame. The first row shows visible-domain previews of the observed HSI, the observed RGB image, SSC-HSR, and the proposed CMF-GSA-wg reconstruction. The second row compares spectra at selected image locations. The observed HSI and CMF-GSA-wg curves cover the full 51-band VIS–NIR range, SSC-HSR is restricted to the 31-band VIS subset, and the RGB observation provides only three broadband measurements.

The approach has limitations. First, the confidence is an energy-based reliability score, not a calibrated posterior probability. Second, if the HSI-to-RGB proxy is spectrally inaccurate or poorly radiometrically matched to the RGB camera, the RGB residual can confuse spectral-response mismatch with geometric misregistration. On real-device data, replacing the raw residual with a ZNCC-based structure cue reduces sensitivity to radiometry, but remains a heuristic reliability cue rather than a calibrated geometric uncertainty. Third, while confidence gating reduces the impact of local flow errors, it misses RGB details where no reliable correspondence exists, leading to local low-resolution fallbacks. That said, stronger HSI-only priors, such as spatial-spectral regularization [5] or HSI single-image super-resolution fallback models, may improve large unreliable regions. Future work should include real-device RGB-HSI benchmarks with clean HR HSI references.

7 Conclusion

We presented a lightweight interpretable framework for RGB-guided hyperspectral super-resolution under real cross-resolution misalignment. The method combines cross-modal flow and GSA with an explicit reliability map used for both confidence-weighted spectral regression and gated detail injection. On Real, the proposed method improves over learned unaligned fusion baselines while remaining substantially faster, with the final gate providing the main reconstruction gain. On a real RGB-HSI setup, the 34-frame Ultris-RPi evaluation further validates the method under real cross-sensor differences, while native-resolution results demonstrate deployment on a full 51-band VIS-NIR cube without retraining.

Acknowledgements

This work was supported by the Institut Carnot Logiciels et Systèmes Intelligents (Carnot LSI) and by the French National Research Agency (ANR) under the France 2030 programme with reference ANR-23-IACL-0006. The computations were performed using the GRICAD infrastructure (GRANT CPER07_13 CIRA, ANR10 LABX56, ANR-10-EQPX-29-01). Data acquisitions were carried out with the support of the Multi-camera Imaging Research and Acquisition (MIRA) Platform (https://mira-imaging.github.io/) of GIPSA-lab, which provided the camera setup and the acquisition and research resources used in this work. The MIRA Platform is developed with the financial support of the Institut Carnot Logiciels et Systèmes Intelligents (Carnot LSI). The authors thank Dr. Amaury Negre for his contributions to the ROS-based acquisition software and to the acquisition of the Ultris-RPi sequence.

References

[1] Bruno Aiazzi, Stefano Baronti, and Massimo Selva. Improving component substitution pansharpening through multivariate regression of ms + pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3230–3239, 2007.

[2] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011.

[3] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2017)*, 36(6):218:1–13, 2017. doi: 10.1145/3130800.3130810. URL http://dx.doi.org/10.1145/3130800.3130810.

[4] Matteo Ciotola, Giuseppe Guarino, Gemine Vivone, Giovanni Poggi, Jocelyn Chanussot, Antonio Plaza, and Giuseppe Scarpa. Hyperspectral pansharpening: Critical review, tools, and future perspectives. *IEEE Geoscience and Remote Sensing Magazine*, 13(1):311–338, 2024.

[5] Laurent Condat, Daichi Kitahara, Andrés Contreras, and Akira Hirabayashi. Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists. *SIAM Review*, 65(2):375–435, 2023.

[6] Kebin Contreras, Mohamad Jouni, Mauro Dalla Mura, and Jorge Bacca. A multimodal hyperspectral dataset of cocoa beans with physicochemical annotation. *Scientific Data*, 2025.

[7] Luigi Di Stefano, Stefano Mattoccia, and Federico Tombari. Zncc-based template matching using bounded partial correlation. *Pattern recognition letters*, 26(14):2129–2134, 2005.

[8] Honey Gupta and Kaushik Mitra. Pyramidal edge-maps and attention based guided thermal super-resolution. In *European Conference on Computer Vision*, pages 698–715. Springer, 2020.

[9] Honey Gupta and Kaushik Mitra. Toward unaligned guided thermal super-resolution. *IEEE Transactions on Image Processing*, 31:433–445, 2021.

[10] Nathan Hagen and Michael W Kudenov. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):090901–090901, 2013.

[11] Mohamad Jouni, Daniele Picone, and Mauro Dalla Mura. Multiple-beam interference spectroscopy: Instrument analysis and spectrum reconstruction. *IEEE Transactions on Instrumentation and Measurement*, 2025.

[12] Zeqiang Lai, Ying Fu, and Jun Zhang. Hyperspectral image super resolution with real unaligned rgb guidance. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):2999–3013, 2025. doi: 10.1109/TNNLS.2023.3340561.

[13] Dimitris G Manolakis, Ronald B Lockwood, and Thomas W Cooley. *Hyperspectral imaging remote sensing: physics, sensors, and algorithms*. Cambridge University Press, 2016.

[14] Pai Chet Ng, Zhixiang Chi, Yannick Verdie, Juwei Lu, and Konstantinos N Plataniotis. Hyper-skin: A hyperspectral dataset for reconstructing facial skin-spectra from rgb images. *Advances in Neural Information Processing Systems*, 36:24158–24170, 2023.

[15] Subir Paul, Vinayaraj Poliyapram, Nevrez İmamoğlu, Kuniaki Uto, Ryosuke Nakamura, and D. Nagesh Kumar. Canopy averaged chlorophyll content prediction of pear trees using convolutional autoencoder on hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:1426–1437, 2020. doi: 10.1109/JSTARS.2020.2983000.

[16] Daniele Picone, Mohamad Jouni, and Mauro Dalla Mura. Spectro-spatial hyperspectral image reconstruction from interferometric acquisitions. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2590–2594, 2024. doi: 10.1109/ICASSP48485.2024.10447567.

[17] Daniele Picone, Mohamad Jouni, and Mauro Dalla Mura. Leveraging pretrained rgb denoisers for hyperspectral image restoration. In *2026 IEEE International Conference on Image Processing (ICIP)*, pages 1–6. IEEE, 2026.

[18] Rafael E Rivadeneira, Angel D Sappa, Chenyang Wang, Junjun Jiang, Zhiwei Zhong, Peilin Chen, and Shiqi Wang. Thermal image super-resolution challenge results-pbvs 2024. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3113–3122. IEEE, 2024.

[19] Bernard Schmitt, Zahira Souidi, Frédérique Duquesnoy, and Frédéric-Victor Donzé. From rgb camera to hyperspectral imaging: a breakthrough in neolithic rock painting analysis. *Heritage Science*, 11(1):1–27, 2023.

[20] Linfeng Tang, Yuxin Deng, Yong Ma, Jun Huang, and Jiayi Ma. Superfusion: A versatile image registration and fusion network with semantic awareness. *IEEE/CAA Journal of Automatica Sinica*, 9(12):2121–2137, 2022. doi: 10.1109/JAS.2022.106082.

[21] Zachary Teed and Jia Deng. RAFT: Recurrent all-pairs field transforms for optical flow. In *European Conference on Computer Vision*, 2020.

[22] Marcel Trammer, Nils Genser, and Jürgen Seiler. Rgb-guided resolution enhancement of ir images. In *2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 1–5. IEEE, 2023.

[23] Weixiao Wan, Bowen Zhang, Marija Vella, João F.C. Mota, and Wei Chen. Robust rgb-guided super-resolution of hyperspectral images via TV^3 minimization. *IEEE Signal Processing Letters*, 29:957–961, 2022. doi: 10.1109/LSP.2022.3159149.

[24] F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical Report CUCS-061-08, Department of Computer Science, Columbia University, Nov 2008.

[25] Mingjin Zhang, Chi Zhang, Qiming Zhang, Jie Guo, Xinbo Gao, and Jing Zhang. Essaformer: Efficient transformer for hyperspectral image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23073–23084, 2023.

[26] Yingkai Zhang, Zeqiang Lai, Tao Zhang, Ying Fu, and Chenghu Zhou. Unaligned rgb guided hyperspectral image super-resolution with spatial-spectral concordance. *International Journal of Computer Vision*, 133:6590–6610, 2025. doi: 10.1007/s11263-025-02466-8.

[27] Shili Zhou, Weimin Tan, and Bo Yan. Promoting single-modal optical flow network for diverse cross-modal flow estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3562–3570, 2022.