

Empowering Your Pansharpening Models with Generalizability: Unified Distribution is All You Need

Yongchuan Cui^{1,2} Peng Liu^{1,2,*} Hui Zhang^{3,*}

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, 100094 Beijing, China

² School of Electronic, Electrical and Communication Engineering,
University of Chinese Academy of Sciences, 101408 Beijing, China

³ School of Engineering Medicine, Beihang University, 100191 Beijing, China

yongchuancui@gmail.com, liupeng202303@aircas.ac.cn, hui.zhang@buaa.edu.cn

Abstract

Existing deep learning-based models for remote sensing pansharpening exhibit exceptional performance on training datasets. However, due to sensor-specific characteristics and varying imaging conditions, these models suffer from substantial performance degradation when applied to unseen satellite data, lacking generalizability and thus limiting their applicability. We argue that the performance drops stem primarily from distributional discrepancies from different sources and the key to addressing this challenge lies in bridging the gap between training and testing distributions. To validate the idea and further achieve a “train once, deploy forever” capability, this paper introduces a novel and intuitive approach to empower any pansharpening models with generalizability by employing a unified distribution strategy (UniPAN). Specifically, we construct a distribution transformation function that normalizes the pixels sampled from different sources to conform to an identical distribution. The deep models are trained on the transformed domain, and during testing on new datasets, the new data are also transformed to match the training distribution. UniPAN aims to train and test the model on a unified and consistent distribution, thereby enhancing its generalizability. Extensive experiments validate the efficacy of UniPAN, demonstrating its potential to significantly enhance the performance of deep pansharpening models across diverse satellite sensors. Codes: <https://github.com/yc-cui/UniPAN>.

1. Introduction

Due to the inherent limitations of sensor imaging technology, advanced remote sensing satellites such as IKONOS, QuickBird, GaoFen-2, and WorldView-3, necessitate a compromise between the acquisition of extensive spectral infor-

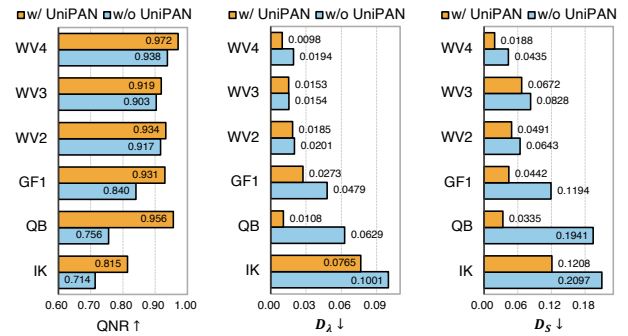


Figure 1. Comparison of generalization metrics w/ and w/o UniPAN trained on WorldView-3 (WV3) using FusionNet [19].

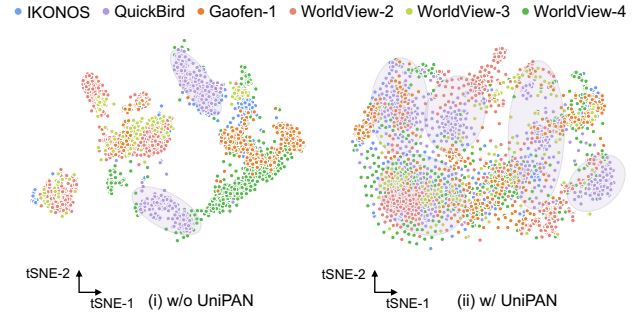


Figure 2. Comparison of tSNE visualization w/ and w/o UniPAN.

mation and maintaining fine spatial details in their captured imagery [20, 49, 68]. Consequently, these satellites typically capture either high spatial resolution single-band panchromatic (PAN) images or low spatial resolution multispectral (LRMS) images. Pansharpening is a crucial technique in remote sensing image processing, aiming to combine PAN and LRMS images to generate fused images with high spatial resolution and abundant spectra (HRMS).

Traditional pansharpening methods have been extensively studied over the past few decades. However, these

*Corresponding author.

methods may introduce spectral distortions, require significant computational resources, and are sensitive to the quality of input data [20, 42, 49, 68]. In contrast, deep learning (DL)-based methods have shown great potential in pansharpening. Notably, various innovative architectures have emerged, starting from early convolutional models [19, 47, 82, 83], progressing to Transformer-based approaches [8, 21, 22, 24, 50, 62, 65], then generative methods [43, 46, 48, 58, 78], and continuing with the emergence of the latest novel designs [9, 23, 31, 53, 86]. The networks are evolving towards larger parameters capacity, stronger representation ability, and increasingly complex and diverse structures.

However, as deep pansharpening thrives, the overly powerful fitting capabilities of these networks have concurrently presented several ongoing concerns, *e.g.* generalizability, interpretability, over-reliance on high-quality data, *etc.* This paper specifically delves into the widely criticized issue of generalizability. Specifically, although deep models achieve excellent performance on training datasets, they often fail to generalize well to new data due to distribution shift such as different noise levels, varying spectral signatures and sensor conditioning characteristics. While models can achieve satisfactory performances on new datasets through retraining or fine-tuning, the process is time-consuming and fine-tuning on every new dataset is impractical and cumbersome. The ideal scenario would involve developing a pansharpening model that can be trained once and deployed effectively across diverse datasets and sensors, thereby providing a *train once, deploy forever* capability.

To this end, this paper introduces a novel approach that aims to empower any pansharpening models with generalizability by employing a unified distribution strategy (dubbed as **UniPAN**). The proposed method aims to bridge the gap between training and testing distributions, thereby improving the model’s ability to generalize to unseen data. The core idea is to normalize the distribution of spatial pixels from the training dataset to conform to a predefined distribution (*e.g.* Gaussian distribution) and deep models are then trained on the transformed domain. This process is model-agnostic and requires no additional trainable parameters. When generalizing to new data, testing datasets are also transformed to match the training distribution, ensuring that the models are evaluated under the same distributional assumptions. Fig. 1 presents the cross-satellite generalization performance of a well-known pansharpening model FusionNet [19] on NBUPansharpRSDData [49] trained w/ and w/o UniPAN. The model utilizing UniPAN demonstrates a significant improvement on full-resolution metrics, which validates the strong generalization ability for real-world scenes. To further elucidate the efficacy of UniPAN, Fig. 2 presents tSNE [64] dimensionality reduction clustering of multispectral images from diverse satellites using NBUPansharpRSDData [49] w/

and w/o UniPAN. The pre-UniPAN visualization reveals distinct clustering patterns, with data points from different satellites forming separate groups, indicative of substantial inter-satellite distributional disparities. In contrast, following the application of UniPAN, a notable transformation is observed: the previously separated clusters become substantially intermingled, demonstrating a significant reduction in distributional discrepancies across different satellite datasets (*e.g.* the purple scatters of QuickBird). This visual evidence underscores UniPAN’s capability to harmonize heterogeneous satellite data, effectively creating a more cohesive and unified representation.

The contributions of this paper could be summarized as follows:

- We propose UniPAN, a model-agnostic approach that employs a unified distribution strategy to empower any deep pansharpening models with generalizability, achieving a *train once, deploy forever* capability. This strategy provides two advantages: 1) *Architectural independence*: Compatible with any existing or future network architecture without structural modifications. 2) *Training efficiency*: Eliminates the need for retraining or fine-tuning when adapting to new sensors.
- UniPAN is implemented via inverse transform sampling, and we prove its equivalence to the optimal solution of optimal transport in the one-dimensional case. UniPAN requires no additional trainable parameters and only needs to be fitted once before training begins, introducing almost no additional computational overhead. It serves as a viable alternative to vanilla data normalization techniques such as min-max and *z*-score normalization, offering a more robust and unified approach to data distribution alignment.
- We conduct extensive experiments using a variety of deep pansharpening models across numerous satellites. The simple yet intuitive and effective strategy provides a universal solution to the generalization challenge in DL-based pansharpening and experimental results demonstrate its effectiveness, showcasing significant improvements in cross-dataset generalization across a wide range of models and satellite sensors.

2. Related Work

2.1. Pansharpening Methods

Traditional Methods. Traditional pansharpening methods can be broadly categorized into three paradigms: component substitution (CS), multi-resolution analysis (MRA), and variational optimization (VO) [20, 42, 49, 68]. CS-based methods [4, 10, 27, 38, 55, 61] operate by substituting the spatial component of the LRMS image with the PAN image. Representative algorithms include Gram-Schmidt (GS) [4, 38], principal component analysis (PCA) [10, 61], and intensity-hue-saturation (IHS) [27, 55]. While computationally effi-

cient, CS methods often introduce spectral distortions due to mismatches between PAN and substituted components. MRA-based methods [1–3, 13, 26, 41] employ multi-scale decomposition tools (*e.g.* wavelets [26], contourlets [13]) to inject high-frequency details from PAN into LRMS. Typical implementations include MTF-GLP [1–3] and SFIM [41]. These methods generally preserve spectral fidelity better than CS approaches but may exhibit limited spatial enhancement. VO-based approaches [7, 52, 63, 67, 74, 75] formulate pansharpening as an optimization problem with handcrafted priors, such as the variational wavelet-based methods [7] and Bayesian approaches [52]. Despite their mathematical elegance, traditional methods often require manual parameter tuning and struggle with complex spatial-spectral relationships [20, 68].

Deep Learning-Based Methods. The advent of deep learning (DL) have revolutionized pansharpening by enabling end-to-end learning of spatial-spectral relationships. Early pioneering convolutional neural networks (CNNs) such as PNN [47], PanNet [82], and MSDCNN [83] established the feasibility of learning fusion rules from a mount of simulated data. Subsequent innovations introduced multi-scale architectures [33, 72, 76], attention mechanisms [39, 73], and kernel customizing [12, 18, 25, 34, 85], *etc.* The CNN-based methods demonstrated superior performance over traditional approaches but faced limitations in modeling long-range dependencies. In contrast, Transformer [24, 45, 65] addressed this limitation with self-attention mechanisms. Although various Transformer architectures specifically designed for pansharpening [8, 21, 22, 50, 62] have achieved impressive performance, they face criticism for quadratic computational complexity w.r.t. sequence length [16]. To address this limitation, recent studies have explored linear-complexity models [16, 31, 53, 86] that maintain performance while improving efficiency. To further enhance fidelity preservation, generative models like generative adversarial networks (GANs) [17, 43, 44, 46, 79] and diffusion models [48, 58, 77, 78, 88] have emerged as promising solutions. Beyond conventional configurations, contemporary research has expanded into more challenging frontiers to address practical constraints, including zero-shot learning paradigms [9, 23, 57, 71], frequency domain integration [30, 40, 84, 89, 90], and interpretable fusion mechanisms [32, 70, 80, 81].

2.2. Generalization in DL-Based Pansharpening

Despite remarkable progress in DL-based pansharpening, generalization capability remains a persistent challenge. The issue manifests conspicuously in two critical scenarios: 1) Models trained under the Wald protocol [69] using spatially degraded data often fail to maintain performance when applied to full-resolution inference, frequently exhibiting severe spectral distortions. 2) Networks optimized for specific

satellite sensors demonstrate substantial performance degradation when deployed on unseen satellite platforms due to distribution shifts in spectral signatures and sensor characteristics. Existing approaches to address these challenges primarily focus on unsupervised learning [14, 15, 46], generalized module design [77], or advanced data augmentation [11]. However, such methods typically require either meticulous architectural redesign or computationally intensive retraining/fine-tuning procedures. These requirements pose significant practical limitations for real-world deployment.

In contrast to model-centric solutions, our approach adopts a fundamentally different perspective by addressing the generalization challenge through data-space intervention. The proposed UniPAN framework introduces a model-agnostic preprocessing technique that operates on raw sensor data. Compared to traditional linear data normalization such as min-max scaling or z -score standardization, our method employs a non-linear mapping to project the input data onto a unified distribution, effectively harmonizing the distribution mismatches from all sensors.

3. Method

3.1. Distribution Transformation

The core challenge in cross-domain pansharpening stems from the distribution discrepancy between source (training) and target (testing) domains. Let $\mathcal{X}_S$ denote the source domain (training data) and $\mathcal{X}_T$ the target domain (testing data). Our objective is to learn a transformation $\mathcal{T} : \mathcal{X} \rightarrow \mathcal{Z}$ that maps both domains to a shared space $\mathcal{Z}$ with unified distribution ρ_t , *e.g.* standard normal distribution.

Inverse Transform Sampling. Given a univariate r.v. $X \sim \rho_s$, we seek a transformation $\mathcal{T} : \mathbb{R} \rightarrow \mathbb{R}$ such that $Z = \mathcal{T}(X) \sim \rho_t$. The inverse transform sampling theorem provides a theoretically grounded solution:

$$\mathcal{T}(x) = F_T^{-1}(F_S(x)), \quad (1)$$

where $F_S(x) = \rho_s(-\infty, x]$ is the cumulative distribution function (CDF) of X , and $F_T^{-1}(u) = \inf\{z \in \mathbb{R} : F_T(z) \geq u\}$ is the inverse CDF of the target distribution. By the universality of the uniform, $U = F_S(X)$ follows a uniform distribution, and $Z = F_T^{-1}(U)$ inherits ρ_t exactly when F_S is continuous. To provide more intuitively analyze of the algorithm’s operational process and efficiency, we establish a fundamental connection between inverse transform sampling and the well-studied optimal transport theory [51, 54].

Optimal Transport Perspective. Given two probability distributions ρ_s (source domain) and ρ_t (target domain) defined on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, we consider the Monge problem [51, 54]:

$$\mathcal{C}(\rho_s, \rho_t) \triangleq \min_{\substack{T : \mathbb{R}^n \rightarrow \mathbb{R}^m \\ T_{\#}\rho_s = \rho_t}} \int_{\mathbb{R}^n} c(x, T(x)) \rho_s(x) dx, \quad (2)$$

where $c : \Omega_s \times \Omega_t \rightarrow \mathbb{R}^+$ is a cost function and for any measurable Borel set $E \in \Omega_s$, $T : \Omega_s \rightarrow \Omega_t$ transfers measures from one space Ω_s to another space Ω_t ,

$$T_{\#}\rho_s(E) = \rho_s(T^{-1}(E)), \quad (3)$$

here $T_{\#}$ is the pushforward operator. The Monge problem seeks a transport plan T from ρ_s to ρ_t that minimizes the transportation cost. The optimal T^* is also known as the Monge map of Eq. (2).

Let $F_S(x) = \rho_s(-\infty, x]$ and $F_T(z) = \rho_t(-\infty, z]$ denote the CDFs of the source and target distributions respectively. In the univariate setting ($n = m = 1, \Omega_s = \Omega_t = \mathbb{R}$), when the source measure ρ_s is continuous and the cost function takes the form $c(x, z) = h(|x - z|)$ with h being strictly convex, by cyclical monotonicity [66], the Monge map T^* satisfies:

$$F_S(x) = F_T(T^*(x)), \quad (4)$$

yielding the explicit expression:

$$T^*(x) = F_T^{-1}(F_S(x)). \quad (5)$$

This remarkable result coincides exactly with the inverse transform sampling formula Eq. (1), *i.e.* $T^* = \mathcal{T}$. The transport cost in this setting [37, 56, 59] can be computed as

$$\mathcal{C}(\rho_s, \rho_t) = \int_0^1 c(F_S^{-1}(u), F_T^{-1}(u)) du. \quad (6)$$

For the L_1 cost function $c(x, z) = |x - z|$, this corresponds to the W_1 Wasserstein distance [36, 59] between the distributions. In the one-dimensional (1D) scenario, UniPAN operates independently on spatial pixels within each spectral band. This implementation can be achieved through interpolation-based inverse transform sampling without framing it as an optimization problem. However, given the inherent correlations between spectral bands and interdependencies among spatial patches, multidimensional extension of this framework may be promising, albeit with increased implementation complexity. The theoretical derivation establishing connections with optimal transport theory serves to facilitate future generalization of UniPAN to multidimensional configurations. Building upon the theoretical foundation of distribution transformation through optimal transport, in the following, we present a practical implementation of 1D UniPAN.

3.2. UniPAN Implementation

We present a practical implementation for 1D finite discrete distributions of UniPAN through a quantile-based distribution transformation algorithm.

Finite Discrete Distribution Transformation. Given pixel values $X = \{x_1, x_2, \dots, x_n\}$ sampled from a spectral band,

Algorithm 1 UniPAN Training and Testing Procedure

Training Phase

Input: Training dataset $\mathcal{D}_{\text{train}} = \{L_i, P_i, H_i\}_{i=1}^N$, inverse CDF F_T^{-1} of target distribution ρ_t , model $\mathcal{M}$

Output: Trained model M

```

1: Initialize transformed LRMS  $\tilde{L}$  and PAN  $\tilde{P}$ 
2: for each band  $b$  of LRMS  $\{L_i\}_{i=1}^N$  & PAN  $\{P_i\}_{i=1}^N$  do
3:   Sample  $n$  pixels  $X_b \leftarrow \{x_1, \dots, x_n\}$  from band  $b$ 
4:   Compute  $m$  quantiles  $Q = \{q_0, \dots, q_{m-1}\}$   $\triangleright$  Eq. (7)
5:   Linspace uniform CDF  $U = \{0, \frac{1}{m-1}, \dots, 1\}$ 
6:   for each pixel  $x$  in band  $b$  do
7:     Forward and backward interp by  $u_{\uparrow}$  and  $u_{\downarrow}$   $\triangleright$  Eq. (9)
8:     Compute eCDF value  $u(x) \leftarrow (u_{\uparrow} + u_{\downarrow}) / 2$   $\triangleright$  Eq. (8)
9:     Transform  $x$  to  $z \leftarrow F_T^{-1}(u(x))$  and assign  $z$  to corresponding band of  $\tilde{L}$  and  $\tilde{P}$   $\triangleright$  Eq. (11)
10:  end for
11: end for
12: while not converged do
13:   Train  $\mathcal{M}$  with  $\mathcal{L} = \|H - \mathcal{M}(\tilde{L}, \tilde{P}) + L\|_1$ 
14: end while
15: return Trained model  $\mathcal{M}$ 

```

Testing Phase

Input: Testing dataset $\mathcal{D}_{\text{test}} = \{L_i, P_i\}_{i=1}^N$, inverse CDF F_T^{-1} of target distribution ρ_t , trained model $\mathcal{M}$

Output: Pansharpening result $\hat{H}^{\text{test}}$

```

16: for each band  $b$  of LRMS & PAN do
17:   Transform  $L, P$  to  $\tilde{L}, \tilde{P}$  using the training transformation
18: end for
19: Generate  $\hat{H}^{\text{test}} \leftarrow \mathcal{M}(\tilde{L}, \tilde{P}) + L$ 
20: return Fused result  $\hat{H}^{\text{test}}$ 

```

we first compute m empirical quantiles $Q = \{q_0, \dots, q_{m-1}\}$ where:

$$q_i = \inf \left\{ x \in X \mid P(X \leq x) \geq \frac{i}{m-1} \right\}, i = \{0, 1, \dots, m-1\}. \quad (7)$$

These quantiles discretize the empirical cumulative distribution function (eCDF) of the source data. The number of quantiles m controls the approximation fidelity.

Given any data point x to be transformed, the next step involves determining its value in the eCDF through interpolation. However, when the source distribution contains repeated values or sparse regions, quantile estimation may yield non-unique or overlapping quantile boundaries. This can result in abrupt jumps in the eCDF, thereby violating the properties of monotonicity and smoothness. Therefore, we use bidirectional interpolation to estimate a robust eCDF value of x :

$$\begin{aligned}
u(x) &= u_{\uparrow}(x)/2 + u_{\downarrow}(x)/2 \\
&= \text{interp}(x, Q, U) / 2 + \\
&\quad (1 - \text{interp}(-x, -Q^{\text{rev}}, -U^{\text{rev}})) / 2,
\end{aligned} \quad (8)$$

where $U = \{0, \frac{1}{m-1}, \dots, 1\}$, Q^{rev} is the reversed quantiles.

QNR↑	$\rho_t = \mathcal{N}(0, 1)$						$\rho_t = \mathcal{U}(0, 1)$					
	GF1	IK	QB	WV2	WV3	WV4	GF1	IK	QB	WV2	WV3	WV4
GF1	0.980(+0.005)	0.946(−0.010)	0.971(+0.000)	0.963(−0.003)	0.964(−0.004)	0.974(−0.000)	0.985(+0.010)	0.961(+0.004)	0.979(+0.007)	0.984(+0.017)	0.982(+0.013)	0.987(+0.013)
IK	0.950(+0.005)	0.912(+0.039)	0.915(+0.045)	0.924(+0.011)	0.916(+0.022)	0.958(+0.011)	0.956(+0.011)	0.919(+0.045)	0.940(+0.070)	0.950(+0.037)	0.945(+0.051)	0.968(+0.021)
QB	0.957(+0.006)	0.936(+0.016)	0.931(+0.047)	0.953(+0.017)	0.944(+0.014)	0.970(+0.016)	0.965(+0.014)	0.944(+0.024)	0.966(+0.082)	0.969(+0.032)	0.969(+0.038)	0.970(+0.017)
WV2	0.880(+0.026)	0.811(+0.029)	0.832(+0.062)	0.946(+0.008)	0.916(+0.012)	0.951(+0.012)	0.872(+0.018)	0.775(−0.006)	0.825(+0.055)	0.944(+0.006)	0.914(+0.011)	0.940(+0.001)
WV3	0.862(−0.006)	0.804(+0.035)	0.795(+0.067)	0.921(−0.011)	0.919(+0.011)	0.950(+0.001)	0.880(+0.011)	0.805(+0.036)	0.813(+0.084)	0.943(+0.010)	0.924(+0.016)	0.952(+0.003)
WV4	0.924(+0.009)	0.898(+0.045)	0.875(+0.036)	0.948(−0.000)	0.929(+0.003)	0.972(+0.006)	0.939(+0.024)	0.889(+0.035)	0.903(+0.064)	0.952(+0.002)	0.935(+0.009)	0.973(+0.007)
$D_\lambda \downarrow$	$\rho_t = \mathcal{N}(0, 1)$						$\rho_t = \mathcal{U}(0, 1)$					
	GF1	IK	QB	WV2	WV3	WV4	GF1	IK	QB	WV2	WV3	WV4
GF1	0.007(−0.003)	0.022(+0.008)	0.005(−0.000)	0.017(+0.004)	0.015(+0.005)	0.011(+0.000)	0.004(−0.005)	0.012(−0.001)	0.004(−0.001)	0.004(−0.008)	0.004(−0.005)	0.004(−0.005)
IK	0.011(−0.000)	0.020(−0.015)	0.018(−0.011)	0.031(+0.004)	0.022(−0.003)	0.014(−0.003)	0.008(−0.003)	0.020(−0.015)	0.013(−0.016)	0.013(−0.014)	0.009(−0.016)	0.010(−0.008)
QB	0.013(−0.000)	0.011(−0.004)	0.015(−0.010)	0.016(−0.004)	0.014(+0.000)	0.010(−0.005)	0.009(−0.003)	0.009(−0.006)	0.007(−0.017)	0.008(−0.013)	0.006(−0.007)	0.012(−0.003)
WV2	0.042(−0.006)	0.066(−0.007)	0.037(−0.019)	0.008(−0.000)	0.014(−0.000)	0.016(−0.002)	0.046(−0.002)	0.086(+0.012)	0.048(−0.008)	0.008(+0.000)	0.015(+0.001)	0.022(+0.002)
WV3	0.050(+0.010)	0.067(−0.011)	0.049(−0.024)	0.027(+0.011)	0.014(−0.000)	0.016(+0.002)	0.042(+0.002)	0.072(−0.006)	0.051(−0.022)	0.013(−0.002)	0.011(−0.002)	0.014(−0.000)
WV4	0.024(−0.000)	0.031(−0.012)	0.029(−0.009)	0.017(+0.008)	0.018(+0.008)	0.005(−0.001)	0.016(−0.008)	0.032(−0.011)	0.022(−0.016)	0.014(+0.004)	0.015(+0.005)	0.004(−0.002)
$D_S \downarrow$	$\rho_t = \mathcal{N}(0, 1)$						$\rho_t = \mathcal{U}(0, 1)$					
	GF1	IK	QB	WV2	WV3	WV4	GF1	IK	QB	WV2	WV3	WV4
GF1	0.012(−0.003)	0.032(+0.003)	0.023(+0.000)	0.020(+0.000)	0.021(+0.000)	0.015(−0.000)	0.010(−0.005)	0.026(−0.002)	0.016(−0.006)	0.010(−0.009)	0.013(−0.007)	0.007(−0.008)
IK	0.039(−0.005)	0.068(−0.026)	0.068(−0.036)	0.046(−0.014)	0.063(−0.019)	0.027(−0.007)	0.035(−0.008)	0.062(−0.033)	0.046(−0.058)	0.037(−0.024)	0.045(−0.036)	0.022(−0.013)
QB	0.030(−0.005)	0.053(−0.012)	0.054(−0.039)	0.031(−0.012)	0.042(−0.014)	0.020(−0.012)	0.025(−0.011)	0.046(−0.018)	0.026(−0.067)	0.023(−0.020)	0.024(−0.031)	0.020(−0.012)
WV2	0.081(−0.022)	0.134(−0.024)	0.137(−0.049)	0.045(−0.008)	0.071(−0.012)	0.032(−0.010)	0.085(−0.018)	0.153(−0.005)	0.133(−0.053)	0.047(−0.006)	0.071(−0.012)	0.039(−0.003)
WV3	0.096(−0.000)	0.141(−0.027)	0.167(−0.050)	0.054(+0.002)	0.067(−0.011)	0.033(−0.004)	0.083(−0.013)	0.134(−0.033)	0.146(−0.072)	0.043(−0.008)	0.064(−0.014)	0.033(−0.003)
WV4	0.053(−0.009)	0.073(−0.036)	0.099(−0.029)	0.035(−0.006)	0.053(−0.011)	0.021(−0.005)	0.045(−0.018)	0.082(−0.027)	0.076(−0.052)	0.036(−0.004)	0.052(−0.012)	0.021(−0.005)
$D_\rho \downarrow$	$\rho_t = \mathcal{N}(0, 1)$						$\rho_t = \mathcal{U}(0, 1)$					
	GF1	IK	QB	WV2	WV3	WV4	GF1	IK	QB	WV2	WV3	WV4
GF1	0.043(+0.003)	0.110(+0.030)	0.026(−0.008)	0.079(+0.032)	0.062(+0.015)	0.041(−0.002)	0.019(−0.020)	0.059(−0.021)	0.022(−0.013)	0.015(−0.032)	0.017(−0.029)	0.017(−0.026)
IK	0.082(−0.006)	0.095(−0.038)	0.090(−0.058)	0.126(+0.024)	0.110(−0.010)	0.061(−0.008)	0.064(−0.024)	0.092(−0.040)	0.066(−0.083)	0.051(−0.049)	0.061(−0.059)	0.045(−0.024)
QB	0.061(−0.003)	0.054(−0.020)	0.075(−0.038)	0.052(−0.026)	0.060(−0.019)	0.044(−0.017)	0.036(−0.028)	0.048(−0.026)	0.029(−0.083)	0.021(−0.057)	0.023(−0.056)	0.028(−0.034)
WV2	0.166(−0.027)	0.198(−0.039)	0.199(−0.059)	0.115(+0.001)	0.130(−0.045)	0.093(−0.026)	0.164(−0.029)	0.235(−0.002)	0.200(−0.058)	0.100(−0.013)	0.126(−0.048)	0.112(−0.007)
WV3	0.205(+0.018)	0.214(−0.041)	0.234(−0.053)	0.231(+0.109)	0.173(+0.011)	0.105(−0.015)	0.176(−0.010)	0.226(−0.029)	0.219(−0.068)	0.117(−0.004)	0.131(−0.030)	0.101(−0.018)
WV4	0.105(−0.001)	0.112(−0.042)	0.115(−0.044)	0.146(+0.080)	0.169(+0.077)	0.038(−0.014)	0.072(−0.034)	0.118(−0.035)	0.105(−0.055)	0.053(−0.012)	0.062(−0.029)	0.035(−0.017)

Table 1. Average results of 12 models. Rows correspond to training satellites, and columns correspond to testing satellites. **Green** indicates performance improvement, while **red** indicates performance degradation. The values in parentheses represent the average delta compared to the baseline without UniPAN.

The function `interp` represents a linear interpolation between the quantile point Q and the reference point U , with the interpolation formula given by:

$$\text{interp}(x, Q, U) = \begin{cases} U[0], & x \leq Q[0] \\ f(x, i, Q, U), & Q[i] < x \leq Q[i+1] \\ U[m-1], & x > Q[m-1] \end{cases} \quad (9)$$

where f computes the Q to U mapping,

$$f(x, i, Q, U) = U[i] + \frac{x - Q[i]}{Q[i+1] - Q[i]}(U[i+1] - U[i]) \quad (10)$$

The last step is to apply inverse CDF of target distribution,

$$\mathcal{T}(x) = F_T^{-1}(u(x)) \quad (11)$$

Training and Testing Procedure. The training and testing procedures of UniPAN are described in Algorithm 1. Here, $L \uparrow$ denotes the upsampled LRMS image using bicubic interpolation to match the size of the PAN image. The model $\mathcal{M}$ is trained to predict the residual of LRMS with an l_1 loss function that minimizes the discrepancy between the predicted HRMS image and the ground truth.

As for the target distribution ρ_t , we test two widely used distributions, *i.e.* standard normal distribution $\mathcal{N}(0, 1)$ and standard uniform distribution $\mathcal{U}(0, 1)$, the transformation $\mathcal{T}(x)$ becomes:

$$\mathcal{T}_n(x) = \Phi^{-1}(F_S(x)), \quad \mathcal{T}_u(x) = F_S(x), \quad (12)$$

where Φ^{-1} denotes the inverse CDF of standard normal distribution. Since the inverse CDF of standard uniform distribution is an identical mapping, $\mathcal{T}_u$ becomes the estimated empirical distribution itself. The standard normal distribution transformation non-linearly warps the input space to match Gaussian statistics, while the uniform transformation directly maps pixel intensities to their empirical quantiles.

4. Experiment

4.1. Setups

Datasets. We conducted experiments utilizing the Red, Green, Blue, and NIR bands of the NBUPansharperSData dataset [49]. The variations in resolution among different satellites, coupled with significant differences in land covers and land uses, render this dataset particularly suitable for assessing the generalization capabilities of pansharpening models. The dataset comprises data from 6 satellites. For each satellite, we partitioned the data into training, validation, and testing sets in a ratio of 7 : 1 : 2.

Implementation Details. To validate the effectiveness and generalization capability of UniPAN, we conducted comparative experiments between baseline (w/o UniPAN) and enhanced (w/ UniPAN) configurations using diverse commonly-recognized models, including FusionNet [19], PanNet [82], MSDDN [28, 29], GPPNN [80], MSD-CNN [83], FeINFN [40], PNN [47], LAGConv [34], SFIIN [89, 90], PreMix [18], UAPN [87], MDCUN [81].

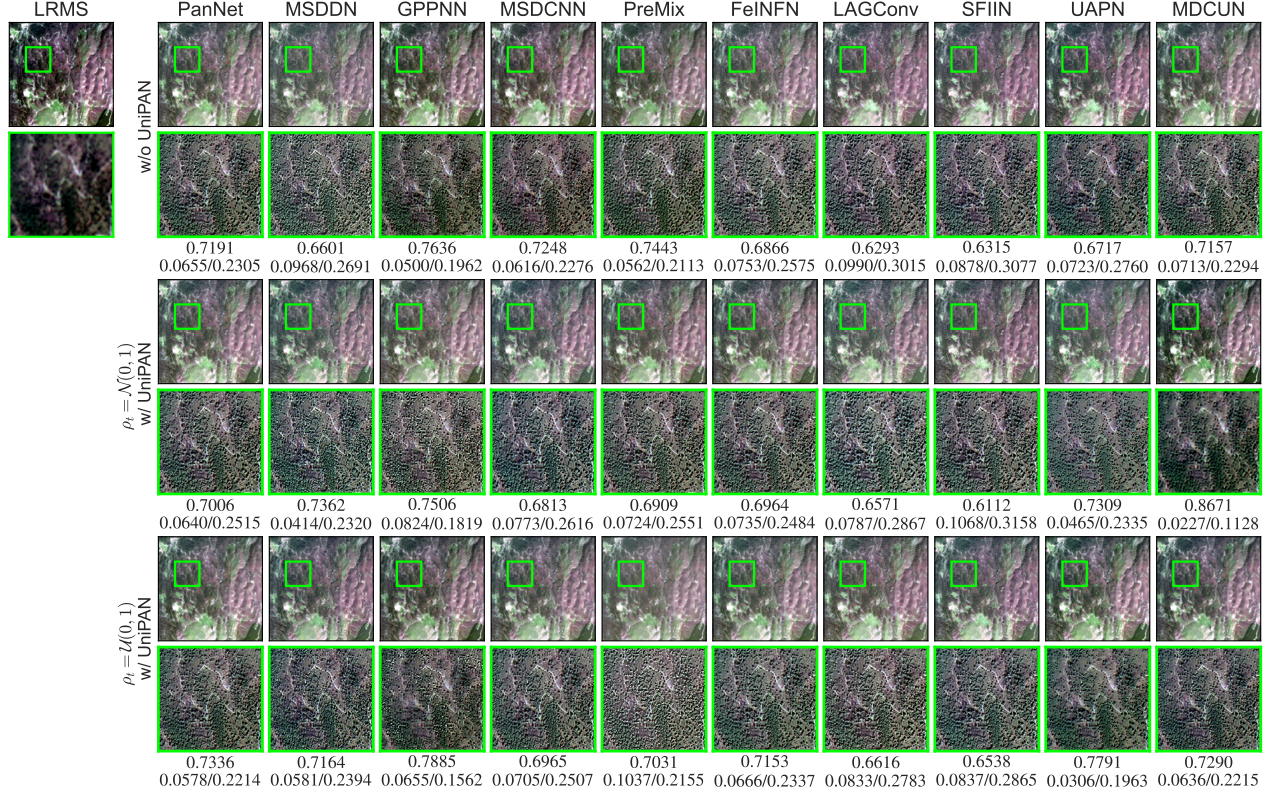


Figure 3. Generalization testing on GaoFen-1, the networks are trained on WorldView-3. The **bright green** box indicates the zoomed-in area. Below the zoomed-in image, the first line is the QNR value, and the second line is the D_λ/D_S value. In most cases, the use of UniPAN yields superior spectral and spatial preservation. For instance, MDCUN [81] exhibits a highly close alignment with the LRMS in spectral characteristics with UniPAN of $\rho_t = \mathcal{N}(0, 1)$.

All implementations followed identical experimental procedures: a batch size of 8 (except for MDCUN [81] and FeINFN [40] which used 4), the Adam [35] optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) with an initial learning rate of 10^{-3} , and StepLR scheduling which reduces the learning rate by a factor of 0.8 every 100 epochs. Training was performed on an NVIDIA GeForce RTX 4090 GPU for 600 epochs using the Wald protocol [69] with l_1 loss, with best model selected based on optimal validation performance. Note that, to prevent data leakage of testing set, the transformation fitting process is conducted on the training set of each satellite.

Evaluation Metrics. The full-reference metrics based on spatial degradation (*e.g.* PSNR, SSIM) can only reflect performance in simulated reduced-resolution experiments and fail to demonstrate effectiveness in real-world full-resolution scenarios [6, 16, 46]. In contrast, no-reference metrics like QNR explicitly quantify model performance in both spectral preservation and spatial enhancement, enabling direct evaluation of fusion results at native resolution while avoiding theoretical biases introduced by downsampling. To better align with practical applications where reference images are unavailable, this paper adopts widely-used no-reference spectral and spatial metrics, *i.e.* QNR [5], D_λ [5], D_S [5],

and D_ρ [60], to comprehensively assess cross-resolution and cross-satellite generalization capabilities.

4.2. Quantitative Evaluation

Tab. 1 presents the average generalization performance of various models enhanced with UniPAN under different target distributions. **Green** indicate metric improvements, while **red** denote performance degradation. Slight yet observable differences exist between the two target distributions. Overall, the uniform distribution achieves better performance than the normal distribution. The normal distribution demonstrated a modest decline in performance across the WV2 and WV3 datasets. This phenomenon may stem from the peaked characteristics of the normal distribution, which could hinder model training, whereas the uniform distribution provides smoother normalization (see Fig. 5 in Sec. 4.3).

Notably, UniPAN with $\rho_t = \mathcal{U}(0, 1)$ achieves significant improvements across most metrics, with only marginal declines in rare cases. For instance, when generalizing from WV3 to QB, UniPAN improves the QNR metric by over 0.084 (from 0.729 to 0.813) averaged on all 12 models, underscoring its effectiveness. Beyond QNR, UniPAN also enhances spectral preservation (D_λ) and spatial consistency



Figure 4. Full-resolution generalization testing on QuickBird. The **bright green** box indicates the zoomed-in area. Below the zoomed-in image, the first line is the QNR value, and the second line is the D_λ/D_S value.

(D_S, D_p) , further validating the efficacy of the unified distribution strategy. These results confirm that aligning heterogeneous satellite data to a unified distribution effectively mitigates domain shifts, enabling robust cross-sensor generalization. However, while UniPAN improves spatial fidelity greatly, it may cause slight spectral distortion, as evidenced by the minor performance degradation observed in the D_λ metric. This is reasonable since the transformation applied to each spectral band maintains the relative spatial structure but might result in minimal loss of spectral information. Nevertheless, such information loss is extremely limited, as the overall enhancement in performance remains predominant.

4.3. Visual Comparison

Comparison of Fusion Results. To illustrate the practical impact of UniPAN, Figs. 3 and 4 compare fusion results w/ and w/o UniPAN. In most cases, UniPAN achieves highly favorable results in both spectral and spatial preservation. Furthermore, the remarkably high QNR values demonstrate its superior capability in balancing spectral and spatial fidelity, *e.g.* in Fig. 3, MDCUN [81] applying UniPAN with normal distribution enhanced the QNR by 0.1514 (from 0.7157 to 0.8671) and aligned fused results closely with the ground truth spectra. In the full-resolution generalization test depicted in Fig. 4, even with minimal visual differences,

UniPAN achieves metrically significant improvements in spectral and spatial preservation. This capability is critical for downstream applications requiring accurate spectral information. Additional visual comparison results are included in the supplementary materials.

Comparison of Data Distribution. Fig. 5 further analyzes data distribution alignment. The kernel density estimation (KDE) curves (top right area and diagonals) reveal that UniPAN successfully transforms the original complex and heterogeneous distributions into a unified target distribution. The quantile-quantile (Q-Q) plots (bottom left area) demonstrate harmonized distributions across satellites after applying UniPAN (The greater the deviation from the $y = x$ line, the more significant the disparity between the two distributions). For example, the original Q-Q plots between WV4 and QB show significant dispersion, whereas post-UniPAN points close tightly around $y = x$. This visual evidence corroborates UniPAN’s ability to bridge distribution gaps.

4.4. Parameter Analysis

UniPAN introduces two hyperparameters: the number of sampled pixels n and the number of quantiles m . Figs. 6 and 7 investigate their impact on performance of FusionNet [19] trained on WV2 and tested on GF1. As n and m vary, the metrics stabilize. Generally, $U(0, 1)$ demonstrates

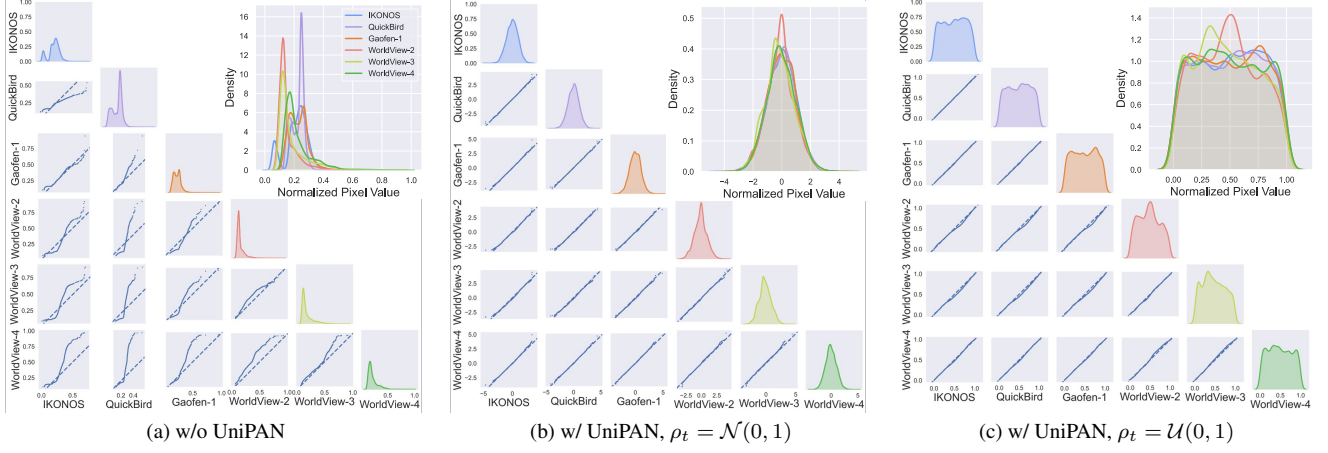


Figure 5. Comparison of data distributions before and after applying UniPAN. The upper-right and diagonals of each subplot display the kernel density estimation (KDE) curves of Green band for each satellite. The scatter plots in the lower-left areas of each subplots represent quantile-quantile (Q-Q) plots of green band, with the dashed line indicating $y = x$. The greater the deviation from the $y = x$ line, the more significant the disparity between the two distributions. Zoom in for better view.

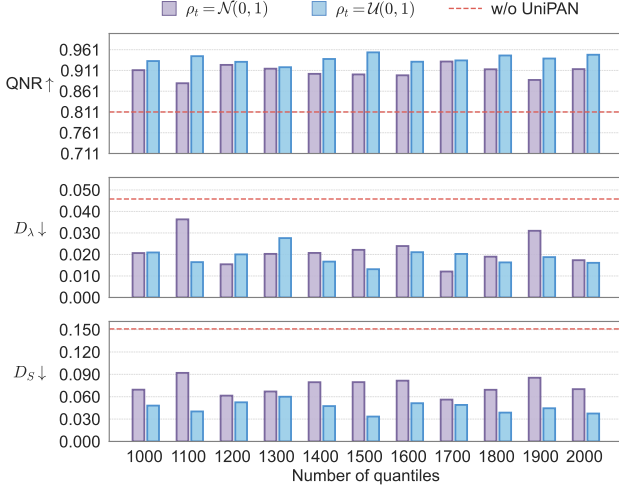


Figure 6. Parameter analysis of number of quantiles, with number of sampled pixels fixed to 10000.

superior performance compared to $\mathcal{N}(0, 1)$. Both configurations achieve statistically significant improvements over baseline methods without applying UniPAN. This parametric robustness demonstrates that optimal performance can be attained without exhaustive hyperparameter tuning, significantly reducing computational overhead.

5. Conclusion and Discussion

We proposed UniPAN, a unified distribution strategy addressing generalization challenges in deep learning-based pansharpening. UniPAN eliminates spectral discrepancies by normalizing multi-source remote sensing imagery into predefined target distributions, effectively bridging the gap between training and testing distributions. Extensive quantitative and visual comparisons across 6 satellite sensors and

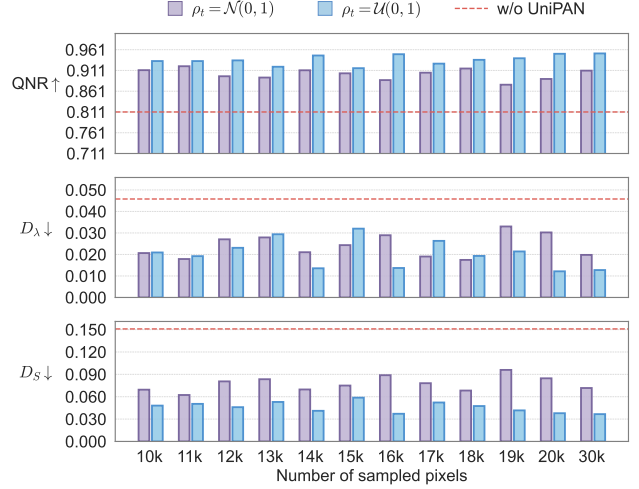


Figure 7. Parameter analysis of number of sampled pixels, with number of quantiles fixed to 1000.

12 models demonstrate that UniPAN significantly enhances cross-sensor generalization. Notably, this enhancement is achieved without introducing any additional trainable parameters or the need for retraining or fine-tuning the models.

While this work aligns theoretically with 1D optimal transport, considering interdependencies of spectral bands or spatial patches, future research could explore extending the method to the multi-dimensional case with optimal transport theories. Additionally, the experimental results indicate that the choice of target distribution has a notable impact on performance, thus it is worthwhile investigating adaptive target domain selection. To further validate the generalizability of the proposed unified distribution strategy, the ideas presented in this work could be extended to other remote sensing image processing tasks, such as multi-modal fusion and hyperspectral pansharpening, *etc.*

Acknowledgment. This research was partially supported by the National Natural Science Foundation of China under Grant 32441114, 41971397, 61731022, and the project Y1H103101A.

References

- [1] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE TGRS*, 40(10):2300–2312, 2002. 3
- [2] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva. An MTF-based spectral distortion minimizing model for pansharpening of very high resolution multispectral images of urban areas. In *2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pages 90–94, 2003.
- [3] Bruno Aiazzi, L. Alparone, Stefano Baronti, Andrea Garzelli, and Massimo Selva. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogramm. Eng. Remote Sens.*, 72:591–596, 2006. 3
- [4] Bruno Aiazzi, Stefano Baronti, and Massimo Selva. Improving component substitution pansharpening through multivariate regression of MS+Pan data. *IEEE TGRS*, 45(10):3230–3239, 2007. 2
- [5] Luciano Alparone, Bruno Aiazzi, Stefano Baronti, Andrea Garzelli, Filippo Nencini, and Massimo Selva. Multispectral and panchromatic data fusion assessment without reference. *Photogramm. Eng. Remote Sens.*, 74(2):193–200, 2008. 6
- [6] Alberto Arienzo, Gemine Vivone, Andrea Garzelli, Luciano Alparone, and Jocelyn Chanussot. Full-resolution quality assessment of pansharpening: Theoretical and hands-on approaches. *IEEE GRSM*, 10(3):168–201, 2022. 6
- [7] Coloma Ballester, Vicent Caselles, and Laura Igual. A variational model for P+XS image fusion. *IJCV*, 69(1):43–58, 2006. 3
- [8] Wele Gedara Chaminda Bandara and Vishal M Patel. Hypertransformer: A textural and spectral feature fusion Transformer for pansharpening. In *CVPR*, pages 1767–1777, 2022. 2, 3
- [9] Qi Cao, Liang-Jian Deng, Wu Wang, Junming Hou, and Gemine Vivone. Zero-shot semi-supervised learning for pansharpening. *Inf. Fusion.*, 101:102001, 2024. 2, 3
- [10] Pat S. Chavez and Andy Yaw Kwarteng. Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens.*, 55:339–348, 1989. 2
- [11] Lihui Chen, Gemine Vivone, Zihao Nie, Jocelyn Chanussot, and Xiaomin Yang. Spatial data augmentation: Improving the generalization of neural networks for pansharpening. *IEEE TGRS*, 61:1–11, 2023. 3
- [12] Zhi-Xuan Chen, Cheng Jin, Tian-Jing Zhang, Xiao Wu, and Liang-Jian Deng. SpanConv: A new convolution via spanning kernel space for lightweight pansharpening. In *IJCAI*, pages 841–847, 2022. Main Track. 3
- [13] Minhyeok Choi, Rae Young Kim, and Myeong-Ryong Nam. Fusion of multispectral and panchromatic satellite images using the curvelet transform. *IEEE GRSL*, 2(2):136–140, 2005. 3
- [14] Matteo Ciotola, Sergio Vitale, Antonio Mazza, Giovanni Poggi, and Giuseppe Scarpa. Pansharpening by convolutional neural networks in the full resolution framework. *IEEE TGRS*, pages 1–1, 2022. 3
- [15] Matteo Ciotola, Giovanni Poggi, and Giuseppe Scarpa. Un-supervised deep learning-based pansharpening with jointly enhanced spectral and spatial fidelity. *IEEE TGRS*, 61:1–17, 2023. 3
- [16] Yongchuan Cui, Peng Liu, Yan Ma, Lajiao Chen, Mengzhen Xu, and Xingyan Guo. Pixel-wise ensembled masked autoencoder for multispectral pansharpening. *IEEE TGRS*, 62:1–22, 2024. 3, 6
- [17] Yongchuan Cui, Peng Liu, Bingze Song, Lingjun Zhao, Yan Ma, and Lajiao Chen. Reconstruction of large-scale missing data in remote sensing images using Extend-GAN. *IEEE GRSL*, 21:1–5, 2024. 3
- [18] Yongchuan Cui, Peng Liu, Yan Ma, Lajiao Chen, Mengzhen Xu, and Xingyan Guo. Pansharpening via predictive filtering with element-wise feature mixing. *ISPRS J. Photogramm. Remote Sens.*, 219:22–37, 2025. 3, 5
- [19] Liang-Jian Deng, Gemine Vivone, Cheng Jin, and Jocelyn Chanussot. Detail injection-based deep convolutional neural networks for pansharpening. *IEEE TGRS*, 59(8):6995–7010, 2021. 1, 2, 5, 7
- [20] Liang-jian Deng, Gemine Vivone, Mercedes E. Paoletti, Giuseppe Scarpa, Jiang He, Yongjun Zhang, Jocelyn Chanussot, and Antonio Plaza. Machine learning in pansharpening: A benchmark, from shallow to deep networks. *IEEE GRSM*, 10(3):279–315, 2022. 1, 2, 3
- [21] Shangqi Deng, Liang-Jian Deng, Xiao Wu, Ran Ran, and Rui Wen. Bidirectional dilation Transformer for multispectral and hyperspectral image fusion. In *IJCAI*, 2023. 2, 3
- [22] Shang-Qi Deng, Liang-Jian Deng, Xiao Wu, Ran Ran, Danfeng Hong, and Gemine Vivone. PSRT: Pyramid shuffle-and-reshuffle Transformer for multispectral and hyperspectral image fusion. *IEEE TGRS*, 61:1–15, 2023. 2, 3
- [23] Renwei Dian, Anjing Guo, and Shutao Li. Zero-shot hyperspectral sharpening. *IEEE TPAMI*, 45(10):12650–12666, 2023. 2, 3
- [24] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021. 2, 3
- [25] Yule Duan, Xiao Wu, Haoyu Deng, and Liang-Jian Deng. Content-adaptive non-local convolution for remote sensing pansharpening. In *CVPR*, pages 27738–27747, 2024. 3
- [26] Andrea Garzelli and Filippo Nencini. Wavelet-based fusion of optical and SAR image data in remote sensing applications. *Int. J. Remote Sens.*, 23(23):5025–5039, 2002. 3
- [27] Morteza Ghahremani and Hassan Ghassemian. Nonlinear IHS: A promising method for pan-sharpening. *IEEE GRSL*, 13(11):1606–1610, 2016. 2

- [28] Xuanhua He, Keyu Yan, Rui Li, Chengjun Xie, Jie Zhang, and Man Zhou. Pyramid dual domain injection network for pan-sharpening. In *ICCV*, pages 12862–12871, 2023. 5
- [29] Xuanhua He, Keyu Yan, Jie Zhang, Rui Li, Chengjun Xie, Man Zhou, and Danfeng Hong. Multiscale dual-domain guidance network for pan-sharpening. *IEEE TGRS*, 61:1–13, 2023. 5
- [30] Xuanhua He, Keyu Yan, Rui Li, Chengjun Xie, Jie Zhang, and Man Zhou. Frequency-adaptive pan-sharpening with mixture of experts. *AAAI*, 38(3):2121–2129, 2024. 3
- [31] Xuanhua He, Ke Cao, Jie Zhang, Keyu Yan, Yingying Wang, Rui Li, Chengjun Xie, Danfeng Hong, and Man Zhou. Pan-Mamba: Effective pan-sharpening with state space model. *Inf. Fusion.*, 115:102779, 2025. 2, 3
- [32] Junming Hou, Xinyang Liu, Chenxu Wu, Xiaofeng Cong, Chentong Huang, Liang-Jian Deng, and Jian Wei You. Bidomain uncertainty gated recursive network for pan-sharpening. *Inf. Fusion.*, 118:102938, 2025. 3
- [33] Cheng Jin, Liang-Jian Deng, Ting-Zhu Huang, and Gemine Vivone. Laplacian pyramid networks: A new approach for multispectral pansharpening. *Inf. Fusion.*, 78:158–170, 2022. 3
- [34] Zi-Rong Jin, Tian-Jing Zhang, Tai-Xiang Jiang, Gemine Vivone, and Liang-Jian Deng. LAGConv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening. *AAAI*, 36(1):1113–1121, 2022. 3, 5
- [35] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [36] Soheil Kolouri, Se Rim Park, Matthew Thorpe, Dejan Slepcev, and Gustavo K. Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE SPM*, 34(4):43–59, 2017. 4
- [37] Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo K. Rohde. Generalized sliced wasserstein distances, 2019. 4
- [38] Craig A Laben and Bernard V Brower. Process for enhancing the spatial resolution of multispectral imagery using pansharpening. 2000. US Patent 6,011,875. 2
- [39] Zhixuan Li, Jinjiang Li, Fan Zhang, and Linwei Fan. CADUI: Cross-attention-based depth unfolding iteration network for pansharpening remote sensing images. *IEEE TGRS*, 61:1–20, 2023. 3
- [40] YuJie Liang, Zihan Cao, Shangqi Deng, Hong-Xia Dou, and Liang-Jian Deng. Fourier-enhanced implicit neural fusion network for multispectral and hyperspectral image fusion. In *NeurIPS*, pages 63441–63465. Curran Associates, Inc., 2024. 3, 5, 6
- [41] J. G. Liu. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *Int. J. Remote Sens.*, 21(18):3461–3472, 2000. 3
- [42] Peng Liu, Jun Li, Lizhe Wang, and Guojin He. Remote sensing data fusion with generative adversarial networks: State-of-the-art methods and future research directions. *IEEE GRSM*, 10(2):295–328, 2022. 2
- [43] Qingjie Liu, Huanyu Zhou, Qizhi Xu, Xiangyu Liu, and Yunhong Wang. PSGAN: A generative adversarial network for remote sensing image pan-sharpening. *IEEE TGRS*, 59(12):10227–10242, 2020. 2, 3
- [44] Xiaobo Liu, Xiang Liu, Haoran Dai, Xudong Kang, Antonio Plaza, and Wenjie Zu. Mun-GAN: a multiscale unsupervised network for remote sensing image pansharpening. *IEEE TGRS*, 61:1–18, 2023. 3
- [45] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical vision Transformer using shifted windows. In *ICCV*, 2021. 3
- [46] Jiayi Ma, Wei Yu, Chen Chen, Pengwei Liang, Xiaojie Guo, and Junjun Jiang. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Inf. Fusion.*, 62:110–120, 2020. 2, 3, 6
- [47] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa. Pansharpening by convolutional neural networks. In *Remote Sens.*, page 594, 2016. 2, 3, 5
- [48] Qingyan Meng, Wenxu Shi, Sijia Li, and Linlin Zhang. Pan-Diff: A novel pansharpening method based on denoising diffusion probabilistic model. *IEEE TGRS*, 61:1–17, 2023. 2, 3
- [49] Xiangchao Meng, Yiming Xiong, Feng Shao, Huanfeng Shen, Weiwei Sun, Gang Yang, Qiangqiang Yuan, Randi Fu, and Hongyan Zhang. A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation. *IEEE GRSM*, 9(1):18–52, 2021. 1, 2, 5
- [50] Xiangchao Meng, Nan Wang, Feng Shao, and Shutao Li. Vision Transformer for pansharpening. *IEEE TGRS*, 60:1–11, 2022. 2, 3
- [51] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781. 3
- [52] Fjola Palsson, Johannes R. Sveinsson, and Magnus O. Ulfarson. A new pansharpening algorithm based on total variation. *IEEE GRSL*, 11(1):318–322, 2014. 3
- [53] Siran Peng, Xiangyu Zhu, Haoyu Deng, Liang-Jian Deng, and Zhen Lei. FusionMamba: Efficient remote sensing image fusion with state space model. *IEEE TGRS*, 2024. 2, 3
- [54] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Found. Trends Mach. Learn.*, 11(5-6):355–607, 2019. 3
- [55] Sheida Rahmani, Melissa Strait, Daria Merkurjev, Michael Moeller, and Todd Wittman. An adaptive IHS pan-sharpening method. *IEEE GRSL*, 7(4):746–750, 2010. 2
- [56] Aaditya Ramdas, Nicolas Garcia, and Marco Cuturi. On wasserstein two sample testing and related families of non-parametric tests, 2015. 4
- [57] Xiangyu Rui, Xiangyong Cao, Yining Li, and Deyu Meng. Variational zero-shot multispectral pansharpening. *IEEE TGRS*, 62:1–16, 2024. 3
- [58] Xiangyu Rui, Xiangyong Cao, Li Pang, Zeyu Zhu, Zongsheng Yue, and Deyu Meng. Unsupervised hyperspectral pansharpening via low-rank diffusion model. *Inf. Fusion.*, 107:102325, 2024. 2, 3
- [59] Filippo Santambrogio. *One-dimensional issues*, pages 59–85. Springer International Publishing, Cham, 2015. 4

- [60] Giuseppe Scarpa and Matteo Ciotola. Full-resolution quality assessment for pansharpening. *Remote Sens.*, 14(8), 2022. 6
- [61] Vijay P. Shah, Nicolas H. Younan, and Roger L. King. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. *IEEE TGRS*, 46(5):1323–1335, 2008. 2
- [62] Xunyang Su, Jinjiang Li, and Zhen Hua. Transformer-based regression network for pansharpening remote sensing images. *IEEE TGRS*, 60:1–23, 2022. 2, 3
- [63] Weiwei Sun, Jun Zhou, Xiangchao Meng, Gang Yang, Kai Ren, and Jiangtao Peng. Coupled temporal variation information estimation and resolution enhancement for remote sensing spatial-temporal-spectral fusion. *IEEE TGRS*, 61:1–18, 2023. 3
- [64] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9(86):2579–2605, 2008. 2
- [65] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017. 2, 3
- [66] Cédric Villani et al. *Optimal transport: old and new*. Springer, 2008. 4
- [67] Gemine Vivone, Miguel Simões, Mauro Dalla Mura, Rocco Restaino, José M. Bioucas-Dias, Giorgio A. Licciardi, and Jocelyn Chanussot. Pansharpening based on semiblind deconvolution. *IEEE TGRS*, 53(4):1997–2010, 2015. 3
- [68] Gemine Vivone, Mauro Dalla Mura, Andrea Garzelli, Rocco Restaino, Giuseppe Scarpa, Magnus O. Ulfarsson, Luciano Alparone, and Jocelyn Chanussot. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE GRSM*, 9(1):53–81, 2021. 1, 2, 3
- [69] Lucien Wald, Thierry Ranchin, and Marc Mangolini. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.*, 63:691–699, 1997. 3, 6
- [70] Hebaixu Wang, Meiqi Gong, Xiaoguang Mei, Hao Zhang, and Jiayi Ma. Deep unfolded network with intrinsic supervision for pan-sharpening. *AAAI*, 38(6):5419–5426, 2024. 3
- [71] Hebaixu Wang, Hao Zhang, Xin Tian, and Jiayi Ma. Zero-Sharp: A universal pansharpening method across satellites for reducing scale-variance gap via zero-shot variation. *Inf. Fusion.*, 101:102003, 2024. 3
- [72] Jun-Da Wang, Liang-Jian Deng, Chen-Yu Zhao, Xiao Wu, Hong-Ming Chen, and Gemine Vivone. Cascadic multireceptive learning for multispectral pansharpening. *IEEE TGRS*, 61:1–16, 2023. 3
- [73] Yingying Wang, Xuanhua He, Yuhang Dong, Yunlong Lin, Yue Huang, and Xinghao Ding. Cross-modality interaction network for pan-sharpening. *IEEE TGRS*, 62:1–16, 2024. 3
- [74] Zhong-Cheng Wu, Ting-Zhu Huang, Liang-Jian Deng, Jin-Fan Hu, and Gemine Vivone. VO+Net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening. *IEEE TGRS*, 60:1–16, 2022. 3
- [75] Jin-Liang Xiao, Ting-Zhu Huang, Liang-Jian Deng, Zhong-Cheng Wu, Xiao Wu, and Gemine Vivone. Variational pansharpening based on coefficient estimation with nonlocal regression. *IEEE TGRS*, 61:1–15, 2023. 3
- [76] Guangxu Xie, Rencan Nie, Jinde Cao, He Li, and Jintao Li. A deep multiresolution representation framework for pansharpening. *IEEE TGRS*, 62:1–16, 2024. 3
- [77] Yinghui Xing, Litao Qu, Shizhou Zhang, Jiapeng Feng, Xi-wei Zhang, and Yanning Zhang. Empower generalizability for pansharpening through text-modulated diffusion model. *IEEE TGRS*, 2024. 3
- [78] Yinghui Xing, Litao Qu, Shizhou Zhang, Kai Zhang, Yanning Zhang, and Lorenzo Bruzzone. CrossDiff: Exploring self-supervised representation of pansharpening via cross-predictive diffusion model. *IEEE TIP*, 2024. 2, 3
- [79] Qizhi Xu, Yuan Li, Jinyan Nie, Qingjie Liu, and Mengyao Guo. UPanGAN: Unsupervised pansharpening based on the spectral and spatial loss constrained generative adversarial network. *Inf. Fusion.*, 91:31–46, 2023. 3
- [80] Shuang Xu, Jianshe Zhang, Zixiang Zhao, Kai Sun, Junmin Liu, and Chunxia Zhang. Deep gradient projection networks for pan-sharpening. In *CVPR*, pages 1366–1375, 2021. 3, 5
- [81] Gang Yang, Man Zhou, Keyu Yan, Aiping Liu, Xueyang Fu, and Fan Wang. Memory-augmented deep conditional unfolding network for pansharpening. In *CVPR*, pages 1778–1787, 2022. 3, 5, 6, 7
- [82] Jie Yang, Xiao-Yang Fu, Yan Hu, Yue Huang, Xing Ding, and John Paisley. PanNet: A deep network architecture for pan-sharpening. *ICCV*, pages 1753–1761, 2017. 2, 3, 5
- [83] Qiangqiang Yuan, Yancong Wei, Xiangchao Meng, Huanfeng Shen, and Liangpei Zhang. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE JSTARS*, 11(3):978–989, 2018. 2, 3, 5
- [84] Jie Zhang, Xuanhua He, Keyu Yan, Ke Cao, Rui Li, Chengjun Xie, Man Zhou, and Danfeng Hong. Pan-sharpening with wavelet-enhanced high-frequency information. *IEEE TGRS*, 62:1–14, 2024. 3
- [85] Chen-Yu Zhao, Tian-Jing Zhang, Ran Ran, Zhi-Xuan Chen, and Liang-Jian Deng. LGPConv: Learnable gaussian perturbation convolution for lightweight pansharpening. In *IJCAI*, pages 4647–4655, 2023. Main Track. 3
- [86] Mengjiao Zhao, Mengting Ma, Xiangdong Li, Xiaowen Ma, Wei Zhang, and Siyang Song. Supervised detail-guided multiscale state space model for pan-sharpening. *IEEE TGRS*, 2024. 2, 3
- [87] Kaiwen Zheng, Jie Huang, Man Zhou, Danfeng Hong, and Feng Zhao. Deep adaptive pansharpening via uncertainty-aware image fusion. *IEEE TGRS*, 61:1–15, 2023. 5
- [88] Yu Zhong, Xiao Wu, Liang-Jian Deng, Zihan Cao, and Hong-Xia Dou. SSDiff: Spatial-spectral integrated diffusion model for remote sensing pansharpening. In *NeurIPS*, pages 77962–77986. Curran Associates, Inc., 2024. 3
- [89] Man Zhou, Jie Huang, Keyu Yan, Hu Yu, Xueyang Fu, Aiping Liu, Xian Wei, and Feng Zhao. Spatial-frequency domain information integration for pan-sharpening. In *ECCV*, pages 274–291, Cham, 2022. Springer Nature Switzerland. 3, 5
- [90] Man Zhou, Jie Huang, Keyu Yan, Danfeng Hong, Xiuping Jia, Jocelyn Chanussot, and Chongyi Li. A general spatial-frequency learning framework for multimodal image fusion. *IEEE TPAMI*, pages 1–18, 2024. 3, 5