OTPNet: ODE-inspired Tuning-free Proximal Network for Remote Sensing Image Fusion

Wei Yu, Zonglin Li, Qinglin Liu, Xin Sun*

School of Computer Science and Technology, Harbin Institute of Technology, China
{20b903014, zonglin.li, qinglin.liu}@stu.hit.edu.cn, sunxintyc@hit.edu.cn

Abstract

Remote sensing image fusion aims to reconstruct a high spatial and spectral resolution image by integrating the spatial and spectral information from multiple remote sensing sensor data. Despite the remarkable progress of deep learning-based fusion methods, most existing methods rely on manual network architecture design and hyperparameter tuning, lacking sufficient interpretability and adaptability. To address this limitation, we propose a novel neural Ordinary Differential Equation (ODE)-inspired tuning-free proximal splitting algorithm, which splits remote sensing image fusion into two optimization problems regularized by deep priors to model the fusion of spatial and spectral. Firstly, based on the physical properties of spatial and spectral information, the two problems are optimized by two proximal splitting operators to iteratively integrate spatial-spectral complementary information, eliminating or suppressing redundant information to reduce fusion errors. Secondly, considering the efficiency of neural ODE in reducing optimization error, we utilize a high-order numerical scheme to customize the proximal operator theoretically without additional handcrafted design and parameter tuning. Finally, by incorporating the numerical scheme as a solver into the proximal optimization algorithm, we derive an ODE-inspired Tuning-free Proximal Network, dubbed OTPNet, which achieves efficient and robust fusion reconstruction. Extensive experiments on nine datasets across three different remote sensing image fusion tasks show that our OTPNet outperforms existing state-of-the-art approaches, which validates the effectiveness of our method.

Introduction

Remote sensing images have been widely used in many visual fields (Berni et al. 2009; Cheng, Han, and Lu 2017; Deng et al. 2018), leading to an increasing demand for high-resolution and high-quality remote sensing images. However, it is impractical to directly obtain remote sensing images with both high spatial and spectral resolution due to physical limitations in sensor design and data acquisition. A practical solution (Hu et al. 2022; Zhou et al. 2022) is to integrate the complementary information from two sensor images, where one sensor typically offers spatial resolution information and the other spectral resolution information. By combining the strengths of both sensors, this solution enables the reconstruction of an image with both high spatial and spectral resolution, providing richer information than either sensor can achieve individually. A well-known fusion task is the pan-sharpening, which involves merging a high spatial resolution PANchromatic (PAN) image and a Low spatial Resolution MultiSpectral (LR-MS) image to generate a High-Resolution MultiSpectral (HR-MS) image with high spatial and spectral resolution. It serves as a typical case to demonstrate our remote sensing fusion method.

Typically, it is assumed that the HR-MS image maintains the rich spectral information and produces the LR-MS images through spatial downsampling and blurring degradation. Simultaneously, the HR-MS image maintains a high spatial resolution and derives the single-band PAN image with a specific spectral response function. Therefore, the pan-sharpening algorithm essentially restores the spatial and spectral degradation information of the input images simultaneously and reduces the introduction of artifacts and fusion noise to produce the required HR-MS. The physical degradation process of the HR-MS image $H \in \mathbb{R}^{HW \times C}$ is usually defined mathematically as follows

$$L = DH, \quad P = HS \tag{1}$$

where $L \in \mathbb{R}^{hw \times C}$ and $P \in \mathbb{R}^{HW \times c}$ represent the degrated LR-MS and PAN images, respectively. $D \in \mathbb{R}^{hw \times HW}$ and $S \in \mathbb{R}^{C \times c}$ denote the spatial degradation matrix and the spectral response matrix, respectively. With the rapid development of deep learning, pan-sharpening technology based on neural networks (Ghassemian 2016; Ma et al. 2020; Wang, Ma, and Zhang 2023) has achieved remarkable results by designing neural networks with different structures to fuse spatial and spectral information from two sensors.

Despite significant advancements in existing deep learning-based methods (Yang et al. 2017; Jin et al. 2022), the majority of these approaches heavily rely on manually designed network architectures and tuned hyperparameters, lacking sufficient interpretability and adaptability. To address this issue, model-driven deep learning algorithms (Xu et al. 2021; Wang et al. 2021) have received more attention, which integrate prior knowledge for specific tasks to formulate an observed optimization problem and transform each iteration of the solution algorithm into a layer of a deep neural network. Their adaptability and effectiveness have

*Corresponding author (sunxintyc@hit.edu.cn).

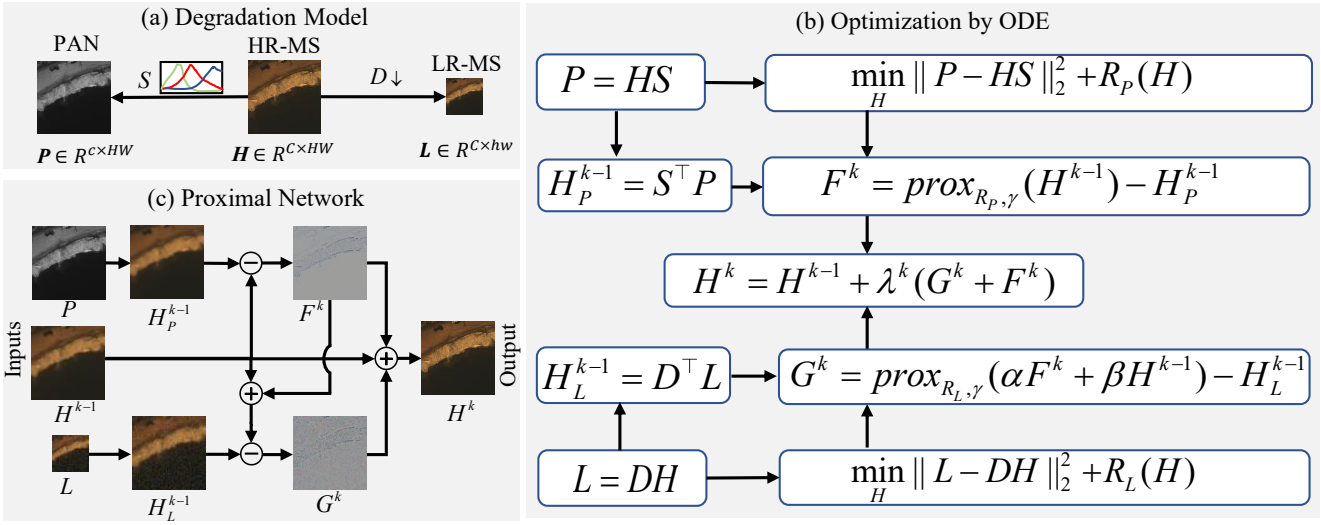

Figure 1: The schematic diagram of the pan-sharpening fusion task. (a) The physical degradation model of the pan-sharpening task; (b) Iterative solution steps for two formulaic proximal optimization problems with the high-order ODE numerical scheme; (c) The derived ODE-inspired tuning-free proximal optimization network.

been demonstrated in various computer vision tasks (Yang and Sun 2018; Wang et al. 2020) and have high theoretical interpretability Therefore, we propose a novel neural Ordinary Differential Equation (ODE)-inspired tuning-free proximal splitting algorithm, which splits remote sensing image fusion into two optimization problems regularized by deep priors and derives these into higher-order ODE forms for iterative solving via deep proximal unfolding networks. Firstly, inspired by the physical properties of spatial and spectral information, the two problems are optimized by two proximal splitting operators to iteratively integrate spatial-spectral complementary information, eliminating or suppressing redundant information to reduce fusion errors. Secondly, considering the efficiency of neural ODE in reducing optimization error, we utilize a high-order numerical scheme to customize the proximal operator theoretically without additional handcrafted design and parameter tuning. Finally, by incorporating the numerical scheme as a solver into the proximal optimization algorithm, we derive an ODE-inspired Tuning-free Proximal Network, dubbed OTPNet, which achieves efficient and robust fusion reconstruction. Our contributions are summarized as follows:

- We propose a tuning-free proximal splitting algorithm to divide the spatial and spectral fusion task into two sub-optimization problems, which customize two corresponding proximal operators to iteratively reconstruct the spatial-spectral information.
- We theoretically establish a formal relationship between the proximal operator and the neural ODE by introducing a higher-order numerical scheme as a solver, and deriving an interpretable proximal unfolding network with theoretical guidance.
- Extensive experiments on nine popular datasets across three different remote sensing image fusion tasks demonstrate that our OTPNet achieves competitive performance across state-of-the-art (SOTA) approaches.

Related Work

Traditional Fusion Methods

The traditional remote sensing image fusion methods have been extensively studied and can be roughly divided into three categories according to the type of fused sensor data: panchromatic image sharpening, multispectral image super-resolution, and multispectral and hyperspectral image fusion. The traditional fusion strategies they employed primarily include methods based on Component Substitution (CS) (Shah, Younan, and King 2008; Ghahremani and Ghassemian 2016), Multi-Resolution Analysis (MRA) (Khan et al. 2008), and Variational Optimization (VO) (Tian et al. 2020; Chen et al. 2015). CS algorithms (Aiazzi, Baronti, and Selva 2007; Gillespie, Kahle, and Walker 1987) enhance the fused MS images by substituting the spatial components with PAN images to improve the fusion speed. MRA approaches (Vivone et al. 2014) obtain more spatial information by spatial filtering and have less spectral distortion. Variational methods (Ballester et al. 2006; Fu et al. 2019) assume that the PAN image is formed by a linear combination of different multi-spectral bands, while the LR-MS results from the blurred transformation of HR-MS images. Recently, model-based methods (Xie et al. 2019; Yan et al. 2022) attracted a lot of attention, which explicitly represent prior information and imaging models to fuse images by formulating a specific optimization problem, which can be intuitively interpreted as deep unfolding network architecture and iterative design. These methods mainly rely on Half Quadratic splitting (Yang et al. 2022) and gradient projection (Xu et al. 2021), which require explicit gradient computation and precise parameter tuning, resulting in poor efficiency and adaptability. Moreover, these methods require numerous handcrafted priors and parameter tuning, which increases the complexity of model design and limits their applicability in practical application.

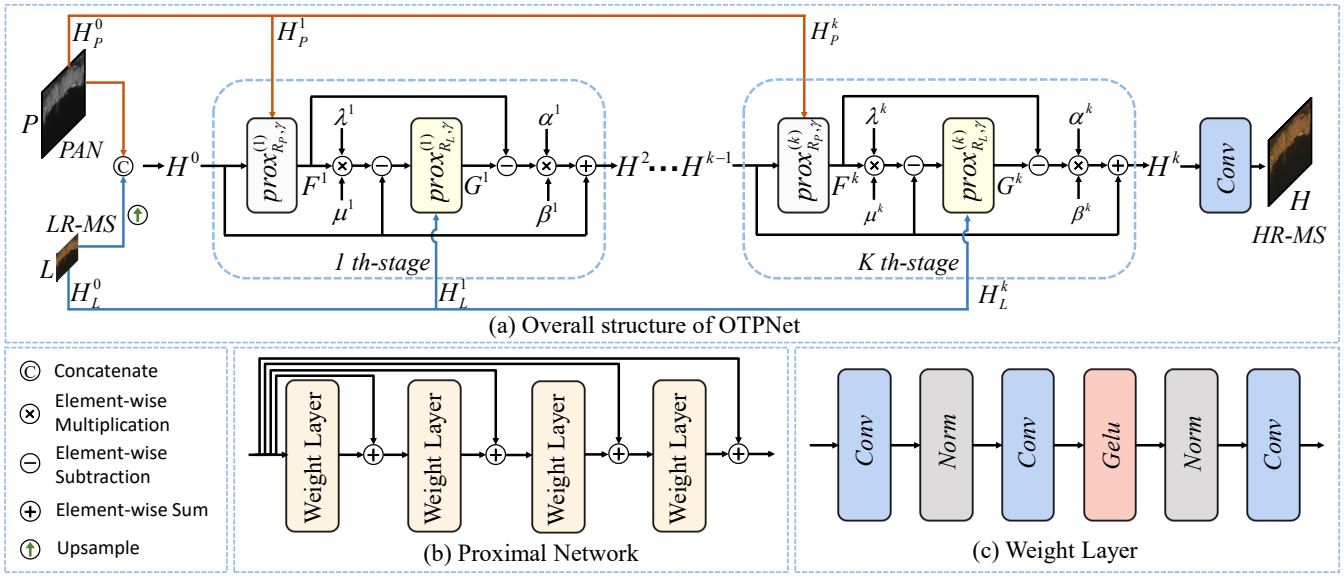

Figure 2: Illustration of the proposed OTPNet. (a) Overall structure of the OTPNet. (b) The structure of the ODE-inspired proximal network. (c) The structure of the weight layer in the proximal network.

Deep Learning-based Fusion Methods

Recently, deep learning-based approaches (Masi et al. 2016; Cai and Huang 2020; Yang et al. 2017; Wang et al. 2021) leverage the powerful learning capabilities of CNNs and Transformers to capture the complex end-to-end mapping functions required for remote sensing image fusion. These methods primarily depend on carefully crafted stacks of neural network layers or attention mechanisms to adaptively fuse spatial and spectral information, capitalizing on their complementary information. MIPSM (Liu et al. 2020) proposes a spatial feature extraction network for the PAN images and injects the details into the LR-MS images by spatial attention mechanism. Meanwhile, researchers explore the dual branch networks (Peng, Guo, and Wu 2022; Yuan et al. 2018) for LR-MS and PAN feature interaction and fusion to further improve the reconstruction performance. TFNet (Liu, Liu, and Wang 2020) extracts spatial and spectral information from PAN and MS images through two networks and fuses them to form compact features, achieving robust and high-performance fusion reconstruction. Swin-STFM (Chen et al. 2022) introduces Swin Transformer to fully utilize its advantages in feature extraction, which greatly improves the quality of fused images. However, these deep learning-based methods lack theoretical validation and interpretability due to the black-box nature of deep neural networks, which hinders a deeper understanding of fusion works and limits their actual generalization performance.

Proposed Method

In this section, we first introduce the tuning-free proximal splitting algorithm. Then, we explain the proximal operator from the perspective of ordinary differential equations. Finally, we derive the ODE-inspired proximal network and provide a theoretical demonstration of the accuracy and reduced error in the higher-order ODE numerical method.

Tuning-free Proximal Algorithm

Based on the physical degradation model of the HR-MS image in Eq. 1, we propose a tuning-free proximal algorithm to solve the pan-sharpening fusion function, which is optimized by two proximal splitting operators to iteratively integrate spatial-spectral complementary information. Specifically, we formulate the HR-MS image fusion reconstruction as an optimization problem by minimizing the following energy function as

$$\min_{H} \frac{1}{2}||L - DH||_2^2 + \frac{1}{2}||P - HS||_2^2 + R(H) \tag{2}$$

where the first two terms indicate the data fidelity terms of the LR-MS and PAN images, respectively. $R(\cdot)$ indicates the prior term to capture the implicit prior through neural network parameterization. We decompose the original energy function into an LR-MS spatial reconstruction problem and a PAN spectral reconstruction problem to enable the full utilization of deep priors. Mathematically, the sub-problem of spatial and spectral reconstruction can be represented as

$$\begin{aligned} &\min_{H} \frac{1}{2}||P - HS||_2^2 + R_P(H) \\ &\min_{H} \frac{1}{2}||L - DH||_2^2 + R_L(H) \end{aligned} \tag{3}$$

where $R_L(\cdot)$ and $R_P(\cdot)$ represent the two deep priors corresponding to the LR-MS and PAN image observations, respectively. To solve Eq. (3), we customize corresponding proximal operators for each of them as follows

$$prox_{h,\varphi}(\mathbf{x}) = \arg\min_{\mathbf{z}} \varphi||\mathbf{z} - \mathbf{x}||_2^2 + h(\mathbf{x}) \tag{4}$$

By applying Eq. 3 in Eq. 4, we derive the explicit formulations for the proximal operators as follows

$$\begin{aligned} prox_{R_P,\gamma}(H) &= \arg\min_{H} \gamma||P - HS||_2^2 + R_P(H) \\ prox_{R_L,\gamma}(H) &= \arg\min_{H} \gamma||L - DH||_2^2 + R_L(H) \end{aligned} \tag{5}$$

Networks	Corresponding ODE Formula	ODE Scheme
ResNet (He et al. 2016)	$\mathcal{N}^{t+1} = \mathcal{N}^t + \mathcal{F}(\mathcal{N}^t)$	Forward Euler
PolyNet (Zhang et al. 2017)	$\mathcal{N}^{t+1} = \mathcal{N}^t + \mathcal{F}(\mathcal{N}^t) + \mathcal{F}(\mathcal{F}(\mathcal{N}^t))$	Backward Euler
LM-ResNet (Lu et al. 2018)	$\mathcal{N}^{t+1} = (1-\Delta t)\mathcal{N}^t + \Delta t\mathcal{F}(\mathcal{N}^{t-1}) + \mathcal{F}(\mathcal{N}^t)$	Linear-MultiStep
FractalNet (Larsson, Maire, and Shakhnarovich 2016)	$\mathcal{N}^{t+1} = \mathcal{N}^t + \mathcal{F}(\mathcal{N}^t) + \mathcal{F}(\Delta t\mathcal{F}(\mathcal{N}^t) + \mathcal{N}^t)$	Runge-Kutta
Second-Order CNNs (Burrage, Lenane, and Lythe 2007)	$\mathcal{N}^{t+1} = 2\mathcal{N}^t - \mathcal{N}^{t-1} + \Delta t^2\mathcal{F}(\mathcal{N}^t)$	Second-Order

Table 1: Neural networks and their corresponding ODE schemes.

where $prox_{R_P,\gamma}(\cdot)$ and $prox_{R_L,\gamma}(\cdot)$ represent the proximal operators for the spatial and spectral reconstruction problems, respectively. To facilitate alternating iterations of the proximal operators for effective reconstruction, we employ the Douglas-Rachford (DR) splitting algorithm (Eckstein and Bertsekas 1992) to solve these non-smooth optimization problems of Eq. (5). By iteratively applying the DR splitting method, we can ensure convergence towards an optimal solution while maintaining computational efficiency. The corresponding update rules of the DR splitting algorithm as

$$\begin{aligned} u^k &= prox_{f,\varphi}(y^{k-1}) \\ w^k &= prox_{g,\varphi}(2u^k - y^{k-1}) \\ y^k &= y^{k-1} + (w^k - u^k) \end{aligned} \tag{6}$$

Building on Eq. (6) and Eq. (5), we incorporate learnable parameters $(\mu, \lambda, \alpha, \beta)$ alongside the residual learning concept (He et al. 2016) to formulate the efficient tuning-free proximal search steps, which expressed as follows

$$\begin{aligned} F^k &= prox_{R_P,\gamma}(H^{k-1}) - H_P^k \,, \\ G^k &= prox_{R_L,\gamma}(\lambda^k F^{k-1} + \mu^k H^{k-1}) - H_L^k \,, \\ H^k &= H^{k-1} + (\alpha^k G^k + \beta^k F^k) \end{aligned} \tag{7}$$

where H^k and H^{k-1} denote the output and input of the k-th iteration to learn intricate spatial-spectral relationships, respectively. H_P^{k-1} and H_L^{k-1} represent the intermediate results of the HR-MS image derived from PAN and LR-MS by updating the DR matrix to eliminate and suppress redundant information, respectively. G^k and F^k indicate the spatial and spectral difference signals to minimize fusion errors and ensure accurate reconstruction. α^k, β^k, μ^k, and λ^k are learnable step size parameters of the k-th iteration.

From Proximal Operator to ODEs

Benefiting from the powerful learning ability of deep networks, recent approaches have extended the traditional proximal operator by integrating it with convolutional neural networks (Xu et al. 2021; Cao, Chen, and Cao 2022) to achieve effective and flexible modeling of complex data structures. As illustrated in Table 1, numerical differential equations have become a common approach for interpreting deep neural networks and guiding the design of network architectures (Lu et al. 2018). Building upon this foundation, we establish a connection between proximal networks and numerical differential equations and design the proximal network via a novel high-order Runge-Kutta (RK) numerical scheme. Inspired by the previous first-order RK (Larsson, Maire, and Shakhnarovich 2016) scheme, the formulation of the t-th layer of our proximal network is expressed as

$$\mathcal{N}^{t+1} = \mathcal{F}(N^t, W^t) + \mathcal{N}^t \tag{8}$$

where $\mathcal{F}(\cdot, W)$ represents the operation of a single layer with learnable parameters W in the proximal network. As $\Delta t \to 0$, the expression converges to

$$\lim_{\Delta t \to 0} \frac{\mathcal{N}_{t+\Delta t} - \mathcal{N}_t}{\Delta t} = \frac{d\mathcal{N}(t)}{dt} = \mathcal{F}(\mathcal{N}(t), t) \tag{9}$$

Consequently, the proximal network can be optimized by ODEs with parameter learning. The initial parameter value x is transformed into a set of features $\mathcal{N}(0)$ by solving the initial value problem (Li, Osborne, and Prvan 2005) of the ODE. This transformation is described as the equation

$$\frac{d\mathcal{N}(t)}{dt} = \mathcal{F}(\mathcal{N}(t), t) \,, \;\; \mathcal{N}(0) = x \tag{10}$$

where $\mathcal{N}(t)$ represents the learned feature map at the t-th layer, which is derived from the initial feature value $\mathcal{N}(0)$ by solving the ODE equation.

ODE-inspired Proximal Network

To fully leverage the advantages of model-driven and network-based methods, we learn the proximal operator of the mentioned tuning-free proximal algorithm using an ODE-inspired proximal network As shown in Table 1, most existing fusion methods (Xu et al. 2021; Cao, Chen, and Cao 2022) adopt the original residual network (He et al. 2016), which is considered as the forward Euler discretization of the ODE equations. For pansharpening fusion tasks, the first-order accuracy and inherent instability of the forward Euler method can lead to insufficient spatial details and spectral distortion. Additionally, the accumulation of numerical errors during fusion may introduce blurring or artifacts, degrading the quality of the high-resolution fused image. Compared to the forward Euler method, the high-order Runge-Kutta (RK) method improves accuracy and stability by effectively reducing numerical errors and enhancing convergence. It can handle larger step sizes without sacrificing accuracy, making it efficient for solving the two optimization problems associated with fusion tasks. Therefore, we adopt the high-order RK method as the foundation for designing our proximal network. The formula for the n-stage iteration

of the RK method is given by

$$\begin{aligned} y_{n+1} &= y_n + h\sum_{i=1}^{n} v_i k_i \\ k_1 &= f(x_n, y_n) \\ k_i &= f(x_n + d_i h,\ y_n + h\sum_{j=1}^{i-1} p_{ij} k_j) \end{aligned} \tag{11}$$

where h denotes the step size, v_i represents the weight coefficients. k_i represents the intermediate estimates utilized to approximate the solution of the differential equation at different stages within each iteration. d_i and p_{ij} denote the intermediate coefficients to define the intermediate points and weights for calculating the k_i. In particular, the fourth-order RK is described as follows

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ k_1 &= hf(x_n,\ y_n) \\ k_2 &= hf(x_n + \frac{h}{2},\ y_n + \frac{k_1}{2}) \\ k_3 &= hf(x_n + \frac{h}{2},\ y_n + \frac{k_2}{2}) \\ k_4 &= hf(x_n + h,\ y_n + k_3) \end{aligned} \tag{12}$$

Building on the aforementioned definitions, we design a novel proximal network that leverages this fourth-order RK approach. Specifically, we incorporate four trainable parameters $(\omega^1, \omega^2, \omega^3, \omega^4)$ to seamlessly integrate the network architecture with tuning-free optimization. This design enhances the interpretability and adaptability of the fusion network by leveraging the fourth-order RK method's structured approach to parameter learning. The integration of fourth-order RK enables more accurate and stable solutions by refining intermediate estimates and balancing accuracy with computational efficiency, which allows the network to better handle the complexities of multi-fusion tasks, as detailed in the following formulation

$$\begin{aligned} \mathcal{N}^{t+1} &= \mathcal{N}^t + \omega^4(k_1 + 2k_2 + 2k_3 + k_4) \\ k_1 &= \mathcal{F}(N^t) \\ k_2 &= \mathcal{F}(N^t + \omega^1 k_1) \\ k_3 &= \mathcal{F}(N^t + \omega^2 k_2) \\ k_4 &= \mathcal{F}(N^t + \omega^3 k_3) \end{aligned} \tag{13}$$

Local Truncation Error Minimization

Local truncation error quantifies the discrepancy between a numerical approximation and the exact solution of a differential equation, which reflects the accuracy of the numerical method. For the fusion task of a given initial value problem

$$\frac{dy}{dt} = f(t, y)\ ,\ \ y(t_0) = y_0 \tag{14}$$

In numerical analysis, minimizing its truncation error is a primary objective, which signifies that the numerical solution more closely approximates the true solution. Then, we apply the forward Euler method to yield the following numerical approximation

$$y(t + h) = y(t) + hf(t, y(t)) \tag{15}$$

By performing a Taylor expansion on the true solution $y(t)$, we derive the following conclusion

$$y(t + h) = y(t) + h * y'(t) + O(h^2) \tag{16}$$

From Eq. (14) and Eq. (16), we can determine that the local truncation error for the forward Euler method is $O(h^2)$. In contrast, the fourth-order RK method achieves a local truncation error of $O(h^5)$, reflecting its superior accuracy due to the higher-order terms in its Taylor series expansion. This indicates that the fourth-order RK method is inherently more suited to the intricate requirements of fusion tasks, offering enhanced precision and stability compared to the previous forward Euler method.

Experiments

Datasets. Remote sensing image fusion is categorized into different tasks based on the varying input data from multiple sensors. Essentially, all of these tasks involve extracting and fusing spatial and spectral information from the input images. In this study, we focus on three widely researched fusion tasks: 1) panchromatic image sharpening (Pan-Sharpening); 2) hyperspectral image super-resolution (HSR); and 3) multispectral image and hyperspectral image fusion (MHF) to validate our fusion method.

Specifically, Pan-Sharpening inputs a high-resolution panchromatic (PAN) image and a low-resolution multispectral (LR-MS) image to output a high-resolution multispectral (HR-MS) image. HSR takes a low-resolution hyperspectral (LR-HS) image and a high-resolution panchromatic (HR-MS) or panchromatic (PAN) image as input to generate a high-resolution hyperspectral (HR-HS) image. MHF inputs a low-resolution hyperspectral (LR-HS) image and a high-resolution multispectral (HR-MS) image to produce a high-resolution hyperspectral (HR-HS) image.

To comprehensively evaluate our approach, we conduct experiments on nine fusion datasets corresponding to these three tasks: Pan-Sharpening task with WorldView-3, GaoFen-2, and QuickBird satellite datasets; HSR task with PaviaCentre, Botswana4, and Chikusei datasets; MHF task with CAVE, Harvard and NTIRE2020 datasets. For all datasets, we allocate 90% of the data to the training set and the remaining 10% to the validation set. Each dataset, despite containing different types of image pairs (e.g., PAN/LR-MS and HR-MS for the WorldView-3 dataset), follows the same processing approach, with sample sizes of 64×64, $C \times 16 \times 16$, and $C \times 64 \times 64$, respectively.

Implementation Details. The proposed OTPNet is implemented in PyTorch and trained using the AdamW optimizer with parameters set to β_1 = 0.9 and β_2 = 0.999. The training is conducted over 120k steps in a multistep schedule. The initial learning rate is set to 2e-4 and is reduced by half every 30k iterations. We adopt a batch size of 64 for all experiments. For additional implementation details and metrics, please refer to the supplementary materials.

Method	GaoFen2				Worldview3				QuickBird			
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓
Brovey (Gillespie, Kahle, and Walker 1987)	37.739	0.947	1.852	2.302	30.514	0.849	5.838	5.869	30.360	0.813	8.557	9.508
GS (Laben and Brower 2000)	37.791	0.947	2.087	2.420	31.288	0.853	6.127	5.524	30.354	0.808	8.672	9.537
IHS (Haydn 1982)	37.879	0.953	1.854	2.365	30.395	0.835	6.497	5.750	30.274	0.806	8.850	9.600
SFIM (Liu 2000)	40.991	0.957	1.669	1.697	32.004	0.867	5.373	5.077	31.122	0.827	8.159	9.462
PANNet (Yang et al. 2017)	46.003	0.987	0.956	0.946	37.540	0.967	3.385	2.613	37.159	0.951	5.018	4.292
SRPPNN (Cai and Huang 2020)	47.673	0.989	0.877	0.783	38.791	0.972	2.958	2.213	36.525	0.944	5.081	4.618
GPPNN (Xu et al. 2021)	45.097	0.981	1.042	1.077	37.971	0.968	3.248	2.446	37.552	0.954	4.811	4.060
MUCNN (Wang et al. 2021)	47.286	0.988	0.878	0.822	38.085	0.967	3.253	2.436	35.146	0.931	5.248	5.467
MDCUN (Yang et al. 2022)	47.787	0.990	0.830	0.758	38.456	0.970	3.084	2.316	36.178	0.944	5.002	4.817
LAGNet (Jin et al. 2022)	47.852	0.990	0.807	0.749	38.869	0.973	2.976	2.205	38.377	0.961	4.509	3.687
PMACNet (Liang et al. 2022)	44.796	0.983	1.388	1.275	38.675	0.972	3.006	2.250	37.523	0.955	4.784	4.099
PANFormer (Zhou et al. 2022)	44.345	0.975	1.393	1.431	37.359	0.965	3.649	2.670	36.124	0.941	5.417	4.828
OTPNet (Ours)	49.340	0.992	0.736	0.646	39.255	0.975	2.850	2.104	38.521	0.962	4.415	3.624

Table 2: Quantitative comparison results of our method and other SOTA Pan-Sharpening methods on three datasets. The optimal and suboptimal results are highlighted in red and blue.

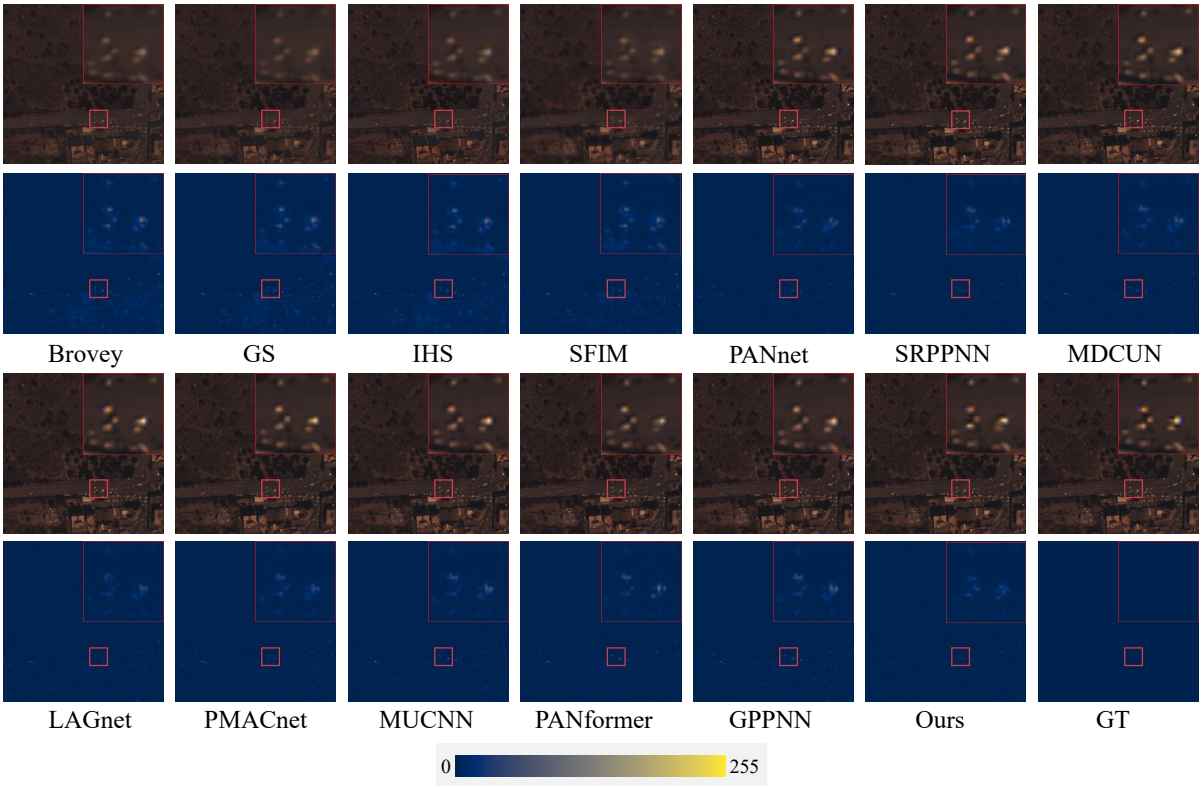

Figure 3: Qualitative results of our method with other SOTA methods on the WorldView3 dataset. The first and third rows display the reconstructed HR-MS images. The second and fourth rows present the absolute error maps. The color bar at the bottom indicates that pixels with small errors are represented in blue, while pixels with large errors are represented in yellow.

Comparison With Other Approaches

Quantitative results. In Table 2, we provide the comparison results between OTPNet and twelve SOTA pan-sharpening methods, including four traditional methods (Brovey (Gillespie, Kahle, and Walker 1987), GS (Laben and Brower 2000), IHS (Haydn 1982) and SFIM (Liu 2000)), two model-driven deep unfolding methods (GPPNN (Xu et al. 2021) and MDCUN (Yang et al. 2022)), and six deep learning-based methods (PANnet (Yang et al. 2017), SRPPNN (Cai and Huang 2020), MUCNN (Wang et al. 2021), LAGnet (Jin et al. 2022), PMACnet (Liang et al. 2022) and PANFormer (Zhou et al. 2022)). Our method achieves the best performance on all metrics across the testing datasets, demonstrating its effectiveness. *More quantitative results of HSR and MHF fusion tasks on the other six fusion datasets in the supplementary materials.*

Qualitative results Figure 3 presents qualitative visualization results between our OTPNet and other methods on the WorldView3 dataset. It can be seen that when the reconstructed images using traditional methods, noticeable blurring can be observed. The absolute error maps show that the errors produced by our method are significantly smaller than those from other SOTA methods. *More visualization results of HSR and MHF tasks in the supplementary materials.*

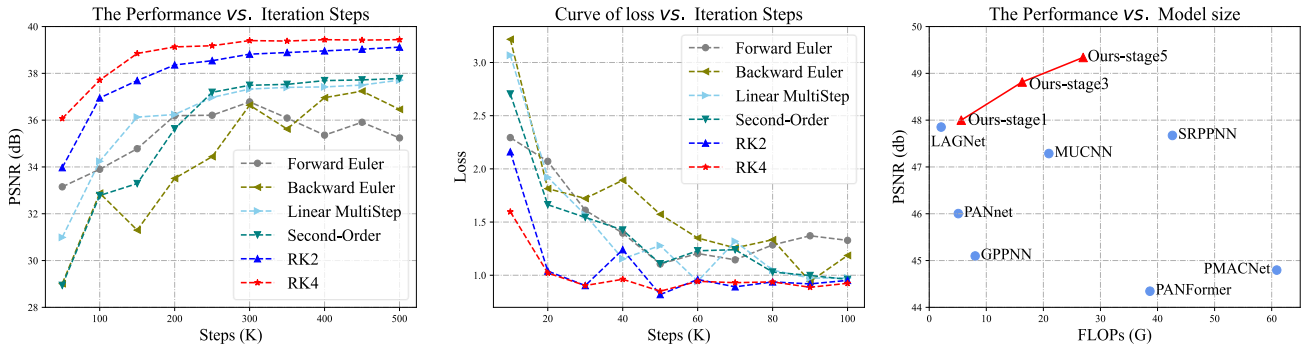

Figure 4: The left and middle figures illustrate the change curves of performance PSNR and training loss for the network designed with different numerical ODE schemes during the iterative process. The right figure presents a comparison of the computational complexity and performance of existing SOTA methods.

Framework	GP	ADMM	HQS	DR(Ours)
PSNR↑	35.833	38.133	39.068	**39.255**
SSIM↑	0.950	0.968	0.974	**0.975**
SAM↓	4.279	3.198	2.889	**2.850**
ERGAS↓	3.366	2.415	2.145	**2.104**

Table 3: Comparison results of different proximal splitting methods for the fusion optimization problem.

Ablation Studies

Effect of Proximal Splitting Algorithm. As illustrated in Table 3, we provide comparison results of our framework using different proximal splitting algorithms including Gradient Projection (GP) (Xu et al. 2021), ADMM (Sun et al. 2016), Half Quadratic Splitting (HQS) (Yang et al. 2022), and Douglas-Rachford (DR) algorithms on the WorldView3 dataset. Our DR proximal algorithm significantly outperforms GP, ADMM, and HQS by 3.4, 1.1, and 0.19 dB in terms of PSNR, respectively. This superior performance is due to our method's ability to adaptively exploit space-spectral complementary information via alternating operator iterations, thus optimizing fusion reconstruction.

Effect of Numerical ODE scheme. As shown in Table 4, we provide comparison results of our framework using different numerical ODE schemes including the forward Euler method (He et al. 2016), the backward Euler method (Zhang et al. 2017), the linear multistep method (Lu et al. 2018), the second-order numerical scheme (Burrage, Lenane, and Lythe 2007), and the adopted Runge-Kutta numerical scheme. Compared to previous first-order numerical schemes, our higher-order method (i.e., RK2 and RK4)demonstrates superior performance, exceeding them by at least 2.2 dB of PSNR, which indicates that the solution of the proximal operator by high-order numerical schemes yields superior results compared to those steered by lower-order schemes. Furthermore, we provide performance and convergence comparisons across various numerical schemes in Figure 4. The performance curve demonstrates quicker accuracy enhancement and smoother progression with our method as iterations increase. The loss curve in the center graph shows our method converging post 40K iterations with minimal oscillation amplitude, signifying improved robustness and efficiency. It highlights the advantages of our proposed method in terms of robustness and computation cost.

Architecture	Forward Euler	Backward Euler	Linear MultiStep	Second Order	RK2	RK4(Ours)
PSNR↑	37.094	37.297	38.117	38.420	39.050	**39.255**
SSIM↑	0.960	0.962	0.968	0.970	0.974	**0.975**
SAM↓	3.709	3.621	3.212	3.135	2.920	**2.850**
ERGAS↓	2.742	2.676	2.410	2.335	2.159	**2.104**

Table 4: Comparison of different numerical ODE schemes.

Framework	Architecture	PSNR↑	SSIM↑	SAM↓	ERGAS↓
✗	✗	38.859	0.972	2.972	2.213
✓	✗	39.067	0.974	2.905	2.151
✗	✓	39.123	0.974	2.895	2.141
✓	✓	**39.255**	**0.975**	**2.850**	**2.104**

Table 5: Ablation experiments of our tuning-free architecture. ✗ represents fixed parameters, ✓ represents tuning-free.

Effect of Tuning-free. To validate the effectiveness and advantages of our tuning-free approach, we present a comparison of three parameter configurations: fixed, sequentially unfixed, and entirely tuning-free, as shown in Table 5. Experimental results indicate that the tuning-free approach outperforms the static fixed parameter strategy. In our framework, the tuning-free strategy adaptively balances the learned spatial and spectral difference signals. Additionally, within the proximal network context, the tuning-free operation enables adaptive step size recalibration.

Conclusion

In this paper, we reinterpret the relationship between ODE and optimization and propose a novel ODE-inspired Tuning-free Proximal splitting Network (OTPNet) for efficient and robust image fusion reconstruction. We first introduce the Douglas-Rachford proximal optimization method and derive a proximal network with two proximal operators for spatial and spectral reconstruction. Next, we integrate a high-order Runge-Kutta ODE scheme into the proximal algorithm to customize the proximal operator without additional parameter tuning. Future work will focus on developing a more general architecture to bridge ODEs and deep neural networks for other multisource and multimodal fusion tasks.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 62072141, in part by the National Natural Science Foundation of Shandong Province under Grants ZR2024QF136.

References

Aiazzi, B.; Baronti, S.; and Selva, M. 2007. Improving component substitution pansharpening through multivariate regression of MS + Pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10): 3230–3239.

Ballester, C.; Caselles, V.; Igual, L.; Verdera, J.; and Rougé, B. 2006. A variational model for P+ XS image fusion. *International Journal of Computer Vision*, 69(1): 43.

Berni, J. A.; Zarco-Tejada, P. J.; Suárez, L.; and Fereres, E. 2009. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Transactions on geoscience and Remote Sensing*, 47(3): 722–738.

Burrage, K.; Lenane, I.; and Lythe, G. 2007. Numerical methods for second-order stochastic differential equations. *SIAM journal on scientific computing*, 29(1): 245–264.

Cai, J.; and Huang, B. 2020. Super-resolution-guided progressive pansharpening based on a deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6): 5206–5220.

Cao, X.; Chen, Y.; and Cao, W. 2022. Proximal PanNet: A Model-Based Deep Network for Pansharpening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 176–184.

Chen, C.; Li, Y.; Liu, W.; and Huang, J. 2015. SIRF: Simultaneous satellite image registration and fusion in a unified framework. *IEEE Transactions on Image Processing*, 24(11): 4213–4224.

Chen, G.; Jiao, P.; Hu, Q.; Xiao, L.; and Ye, Z. 2022. Swin-STFM: Remote sensing spatiotemporal fusion using Swin transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–18.

Cheng, G.; Han, J.; and Lu, X. 2017. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10): 1865–1883.

Deng, L.; Mao, Z.; Li, X.; Hu, Z.; Duan, F.; and Yan, Y. 2018. UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. *ISPRS journal of photogrammetry and remote sensing*, 146: 124–136.

Eckstein, J.; and Bertsekas, D. P. 1992. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical programming*, 55: 293–318.

Fu, X.; Lin, Z.; Huang, Y.; and Ding, X. 2019. A variational pan-sharpening with local gradient constraints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10265–10274.

Ghahremani, M.; and Ghassemian, H. 2016. Nonlinear IHS: A promising method for pan-sharpening. *IEEE Geoscience and Remote Sensing Letters*, 13(11): 1606–1610.

Ghassemian, H. 2016. A review of remote sensing image fusion methods. *Information Fusion*, 32: 75–89.

Gillespie, A. R.; Kahle, A. B.; and Walker, R. E. 1987. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. *Remote Sensing of Environment*, 22(3): 343–365.

Haydn, R. 1982. Application of the IHS color transform to the processing of multisensor data and image enhancement. In *Proc. of the International Symposium on Remote Sensing of Arid and Semi-Arid Lands, Cairo, Egypt, 1982*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hu, J.-F.; Huang, T.-Z.; Deng, L.-J.; Dou, H.-X.; Hong, D.; and Vivone, G. 2022. Fusformer: A transformer-based fusion network for hyperspectral image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Jin, Z.-R.; Zhang, T.-J.; Jiang, T.-X.; Vivone, G.; and Deng, L.-J. 2022. LAGConv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 1113–1121.

Khan, M. M.; Chanussot, J.; Condat, L.; and Montanvert, A. 2008. Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique. *IEEE Geoscience and Remote Sensing Letters*, 5(1): 98–102.

Laben, C. A.; and Brower, B. V. 2000. Process for enhancing the spatial resolution of multispectral imagery using pansharpening. US Patent 6,011,875.

Larsson, G.; Maire, M.; and Shakhnarovich, G. 2016. Fractalnet: Ultra-deep neural networks without residuals. *arXiv preprint arXiv:1605.07648*.

Li, Z.; Osborne, M. R.; and Prvan, T. 2005. Parameter estimation of ordinary differential equations. *IMA Journal of Numerical Analysis*, 25(2): 264–285.

Liang, Y.; Zhang, P.; Mei, Y.; and Wang, T. 2022. Pmacnet: Parallel multiscale attention constraint network for pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Liu, J. 2000. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of remote sensing*, 21(18): 3461–3472.

Liu, L.; Wang, J.; Zhang, E.; Li, B.; Zhu, X.; Zhang, Y.; and Peng, J. 2020. Shallow–deep convolutional network and spectral-discrimination-based detail injection for multispectral imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 1772–1783.

Liu, X.; Liu, Q.; and Wang, Y. 2020. Remote sensing image fusion based on two-stream fusion network. *Information Fusion*, 55: 1–15.

Lu, Y.; Zhong, A.; Li, Q.; and Dong, B. 2018. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *International Conference on Machine Learning*, 3276–3285. PMLR.

Ma, J.; Yu, W.; Chen, C.; Liang, P.; Guo, X.; and Jiang, J. 2020. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion*, 62: 110–120.

Masi, G.; Cozzolino, D.; Verdoliva, L.; and Scarpa, G. 2016. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7): 594.

Peng, S.; Guo, C.; and Wu, X. 2022. Source-Aware Spatial-Spectral-Integrated Double U-Net for Image Fusion. *arXiv preprint arXiv:2212.06466*.

Shah, V. P.; Younan, N. H.; and King, R. L. 2008. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. *IEEE transactions on geoscience and remote sensing*, 46(5): 1323–1335.

Sun, J.; Li, H.; Xu, Z.; et al. 2016. Deep ADMM-Net for compressive sensing MRI. *Advances in neural information processing systems*, 29.

Tian, X.; Chen, Y.; Yang, C.; Gao, X.; and Ma, J. 2020. A variational pansharpening method based on gradient sparse representation. *IEEE Signal Processing Letters*, 27: 1180–1184.

Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G. A.; Restaino, R.; and Wald, L. 2014. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5): 2565–2586.

Wang, H.; Xie, Q.; Zhao, Q.; and Meng, D. 2020. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3103–3112.

Wang, Y.; Deng, L.-J.; Zhang, T.-J.; and Wu, X. 2021. SSconv: Explicit spectral-to-spatial convolution for pansharpening. In *Proceedings of the 29th ACM International Conference on Multimedia*, 4472–4480.

Wang, Z.; Ma, Y.; and Zhang, Y. 2023. Review of pixel-level remote sensing image fusion based on deep learning. *Information Fusion*, 90: 36–58.

Xie, Q.; Zhou, M.; Zhao, Q.; Meng, D.; Zuo, W.; and Xu, Z. 2019. Multispectral and hyperspectral image fusion by MS/HS fusion net. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1585–1594.

Xu, S.; Zhang, J.; Zhao, Z.; Sun, K.; Liu, J.; and Zhang, C. 2021. Deep gradient projection networks for pansharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1366–1375.

Yan, K.; Zhou, M.; Zhang, L.; and Xie, C. 2022. Memory-Augmented Model-Driven Network for Pansharpening. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, 306–322. Springer.

Yang, D.; and Sun, J. 2018. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *Proceedings of the european conference on computer vision (ECCV)*, 702–717.

Yang, G.; Zhou, M.; Yan, K.; Liu, A.; Fu, X.; and Wang, F. 2022. Memory-augmented deep conditional unfolding network for pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1788–1797.

Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; and Paisley, J. 2017. PanNet: A deep network architecture for pansharpening. In *Proceedings of the IEEE international conference on computer vision*, 5449–5457.

Yuan, Q.; Wei, Y.; Meng, X.; Shen, H.; and Zhang, L. 2018. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3): 978–989.

Zhang, X.; Li, Z.; Change Loy, C.; and Lin, D. 2017. Polynet: A pursuit of structural diversity in very deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 718–726.

Zhou, M.; Huang, J.; Fang, Y.; Fu, X.; and Liu, A. 2022. Pan-sharpening with customized transformer and invertible neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 3553–3561.