Information Fusion 101 (2024) 102001

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Zero-shot semi-supervised learning for pansharpening

Qi Cao [a], Liang-Jian Deng [b,*], Wu Wang [b], Junming Hou [c], Gemine Vivone [d,e]

[a] Yingcai Honors College, University of Electronic Science and Technology of China, Xiyuan Road 2006, Chengdu, 611731, China
[b] School of Mathematical Sciences, University of Electronic Science and Technology of China, Xiyuan Road 2006, Chengdu, 611731, China
[c] State Key Laboratory of Millimetre Waves, School of Information Science and Engineering, Southeast University, No. 2, Southeast University Road, Nanjing, 211189, China
[d] Institute of Methodologies for Environmental Analysis, CNR-IMAA, Tito, 85050, Italy
[e] National Biodiversity Future Center, NBFC, Palermo, 90133, Italy

ARTICLE INFO

Keywords:
Pansharpening
Convolutional neural network
Zero-shot learning
Semi-supervised learning
Multispectral image fusion
Remote sensing

ABSTRACT

Pansharpening refers to fusing a low-resolution multispectral image (LRMS) and a high-resolution panchromatic (PAN) image to generate a high-resolution multispectral image (HRMS). Traditional pansharpening methods use a single pair of LRMS and PAN to generate HRMS at full resolution, but they fail to generate high-quality fused products due to the assumption of a (often inaccurate) linear relationship between the fused products. Convolutional neural network methods, *i.e.*, supervised and unsupervised learning approaches, can model any arbitrary non-linear relationship among data, but performing even worse than traditional methods when testing data are not consistent with training data. Moreover, supervised methods rely on simulating reduced resolution data for training causing information loss at full resolution. Unsupervised pansharpening suffers from distortion due to the lack of reference images and inaccuracy in the estimation of the degradation process. In this paper, we propose a zero-shot semi-supervised method for pansharpening (ZS-Pan), which only requires a single pair of PAN/LRMS images for training and testing networks combining both the pros of supervised and unsupervised methods. Facing with challenges of limited training data and no reference images, the ZS-Pan framework is built with a two-phase three-component model, *i.e.*, the reduced resolution supervised pre-training (RSP), the spatial degradation establishment (SDE), and the full resolution unsupervised generation (FUG) stages. Specifically, a special parameter initialization technique, a data augmentation strategy, and a non-linear degradation network are proposed to improve the representation ability of the network. In our experiments, we evaluate the performance of the proposed framework on different datasets using some state-of-the-art (SOTA) pansharpening approaches for comparison. Results show that our ZS-Pan outperforms these SOTA methods, both visually and quantitatively. The code is available at https://github.com/coder-qicao/ZS-Pan.

1. Introduction

Remote sensing images usually require high spatial resolution that is essential in many fields, *e.g.*, forecasting, agriculture, and environmental observation [1]. The enhancement of spatial and spectral resolutions of remote sensing products by improving satellite hardware [2–4] is a hard task when we have as constraint the preservation of the signal-to-noise ratio (SNR). As a result, many commercial sensors, including WorldView-3 (WV3) and WorldView-2 (WV2), produce two images with complementary features: a low-resolution multispectral (LRMS) image, retaining spectral information; and a high spatial resolution panchromatic (PAN) image, *i.e.*, a monochromatic data with a finer resolution than the MS counterpart. Pansharpening, which refers to the fusion of an MS and a PAN image, has as goal to build a high-resolution multispectral image (HRMS) combining the best features of the acquired LRMS/PAN pair, as illustrated in Fig. 1.

In recent years, many pansharpening algorithms [7–10] have been put forth to extract the spectral information from the MS image and the spatial information from the PAN image, and to produce an image that effectively combines them. They can be roughly categorized into four classes [11–13], *i.e.* (i) component substitution (CS) methods, (ii) multi-resolution analysis (MRA) techniques, (iii) variational optimization (VO) approaches, (iv) deep learning (DL) methods.

CS [14–16], MRA [17–19], and VO [20–22] approaches are three conventional pansharpening classes that are (often) heavily reliant on linear mathematical modeling and optimization. For instance, the

* Corresponding author.
E-mail addresses: 2020080601007@std.uestc.edu.cn (Q. Cao), liangjian.deng@uestc.edu.cn (L. Deng), wangwu@uestc.edu.cn (W. Wang), junming_hou@seu.edu.cn (J. Hou), gemine.vivone@imaa.cnr.it (G. Vivone).

https://doi.org/10.1016/j.inffus.2023.102001
Received 15 June 2023; Received in revised form 29 August 2023; Accepted 30 August 2023
Available online 4 September 2023

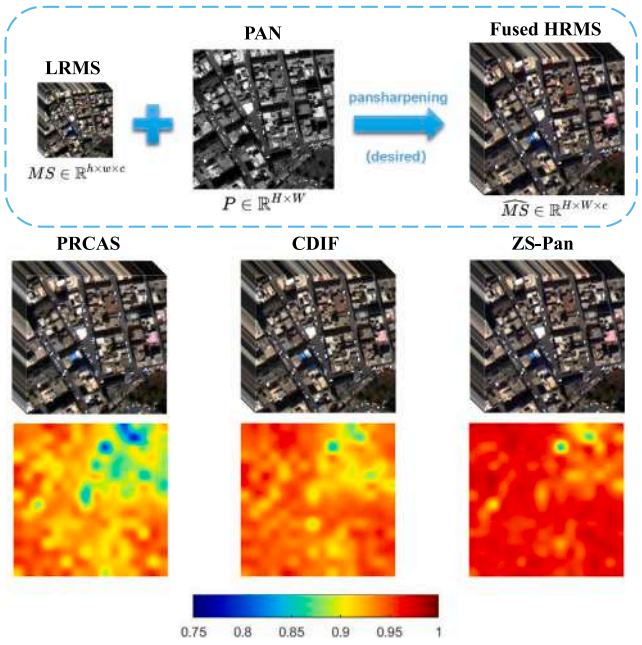


Fig. 1. The first row: flowchart for pansharpening on an 8-band WV3 satellite data. The second row: the pansharpened HRMS images provided by three pansharpening approaches, *i.e.* PRACS [5] (HQNR/Elaboration time = 0.9365/0.47 s), CDIF [6] (0.9509/118.22 s), and the proposed ZS-Pan (0.9666/134.26 s). The third row: the HNQR maps for the three methods.

CS-based Gram–Schmidt spectral sharpening method applies the GS orthogonalization to the MS image [23], the MRA-based generalized Laplacian pyramid method designs MS sensor's modulation transfer function (MTF) matched filters [24] to get high performance [25]. The VO-based methods view the pansharpening task as an ill-posed inverse optimization problem, including Bayesian methods [26,27], variational approaches [28,29], metaheuristics-based approaches [30,31], and compressed sensing techniques [32,33]. These approaches are based on a solid mathematical foundation, and have the benefit of using only a single pair of LRMS and PAN to generate HRMS. However, because of their (often not valid) assumption of linear relationship among LRMS, PAN, and HRMS [34], many of them exhibit spectral and spatial distortions implying reduced performance on their outcomes.

DL methods, such as, [35–37], leverage on the convolutional neural network's effective feature extraction capabilities to achieve high performance. DL is widely used in many computer vision fields, *i.e.,* image super-resolution [38,39], image segmentation [40,41], and image deraining [42,43]. Because there is no reference at full resolution, the way to deal with this problem can be mainly divided into two categories, *i.e.*, supervised learning and unsupervised learning methods. As for supervised learning methods, the training of networks relies on simulating data at reduced resolution, and the trained network at reduced resolution is used at full resolution for testing, thus generating the HRMS. However, the training at reduced resolution can distort the original features at full resolution, and, thus, supervised pansharpening usually performs worse in full resolution experiments than reduced resolution ones [44]. Unsupervised pansharpening [45,46] has recently drawn attention as a solution for improving full resolution performance. The training of unsupervised networks is done directly at full resolution, and the training is based on modeling the degradation process to compute the loss function. However, the (often linear) estimate of the degradation process exploited by these techniques is inaccurate [45,47]. Moreover, these approaches still need a lot of training data, and when the training data are not consistent with test data, their performance can be even worse than traditional methods. Zero-shot learning (introduced in computer vision [48,49]) can use the same image for both the training and testing phases. The zero-shot approaches can be trained relatively quickly because of the minimal size of the training data. Furthermore, because training and testing are performed on the same image, no additional simulated images are required and the training and testing data are totally consistent.

The proposed zero-shot semi-supervised learning for pansharpening (ZS-Pan) attempts to address the drawbacks of traditional and DL pansharpening. ZS-Pan exploits a single pair of LRMS and PAN images as input for the non-linear network exploring their original features. The challenges of applying zero-shot learning include limited training data and the lack of reference images. To deal with these challenges, the proposed ZS-Pan framework is built with three-dependent components, *i.e.*, the reduced resolution supervised pre-training (RSP), the spatial degradation establishment (SDE), and the full-resolution unsupervised generation (FUG) stages. The contributions of this work are as follows:

- We propose a zero-shot semi-supervised learning for the task of multispectral pansharpening (ZS-Pan). Any pansharpening network can use ZS-Pan as a plug-and-play module to be trained at full resolution and using a unique LRMS/PAN pair without requiring labeled data. As far as we know, this is the first attempt to apply the zero-shot semi-supervised learning strategy for pansharpening.
- A two-phase three-component semi-supervised model is designed to deal with the challenges of limited training data and no reference images. More specifically, in the RSP stage, a supervised training is conducted only on the available pair of LRMS and PAN images. In the SDE stage, an MS2PAN-Net is designed to learn the non-linear spatial degradation process. Finally, in the FUG stage, an unsupervised training is performed supported by the above-mentioned two stages to get the HRMS image.

The results show that the proposed ZS-Pan overcomes traditional state-of-the-art (SOTA) approaches[1] both qualitatively and quantitatively. Furthermore, we compare our method with some SOTA supervised and unsupervised pansharpening methods, demonstrating its advantages when small-scale training data are used. An ablation study is also carried out to reveal the crucial role of each component of ZS-Pan.

The paper is organized as follows. The related works and motivations will be introduced in Section 2. The challenges in applying zero-shot to pansharpening and how they are overcome will be detailed in Section 3. Section 4 will be devoted to the experimental results and the related discussions. Finally, conclusions will be drawn in Section 5.

2. Related works and motivations

The proposed ZS-Pan semi-supervised framework belongs to the DL class. Because there is no natural reference for pansharpening, the existing paradigms can be divided into two categories, *i.e.,* supervised and unsupervised. In this section, we will introduce first existing supervised and unsupervised learning strategies for pansharpening, and the zero-shot learning in other computer vision fields. Afterwards, we will point out the motivations under this work. The frameworks of the different pansharpening strategies are depicted in Fig. 2.

2.1. Supervised learning for pansharpening

Thanks to the powerful feature extraction ability of convolutional neural networks, the use of DL is recently a hot topic for pansharpening. The core idea of the supervised pansharpening is to train the network at reduced resolution using simulating data. The LRMS and PAN images

[1] For fair comparison, we mainly compare our method with traditional approaches, since they also only require a single pair of PAN/LRMS images as input without any large-scale training.

are downsampled to reach a reduced resolution and used as input. Instead, the original LRMS image is exploited as reference. Afterwards, the trained network at reduced resolution is used at full resolution for testing, thus generating the HRMS image. One of the first convolutional neural networks for pansharpening has been developed in [50], the so-called PNN. PNN applies a simple three-layer fully-convolutional model with rectified linear unit (ReLU) activations. In [36], Yang et al. designed a deep convolutional neural network (CNN), called PanNet. PanNet is divided into two parts, one is for the preservation of spatial details, while the other is to retain spectral information. To better extract this information, a deep residual network has been employed using four ResNet blocks [35] with skip connection to deepen the network depth. In [51], Deng et al. explored the combination of CNN methods and traditional fusion schemes, *i.e.*, CS and MRA, to address the task of pansharpening and create a new CNN-based architecture, called FusionNet.

However, two limitations exist in most of the supervised methods: (1) full resolution distortion, *i.e.*, due to the absence of HRMS images, the training of the network is conducted at reduced resolution, thus neglecting features at original (full) resolution; (2) large-scale dataset dependency, *i.e.*, the training of CNN-based networks requires a huge amount of data implying the use of high-performance equipments and long training times. Moreover, when training data are not consistent with testing data, these supervised networks can get lower performance.

2.2. Unsupervised learning for pansharpening

To overcome the shortcomings of supervised learning approaches, some pansharpening methods [45,52,53] based on unsupervised learning have been proposed. Unsupervised learning implies that the training no longer depends on simulating high-resolution MS images, but, instead, depends on the PAN and MS images themselves (indicating that the network can be trained at full resolution). The main problem, in this case, is that the loss function does not exploit any reference data and should be computed using the input LRMS and PAN data, and the fused HRMS cube. Thus, to work in an unsupervised way, the following questions have to be answered: (1) how to model the spectral relationship between the HRMS and LRMS images? (2) how to model the spatial relationship between the HRMS and PAN images? About the first question, the HRMS images are usually downsampled to the LRMS resolution using this latter for comparison. Instead, for the second question, linear models are often exploited. For instance, Ma et al. [45] considered the spectral average of the HRMS bands to be compared with the PAN image. In [52], Luo et al. combined the HRMS bands through a linear model with the related coefficients calculated by the minimization of the mean squared error between the down-sampled version of the PAN image and the LRMS cube. In [53], the spatial relationship has been modeled exploiting traditional methods with high spatial fidelity. Non-linear and multi-stage models are usually preferred by other unsupervised approaches, because they relax the linear assumption. For example, GTP-PNet [54] designed a TNet to establish a gradient connection between HRMS and PAN images. SUFNet [55] proposed a cross-scale learning and a two-stage comparison. Zhang et al. [56] also designed P2Net and STNet for non-linear spatial relation and multi-stage training in P2Sharpen. Although these methods use more reasonable non-linear relationships, they use complicated loss functions, *e.g.*, five loss functions in P2Sharpen and four loss functions in SUFNet, making the parameter adjustment process too hard and, thus, good results are difficult to be obtained.

Some limitations still exist in unsupervised pansharpening methods: (1) the absence of reference data leading to a hard design of proper loss functions and a high sensitivity of the approaches to the selected loss and the related hyperparameters, thus resulting in a complicated tuning phase; (2) the large-scale dataset dependency of the training, even requiring more computational resources than supervised learning because it is done at full resolution; (3) the inconsistency between training and testing data can lead to non-ideal pansharpened results, as for supervised techniques.

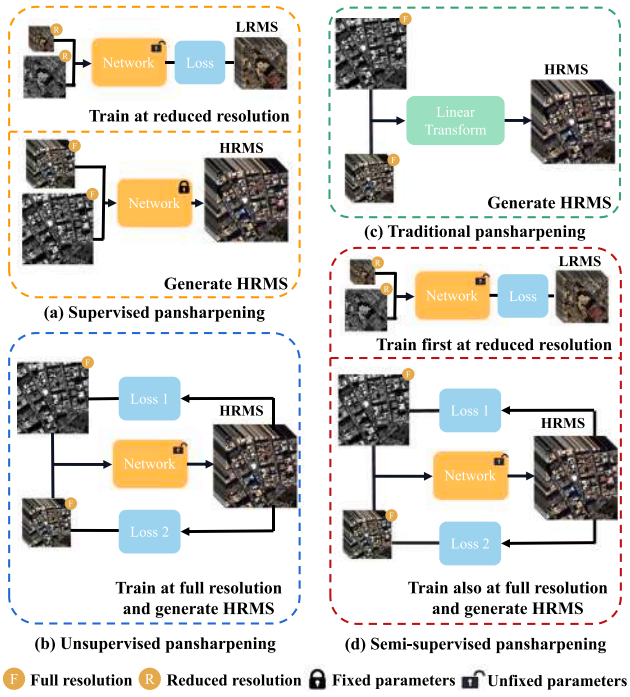


Fig. 2. The comparison of the different pansharpening paradigms: (a) supervised, (b) unsupervised, (c) traditional, and (d) semi-supervised pansharpening.

2.3. Zero-shot learning

Zero-shot learning relies upon a single image, where the training and testing are performed on the same image. The zero-shot approach does not require a huge amount of training data, thus the training is very light. The zero-shot learning method in the image processing field has been presented first for image super-resolution. In [48], Shocher et al. stated that the visual entropy inside a single image is much smaller than in a general external collection of images, thus training the network using a single image can be possible. Indeed, the core of zero-shot super-resolution is an implicit cross-scale patch matching approach using a lightweight network. Afterwards, zero-shot has been attempted many times for super-resolution and even in other research fields. For image restoration, Wang et al. proposed a zero-shot learning strategy using a denoising diffusion null-space model [57]. In the field of remote sensing and multispectral image sharpening, Nguyen et al. applied zero-shot learning for Sentinel-2 sharpening with a skipped connection CNN outperforming SOTA methods [58].

Anyway, there are still some difficulties in applying zero-shot learning to pansharpening: (1) how to build spatial and spectral relationships? There is no reference image for zero-shot learning, thus, the typical challenges in unsupervised learning also exist in the zero-shot task. (2) how to settle the problem of limited data? The input of zero-shot pansharpening is only a single LRMS/PAN pair. How to generate a high-quality HRMS image using a limited amount of data is another relevant issue.

2.4. Motivations

DL-based pansharpening has a large dataset dependency, but traditional pansharpening methods remind us that small data can also generate high-quality fused products. Although data-driven methods produce satisfactory results in some fields, the poor performance of DL pansharpening when training and testing data are not consistent makes us think how to optimize training data to avoid this problem. Model-based methods hardly overcome DL approaches because of their

Table 1
Comparison of the different pansharpening paradigms.

	Traditional	Supervised	Unsupervised	ZS-Pan
Non-linear transformation	×	✓	✓	✓
Reduced resolution training	×	✓	×	✓
Full resolution training	×	×	✓	✓
Consistent training and testing data	✓	×	×	✓
Small-scale dataset	✓	×	×	✓

assumption of linear relationship among LRMS, PAN, and HRMS, but they still generate excellent visual results, and they are more flexible because they do not need a huge amount of data and they do not (usually) require any simulation step. Thus, the use of small-scale datasets and non-linear models can be a good solution for pansharpening.

Supervised and unsupervised pansharpening both have their benefits and drawbacks. Supervised pansharpening cannot explore the information at full resolution, but it has a reference image to guide the training process. Unsupervised pansharpening explores full resolution features, but the lack of reference makes the training difficult. Thus, a combination of these two strategies can improve the representation ability of networks.

As a result, we compare the pros and cons of the different pansharpening paradigms, as shown in Table 1, and we propose a zero-shot semi-supervised framework for pansharpening. The input data for our ZS-Pan are a single LRMS/PAN pair. The original (full resolution) features can be explored by non-linear neural networks.

There are some benefits in applying zero-shot semi-supervised learning to pansharpening: (1) consistent training and testing data, *i.e.*, the training and testing data of the zero-shot pansharpening are both the single LRMS/PAN pair to be fused; (2) cross-scale training, *i.e.*, the training of our approach is performed both at reduced and at full resolution, which means that the reference at reduced resolution can help to improve the representation ability, even without neglecting the information at full (original) resolution.

3. The proposed method

This section is devoted to presenting the ZS-Pan framework and the techniques we use to deal with the zero-shot issues. Firstly, we will introduce the notation and the general fusion framework for DL-based pansharpening. Then, the two phases with three components of the ZS-Pan method will be detailed.

3.1. Notation and DL-based pansharpening

The notation used in this paper is presented first. LRMS images are defined as $\mathbf{MS} \in \mathbb{R}^{h\times w\times c}$, while PAN images are defined as $\mathbf{P} \in \mathbb{R}^{H\times W}$. c is the number of the LRMS spectral bands, h and w denote the height and width of the LRMS images, respectively, while H and W represent the height and width of the PAN images, respectively. The fused HRMS images are denoted as $\widehat{\mathbf{MS}} \in \mathbb{R}^{H\times W\times c}$. The downsampled LRMS and PAN images, which are called reduced resolution multispectral images (RRMS) and reduced resolution panchromatic images (RRPAN), are denoted as $\widetilde{\mathbf{MS}} \in \mathbb{R}^{h/r\times w/r\times c}$ and $\widetilde{\mathbf{P}} \in \mathbb{R}^{h\times w\times c}$, respectively. Instead, r stands for the resolution ratio between PAN and MS.

About the pansharpening methods based on deep learning, the core idea is to estimate a fusion function. The inputs of this function are $\mathbf{MS} \in \mathbb{R}^{h\times w\times c}$ and $\mathbf{P} \in \mathbb{R}^{H\times W}$, and the related output is $\widehat{\mathbf{MS}} \in \mathbb{R}^{H\times W\times c}$. The fusion equation is as follows:

$$\widehat{\mathbf{MS}} = \mathcal{N}(\mathbf{P}, \mathbf{MS}; \theta), \tag{1}$$

where $\mathcal{N}(\cdot)$ refers to the fusion function, and θ represents the parameters of the function to be estimated.

The training of pansharpening networks is an optimization process. In case of supervised pansharpening, the training process using simulated (reference) ground-truth (GT) data can be represented as:

$$\min_{\theta} D\left(\mathbf{GT} - \mathcal{N}(\mathbf{P}, \mathbf{MS}; \theta)\right), \tag{2}$$

where $D(\cdot)$ is a function to measure the distance (*e.g.*, ℓ_1 or ℓ_2 norms) between the outcome of the network, $\widehat{\mathbf{MS}}$, and the reference image, **GT**.

3.2. Overall framework

The ZS-Pan framework is a two-phase semi-supervised framework with three components, *i.e.*, RSP, SDE, and FUG. In the first phase, RSP and SDE are performed simultaneously. In the second phase, FUG is executed to generate HRMS. Subsequently, we will delve into a detailed discussion of the three components of ZS-Pan. The complete ZS-Pan approach is illustrated in Fig. 3.

3.3. RSP

As stated in Section 2.3, one of the issues for zero-shot learning is to generate a high-quality HRMS with a reduced number of data. To this aim, we propose a RSP strategy. In this strategy, we train the proposed zero-shot network (ZS-Net) first at reduced resolution using the LRMS as reference, see Fig. 4. Afterwards, we initialize the ZS-Net at full resolution with the parameters trained at reduced resolution.

In the following, we will present each step of RSP with the related equations. We downsample first (using MTF-matched filters) the original LRMS and PAN images to get RRMS and RRPAN, where the decimation rate is equal to the resolution ratio between PAN and MS:

$$\begin{aligned} \widetilde{\mathbf{MS}} &= \mathcal{MTF}(\mathbf{MS}), \\ \widetilde{\mathbf{P}} &= \mathcal{MTF}(\mathbf{P}), \end{aligned} \tag{3}$$

where $\mathcal{MTF}$ represents the MTF-matched filter function plus decimation. We perform data augmentation, *i.e.*, cropping the original images into five pieces and flipping them with mirror symmetric transformation, on RRMS, RRPAN, and LRMS, as follows:

$$\begin{aligned} &\widetilde{\mathbf{MS}}_1, \widetilde{\mathbf{MS}}_2, \dots, \widetilde{\mathbf{MS}}_n = \mathcal{DA}(\widetilde{\mathbf{MS}}), \\ &\widetilde{\mathbf{P}}_1, \widetilde{\mathbf{P}}_2, \dots, \widetilde{\mathbf{P}}_n = \mathcal{DA}(\widetilde{\mathbf{P}}), \\ &\mathbf{MS}_1, \mathbf{MS}_2, \dots, \mathbf{MS}_n = \mathcal{DA}(\mathbf{MS}), \end{aligned} \tag{4}$$

where $\mathcal{DA}$ denotes the data augmentation process, $\widetilde{\mathbf{MS}}_i$, $\widetilde{\mathbf{P}}_i$, and $\mathbf{MS}_i$ represent the ith augmented data, respectively, and n refers to the amount of the augmented data. Afterwards, we train ZS-Net with the augmented data optimizing the following function:

$$\min_{\theta_{ZS}} \|\mathbf{MS}_i - \mathcal{N}(\widetilde{\mathbf{P}}_i, \widetilde{\mathbf{MS}}_i; \theta_{ZS})\|_2, \tag{5}$$

where $\|\cdot\|_2$ is the ℓ_2 norm, and the weights of ZS-Net are indicated with θ_{ZS}. The trained weights (*i.e.*, θ_{ZS}) are saved for the final stage.

Supervised learning leverages on reference images. The training done in this way puts the network in a more favorable state for applying the zero-shot method. By initially conducting supervised training at reduced resolution, the network's representation pattern can be initialized with the aid of reference data. Subsequently, unsupervised training at full resolution serves as fine-tuning, enabling the network to be effective with original (full resolution) images. The superiority of this approach with respect to just conducting unsupervised learning is proved in Section 4.4.

Data augmentation is proposed for the reduced resolution supervised training in our ZS-Pan. It should be noted that the data augmentation strategy is just applied to the LRMS and PAN pair, ensuring that ZS-Pan remains consistent with the requirement of zero-shot learning. As ZS-Net is initially trained on the LRMS and PAN images at reduced

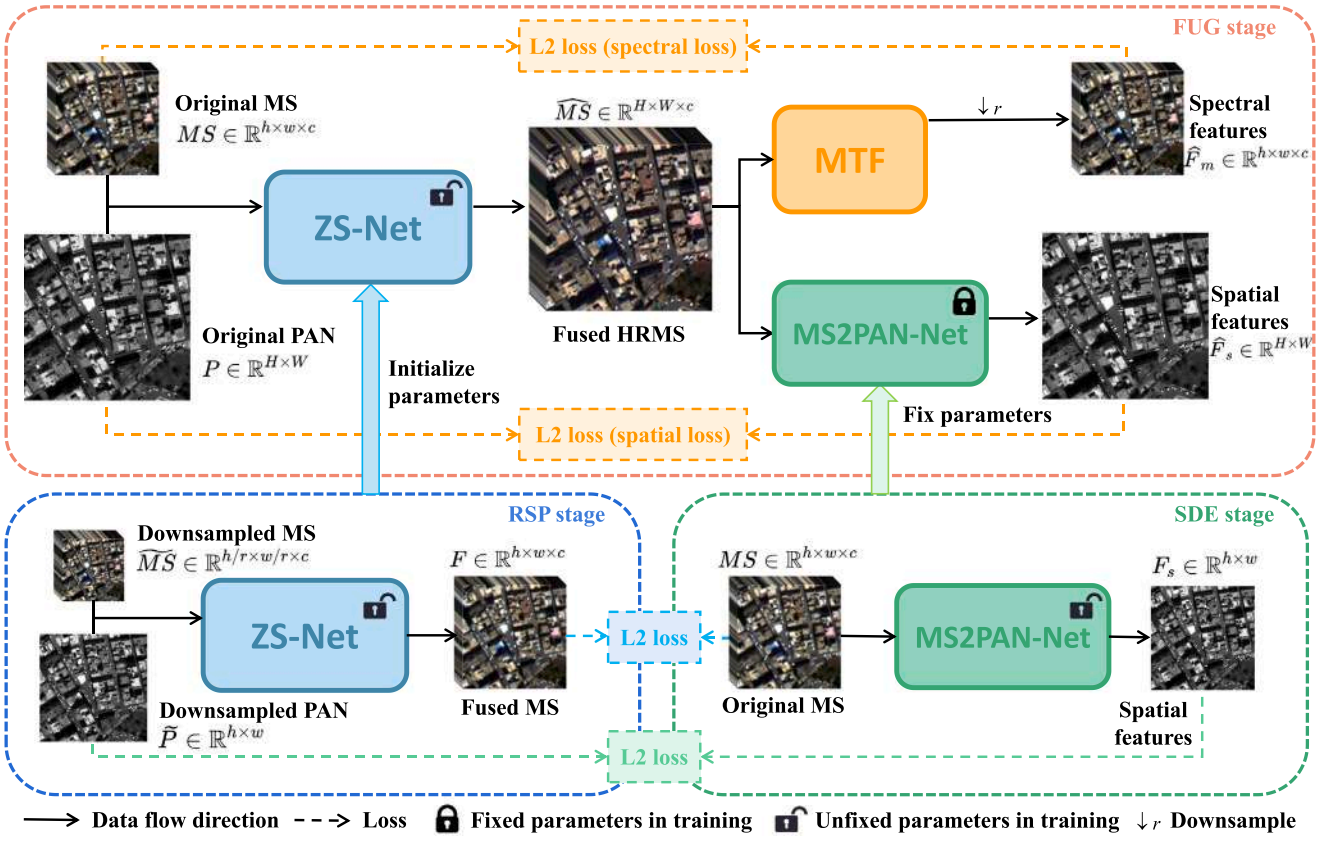


Fig. 3. The flowchart of the ZS-Pan framework. SDE and FUG are simultaneously performed first, and then the trained parameters of ZS-Net and MS2PAN-Net are transferred to the FUG stage. FUG generates the final (HRMS) pansharpened product. The "locked" and "unlocked" symbols indicate the fixed and learnable parameters during the training, respectively.

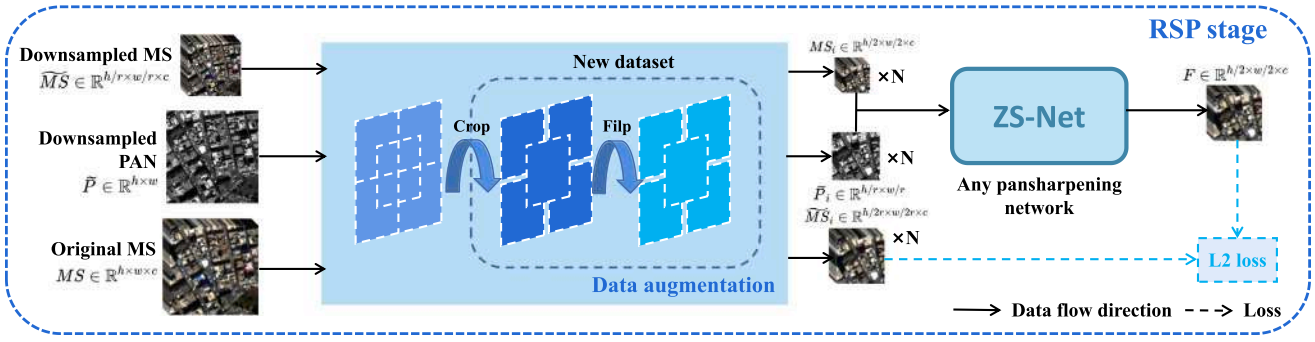


Fig. 4. The flowchart of the RSP stage. "$\times N$" is the multiplicative factor due to data augmentation.

resolution, applying this network to the original (full resolution) LRMS and PAN images helps prevent overfitting. Data augmentation serves to expand the dataset size, facilitating ZS-Net in learning a more generalized network rather than becoming biased towards specific patterns, thus mitigating overfitting. However, for the production of higher-quality HRMS images, data augmentation should only be applied during reduced resolution training and not at full resolution. We refer to this training phase as the RSP stage.

3.4. SDE

As stated in Section 2.3, the zero-shot learning has to address the problem of generating spatial and spectral degradation processes, which are used to compute the loss function. Thus, how to model the spectral relationship between HRMS and LRMS images and how to model the spatial relationship between HRMS and PAN images is a problem to consider during the design phase.

About the spatial relationship, we train a non-linear network (called MS2PAN) in the SDE stage, as shown in Fig. 5. We present first each step of the SDE with the related equations. Again, the PAN image is downsampled (using MTF-based filters) to get RRPAN. Thus, we have:

$$\widetilde{\mathbf{P}} = \mathcal{MTF}(\mathbf{P}). \tag{6}$$

Afterwards, we extract the spatial features of the LRMS exploiting the channel weighted sum block (CWSB) module. CWSB extracts spatial details by a weighted sum including all the LRMS channels, even considering sigmoid functions to account for non-linearity. Thus, we

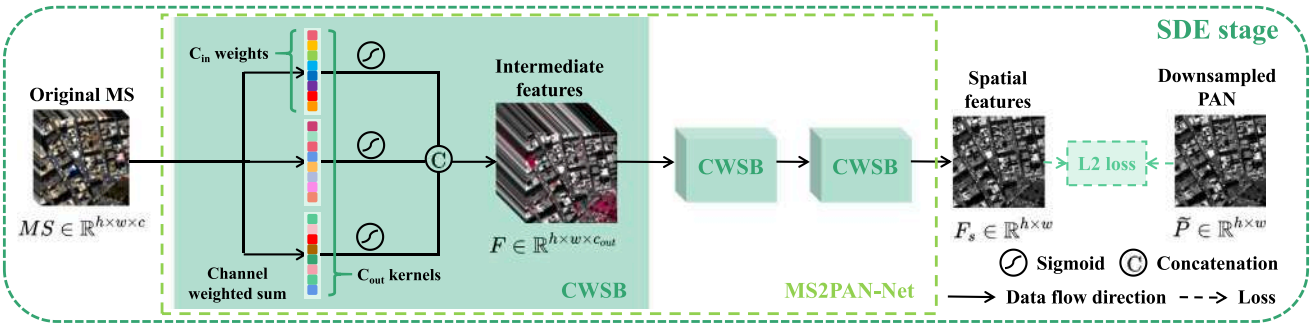


Fig. 5. The flowchart of SDE stage. c_{in} and c_{out} denote the number of channels before and after the CWSB, respectively.

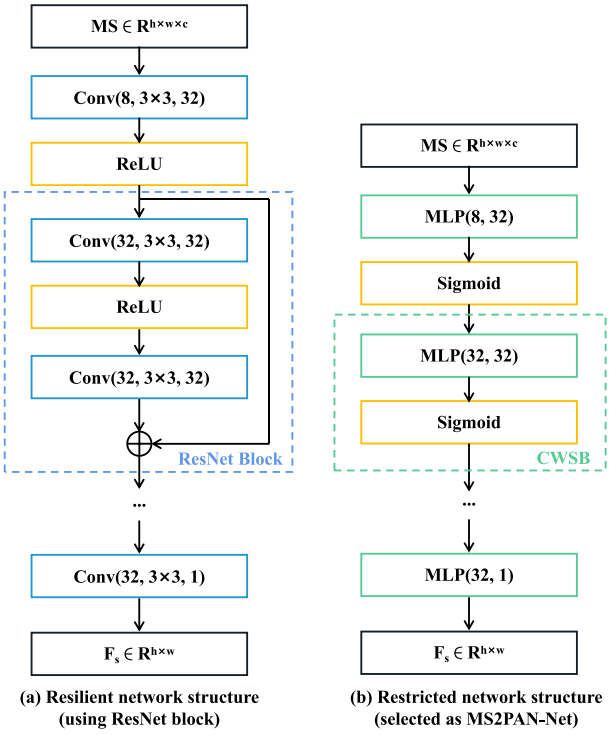


Fig. 6. Comparison of two kinds of network architectures discussed in the SDE stage: (a) resilient network (with more parameters); (b) restricted network (with fewer parameters).

have:

$$\begin{aligned}\mathbf{F}_1 &= \mathcal{CWSB}(\mathbf{MS}),\\ \mathbf{F}_2 &= \mathcal{CWSB}(\mathbf{F}_1),\\ &\cdots\\ \mathbf{F}_i &= \mathcal{CWSB}(\mathbf{F}_{i-1}),\end{aligned} \tag{7}$$

where $\mathcal{CWSB}$ denotes the CWSB module and $\widetilde{\mathbf{F}}_i$ is the ith extracted feature. The MS2PAN-Net consists of several CWSBs. LRMS images are fed into the MS2PAN-Net to extract the spatial features, *i.e.*:

$$\mathbf{F}_s = \mathcal{N}(\mathbf{MS}; \theta_{M2P}), \tag{8}$$

where $\mathbf{F}_s$ denotes the extracted spatial features and the weights of the MS2PAN-Net are indicated as θ_{M2P}. The cost function to be optimized for MS2PAN-Net is:

$$\min_{\theta_{M2P}} \|\widetilde{\mathbf{P}} - \mathbf{F}_s\|_2. \tag{9}$$

θ_{M2P} is saved to be used in the final stage.

As for the previous unsupervised pansharpening approaches, linear functions are commonly applied to HRMS images to model the HRMS-PAN relationship. However, a simple linear function might cause distortion during the extraction of the spatial details from the HRMS image [34], while a non-linear function might be more accurate in getting the spatial information. Therefore, we consider the training of the non-linear MS2PAN network.

It is worth to be noted that we train the MS2PAN-Net on LRMS images to make this network also applicable for HRMS images, thus overfitting should be avoided in this phase. To settle the problem of overfitting, the trade-off between a deep resilient and a restricted network should be considered. A deep resilient network (with more parameters) has a better ability to represent a general (non-linear) transformation, but overfitting can significantly reduce the performance of such networks, especially for the zero-shot task. Instead, a simple restricted network has a lower representation ability, but with a reduced overfitting phenomenon. The comparison of these two kinds of networks is shown in Fig. 6. CWSB modules are selected to build the restricted network. The experiments in Section 4.4 show better performance of this latter network with respect to more complex solutions for the zero-shot task. As described in Section 3.3, data augmentation has also been applied before the training to avoid overfitting. Considering that the MS2PAN network (MS2PAN-Net) is used to extract spatial features, this stage is called SDE.

3.5. FUG

The first two components (*i.e.*, RSP and SDE) are exploiting during the first phase working at reduced resolution. Instead, the FUG component is used during the second phase of the ZS-Pan framework. The network architecture for the ZS-Net can be any state-of-the-art DL-based pansharpening architecture, such as, FusionNet [51] or any other recent development in this field.

To train the ZS-Net in an unsupervised way, spatial and spectral losses should be defined. For the spatial loss, the MS2PAN-Net, which has already been trained in the SDE stage, can be applied. HRMS is fed into the MS2PAN-Net to extract spatial features:

$$\widehat{\mathbf{F}}_s = \mathcal{N}(\widehat{\mathbf{MS}}; \theta_{M2P}), \tag{10}$$

where the trained weights of MS2PAN-Net in the SDE stage are indicated with θ_{M2P} and $\widehat{\mathbf{F}}_s$ represents the extracted spatial features. The trained weights (θ_{M2P}) should not change in the FUG stage. The spatial loss is computed by comparing $\widehat{\mathbf{F}}_s$ with the PAN image:

$$\mathcal{L}_{spatial} = \|\mathbf{P} - \widehat{\mathbf{F}}_s\|_2, \tag{11}$$

As spectral loss, we adopt a widely-used ℓ_2 distance between the LRMS and the downsampled version of the HRMS image. Thus, we have:

$$\mathcal{L}_{spectral} = \|\mathbf{MS} - \mathcal{MTF}(\widehat{\mathbf{MS}})\|_2, \tag{12}$$

where $\mathcal{MTF}$ indicates the operation of MTF-based filtering plus decimation.

Hence, the ZS-Net can be trained in an unsupervised manner. The overall loss is simply a weighted sum of the spatial and spectral losses:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{spatial} + \lambda_2 \mathcal{L}_{spectral}, \tag{13}$$

where λ_1 and λ_2 are two weighting coefficients.

To train ZS-Net in a zero-shot way, it should be initialized using the weights obtained in the RSP stage. Thus, the weights (indicated with θ_{ZS}) are properly initialized as described above, and they can be fine-tuned during the FUG stage, driving the process with the full resolution loss in (13). Finally, the output of this procedure is the outcome of the proposed pansharpening approach.

3.6. MTF

In the proposed ZS-Pan, MTF-matched filters are applied to downsample the PAN and MS. MTF-matched filters are blur filters designed to match MS sensors' MTFs. They usually have a Gaussian-like shape, where the unique free parameter (the standard deviation) is properly set to get the matching. To this aim, gains at Nyquist frequency are exploited (because usually distributed by remote sensing sensors providers) to define these Gaussian filters obtaining the desired matching. The use of MTF-matched filters is a widespread practice in remote sensing pansharpening. The interested readers can refer to the related literature [24,59] for more details.

Overall, we design the ZS-Pan model with the aim of addressing challenges as mentioned in Section 2.3: (1) The RSP and FUG stages help to solve the "limited dataset" problem with cross-scale training and data augmentation. (2) The SDE and FUG stages support the solution to the "spatial/spectral relationship" problem with establishing a non-linear degradation relationship and applying MTF-matched filters.

4. Experimental results

In this section, we compare first the suggested strategy with some current SOTA pansharpening methods. The experiment settings will also be described. Afterwards, we will assess the performance at full resolution comparing our ZS-Pan with SOTA traditional methods (CS, MRA, VO) and some unsupervised and supervised DL-based methods to prove the strength of our approach using small-scale datasets. After that, we complete the performance assessment using reduced resolution datasets. Finally, ablation studies and further discussions will be provided to the readers.

4.1. Experiment settings

The settings for the experiments will be discussed in this section together with the selected datasets, the considered benchmark, the quality indexes, and the training parameter settings.

4.1.1. Datasets

The datasets were collected by the WorldView-2 (WV2) and WorldView-3 (WV3) sensors, two sensors that are widely used for comparison. Eight LRMS bands (red, green, blue, near-infrared 1, coastal, yellow, red edge, and near-infrared 2) and a high-resolution PAN channel are acquired by WV2. The spatial resolution ratio is equal to 4 since the PAN and LRMS images have a spatial resolution of 0.5 m and 2 m, respectively. The radiometric resolution is 11 bits. Instead, WV3 provides a different spatial resolution for PAN and LRMS sensors, that is 0.3 m and 1.2 m, respectively, retaining the other features of WV2.

All the used data (*i.e.*, the PanCollection dataset [60]) are publicly available. Details and related data can be found at.[2] At full resolution, for 8-bands (WV2 and WV3) data, we have a size of 512 × 512 and 128 × 128 × 8 for PAN and LRMS images, respectively, to get HRMS data with a size of 512 × 512 × 8. At reduced resolution, we have a size of 256 × 256 and 64 × 64 × 8 for PAN and LRMS images, respectively. These reduced resolution LRMS and PAN images are simultaneously blurred and downsampled according to Wald's protocol [61] starting from full resolution data.

4.1.2. Benchmark

For fair comparison, many SOTA approaches belonging to the CS, MRA, and VO classes are employed. The selection of the approaches in these classes is mainly based on the ranking in a recent review paper about the pansharpening full resolution assessment [62]. Moreover, we added some latest unsupervised and supervised DL-based techniques. More details can be found in Section 4.3.

4.1.3. Quality assessment

The used quality indexes at reduced resolution are: the spectral angle mapper (SAM) [63], the relative dimensionless global error in synthesis (ERGAS) [61], the spatial correlation coefficient (SCC) [64], the universal image quality index for 8-band images (Q8) [65], and the structural similarity index metric (SSIM) [66] averaged along the spectral bands. Optimal values for Q8, SSIM, and SCC are 1, whereas they are 0 for SAM and ERGAS.

For the full resolution assessment, the quality with no reference (QNR) [67] index is widely used [62]. However, because of its well-known inconsistencies [62], the hybrid QNR (HQNR) [62] index has been exploited in this work. The HQNR index consists of two distortion metrics, the spatial one, D_s, and the spectral one, D_λ. The HQNR has an ideal value of 1, instead, D_s and D_λ have an ideal value of 0.

4.1.4. Training platform and parameters setting

The proposed network is coded with Python 3.8.2 and Pytorch 1.11.0, and it is trained with an NVIDIA GPU GeForce RTX 3060. We use the ADAM optimizer, in which the betas and weight decay are set to (0.9, 0.999) and 0, respectively. Because of the peculiarities of the zero-shot learning, we set the batch size to 1. The learning rate is set to 0.0005. For the three stages, we minimize the loss function in (9) for 100 epochs, the loss function in (5) for 150 epochs, and the loss function in (13) for 50 epochs. Finally, FusionNet [51] is always used in our ZS-Net, if not explicitly stated otherwise.

4.2. Full resolution assessment

This section is devoted to the performance assessment at full resolution of the proposed framework.

4.2.1. Comparison with traditional methods

Traditional pansharpening techniques, which are quick and effective, represent an ideal benchmark because they do not require any reference for training and, thus, a single pair of LRMS and PAN images can produce the HRMS, consistently with our zero-shot training. A number of representative techniques are selected for our benchmark. Thus, we have:

EXP: MS image interpolated by a polynomial kernel with 23 coefficients [68].

CS methods:

- BT-H: Brovey transform with haze correction pansharpening approach [69].
- BDSD-PC: band-dependent spatial details with physical constraints pansharpening approach [16].
- C-GSA: Gram–Schmidt adaptive with clustering pansharpening approach [19].
- PRACS: partial replacement adaptive component substitution pansharpening approach [5].

[2] https://github.com/liangjiandeng/PanCollection

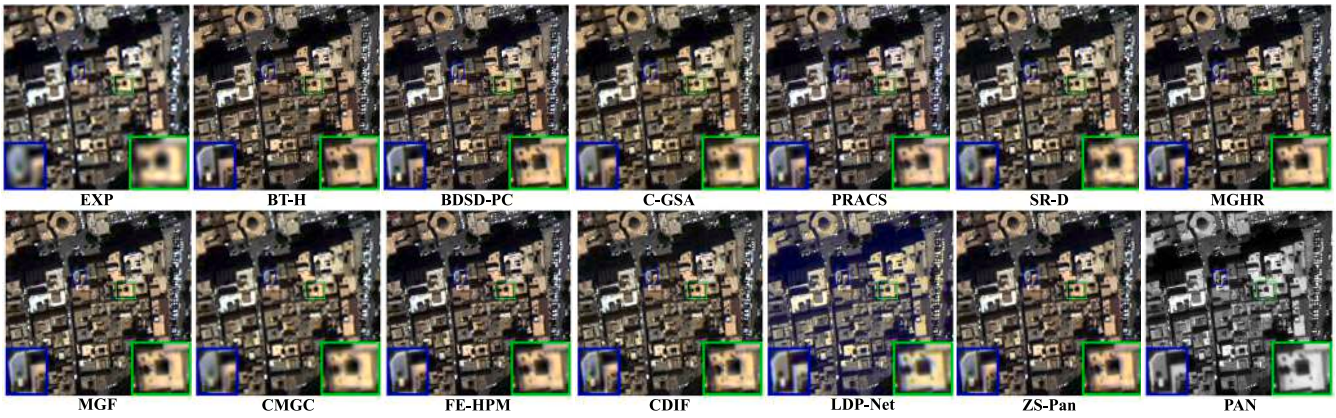


Fig. 7. Visual comparison in natural colors of the most representative 11 approaches on a full resolution WV3 example.

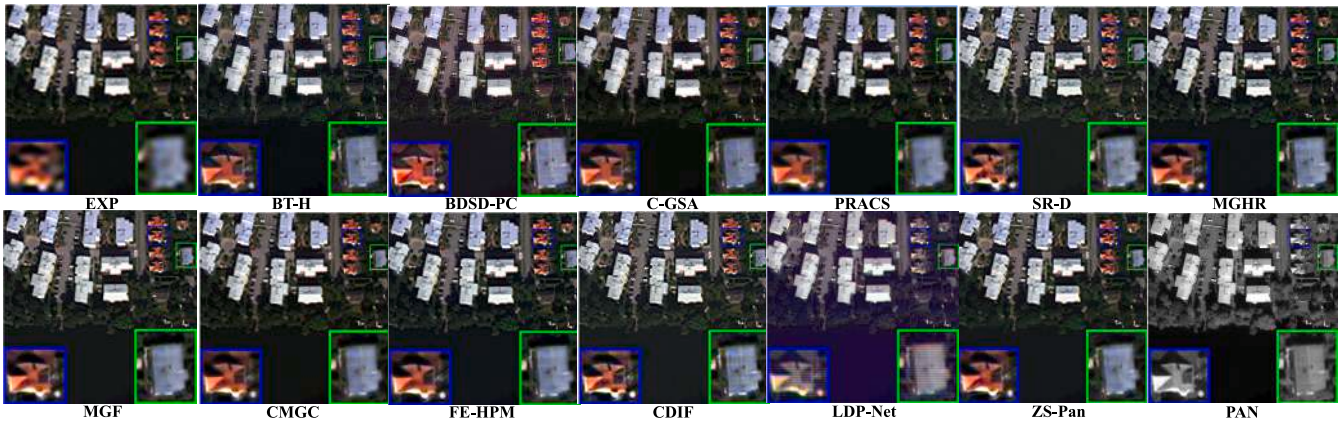


Fig. 8. Visual comparisons in natural colors of the most representative 11 approaches on a full resolution WV2 example.

MRA methods:

- MTF-GLP-HPM-R (MGHR): MTF-GLP-HPM [24,70] with regression-based spectral matching pansharpening approach [71].
- MTF-GLP-FS (MGF): MTF-GLP [24,68] with full scale regression-based injection model pansharpening approach [72].
- C-MTF-GLP-CBD (CMGC): MTF-GLP-CBD [24,68,72] with local parameter estimation exploiting clustering pansharpening approach [19]

.

VO methods:

- FE-HPM: filter estimation based on a semiblind deconvolution framework and HPM injection model pansharpening approach [73].
- SR-D: sparse representation of injected details pansharpening approach [33].
- CDIF: context-aware details injection fidelity for variational pansharpening approach [6].

Regarding to the WV3 dataset, Table 2 reports the quantitative evaluation for all the compared approaches. Our ZS-Pan is the best approach according to the overall HQNR index. Moreover, ZS-Pan has the best D_λ value suggesting a great ability of the proposed approach to be spectral consistent with respect to the original LRMS image. Furthermore, the standard deviation (std) of the HQNR for our approach gets the second-smallest values for all the indexes, thus demonstrating the robustness of the proposed ZS-Pan.

Table 2
Average quantitative comparisons on 20 full resolution WV3 examples.

Name	D_λ (± std)	D_s (± std)	HQNR (± std)
EXP	0.0401 ± 0.0102	0.0813 ± 0.0318	0.8821 ± 0.0374
BT-H	0.0561 ± 0.0228	0.0810 ± 0.0374	0.8682 ± 0.0540
BDSD-PC	0.0683 ± 0.0244	0.0730 ± 0.0356	0.8645 ± 0.0539
C-GSA	0.0472 ± 0.0200	0.0583 ± 0.0340	0.8979 ± 0.0490
PRACS	0.0449 ± 0.0152	0.0455 ± 0.0241	0.9119 ± 0.0365
MGF	0.0389 ± 0.0121	0.0630 ± 0.0284	0.9009 ± 0.0378
MGHR	0.0381 ± 0.0113	0.0630 ± 0.0289	0.9016 ± 0.0375
CMGC	0.0362 ± 0.0101	0.0287 ± 0.0145	0.9363 ± 0.0217
FE-HPM	0.0401 ± 0.0143	0.0661 ± 0.0328	0.8968 ± 0.0432
SR-D	0.0344 ± 0.0084	**0.0236 ± 0.0057**	0.9429 ± 0.0119
CDIF	0.0317 ± 0.0075	0.0305 ± 0.0152	0.9389 ± 0.0213
LDP-Net	0.1037 ± 0.0341	0.1055 ± 0.0434	0.8038 ± 0.0641
ZS-Pan	**0.0185 ± 0.0060**	0.0279 ± 0.0141	**0.9542 ± 0.0188**

Best results are in boldface. Second-best results are underlined.

Table 3 reports the quantitative results for the WV2 dataset. As for WV3 data, ZS-Pan is the best approach (showing the highest HQNR value). Again, the lowest value for the D_λ metric indicates that ZS-Pan has a high spectral fidelity.

The visual analysis further corroborates these numerical results. The visual performance on a full resolution WV3 dataset is shown in Fig. 7, with ZS-Pan exhibiting very high spectral and spatial fidelity. Instead, Fig. 8 depict the performance on WV2 data, corroborating the visual appearance and features of ZS-Pan shown in the WV3 test case.

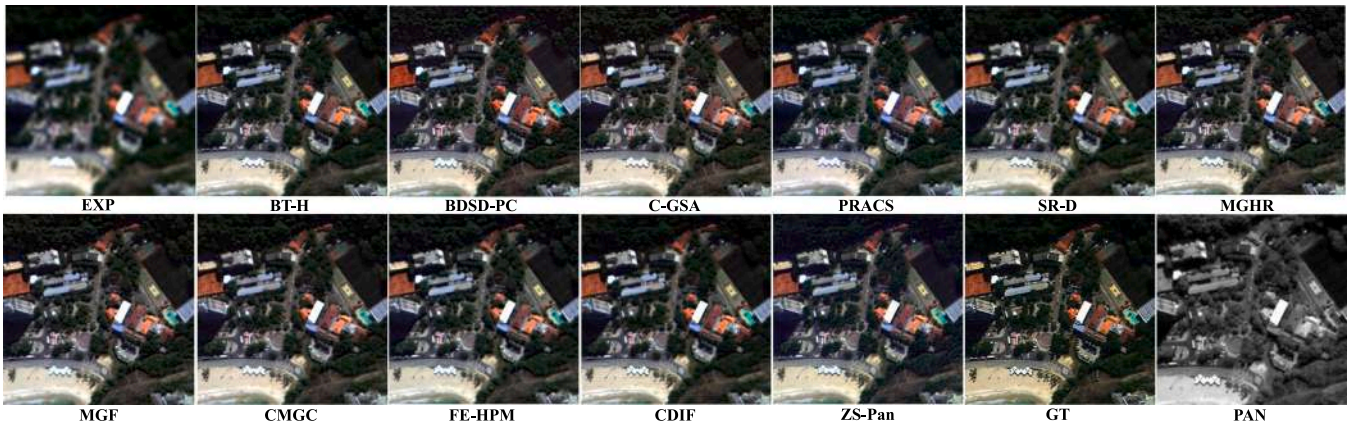


Fig. 9. Visual comparisons in natural colors of the most representative 10 approaches on a reduced resolution WV3 example.

Table 3
Average quantitative comparisons on 20 full resolution WV2 examples.

Name	D_λ (± std)	D_s (± std)	HQNR (± std)
EXP	0.0515 ± 0.0084	0.0599 ± 0.0127	0.8917 ± 0.0149
BT-H	0.0553 ± 0.0173	0.0858 ± 0.0164	0.8638 ± 0.0287
BDSD-PC	0.0944 ± 0.0204	0.0386 ± 0.0178	0.8708 ± 0.0316
C-GSA	0.0541 ± 0.0157	0.0796 ± 0.0125	0.8708 ± 0.0237
PRACS	0.0503 ± 0.0125	0.0324 ± 0.0085	0.9190 ± 0.0175
MGF	0.0507 ± 0.0220	0.0674 ± 0.0194	0.8855 ± 0.0343
MGHR	0.0436 ± 0.0074	0.0756 ± 0.0247	0.8843 ± 0.0285
CMGC	0.0437 ± 0.0074	0.0576 ± 0.0121	0.9013 ± 0.0169
FE-HPM	0.0650 ± 0.0445	0.0792 ± 0.0249	0.8617 ± 0.0599
SR-D	0.0443 ± 0.0079	**0.0263 ± 0.0077**	0.9305 ± 0.0125
CDIF	0.0378 ± 0.0053	0.0348 ± 0.0054	0.9287 ± 0.0090
LDP-Net	0.1311 ± 0.0812	0.0718 ± 0.0458	0.8090 ± 0.1049
ZS-Pan	**0.0285 ± 0.0152**	0.0386 ± 0.0155	**0.9341 ± 0.0260**

Best results are in boldface. Second-best results are underlined.

Table 4
Average quantitative comparisons on 20 full resolution WV3 examples.

Name	D_λ (± std)	D_s (± std)	HQNR (± std)
PNN	0.0399 ± 0.0116	0.0428 ± 0.0143	0.9192 ± 0.0240
PanNet	0.0395 ± 0.0119	0.0470 ± 0.0207	0.9156 ± 0.0303
MSDCNN	0.0407 ± 0.0120	0.0467 ± 0.0194	0.9147 ± 0.0293
BDPN	0.0472 ± 0.0160	0.0459 ± 0.0187	0.9093 ± 0.0321
DiCNN	0.0487 ± 0.0148	0.0462 ± 0.0171	0.9076 ± 0.0288
FusionNet	0.0424 ± 0.0121	0.0364 ± 0.0133	0.9228 ± 0.0215
LagNet	0.0482 ± 0.0184	0.0418 ± 0.0148	0.9122 ± 0.0279
ZS-Pan	**0.0185 ± 0.0060**	**0.0279 ± 0.0141**	**0.9542 ± 0.0188**

Best results are in boldface. Second-best results are underlined.

Table 5
Average results of ZS-Pan used with different supervised DL-based methods on 20 full resolution WV3 examples.

Name	D_λ (± std)	D_s (± std)	HQNR (± std)
w FusionNet	0.0185 ± 0.0060	0.0279 ± 0.0141	0.9542 ± 0.0188
w PanNet	0.0193 ± 0.0056	0.0303 ± 0.0156	0.9511 ± 0.0203
w PNN	0.0187 ± 0.0061	0.0180 ± 0.0101	0.9637 ± 0.0148

4.2.2. Comparison with unsupervised DL methods

Unsupervised pansharpening represents a family of SOTA solutions that can be trained at full resolution. Although some approaches consider large-scale datasets, we can train unsupervised methods with a single LRMS/PAN pair as done for ZS-Pan.

The models for unsupervised pansharpening can be roughly divided into two categories: (1) pansharpening based on GAN models [74], *i.e.*, PanGAN [45] and PGMAN [47]. (2) pansharpening based on proposing loss functions to model the degradation process, *i.e.*, LDP-Net [75] and Z-PNN [44]. Because of the characteristics of GAN models, input data should be large enough to train both the generator and the discriminator. As a result, GAN-based methods cannot be selected for our experiments. Instead, for the second category, LDP-Net[3] is exploited as benchmark.

The comparison with unsupervised approaches is reported in Tables 2, 3. It can be remarked that unsupervised pansharpening fails to generate high-quality pansharpened products when small-sized samples are used to train the network, resulting in worse outcomes with respect to the proposed ZS-Pan (see Fig. 7 and Fig. 8).

4.2.3. Comparison with supervised DL methods

To demonstrate the superiority of our method at full resolution, we chose some SOTA DL-based supervised methods for comparison. It is worth to be remarked that supervised methods are trained with a huge amount of reduced resolution images, instead, our method is trained with a single LRMS/PAN pair.

The selected approaches are as follows:

- PNN[4]: pansharpening via convolutional neural networks (CNNs) [50]
- PanNet[5]: CNN in the high-frequency domain for pansharpening [36]
- DiCNN[6]: CNN based on detail injection pansharpening method [76]
- MSDCNN[7]: CNN based on multi-scale and multi-depth pansharpening method [77]
- BDPN[8]: pansharpening method based on bidirectional networks [78]
- FusionNet[9]: deep CNN inspired by traditional CS and MRA methods [51]
- LagNet[10]: CNN panchromatic sharpening based on local content adaptation [79]

The results are shown in Table 4 and Fig. 11. Although supervised methods rely on a huge amount of training data, the quantitative results

[3] Code link: https://github.com/suifenglian/LDP-Net.

[4] Code link: http://openremotesensing.net/kb/codes/pansharpening/.
[5] Code link: https://xueyangfu.github.io/.
[6] Code link: http://openremotesensing.net/kb/codes/pansharpening/.
[7] Code link: https://github.com/liangjiandeng/DLPan-Toolbox.
[8] Code link: https://github.com/liangjiandeng/DLPan-Toolbox.
[9] Code link: https://github.com/liangjiandeng/FusionNet.
[10] Code link: https://github.com/liangjiandeng/LAGConv.

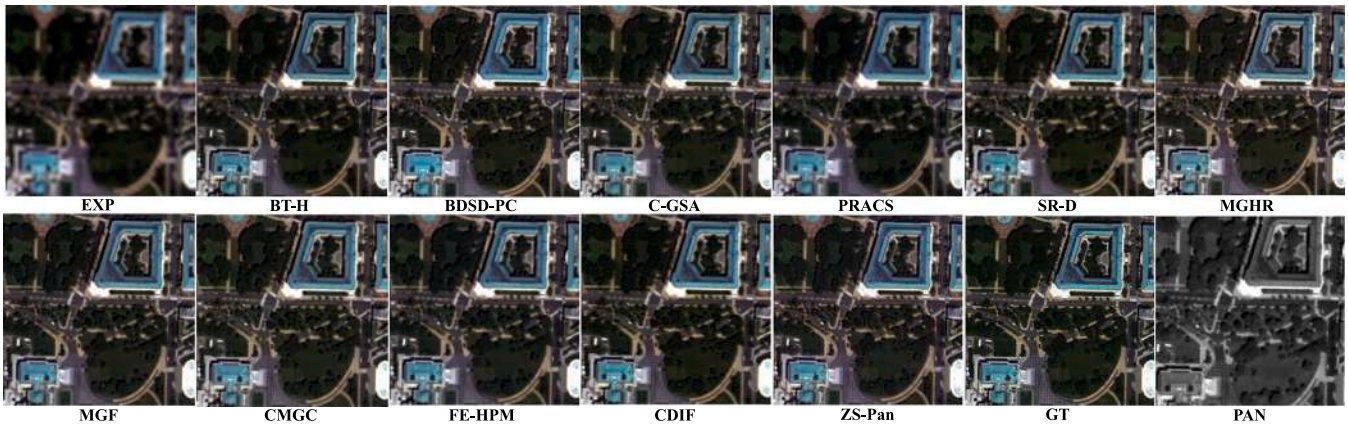


Fig. 10. Visual comparisons in natural colors of the most representative 10 approaches on a reduced resolution WV2 example.

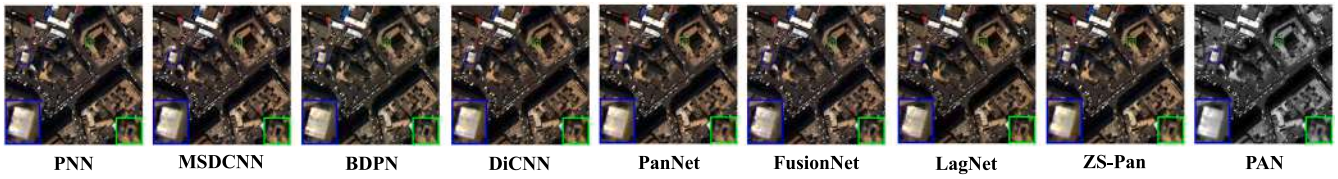


Fig. 11. Visual comparisons in natural colors of 7 DL-based approaches on a full resolution WV3 example.

Fig. 12. Visual results in natural colors of the comparison using different supervised DL-based solutions in the proposed framework and the performed ablation study.

are still worse than our ZS-Pan. These results show the importance of the consistency between training and testing data.

4.2.4. ZS-Pan with DL methods

As mentioned in Section 3.5, the proposed ZS-Net framework can include any pansharpening network to be trained at full resolution in a zero-shot learning manner. Thus, we chose some SOTA supervised pansharpening methods, *i.e.*, PNN [50], PanNet [36] and FusionNet [51], to analyze their performance when considered in our framework. Table 5 shows the related outcomes on WV3 data. High performance (HQNR values always greater than 0.92) can be remarked for all the configurations, demonstrating that our framework can be well-integrated with supervised DL-based techniques.

4.3. Reduced resolution assessment

This section is devoted to the performance assessment at reduced resolution, where the ground-truth (GT) image is used as reference. SOTA pansharpening methods from many categories are considered, including conventional techniques (CS, MRA, and VO techniques).

Table 6 reports the quantitative evaluation for WV3 data. According to the ERGAS and SCC indexes, our technique outperforms conventional pansharpening methods. Fig. 9 depicts the visual performance on a reduced resolution WV3 dataset, showing excellent spectral and spatial accuracies for ZS-Pan.

Table 7 reports the quantitative results on the WV2 dataset. The best Q8, ERGAS, and SCC values indicate that ZS-Pan can generate high-quality pansharpened products, even corroborated by Fig. 10.

Table 6
Average quantitative comparisons on 20 reduced resolution WV3 examples.

Name	Q8 (± std)	SAM (± std)	ERGAS (± std)	SCC (± std)	SSIM (± std)
EXP	0.6300 ± 0.0971	5.7534 ± 1.7829	7.1220 ± 1.7543	0.7434 ± 0.0268	0.7878 ± 0.0831
BT-H	**0.8337 ± 0.0992**	4.8734 ± 1.3442	4.5496 ± 1.4193	0.9253 ± 0.0230	**0.9232 ± 0.0217**
BDSD-PC	0.8277 ± 0.0943	5.4024 ± 1.7304	4.6766 ± 1.5393	0.9075 ± 0.0392	0.9173 ± 0.0268
C-GSA	0.8155 ± 0.0914	5.6706 ± 1.6653	4.8733 ± 1.4762	0.8925 ± 0.0391	0.9050 ± 0.0278
PRACS	0.7842 ± 0.1097	5.5403 ± 1.7815	5.3383 ± 1.5839	0.8791 ± 0.0539	0.8954 ± 0.0286
MGF	0.8286 ± 0.0998	5.2971 ± 1.6714	5.1301 ± 2.6770	0.8909 ± 0.1104	0.9135 ± 0.0264
MGHR	0.8254 ± 0.0904	5.2791 ± 1.6758	4.6776 ± 1.5189	0.9007 ± 0.0439	0.9153 ± 0.0262
CMGC	0.8177 ± 0.0880	5.5562 ± 1.5772	4.8331 ± 1.4962	0.8960 ± 0.0369	0.9080 ± 0.0253
FE-HPM	0.8277 ± 0.0945	5.1884 ± 1.5675	4.6430 ± 1.3013	0.9156 ± 0.0230	0.9107 ± 0.0242
SR-D	0.8262 ± 0.0960	4.9190 ± 1.3776	4.6397 ± 1.3794	0.9166 ± 0.0212	0.9064 ± 0.0246
CDIF	0.8322 ± 0.1032	**4.8548 ± 1.4788**	4.5029 ± 1.5338	0.9163 ± 0.0298	0.9187 ± 0.0242
ZS-Pan	0.8118 ± 0.1099	5.3000 ± 1.2026	**4.4397 ± 1.1382**	**0.9339 ± 0.0193**	0.9206 ± 0.0195

Best results are in boldface. Second-best results are underlined.

4.4. Ablation study

This section is devoted to some ablation studies to investigate the effect of each component of the ZS-Pan framework. For simplicity, we consider the WV3 dataset as reference. Table 8 reports the results of this study and Fig. 12 depicts the related visual results. It is clear that the proposed ZS-Pan shows the highest quantitative performance and the best visual effects.

4.4.1. The effect of RSP

To investigate whether the RSP stage contributes to the final result, we remove the RSP stage from the ZS-Pan framework. Table 8 presents the quantitative outcomes for ZS-Pan with and without RSP (w/o RSP). It can be observed that performance is strongly reduced by removing

Table 7
Average quantitative comparisons on 20 reduced resolution WV2 examples.

Name	Q8 (± std)	SAM (± std)	ERGAS (± std)	SCC (± std)	SSIM (± std)
EXP	0.6400 ± 0.0771	6.5295 ± 0.8808	6.7679 ± 0.7981	0.7340 ± 0.0210	0.7047 ± 0.0642
BT-H	0.8286 ± 0.1014	5.8910 ± 0.7457	4.3980 ± 0.5754	0.9173 ± 0.0097	0.8737 ± 0.0193
BDSD-PC	0.8431 ± 0.1040	6.1427 ± 0.8911	4.2525 ± 0.6993	0.9118 ± 0.0180	0.8809 ± 0.0170
C-GSA	0.8216 ± 0.1005	6.2450 ± 0.8733	4.5209 ± 0.6497	0.8920 ± 0.0203	0.8611 ± 0.0218
PRACS	0.7657 ± 0.0966	6.3163 ± 0.8683	5.3044 ± 0.7645	0.8587 ± 0.0242	0.8128 ± 0.0305
MGF	0.8242 ± 0.1001	6.3521 ± 0.9353	4.5549 ± 0.7608	0.8901 ± 0.0362	0.8594 ± 0.0216
MGHR	0.8251 ± 0.1007	6.1865 ± 0.9114	4.4545 ± 0.6859	0.8945 ± 0.0235	0.8645 ± 0.0208
CMGC	0.8224 ± 0.0999	6.2256 ± 0.8394	4.5236 ± 0.7200	0.8912 ± 0.0213	0.8604 ± 0.0222
FE-HPM	0.8294 ± 0.0998	6.0927 ± 0.8097	4.4228 ± 0.5679	0.9087 ± 0.0097	0.8657 ± 0.0227
SR-D	0.8250 ± 0.0999	5.8779 ± 0.7643	4.5226 ± 0.6219	0.9011 ± 0.0103	0.8570 ± 0.0202
CDIF	0.8410 ± 0.1036	**5.6297 ± 0.7127**	4.2655 ± 0.5800	0.9097 ± 0.0139	0.8759 ± 0.0172
ZS-Pan	**0.8441 ± 0.1054**	6.1298 ± 0.7838	**4.2345 ± 0.5186**	**0.9202 ± 0.0123**	**0.8852 ± 0.0160**

Best results are in boldface. Second-best results are underlined.

Table 8
Average results of the ablation study for our ZS-Pan framework on 20 full resolution WV3 examples.

Name	D_λ (± std)	D_s (± std)	HQNR (± std)
ZS-Pan	**0.0185 ± 0.0060**	**0.0279 ± 0.0141**	**0.9542 ± 0.0188**
w/o RSP	0.0413 ± 0.0145	0.0552 ± 0.0258	0.9060 ± 0.0341
w/o MS2PAN	0.0218 ± 0.0078	0.0415 ± 0.0218	0.9377 ± 0.0283
w/o MTF	0.0263 ± 0.0085	0.0622 ± 0.0298	0.9133 ± 0.0348
w/o DA	0.0219 ± 0.0076	0.0281 ± 0.0193	0.9508 ± 0.0251

Best results are in boldface. Second-best results are underlined.

Table 9
Average computational times for six traditional methods and our ZS-Pan. The size of the PAN is 512 × 512. The unit is seconds.

CS		MRA		VO		DL
GSA	PRACS	MGF	CMGC	SR-D	CDIF	ZS-Pan
1.73	0.47	0.27	1.93	3.60	118.22	134.26

the RSP module, thus demonstrating that it can help to improve the zero-shot performance in the FUG stage.

4.4.2. The effect of MS2PAN-net

As mentioned in Section 3.4, there are two ways to build MS2PAN-Net, see Fig. 6. To prove the validity of our choice, we replace the MS2PAN-Net with the method represented in Fig. 6. As shown in Table 8 and Fig. 12, ZS-Pan without MS2PAN-Net (w/o MS2PAN) has lower performance with respect to the original ZS-Pan with MS2PAN-Net, which verifies that a simple restricted network is a good solution to avoid overfitting, even boosting the performance.

4.4.3. The effect of MTF

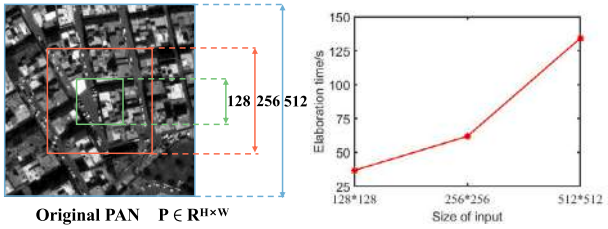
In this section, the role of MTF filters is analyzed. Indeed, we replace MTF filters with bicubic ones to investigate the validity of MTF-based (selected) filters. We denoted the ZS-Pan without MTF as w/o MTF. The quantitative results reported in Table 8 demonstrate that w/o MTF yields the second-worst performance, also corroborated by Fig. 12.

4.4.4. The effect of data augmentation

In RSP, data augmentation is exploited. To investigate the effect of this module, an ablation study without data augmentation in the RSP stage is performed (denoted as w/o DA). The quantitative results are reported in Table 8. The higher D_s value obtained by w/o DA demonstrates that data augmentation does not negatively affect the spatial detail quality. The lower HQNR value instead indicates that data augmentation is needed to get more accurate results.

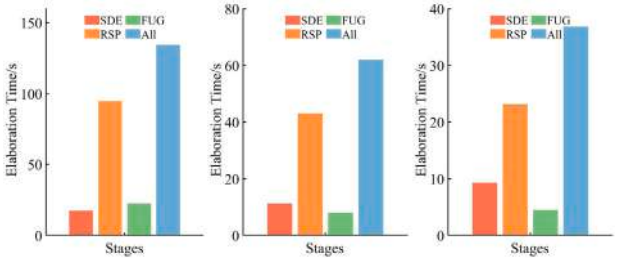
4.5. Discussions

Based on the previously shown results, it is clear that ZS-Pan obtains good pansharpened products. In this section, we will discuss about the training time of the ZS-Pan framework and the hyperparameters used in the loss function.



(a) Different input PAN size (b) Elaboration time with different PAN

Fig. 13. The processing time of ZS-Pan considering different PAN data sizes.



(a) PAN size of 512 (b) PAN size of 256 (c) PAN size of 128

Fig. 14. The computational load of the different parts of ZS-Pan.

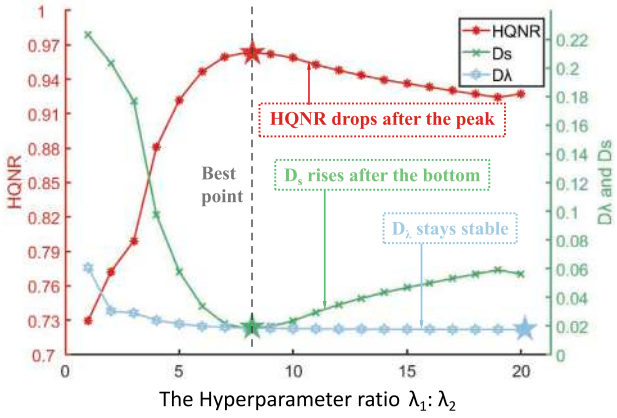


Fig. 15. The line chart of quality index and hyperparameters.

4.5.1. Training time

Fig. 13 reports the training times of ZS-Pan on 8-bands data (*i.e.*, WV3) varying the PAN size. The training time for 8-bands data with 512 × 512 pixels is 134.26 s, with 256 × 256 pixels is 61.87 s, and with 128 × 128 pixels is 36.76 s. The computation times are reported in Table 9 comparing the proposed approach with traditional methods. Our result (134.26 s) is slightly higher than that of the slowest VO method, *i.e.*, CDIF, which is 118.22 s. It is worth to be noted that DL-based pansharpening methods require hours to be trained, *e.g.*, 25 h for PNN [50], 6 h for TDNet [80], and 2 h for FusionNet [51], thus proving the efficiency of our method. Moreover, Fig. 14 shows the computational load of the different parts of our ZS-Pan. RSP requires more time to be trained because of the adopted data augmentation strategy.

4.5.2. Loss function hyperparameters

As described in Section 3.5, two hyperparameters (λ_1 and λ_2) weigh two sub-loss functions. Hence, the ratio between λ_1 and λ_2 determines the importance of one loss function with respect to the other. The

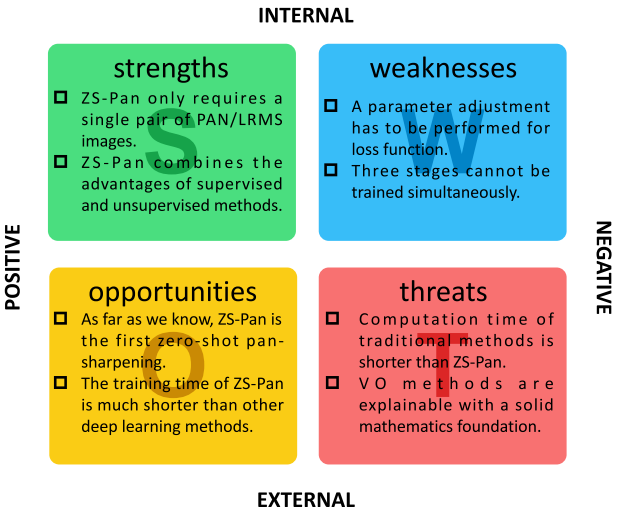


Fig. 16. SWOT analysis for our ZS-Pan.

higher the ratio is, the more important the spectral loss is, while the lower the ratio is, the more important the spatial loss is. Fig. 15 shows the changes in the D_s, D_λ, and HQNR indexes varying $\lambda_1 : \lambda_2$ on WV3 data. When the ratio is lower than 8, HQNR grows, while D_s and D_λ decrease. However, when the ratio is higher than 8, HQNR starts decreasing because of the spatial loss leads to increase D_s. Thus, we chose a ratio between the two λ coefficients equal to 8 for training our ZS-Pan.

4.5.3. Improvement analysis

In this section, we propose the analysis of the strengths, weaknesses, opportunities, and threats (SWOT) for our ZS-Pan. These latter are summed up in Fig. 16 aiding the readers in catching in a quick way the pros and cons of the proposed methodology.

5. Conclusions

In this paper, we investigated a new training strategy for DL-based pansharpening. In particular, we studied a two-phase three-stage model for zero-shot semi-supervised pansharpening (ZS-Pan), including the reduced resolution supervised pre-training (RSP), the spatial degradation establishment (SDE), and the full-resolution unsupervised generation (FUG) stages. Afterwards, the ZS-Pan framework has been assessed on real WV2 and WV3 data. ZS-Pan yielded the best quantitative and visual performance compared with many SOTA techniques. Ablation studies and further discussions demonstrated the high performance of the proposed approach for the pansharpening problem.

CRediT authorship contribution statement

Qi Cao: Methodology, Software, Writing – original draft. **Liang-Jian Deng:** Methodology, Supervision, Writing – review & editing. **Wu Wang:** Software. **Junming Hou:** Writing – review & editing. **Gemine Vivone:** Writing – review & editing.

Declaration of competing interest

This is No Conflict of Interest.

Data availability

The data that has been used is confidential.

Acknowledgments

This research is supported by National Natural Science Foundation of China (12271083), Natural Science Foundation of Sichuan Province (2022NSFSC0501), and National Key Research and Development Program of China (Grant No. 2020YFA0714001).

References

[1] Q. Xu, Y. Li, M. Zhang, W. Li, COCO-net: A dual-supervised network with unified ROI-loss for low-resolution ship detection from optical satellite image sequences, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–15.
[2] Y. Yang, H. Lu, S. Huang, W. Tu, Remote sensing image fusion based on fuzzy logic and salience measure, IEEE Geosci. Remote Sens. Lett. 17 (11) (2019) 1943–1947.
[3] Y. Yang, L. Wu, S. Huang, W. Wan, W. Tu, H. Lu, Multiband remote sensing image pansharpening based on dual-injection model, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 1888–1904.
[4] Y.-W. Zhuo, T.-J. Zhang, J.-F. Hu, H.-X. Dou, T.-Z. Huang, L.-J. Deng, A deep-shallow fusion network with multidetail extractor and spectral attention for hyperspectral pansharpening, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15 (2022) 7539–7555.
[5] J. Choi, K. Yu, Y. Kim, A new adaptive component-substitution-based satellite image fusion by using partial replacement, IEEE Trans. Geosci. Remote Sens. 49 (1) (2010) 295–309.
[6] J.-L. Xiao, T.-Z. Huang, L.-J. Deng, Z.-C. Wu, G. Vivone, A new context-aware details injection fidelity with adaptive coefficients estimation for variational pansharpening, IEEE Trans. Geosci. Remote Sens. (2022).
[7] C. Jin, L.-J. Deng, T.-Z. Huang, G. Vivone, Laplacian pyramid networks: A new approach for multispectral pansharpening, Inf. Fusion 78 (2022) 158–170.
[8] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J.-F. Hu, G. Vivone, VO+Net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–16.
[9] T. Xu, T.-Z. Huang, L.-J. Deng, N. Yokoya, An iterative regularization method based on tensor subspace representation for hyperspectral image super-resolution, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–16.
[10] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J. Huang, J. Chanussot, G. Vivone, LRTCFPan: Low-rank tensor completion based framework for pansharpening, IEEE Trans. Image Process. 32 (2023) 1640–1655.
[11] X. Meng, H. Shen, H. Li, L. Zhang, R. Fu, Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges, Inf. Fusion 46 (2019) 102–113.
[12] G. Vivone, M. Dalla Mura, A. Garzelli, R. Restaino, G. Scarpa, M.O. Ulfarsson, L. Alparone, J. Chanussot, A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods, IEEE Geosci. Remote Sens. Mag. 9 (1) (2020) 53–81.
[13] G. Vivone, M. Dalla Mura, A. Garzelli, F. Pacifici, A benchmarking protocol for pansharpening: Dataset, preprocessing, and quality assessment, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14 (2021) 6102–6118.
[14] P. Kwarteng, A. Chavez, Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis, Photogramm. Eng. Remote Sens. 55 (1) (1989) 339–348.
[15] A. Garzelli, F. Nencini, L. Capobianco, Optimal MMSE pan sharpening of very high resolution multispectral images, IEEE Trans. Geosci. Remote Sens. 46 (1) (2007) 228–236.
[16] G. Vivone, Robust band-dependent spatial-detail approaches for panchromatic sharpening, IEEE Trans. Geosci. Remote Sens. 57 (9) (2019) 6421–6433.
[17] R. Restaino, G. Vivone, M. Dalla Mura, J. Chanussot, Fusion of multispectral and panchromatic images based on morphological operators, IEEE Trans. Image Process. 25 (6) (2016) 2882–2895.
[18] G. Vivone, R. Restaino, J. Chanussot, Full scale regression-based injection coefficients for panchromatic sharpening, IEEE Trans. Image Process. 27 (7) (2018) 3418–3431.
[19] R. Restaino, M. Dalla Mura, G. Vivone, J. Chanussot, Context-adaptive pansharpening based on image segmentation, IEEE Trans. Geosci. Remote Sens. 55 (2) (2016) 753–766.
[20] X.X. Zhu, R. Bamler, A sparse image fusion algorithm with application to pan-sharpening, IEEE Trans. Geosci. Remote Sens. 51 (5) (2012) 2827–2836.
[21] L.-J. Deng, G. Vivone, W. Guo, M. Dalla Mura, J. Chanussot, A variational pansharpening approach based on reproducible kernel Hilbert space and heaviside function, IEEE Trans. Image Process. 27 (9) (2018) 4330–4344.
[22] L.-J. Deng, M. Feng, X.-C. Tai, The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior, Inf. Fusion 52 (2019) 76–89.
[23] C.A. Laben, B.V. Brower, Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening, in: Google Patents, US Patent 6,011,875, 2000.

[24] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, M. Selva, MTF-tailored multiscale fusion of high-resolution MS and pan imagery, Photogramm. Eng. Remote Sens. 72 (5) (2006) 591–596.

[25] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, in: Readings in Computer Vision, Elsevier, 1987, pp. 671–679.

[26] X. He, L. Condat, J.M. Bioucas-Dias, J. Chanussot, J. Xia, A new pansharpening method based on spatial and spectral sparsity priors, IEEE Trans. Image Process. 23 (9) (2014) 4160–4174.

[27] T. Wang, F. Fang, F. Li, G. Zhang, High-quality Bayesian pansharpening, IEEE Trans. Image Process. 28 (1) (2018) 227–239.

[28] L.-J. Deng, M. Feng, X.-C. Tai, The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior, Inf. Fusion 52 (2019) 76–89.

[29] C. Ballester, V. Caselles, L. Igual, J. Verdera, B. Rougé, A variational model for P+ XS image fusion, Int. J. Comput. Vis. 69 (1) (2006) 43.

[30] C.S. Yilmaz, V. Yilmaz, O. Gungor, J. Shan, Metaheuristic pansharpening based on symbiotic organisms search optimization, ISPRS J. Photogramm. Remote Sens. 158 (2019) 167–187.

[31] C.S. Yilmaz, V. Yilmaz, O. Gungor, A theoretical and practical survey of image fusion methods for multispectral pansharpening, Inf. Fusion 79 (2022) 1–43.

[32] S. Li, B. Yang, A new pan-sharpening method using a compressed sensing technique, IEEE Trans. Geosci. Remote Sens. 49 (2) (2010) 738–746.

[33] M.R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura, J. Chanussot, A pansharpening method based on the sparse representation of injected details, IEEE Geosci. Remote Sens. Lett. 12 (1) (2014) 180–184.

[34] R. Wen, L.-J. Deng, Z.-C. Wu, X. Wu, G. Vivone, A novel spatial fidelity with learnable nonlinear mapping for panchromatic sharpening, IEEE Trans. Geosci. Remote Sens. (2023).

[35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[36] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, J. Paisley, PanNet: A deep network architecture for pan-sharpening, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5449–5457.

[37] R. Ran, L.-J. Deng, T.-X. Jiang, J.-F. Hu, J. Chanussot, G. Vivone, GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution, IEEE Trans. Cybern. (2023) 1–14.

[38] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.

[39] T. Xu, T.-Z. Huang, L.-J. Deng, X.-L. Zhao, J. Huang, Hyperspectral image superresolution using unidirectional total variation with tucker decomposition, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 4381–4398.

[40] G. Lin, C. Shen, A. Van Den Hengel, I. Reid, Efficient piecewise training of deep structured models for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3194–3203.

[41] G. Lin, A. Milan, C. Shen, I. Reid, Refinenet: Multi-path refinement networks for high-resolution semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1925–1934.

[42] G. Wang, C. Sun, A. Sowmya, Erl-net: Entangled representation learning for single image de-raining, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 5644–5652.

[43] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, L.-J. Deng, Y. Wang, A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors, in: Proceedings of the Ieee Conference on Computer Vision and Pattern Recognition, 2017, pp. 4057–4066.

[44] M. Ciotola, S. Vitale, A. Mazza, G. Poggi, G. Scarpa, Pansharpening by convolutional neural networks in the full resolution framework, IEEE Trans. Geosci. Remote Sens. (2022) 1.

[45] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, J. Jiang, Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion, Inf. Fusion 62 (2020) 110–120.

[46] Q. Xu, Y. Li, J. Nie, Q. Liu, M. Guo, UPanGAN: Unsupervised pansharpening based on the spectral and spatial loss constrained generative adversarial network, Inf. Fusion 91 (2023) 31–46.

[47] H. Zhou, Q. Liu, Y. Wang, PGMAN: An unsupervised generative multiadversarial network for pansharpening, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14 (2021) 6316–6327.

[48] A. Shocher, N. Cohen, M. Irani, Zero-shot super-resolution using deep internal learning, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3118–3126.

[49] B. Li, Y. Gou, J.Z. Liu, H. Zhu, J.T. Zhou, X. Peng, Zero-shot image dehazing, IEEE Trans. Image Process. 29 (2020) 8457–8466.

[50] G. Masi, D. Cozzolino, L. Verdoliva, G. Scarpa, Pansharpening by convolutional neural networks, Remote Sens. 8 (7) (2016) 594.

[51] L.-J. Deng, G. Vivone, C. Jin, J. Chanussot, Detail injection-based deep convolutional neural networks for pansharpening, IEEE Trans. Geosci. Remote Sens. 59 (8) (2020) 6995–7010.

[52] S. Luo, S. Zhou, Y. Feng, J. Xie, Pansharpening via unsupervised convolutional neural networks, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 4295–4310.

[53] J. Gao, J. Li, X. Su, M. Jiang, Q. Yuan, Deep image interpolation: A unified unsupervised framework for pansharpening, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 609–618.

[54] H. Zhang, J. Ma, GTP-PNet: A residual learning network based on gradient transformation prior for pansharpening, ISPRS J. Photogramm. Remote Sens. 172 (2021) 223–239.

[55] Q. Liu, X. Meng, F. Shao, S. Li, Supervised-unsupervised combined deep convolutional neural networks for high-fidelity pansharpening, Inf. Fusion 89 (2023) 292–304.

[56] H. Zhang, H. Wang, X. Tian, J. Ma, P2Sharpen: A progressive pansharpening network with deep spectral transformation, Inf. Fusion 91 (2023) 103–122.

[57] Y. Wang, J. Yu, J. Zhang, Zero-shot image restoration using denoising diffusion null-space model, 2022, arXiv preprint arXiv:2212.00490.

[58] H.V. Nguyen, M.O. Ulfarsson, J.R. Sveinsson, J. Sigurdsson, Zero-shot sentinel-2 sharpening using a symmetric skipped connection convolutional neural network, in: IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2020, pp. 613–616.

[59] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G.A. Licciardi, R. Restaino, L. Wald, A critical comparison among pansharpening algorithms, IEEE Trans. Geosci. Remote Sens. 53 (5) (2014) 2565–2586.

[60] L.-J. Deng, G. Vivone, M.E. Paoletti, G. Scarpa, J. He, Y. Zhang, J. Chanussot, A. Plaza, Machine learning in pansharpening: A benchmark, from shallow to deep networks, IEEE Geosci. Remote Sens. Mag. 10 (3) (2022) 279–315.

[61] L. Wald, Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions, Presses des MINES, 2002.

[62] A. Arienzo, G. Vivone, A. Garzelli, L. Alparone, J. Chanussot, Full-resolution quality assessment of pansharpening: Theoretical and hands-on approaches, IEEE Geosci. Remote Sens. Mag. (2022).

[63] R.H. Yuhas, A.F. Goetz, J.W. Boardman, Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm, in: JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop, 1992.

[64] J. Zhou, D.L. Civco, J.A. Silander, A wavelet transform method to merge landsat TM and SPOT panchromatic data, Int. J. Remote Sens. 19 (4) (1998) 743–757.

[65] A. Garzelli, F. Nencini, Hypercomplex quality assessment of multi/hyperspectral images, IEEE Geosci. Remote Sens. Lett. 6 (4) (2009) 662–665.

[66] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[67] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, M. Selva, Multispectral and panchromatic data fusion assessment without reference, Photogramm. Eng. Remote Sens. 74 (2) (2008) 193–200.

[68] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis, IEEE Trans. Geosci. Remote Sens. 40 (10) (2002) 2300–2312.

[69] S. Lolli, L. Alparone, A. Garzelli, G. Vivone, Haze correction for contrast-based multispectral pansharpening, IEEE Geosci. Remote Sens. Lett. 14 (12) (2017) 2255–2259.

[70] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, J. Chanussot, Contrast and error-based fusion schemes for multispectral image pansharpening, IEEE Geosci. Remote Sens. Lett. 11 (5) (2013) 930–934.

[71] G. Vivone, R. Restaino, J. Chanussot, A regression-based high-pass modulation pansharpening approach, IEEE Trans. Geosci. Remote Sens. 56 (2) (2017) 984–996.

[72] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, L.M. Bruce, Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest, IEEE Trans. Geosci. Remote Sens. 45 (10) (2007) 3012–3021.

[73] G. Vivone, M. Simões, M. Dalla Mura, R. Restaino, J.M. Bioucas-Dias, G.A. Licciardi, J. Chanussot, Pansharpening based on semiblind deconvolution, IEEE Trans. Geosci. Remote Sens. 53 (4) (2014) 1997–2010.

[74] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.

[75] J. Ni, Z. Shao, Z. Zhang, M. Hou, J. Zhou, L. Fang, Y. Zhang, LDP-net: An unsupervised pansharpening network based on learnable degradation processes, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15 (2022) 5468–5479.

[76] L. He, Y. Rao, J. Li, J. Chanussot, A. Plaza, J. Zhu, B. Li, Pansharpening via detail injection based convolutional neural networks, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (4) (2019) 1188–1204.

[77] Q. Yuan, Y. Wei, X. Meng, H. Shen, L. Zhang, A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (3) (2018) 978–989.

[78] Y. Zhang, C. Liu, M. Sun, Y. Ou, Pan-sharpening using an efficient bidirectional pyramid network, IEEE Trans. Geosci. Remote Sens. 57 (8) (2019) 5549–5563.

[79] Z.-R. Jin, T.-J. Zhang, T.-X. Jiang, G. Vivone, L.-J. Deng, LAGConv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening, in: AAAI Conference on Artificial Intelligence, AAAI, 2022.

[80] T.-J. Zhang, L.-J. Deng, T.-Z. Huang, J. Chanussot, G. Vivone, A triple-double convolutional neural network for panchromatic sharpening, IEEE Trans. Neural Netw. Learn. Syst. (2022).