FOTA: FORWARD-ONLY TEST-TIME ADAPTATION FOR MULTISPECTRAL AND HYPERSPECTRAL IMAGE FUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multispectral and hyperspectral image fusion (MHIF) is an important approach for obtaining high-resolution hyperspectral images. However, in real-world applications, out-of-distribution (OOD) shifts caused by unknown, target-specific physical degradations lead to severe performance degradation in existing pre-trained MHIF models. Most existing test-time adaptation (TTA) methods rely on proxy losses or explicit degradation loops that require computationally intensive iterative backpropagation, which incurs high computational overhead and makes models prone to overfitting in few-shot scenarios. To address this bottleneck, we propose a novel backpropagation-free, architecture-agnostic framework: **Forward-Only Test-Time Adaptation (FOTA)**. Specifically, we first project high-dimensional features into a compact low-rank subspace to reduce the parameter scale. To overcome the slow convergence of zeroth-order optimization, we utilize the analytical closed-form solution of Bures-Wasserstein optimal transport under the multivariate Gaussian assumption for initialization, achieving rapid distribution alignment. Furthermore, we construct a structured prior from the physical observation differences of high-resolution multispectral images to directly guide the search direction. Finally, the algorithm iteratively refines the low-rank prompt via forward-only zeroth-order optimization, ensuring physical imaging consistency without backpropagation. Experiments on five benchmarks demonstrate that FOTA achieves leading adaptation performance. Our code will be released.

1 INTRODUCTION

Unlike conventional RGB images, hyperspectral images (HSIs) comprise tens to hundreds of contiguous spectral bands, enabling fine-grained material identification in precision tasks such as food safety inspection Feng & Sun (2012), remote sensing reconnaissance Shimoni et al. (2019), and crop pest and disease control Zhang et al. (2019). However, due to physical constraints of imaging sensors, HSIs typically suffer from limited spatial resolution. the MHIF task aims to integrate the spatial details of high-resolution multispectral images with the abundant spectral characteristics of low-resolution hyperspectral images, thereby synthesizing high-resolution hyperspectral images (HR-HSIs). Despite the remarkable success of deep learning-based fusion models Xu et al. (2026); Li et al. (2026); Zhu et al. (2025), they typically assume that training and testing data are independent and identically distributed (i.i.d.). In real-world deployments, testing data inevitably undergoes distribution shifts due to hardware variations in sensors, atmospheric turbulence, or dynamic changes in lighting conditions Li et al. (2025d). Such domain discrepancies often cause a catastrophic performance drop in pre-trained fusion models.

To address this challenge, Test-Time Adaptation (TTA) has emerged as a promising paradigm, allowing models to update on-the-fly during the inference stage using only unlabeled test data. Traditional TTA methods (e.g., TENT Wang et al. (2020), EATA Niu et al. (2022)) generally rely on backpropagation (BP) to update a subset of model parameters, such as Batch Normalization layers. However, for high-dimensional HSI data, continuous BP incurs prohibitive memory and computational overhead and is prone to catastrophic forgetting of source domain knowledge Li et al. (2025c). More critically, in practical aerospace or edge computing scenarios, models are often heavily quantized or hard-coded to accelerate inference, rendering any weight-update or BP-based TTA

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

method infeasible on physical devices. In response to these limitations, BP-free or forward-only TTA methods (e.g., FOA Niu et al. (2024), FOZO Wang & Wang (2026)) have been proposed. These methods bypass gradient computation by using zeroth-order (ZO) optimization to learn prompts at the input level, demonstrating great potential for lightweight deployment. Nevertheless, these approaches are almost exclusively tailored for image classification, driven primarily by prediction entropy minimization. For pixel-level dense regression tasks like MHIF—which lack explicit uncertainty metrics and demand texture details and spectral fidelity—classification entropy becomes entirely ineffective. For pixel-level dense regression tasks like MHIF, which lack explicit uncertainty metrics and demand texture details and spectral fidelity, classification entropy becomes entirely ineffective.

Concurrently, a series of TTA methods has emerged for pixel-level vision tasks such as image super-resolution Deng et al. (2023); Tang & Yang (2024) and restoration Gou et al. (2024); Guo et al. (2025). For instance, some works focus on adapting to complex degradations at test time, such as second-order degradation and reconstruction Deng et al. (2023) and degradation adaptation for open-set restoration Gou et al. (2024). To construct effective supervisory signals in the absence of ground truth, researchers have proposed diverse innovative paradigms. For instance, IODA Tang & Yang (2024) uses instance masks from the Segment Anything Model (SAM) to extract instance-specific Alpha-CLIP features for target-domain adaptation in one-shot super-resolution; SPA Niu et al. (2025) uses Fourier-domain low-frequency masking and high-frequency noise injection to improve prediction consistency between augmented views; and DCTTA Tang et al. (2026) uses a diffusion-based degradation generator to produce redegraded images and enforce restoration consistency under unseen degradations. However, HSI data is characterized by continuous 3D spatial-spectral structures and scarce sample availability. Directly deploying these RGB-tailored methods onto such tasks inevitably leads to the curse of dimensionality. To address this, recent efforts have adapted TTA to HSIs. In particular, UTAL Zhang et al. (2024) uses a mutual-guiding fusion module to learn a shared spatial-spectral prior and adapts it through a self-guiding adaptation module and degradation estimation network. THSI Li et al. (2025c) develops a self-training framework that refines pseudo-labels and incorporates Spectral Mixup to enhance data diversity. Despite their task-specific tailoring, these methods still fundamentally rely on iterative backpropagation to update network weights. On high-dimensional HSI data with tens to hundreds of spectral bands, continuous gradient computation incurs prohibitive memory and latency overhead, while risking catastrophic forgetting and adaptation collapse in few-shot settings. The illustration figure is show in Fig. 1.

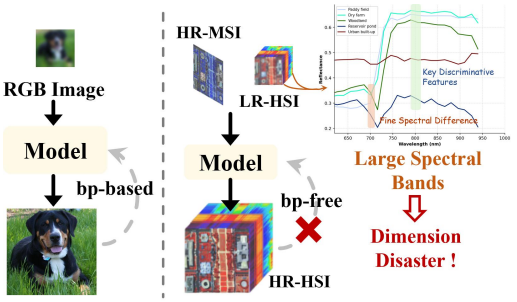


Figure 1: The illustration of our FOTA against before methods. Unlike gradient-based methods ill-suited for hyperspectral data, our framework is tailored for MHIF, preserving rich spectral signatures while remaining entirely BP-free.

To fill this critical gap, we propose a novel framework: **Forward-Only Test-Time Adaptation (FOTA)**. To the best of our knowledge, this is the first BP-free TTA paradigm in MHIF. Our core idea is to model TTA as a physics-driven continuous affine prompt optimization problem within a tensor low-rank subspace. By leveraging the analytical properties of the Bures-Wasserstein (BW) distance under multivariate Gaussian assumptions, we can instantaneously compute a closed-form affine mapping matrix at test time, providing a zero-latency geometric initialization for feature alignment. Subsequently, to correct statistical biases from single-sample estimation and achieve fine-grained local adaptation, we discard traditional cross-entropy in favor of a backpropagation-free (BP-free) physical degradation consistency objective as the fitness function. Guided by this physical prior, the zeroth-order evolution algorithm requires only a few forward passes to converge to the optimal residual prompt within a minimal low-rank parameter space, achieving pixel-level reconstruction that bridges the distribution gap.

Main Novelty and Contributions: 1) We propose FOTA, the first BP-free TTA framework for the pixel-level regression MHIF task that is entirely BP-free and architecture-agnostic. 2) We introduce a novel initialization strategy leveraging the analytical properties of the BW distance. This provides an instantaneous geometric alignment at the beginning. 3) By synergizing physical

degradation consistency with a low-rank subspace, we design an efficient zeroth-order evolution mechanism. 4) We construct nine realistic corruption scenarios tailored to the intrinsic characteristics of hyperspectral data, and validate the efficacy of the proposed method across challenging cross-scene and cross-temporal benchmarks.

2 Related Work

BP-based TTA Methods. Backpropagation-based Test-Time Adaptation (TTA) methods typically update model parameters, such as the affine parameters of Batch Normalization layers, by minimizing unsupervised objectives during inference. Wang et al. (2020) pioneered this field with TENT, which utilizes entropy minimization to reduce prediction uncertainty on test samples. Subsequent works have introduced various regularization and sample selection mechanisms to enhance stability and prevent catastrophic forgetting. For instance, SAR Niu et al. (2023) employs sharpness-aware optimization and filters unreliable samples, while EATA Niu et al. (2022) utilizes weight anchoring and Fisher information to preserve source knowledge. However, a major challenge in entropy-based adaptation is the collapse problem, where the model produces trivial one-hot outputs to minimize the objective. To address this, ZeroSiam Chen et al. (2026) introduces an efficient asymmetric Siamese architecture. By incorporating a learnable predictor and a stop-gradient operation, ZeroSiam prevents collapse through asymmetric divergence alignment, providing a more robust learning signal for online optimization without requiring labels.

Forward-Only TTA Methods. To alleviate the high memory and computational overhead of gradient backpropagation, BP-free TTA methods have emerged as a promising alternative for deployment on resource-constrained devices. Early approaches like AdaBN Li et al. (2016) and T3A Iwasawa & Matsuo (2021) focus on calibrating statistics or adjusting the classifier layer, but their adaptation capacity is often limited as they do not optimize the model's core representation. Recent research has shifted toward a "forward-only optimization" paradigm. FOA Niu et al. (2024) initiates the forward-only TTA paradigm by employing a derivative-free Covariance Matrix Adaptation Evolution Strategy (CMA-ES) to optimize input prompts alongside an activation shifting scheme. However, its reliance on dozens of forward evaluations per step incurs substantial computational overhead. To reduce this query cost, ZOA Deng et al. (2025) introduces two-point zeroth-order optimization into TTA for the first time, achieving continual adaptation with only two forward passes per sample by integrating a domain knowledge management scheme. Building upon this, FOZO Wang & Wang (2026) explicitly formalizes Simultaneous Perturbation Stochastic Approximation (SPSA) for prompt optimization and designs a dynamically decaying perturbation scale to counter estimation variance under streaming distribution shifts. Going beyond prompt tuning, ZOTTA Zhang et al. (2026b) optimizes internal model weights by reducing parameter dimensions via layer selection and smoothing the loss landscape to ensure reliable zeroth-order gradient estimates. Although derivative-free algorithms such as CMA-ES and SPSA are theoretically task-agnostic and hold promise for pixel-level tasks, existing frameworks rely on classification-tailored entropy objectives that fail in dense regression scenarios.

Pixel-level TTA Methods. To maintain geometric structures and texture fidelity in dense prediction scenarios, several specialized TTA frameworks have been developed. SRTTA Deng et al. (2023) introduces a second-order reconstruction framework that constructs self-supervised pseudo-pairs through a degradation classifier and employs an Adaptive Parameter Preservation strategy to prevent catastrophic forgetting. TAO Gou et al. (2024) leverages a pre-trained degradation-agnostic diffusion model and a test-time degradation adapter to dynamically align the generative domain with specific test-sample shifts. IODA Tang & Yang (2024) addresses the limitations of one-shot adaptation by expanding the target domain distribution using SAM-guided instance masks and Alpha-CLIP latent space direction alignment. SPA Niu et al. (2025) establishes a versatile self-bootstrapping paradigm that utilizes geometry-preserving Fourier-domain augmentations, such as low-frequency amplitude masking and high-frequency noise injection, to provide reliable learning signals for pixel-level tasks. Targeting multi-degradation shifts in all-in-one restoration, DCTTA Tang et al. (2026) synthesizes redegraded pseudo-pairs and applies degradation-guided consistency while selectively updating degradation-sensitive parameters. Specifically for hyperspectral imaging, UTAL Zhang et al. (2024) mitigates unknown degradations by adapting a content-agnostic prior shared across different images through an unsupervised reconstruction loss. Building upon this, THSI Li et al. (2025c) develops a

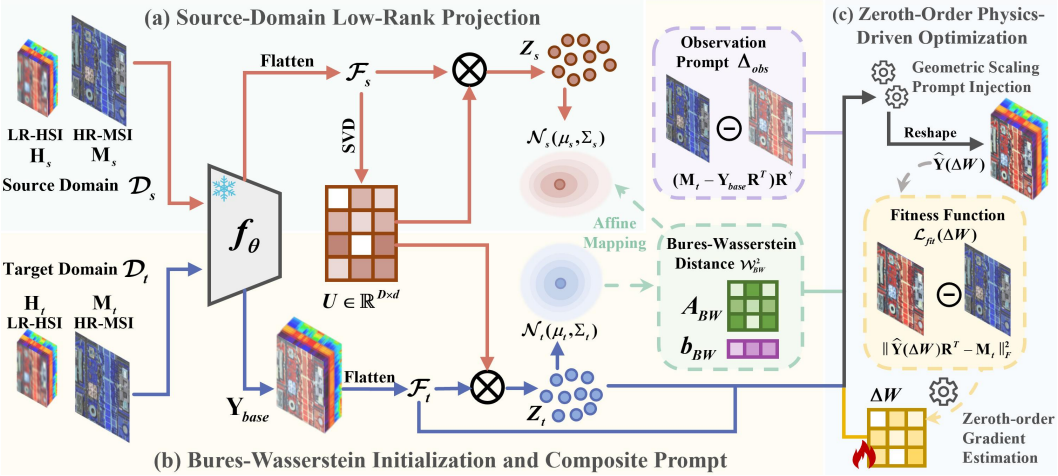


Figure 2: The pipeline and framework of our proposed FOTA. (a) projects high-dimensional hyperspectral features into a compact low-rank subspace to boost TTA efficiency; (b) establishes a well-conditioned initial solution alongside the residual prompt to be optimized; and (c) injects directional perturbations during TTA to guide the zeroth-order optimization trajectory.

teacher-student self-training framework incorporating Spectral Mixup to enhance training diversity while preserving essential spatial-spectral consistency.

3 Methodology

The overall framework of our method is illustrated in Fig. 2.

3.1 Preliminary: Bures-Wasserstein Distance

In optimal transport (OT) theory, the 2-Wasserstein distance is widely used to quantify the discrepancy between two probability distributions. Under the assumption that feature distributions follow multivariate Gaussian distributions $\mathcal{N}(\mu, \Sigma)$—where μ denotes the mean vector and Σ represents the symmetric positive-definite covariance matrix—the 2-Wasserstein distance between the source-domain Gaussian distribution $\mathcal{N}_s(\mu_s, \Sigma_s)$ and the target-domain Gaussian distribution $\mathcal{N}_t(\mu_t, \Sigma_t)$ admits a closed-form expression known as the Bures-Wasserstein distance:

$$\mathcal{W}_{BW}^2(\mathcal{N}_t, \mathcal{N}_s) = \|\mu_s - \mu_t\|_2^2 + \mathrm{Tr}\left(\Sigma_s + \Sigma_t - 2\left(\Sigma_s^{1/2}\Sigma_t\Sigma_s^{1/2}\right)^{1/2}\right), \tag{1}$$

where $\|\cdot\|_2$ denotes the Euclidean norm, $\mathrm{Tr}(\cdot)$ is the trace operator, and $(\cdot)^{1/2}$ denotes the principal matrix square root of a symmetric positive semi-definite matrix.

Furthermore, the optimal transport map that pushes the target distribution $\mathcal{N}_t$ toward the source distribution $\mathcal{N}_s$ also admits an analytical formulation Gabriel & Marco (2019). This optimal transport map is characterized as an affine transformation $T(z) = Az + b$, where z denotes the feature vector to be adapted. To guarantee numerical stability and preserve mapping symmetry, the optimal transformation matrix A and bias vector b are formulated as:

$$A = \Sigma_t^{-1/2}\left(\Sigma_t^{1/2}\Sigma_s\Sigma_t^{1/2}\right)^{1/2}\Sigma_t^{-1/2}, \quad b = \mu_s - A\mu_t. \tag{2}$$

This solution establishes an analytical geometric baseline for feature distribution alignment.

3.2 Problem Formulation and Motivation

The MHIF task aims to merge a high-resolution multispectral image (HR-MSI) $\mathbf{M} \in \mathbb{R}^{H\times W\times c}$ with a low-resolution hyperspectral image (LR-HSI) $\mathbf{H} \in \mathbb{R}^{h\times w\times C}$ to reconstruct a high-resolution

hyperspectral image (HR-HSI) $\mathbf{Y} \in \mathbb{R}^{H\times W\times C}$. Here, (H, W, C) and (h, w, c) denote spatial height, width, and spectral bands, respectively, with $h \ll H$, $w \ll W$, and $c \ll C$. In the TTA setting, we are provided with a pre-trained fusion network f_θ, having been optimized on a clean source domain $\mathcal{D}_s = \{(\mathbf{M}_s, \mathbf{H}_s)\}$. During test-time deployment on a target domain $\mathcal{D}_t = \{(\mathbf{M}_t, \widetilde{\mathbf{H}}_t)\}$, the observed target HSI $\widetilde{\mathbf{H}}_t$ is considerably more prone to unknown physical degradations due to hardware fabrication limits and physical sampling trade-offs of HSI sensors, inducing severe single-sided distribution shifts. Our goal is to adapt the pre-trained model to the target domain $\mathcal{D}_t$.

Unlike classification tasks that typically exploit predictive entropy for adaptation, MHIF is essentially a continuous pixel-level regression task over a high dynamic range, lacking explicit confidence scores or predictive uncertainty metrics. In addition, performing continuous backpropagation across high-dimensional HSI tensors incurs prohibitively high GPU memory overhead and readily triggers catastrophic forgetting. To circumvent these challenges, we propose injecting continuous affine prompts at the feature level and optimizing them exclusively via forward passes.

However, hyperspectral imagery exhibits distinctive three-dimensional (3D) characteristics characterized by the tight intertwining of spatial and spectral dimensions, while realistic HSI datasets typically feature very scarce samples (often only 1 to 40 images) Li et al. (2025c). Such properties cause conventional feature adaptation techniques to suffer from the severe "curse of dimensionality." If one directly employs zeroth-order optimization Deng et al. (2025) to blindly explore optimal continuous affine prompts within such an expansive parameter space, the massive search volume leads to sluggish convergence or even optimization divergence. To address this bottleneck, we incorporate Bures-Wasserstein optimal transport theory. Leveraging the analytical closed-form solution of the BW distance under the multivariate Gaussian assumption, we can directly compute the optimal affine transformation matrix that minimizes the distribution discrepancy between the target and source domains. This closed-form solution geometrically narrows the domain gap, providing a well-conditioned starting point for high-dimensional feature prompt alignment.

3.3 Source-Domain Low-Rank Projection

In accordance with zeroth-order optimization theory Deng et al. (2025); Wang & Wang (2026), it is necessary to pre-compute and store statistical priors of the source-domain features. To ensure the parameter efficiency, we project the high-dimensional features into a compact low-rank tensor subspace. Given the pre-trained model f_θ, during the forward pass on clean source-domain data, we extract the final representations and flatten them along the spatial dimension into a feature matrix $\mathcal{F}_s \in \mathbb{R}^{N\times D}$, where $N = H \times W$ and D denotes the feature channel dimension. We compute the covariance matrix of these features and perform Truncated Singular Value Decomposition (Truncated SVD) to extract the dominant directions of variation. By selecting the orthogonal singular vectors corresponding to the top-d largest singular values (we set $d = 16$ in our experiments), we construct an orthogonal projection matrix $U \in \mathbb{R}^{D\times d}$ satisfying $U^T U = I_d$ with $d \ll D$.

Subsequently, we project the source-domain feature matrix $\mathcal{F}_s$ into this low-rank subspace to obtain a compact representation $z_s = \mathcal{F}_s U \in \mathbb{R}^{N\times d}$. Specifically, we compute the mean vector μ_s and covariance matrix Σ_s of the low-rank features z_s:

$$\mu_s = \frac{1}{N}\sum_{i=1}^{N} z_{s,i} \in \mathbb{R}^d, \quad \Sigma_s = \frac{1}{N-1}\sum_{i=1}^{N}(z_{s,i} - \mu_s)(z_{s,i} - \mu_s)^T + \epsilon I \in \mathbb{R}^{d\times d}, \tag{3}$$

where $z_{s,i} \in \mathbb{R}^d$ denotes the i-th spatial pixel vector of the low-rank feature matrix z_s, $\epsilon = 1 \times 10^{-6}$ is an infinitesimal positive scalar to prevent the covariance matrix from becoming ill-conditioned or singular, and $I \in \mathbb{R}^{d\times d}$ is the identity matrix. Consequently, U and (μ_s, Σ_s) form lightweight geometric priors transferred to analytically guide test-time target feature alignment.

3.4 Bures-Wasserstein Initialization and Composite Prompt

Here, we employ the Bures-Wasserstein metric to establish the connection between the source and target domains, and introduce physics-informed prompts to facilitate initialization.

For an incoming test sample pair $\{\mathbf{M}_t, \widetilde{\mathbf{H}}_t\}$, we feed it into the pre-trained model f_θ to extract the baseline prediction $\mathbf{Y}_{base}$ directly from the forward pass. We obtain its flattened feature matrix

$\mathcal{F}_t \in \mathbb{R}^{N\times D}$ and project it into the low-rank subspace via $z_t = \mathcal{F}_t U \in \mathbb{R}^{N\times d}$. Then, using Eq. (2), we estimate the empirical mean μ_t and covariance Σ_t of the test data in the low-rank subspace on the fly. Substituting these statistics into the affine mapping $T(z) = Az + b$, the closed-form optimal transformation matrix $A_{BW} \in \mathbb{R}^{d\times d}$ and bias vector $b_{BW} \in \mathbb{R}^d$ that map the target distribution $\mathcal{N}(\mu_t, \Sigma_t)$ to the source distribution $\mathcal{N}(\mu_s, \Sigma_s)$ are computed as:

$$A_{BW} = \Sigma_t^{-1/2}\left(\Sigma_t^{1/2}\Sigma_s\Sigma_t^{1/2}\right)^{1/2}\Sigma_t^{-1/2}, \quad b_{BW} = \mu_s - A_{BW}\mu_t. \tag{4}$$

A_{BW} provides optimal statistical distribution alignment in a global geometric sense. However, global transformations cannot restore localized high-frequency spatial edge details at the pixel level. To compensate for the loss of fine-grained spatial textures, we exploit the undegraded physical observation from the HR-MSI to construct an observation prompt Δ_{obs}.

Specifically, we project the $\mathbf{Y}_{base}$ into the multispectral domain by multiplying it with the spectral response function matrix (SRF, denoted by $\mathbf{R} \in \mathbb{R}^{c\times C}$, which is determined by sensor hardware calibration and generally known a priori), and compute its spatial-spectral residual relative to the target HR-MSI observation $\mathbf{M}_t$, given by $(\mathbf{M}_t - \mathbf{Y}_{base}\mathbf{R}^T)$. Subsequently, we introduce a ridge-regularized back-projection pseudoinverse operator $\mathbf{R}^\dagger = \mathbf{R}^T(\mathbf{R}\mathbf{R}^T + \lambda\mathbf{I})^{-1} \in \mathbb{R}^{C\times c}$ (where $\lambda = 1\times 10^{-2}$ is a smoothing regularization) regression parameter to back-project this into the high-dimensional spectral feature space, yielding an explicit physical correction term:

$$\Delta_{obs} = (\mathbf{M}_t - \mathbf{Y}_{base}\mathbf{R}^T)\mathbf{R}^\dagger \in \mathbb{R}^{N\times D}. \tag{5}$$

3.5 Zeroth-Order Physics-Driven Optimization

Equipped with the composite initialization comprising the Bures-Wasserstein closed-form solution and the physical observation prompt, we dynamically learn a low-rank prompt matrix $\Delta W \in \mathbb{R}^{d\times d}$ during forward inference. After performing geometric scaling and prompt injection on the low-rank manifold, the adapted features are remapped back to the high-dimensional space to rectify the baseline fusion prediction. This transition process is formally formulated as:

$$\mathcal{F}_{adapted}(\Delta W) = \mathcal{F}_t + \left[z_t A_{BW}^T + b_{BW}\mathbf{1}^T - z_t + z_t\Delta W^T\right]U^T + \gamma_{obs}\cdot\Delta_{obs}, \tag{6}$$

where $\mathbf{1} \in \mathbb{R}^N$ is an all-ones vector, and $\gamma_{obs} = 0.45$ is a scalar hyperparameter weighting the physical observation prompt term. Subsequently, the adapted feature matrix $\mathcal{F}_{adapted}(\Delta W)$ is reconstructed into the candidate spatial-spectral output $\hat{\mathbf{Y}}(\Delta W) \in \mathbb{R}^{H\times W\times C}$.

To optimize the low-rank prompt ΔW without backpropagation, we formulate an unsupervised objective grounded in physical degradation consistency. According to the sensor degradation mechanism, the fitness loss function $\mathcal{L}_{fit}$ for test-time zeroth-order optimization Deng et al. (2025) is defined as the residual energy between the output and the input HR-MSI:

$$\mathcal{L}_{fit}(\Delta W) = \|\hat{\mathbf{Y}}(\Delta W)\mathbf{R}^T - \mathbf{M}_t\|_F^2, \tag{7}$$

where $\hat{\mathbf{Y}}(\Delta W)\mathbf{R}^T$ is the spectral degradation operator and $\|\cdot\|_F$ denotes the Frobenius norm.

We employ randomized Gaussian zeroth-order gradient estimation based on two-point evaluation to approximate the gradient of the objective function with respect to ΔW. At iteration g, we independently sample $P = 8$ symmetric perturbation direction matrices from a standard normal distribution: $\mathbf{Z}_i \sim \mathcal{N}(0, \mathbf{I}_{d\times d})$, $i \in \{1, \ldots, P\}$.

By executing forward evaluations along these perturbation directions, we estimate the empirical zeroth-order gradient $\hat{\nabla}\mathcal{L}_{fit}$ via finite differences:

$$\hat{\nabla}\mathcal{L}_{fit}(\Delta W^{(g)}) = \frac{1}{P}\sum_{i=1}^{P}\Big(\mathcal{L}_{fit}(\Delta W^{(g)} + \mathbf{Z}_i) - \mathcal{L}_{fit}(\Delta W^{(g)})\Big)\mathbf{Z}_i. \tag{8}$$

The update rule for the residual prompt matrix follows the zeroth-order gradient descent scheme:

$$\Delta W^{(g+1)} = \Delta W^{(g)} - \beta\hat{\nabla}\mathcal{L}_{fit}(\Delta W^{(g)}), \tag{9}$$

where β is the adaptation learning rate and set to 0.05. We set the iteration count G to 40 to balancing the performance and cost.

Table 1: Quantitative results on the **CAVE** dataset, where corruptions are exclusively applied to the LR-HSI. Abbreviations include G-Noise (Gaussian Noise), S-Noise (Stripe Noise), G-Blur (Gaussian Blur), and M-Blur (Motion Blur). NoCorr denotes performance on clean images. For each method, the first row shows PSNR (↑) and the gray-shaded row shows SAM (↓). Small colored values represent the improvement relative to the non-adapted state. Bold and underlined colored values indicate the **best** and second-best improvements, respectively.

Method	NoCorr	Environment			Sensor			Spatial			Average
		Haze	Brightness	Cloud	G-Noise	S-Noise	Band-Drop	G-Blur	M-Blur	Turbulence	
SRTTA	48.98	31.86 **+3.18**	40.33 +2.55	42.18 +2.30	39.18 +1.57	44.01 **+1.62**	42.66 +0.15	41.34 +1.89	37.52 +2.26	39.80 +1.38	39.88 +1.87
	2.23	16.74 +1.23	10.40 +0.76	3.52 +0.28	14.57 +1.31	4.82 +0.58	7.31 +0.00	4.64 +0.18	6.03 +0.50	4.70 +0.14	8.08 +0.55
IODA	48.30	33.54 +1.84	38.72 +2.59	42.53 +1.02	39.18 +0.89	44.03 +0.93	42.59 +0.18	42.07 +1.30	39.75 +1.61	42.01 +1.40	40.49 +1.31
	2.40	17.79 **-0.52**	9.97 **-1.03**	4.00 -0.12	15.54 -0.46	4.64 **-0.16**	7.26 -0.11	4.13 -0.28	4.95 -0.44	3.75 -0.41	8.00 -0.39
TAO	46.42	34.83 +0.33	39.71 +0.02	43.84 +0.14	38.79 +0.19	44.56 +0.12	41.79 +0.08	41.32 +0.27	41.16 +0.42	42.81 +0.33	40.98 +0.21
	2.88	17.54 -0.27	8.71 +0.30	4.57 -0.05	14.65 -0.58	4.14 -0.05	9.01 -0.07	4.26 -0.11	4.83 -0.23	3.73 -0.15	7.94 -0.14
SPA	48.43	25.83 +0.11	29.64 +0.25	33.85 +0.00	35.67 +0.27	39.35 +0.04	42.39 +0.03	36.52 +0.04	32.40 +0.03	35.74 +0.01	34.60 +0.09
	2.50	17.91 -0.07	13.32 -0.19	3.80 +0.00	14.33 -0.13	4.61 -0.02	7.42 +0.05	4.68 -0.02	5.60 -0.01	4.50 -0.00	8.46 -0.04
DCTTA	47.83	24.00 +0.03	27.33 +0.05	32.09 +0.02	33.89 +0.07	37.67 +0.05	42.11 +0.06	34.43 +0.09	30.40 +0.06	34.07 +0.09	32.89 +0.06
	2.55	19.88 -0.02	15.71 -0.03	4.08 -0.05	15.68 -0.02	5.39 -0.05	7.44 -0.05	5.17 -0.03	6.10 -0.04	4.64 -0.32	9.34 -0.07
UTAL	44.55	34.99 -1.43	39.78 -0.60	42.50 +0.80	39.73 +1.02	44.00 +0.96	42.05 +0.14	42.64 +1.36	41.45 +1.42	42.56 +1.10	**41.07** +0.53
	3.18	15.56 +4.08	8.55 +2.20	3.74 -0.03	8.02 -0.48	4.11 +0.03	6.91 +0.19	3.89 -0.32	4.47 -0.39	3.75 -0.19	**6.56** +0.57
THSI	42.91	25.55 +2.50	29.20 +2.55	32.77 +2.45	34.41 **+1.78**	36.40 +1.16	40.28 **+0.62**	31.18 +2.01	29.07 +2.55	32.23 **+2.50**	32.34 +2.01
	3.42	17.84 -0.12	13.93 -0.51	4.30 **-0.55**	15.67 **-1.17**	4.91 +0.43	7.81 **-0.28**	5.76 -0.40	6.37 **-0.69**	4.52 **-0.65**	9.01 **-0.44**
Ours	46.74	34.41 **+3.18**	39.59 **+2.67**	42.68 **+2.62**	38.59 +1.46	43.98 +0.92	42.16 +0.25	42.78 **+2.05**	41.32 **+3.22**	42.91 +2.48	40.93 **+2.09**
	2.82	17.31 -0.20	11.27 +1.09	3.80 -0.00	14.32 -0.89	4.61 +0.09	7.52 -0.14	4.28 **-0.44**	5.00 -0.64	3.92 -0.46	8.00 -0.18

4 EXPERIMENTS

4.1 EXPERIMENTAL CONFIGURATION

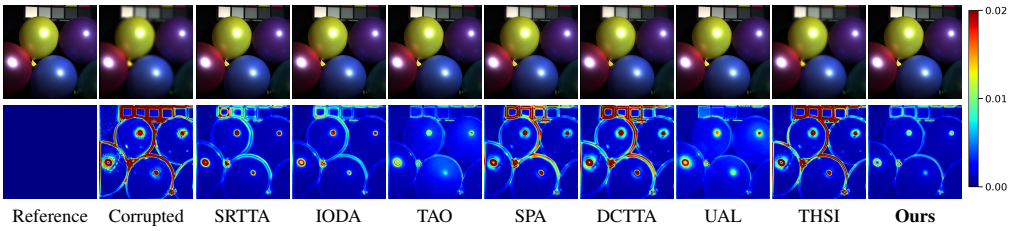


Figure 3: Visual comparison on the CAVE dataset under Gaussian blur corruption (severity 5). The top row illustrates the pseudo-color reconstructed images, while the bottom row depicts the corresponding absolute difference maps compared to the reference.

Datasets and Models. We conducted experiments on five benchmarks. On the CAVE Yasuma et al. (2010) and Harvard Chakrabarti & Zickler (2011) datasets, we construct nine distinct corruption scenarios tailored to the intrinsic physical properties of HSIs, with details provided in Appendix F. In the main manuscript, corruptions are only applied to input LR-HSIs. For remote sensing scenarios, we evaluate our method on three realistic dataset pairs: two cross-scene pairs—Pavia Center → Pavia University and HyRANK Dioni → Loukia—along with a cross-temporal pair, Houston 2013 → Houston 2018. The details of our datasets are provided in Appendix C. Following standard practice, we simulated LR-HSIs using a Gaussian blur kernel of size 5×5 with $\sigma = 2$, followed by bilinear downsampling. HR-MSIs were generated using the Spectral Response Function (SRF) of the Gaofen-1 satellite Chen et al. (2022). For all datasets, we directly feed the full-size LR-HSI and HR-MSI pairs into the model during inference without partitioning them into patches.

We comprehensively collected various open-source pixel-level TTA methods for comparison. These include SRTTA Deng et al. (2023) and IODA Tang & Yang (2024) for RGB image super-resolution,

Table 2: : Comparison on the remote-sensing cross-domain datasets. For each comparison method the first row is the method itself and the second row applies our FOTA to their source networks; both rows of a pair share the same non-adapted baseline. Improvements by Ours are highlighted in bold.

Method	Houston 2013 → 2018				Pavia Center → University				HyRANK Dioni → Loukia			
	PSNR↑	SAM↓	SSIM↑	ERGAS↓	PSNR↑	SAM↓	SSIM↑	ERGAS↓	PSNR↑	SAM↓	SSIM↑	ERGAS↓
SRTTA	44.967+0.281	2.453-0.017	0.9922+0.0003	1.428-0.022	43.789+0.044	1.922-0.002	0.9853+0.0000	1.278-0.003	50.096+0.715	1.159-0.042	0.9942+0.0009	3.616-0.141
Ours	45.331**+0.645**	2.368**-0.102**	0.9926**+0.0007**	1.416**-0.034**	43.792**+0.047**	1.921**-0.003**	0.9853**+0.0000**	1.275**-0.006**	49.719**+0.338**	1.167**-0.034**	0.9937**+0.0004**	3.615**-0.142**
IODA	38.255+1.176	4.510-0.336	0.9702+0.0076	3.176-0.170	41.924+0.214	2.199-0.026	0.9824+0.0004	1.478-0.035	49.956+0.110	1.135-0.012	0.9943+0.0002	3.298-0.093
Ours	40.066**+2.987**	4.308**-0.538**	0.9789**+0.0163**	2.941**-0.405**	42.189**+0.479**	2.162**-0.063**	0.9828**+0.0008**	1.422**-0.091**	50.073**+0.227**	1.124**-0.023**	0.9944**+0.0003**	3.293**-0.098**
TAO	34.957+0.440	6.824-0.462	0.9704+0.0072	4.545-0.140	42.018+0.202	2.201-0.025	0.9826+0.0003	1.468-0.034	37.681+3.160	4.876-0.947	0.9535+0.0170	13.247-4.581
Ours	35.407**+0.890**	6.669**-0.617**	0.9713**+0.0081**	4.458**-0.227**	42.185**+0.369**	2.182**-0.044**	0.9829**+0.0006**	1.432**-0.070**	36.418**+1.897**	5.289**-0.534**	0.9263**-0.0102**	12.288**-5.540**
SPA	36.223+1.268	6.121-0.600	0.9602+0.0070	4.026-0.207	41.791+0.129	2.293-0.020	0.9819+0.0003	1.492-0.020	45.683+0.567	1.603-0.066	0.9861+0.0018	6.272-0.639
Ours	37.199**+2.244**	6.059**-0.662**	0.9681**+0.0149**	3.941**-0.292**	41.917**+0.255**	2.276**-0.037**	0.9821**+0.0005**	1.464**-0.048**	46.925**+1.809**	1.501**-0.168**	0.9888**+0.0045**	6.761**-0.150**
DCTTA	39.092+1.625	4.447-0.733	0.9838+0.0046	2.885-0.578	41.704+0.042	2.312-0.001	0.9816+0.0000	1.509-0.003	44.888+0.137	1.987-0.017	0.9829+0.0010	7.924-2.765
Ours	38.514**+1.047**	4.797**-0.383**	0.9822**+0.0030**	3.302**-0.161**	41.914**+0.252**	2.276**-0.037**	0.9821**+0.0005**	1.465**-0.047**	45.339**+0.588**	1.927**-0.077**	0.9839**+0.0020**	10.418**-0.271**
UTAL	39.644+1.729	3.840-1.126	0.9872+0.0043	2.656-0.649	42.530+0.127	2.103-0.016	0.9835+0.0002	1.422-0.020	43.062+1.267	2.201-0.339	0.9785+0.0048	6.233-0.574
Ours	38.189**+0.274**	4.767**-0.199**	0.9837**+0.0008**	3.267**-0.038**	42.675**+0.272**	2.086**-0.033**	0.9837**+0.0004**	1.396**-0.046**	43.474**+1.679**	2.316**-0.224**	0.9834**+0.0097**	6.023**-0.784**
THSI	31.853+2.244	6.092-1.896	0.8758+0.0472	9.051-2.511	36.982+0.839	3.267-0.168	0.9681+0.0036	2.516-0.284	48.731+0.547	1.233-0.034	0.9925+0.0006	4.190-0.377
Ours	33.109**+3.500**	5.653**-2.335**	0.9098**+0.0812**	7.729**-3.834**	38.350**+2.207**	3.058**-0.377**	0.9729**+0.0084**	2.086**-0.714**	49.167**+0.983**	1.219**-0.048**	0.9931**+0.0012**	4.538**-0.029**

TAO Gou et al. (2024) for RGB reconstruction, the pixel-level TTA model SPA Niu et al. (2025) and the source-free TTA method DCTTA Tang et al. (2026). Additionally, we included TTA models specifically designed for hyperspectral image super-resolution, such as THSI Li et al. (2025c) and UTAL Zhang et al. (2024). Details about benchmarks are provided in Appendix D.

Implementation Details. We used a batch size of 1 and a learning rate of 1×10^{-4} to pre-train source models across all datasets. By default, our source model is instantiated as EDSR Lim et al. (2017) for CAVE and Harvard, UNet Ronneberger et al. (2015) for Houston and Pavia, and SwinIR Liang et al. (2021) for HyRANK. During training for all models, the HR-MSI is directly concatenated with the upsampled LR-HSI at the input stage and fed into the network without any specialized preprocessing. Experiments were conducted on an Ubuntu 22.04 server equipped with a NVIDIA L40 GPU, using Python 3.10.18 and PyTorch 2.6.0.

4.2 Results and Discussion

CAVE and Harvard Datasets. In the main manuscript, we report the PSNR and SAM metrics on the CAVE dataset in Table 1. Our method achieves the highest PSNR gain and ranks third in SAM improvement. It is worth noting that since competing methods rely on their own distinct backbones, the effectiveness of TTA is partially governed by the initial backbone performance under clean conditions as well as its inherent robustness against corruptions. Visual results are presented in Fig. 3, where our approach clearly yields the lowest reconstruction error. Additional results regarding SSIM and ERGAS improvements on this benchmark are provided in Table 12 of the Appendix. For the Harvard dataset, two detailed comparison tables are reported in Table 8 and Table 11 of the Appendix, proving that our method attains the best improvements across PSNR, SSIM, and ERGAS.

Houston, Pavia and HyRANK Datasets. The results for the cross-temporal and cross-geographic cross-domain scenarios are reported in Table 2. To eliminate the confounding effect of source backbones on TTA performance as discussed earlier, we fix the source backbone across all competing methods, enabling a rigorous and fair comparison between their respective TTA mechanisms and our proposed approach. As can be observed from the table, our method exhibits broad applicability to the MHIF task, achieving notable improvements over existing TTA methods in most cases across evaluation metrics. Furthermore, our approach incurs significantly lower test-time adaptation computational overhead, as evidenced by the complexity analysis in Table 5 of the Appendix.

4.3 Ablation Study and Sensitivity Analysis

Table 3 reports the ablation results across the nine corruption scenarios on CAVE. Disabling OT and ZO reflects the removal of our contributions in Sec. 3.4 and Sec. 3.5, respectively, verifying the efficacy of both designs. Moreover, *Eq.(5)/Eq.(7) on LR-HSI and HR-MSI* evaluates joint projection onto both HSI and MSI spaces, while *Eq.(5)/Eq.(7) on LR-HSI (no HR-MSI)* confines

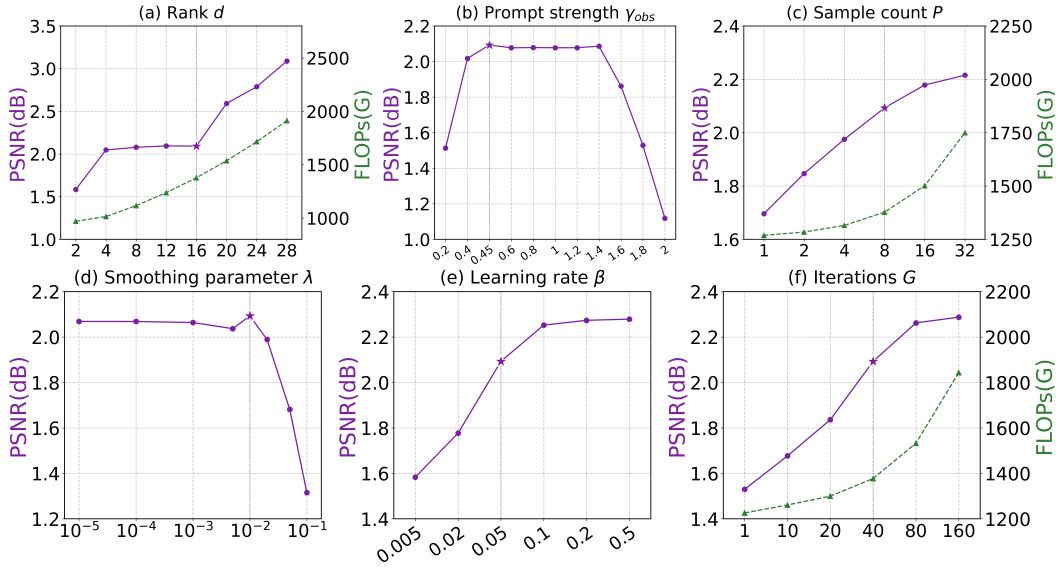


Figure 4: Hyperparameter sensitivity analysis of our proposed FOTA on the CAVE dataset.

the supervision purely to the corrupted LR-HSI domain. The results show that joint adaptation offers negligible improvement alongside increased computation, whereas omitting HR-MSI causes substantial performance collapse. These findings confirm the necessity of grounding adaptation on the trustworthy structural cues from HR-MSI.

Table 3: Ablation study on the CAVE dataset.

Variant	MSI	ZO	OT	CAVE PSNR	SAM	SSIM	ERGAS
MSI observation	✓			+1.154	-0.092	+0.0039	-0.361
+ OT	✓		✓	+1.678	-0.117	+0.0053	-0.505
+ ZO	✓	✓		+1.629	-0.078	+0.0052	-0.492
Ours	✓	✓	✓	+2.093	-0.176	+0.0228	-1.149
Eq. (5)/Eq. (7) on LR-HSI and HR-MSI				+2.098	-0.178	+0.0228	-1.151
Eq. (5)/Eq. (7) on LR-HSI (no HR-MSI)				-8.357	+2.017	-0.1093	+6.824

Sensitivity analysis is shown in Fig. 4. Increasing the rank d of the low-rank subspace in Eq. (3), the perturbation sample count P in Eq. (8), and the iteration count G in Eq. (9) monotonically boosts TTA performance; however, this comes at the expense of substantially increased computational FLOPs. As a favorable trade-off between reconstruction accuracy and inference efficiency, we set $d = 16$, $P = 8$, and $G = 40$ as our default configuration. Regarding the prompt strength γ_{obs} in Eq. (6) and the ridge regularization parameter λ of the back-projection operator, we empirically adopt their optimal values, namely $\gamma_{obs} = 0.45$ and $\lambda = 10^{-2}$. For the learning rate β in Eq. (9), we fix it to a conservative value of 0.05, rather than the empirical optimum, as larger learning rates destabilize the effects of other hyperparameters.

5 Conclusion

In this paper, we propose Forward-Only Test-Time Adaptation (FOTA), the first BP-free, architecture-agnostic TTA framework specifically tailored for the MHIF task. FOTA models TTA as a physics-driven continuous prompt optimization within a compact tensor low-rank subspace. By leveraging the closed-form BW optimal transport, our framework achieves instantaneous geometric feature alignment, while a physical degradation consistency objective guides forward-only zeroth-order optimization to refine the prompts. Extensive experiments across synthetic corruptions and real-world cross-scene and cross-temporal benchmarks demonstrate that FOTA outperforms existing TTA competitors while incurring minimal computational overhead. In future research, efforts by our team and the community could focus on the following directions: 1) extending the proposed framework to a broader spectrum of hyperspectral image processing tasks; 2) constructing larger-scale training benchmarks with authentic physical degradations to bolster the intrinsic generalization of source models; and 3) standardizing backbone architectures across future adaptation benchmarks to facilitate rigorous and fair comparisons among test-time adaptation algorithms.

AI USE STATEMENT

In this work, we used generative AI tools for assisting with translation, cleaning and reformatting datasets, and interpreting results. We have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, and implementing methods, and the rest of the required disclosure tasks (supporting qualitative and thematic data analysis) are not applicable to this work. Additionally, we used generative AI tools for brainstorming, suggesting a structure for a research paper, summarizing or analysing existing literature, and drafting parts of a research paper. We have reviewed all AI-assisted work. Specifically, all data cleaning scripts and reformatted data were manually verified to ensure data integrity without unintended modifications; all AI-assisted interpretations of experimental results were critically cross-checked against raw metric outputs, with final conclusions drawn independently by the authors; all literature summarized by LLMs was manually traced back to its original source publication to avoid hallucinated citations; and all AI-assisted translations and drafted text were thoroughly revised and proofread by the authors to guarantee technical and linguistic accuracy. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Jón Atli Benediktsson, Jón Aevar Palmason, and Johannes R Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):480–491, 2005.

Nassim Ait Ali Braham, Conrad M Albrecht, Julien Mairal, Jocelyn Chanussot, Yi Wang, and Xiao Xiang Zhu. Spectralearth: Training hyperspectral foundation models at scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *CVPR 2011*, pp. 193–200. IEEE, 2011.

Guohao Chen, Shuaicheng Niu, Chen Deyu, Jiahao Yang, Zitian Zhang, Mingkui Tan, Pengcheng Wu, and Zhiqi Shen. Zerosiam: An efficient asymmetry for test-time entropy optimization without collapse. In *International Conference on Learning Representations*, volume 2026, pp. 144893–144925, 2026.

Liangfu Chen, Husi Letu, Meng Fan, Huazhe Shang, Jinhua Tao, Laixiong Wu, Ying Zhang, Chao Yu, Jianbin Gu, Ning Zhang, et al. An introduction to the chinese high-resolution earth observation system: Gaofen-1˜ 7 civilian satellites. *Journal of Remote Sensing*, 2022.

Christian Debes, Andreas Merentitis, Roel Heremans, Jürgen Hahn, Nikolaos Frangiadakis, Tim Van Kasteren, Wenzhi Liao, Rik Bellens, Aleksandra Pižurica, Sidharta Gautama, et al. Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2405–2418, 2014.

Fabio Dell'Acqua, Paolo Gamba, Alessio Ferrari, Jon Aevar Palmason, Jón Atli Benediktsson, and Kolbeinn Árnason. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geoscience and Remote Sensing Letters*, 1(4):322–326, 2004.

Zeshuai Deng, Zhuokun Chen, Shuaicheng Niu, Thomas Li, Bohan Zhuang, and Mingkui Tan. Efficient test-time adaptation for super-resolution with second-order degradation and reconstruction. *Advances in Neural Information Processing Systems*, 36:74671–74701, 2023.

Zeshuai Deng, Guohao Chen, Shuaicheng Niu, Hui Luo, Shuhai Zhang, Yifan Yang, Renjie Chen, Wei Luo, and Mingkui Tan. Test-time model adaptation for quantized neural networks. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 7258–7267, 2025.

Renwei Dian, Shutao Li, Anjing Guo, and Leyuan Fang. Deep hyperspectral image sharpening. *IEEE Transactions on Neural Networks and Learning Systems*, 29(11):5345–5355, 2018. doi: 10.1109/TNNLS.2018.2798162.

Yao-Ze Feng and Da-Wen Sun. Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in food science and nutrition*, 52(11):1039–1058, 2012.

Peyré Gabriel and Cuturi Marco. Computational optimal transport with applications to data sciences. *Foundations and trends® in machine learning*, 11(5-6):355–607, 2019.

Yuanbiao Gou, Haiyu Zhao, Boyun Li, Xinyan Xiao, and Xi Peng. Test-time degradation adaptation for open-set image restoration. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 16167–16177. PMLR, 2024.

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, pp. 222–241. Springer, 2024.

Jiayi Guo, Junhao Zhao, Chaoqun Du, Yulin Wang, Chunjiang Ge, Zanlin Ni, Shiji Song, Humphrey Shi, and Gao Huang. Everything to the synthetic: Diffusion-driven test-time adaptation via synthetic-domain alignment. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30503–30513. IEEE, 2025.

Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *20th International Conference on Pattern Recognition*, pp. 2366–2369. IEEE, 2010.

Yusuke Iwasawa and Yutaka Matsuo. Test-time classifier adjustment module for model-agnostic domain generalization. *Advances in Neural Information Processing Systems*, 34:2427–2440, 2021.

Konstantinos Karantzalos, Christina Karakizi, Zacharias Kandylakis, and Georgia Antoniou. HyRANK: The New ISPRS Hyperspectral Benchmark. Scientific initiatives final report, International Society for Photogrammetry and Remote Sensing (ISPRS), 2018.

Abdolraheem Khader, Guanyiman Fu, Jue Zhang, Jingxiang Yang, Jianfeng Lu, and Liang Xiao. Tm-moe: A degradation-aware transformer-mamba moe for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2026.

Baisong Li, Xingwang Wang, and Haixiao Xu. Pif-net: Ill-posed prior guided multispectral and hyperspectral image fusion via invertible mamba and fusion-aware lora. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 5964–5972, 2026.

Jiaxin Li, Ke Zheng, Lianru Gao, Zhu Han, Zhi Li, and Jocelyn Chanussot. Enhanced deep image prior for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–18, 2025a.

Jingtao Li, Yingyi Liu, Xinyu Wang, Yunning Peng, Chen Sun, Shaoyu Wang, Zhendong Sun, Tian Ke, Xiao Jiang, Tangwei Lu, et al. Hyperfree: A channel-adaptive and tuning-free foundation model for hyperspectral remote sensing imagery. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 23048–23058, 2025b.

Ke Li, Luc Van Gool, and Dengxin Dai. Test-time training for hyperspectral image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(9):7231–7242, 2025c. doi: 10.1109/TPAMI.2024.3461807.

Xiang Li, Yong Tao, Siyuan Zhang, Siwei Liu, Zhitong Xiong, Chunbo Luo, Lu Liu, Mykola Pechenizkiy, Xiao Xiang Zhu, and Tianjin Huang. REOBench: Benchmarking robustness of earth observation foundation models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2025d.

Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779*, 2016.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *2021 IEEE/CVF international conference on computer vision workshops (ICCVW)*, pp. 1833–1844. IEEE, 2021.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pp. 1132–1140. IEEE, 2017.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *International conference on machine learning*, pp. 16888–16905. PMLR, 2022.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. *arXiv preprint arXiv:2302.12400*, 2023.

Shuaicheng Niu, Chunyan Miao, Guohao Chen, Pengcheng Wu, and Peilin Zhao. Test-time model adaptation with only forward passes. *arXiv preprint arXiv:2404.01650*, 2024.

Shuaicheng Niu, Guohao Chen, Peilin Zhao, Tianyi Wang, Pengcheng Wu, and Zhiqi Shen. Self-bootstrapping for versatile test-time adaptation. In *Proceedings of the 42nd International Conference on Machine Learning*. PMLR, 2025.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.

Michal Shimoni, Rob Haelterman, and Christiaan Perneel. Hyperspectral imaging for military and security applications: Combining myriad processing and sensing techniques. *IEEE Geoscience and Remote Sensing Magazine*, 7(2):101–117, 2019.

Lijuan Su, Yuxiao Sui, and Yan Yuan. An unmixing-based multi-attention gan for unsupervised hyperspectral and multispectral image fusion. *Remote Sensing*, 15(4):936, 2023.

Le Sun, Jianxiao Zhou, Qiaolin Ye, Zebin Wu, Qiao Chen, Zhongqi Xu, and Liyong Fu. Mdc-fusformer: Multiscale deep cross-fusion transformer network for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

Ni Tang, Shenghao Nie, Xiaotong Luo, Yuan Xie, and Yanyun Qu. Degradation-consistent test-time adaptation for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15476–15485, 2026.

Zai-Zuo Tang and Yu-Bin Yang. Ioda: instance-guided one-shot domain adaptation for super-resolution. *Advances in Neural Information Processing Systems*, 37:117291–117314, 2024.

Lucien Wald. Quality of high resolution synthesised images: Is there a simple criterion? In *Third conference" Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images"*, pp. 99–103. SEE/URISCA, 2000.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020.

Di Wang, Meiqi Hu, Yao Jin, Yuchun Miao, Jiaqi Yang, Yichu Xu, Xiaolei Qin, Jiaqi Ma, Lingyu Sun, Chenxing Li, et al. Hypersigma: Hyperspectral intelligence comprehension foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Xingyu Wang and Tao Wang. Fozo: Forward-only zeroth-order prompt optimization for test-time adaptation. *arXiv preprint arXiv:2603.04733*, 2026.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Honghui Xu, Chuangjie Fang, Yiqun Meng, Jiawei Jiang, Sixian Chan, Shiqing Zhang, and Jianwei Zheng. Tptransformer: Tensor-tensor product transformer for hyperspectral image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1670–1679, 2026.

Jin Xu, Yuanchao Su, Xu Sun, Jiaxin Li, Haiwei Li, Jianjian Gao, Xiaohua Feng, and Ronghua Liu. Fusgat: Graph attention-based fusion network for unsupervised hyperspectral image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 22:1–5, 2025.

Yonghao Xu, Bo Du, Liangpei Zhang, Daniele Cerra, Miguel Pato, Emiliano Carmona, Saurabh Prasad, Naoto Yokoya, Ronny Hänsch, and Bertrand Le Saux. Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 ieee grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(6):1709–1724, 2019.

Jun Yan, Kai Zhang, Qinzhu Sun, Chiru Ge, Wenbo Wan, Jiande Sun, and Huaxiang Zhang. Spatial–spectral unfolding network with mutual guidance for multispectral and hyperspectral image fusion. *Pattern Recognition*, 161:111277, 2025.

Jingxiang Yang, Yong-Qiang Zhao, and Jonathan Cheung-Wai Chan. Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. *Remote Sensing*, 10(5):800, 2018.

Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010.

Roberta H Yuhas, Alexander FH Goetz, and Joe W Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop*, 1992.

Jingcheng Zhang, Yanbo Huang, Ruiliang Pu, Pablo Gonzalez-Moreno, Lin Yuan, Kaihua Wu, and Wenjiang Huang. Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and electronics in agriculture*, 165:104943, 2019.

Lei Zhang, Jiangtao Nie, Wei Wei, and Yanning Zhang. Unsupervised test-time adaptation learning for effective hyperspectral image super-resolution with unknown degeneration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):5008–5025, 2024.

Mingjin Zhang, Zhongkai Yang, Jie Guo, and Yunsong Li. Mosymamba: Modal-synergy mamba network for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2026a.

Ronghao Zhang, Shuaicheng Niu, Qi Deng, Yanjie Dong, Jian Chen, and Runhao Zeng. Zotta: Test-time adaptation with gradient-free zeroth-order optimization. *arXiv preprint arXiv:2603.14254*, 2026b.

Jian Zhu, He Wang, Yang Xu, Zebin Wu, and Zhihui Wei. Self-learning hyperspectral and multispectral image fusion via adaptive residual guided subspace diffusion model. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17862–17871. IEEE, 2025.

A SUPPLEMENTARY MATERIAL OUTLINE

The outline of the supplementary material is as follows:

1. Section B presents the **related work** for the Multispectral and Hyperspectral Image Fusion (MHIF) task.
2. Section C introduces the **datasets** used in our paper.
3. Section D provides a brief overview of the **benchmarks** used for the comparative experiments.
4. Section E details the mathematical definitions of the **evaluation metrics** employed for the quantitative assessment of the experimental outcomes.
5. Section F provides detailed formulations and visualizations of the **corruptions** tailored to hyperspectral imagery.
6. Section G presents the **complexity analysis** of the proposed method.
7. Section H investigates the architecture-agnostic property of our proposed FOTA across **diverse backbones**.
8. Section I evaluates the generalization capability of different TTA methods across varying super-resolution **scaling factors**.
9. Section K investigates the robustness of adaptation performance under varying **spectral response functions (SRFs) and MSI distortions**.
10. Section L provides more **ablation studies** on BW-initialization of Eq. (4).
11. Section J provides additional **quantitative and visual results**.

B RECENT PROGRESS IN DEEP LEARNING-BASED MHIF

In recent years, deep learning-based methods have significantly advanced the MHIF task due to their powerful representation capability. Early approaches mainly focused on convolutional neural networks (CNNs). Under the supervised learning paradigm, additional sensor parameters are generally not required. DHSIS Dian et al. (2018) combines a physical imaging model with CNN-based nonlinear mapping for HR-HSI reconstruction. TwoCNNFu Yang et al. (2018) employs a dual-branch CNN to separately capture spectral and spatial information from HSIs and MSIs. Transformer-based methods have subsequently been introduced to overcome the limited receptive fields of CNNs and model long-range dependencies. MDCFormer Sun et al. (2024) utilizes a multiscale spatial Transformer with spectral attention for global feature modeling, and SMGU-Net Yan et al. (2025) proposes a spatial-spectral mutual guidance mechanism to improve information exchange between different imaging modalities. More recently, Mamba-based methods have been explored for efficient long-range spatial-spectral modeling with linear computational complexity. To handle cross-modal interactions in state space, MosyMamba Zhang et al. (2026a) introduces a synergistic scanning mechanism to jointly encode spatial details and spectral signatures. Meanwhile, TM-MOE Khader et al. (2026) combines 3-D Mamba scanning with mixture-of-experts routing to efficiently model spatial-spectral dependencies across modalities. For the unsupervised methods, some enable real-time training; however, they typically require physical imaging priors, such as the point spread function (PSF) and spectral response function (SRF), to construct spatial and spectral degradation losses Li et al. (2025a). Alternatively, other approaches decompose the hyperspectral image into two physically meaningful matrices, such as the abundance and endmember matrices, through tensor or matrix factorization Su et al. (2023); Xu et al. (2025).

C DATASETS

This section details the benchmark datasets. All referenced datasets are publicly accessible online.

(1) **CAVE:** Yasuma et al. (2010) Acquired using a tunable filter and a cooled CCD camera under controlled illumination, this dataset contains 32 indoor scenes of various materials and objects. Each scene provides a spatial resolution of 512×512 pixels with 31 spectral bands covering the 400 to 700



(a) balloons (b) beads (c) clay (d) fake/real food (e) lemon slices



(f) imgf3 (g) imgf4 (h) imgf5 (i) imgf6 (j) imgf7

Figure 5: False-color composite visualizations of the hyperspectral datasets: the first row shows samples from the CAVE dataset ((a)–(e), bands 25-15-5), and the second row shows samples from the Harvard dataset ((f)–(j), bands 23-13-3).



(a) Houston 2013



(b) Houston 2018



(c) Pavia Center



(d) Pavia University



(e) Dioni



(f) Loukia

Figure 6: False-color composite visualizations of the hyperspectral datasets: (a) Houston 2013 and (b) Houston 2018 (bands 19-12-6); (c) Pavia Center and (d) Pavia University (bands 52-28-7); (e) Dioni and (f) Loukia (bands 22-13-5).

nm range at 10 nm intervals. We select 12 commonly used images as the test set, while the remaining 20 images are used for training. Subfigures (a)–(e) in Fig. 5 illustrate representative samples from the CAVE dataset.

(2) **Harvard:** Chakrabarti & Zickler (2011) The Harvard dataset comprises 50 hyperspectral images of diverse indoor and outdoor scenes captured under natural daylight. Each image contains 31 spectral bands ranging from 400 to 700 nm with a 10 nm sampling interval, featuring a spatial resolution of 1040×1392. In our experiments, 40 samples are selected for training, while the remaining 10 are reserved for testing. Subfigures (f)–(j) in Fig. 5 depict exemplary samples from the Harvard dataset.

(3) **Houston 2013:** Debes et al. (2014) Gathered by the ITRES CASI-1500 sensor over the University of Houston campus and neighboring urban area by *the National Center for Airborne Laser Mapping (NCALM)*, this dataset was featured in the 2013 IEEE GRSS Data Fusion Contest. The hyperspectral

image encompasses 349×1905 pixels with 144 spectral bands covering the 364 to 1046 nm range, achieving a spatial resolution of 2.5 m and a spectral resolution of approximately 10 nm.

(4) **Houston 2018:** Xu et al. (2019) Collected by the ITRES CASI-1500 sensor over the same University of Houston region by *the National Center for Airborne Laser Mapping (NCALM)*, this scene was released for the 2018 IEEE GRSS Data Fusion Contest. The original hyperspectral image comprises 1202×4172 pixels with 48 spectral bands spanning the 380 to 1050 nm range at 1 m spatial resolution. After clipping and resampling to match the ground truth extent, the processed image retains 601×2384 pixels. Houston 2013 $\rightarrow$ Houston 2018: This pair is a standard benchmark for assessing cross-temporal adaptation in HSI classification task . By convention, we utilize Houston 2013 as the source domain and Houston 2018 as the target domain to evaluate the model's robustness against temporal variations. Fig. 6 (a)–(b) show the false-color visualizations of Houston 2013 and Houston 2018.

(5) **Pavia Center:** Benediktsson et al. (2005) This dataset was captured by the ROSIS instrument over the center of Pavia, Italy. After cropping out regions lacking information, the image size is 1096×1096 pixels. It features a spatial resolution of 1.3 m and encompasses 102 spectral bands ranging from 0.43 to 0.86 μm.

(6) **Pavia University:** Dell'Acqua et al. (2004) This scene was recorded by the ROSIS sensor over the University of Pavia, Italy. Discarding uninformative samples yields a spatial dimension of 610×340 pixels. It provides 103 spectral bands within the 0.43 to 0.86 μm range, possessing a spatial resolution of 1.3 m. Pavia Center $\rightarrow$ Pavia University: This pair is frequently used to verify the cross-spatial scene adaptation capability of algorithms in the HSI classification task. We select them for our cross-domain task. In our experiments, Pavia Center serves as the source domain, while Pavia University is designated as the target domain. Following standard practice, the final spectral band of the Pavia University dataset is removed to maintain spectral consistency, allowing the model to be directly deployed on both datasets without architectural adaptation. The false-color representations of Pavia Center and Pavia University are depicted in Fig. 6 (c)–(d).

(7) **Dioni:** Karantzalos et al. (2018) This scene belongs to the HyRANK hyperspectral satellite dataset, which was developed within the framework of the ISPRS Scientific Initiatives by the Remote Sensing Laboratory of the National Technical University of Athens. It was acquired by the Hyperion imaging spectrometer onboard NASA's Earth Observing-1 (EO-1) satellite, which records the visible, near-infrared and short-wave infrared spectrum, nominally from 400 to 2500 nm. Discarding the uncalibrated, overlapping and water-vapour-affected channels from the 242 potential Hyperion bands leaves 176 bands per pixel. The image contains 250×1376 pixels at a spatial resolution of 30 m, and its reference map provides annotated samples for 12 of the 14 land-cover classes defined by the benchmark.

(8) **Loukia:** Karantzalos et al. (2018) Recorded by the same Hyperion sensor over a second, geographically distinct region, this scene provides 249×945 pixels with the same 30 m spatial resolution and the identical 176 spectral bands. Its reference map contains all 14 land-cover classes. Dioni $\rightarrow$ Loukia: Unlike the Pavia and Houston pairs, which consist of two acquisitions of the same urban area, the two HyRANK scenes cover two geographically distinct regions, so this pair exposes the model to a change of land cover in addition to a change of acquisition conditions. It therefore serves as a cross-region benchmark for HSI-MSI fusion, complementing the cross-spatial and cross-temporal protocols above. In our experiments, Dioni serves as the source domain and Loukia is designated as the target domain. Fig. 6 (e)–(f) display false-color composites for the Dioni and Loukia datasets.

D THE BENCHMARKS

In this section, we provide a brief overview of the competing methods used for comparison in our experiments, including traditional fusion algorithms and recent deep learning-based approaches.

- **SRTTA (NeurIPS 2023)** Deng et al. (2023): We adapt the SRTTA framework for test-time adaptation on the MHIF task. Specifically, a pre-trained multi-label degradation classifier $C(\cdot)$ is first employed to predict the specific degradation categories (e.g., blur or noise) present in the input LR-HSI. Based on these predictions, we implement a second-order degradation scheme to synthesize a further degraded version of the LR-HSI, denoted as x_{sd}, which serves as pseudo-paired data $\{x_{sd}, x\}$ for self-supervised learning. The network

is then optimized by minimizing a combined second-order reconstruction loss: (i) a self-supervised adaptation loss $\mathcal{L}_s$ that encourages the model to reconstruct the original LR-HSI x from the second-order version x_{sd} at the feature level; (ii) an adaptation consistency loss $\mathcal{L}_a$ that constrains the adapted features of x_{sd} to remain consistent with the original pre-trained features of x, thereby preventing model collapse. We set the trade-off hyperparameter α to 1.0 and update the ResBlocks of the fusion backbone using the Adam optimizer with a learning rate of 5×10^{-5} for $S = 10$ iterations per sample. Furthermore, to mitigate catastrophic forgetting, the Adaptive Parameter Preservation (APP) strategy is applied to freeze the top 50% ($\rho = 0.50$) of the most important parameters identified via the Fisher information matrix .

- **IODA (NeurIPS 2024)** Tang & Yang (2024): We implement the instance-guided target domain distribution expansion strategy on our MHIF TTA task. Specifically, the Segment Anything Model (SAM) is utilized to segment each instance within the input LR-HSI to generate a collection of instance masks $Mask_{ALL}$. These masks are randomly concatenated and used as region-range masks for an Alpha-CLIP image encoder, enabling the model to generate diverse instance-specific features focused on different image areas. We further incorporate random occlusion masks to enhance distribution diversity. Domain adaptation is achieved by minimizing the $L_{direction}$ loss, which enforces collinearity between the SR cross-domain direction vector and the LR cross-domain direction vector in the Alpha-CLIP latent space, providing pixel-level granularity for texture detail reconstruction. Following the original implementation, the target network M_T is initialized with source domain weights and updated for each test sample to mitigate the negative impact of domain gaps.

- **TAO (ICML 2024)** Gou et al. (2024): TAO utilizes a pre-trained, degradation-agnostic image diffusion model (PDM) as a foundation to generate clean images from the degraded input LR-HSI. The adaptation consists of two core components: (i) a test-time degradation adapter (TDA), implemented as a four-layer convolutional network, which is optimized to align the generative domain of the PDM with the specific degraded domain of the test sample; and (ii) an adapter-guided image restoration (AIR) module that dynamically adjusts supervision strategies across different denoising stages. Specifically, the adapter is updated once at each denoising step by minimizing a combined objective of reconstruction loss $\mathcal{L}_{rec}$, perceptual loss $\mathcal{L}_{pec}$, and adversarial loss $\mathcal{L}_{adv}$. The denoising process follows a three-stage heuristic supervision: global content reconstruction (steps 999–700), perceptual and adversarial alignment (steps 700–50), and total variation (TV) regularization for fine details (steps 50–0). We use the Adam optimizer with a learning rate of 1×10^{-3} and perform adaptation on a single test sample to recover high-fidelity texture details while maintaining the integrity of the spectral information.

- **SPA (ICML 2025)** Niu et al. (2025): The method establishes a self-bootstrapping paradigm that optimizes prediction consistency between the original image (strong view) and its deteriorated version (weak view). To make SPA applicable to the pixel-level fusion task, we employ geometry-preserving augmentations that maintain object sizes and locations while providing sufficient learning signals: (i) random low-frequency amplitude masking in the Fourier domain, which supplies rich signals by altering high-power frequency components; and (ii) high-frequency noise injection to enhance information power across all spatial frequencies. Since MHIF is a regression task, we utilize the L_1 loss to align the predictions of the two views. Following the configurations for pixel-level tasks, we set the mask ratio $m = 0.2$ and the noise factor $\gamma = 0.1$. The model is updated using the Adam optimizer for $S = 10$ iterations per test sample to achieve a task-agnostic and stable adaptation.

- **DCTTA (CVPR 2026)** Tang et al. (2026): DCTTA adapts a teacher-student framework to achieve source-free test-time adaptation under unknown degradation shifts. Specifically, geometric augmentations are applied to the input to construct an ensemble pseudo-label $\overline{y}$ via the teacher network, and a diffusion-based generator (RDDM) simulates the residual discrepancy to synthesize a redegraded counterpart x_{sd}. The student network is optimized by minimizing a combined objective comprising a self-supervised consistency loss $\mathcal{L}_s$ between outputs $\{y_{in}, y_{sd}\}$ and an adaptation consistency loss $\mathcal{L}_a$ between y_{sd} and $\overline{y}$, each combining pixel-wise L_1 and perceptual constraints. We update the student network using the Adam optimizer with a learning rate of 2×10^{-4} while updating the teacher via EMA with decay $\eta = 0.95$. Furthermore, the Test-Time Important Parameter Selection (TIPS) module freezes

the top 40% ($\rho = 0.40$) critical parameters computed via the Fisher information matrix to prevent catastrophic forgetting.

- **UTAL (TPAMI 2024)** Zhang et al. (2024): UTAL mitigates the impact of unknown degradation by adapting a content-agnostic prior, which is pre-trained on extensive synthetic data, to the private characteristics of each specific latent high-resolution hyperspectral image (HR-HSI) at test time. The framework utilizes a mutual-guiding fusion module (MGM) to implicitly learn the spatial-spectral relationship between the observed images. During the test-time adaptation phase, the network parameters are refined by minimizing an unsupervised reconstruction loss $\mathcal{L}_{rec}$, which enforces consistency between the observed LR-HSI/HR-MSI and the degraded versions of the estimated HR-HSI under the estimated or given degradation model. Following the original implementation, we use the Adam optimizer with a learning rate of 1×10^{-5} and perform the adaptation for 100 iterations per test sample to achieve image-specific prior refinement.
- **THSI (TPAMI 2025)** Li et al. (2025c):THSI employs a teacher-student self-training scheme where a weight-averaged teacher model Θ' generates pseudo-labels to guide the student model Θ. To ensure an accurate LR-HR relationship, the adaptation process minimizes the consistency loss $\mathcal{L}(\Theta'_t(x), \Theta_t(\downarrow \langle \Theta_t(x) \rangle))$, where the pseudo-labels are synthesized from the teacher's prediction and the input is its downsampled version. For each test instance, the student model is updated for $T = 20$ epochs using the Adam optimizer with a learning rate of 10^{-5}. The teacher model is updated via exponential moving average (EMA) with a smoothing factor $\alpha = 0.99$. Furthermore, to enhance data diversity at test time, the Spectral Mixup augmentation is implemented with a fidelity balance parameter $\lambda = 0.5$, synthesizing virtual samples through convex combinations of spectral bands. The total training objective combines the L_1 loss and the spatial-spectral total variation (SSTV) loss to preserve both spatial structures and spectral consistency during fusion.

E Quantitative Evaluation Metrics

To comprehensively evaluate the performance of the competing methods, we employ five widely adopted quantitative metrics.

(1) **Peak Signal-to-Noise Ratio (PSNR):** PSNR is a fundamental metric for assessing spatial reconstruction quality Hore & Ziou (2010). It evaluates the ratio between the maximum possible power of a signal and the power of corrupting noise, defined for the c-th channel as:

$$\mathrm{PSNR} = 10 \cdot \log_{10} \left(\frac{\max(\mathbf{R}_c)^2}{\frac{1}{HW} \|\mathbf{R}_c - \mathbf{Z}_c\|_2^2} \right), \tag{1}$$

where $\mathbf{R}_c$ and $\mathbf{Z}_c$ represent the c-th spectral band of the ground-truth reference and the fused image, respectively, and $\|\cdot\|_2$ denotes the ℓ_2 norm. The overall PSNR of the image is typically calculated by averaging across all spectral channels. A higher PSNR value indicates superior spatial fidelity.

(2) **Structural Similarity (SSIM):** SSIM evaluates the structural consistency between two images, aligning closely with the perception of the human visual system Wang et al. (2004). The metric is formulated as:

$$\mathrm{SSIM} = [l(\mathbf{R}, \mathbf{Z})]^\alpha \cdot [c(\mathbf{R}, \mathbf{Z})]^\beta \cdot [s(\mathbf{R}, \mathbf{Z})]^\gamma, \tag{2}$$

$$l(\mathbf{R}, \mathbf{Z}) = \frac{2\mu_{\mathbf{R}}\mu_{\mathbf{Z}} + C_1}{\mu_{\mathbf{R}}^2 + \mu_{\mathbf{Z}}^2 + C_1}, \tag{3}$$

$$c(\mathbf{R}, \mathbf{Z}) = \frac{2\sigma_{\mathbf{R}}\sigma_{\mathbf{Z}} + C_2}{\sigma_{\mathbf{R}}^2 + \sigma_{\mathbf{Z}}^2 + C_2}, \tag{4}$$

$$s(\mathbf{R}, \mathbf{Z}) = \frac{\sigma_{\mathbf{RZ}} + C_3}{\sigma_{\mathbf{R}}\sigma_{\mathbf{Z}} + C_3}, \tag{5}$$

where l, c, and s model the similarities in luminance, contrast, and spatial structure, respectively. The weighting exponents α, β, γ are conventionally set to 1. The parameters $\mu_{\mathbf{R}}$ and $\mu_{\mathbf{Z}}$ represent the mean values, $\sigma_{\mathbf{R}}$ and $\sigma_{\mathbf{Z}}$ are the standard deviations, and $\sigma_{\mathbf{RZ}}$ denotes the covariance between $\mathbf{R}$ and $\mathbf{Z}$. C_1, C_2, C_3 are small constants added for numerical stability. A higher SSIM value reflects better structural preservation.

(3) **Spectral Angle Mapper (SAM):** SAM is a critical metric for quantifying spectral distortion Yuhas et al. (1992). It treats the spectral signature of each pixel as a high-dimensional vector and measures the angle between the reference and the reconstructed spectra:

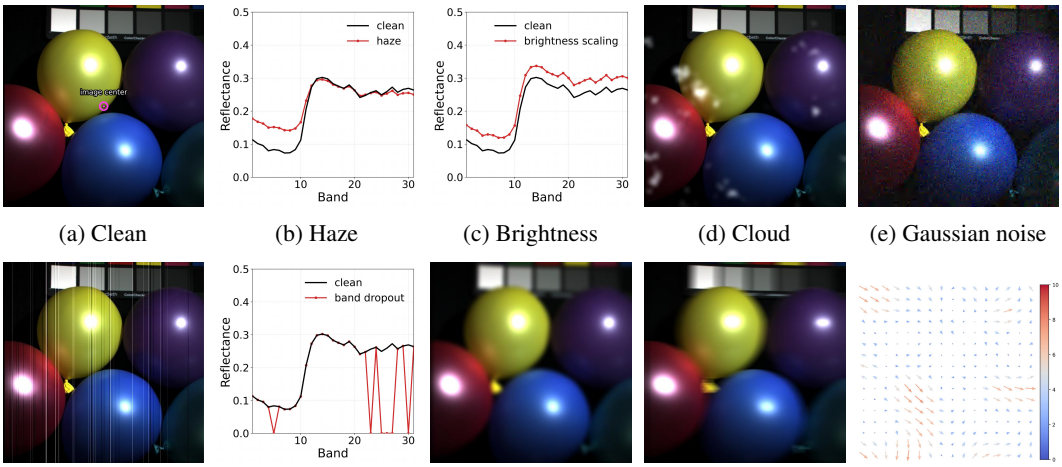
$$\mathrm{SAM} = \cos^{-1}\left(\frac{\sum_{c=1}^{C}\mathbf{r}_c\mathbf{z}_c}{\sqrt{\sum_{c=1}^{C}\mathbf{r}_c^2}\sqrt{\sum_{c=1}^{C}\mathbf{z}_c^2}}\right), \tag{6}$$

where $\mathbf{r}_c$ and $\mathbf{z}_c$ are the spectral vectors of the reference and predicted images. A smaller SAM value (closer to zero) signifies higher spectral fidelity.

(4) **Relative Dimensionless Global Error in Synthesis (ERGAS):** ERGAS offers a comprehensive global assessment of the fused image quality, taking into account both the mean squared error and the mean radiance of each band Wald (2000). It is mathematically defined as:

$$\mathrm{ERGAS} = \frac{100}{r}\sqrt{\frac{1}{C}\sum_{c=1}^{C}\frac{\|\mathbf{R}_c - \mathbf{Z}_c\|_2^2}{\mu^2\left(\mathbf{R}_c\right)}}, \tag{7}$$

where r denotes the spatial downsampling ratio between the low-resolution and high-resolution images, and $\mu\left(\mathbf{R}_c\right)$ is the mean value of the c-th reference band. Lower ERGAS values indicate better overall fusion performance.



(a) Clean (b) Haze (c) Brightness (d) Cloud (e) Gaussian noise

(f) Stripe noise (g) Band dropout (h) Gaussian blur (i) Motion blur (j) Turbulence field

Figure 7: Visualization results of various types of corruption applied to the HSIs are presented. For corruptions that intuitively impact the spectral dimension, we represent them using the spectral signatures of the center pixel. In contrast, the turbulence distortion is illustrated through the pixel displacement field in the spatial domain. All results reflect a severity level of S5, which is used as the basic level in the CAVE and Harvard datasets. The visualized images are obtained from the `balloons` picture in the CAVE dataset.

F CORRUPTIONS

In this section, we introduce the various types and severity levels of corruptions applied to the CAVE and Harvard datasets. To simulate the common corruptions encountered by hyperspectral images in real-world scenarios, we consider three major categories of degradation: environmental degradations (haze, brightness change, and cloud), sensor distortions (Gaussian noise, stripe noise, and random band dropout), and spatial degradations (Gaussian blur, motion blur, and turbulence distortion). Note that most corruption types and their associated parameters are adapted from REOBench Li et al. (2025d). We further adjust these parameters according to the unique characteristics of hyperspectral images and introduce new degradation types across the spectral bands. Detailed parameter settings are summarized in Table 4. It should be noted that the max drift parameter in

atmospheric turbulence denotes the maximum spatial displacement of an individual pixel induced by aerodynamic disturbances and refractive fluctuations. In addition, qualitative visual comparisons on the CAVE dataset are presented in Fig. 7. For a fair comparison, all corruption experiments on both CAVE and Harvard datasets are conducted under severity level 5 (S5).

Table 4: Configurations and severity levels of corruption types tailored to hyperspectral imagery. Most parameters are directly inherited from REOBench Li et al. (2025d), with specific modifications made to account for the unique characteristics of hyperspectral images.

Corruption Type	**Parameter**	**S1**	**S2**	**S3**	**S4**	**S5**
Haze	intensity	0.200	0.225	0.250	0.275	0.300
Brightness change	b / c	$+0.000/1.000$	$+0.013/0.987$	$+0.025/0.975$	$+0.038/0.962$	$+0.050/0.950$
Cloud	threshold	0.950	0.893	0.835	0.778	0.720
Gaussian Noise	σ	0.04	0.05	0.06	0.07	0.08
Stripe Noise	amplitude	0.02	0.04	0.07	0.11	0.16
Band Dropout	ratio	0.02	0.04	0.07	0.12	0.18
Gaussian Blur	kernel size	5×5	9×9	13×13	17×17	21×21
Motion Blur	kernel size	5×5	9×9	15×15	23×23	31×31
Turbulence	max drift	4.0	5.0	6.0	7.0	8.0

G Model Complexity Analysis

We conduct a complexity analysis across different TTA methods, with the results summarized in Table 5. Note that since the existing comparison methods are trained independently and lack a unified source model architecture, their TTA performance is inherently influenced by the complexity of the underlying source models. Furthermore, owing to the limited exploration of test-time adaptation algorithms in hyperspectral image super-resolution or MHIF tasks, extending methods originally tailored for RGB tasks to the MHIF scenario may also induce performance bias. Overall, our bp-free method ranks first with the fewest updated parameters, lower peak memory consumption, and reduced inference time. This demonstrates that our model is more viable for practical real-world deployments, particularly on resource-constrained satellite-borne or UAV-borne onboard computing units.

Table 5: Complexity analysis on the CAVE dataset using the `balloons` test image (LR-HSI size $128 \times 128 \times 31$ + HR-MSI size $512 \times 512 \times 3 \rightarrow$ HR-HSI size $512 \times 512 \times 31$). Source-model metrics summarize the backbone scale, while test-time adaptation metrics report the additional overhead of one adaptation pass on that image. "Updated Params" denotes the parameters involved in gradient updates. Peak memory is measured on a GPU.

Method	**Source Model**		**Test-Time Adaptation**			
	Params (M)	FLOPs (G)	FLOPs (G)	Updated Params (M)↓	Peak Memory (MB)↓	Time (s)↓
SRTTA	1.256	657.4	5,832	1.219	4,746	2.0
IODA	0.244	125.6	4,928	0.244	4,775	2.5
TAO	2.944	1,819.3	13,494	3.710	5,139	3.0
SPA	0.826	479.2	2,399	0.004	1,637	1.9
DCTTA	35.53	1,401.4	54,927	11.01	14,437	15.6
UAL	7.098	6,825.6	345,946	0.044	9,940	250.4
THSI	0.631	10,232.3	178,926	0.631	13,525	8.6
Ours	1.256	657.4	**1,381**	**0**	**581**	**1.2**

Table 6: Quantitative comparison of our FOTA across different **backbones** on the CAVE dataset. NoCorr denotes clean images. For each method the first row shows PSNR (↑) and the gray-shaded row shows SAM (↓). Small colored values are the improvement over the non-adapted state.

Method	NoCorr	Environment			Sensor			Spatial			Average
		Haze	Brightness	Cloud	G-Noise	S-Noise	Band-Drop	G-Blur	M-Blur	Turbulence	
Standard Backbones:											
MambaIR	46.93	35.49 **+11.52**	37.74 **+10.48**	42.51 +10.67	37.05 +3.17	43.81 +5.28	42.48 **+0.69**	42.29 +5.07	40.76 +8.49	42.75 +7.16	40.54 +6.95
	2.63	15.27 **-4.79**	11.83 **-4.02**	4.83 +0.54	16.09 +0.44	4.78 -0.76	7.35 -0.16	5.13 +0.17	6.27 +0.25	4.77 +0.04	8.48 -0.92
SwinIR	48.64	35.16 +11.04	37.02 +9.44	42.76 **+10.77**	38.97 **+4.79**	43.92 **+6.03**	42.80 +0.40	42.50 **+7.62**	40.11 **+9.53**	42.56 **+8.52**	40.65 **+7.57**
	2.43	16.20 -3.43	12.78 -2.59	4.41 +0.42	13.64 -1.79	4.84 -0.17	7.26 -0.08	5.28 +0.26	6.92 +0.94	4.95 +0.52	8.48 -0.66
EDSR (Ours)	46.74	34.41 +3.18	39.59 +2.67	42.68 +2.62	38.59 +1.46	43.98 +0.92	42.16 +0.25	42.78 +2.05	41.32 +3.22	42.91 +2.48	40.93 +2.09
	2.82	17.31 -0.20	11.27 +1.09	3.80 -0.00	14.32 -0.89	4.61 +0.09	7.52 -0.14	4.28 -0.44	5.00 -0.64	3.92 -0.46	8.00 -0.18
Foundation Models:											
HyperSIGMA	37.35	30.56 +7.52	33.24 +6.20	36.65 +6.34	32.58 +2.25	37.12 +3.40	36.98 +0.42	35.85 +4.58	35.12 +5.12	35.55 +4.40	34.85 +4.47
	2.28	5.82 -4.42	4.36 -3.25	2.52 -1.10	4.65 -1.75	2.44 -1.02	2.65 -0.22	2.25 -2.45	2.15 -2.82	1.42 -1.96	3.14 -2.11
HyperFree	36.50	29.85 +6.95	32.36 +5.88	35.75 +5.92	31.78 +2.10	36.25 +3.15	36.02 +0.38	34.90 +4.25	34.15 +4.80	34.22 +4.10	33.92 +4.17
	2.45	6.15 -4.55	4.60 -3.40	2.75 **-1.15**	4.90 **-1.85**	2.62 **-1.08**	2.85 **-0.25**	2.42 **-2.55**	2.30 **-2.95**	1.56 **-2.11**	3.35 **-2.21**
SpectralEarth	38.10	31.28 +7.80	34.02 +6.45	37.45 +6.62	33.35 +2.35	37.95 +3.52	37.78 +0.45	36.65 +4.78	35.95 +5.35	36.69 +4.62	35.68 +4.66
	2.12	5.35 -4.45	3.90 -3.30	2.30 -1.08	4.25 -1.78	2.22 -0.98	2.45 -0.20	2.02 -2.50	1.95 -2.90	1.10 -1.98	2.82 -2.13

H Generalization Ability Across Different Source Model Backbones

We conduct generalization experiments for our proposed TTA method across different source model backbones, and the results are reported in Table 6. The standard backbones include the lightweight Mamba-based model MambaIR Guo et al. (2024), the Transformer-based SwinIR Liang et al. (2021), and our default CNN-based backbone EDSR Lim et al. (2017). It can be observed that under the initial no corruption setting, the three backbones exhibit comparable performance, achieving similar final average PSNR and SAM metrics. Furthermore, the higher TTA performance gains achieved by MambaIR and SwinIR suggest that models with more complex architectures are more susceptible to performance degradation induced by domain shifts. Regarding the foundation models for hyperspectral imagery—namely HyperSIGMA Wang et al. (2025), HyperFree Li et al. (2025b), and SpectralEarth Braham et al. (2025)—our proposed TTA method consistently enhances their performance on specific downstream tasks. Because these three foundation models are primarily tailored for remote sensing benchmarks and have not been specialized for image fusion tasks on ground-level datasets, their source model super-resolution performance under the no corruption scenario is relatively modest. Specifically, we adopt the HyperSIGMA-Base variant as the source model for HyperSIGMA, the ViT-Base variant for HyperFree, and the ViT-Base variant for SpectralEarth; fine-tuning details follow the original protocols described in their respective papers.

I Generalization Ability Across Different Fusion or Super-Resolution Scales

We conduct generalization experiments across multiple magnification scales on the CAVE dataset, as shown in Table 7. Note that all models are pre-trained exclusively at the $\times 4$ scale. As can be seen, the advantages of our TTA method vary across different magnification scales: at the extreme $\times 16$ scale, our approach achieves the highest PSNR gain. In contrast, under the in-distribution $\times 4$ setting, UTAL yields the largest PSNR improvement, while THSI attains the best PSNR gain at the $\times 8$ scale. Both of these competing methods are explicitly designed for hyperspectral tasks and heavily rely on parameter updates. Although our method achieves the third-best performance gain under these settings, it is entirely backpropagation-free, representing an effective trade-off between computational complexity and adaptation performance.

Table 7: Source model and zero-shot test-time adaptation on **CAVE** under magnification shift. Every method is pretrained at $\times 4$ and tested at $\times 4$, $\times 8$, and $\times 16$; the LR-HSI of the target scale is produced with that scale's blur kernel. Colored values represent adapted performance with the gain over its own non-adapted baseline.

Method	Metric	Test-time magnification (pretrained at $\times 4$)		
		$\times 4$	$\times 8$	$\times 16$
SRTTA	PSNR ↑	49.03 +0.05	42.68 +0.32	33.81 +0.12
	SSIM ↑	0.9964 +0.0000	0.9882 +0.0004	0.9553 +0.0011
	SAM ↓	2.23 -0.01	3.68 +0.01	6.26 +0.00
	ERGAS ↓	1.058 -0.008	2.111 -0.069	5.568 -0.077
IODA	PSNR ↑	48.50 +0.21	43.90 +0.97	38.55 +1.52
	SSIM ↑	0.9961 +0.0002	0.9905 +0.0023	0.9753 +0.0105
	SAM ↓	2.33 -0.07	3.47 -0.19	5.31 -0.54
	ERGAS ↓	1.123 -0.031	1.848 -0.229	3.368 -0.687
TAO	PSNR ↑	46.49 +0.08	44.17 +0.14	41.19 +0.27
	SSIM ↑	0.9947 +0.0001	0.9907 +0.0012	0.9827 +0.0042
	SAM ↓	2.86 -0.03	3.73 -0.08	5.07 -0.14
	ERGAS ↓	1.418 -0.014	1.807 -0.050	2.527 -0.133
SPA	PSNR ↑	45.91 -2.51	39.70 +0.19	31.28 +0.12
	SSIM ↑	0.9923 -0.0037	0.9824 +0.0005	0.9371 +0.0005
	SAM ↓	3.46 +0.96	4.03 +0.17	6.40 +0.25
	ERGAS ↓	1.564 +0.413	2.914 -0.059	7.458 -0.121
DCTTA	PSNR ↑	47.93 +0.10	37.75 +0.19	29.28 +0.10
	SSIM ↑	0.9956 +0.0001	0.9762 +0.0011	0.9176 +0.0029
	SAM ↓	2.48 -0.07	4.20 -0.04	6.63 -0.15
	ERGAS ↓	1.200 -0.015	3.590 -0.084	9.331 -0.115
UTAL	PSNR ↑	46.31 **+1.75**	44.68 +1.58	41.79 +1.56
	SSIM ↑	0.9945 +0.0020	0.9923 +0.0029	0.9870 +0.0060
	SAM ↓	2.90 -0.28	3.41 -0.34	4.45 -0.55
	ERGAS ↓	1.447 **-0.333**	1.693 -0.371	2.310 -0.572
THSI	PSNR ↑	43.85 +0.94	33.70 **+2.54**	28.14 +2.60
	SSIM ↑	0.9901 **+0.0021**	0.9515 **+0.0287**	0.8874 **+0.0627**
	SAM ↓	3.11 **-0.30**	4.69 **-0.60**	6.67 **-1.05**
	ERGAS ↓	1.942 -0.233	5.826 **-2.017**	10.731 **-3.857**
Ours	PSNR ↑	47.01 +0.27	43.91 +1.42	39.89 **+3.18**
	SSIM ↑	0.9939 +0.0006	0.9891 +0.0033	0.9774 +0.0117
	SAM ↓	2.75 -0.07	3.73 -0.28	5.53 -0.49
	ERGAS ↓	1.368 -0.068	1.893 -0.348	2.944 -1.395

J Generalization Experiments and Extended Quantitative and Qualitative Evaluations on the Harvard Dataset

Tables 8 and 11 report the performance of various TTA methods across nine corruption scenarios on the Harvard dataset. As can be observed, our proposed method achieves a substantial PSNR gain of 5.24 dB, remarkably outperforming the second-best method (THSI with 2.86 dB), while also securing the top rank in both SSIM and ERGAS improvements. Although our SAM improvement is slightly inferior to those of TAO, UTAL, and THSI, we attribute this to the drastic degradation of our source model when directly applied to the Brightness shift and Gaussian Noise corruptions, where unadapted SAM plummets to 17.73 and 18.08, respectively. Under such catastrophic degradation, the spectral fidelity is severely compromised, naturally bottlenecking the headroom for test-time adaptation. Regarding the CAVE benchmark, the SSIM and ERGAS improvements are detailed in Table 12. Our framework attains the most significant ERGAS reduction on CAVE, surpassing the second-place THSI by 0.13, while ranking closely behind THSI in SSIM by merely 0.003. Crucially, our proposed TTA method dramatically slashes test-time FLOPs without requiring iterative backpropagation to update the source model parameters, thereby spearheading the practical adoption of BP-free TTA paradigms in the hyperspectral imaging domain.

Table 8: Quantitative results of various methods on the **Harvard** dataset, where corruptions are exclusively applied to the LR-HSI. Abbreviations include G-Noise (Gaussian Noise), S-Noise (Stripe Noise), G-Blur (Gaussian Blur), and M-Blur (Motion Blur). NoCorr denotes performance on clean images. For each method, the first row shows **PSNR** (↑) and the gray-shaded row shows **SAM** (↓). Small colored values represent the improvement relative to the non-adapted state. Bold and underlined colored values indicate the **best** and second-best improvements, respectively.

		Environment			Sensor			Spatial			Avg. Δ
Method	NoCorr	Haze	Brightness	Cloud	G-Noise	S-Noise	Band-Drop	G-Blur	M-Blur	Turbulence	
SRTTA	50.73	38.06 +2.79	46.18 +3.91	34.00 +1.89	32.73 +1.49	34.06 +1.96	34.02 +1.89	34.95 +2.56	34.75 +2.29	47.87 +4.75	+2.61
	2.65	18.42 -1.23	4.70 -0.36	6.79 +0.07	16.45 +1.46	7.40 +0.19	6.81 +0.03	6.83 +0.40	6.74 +0.31	4.83 -0.29	+0.06
IODA	50.54	34.26 +1.43	36.30 +2.07	38.28 +2.01	35.93 +1.46	38.18 +2.00	38.23 +2.00	38.67 +2.05	38.72 +2.05	40.40 +2.04	+1.90
	2.68	21.95 -1.98	6.95 +0.13	5.89 **-0.73**	15.22 -1.13	6.45 **-0.77**	5.97 **-0.72**	5.65 **-0.70**	5.65 **-0.69**	5.11 **-0.73**	-0.81
TAO	45.64	35.62 +1.15	40.36 +2.02	42.66 +0.75	38.91 +0.49	42.29 +0.73	42.60 +0.75	43.31 +0.85	43.28 +0.81	43.63 +0.77	+0.92
	4.72	17.59 **-4.90**	5.12 -0.73	6.19 -0.19	13.84 -1.12	6.94 -0.23	6.15 -0.21	5.86 -0.22	5.91 -0.18	5.98 -0.16	**-0.88**
SPA	50.16	35.48 +0.33	42.85 +0.25	34.12 +0.25	32.54 +0.42	34.08 +0.33	34.04 +0.24	34.86 +0.36	34.72 +0.32	45.45 +0.15	+0.29
	2.77	19.35 -0.25	5.04 -0.14	6.69 -0.11	15.12 -0.28	7.14 -0.17	6.75 -0.13	6.32 -0.20	6.33 -0.18	4.98 -0.12	-0.18
DCTTA	50.55	27.28 +0.11	30.26 +0.34	32.49 +0.17	31.75 +0.18	33.49 +0.25	33.91 +0.22	34.63 +0.12	35.00 +0.12	34.10 +0.23	+0.19
	2.69	25.30 +0.06	16.24 +0.17	6.59 -0.17	17.76 +0.09	8.07 -0.10	6.34 -0.18	5.48 -0.13	5.46 -0.14	5.89 -0.17	-0.06
UTAL	50.00	35.95 +3.25	40.25 +3.52	42.89 +2.70	41.71 +2.95	43.44 +2.65	43.38 +1.70	45.97 +2.45	46.10 +2.40	45.39 +2.50	+2.68
	2.79	19.31 -2.35	12.86 **-1.75**	5.11 -0.45	10.34 **-1.30**	6.22 -0.55	5.03 -0.25	4.31 -0.34	4.32 -0.30	4.52 -0.36	-0.85
THSI	44.86	34.70 +3.45	37.60 +3.50	40.45 +2.85	36.55 +2.75	39.65 +2.45	41.29 +1.79	40.75 +2.95	40.10 +3.00	42.20 +3.00	+2.86
	3.35	21.30 -2.15	9.75 -1.45	6.90 -0.65	13.55 -1.25	8.05 -0.55	7.50 -0.30	6.68 -0.52	6.85 -0.55	5.60 -0.50	**-0.88**
Ours	50.28	32.36 **+3.66**	36.96 **+4.57**	40.05 **+5.31**	36.29 **+3.06**	40.95 **+5.32**	41.59 **+5.45**	43.96 **+6.69**	44.08 **+6.52**	43.09 **+6.56**	**+5.24**
	2.75	27.92 -1.75	17.73 -0.39	6.02 -0.52	18.08 -0.52	7.35 -0.57	5.62 -0.47	4.73 -0.54	4.72 -0.53	5.09 -0.54	-0.65

Visual comparisons on the cross-temporal remote sensing benchmark, Houston 2018, are illustrated in Fig. 8. As observed from the difference maps, our proposed method consistently enhances the performance of source models across different comparative TTA paradigms. Notably, the visual improvements yielded by our approach on the source backbones of IODA and THSI are substantially more pronounced than those achieved by their original adaptation mechanisms. In addition, Fig. 9 presents the adaptation results on the Pavia University benchmark; the zoomed-in RGB patches of THSI and our method clearly demonstrate that our FOTA preserves fine-grained texture details with superior fidelity. Finally, Fig. 10 showcases the test-time adaptation efficacy on the Loukia scene of the HyRANK dataset. Here, the Mean Relative Absolute Error (MRAE) map serves as a pixel-wise spatial error heatmap designed to delineate the spatial distribution of relative percentage errors between the reconstructed image and the ground-truth reference. As depicted in the MRAE error maps, our approach achieves a particularly salient improvement on the source model of TAO, markedly suppressing residual artifacts.

K Robustness Experiments on the SRF and HR-MSI Corruptions

Table 9 showcases the TTA performance across varying SRFs. Note that the SRF utilized by the pre-trained source model to simulate HR-MSIs originates from the Gaofen-1 satellite, as previously described. As observed, our FOTA achieves the overall second-best performance gains across diverse SRFs. In contrast, for MSI corruptions, our method exhibits sub-optimal adaptation, as the proposed fitness function is inherently premised on high-quality HR-MSIs. Nevertheless, it is worth emphasizing that in real-world scenarios, HR-MSIs are significantly less susceptible to environmental disturbances compared to LR-HSIs.

L More ablation studies on BW-initialization of Eq. (4)

The detailed ablation results for Eq. (4) are summarized in Table 10. Specifically, mean alignment aligns only the first-order statistics of the source and target features to match their distribution centers, whereas covariance alignment matches their feature covariances. Random initialization denotes

Table 9: Ablation and comparison of adaptation performance under varying Spectral Response Functions (SRF) and MSI distortions on the CAVE dataset. For each method, the first row indicates the adapted PSNR (↑) and the second gray-shaded row reports the adapted SAM (↓). Small colored superscripts denote the relative adaptation gains. Bold and underlined values highlight the **best** and second-best gains, respectively.

Method	Metric	Spectral Response Function (SRF) Shift			MSI Corruption		
		Canon-EOS10D	NikonD700	SonyDCZV1B	Haze	G-Blur	G-Noise
SRTTA	PSNR (↑)	40.986 +0.968	49.039 +0.058	42.044 +0.928	30.759 -0.382	27.803 -0.250	21.831 -0.365
	SAM (↓)	3.346 -0.430	2.231 -0.001	3.276 -0.411	14.992 +0.358	6.933 +0.190	45.316 +0.852
IODA	PSNR (↑)	40.522 +1.698	48.505 +0.208	41.449 +1.507	23.880 -0.424	27.689 -0.176	21.270 **+1.022**
	SAM (↓)	3.239 -0.645	2.331 -0.067	3.205 -0.580	21.382 +0.462	6.756 +0.127	37.048 -9.312
TAO	PSNR (↑)	36.958 +1.561	46.499 +0.084	37.515 +1.460	23.260 -0.256	27.366 -0.043	22.220 -0.651
	SAM (↓)	5.093 -0.563	2.846 -0.036	4.997 -0.630	22.154 +0.650	7.221 +0.091	33.893 -1.383
SPA	PSNR (↑)	39.933 -1.867	45.840 -2.589	40.642 -1.690	32.814 -0.133	29.014 -0.205	21.644 -0.204
	SAM (↓)	4.455 +1.052	3.320 +0.825	4.441 +1.031	14.327 -0.222	6.619 +0.430	47.726 +0.630
UTAL	PSNR (↑)	44.675 **+9.039**	46.285 **+1.731**	44.796 **+8.361**	23.109 -0.501	27.341 -0.123	23.409 -1.113
	SAM (↓)	3.145 **-1.492**	2.890 -0.290	3.129 **-1.464**	21.792 +1.037	7.144 +0.206	32.574 +2.265
THSI	PSNR (↑)	42.876 +0.614	43.854 +0.942	42.961 +0.631	33.136 -6.185	31.849 -1.182	26.423 +0.446
	SAM (↓)	3.259 -0.241	3.112 **-0.303**	3.251 -0.242	14.608 +9.976	5.189 +0.382	23.782 -7.888
DCTTA	PSNR (↑)	44.529 +0.209	47.935 +0.102	44.811 +0.204	37.925 **-0.002**	29.699 **-0.027**	22.735 +0.736
	SAM (↓)	3.052 -0.079	2.485 -0.065	3.047 -0.086	6.866 **-0.295**	6.138 **-0.030**	35.957 **-12.342**
Ours	PSNR (↑)	39.073 +2.352	47.009 +0.266	39.752 +2.158	25.117 -0.142	27.862 -0.198	22.415 -0.274
	SAM (↓)	5.240 +0.387	2.750 -0.068	4.774 -0.015	21.070 +0.231	6.811 +0.380	35.014 +0.524

Table 10: Comparison of different initialization strategies on the CAVE dataset.

Initialization Strategy	MSI	ZO	OT	CAVE			
				PSNR	SAM	SSIM	ERGAS
Mean alignment	✓	✓		+1.789	-0.093	+0.0135	-1.03
Covariance alignment	✓	✓		+1.650	-0.085	+0.0128	-0.976
Random initialization	✓	✓		+1.561	-0.139	+0.0143	-0.670
Ours	✓	✓	✓	**+2.093**	-0.176	**+0.0228**	**-1.149**

randomly initializing the transformation parameters and relying entirely on the TTA process to learn the alignment. As evidenced by the results in the table, our proposed Bures-Wasserstein (BW) initialization consistently yields superior performance, validating its effectiveness in establishing a geometrically principled starting point for adaptation.

Table 11: Quantitative results of various methods on the **Harvard** dataset, where corruptions are exclusively applied to the LR-HSI. Abbreviations include G-Noise (Gaussian Noise), S-Noise (Stripe Noise), G-Blur (Gaussian Blur), and M-Blur (Motion Blur). NoCorr denotes performance on clean images. For each method, the first row shows **SSIM** (↑) and the gray-shaded row shows **ERGAS** (↓). Small colored values represent the improvement relative to the non-adapted state. Bold and underlined colored values indicate the **best** and second-best improvements, respectively.

		Environment			Sensor			Spatial			Avg. Δ
Method	NoCorr	Haze	Brightness	Cloud	G-Noise	S-Noise	Band-Drop	G-Blur	M-Blur	Turbulence	
SRTTA	0.9938	0.9708 +0.009	0.9782 +0.007	0.9541 +0.011	0.8681 +0.027	0.9491 +0.011	0.9549 +0.011	0.9578 +0.011	0.9577 +0.010	0.9806 +0.002	+0.014
	1.447	4.568 -1.064	3.817 -0.751	8.293 -1.337	20.449 -1.844	8.518 -1.340	6.718 -1.470	6.207 -1.679	6.282 -1.539	3.399 -0.317	-1.535
IODA	0.9931	0.9568 +0.008	0.9784 +0.007	0.9704 +0.014	0.8917 +0.019	0.9654 +0.014	0.9704 **+0.014**	0.9734 +0.013	0.9735 **+0.013**	0.9804 +0.002	+0.014
	1.642	6.531 -1.405	3.383 -0.753	6.707 -1.361	19.410 -1.289	7.266 -1.239	5.188 -1.433	4.869 -1.382	4.860 -1.363	3.374 -0.315	-1.344
TAO	0.9872	0.9631 +0.001	0.9872 +0.001	0.9804 +0.003	0.9106 +0.009	0.9755 +0.004	0.9811 +0.003	0.9832 +0.003	0.9828 +0.003	0.9856 +0.001	+0.004
	2.820	3.751 -0.124	2.012 -0.054	4.225 -0.577	11.441 -2.053	4.485 -0.585	3.256 -0.589	3.068 -0.585	3.100 -0.545	2.124 -0.041	-0.822
SPA	0.9933	0.9612 +0.003	0.9741 +0.002	0.9512 +0.003	0.8602 +0.004	0.9482 +0.003	0.9523 +0.002	0.9551 +0.003	0.9542 +0.003	0.9798 +0.001	+0.003
	1.585	5.542 -0.280	4.065 -0.145	8.590 -0.215	21.080 -0.420	8.955 -0.255	7.330 -0.175	6.762 -0.245	6.872 -0.235	3.485 -0.120	-0.243
DCTTA	0.9936	0.7440 +0.004	0.8019 +0.009	0.9458 +0.003	0.8027 +0.005	0.9295 +0.004	0.9572 +0.003	0.9634 +0.002	0.9640 +0.001	0.9579 +0.003	+0.004
	1.477	88.922 -0.493	46.652 -0.788	28.645 -0.215	28.954 -0.370	17.066 -0.224	6.828 -0.316	6.237 -0.183	6.042 -0.154	6.694 -0.260	-0.334
UTAL	0.9931	0.9061 **+0.052**	0.9423 +0.038	0.9863 +0.009	0.9513 +0.027	0.9753 +0.009	0.9865 +0.003	0.9913 +0.005	0.9913 +0.005	0.9892 +0.005	+0.017
	1.642	29.953 -7.250	16.488 -4.350	9.535 -1.950	10.895 -2.550	6.924 -1.450	3.059 -0.420	2.595 -0.380	2.561 -0.360	2.795 -0.410	-2.124
THSI	0.9852	0.8670 +0.045	0.8990 +0.038	0.9570 +0.015	0.9170 +0.032	0.9620 +0.014	0.9740 +0.006	0.9660 **+0.014**	0.9610 **+0.013**	0.9800 +0.012	+0.021
	2.184	8.600 -3.850	6.900 -2.950	6.300 -1.950	7.450 -2.150	4.750 -1.450	3.400 -0.750	3.050 -1.150	3.200 -1.150	2.820 -0.980	-1.820
Ours	0.9932	0.8322 +0.031	0.8942 **+0.049**	0.9747 **+0.016**	0.8572 **+0.033**	0.9635 **+0.020**	0.9811 **+0.014**	0.9855 +0.012	0.9853 +0.012	0.9822 **+0.014**	**+0.022**
	1.524	67.052 **-17.699**	34.253 **-9.306**	21.184 **-5.934**	22.682 **-4.949**	12.940 **-2.946**	3.408 **-2.210**	2.781 **-2.146**	2.777 **-2.039**	3.060 **-2.332**	**-5.507**

Table 12: Quantitative results of various methods on the **CAVE** dataset, where corruptions are exclusively applied to the LR-HSI. Abbreviations include G-Noise (Gaussian Noise), S-Noise (Stripe Noise), G-Blur (Gaussian Blur), and M-Blur (Motion Blur). NoCorr denotes performance on clean images. For each method, the first row shows **SSIM** (↑) and the gray-shaded row shows **ERGAS** (↓). Small colored values represent the improvement relative to the non-adapted state. Bold and underlined colored values indicate the **best** and second-best improvements, respectively.

		Environment			Sensor			Spatial			Avg. Δ
Method	NoCorr	Haze	Brightness	Cloud	G-Noise	S-Noise	Band-Drop	G-Blur	M-Blur	Turbulence	
SRTTA	0.9964	0.9177 +0.068	0.9693 +0.039	0.9877 **+0.012**	0.9222 +0.042	0.9876 **+0.009**	0.9900 -0.000	0.9829 +0.003	0.9640 +0.003	0.9782 +0.002	+0.020
	1.066	8.079 -2.759	3.056 -1.789	2.596 -0.843	3.354 **-0.982**	1.881 -0.423	2.339 -0.036	2.465 -0.526	3.770 -0.955	2.864 -0.414	-0.970
IODA	0.9959	0.8956 +0.060	0.9346 +0.052	0.9877 +0.004	0.9240 +0.025	0.9868 +0.005	0.9898 -0.000	0.9863 +0.004	0.9781 +0.010	0.9871 **+0.006**	+0.018
	1.154	6.794 -1.740	3.730 -1.298	2.447 -0.340	3.357 -0.493	1.882 -0.219	2.347 -0.042	2.259 -0.369	2.904 -0.645	2.236 -0.432	-0.620
TAO	0.9946	0.9629 +0.009	0.9849 +0.003	0.9922 +0.001	0.9698 +0.004	0.9905 +0.000	0.9885 +0.000	0.9883 +0.002	0.9845 +0.005	0.9899 +0.002	+0.003
	1.432	4.396 -0.161	2.526 -0.019	1.975 -0.056	2.364 -0.076	1.755 -0.030	2.296 -0.020	1.994 -0.090	2.246 -0.170	1.874 -0.119	-0.082
SPA	0.9960	0.6634 +0.003	0.7411 +0.005	0.9639 +0.000	0.8257 +0.006	0.9568 +0.000	0.9901 -0.000	0.9690 +0.000	0.9441 +0.000	0.9712 +0.000	+0.002
	1.151	18.046 -0.239	11.033 -0.274	7.009 -0.003	5.666 -0.187	3.334 -0.010	2.410 -0.003	4.133 -0.023	6.543 -0.024	4.481 -0.006	-0.085
DCTTA	0.9955	0.6221 +0.001	0.6927 +0.001	0.9576 +0.000	0.7862 +0.001	0.9419 +0.000	0.9897 +0.000	0.9575 +0.001	0.9264 +0.001	0.9644 +0.003	+0.001
	1.215	22.001 -0.083	14.117 -0.066	8.461 -0.018	7.029 -0.052	4.011 -0.019	2.473 -0.013	5.205 -0.057	8.165 -0.061	5.351 -0.069	-0.049
UTAL	0.9925	0.9351 -0.025	0.9752 -0.000	0.9911 +0.002	0.9809 +0.012	0.9905 +0.003	0.9888 +0.001	0.9905 +0.004	0.9873 +0.005	0.9905 +0.003	+0.000
	1.780	6.048 +1.405	2.966 +0.292	2.057 -0.214	2.155 -0.321	1.827 -0.225	2.439 -0.022	1.898 -0.375	2.145 -0.472	1.945 -0.291	-0.025
THSI	0.9880	0.6482 +0.059	0.7273 **+0.058**	0.9572 +0.010	0.8082 **+0.043**	0.9311 +0.007	0.9842 **+0.002**	0.9228 **+0.045**	0.8998 +0.003	0.9494 **+0.006**	**+0.026**
	2.175	17.940 -2.188	11.198 **-1.926**	7.593 **-1.582**	6.287 -0.613	4.563 **-1.351**	2.976 **-0.212**	7.711 **-1.642**	9.656 **-2.384**	6.807 **-1.324**	-1.012
Ours	0.9933	0.9140 **+0.077**	0.9681 +0.055	0.9886 +0.007	0.9266 +0.039	0.9874 +0.004	0.9858 -0.000	0.9863 +0.006	0.9810 **+0.012**	0.9865 **+0.006**	+0.023
	1.436	5.952 **-3.733**	3.305 -1.658	2.328 -1.148	3.483 -0.959	1.932 -0.275	2.493 -0.085	2.151 -0.592	2.528 -1.160	2.115 -0.725	**-1.148**

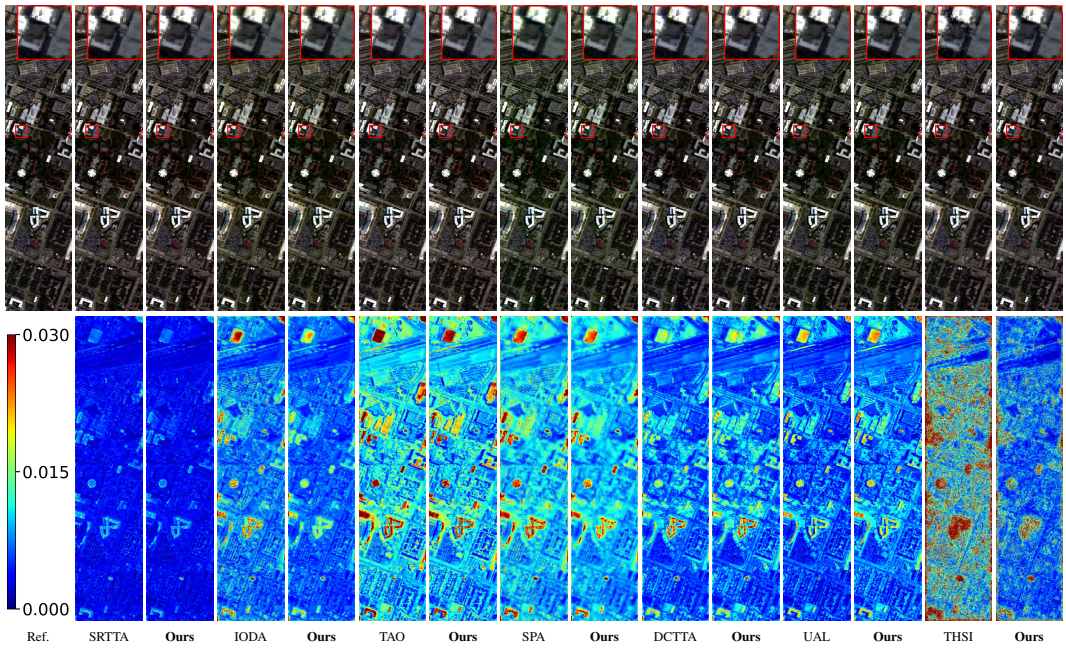


Figure 8: Visual comparison on the Houston 2018 dataset. The top row shows the pseudo-color images, and the bottom row illustrates the corresponding difference maps relative to the reference.

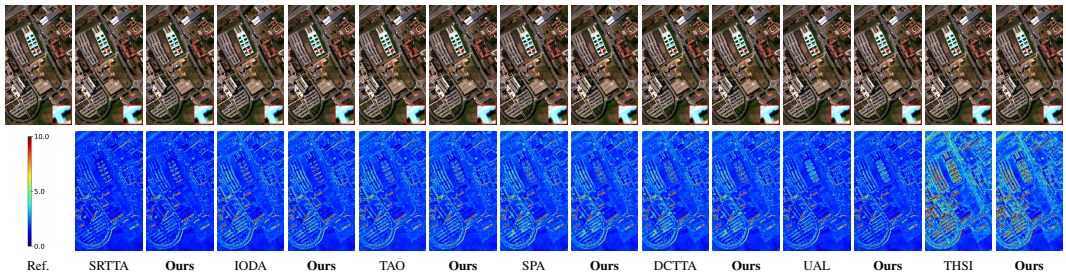


Figure 9: Visual comparison on the Pavia University dataset. The top row shows the pseudo-color images, and the bottom row illustrates the corresponding SAM error maps relative to the reference.

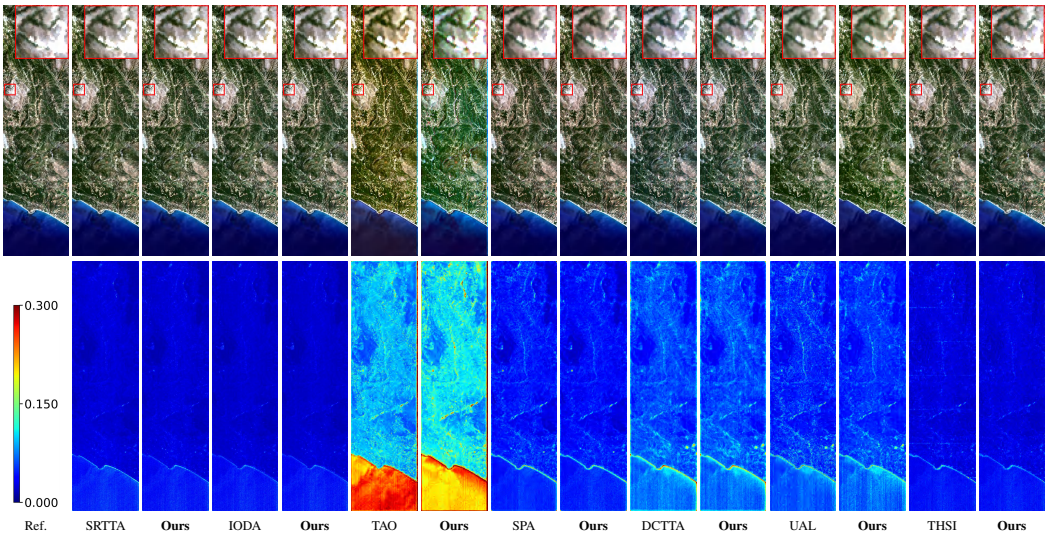


Figure 10: Visual comparison on the HyRANK Loukia dataset. The top row shows the pseudo-color reconstructed images, and the bottom row illustrates the corresponding MRAE error maps relative to the reference.