remote sensing

MDPI

Article

EBIFusion: An Enhancement Module Based on Error Backtracking for Remote Sensing Image Fusion

Jingjing Ma [1], Quanyi Dong [1], Qi Xue [1], Kai Wang [2], Zhangnan Li [2], Shuhe Zhao [1,3,*] and Fanghong Ye [4]

[1] Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China; 502023270135@smail.nju.edu.cn (J.M.); 502024270079@smail.nju.edu.cn (Q.D.); 502025270136@smail.nju.edu.cn (Q.X.)
[2] School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China; wangkai@smail.nju.edu.cn (K.W.); zhangnanli@nju.edu.cn (Z.L.)
[3] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
[4] Land Satellite Remote Sensing Application Center, Ministry of Natural Resources of China, Beijing 100048, China; yefh@lasac.cn
* Correspondence: zhaosh@nju.edu.cn

Highlights

What are the main findings?

- EBIFusion is a lightweight error backtracking module for remote sensing image fusion based on deep learning. The module improves the fusion quality of GF-6, increases QNR by 2.22%, and reduces spectral and spatial distortion.
- Experiments across datasets show that EBIFusion generally improves spatial detail and spectral fidelity across datasets.
- This module can be embedded into supervised and unsupervised networks, including GPPNN, PSGAN, PNN, and UCGAN.

What are the implications of the main findings?

- Prediction error reuse is a feasible strategy to improve the output image quality of the remote sensing image fusion model without redesigning the backbone model.
- EBIFusion provides a model-independent optimization component for improving the generation of high-resolution multispectral images.
- Further verification is still needed for computational efficiency, wider sensor conditions, and more diverse real-world scenarios.

Abstract

High-resolution multispectral images are of great value in various fields. However, the physical limitations of satellite sensors hinder the simultaneous acquisition of high spatial resolution and high spectral resolution images. Deep learning has become a powerful tool for remote sensing image fusion, but its full potential has not been fully utilized. In order to maximize the quality of the high-resolution multispectral images generated by the improved model, this paper proposes a module called EBIFusion, which introduces an error backtracking mechanism to improve fusion performance. The module uses the intermediate results of the deep learning model to capture the information lost during the high-resolution image generation process, thereby guiding the optimization of model training. The experimental results on the GF-6 dataset show that the QNR index is increased by 2.22% after the introduction of the module. In addition, the spatial and spectral quality has been continuously improved on multiple datasets, including QuickBird and GF-2.

Check for updates

Academic Editor: Andrea Garzelli

Received: 7 May 2026
Revised: 12 June 2026
Accepted: 29 June 2026
Published: 1 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

The optimized models, such as PSGAN, PNN, GPPNN, and UCGAN, show stronger spatial details and spectral fidelity. The main contribution of this paper is to propose an error backtracking framework for recovering and compensating for the lost information in the process of remote sensing image fusion based on deep learning. Based on this, a lightweight and model-independent enhancement module EBIFusion is designed, which can be integrated into different deep learning fusion architectures. At the same time, the generalization ability of the module in multiple datasets and network paradigms is verified. In summary, the error backtracking module enhances the quality of the generated high-resolution multispectral images. In addition, it is not limited to specific models and data and can be used as the basis for a versatile and effective optimization component to improve the availability of high-resolution multispectral images.

Keywords: image fusion; deep learning; remote sensing; error backtracking; high resolution

1. Introduction

In recent years, with the rapid development of remote sensing technology, high-resolution remote sensing images have played a significant role in fields such as land use monitoring [1], environmental monitoring and protection [2], urban and rural planning [3], and disaster early warning [4]. Compared with a multispectral low-resolution image and a panchromatic image, a high-resolution multispectral image shows more spatial details and texture features. It offers more comprehensive surface information and spectral characteristics, and is capable of achieving superior spectral color retention [5]. The use of high-resolution multispectral images obtained through image fusion technology is highly significant in fields such as geological exploration, environmental monitoring, precision agricultural analysis, national defense security, and others [6], bringing tangible economic and social benefits with a wide range of potential applications [7].

However, the low-cost acquisition of high-quality, high-resolution remote sensing image data is still a major challenge due to the limitations of the signal-to-noise ratio and the diffraction limit of piggyback satellites. When aiming to maintain an optimal signal-to-noise ratio, there is a fundamental discrepancy between the desired spectrum and the instantaneous field of view (IFOV) [8]. Remote sensing image fusion is a technology that fuses remote sensing image data with complementary features to obtain high-resolution remote sensing images. Currently, numerous satellites launched by countries worldwide are equipped with diverse imaging sensors, enabling the acquisition of low-resolution multispectral and panchromatic images simultaneously. Accordingly, the Pansharpening technology emerged. Its objective is to generate high-resolution multispectral images by using low-resolution multispectral images and high-resolution panchromatic images [9]. Panchromatic images offer a high level of detail in surface representation due to their high spatial resolution, whereas multispectral images provide comprehensive color information as a result of their high spectral resolution [10]. The spatial detail information and rich spectral feature information can be combined in the same image, overcoming the mutual restriction between spectrum and spatial resolution caused by the physical characteristics of satellite sensors to a certain extent. It provides an efficient and cost-effective approach for acquiring high-resolution remote sensing images [11].

At present, the full-color sharpening methods are mainly classified into four categories: component substitution (CS), multiresolution analysis (MRA), variational optimization (VO), and machine learning (ML) [12]. Among these taxonomic frameworks, deep learning-based methods have revolutionized the field of image generation in recent years, leveraging

their robust feature extraction capabilities and nonlinear modeling power to emerge as a prominent research focus. Deep learning-based remote sensing image fusion methods can process massive data and extract valuable information, achieving a balance between spectral and spatial information to some extent [13]. For instance, DR-NET employs the Swin Transformer model for downsampling processing to reduce detail loss during the downsampling process [14], while PanDiff injects mutually calibrated PAN and MS images through a modal intercalibration module (MIM) [15]. Therefore, this paper focuses on exploring deep learning-based remote sensing image fusion methods.

In accordance with the supervised paradigm, image fusion methods based on deep learning are classified into two categories: supervised and unsupervised schemes. The assumption of scale invariance of supervised methods may not be applicable in actual remote sensing data. This presents a challenge to reducing the generalization performance of resolution training models. In contrast, unsupervised methods face accuracy challenges between spatial degradation models that constrain spectral retention and spectral degradation models that constrain spatial structure retention [16]. Further research is needed to improve it. The two aforementioned schemes typically employ convolutional neural networks (CNN) and generative adversarial networks (GAN) as their primary network architectures.

Pan-sharpening neural networks (PNNs) represents the first instance of the introduction of CNN to the extraction and fusion of panchromatic and multispectral image information. Despite the slow convergence and poor generalization exhibited by this approach, it remains a classical and effective image fusion network [17]. Next, there was a rapid advancement in the field of image fusion based on convolutional neural networks. PanNet introduces residual learning and ResNet to mitigate spectral distortion [18]. However, the model is characterized by a high number of parameters and a complex structure. Subsequent research has gradually expanded the network structure, with notable contributions, including TFNet [19], MSDCNN [20], BDPN [21], and others. These approaches leverage multi-scale information to enhance spatial quality and nonlinear fitting ability. Nevertheless, the outcomes continue to exhibit shortcomings, such as local spatial structure smoothing, spectral distortion, and limited generalizability. However, the aforementioned methods do not consider the internal relationship between the input data and the fusion results. GPPNN explores the spatial and spectral degradation process and is used a priori to guide the optimization of the neural network [22]. The fusion performance is greatly improved. PSGAN introduces the generation countermeasure network for remote sensing image fusion for the first time. The GAN framework is employed to enhance fusion performance, with the fusion results approximating the target distribution to guarantee the quality of information retention [23]. Since then, the field of deep learning image fusion has seen a continuous exploration of the generation of confrontation networks. However, this approach does present certain challenges, as evidenced by MSGAN [24], DI-GAN [25], and other similar models. In some cases, these models have been observed to generate unrealistic image details. As a result, it is imperative to further optimize the ability and stability of the model. Drawing inspiration from the cyclic consistency observed in CycleGAN regarding image style migration, UCGAN emulates the spectral and spatial degradation models through a coarser upsampling and downsampling process, thereby achieving the cyclic consistency of the generator. However, this limits the fusion quality to some extent. The current deep learning model is constrained by its structural limitations, which impede its generalization capabilities. Consequently, the fused image exhibits spectral distortion in visual color and blurring of spatial texture and local details.

In order to further explore the model potential and optimize the generation of high-resolution multispectral images, we propose a generic deep learning image fusion model optimization module driven by error backtracking. The results obtained from deep learning

are used as intermediate variables, and the constructed error capture residual space and spectral information are used to compensate for the initial prediction results to obtain the final high-resolution image (i.e., error backtracking). It further improves the quality of the fusion result image by capturing the information lost when the model generates the high-resolution image. Under the premise of maintaining global consistency, the fused remote sensing images with high spatial and multispectral resolution are locally clearer and have higher color fidelity to meet the application requirements. Meanwhile, in order to more comprehensively study the impact of the error backtracking module on the performance of the fusion results, the existing evaluation methods are improved.

The main contributions of this paper are summarized as follows:

(1) A novel error backtracking framework is proposed for deep-learning-based remote sensing image fusion. By reusing the initial fusion result as an intermediate variable and performing secondary prediction, the framework captures and compensates for spatial–spectral information lost during the image generation process.
(2) Based on the proposed framework, a lightweight and model-agnostic enhancement module, termed EBIFusion, is developed. The module can be seamlessly integrated into both supervised and unsupervised fusion networks, improving fusion quality without modifying the original network architecture.

2. Methodology

2.1. Error Backtracking Module Ebifusion and Its Rationale

In deep learning, the task of generating high-resolution multispectral (i.e., HRMS) images typically involves extracting information from low-resolution multispectral (i.e., LRMS) and panchromatic (i.e., PAN) images. During the initial prediction phase, the model generates a simulated high-resolution multispectral image based on these inputs, which is regarded as a pseudo-ground truth. However, due to model limitations, this prediction may contain errors, such as insufficient spatial resolution or spectral information bias. To capture these errors, we employ the following approach: the predicted HRMS image is downsampled and re-input into the model alongside the original PAN image to generate a secondary prediction of the HRMS image. By processing these two images, we can compute an error matrix k that reflects both spatial and spectral discrepancies present in the initial prediction. Using the constructed error to capture the residual spatial and spectral information to compensate for the initial prediction results to obtain the final high-resolution image is called error backtracking (i.e., EB).

Given the considerable interference of the complex scale effect on the error mode of the model, it is not conducive to studying the errors in model predictions. The module differs from the previous residual module, which directly samples the original low-resolution multispectral image and then downsamples the construction error. In contrast, this module employs the generated high-resolution multispectral image predicted by the model for the first time to input the trained model for secondary prediction, thereby obtaining the secondary prediction high-resolution multispectral image to construct the error. The utilization of the available information allows for the effective exploitation of the error feedback mechanism and the intrinsic characteristics of deep learning. When the model is applied to similar data distributions, it is likely to produce comparable error patterns. Therefore, the learned error correction can be reused to improve prediction accuracy. In theory, this approach can be used to achieve accurate high-resolution multispectral image prediction.

Formula (1) is employed to extract the error. This simultaneously captures both the lost spatial and spectral information, and thus allows the error matrix k to be calculated. The error matrix k captures the lost information in space and contains multivariate information

that has been ignored in the spectral data. The error matrix k not only correlates with the discrepancy between the predicted and actual results but also encompasses information that the model is unable to discern. This information is extracted by the spectral–spatial error feature extraction block and plays a key role in the optimization process.

$$k = HRMS - HRMS' \ (k \in R) \tag{1}$$

where k is the error matrix, HRMS is the initially predicted high-resolution image, and HRMS′ represents the secondary prediction generated through downsampling of HRMS and PAN.

From a mathematical perspective, this process can be considered an iterative correction of model prediction errors, equivalent to performing an adaptive correction operation. This operation enables adjustments based on the model's specific performance in different spatial regions and spectral bands. Assuming the model's prediction error at a particular spatial location (x, y) and spectral band c can be expressed as $\varepsilon(x, y, c)$, the initial prediction HRMS can be viewed as the sum of the ground truth (i.e., GT) and an error field ε:

$$GT(x, y, c) = HRMS(x, y, c) + \varepsilon(x, y, c) \tag{2}$$

When we obtain the secondary prediction HRMS′ through downsampling and reprediction, this prediction is similarly affected by errors. However, due to the downsampling operation, these errors manifest differently. Therefore, the error matrix k essentially represents the difference between these two types of errors:

$$k(x, y, c) = \varepsilon_\{initial\}(x, y, c) - \varepsilon_\{secondary\}(x, y, c) \tag{3}$$

The error matrix k serves as a quantitative representation of prediction errors, encoding the spatial and spectral information omitted by the model during initial prediction. The error matrix k quantifies the spatial and spectral information missed by the model's initial prediction. After obtaining the error matrix, the initial prediction of HRMS is optimized through the error compensation block to generate a more accurate high-resolution image, to make the weights of over-compensated and noisy channels close to 0 and the weights of channels with truly missing information close to 1, thereby improving spectral fidelity and suppressing artifacts. The error compensation module receives the 4-band error feature k at once and completes the refinement in a single forward process. First, k is duplicated and simultaneously fed into two parallel branches—one is a 1×1 grouped convolution with 4 groups to capture cross-band spectral correlations, and the other is a 3×3 depth-wise separable convolution to encode local spatial discrepancies. The outputs of the two branches are element-wise summed to form a joint spatial–spectral representation. Subsequently, this representation undergoes global average pooling to compress each channel into a 4-dimensional vector. Each channel is globally average pooled into a 4-dimensional vector z through Formula (4) as follows:

$$z_c = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i, j) \tag{4}$$

which compresses z to one dimension with a 1×1 convolution (equivalent to FC(4→1)) followed by a ReLU activation, restores it to four dimensions with a second 1×1 convolution (equivalent to FC(1→4)), and applies a sigmoid to obtain channel-wise gating weights s given by

$$s = \sigma(W_2 \delta(W_1 z)) \tag{5}$$

where $W_1 \in R^{1\times4}$ and $W_2 \in R^{4\times1}$ are the weight matrices, $\delta(\cdot)$ denotes ReLU and $\sigma(\cdot)$ denotes Sigmoid. The weights are broadcast multiplied channel-by-channel to rescale F as follows:

$$\hat{F}_c(i,j) = s_c F_c(i,j),\ c = 1,\ldots,4 \tag{6}$$

After activation by Sigmoid, channel-wise gating weights are obtained, and channel-wise re-calibration is performed through broadcast multiplication to suppress over-compensation and highlight the band with the largest residual. The re-weighted features are mapped back to four channels by a final 1×1 convolution, stabilized in distribution by batch normalization, and introduced with nonlinearity by ReLU, enabling the network to learn both positive and negative residual corrections simultaneously. Stabilized by batch normalization and rectified by ReLU, they yield the residual compensation map. Although k does not exactly equal the true error field ε, after correction, it can continuously approach it to simulate the error. Theoretically, this method can eliminate systematic errors and mitigate their impacts to some extent, thereby improving output accuracy by reducing blurring effects and spectral distortion. Finally, it is added to the initial 4-band high-resolution multispectral image HRMS to obtain the refined result HRMS_EB.

However, the effectiveness of this approach relies heavily on the accuracy of the error estimation model and the consistency of the training data with real-world scenarios. In practice, residual errors may still persist due to insufficient modeling of complex noise patterns or mismatches between simulated and actual conditions. Additionally, the computational overhead introduced by the error compensation process must be carefully managed to ensure efficient deployment in resource-constrained environments.

In summary, the error backtracking module can be regarded as an optimal feedback compensation mechanism based on error compensation, which exhibits high adaptability, versatility, and strong generalization ability. As a prediction improvement module based on deep learning, it can be applied to different image data and provide relatively accurate predictions. The error backtracking module can be used to improve the accuracy of a trained deep learning model by leveraging the similarity between the model's prediction loss in similar situations. This approach allows the model's error to be applied to a specific case, thereby enhancing the accuracy of the prediction. The overall efficiency and accuracy of the model are consequently improved.

2.2. Deep Learning Image Fusion Model Based on the Error Backtracking Module

In order to assess the viability of the error backtracking module in a range of deep learning frameworks, this paper has selected four models, namely GPPNN, PSGAN, UCGAN, and PNN, which have been demonstrated to be stable and highly cited [9]. The generalization of the module is investigated through the implementation of supervised and unsupervised image fusion. The following provides a brief overview of the models.

A generative adversarial network is a network that has been trained by a generator and a discriminator through the application of game theory. PSGAN represents the inaugural GAN-based approach to generating a countermeasure network for remote sensing image fusion. It employs a confrontation training module to generate a high-resolution multispectral remote sensing image with a discriminator [23]. UCGAN represents an unsupervised generative countermeasure framework. It acquires knowledge from full-scale images devoid of ground truth. Firstly, the modal features of PAN and MS images are extracted by a double-flow generator, then fused in the feature domain, and subsequently, a full-color sharpened image is reconstructed. Moreover, a novel hybrid loss mechanism, based on cyclic consistency and antagonism, has been introduced to enhance performance [26]. The PNN is the inaugural remote sensing image fusion method based on a deep learning convolutional neural network (CNN). It remains a prominent image

fusion network, frequently utilized and regarded as a paragon of CNN deep learning image fusion [17]. GPPNN addresses two optimization issues through depth a priori regularization, which are responsible for the generation models of panchromatic images and low-resolution multispectral images, respectively. Subsequently, the gradient projection algorithm is employed to address these two issues, with the iterative step extended to encompass two network blocks. By alternately superimposing the two blocks, a novel network is constructed—a panchromatic sharpening neural network based on gradient projection [22].

The aforementioned models were analyzed using GF-6 as a data source. Table 1 illustrates the specific preprocessing, training, testing, and final image size for each model, tailored to its specific requirements. In regard to the supervisory network, the image is subjected to preprocessing in accordance with the Wald protocol [27]. The multispectral and panchromatic images are downsampled four times, resulting in 32 m and 8 m resolution images, respectively. Following downsampling, the original GF-6 image is reduced in size, resulting in a reduction in the amount of information contained in each image and, consequently, a limitation in the features that the model can receive.

Table 1. Model information table.

Model Name	Supervised Model		Unsupervised Model	
	CNN	GAN	CNN	GAN
	GPPNN	PSGAN	UCGAN	PNN
Train shape/resolution	(1536, 1536, 4) 32 m	(100, 100, 4) 32 m	(100, 100, 4) 8 m	(640, 64, 4) 8 m
Test shape/resolution	(1536, 1536, 4) 32 m	(100, 100, 4) 32 m	(100, 100, 4) 8 m	(100, 100, 4) 8 m
Generate shape/resolution	(6144, 6144, 4) 8 m	(400, 400, 4) 8 m	(400, 400, 4) 2 m	(400, 400, 4) 2 m

In Figure 1, first and foremost, the high-resolution prediction of the image is accomplished through the implementation of a deep learning generative adversarial network. Specifically, the panchromatic image (PAN) is replicated four times in the channel dimension, and the low-resolution multispectral image is sampled four times. The 4× downsampling factor is not arbitrarily chosen, but is determined by the intrinsic resolution ratio between the panchromatic (PAN) and multispectral (MS) images. Specifically, PAN images have a spatial resolution that is four times higher than that of MS images in the datasets used. Following the Wald protocol, the MS and PAN images are downsampled accordingly to simulate the degradation process while preserving their original resolution relationship. Therefore, the 4× factor ensures consistency with real imaging conditions and is widely adopted in pansharpening studies. Using larger factors would break this physical correspondence and lead to unrealistic degradation patterns, which may negatively affect the validity of the evaluation. This setting is consistent across all compared methods to ensure fair evaluation. The processed panchromatic image and multispectral image are mapped to the feature domain through the EmbNet module, thereby deriving embedding vectors of higher dimensions. Subsequently, FusionNet is employed to establish a connection between the two embedded vectors, thereby generating a fusion embedded vector. Ultimately, RstNet remaps the embedded vector back to the image domain, completes the training process, and outputs a high-resolution image, designated as HRMS, which is presumed to be a simulated real ground image. Then, HRMS is downsampled and reintroduced to the model, accompanied by the original panchromatic image. At this juncture, the secondary predicted high-resolution image, designated HRMS ′, is generated and utilized

to calculate and simulate the discrepancy between the actual ground image and HRMS'. Then, a 4-band error matrix k is obtained, which is used to quantify the spatial–spectral residuals missed by the initial high-resolution multispectral image (HRMS). Subsequently, the error compensation block optimizes it through a single forward pass: two parallel convolutions capture cross-band correlations and local spatial discrepancies, a lightweight gating network suppresses over-compensation while highlighting truly missing information, and the resulting residual correction is added to the initial HRMS to generate the final, spectrally faithful image HRMS_EB.

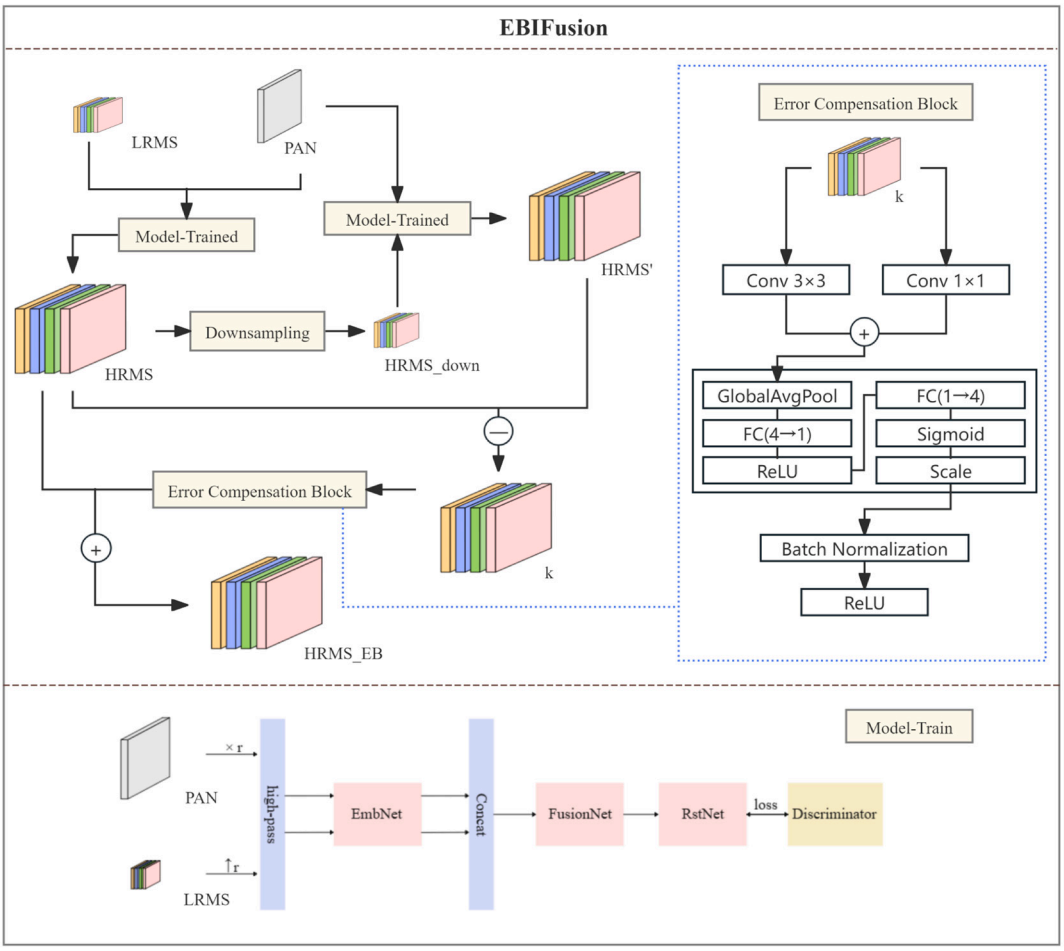


Figure 1. Model structure of EBIFusion.

2.3. Evaluation Index Framework

The evaluation index of remote sensing image fusion parameters can be classified into two categories: a non-reference and a full-reference index. The non-reference index does not rely on the full-resolution multispectral image as a reference, whereas the full-reference index employs the full-resolution multispectral image as a reference image to assess the quality of the fused image. This paper selects the effective index, which has been demonstrated through extensive practice to evaluate the accuracy of the fused image and is commonly utilized to assess the quality of high-resolution multispectral images. Detailed information about the indices is found in Table 2.

Table 2. Remote sensing image fusion parameter index.

Type	Name	Evaluate Features	Best Situation
No-Reference Metrics	D_λ	Spectral distortion, a measure of the difference in similarity between multispectral and panchromatic sharpened images [28].	0
	D_s	Space distortion, assess differences in spatial characteristics [28].	0
	QNR	Quality with No Reference, Comprehensive consideration of spectral and spatial distortion [29].	1
	SD	Standard Deviation, indicating the degree of dispersion of each band of the panchromatic sharpening image [30].	↑
	SF	Spatial Frequency, describe the richness of the details in the image [31].	↑
	FCC	Filtered Correlation Coefficients, to measure the correlation of features between multispectral images and panchromatic sharpened images [32].	1
Full-Reference Metrics	PSNR	peak signal-to-noise ratio, to reflect the distortion at the pixel level [33].	↑
	SSIM	structural similarity index measure, to reflect the structural similarity between the fused image and the source image [31].	1
	CC	correlation coefficient, to measure the linear correlation between the fused image and the source image.	1
	SAM	spectral angular mapper, optical spectrum quality reflecting the fusion result [34].	0
	ERGAS	Erreur Relative Globale Adimensionnelle de Synthese, global indicator for measuring the average error and dynamic range variation between the fusion result and the full-resolution multispectral image [35].	↓

↑ indicates that a higher value represents better performance, whereas ↓ indicates that a lower value represents better performance.

Among the nonreference evaluation indicators used, D_λ is used to evaluate the degree of spectral distortion introduced in the fusion process, D_s is used to measure the spatial information retention ability inherited from the panchromatic image, and QNR integrates spectral distortion and spatial distortion into the evaluation framework. Therefore, it is widely considered to be one of the most representative comprehensive indicators in pan-sharpening quality evaluation. SD and SF are used to characterize the image contrast and spatial detail richness, respectively. The larger the value is, the more texture information and edge structure the fused image retains. FCC is used to evaluate the consistency between the high-frequency spatial information of the fused image and the panchromatic image. However, since FCC only focuses on spatial correlation and cannot independently reflect the overall quality of the fused image, it should be combined with spectral indicators and

comprehensive quality indicators for comprehensive analysis. Therefore, the fused image quality should be evaluated from three complementary dimensions: spectral fidelity, spatial detail retention ability, and comprehensive fusion quality. In this paper, D_λ, D_s, and QNR are used as the core indicators to evaluate the fusion performance and verify fusion quality. SD, SF, and FCC are used as auxiliary indicators for observation, which are used to further analyze the characteristics of image contrast, texture richness, and spatial consistency.

2.4. Experimental Environment

The computational experiment in this paper is completed based on the heterogeneous computing architecture, and the hardware platform configuration is NVIDIA GeForce RTX 4090 The computing node is equipped with an AMD EPYC 7763 processor (64 cores and 128 threads) and 512 GB DDR4 Memory, through the PCIe 5.0 × 16 bus, realizes CPU-GPU heterogeneous communication with a bandwidth of up to 128 GB/s. Software environment adoption Ubuntu 22.04.3 LTS operating system (Linux kernel version 5.15.0-86), in conjunction with NVIDIA driver version 535.129.03 and the CUDA 12.2 toolkit, as well as the cuDNN 8.9.5 acceleration library, are used to optimize tensor operations. Selection of Programming frameworks Use Python 3.9.18 as the main control language, integrate the PyTorch 2.1.0 deep learning library and enable Torch.compile dynamic State diagram compilation features. The I/O processing of large-scale remote sensing images is implemented through the GDAL 3.6.4 library to achieve multi-threaded GeoTIFF solutions The code effectively supports the batch processing requirements of large pixel-level images.

3. Experiment Results

The image fusion process is conducted on UCGAN, an unsupervised generative adversarial network that has been demonstrated to be a sophisticated and effective approach. The high-resolution multispectral images generated prior to and following the implementation of the error backtracking module undergo visual and quantitative assessment.

3.1. Dataset and Environment Configuration

GF-6 is a multispectral high-resolution remote sensing satellite. It is mainly used for remote sensing data acquisition and application in the fields of agriculture, forestry, disaster prevention and mitigation, and others [36]. The panchromatic band resolution of GF-6 is 2 m, while the resolution of the multispectral band is 8 m. PanCollection is the latest open-source dataset and tool set dedicated to the study of panchromatic sharpening of remote sensing images, filling the gap in the field of remote sensing lacking standardized panchromatic sharpening datasets [37]. Therefore, in addition to the above GF-6 dataset, we conducted experiments on several datasets, including PanCollection. Specifically, we selected QuickBird and GF-2 in the open-source dataset Pancollection, as well as the GF-1 dataset, which are three important satellite datasets widely used in the remote sensing field. GF-1 has a panchromatic resolution of 2 m and a multispectral resolution of 8 m, while GF-2 has a panchromatic image resolution of 0.8 m and a multispectral image resolution of 4 m, covering four bands of blue, green, red, and near infrared.

QuickBird is a commercial remote sensing satellite with a panchromatic resolution of 0.61 m and a multispectral resolution of 2.44 m.

The model training described in this paper is conducted on a server equipped with an NVIDIA 4090 graphics card. The graphics card has 64 GB of graphics memory and is capable of supporting parallel computing. It is an appropriate method for training and testing batch processing of large remote sensing images. All models are executed within a Linux environment. The operating system utilized is Ubuntu 20.04 LTS, and Python 3.8 is employed as the principal programming language.

3.2. Performance of the Error Backtracking Module Under the Gf-6 Dataset

3.2.1. Results Description

This study compares the overall visual effects of high-resolution multispectral images generated before and after the application of an error backtracking module under different feature types. The objective is to observe the visual performance with the naked eye at a more detailed level, including aspects such as color, spatial detail, and texture. All datasets are evaluated independently without applying weighted aggregation. This avoids bias caused by differences in dataset characteristics and ensures a fair comparison across datasets. The same experimental settings and evaluation metrics are consistently applied across all datasets. The results are shown in Figure 2.

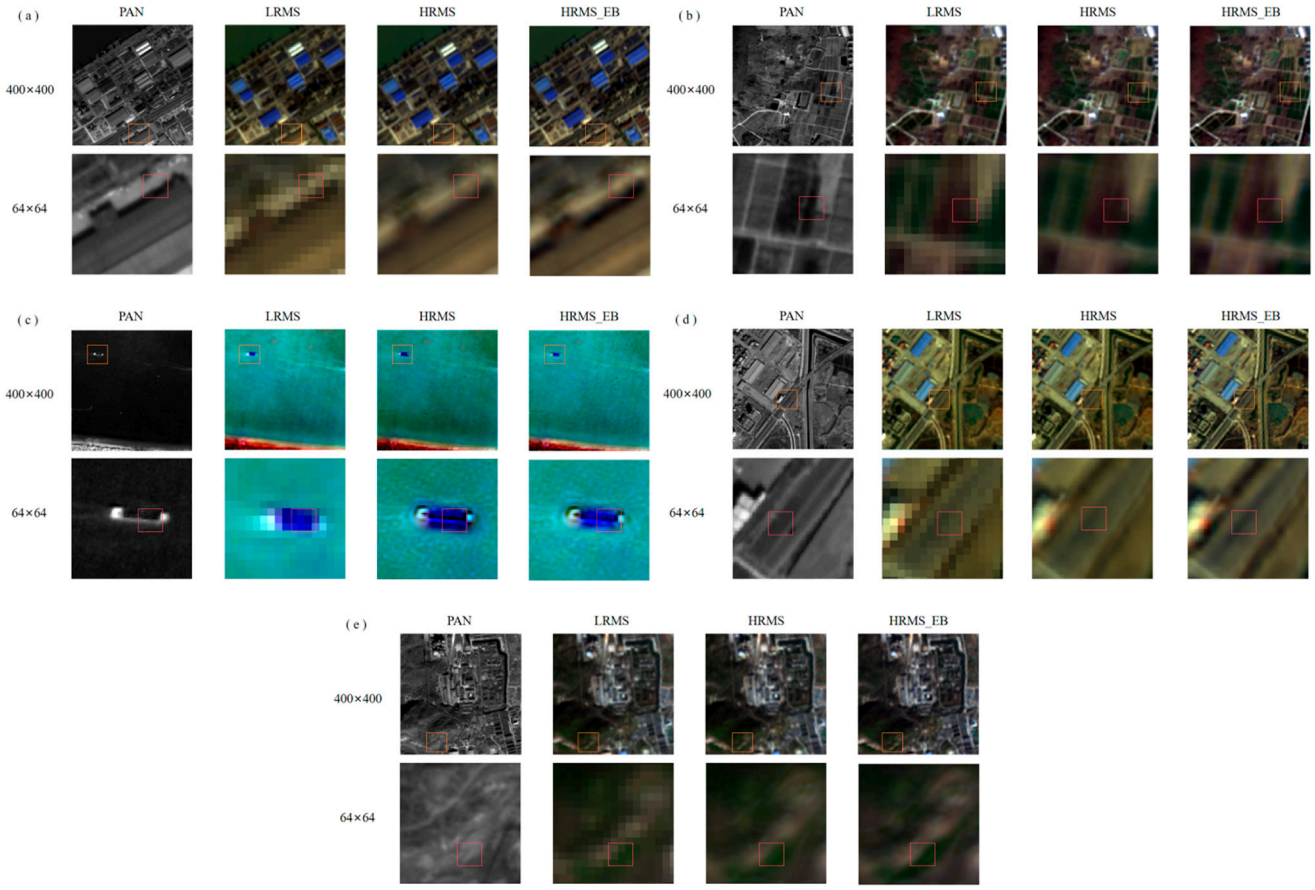


Figure 2. The original panchromatic image, the low-resolution multispectral image, the original fusion image, and the image generated by the error backtracking module are compared horizontally under the background of five kinds of terrestrial features: (**a**) a map of urban and building areas, (**b**) a map of farmland and cultivated land, (**c**) a map of water bodies and ships, (**d**) a map of traffic and roads, and (**e**) a map of forest and woodland. The size of the first column of images in each image is 400 × 400, and the second column of images is an enlarged image of the yellow box area in the first column of images, with a size of 64 × 64.

In Figure 2a, from the visual effect for urban and building areas, it can be seen that the original fusion image HRMS and the image HRMS_EB generated by us are generally relatively natural and coordinated, exhibiting no abrupt sense or obvious splicing trace. However, the image HRMS_EB generated by our method exhibits a color profile more closely aligned with the low-resolution image LRMS. For urban and building areas, under the pixel window size of 400 × 400, from a spatial point of view, the two fused high-resolution images can clearly distinguish the larger texture and outline, but the image we generate can better retain the details of the city and the building. At the resolution of

64 × 64, the spatial information of the panchromatic image is the clearest, followed by our generated image. When we observe the red box area, the spatial color of the HRMS image is blurred, while the HRMS_EB still maintains the clarity that can be distinguished by the naked eye. Through visual inspection, we can distinguish buildings, building shadows, and roads; the processing of mixed pixels is more scientific, and the distribution of light and shadow is more reasonable. For farmland and cultivated land, the original fusion result image is blurred under the size of 400 × 400, while our image is sharper, the color of farmland and cultivated land is natural and coordinated in color consistency, and there is no obvious color difference or color mismatch, but when enlarged to the size of 64 × 64, the fusion image we produce has better color performance, clearer edges and less distortion paste. In the image, the texture features of farmland and cultivated land are preserved, the boundary is clearly discernible, and there is no blurred or jagged edge, especially at the junction of land evenness and crop distribution, and the shape of the field is more accurate.

For water bodies and ships, the water body of the fused image looks clear or translucent, and can reflect the surrounding scenery, without any obvious sense of separation when observed at a scale of 400 × 400. The images are both capable of reflecting the surrounding scenery in a manner that does not create an obvious sense of separation. However, the ripples and direction of the water flow in the generated image HRMS_EB are more closely aligned with the low-resolution multispectral LRMS image in terms of color. Furthermore, the spatial texture of the ship's moving wake is more consistent with that of the panchromatic image, PAN. Furthermore, the transition between the coastline and the water body is more harmonious than in the original fused image. Upon enlarging the image to a size of 64 × 64, deficiencies in color balance are observed in the original fusion result image. The chromatic composition of the water body and the ship at the junction is conspicuously incongruous, exhibiting abrupt and unnatural hues and jagged textures.

For transportation and roads, it can be observed that the original fusion result image exhibits a modest degree of noise and distortion when viewed at a scale below 400 × 400. The yellow box should be enlarged to a size of 64 × 64, and the red box area should be observed. It is evident that the texture of the road in our image is more complete, the flatness is higher, the color and texture of the road are relatively uniform, and the road is clearly discernible. However, the original fusion result image results in a blurring of the road boundary and discoloration of features in the mixed pixel unmixing of the road boundary and the road.

For forest and woodland, from the overall visual effect, the color and tone of the fused image are consistent with the LR color of the low-resolution image, which is dark green and blends with the surrounding features, roads, and buildings. There is no obvious sense of fragmentation, and the colors and textures inside the forest and woodland are relatively uniform, with no obvious abruptness or anomalies, but our generated image is intuitively more coherent in the forest area. Under the size of 64 × 64, the original image of the dense and sparse vegetation areas is relatively blurred, and compared with the panchromatic image PAN, the boundary is more discernible in our generated image.

3.2.2. Index Analysis Results

As illustrated in Table 3, the D_λ value of HRMS_EB is less than that of HRMS, exhibiting a reduction of approximately 33.2170%. This indicates that HRMS_EB demonstrates superior spectral retention properties. The D_s value of HRMS_EB is 0.00253 less than that of HRMS, indicating that HRMS_EB is more similar to panchromatic images in spatial features such as edges and textures. In comparison to HRMS, HRMS_EB has demonstrated a notable enhancement of 0.021 in the QNR index, indicating an elevated signal-to-noise ratio and superior quality. The SD value of HRMS_EB is greater than that of HRMS, with a

relative increase of 5.7069%. The SF value of HRMS_EB is higher than that of HRMS, which has been optimized by 21.4852%. Consequently, the image exhibits greater detail. The FCC value of HRMS_EB is less than that of HRMS, with a reduction of 0.0821. The FCC value of the fusion image exhibits some differences from the original image with respect to specific features. However, this does not necessarily indicate a decline in overall image quality.

Table 3. Evaluation results of reference-free indicators of high-resolution multispectral images generated by UCGAN before and after the error backtracking module.

	D_λ	D_s	QNR	SD	SF	FCC
HRMS	0.0076	0.0651	0.9278	159.2350	29.6780	0.5677
HRMS_EB	0.0051	0.0468	0.9485	168.3224	36.0544	0.4856
Optimization Percentages	33.2170%	28.1533%	2.2239%	5.7069%	21.4852%	\

3.3. Performance of the Error Backtracking Module on Different Datasets

3.3.1. Results Description

In Figure 3, the high-resolution multispectral image HRMS_EB generated using the error backtracking module is significantly closer in color to the original compared to HRMS. The color representation of various ground object types, such as shadows, water bodies, and buildings, is more natural, greatly preserving the spectral information of the multispectral image. At the same time, HRMS_EB retains more details of the PAN image, making the overall details clearer, especially in the processing of object boundaries and contour lines, such as roads, water body boundaries, and building contours.

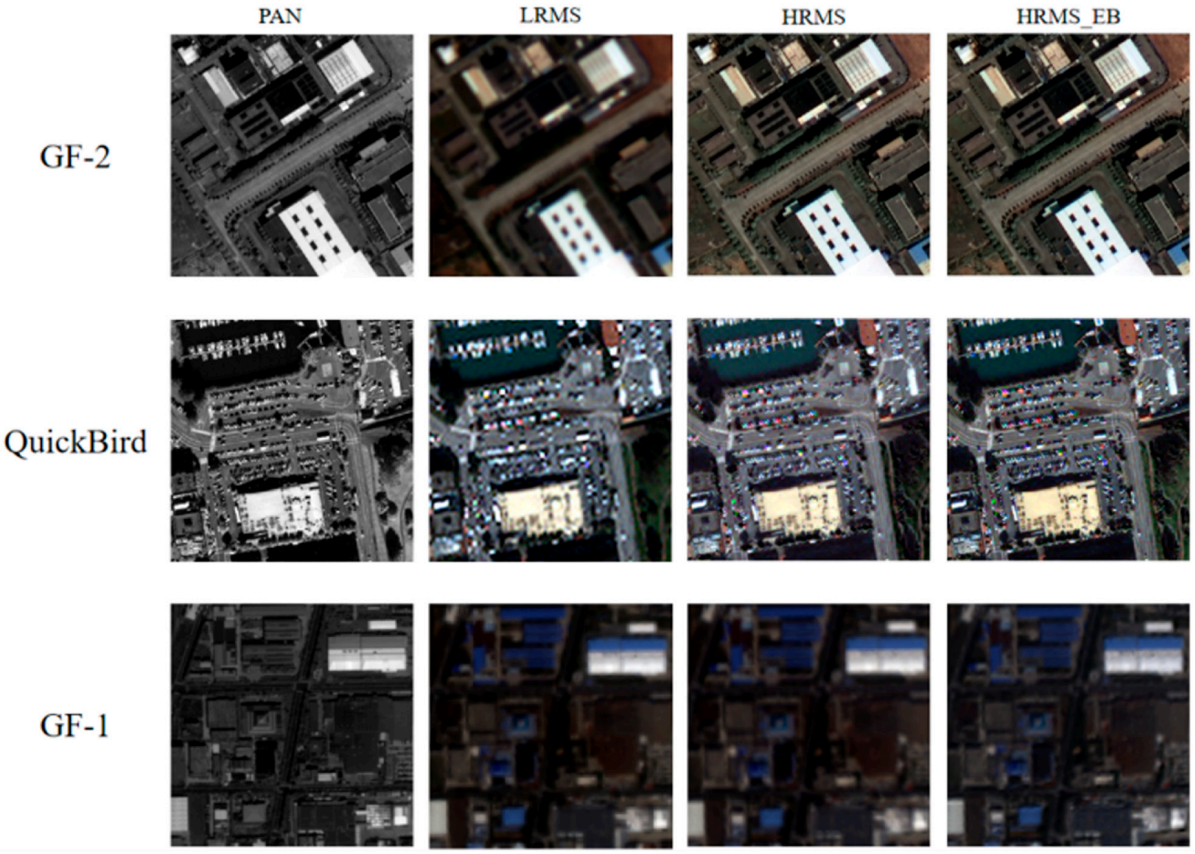


Figure 3. Results of remote sensing image fusion using the error backtracking module on GF-2, QuickBird, and GF-1 datasets.

3.3.2. Index Analysis Results

Additionally, we conducted quantitative evaluations on the remote sensing image fusion results of the above datasets using the same evaluation indices as mentioned earlier. The results are shown in Table 4.

Table 4. Quantitative evaluation results of fusion images using the error backtracking module on GF-2, QuickBird, and GF-1 datasets.

		D_λ	D_s	QNR	SD	SF	FCC
GF-2	HRMS	0.0175	0.0875	0.8964	115.0636	34.9046	0.8160
	HRMS_EB	0.0063	0.0492	0.9346	144.3506	45.4475	0.7421
QuickBird	HRMS	0.0810	0.0087	0.9270	88.3589	26.2061	0.6424
	HRMS_EB	0.0317	0.0174	0.9516	117.8688	37.2020	0.8112
GF-1	HRMS	0.0006	0.0676	0.9319	444.4545	51.9679	0.5215
	HRMS_EB	0.0034	0.0499	0.9470	455.9922	61.1187	0.4513

As shown in Table 4, on the GF-2 satellite data, HRMS_EB has a 64% reduction in D_λ compared to HRMS, indicating an improvement in spectral fidelity. D_s is reduced by 44%, proving better retention of spatial details. The comprehensive QNR also increases by 4.3%. Meanwhile, HRMS_EB has higher SF and SD, further demonstrating its higher spatial clarity and richer spectral information. However, FCC decreases, possibly due to edge enhancement. On the QuickBird satellite data, HRMS_EB has a 61% reduction in D_λ compared to HRMS. Although D_s slightly increases, QNR still improves by 2.6%, SF increases by 41%, SD increases by 33%, and FCC significantly increases by 26%, indicating that the error backtracking module's significant improvement in D_λ on QuickBird satellite data dominates the QNR improvement, showing a comprehensive advantage. For GF-1 satellite data, D_λ increased from 0.0006 to 0.0034, with a slight decrease in spectral fidelity. However, D_s decreased by 26%, and the overall QNR improved by 1.6%. Meanwhile, both SF and SD slightly increased, further supporting the improvement in visual consistency and demonstrating the effectiveness of the error backtracking module on GF-2 satellite data. In summary, the error backtracking module has enhanced the spectral and spatial quality of GF-2, QuickBird, and GF-1 data, performing well across multiple datasets and proving its adaptability, applicability, and feasibility.

4. Discussion

4.1. Discussion on the Results of the Error Backtracking Module in the Ucgan Model

The high-resolution multispectral images generated using the error backtracking module exhibit improved visual performance in terms of color, spatial details, and textures across various land cover types. This approach mitigates the issues of overexposure or underexposure observed in the original fused images and outperforms them in capturing building shadows. The delineation of farmland and cultivated land boundaries is more precise, facilitating accurate identification and analysis of agricultural areas. For water bodies and ships, the generated images demonstrate more reasonable light and shadow distribution, with reflections exhibiting colors and positions better aligned with the hull structures. Overall, the error backtracking module proves effective for image fusion under common feature contexts, achieving superior global visual quality and local color space detail representation.

Evaluation metrics further validate the efficacy of this module in deep learning-based image fusion. The HRMS_EB outperforms the nonerror backtracking HRMS across

four evaluation indices: D_λ, D_S, QNR, SD, and SF. Specifically, the error backtracking module achieves a 33.2170% reduction in D_λ, a 28.1533% reduction in D_S, and a 2.2239% significant improvement in QNR, highlighting its exceptional spectral preservation and spatial detail enhancement capabilities. The generated images exhibit stronger spectral retention, closer alignment with panchromatic images in edges and textures, and higher signal-to-noise ratios. However, a slight decline in the FCC suggests minor trade-offs in feature correlation due to the high retention of spectral and spatial information.

In conclusion, the error backtracking module significantly enhances the performance of the UCGAN model under adversarial training, comprehensively improving the original model's capabilities.

4.2. Performance of Error Backtracking Module in Different Deep Learning Fusion Networks

In order to further verify the effectiveness of error backtracking strategies on different models and further study the performance of deep learning supervised fusion networks and deep learning unsupervised fusion networks, we selected two GAN networks and two CNN networks to conduct extensive experiments.

4.2.1. Performance of Error Backtracking Module Under Deep Learning Supervised Fusion Networks

GPPNN, as a supervision network under convolutional neural networks, has been widely lauded for its suitability in fusion image generation when there are limited samples. PSGAN, as a classic network under the generative adversarial network, has also been demonstrated by numerous studies to be a stable and effective image fusion network. Accordingly, this paper selects these two models, which are representative of CNN and GAN, in accordance with the Wald protocol. The generalization of the module under different models was evaluated by evaluating the fusion images generated before and after the error backtracking module was used to compensate for the prediction results of the two models and optimize them. The architecture diagram of GPPNN and PSGAN with access error backtracking module is shown in Figure 4.

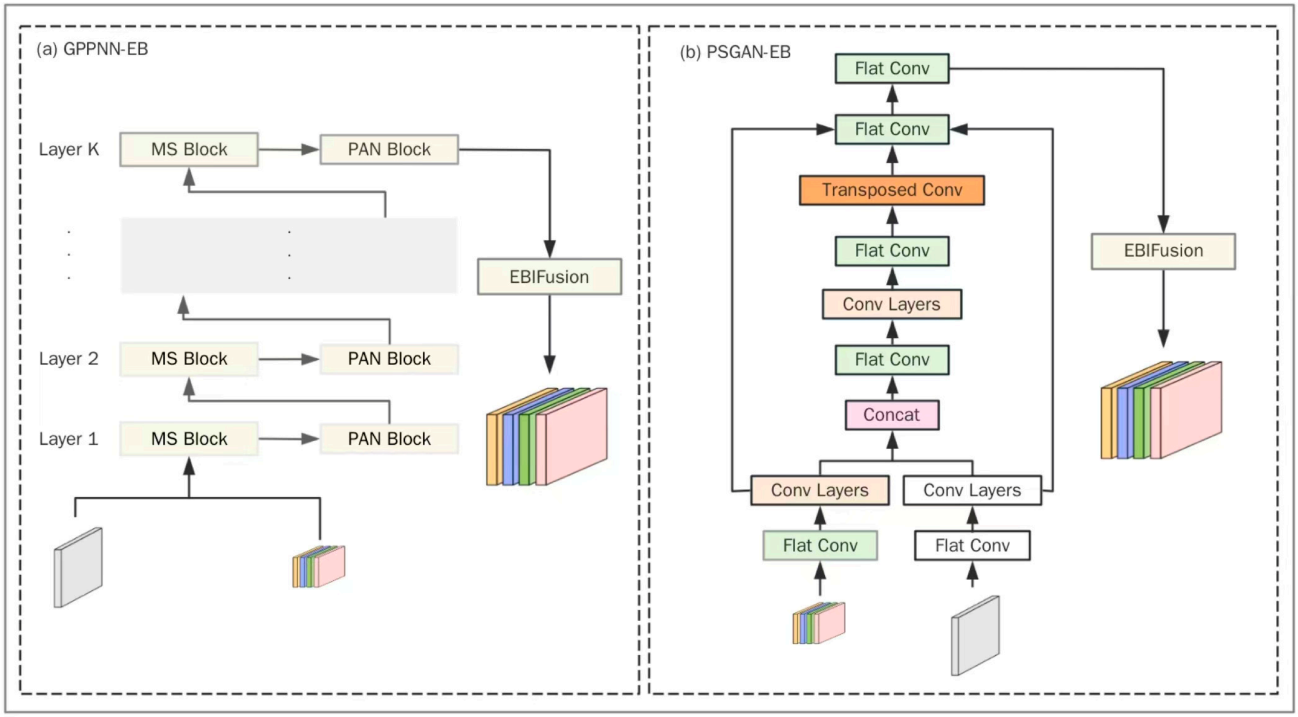


Figure 4. The architecture diagram of GPPNN and PSGAN with access error backtracking module.

As shown in Figure 5, the results of the visual evaluation of the 2 m resolution ground truth image, the PSGAN original fusion result image, and the image generated by PSGAN using the error retrospective module in a side-by-side comparison are shown in Figure 6.

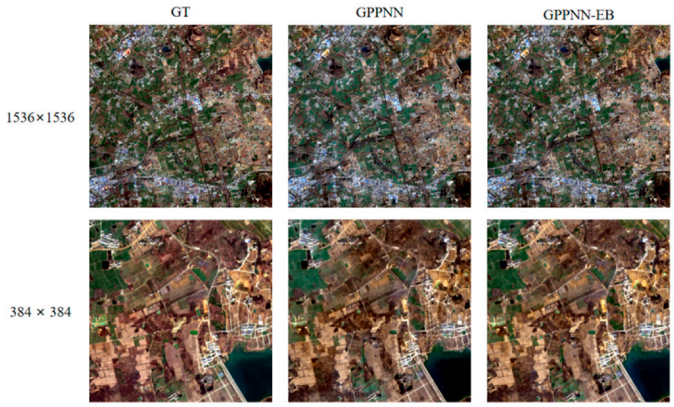


Figure 5. Results of the fusion images of GPPNN and GPPNN-EB.

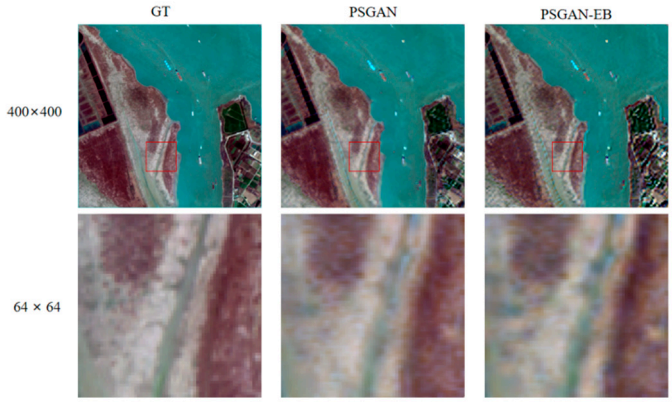


Figure 6. Results of the fusion images of PSGAN and PSGAN-EB. (The red box represents the enlarged area, the same below).

It can be seen from Figure 6 that, on the whole, the hue of the three images is relatively uniform, and the image generated by using the error backtracking module is clearer than the water flow line of the original UCGAN, which is consistent with GT. For the local enlarged image, the color of the original GPPNN is darker than that of GT; for the soil region, the color of the original GPPNN is brown, and the color of GT is more red; and for the water region, the color of the original GPPNN is darker, and the images generated by using the error backtracking module are more similar to the color of GT.

GPPNN is suitable for large-image fusion. Table 5 shows the results of evaluating the images generated under GPPNN before and after using the error backtracking module. According to Table 5, the image generated by the error backtracking module is superior to the original GPPNN fusion image under PSNR, SSIM, CC, and ERGAS indices, indicating that the error backtracking module is helpful to reduce the degree of image distortion and maintain the visual quality of the image. For SAM, although the mean is slightly lower than the original image, it is better than the original GPPNN fusion image on more than half of the images. Therefore, we can assume that the performance of image spectral information is the same before and after using the error backtracking module. To sum up, the error backtracking module shows significant advantages in the supervisory network GPPNN under the convolutional neural network and significantly improves the comprehensive performance of fused images.

Table 5. Results of the quantitative evaluation of the fusion images of GPPNN and GPPNN-EB.

	PSNR		SSIM		CC		SAM		ERGAS	
	GPPNN	GPPNN-EB	GPPNN	GPPNN-EB	GPPNN	GPPNN-EB	GPPNN	GPPNN-EB	GPPNN	GPPNN-EB
1	31.3780	33.0709	0.9362	0.9412	0.9430	0.9500	0.0509	0.0459	3.6104	3.0717
2	25.4800	30.1577	0.8884	0.9300	0.9443	0.9464	0.0462	0.0450	7.2917	4.1934
3	32.5452	33.1216	0.9314	0.9339	0.9333	0.9437	0.0563	0.0500	2.9492	2.7941
4	27.7597	32.7756	0.9275	0.9388	0.9423	0.9483	0.0431	0.0651	4.1196	2.3579
5	29.3958	30.5595	0.9200	0.9271	0.9640	0.9680	0.0620	0.0584	4.9622	4.4092
6	27.5902	30.0743	0.9033	0.9244	0.9549	0.9583	0.0626	0.0576	6.3973	4.8230
7	27.1326	29.8389	0.9086	0.9299	0.9703	0.9737	0.0598	0.0611	6.4907	4.6632
8	25.7432	27.8583	0.8819	0.9163	0.9726	0.9769	0.0531	0.0565	8.4382	6.4893
mean	28.3780	30.9321	0.9122	0.9302	0.9531	0.9582	0.0542	0.0550	5.5324	4.1002

The evaluation results of the fusion images before and after using the error backtracking module under the PSGAN model are shown in the following Table 6. It can be seen from the data that the average PSNR value of the fused image PSGAN-EB using the error backtracking module is 5.9203 higher than that of the original PSGAN fusion image, indicating that the error backtracking module has advantages in reducing the degree of image distortion. At the same time, the error backtracking module is helpful to preserve the spectral information of the image, and the SAM value of PSGAN-EB is reduced by 0.1014 compared with 0.2405 of PSGAN. In addition, the ERGAS value of PSGAN-EB is significantly reduced, indicating that there is a lower error in image quality using the error backtracking module.

Table 6. Results of the quantitative evaluation of the fusion images of PSGAN and PSGAN-EB.

	PSNR	SSIM	CC	SAM	ERGAS
PSGAN	23.6518	0.7096	0.8183	0.2405	9.4243
PSGAN-EB	29.5722	0.8576	0.9198	0.1170	3.5239
Optimization Amount	5.9203	0.1480	0.1014	0.1235	5.9004

In summary, for the deep learning supervised image fusion network, the image generated by using the error backtracking module is closer to the real color and more expressive in details. In terms of visual effects, the error backtracking module shows significant advantages in the deep learning supervised image fusion network. At the same time, the generated image also showed all-around improvement under the image fusion reference index, indicating that the error backtracking module has excellent performance in reducing the degree of image distortion and maintaining the visual quality of the image. Compared with the original model without the error backtracking module, the comprehensive performance of the fusion image generated by the improved model has significantly improved. It comprehensively improves the performance of the image fusion model of the deep learning supervised network.

4.2.2. Performance of Error Backtracking Module Under Deep Learning Unsupervised Fusion Networks

The PNN model represents a significant advancement in the field of remote sensing image fusion, offering a simple yet effective classical deep learning convolutional neural network with a stable training process. Similarly, the UCGAN model has emerged as a leading approach for high-quality remote sensing image fusion, consistently cited in the literature over the past year. Both models demonstrate the capacity to generate superior

high-resolution multispectral images in unsupervised settings. Accordingly, this section assesses the viability of the module in an unsupervised setting through an error backtracking approach on the aforementioned models.

As shown in Figure 7, from a visual perspective, the image generated directly by the PNN model is relatively monochromatic, whereas the image generated by the PNN-EB model exhibits a notable enhancement in detail contrast, with improved performance of the detailed image components. Additionally, the color rendering brightness is more vibrant, which more accurately depicts the structural information and color hierarchy of the image. The image generated by the UCGAN model is significantly blurrier and more flesh-colored in comparison to that of the UCGAN-EB model. Additionally, the color of the image generated by the UCGAN-EB model is lighter brown, and the rendering is also richer and more realistic, which is more similar to the low-resolution multispectral image. The image generated by the UCGAN model is blurred and flesh-colored. In contrast, the color of the image generated by the UCGAN-EB model is light brown, and the rendering is richer and more realistic, which is more similar to that of the low-resolution multispectral image. Additionally, the details of the field intersections are clearer, which preserves the spatial texture of the panchromatic image more accurately.

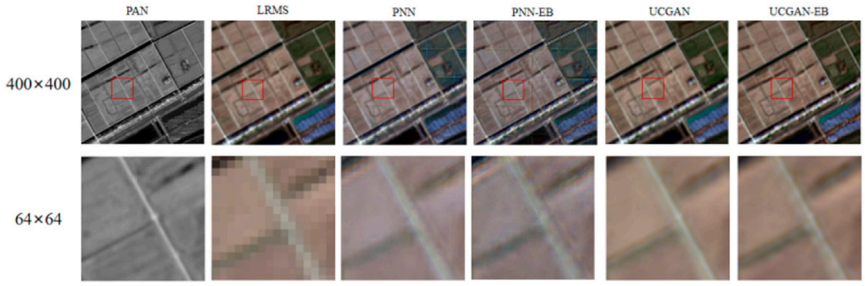


Figure 7. From left to right, they are the original panchromatic image, the low-resolution multispectral image, the PNN model fused high-resolution multispectral image, the high-resolution multispectral image generated by the PNN using the error backtracking module (i.e., PNN-EB), the UCGAN model fused high-resolution multispectral image, and the high-resolution multispectral image generated by UCGAN using the error backtracking module (i.e., UCGAN-EB).

As shown in Table 7, the D_λ and D_s of PNN-EB and UCGAN-EB models are smaller than those of the original models, indicating that the application of the error backtracking module in the model makes the generated image better in spectral retention. The application of the error backtracking module in the two models can make the image closer to the panchromatic image in spatial characteristics, such as edge and texture. After using the error backtracking module, the QNR indices were significantly improved, from 0.8766 to 0.8983 and from 0.9278 to 0.9485, respectively, indicating that the image signal-to-noise ratio was higher and the quality was better after adding the error backtracking module. According to the results, after using the error backtracking module, the fusion images generated by the unsupervised deep learning network show obvious improvement.

Table 7. Quantitative evaluation results of fusion images of PNN and PNN-EB, UCGAN and UCGAN-EB.

	D_λ	D_s	QNR	SD	SF	FCC
PNN	0.0623	0.0652	0.8766	190.3680	59.0158	0.0161
PNN-EB	0.0487	0.0557	0.8983	211.1997	79.2520	0.0186
UCGAN	0.0076	0.0651	0.9278	159.2350	29.6780	0.5677
UCGAN-EB	0.0051	0.0468	0.9485	168.3224	36.0544	0.4856

To sum up, for deep learning unsupervised networks, fusion images generated by the error backtracking module have better visual performance in terms of color, spatial detail, texture, and comprehensive indices without reference parameters. At the same time, by comparing PNN-EB with UCGAN, it can be seen that the use of the module improves the performance of a deep-learning image fusion network with a simple structure in a simple, efficient, and feasible way to a certain extent, and the performance of the improved traditional network is even better than that of the current advanced UCGAN model in terms of parameters such as D_s. The improved PNN-EB model with the error backtracking module and the advanced UCGAN model both obtain three optimal or sub-optimal results in terms of parameter indices. At the same time, compared with the learning and improvement of the complex training network from PNN to UCGAN, the introduction of the error backtracking module to improve the prediction network has achieved a good fusion quality improvement at a small time cost without repeating the training model.

To evaluate the practical feasibility of the proposed EBIFusion module, we further analyzed its computational overhead. The error compensation block mainly consists of lightweight operations, introducing only a negligible number of additional trainable parameters. Theoretically, the parameter increment is less than 0.001 M and can therefore be considered insignificant relative to the backbone fusion networks. The main computational overhead originates from the additional forward inference required to generate the secondary prediction for error estimation rather than from the error compensation block itself. Despite this additional computation, the lightweight design of the proposed module ensures high computational efficiency. Experimental results indicate that the running time of the EB-enhanced network increases by less than 1% compared with the original network, while noticeable improvements in both spectral fidelity and spatial detail preservation are consistently achieved across different datasets and model architectures. Therefore, the performance gains obtained by EBIFusion substantially outweigh its computational cost, demonstrating a favorable trade-off between fusion accuracy and efficiency and confirming its suitability for practical remote sensing image fusion applications.

4.3. Outlook on the Performance of the Error Backtracking Module

Applying the error backtracking module to deep learning image fusion models helps gradually improve the performance of the generated high-resolution fused images, especially in terms of spectral and spatial stability. Moreover, this method is not limited by the model itself. Additionally, this module has achieved good results in image fusion across various remote sensing datasets. Nevertheless, there are some inherent limitations to this approach and potential for further improvement. While the error backtracking module is highly effective in deep learning networks, its underlying principle is to capture information lost in the model to compensate for the generated high-resolution images. Consequently, it is not directly applicable to fusion methods based on component substitution, such as the GS model, Brovey model, and others. Furthermore, given the inherent instability of the novel frame, this study has not been conducted within the Mamba framework [38], which has yet to undergo sufficient testing and expansion. Another limitation of this study lies in the insufficiently detailed quantitative evaluation of the computational overhead introduced by the error backtracking module. EBIFusion is designed as an enhancement module for existing deep learning fusion models, and the overall efficiency of the complete fusion framework is primarily determined by the selected backbone network, input image size, implementation architecture, and hardware environment. Although this module does not alter the training process of the original backbone network, it introduces additional operations during inference, including secondary prediction and residual compensation. Therefore, building on the analysis of the increased parameter count in EBIFusion and the

comparison of runtime between the original network and the EB-enhanced network presented herein, future work will systematically evaluate FLOPs consumption, GPU memory usage, and inference latency under a unified benchmarking environment to further assess the module's feasibility for practical deployment. Concurrently, the module pursues simple generality, to a certain extent, ignoring that the distribution of image spatial information is very complex and that the spectral information may be distorted during the injection process. Therefore, when applying this module to a specific model, one can design an injection coefficient based on local areas. This approach constructs an objective function that contains both spatial and spectral information constraints to achieve the final fusion result. This method helps to avoid the spectral distortion that can occur from injecting too much or too little spatial information. It must be ensured that the fusion result does not have spectral distortion while improving spatial resolution, so that the fusion result after error backtracking does not have spectral distortion while improving the spatial resolution. The initial prediction results can also be combined with the calculated coefficients to form an optimization term, creating a new embedded model. Through iterative updates, the accuracy of the compensation coefficients and the quality of the fusion results may be gradually improved.

5. Conclusions

This paper analyzes the currently available mature deep learning remote sensing image fusion algorithms and proposes an error backtracking-driven optimization module for deep learning image fusion models. Firstly, the error backtracking module is implemented on the unsupervised GAN network. Subsequently, a detailed comparison and analysis of specific image fusion application scenarios were conducted through visual assessment and quantitative evaluations. The study assessed both the overall visual quality and the finer details of the optimized high-resolution multispectral image. The results indicated that the image utilizing an error backtracking module demonstrated superior overall visual effects and local color space detail compared to the image without such a module. The QNR of the fused image generated by the error backtracking module exhibited an increase of 2.22%. In addition, the error backtracking module shows robust performance across different datasets and models. Using the error backtracking module on GF-6, GF-1, and PanCollection datasets, including QuickBird and GF-2 images, can improve the quality of remote sensing image fusion and obtain comprehensive optimal results in spectral and spatial terms. Furthermore, for the deep learning image fusion supervised networks GPPNN and PSGAN, the images generated by using the error backtracking module are better than the fused images of the original GPPNN and PSGAN in terms of PSNR, SSIM, CC, and ERGAS indices. For the deep learning image fusion unsupervised network, the QNR after using the error backtracking module is significantly improved compared to both PNN and UCGAN models, and the color as well as the details of the generated fused images are richer than the original model. In conclusion, the error backtracking module is not restricted by the model and is effective on different image datasets. It effectively enhances the quality of image fusion while maintaining a high signal-to-noise ratio and optimizing feature correlation.

The implementation of the error backtracking module facilitates the retrieval of information lost during the model's generation of high-resolution images. This recovered information is then utilized to enhance the quality of the generated high-resolution images. Importantly, this process neither extends the model training duration nor disrupts the training workflow. The spatial and spectral characteristics of the resulting high-resolution multispectral images are significantly improved, as evidenced by visual assessment, quantitative, and application-oriented metrics. This approach represents a straightforward,

effective deep learning application in remote sensing. It can be seamlessly incorporated into most models as a dedicated optimization module. In the domain of deep learning image fusion for remote sensing, this module has shown remarkable versatility and effectiveness across various evaluation dimensions, highlighting its substantial potential for practical applications.

Author Contributions: Conceptualization, J.M. and Q.D.; Methodology, J.M.; Validation, Q.D., Q.X. and Z.L.; Formal analysis, Z.L.; Data curation, Q.X. and K.W.; Writing—original draft, J.M.; Writing—review & editing, K.W., S.Z. and F.Y.; Visualization, Q.X.; Supervision, S.Z. and F.Y.; Project administration, Z.L., S.Z. and F.Y.; Funding acquisition, K.W. and S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of China under Grant 42271321, and the Nanjing University Integrated Research Platform of the Ministry of Education-Top Talents Program 2026300482, the Fundamental Research Funds for the Central Universities under Grant 020914380174.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available due to data usage agreements with the data providers. However, they can be made available from the corresponding author (zhaosh@nju.edu.cn) upon reasonable request and with permission from the data providers.

Acknowledgments: We would like to thank the China Centre for Resources Satellite Data and Application (CRESDA) for providing the GF-1, GF-2, and GF-6 satellite images and the developers of the PanCollection dataset for making their data publicly available; Generative AI Use Disclosure; No generative artificial intelligence (AI) tools were used in the writing, editing, data analysis, or figure generation of this manuscript. All text, analysis, and conclusions were produced entirely by the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mulla, D.J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [CrossRef]
2. Navalgund, R.R.; Jayaraman, V.; Roy, P. Remote sensing applications: An overview. *Curr. Sci.* **2007**, *93*, 1747–1766.
3. Yin, J.; Dong, J.; Hamm, N.A.; Li, Z.; Wang, J.; Xing, H.; Fu, P. Integrating remote sensing and geospatial big data for urban land use mapping: A review. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102514. [CrossRef]
4. Ge, J.; Tang, H.; Yang, N.; Hu, Y. Rapid identification of damaged buildings using incremental learning with transferred data from historical natural disaster cases. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 105–128. [CrossRef]
5. Liu, S.; Zhao, H.; Du, Q.; Bruzzone, L.; Samat, A.; Tong, X. Novel cross-resolution feature-level fusion for joint classification of multispectral and panchromatic remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5619314. [CrossRef]
6. Nandasena, W.; Brabyn, L.; Serrao-Neumann, S. Using remote sensing for sustainable forest management in developing countries. In *The Palgrave Handbook of Global Sustainability*; Springer: Cham, Switzerland, 2022; pp. 1–22.
7. Zhang, H.; Ma, J.; Chen, C.; Tian, X. NDVI-Net: A fusion network for generating high-resolution normalized difference vegetation index in remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 182–196. [CrossRef]
8. Zhang, H.; Xu, H.; Tian, X.; Jiang, J.; Ma, J. Image fusion meets deep learning: A survey and perspective. *Inf. Fusion* **2021**, *76*, 323–336. [CrossRef]
9. Tang, L.; Zhang, H.; Xu, H.; Ma, J. Deep learning-based image fusion: A survey. *J. Image Graph.* **2023**, *28*, 3–36. (In Chinese) [CrossRef]
10. Sara, D.; Mandava, A.K.; Kumar, A.; Duela, S.; Jude, A. Hyperspectral and multispectral image fusion techniques for high resolution applications: A review. *Earth Sci. Inform.* **2021**, *14*, 1685–1705. [CrossRef]
11. Zhang, J.; Liu, F.; Mei, Y.; Tang, Y. Progress in Spatial Data Fusion: From Classic Approaches to Extended Methods. *Geomat. Inf. Sci. Wuhan Univ.* **2017**, *42*, 1616–1628.

12. Vivone, G.; Mura, M.D.; Garzelli, A.; Restaino, R.; Scarpa, G.; Ulfarsson, M.O.; Alparone, L.; Chanussot, J. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geosci. Remote Sens. Mag.* **2020**, *9*, 53–81. [CrossRef]
13. Zhang, K.; Zhang, F.; Wan, W.; Yu, H.; Sun, J.; Del Ser, J.; Elyan, E.; Hussain, A. Panchromatic and multispectral image fusion for remote sensing and earth observation: Concepts, taxonomy, literature review, evaluation methodologies and challenges ahead. *Inf. Fusion* **2023**, *93*, 227–242. [CrossRef]
14. Su, X.; Li, J.; Hua, Z. Transformer-based regression network for pansharpening remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5407423. [CrossRef]
15. Meng, Q.; Shi, W.; Li, S.; Zhang, L. PanDiff: A novel pansharpening method based on denoising diffusion probabilistic model. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5611317. [CrossRef]
16. Bai, T.; Wang, L.; Yin, D.; Sun, K.; Chen, Y.; Li, W.; Li, D. Deep learning for change detection in remote sensing: A review. *Geo-Spat. Inf. Sci.* **2023**, *26*, 262–288. [CrossRef]
17. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by convolutional neural networks. *Remote Sens.* **2016**, *8*, 594. [CrossRef]
18. Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; Paisley, J. PanNet: A deep network architecture for pan-sharpening. In *Proceedings of the IEEE International Conference on Computer Vision*; IEEE: Piscataway, NJ, USA, 2017; pp. 5449–5457.
19. Liu, X.; Liu, Q.; Wang, Y. Remote sensing image fusion based on two-stream fusion network. *Inf. Fusion* **2020**, *55*, 1–15. [CrossRef]
20. Wei, Y.; Yuan, Q.; Meng, X.; Shen, H.; Zhang, L.; Ng, M. Multi-scale-and-depth convolutional neural network for remote sensed imagery pan-sharpening. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*; IEEE: Piscataway, NJ, USA, 2017; pp. 3413–3416.
21. Chen, Y.; Liu, H.; Fang, F. GF-CSTNet: A Method for Pan-Sharpening Remote Sensing Images by Integrating CSPNet and Transformer. 2024. Available online: https://assets-eu.researchsquare.com/files/rs-4113686/v1_covered_87dbd7fd-32e1-4147-aec0-4738fe6e0024.pdf (accessed on 28 June 2026).
22. Xu, S.; Zhang, J.; Zhao, Z.; Sun, K.; Liu, J.; Zhang, C. Deep gradient projection networks for pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2021; pp. 1366–1375.
23. Liu, Q.; Zhou, H.; Xu, Q.; Liu, X.; Wang, Y. PSGAN: A generative adversarial network for remote sensing image pan-sharpening. In *IEEE Transactions on Geoscience and Remote Sensing*; IEEE: Piscataway, NJ, USA, 2020; Volume 59, pp. 10227–10242.
24. Daihong, J.; Sai, Z.; Lei, D.; Yueming, D. Multi-scale generative adversarial network for image super-resolution. *Soft Comput.* **2022**, *26*, 3631–3641. [CrossRef]
25. Benzenati, T.; Kessentini, Y.; Kallel, A. Pansharpening approach via two-stream detail injection based on relativistic generative adversarial networks. *Expert Syst. Appl.* **2022**, *188*, 115996. [CrossRef]
26. Zhou, H.; Liu, Q.; Weng, D.; Wang, Y. Unsupervised cycle-consistent generative adversarial networks for pan sharpening. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5408814. [CrossRef]
27. Wald, L.; Ranchin, T.; Mangolini, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 691–699.
28. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A.; Nencini, F.; Selva, M. Multispectral and panchromatic data fusion assessment without reference. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200. [CrossRef]
29. Al-Wassai, F.A.; Kalyankar, N.V. A Novel Metric Approach Evaluation for the Spatial Enhancement of Pan-Sharpened Images. *arXiv* **2012**, arXiv:1207.5064.
30. Yang, B.; Jing, Z.-L.; Zhao, H.-T. Review of pixel-level image fusion. *J. Shanghai Jiaotong Univ. (Sci.)* **2010**, *15*, 6–12. [CrossRef]
31. Zhou, J.; Civco, D.L.; Silander, J.A. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. *Int. J. Remote Sens.* **1998**, *19*, 743–757. [CrossRef]
32. Jagalingam, P.; Hegde, A.V. A review of quality metrics for fused image. *Aquat. Procedia* **2015**, *4*, 133–142. [CrossRef]
33. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
34. Yuhas, R.H.; Goetz, A.F.; Boardman, J.W. 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*; JPL: Pasadena, CA, USA, 1992.
35. Wald, L. Quality of high resolution synthesised images: Is there a simple criterion? In *Third Conference "Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images"*; SEE/URISCA: Nice, France, 2000; pp. 99–103.
36. Deng, Z.; Lu, Z.; Wang, G.; Wang, D.; Ding, Z.; Zhao, H.; Xu, H.; Shi, Y.; Cheng, Z.; Zhao, X. Extraction of fractional vegetation cover in arid desert area based on Chinese GF-6 satellite. *Open Geosci.* **2021**, *13*, 416–430. [CrossRef]

37. Deng, L.; Ran, R.; Wu, X.; Zhang, T. CNN-based remote sensing pan-sharpening: A critical review. *J. Image Graph.* **2023**, *28*, 57–79. [CrossRef]
38. He, X.; Cao, K.; Zhang, J.; Yan, K.; Wang, Y.; Li, R.; Xie, C.; Hong, D.; Zhou, M. Pan-mamba: Effective pan-sharpening with state space model. *Inf. Fusion* **2025**, *115*, 102779. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.