remote sensing

MDPI

Article

Spectral-Spatial Residual Network for Fusing Hyperspectral and Panchromatic Remote Sensing Images

Rui Zhao and Shihong Du *

School of Earth and Space Sciences, Peking University, Beijing 100091, China; zr_winton@pku.edu.cn
* Correspondence: dshgis@hotmail.com

Abstract: Fusing hyperspectral and panchromatic remote sensing images can obtain the images with high resolution in both spectral and spatial domains. In addition, it can complement the deficiency of high-resolution hyperspectral and panchromatic remote sensing images. In this paper, a spectral–spatial residual network (SSRN) model is established for the intelligent fusion of hyperspectral and panchromatic remote sensing images. Firstly, the spectral–spatial deep feature branches are built to extract the representative spectral and spatial deep features, respectively. Secondly, an enhanced multi-scale residual network is established for the spatial deep feature branch. In addition, an enhanced residual network is established for the spectral deep feature branch This operation is adopted to enhance the spectral and spatial deep features. Finally, this method establishes the spectral–spatial deep feature simultaneity to circumvent the independence of spectral and spatial deep features. The proposed model was evaluated on three groups of real-world hyperspectral and panchromatic image datasets which are collected with a ZY-1E sensor and are located at Baiyangdian, Chaohu and Dianchi, respectively. The experimental results and quality evaluation values, including RMSE, SAM, SCC, spectral curve comparison, PSNR, SSIM ERGAS and *Q* metric, confirm the superior performance of the proposed model compared with the state-of-the-art methods, including AWLP, CNMF, GIHS, MTF_GLP, HPF and SFIM methods.

Keywords: hyperspectral; panchromatic; image fusion; deep learning; residual network

Citation: Zhao, R.; Du, S. Spectral-Spatial Residual Network for Fusing Hyperspectral and Panchromatic Remote Sensing Images. *Remote Sens.* **2022**, *14*, 800. https://doi.org/10.3390/rs14030800

Academic Editor: Lefei Zhang

Received: 30 November 2021
Accepted: 5 February 2022
Published: 8 February 2022

1. Introduction

With the successful launch of a variety of remote sensing satellites, remote sensing images with different spatial and spectral resolutions from multiple sources have been acquired [1]. There is a certain degree of complementarity among these image data. How to effectively integrate these image data to obtain more abundant image information has become an urgent issue to be solved [2]. Remote sensing image fusion aims at obtaining more accurate and richer information than any single image data. In addition, it generates composite image data with new spatial, spectral and temporal features from the complementary multi-source remote sensing image data in space, time and spectrum [3]. Hyperspectral and panchromatic images are two important types of remote sensing images. The hyperspectral images usually contain abundant spectral information and consist of hundreds or thousands of spectral bands, while the panchromatic images usually have only one spectral band but contain detailed spatial information on ground objects. During the imaging process, due to the limited energy acquired by remote sensing image sensors, for the sake of maintaining high spectral resolution, the spatial resolution of hyperspectral remote sensing images is usually low. Similarly, panchromatic remote sensing images usually have high spatial resolution but low spectral resolution. As a result, the spatial details of ground objects cannot be reflected well in hyperspectral remote sensing images [4], and panchromatic remote sensing images can provide spatial details of ground objects but usually have insufficient spectral information [5]. The fusion of hyperspectral and panchromatic remote sensing images can obtain images with both high spectral and high

spatial resolutions and can complement each other. The fused images can be used in a variety of applications, such as urban target detection [6], ground object classification [7], spectral decomposition [8], etc.

1.1. Existing Fusion Methods

The existing data fusion methods mainly fall into the following levels [9]. (1) Signal level. This type of data fusion process inputs and outputs raw data. Data fusion at this level is conducted immediately after the data are gathered from sensors and based on signal methods. (2) Pixel level. This type of data fusion usually aims at image fusion. Data fusion at this level processes original pixels in raw image data collected from image sensors. (3) Feature level. At this level, both the input and output of the data fusion process are features. Thus, the data fusion process addresses a set of features to improve and refine for obtaining new features. (4) Decision level. This level obtains a set of features as input and provides a set of decisions as output, which is also known as decision fusion.

Remote sensing image fusion is a specific issue in the data fusion field and the fusion algorithms mainly fall into several groups:

(1) The fusion method based on component substitution (CS) relies on a component of multispectral or hyperspectral images, which is substituted by a high spatial resolution remote sensing image. CS fusion methods include the Intensity-Hue-Saturation (IHS) method [10–12] and Principal Component Analysis (PCA) method [13–15], in which PCA components are chosen and fused into new data [16], etc. The shortcoming of this kind of fusion method [17,18] is that the spectral information of the fused images is distorted due to the mismatch between the high spatial resolution images and the spectral range of the hyperspectral spectrum.

(2) The image fusion methods based on multi-scale and resolution analysis (MRA) firstly obtain spatial details through multi-scale decomposition of high spatial resolution images, and then inject them into multi-spectral or hyperspectral remote sensing images. For MRA fusion methods, extractors of spatial details mainly include Decimated Wavelet Transform (DWT), Non-Decimated Wavelet Transform (UDWT) [19], Undecimated Wavelet Transform (UWT) [20], Laplacian Pyramid (LP) [21], Inseparable Transform Based on Curve Wave [22], Inseparable Transform Based on Configure Wave [23], etc. The disadvantages of MRA fusion methods are as follows. The design of spatial filters is usually complex, making the methods difficult to implement and comprehensive in the computational complexity [24].

(3) Bayesian fusion methods rely on the use of a posterior distribution in observed multispectral or hyperspectral images and high spatial resolution images. Selecting appropriate prior information can solve the inverse ill-posed problem in the fusion process [25]. Therefore, this kind of method can intuitively explain the fusion process through the posterior distribution. Since fusion problems are usually ill-conditioned, Bayesian methods provide a convenient way to regularize the problem by defining an appropriate prior distribution for the scenarios of interest. According to this strategy, many scholars have designed different Bayesian estimation methods [26]. The main disadvantage of Bayesian fusion methods is that the prior and posterior information are required for the fusion process, but this information may not be available for all scenes.

(4) The fusion method based on matrix decomposition (i.e., the variational model based fusion method) assumes that hyperspectral images can be modeled as the product of spectral primitives and correlation coefficient matrix. The spectral primitives represent the spectral information in the hyperspectral images and can be divided into sparse expression [27–29] and low-rank expression [30] in the variational image fusion methods. The spectral primitive form of sparse expression has a complete dictionary and assumes that each spectrum is a linear combination of the several dictionary atoms. The atoms here are usually based on a complete dictionary achieved with low spatial resolution of hyperspectral images by sparse dictionary learning methods, such as K-SVD [31], online dictionaries [32], nonnegative dictionary learning, etc. The shortcoming of this kind is that

it needs to make sparse or low-rank a priori assumptions for hyperspectral images, which cannot fully cover all scenes and has certain limitations on the image fusion method.

(5) The fusion methods based on deep learning (DL). Recently, deep learning has gradually become a research hotspot and mainstream development direction in the field of artificial intelligence. DL has made achievements in many fields, such as computer vision [33–35], natural language processing [36], search engine [37], speech recognition [38] and so on. The DL fusion methods are regarded as a new trend, and they train a network model to describe the mapping relationships between hyperspectral images, panchromatic images and target fusion images [39]. Existing DL fusion methods include Pansharpening Neural Network (PNN) [40], Deep Residual Pansharpening Neural Network (DRPNN) [41], Multiscale and Multidepth Convolutional Neural Network (MSDCNN) [42], etc. These methods can usually obtain better spectral fidelity, but spatial enhancement is not enough in image fusion results. The current DL based image fusion methods usually lack deep feature enhancement, and the current DL based image fusion methods usually regard spatial and spectral features as individual units. For existing remote sensing image fusion, it lacks a fusion process on the meter level because the majority of remote sensing image fusion methods lack spatial feature enhancement. Then, the fusion results in existing research are usually on the 10-m level. Meanwhile, it is a lack of interaction for correspondences between spatial and spectral restoration in existing studies, leading to spectral deformation in specific spatial domain.

1.2. Brief Introduction of Proposed Method

In this paper, we utilize the DL method to establish an intelligent fusion model, in which residual networks are implied in the spatial and spectral deep feature branches, respectively. With the residual networks implied in the proposed method, spectral and spatial deep features will be adjusted and enhanced. Specially, multi-scale residual enhancement is utilized on a spatial deep feature branch. Then, the spatial deep feature will be enhanced to a great degree. This ensures the fusion result of the proposed method on the meter level. Then, the proposed method establishes spectral–spatial deep feature simultaneity. This operation is adopted to circumvent the independence of spectral and spatial deep features.

In this paper, the proposed DL method is used to carry out intelligent fusion for hyperspectral and panchromatic images at the meter level. DL methods can extract powerful spectral and spatial deep features from images, effectively maintain the spectral and spatial features of the original images in the fusion process and adjust the learned deep features through some operations such as enhanced residual spatial and spectral networks, spectral–spatial deep feature simultaneity, etc. After the model training, the time complexity of the model can be reduced in the process of generating the fusion image. At the same time, there is no need to establish a priori information or assumptions for hyperspectral and high spatial resolution images. In this paper, the DL method is used to study the meter level intelligent fusion method for hyperspectral and panchromatic images. The spectral–spatial residual network (SSRN) is implied to model the DL method. Firstly, the convolutional deep network is used to extract deep features from hyperspectral and panchromatic images, respectively. In addition, it sets up respective spatial and spectral deep feature branches. This operation is adopted to extract the representative spectral and spatial deep features, respectively. Then the feature level of the DL foundation model is established to build one-to-one correspondence between the spatial and spectral deep feature branches. In addition, the one-to-one correspondent convolutions have the same sizes and dimensions. The current DL image fusion methods usually lack deep feature enhancement. Then, residual networks are implied in the spatial and spectral deep feature branches, respectively. With the residual networks implied in the SSRN method, spectral and spatial deep features will be adjusted and enhanced. The proposed SSRN method establishes feature enhancement for spectral and spatial deep features. Especially, the residual network for the spatial deep feature branch has a multi-scale structure and will maintain a multi-scale spatial deep feature in the proposed SSRN method. At the same

time, the residual network for the spectral deep feature branch will maintain spectral deep feature enhancement in the proposed SSRN method. At this point, the established spatial feature branch is independent of spectral feature, and the current DL image fusion methods usually regard spatial and spectral features as individual units. To integrate convolution of spectral and spatial features, it will be one-to-one correspondence with the same size and dimension. In addition, the spatial convolution for the same piece of spectral convolution will be superposed with the spectral deep convolution. Then, the proposed SSRN method establishes spectral–spatial deep feature simultaneity. This operation is adopted to circumvent the independence of spectral and spatial deep features. This completes the construction of the entire deep learning network.

In this paper, a novel spectral–spatial residual network fusion model is proposed. The main contributions and novelties of the proposed methods are demonstrated as follows:

1. Spectral–spatial one-to-one layer establishment: After spatial deep feature layers are extracted from low level to high level, and spectral deep feature layers are extracted from high level to low level, the spatial and spectral deep feature layers come into being corresponding one-to-one layers with same sizes and dimensions for follow-up operation.
2. Spectral–spatial residual network enhancement: Residual networks are implied in the spatial and spectral deep feature branches, respectively. With the residual networks implied in the SSRN method, spectral and spatial deep features will be adjusted and enhanced. The proposed SSRN method establishes feature enhancement for spectral and spatial deep features. Especially, the residual network for the spatial deep feature branch has a multi-scale structure and will maintain multi-scale spatial deep feature in the proposed SSRN method. At the same time, the residual network for the spectral deep feature branch will maintain spectral deep feature enhancement in the proposed SSRN method.
3. Spectral–spatial deep features simultaneity: To integrate convolution of spectral and spatial features, it will be one-to-one correspondence with the same size and dimension. In addition, the spatial convolution for the same piece of spectral convolution will be superposed with the spectral deep convolution. Then, the proposed SSRN method establishes spectral–spatial deep feature simultaneity. This operation is adopted to circumvent the independence of spectral and spatial deep features.

1.3. Paragraph Arrangement

The rest of this paper is organized as follows: Section 2 gives a detailed description of the proposed method. In Section 3, the experimental results are presented. In Section 4, the discussion of the experimental results is presented. The conclusions are summarized in Section 5.

2. Proposed Method

In this paper, the convolutional neural network (CNN) [43], which is an issue of the deep learning method, is adopted to establish a deep network fusion model for the intelligent fusion of hyperspectral and panchromatic images. CNN is a kind of feedforward neural network that includes convolutional computation and is one of the representative deep learning algorithms. CNN has the ability of representational learning and can carry out translational invariance analysis on input information according to its hierarchical structure. It also builds an imitation biological visual perception mechanism and can undertake supervised learning and unsupervised learning. The difference between CNN and an ordinary neural network is that CNN contains a feature extractor composed of a convolutional layer and a subsampling layer. In the convolutional layer of the CNN, a neuron is only connected with some neighboring layer neurons and usually contains several feature planes. Each feature plane is composed of some matrix arranged neurons, and the neurons in the same feature plane share the same weight, and the shared weight is the convolutional kernel. Subsampling, also known as pooling, usually has two forms:

mean subsampling and maximum subsampling. Subsampling can be considered as a special convolution process. Convolution kernel and subsampling greatly simplify the complexity of the model and reduce the parameters of the model. Before a layer of the feature map, one can learn the convolution kernels of convolution operation. The output of the convolution results after activation function forms a layer of neurons, which constitute the layer feature map. Each neuron input which is connected to the local receptive field of a layer extracts the local features. Once the local features are extracted, the location of the relationship between it and other characteristics were determined.

The proposed deep network model for the fusion of hyperspectral and panchromatic images consists of three parts: (1) spectral–spatial deep feature branches, (2) enhanced multi-scale residual network of spatial feature branch and residual network of spectral feature branch and (3) spectral–spatial deep feature simultaneity. The latter two operations aim at adjusting the deep features learned from the DL network by which the deep features are more representational and integrated.

The schematic diagram of the establishment process of the spectral and spatial deep feature branches of hyperspectral and panchromatic images is shown in Figure 1. Convolution operation is the core process in establishing spectral–spatial deep feature branches. The function of the convolution operation is to extract the features from the input data, and it contains multiple convolution kernels. Each element of the convolution kernel corresponds to a weight coefficient and a deviation quantity, which is like the neurons of a feedforward neural network. Each neuron in the convolutional layer relates to multiple neurons in the area close to the previous layer. The size of the area depends on the size of the convolutional nucleus, which is also called the receptive field and analogous to the receptive field of visual cortex cells. When the convolution kernel is working, it will sweep the input features regularly and sum the input features by matrix element multiplication in the receptive field and overlie the deviation. In Figure 1, PAN represents the panchromatic image, HSI represents the hyperspectral image, conv represents the convolution operation, pool represents the pooling operation and up represents the up-sampling operation.

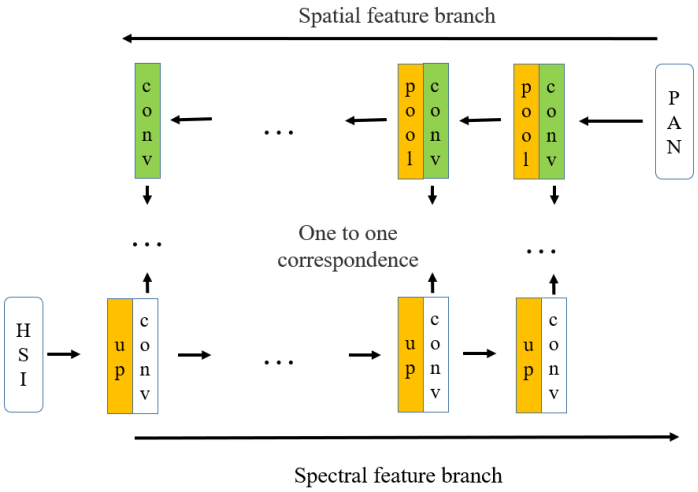


Figure 1. Schematic diagram of spectral–spatial deep feature branch.

Hyperspectral images contain rich and detailed spectral information, while panchromatic images contain relatively rich spatial information. For hyperspectral and panchromatic images, the CNN is used to extract spectral and spatial deep features, respectively, and the two basic deep feature branches are established, which are spectral and spatial deep feature branches, respectively. In the branch of spatial deep feature, the panchromatic image is convoluted and pooled layer by layer to form spatial deep convolution features. Then spatial deep features are extracted layer by layer from panchromatic images and the establishment is completed for the spatial deep feature branch. The convolution operation on the branch of spatial deep feature is shown in Equation (1):

$$\begin{aligned} X_{spatial}^{l+1}(i,j) &= \left[X_{spatial}^{l} \otimes w_{spatial}^{l+1}\right](i,j) \\ &+b_{spatial} = \sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[X_{spatial,k}^{l}(s_0 i + x, s_0 j + y) w_{spatial,k}^{l+1}(x,y)\right] + b_{spatial} \quad (i,j) \\ &\in \{0,1,\ldots,L_{l+1}\}\ L_{l+1} = \frac{L_l + 2p - f}{s_0} + 1 \end{aligned} \tag{1}$$

where $\otimes$ is the convolution operation, $b_{spatial}$ is the deviation value, $X_{spatial}^{l}$ and $X_{spatial}^{l+1}$ represent the input and output of the $l+1$ level convolution on the branch of spatial deep feature, L_{l+1} is the size of $X_{spatial}^{l+1}$, $X_{spatial}^{l+1}(i,j)$ is the pixel corresponding to the feature map, K is the number of channels of the feature map, f is the size of the convolution kernel, s_0 is the convolution step size and p is the number of filling layers. The pooling operation on the branch of spatial deep feature is shown in Equation (2):

$$X_{spatial,k}^{l+1}(i,j) = \left[\sum_{x=1}^{f}\sum_{y=1}^{f} X_{spatial,k}^{l}(s_0 i + x, s_0 j + y)^p\right]^{\frac{1}{p}} \tag{2}$$

where $X_{spatial}^{l}$ and $X_{spatial}^{l+1}$ represent the input and output of the $l+1$ level convolution on the branch of spatial deep feature, step length is s_0, pixel (i,j) has the same meaning as the convolution layer; when $p = 1$, L_p pooling takes the mean value in the pooled area, which is called average-pooling; when $p \to \infty$, L_p pooling takes the maximum value in the region, which is called max-pooling. The max-pooling method is adopted in this paper. At the same time, in the branch of spectral deep feature, the hyperspectral images are subjected to the up-sampling and convolution operation layer by layer to form the spectral deep convolution features. Then spectral deep features are extracted layer by layer from the hyperspectral remote sensing image and the establishment is completed for the spectral deep feature branch. The convolution operation on the spectral deep feature branch is shown in Equation (3):

$$\begin{aligned} X_{spectral}^{l+1}(i,j) &= \left[X_{spectral}^{l} \otimes w_{spectral}^{l+1}\right](i,j) + b_{spectral} \\ &= \sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[X_{spectral,k}^{l}(s_0 i + x, s_0 j + y) w_{spectral,k}^{l+1}(x,y)\right] + b_{spectral} \end{aligned} \tag{3}$$

where $X_{spectral}^{l}$ and $X_{spectral}^{l+1}$ represent the input and output of the $l+1$ layer convolution on the spectral deep feature branch, the other parameters are similar to those of the convolution operation on the spatial feature branch.

Trilinear interpolation is used for the up-sampling procedure. Assuming that we want to know the value of the unknown function f at point $P = (x,y)$; meanwhile, supposing we know the values of the function f at four points $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$, $Q_{22} = (x_2, y_2)$, then we implement linear interpolation in the x direction, as shown in Equation (4):

$$\begin{aligned} f(x,y_1) &\approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \\ f(x,y_2) &\approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \end{aligned} \tag{4}$$

and then we implement linear interpolation in the y direction and obtain $f(x,y)$, as shown in Equation (5):

$$\begin{aligned} f(x,y) &= \frac{y_2 - y}{y_2 - y_1} f(x,y_1) + \frac{y - y_1}{y_2 - y_1} f(x,y_2) \\ &= \frac{1}{(x_2 - x_1)(y_2 - y_1)} \begin{bmatrix} x_2 - x & x - x_1 \end{bmatrix} \begin{bmatrix} f(Q_{11}) & f(Q_{12}) \\ f(Q_{21}) & f(Q_{22}) \end{bmatrix} \begin{bmatrix} y_2 - y \\ y - y_1 \end{bmatrix} \end{aligned} \tag{5}$$

There are the same number of levels in the two basic feature branches of the deep network, and the two basic branches of each layer correspond one-to-one. According to the same set of width ratio and spectral–spatial feature hierarchy between hyperspectral and

panchromatic images, the level number of basic feature branches of two deep network is established. For panchromatic images, pooling down sampling between layers is used to obtain spatial convolution feature blocks of different sizes and dimensions for each layer. For hyperspectral images, the spectral convolution feature block with the same size as the corresponding spatial feature branch block is obtained by using up-sampling between layers. In this way, two basic spectral and spatial deep feature branches with corresponding feature blocks of the same size and dimension are established.

The schematic diagram of multi-scale residual enhancement of spatial deep features and residual enhancement for spectral deep features is shown in Figure 2. In Figure 2, the plus sign represents the residual block. This paper adopts residual network structure to adjust the spatial deep features in the spatial feature branch and then carries out multi-scale spatial information feature enhancement. Meanwhile, the proposed SSRN method also adopts residual network structure to enhance the spectral deep features in the spectral feature branch. Residual network allows original information transferring to the latter layers. For residual network, denoting desired underlying mapping function $H(*)$, we let the stacked nonlinear layers fit another mapping as Equation (6):

$$F(x') = H(x|x') - x \tag{6}$$

where $F(x')$ is the residual part and x is the mapping part. x' is residual variable and x is mapping variable. The original mapping is recast into Equation (7):

$$H(x|x') = x + F(x') \tag{7}$$

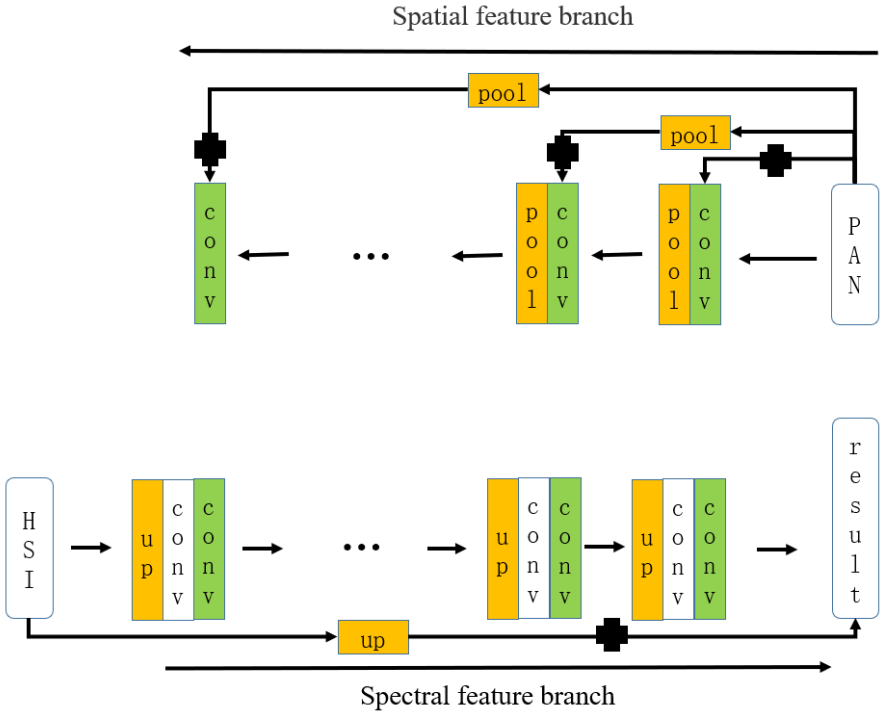


Figure 2. Schematic diagram of multi-scale residual enhancement for spatial deep features and residual enhancement for spectral deep features.

The formulation of $x + F(x')$ can be realized by feedforward neural networks with shortcut connections. The shortcut connections simply perform identity mapping, and their outputs are added to the outputs of the stacked layers. Identity shortcut connections add neither extra parameter nor computational complexity [44]. For convolutional residual network with Equation (7), there is a convolutional operation in residual part $F(x')$ and an identity mapping with mapping part x. Convolutional residual network is adopted in the

proposed SSRN method. Establishing the residuals of the network structure is based on the spatial feature branch at each level of convolution to join the original panchromatic image information to complete. Firstly, according to the size of the features of each convolution layer, pooling operation is adopted on panchromatic images. On the basis of the dimensions of each convolution layer, panchromatic images after pooling dimension are put on the stack, then the convolution results of each layer are added together to construct the spatial deep feature residual-enhanced network structure. Due to the different size and dimension of the convolution features in each layer, the sizes and dimensions of each convolution layer that joins the residuals of the network structure are not the same. Then the original panchromatic images in the spatial information in the form of multi-scale can join into the layers of convolution, which is formed of multi-scale spatial deep enhanced residual. In this way, the spatial features of panchromatic images are highlighted. In this paper, multi-scale residuals are added to the spatial deep feature branch of panchromatic images, which can enhance the spatial deep feature branch of panchromatic images to a certain extent. At the same time, residual network structure is also adopted in the spectral deep feature branch. In the spectral branch, residual network is only adopted in the final output convolution layer to enhance the spectral deep feature. In this procedure, trilinear interpolation up-sampling is adopted firstly to come up to the same spatial size as the final output convolution layer. Then, the up-sampling block is added to the final output convolution layer. The above representation is the procedure of spatial–spectral residual network operation. When the numbers of convolution layers on the spatial feature branch of panchromatic image and the spectral feature branch of hyperspectral image increase, the gradient will gradually disappear in the back propagation, and the underlying parameters cannot be effectively updated, resulting in the phenomenon of gradient disappearance. Or, when taking the derivative of the activation function, if the derivative is greater than 1, then when the number of layers increases, the gradient and novelty finally obtained will increase exponentially, resulting in the phenomenon of gradient explosion. The problem of gradient explosion or gradient vanishing will make the model training difficult to converge. By using the residual network, the gradient updated information can be transmitted from the upper level to the lower level during the back propagation, and the gradient explosion or gradient disappearance can be prevented. Therefore, adding the multi-scale residual enhancement network structure to the spatial deep feature branch and residual enhancement network structure to the spectral deep feature branch cannot only enhance the convolution feature on the spatial and spectral deep feature branch to a certain extent but also prevent the occurrence of gradient explosion or gradient disappearance. The pooling operation of spatial feature branch residual enhancement on panchromatic images is shown in Equation (8):

$$X_{pan}^{l}(i,j) = \left[\sum_{x=1}^{f} \sum_{y=1}^{f} X_{pan}(s_0 i + x, s_0 j + y)^p \right]^{\frac{1}{p}} \tag{8}$$

where X_{pan} is the original panchromatic image, X_{pan}^{l} is the multi-scale residuals on the convolution of the l layer that are enhanced by pooling results. Other parameters are like the pooling operation mentioned above. The result stacking operation after pooling is shown in Equation (9):

$$X_{pan,cat}^{l} = Cat_{i=1}^{k}\left(X_{pan,i}^{l}\right) \tag{9}$$

where $Cat(*)$ is the stack from the dimension i to k, and k is the dimension of the convolution of the corresponding l layer. Residual network is constituted with a series of residual block, which is constituted with mapping and residual parts. The operation of the residual enhancement network structure is shown in Equation (10):

$$X_{spatial,resi}^{l} = H\left(X_{spatial}^{l} \middle| X_{pan,cat}^{l}\right) = X_{spatial}^{l} + F(X_{pan,cat}^{l}) \tag{10}$$

where $X^l_{spatial}$ is the result of the convolution at the l level on the spatial feature branch, and it is the mapping part for identity mapping in the residual block, $F(X^l_{pan,cat})$ is the residual part in the residual block according to Equation (7), $X^l_{spatial,resi}$ is the result of enhanced residuals of the convolution of the l layer on the spatial feature branch. For convolutional residual network, $F(X^l_{pan,cat})$ is represented with a convolution operation as Equation (11):

$$\begin{aligned} F(X^l_{pan,cat})(i,j) &= \left[X^l_{pan,cat} \otimes w^l_{pan,cat}\right](i,j) + b_{pan,cat} \\ &= \sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[X^l_{pan,cat}(s_0 i + x, s_0 j + y) w^l_{pan,cat}(x,y)\right] + b_{pan,cat} \end{aligned} \tag{11}$$

where $w^l_{pan,cat}$ is the convolution weight of the l residual part, $b_{pan,cat}$ is the biases of the l residual part. The other parameters are similar to those of the convolution operation on the spatial and spectral feature branch. Then, $F(X^l_{pan,cat})$ is represented in Equation (12):

$$F(X^l_{pan,cat}) = arr^{n\_spatial}_{i=1}\left(arr^{m\_spatial}_{j=1}\left(\sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[X^l_{pan,cat}(s_0 i + x, s_0 j + y) w^l_{pan,cat}(x,y)\right] + b_{pan,cat}\right)\right) \tag{12}$$

where $arr^n_{i=1}(*)$ and $arr^m_{j=1}(*)$ are the array formation of all of pixel $F(X^l_{pan,cat})(i,j)$, then the array $F(X^l_{pan,cat})$ is formulated with $arr^n_{i=1}\left(arr^m_{j=1}\left(F(X^l_{pan,cat})(i,j)\right)\right)$ here, $n\_spatial$ and $m\_spatial$ are the width and height of $F(X^l_{pan,cat})$. By substituting Equation (12) into Equation (10), $X^l_{spatial,resi}$ is obtained as Equation (13):

$$X^l_{spatial,resi} = X^l_{spatial} + arr^{n\_spatial}_{i=1}\left(arr^{m\_spatial}_{j=1}\left(\sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[X^l_{pan,cat}(s_0 i + x, s_0 j + y) w^l_{pan,cat}(x,y)\right] + b_{pan,cat}\right)\right) \tag{13}$$

The trilinear interpolation of the original hyperspectral image for residual network on spectral feature branch is the same as Equations (4) and (5). After up-sampling, the residual network on spectral feature branch is shown in Equation (14):

$$X^L_{spectral,resi} = H\left(X^L_{spectral}\middle|X_{HSI,up-sampling}\right) = X^L_{spectral} + F(X_{HSI,up-sampling}) \tag{14}$$

where $X_{HSI,up-sampling}$ is the original hyperspectral image after up-sampling, $X^L_{spectral}$ is the final output convolution layer and it is the mapping part for identity mapping in the residual block. In addition, $F(X_{HSI,up-sampling})$ is the residual part in the residual block according to Equation (7). $X^L_{spectral,resi}$ is the result of residual network on spectral deep feature branch. For convolutional residual network, $F(X_{HSI,up-sampling})$ is represented with a convolution operation as Equation (15):

$$\begin{aligned} F(X_{HSI,up-sampling})(i,j) &= \left[X_{HSI,up-sampling} \otimes w_{HSI,up-sampling}\right](i,j) + b_{HSI,up-sampling} \\ &= \sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[X_{HSI,up-sampling}(s_0 i + x, s_0 j + y) w_{HSI,up-sampling}(x,y)\right] + b_{HSI,up-sampling} \end{aligned} \tag{15}$$

where $w_{HSI,up-sampling}$ is the convolution weight of the residual part of the spectral residual network, $b_{HSI,\ up-sampling}$ is the biases of the residual part of the spectral residual network. The other parameters are similar to those of the convolution operation on the spatial and spectral feature branch. Then, $F(X_{HSI,up-sampling})$ is represented as Equation (16):

$$\begin{aligned} &F(X_{HSI,up-sampling}) \\ &= arr^{n\_spectral}_{i=1}\left(arr^{m\_spectral}_{j=1}\left(\sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[X_{HSI,up-sampling}(s_0 i + x, s_0 j + y) w_{HSI,up-sampling}(x,y)\right]\right.\right. \\ &\left.\left. + b_{HSI,up-sampling}\right)\right) \end{aligned} \tag{16}$$

where $arr^n_{i=1}(*)$ and $arr^m_{j=1}(*)$ are the array formation of all of pixel $F(X_{HSI,up-sampling})(i,j)$, then the array $F(X_{HSI,up-sampling})$ is formulated with $arr^n_{i=1}\left(arr^m_{j=1}\left(F(X_{HSI,up-sampling})(i,j)\right)\right)$

here, $n\_spectral$ and $m\_spectral$ are the width and height of $F(X_{HSI,up-sampling})$. By substituting Equation (16) into Equation (14), $X^L_{spectral,resi}$ is obtained as Equation (17):

$$X^L_{spectral,resi} = X^L_{spectral} + arr_{i=1}^{n\_spectral}\left(arr_{j=1}^{m\_spectral}\left(\sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[X_{HSI,up-sampling}(s_0 i + x, s_0 j + y)w_{HSI,up-sampling}(x,y)\right] + b_{HSI,up-sampling}\right)\right) \quad (17)$$

The schematic diagram of alignment for spectral and spatial deep features is shown in Figure 3. The deep network established in this paper, after constructing the spectral and spatial deep feature branches, establishing multi-scale residual enhancement on the spatial feature branch and establishing residual enhancement on the spectral feature branch, builds the deep feature linkage of the spectral and the spatial feature branches. In addition, it adjusts the compatibility of the spatial and spectral deep features and makes the spatial and spectral deep features in the network more representational. Upon the completion of the spectral and spatial features in deep branches and multi-scale enhanced residual features in spatial and spectral feature branches, from low-level to high-level, spatial convolution correspond to the spectral convolution features from senior to junior. Both the existence of one-to-one correspondence with same sizes and dimensions are of the same convolution features. To make the spectral and spatial feature branches work together, the convolution feature blocks corresponding to each other with the same size and dimension are accumulated on the spectral feature branch, such that the corresponding convolution feature blocks in the spatial and spectral feature branches can be fused and interact. In this way, the obtained spectral and spatial accumulated features can have stronger feature characterization ability. After the establishment of spectral and spatial deep feature branches, the spatial deep feature multi-scale residual enhancement and spectral deep feature residual enhancement, and the spectral and spatial deep feature branches joint, the fusion image is finally output.

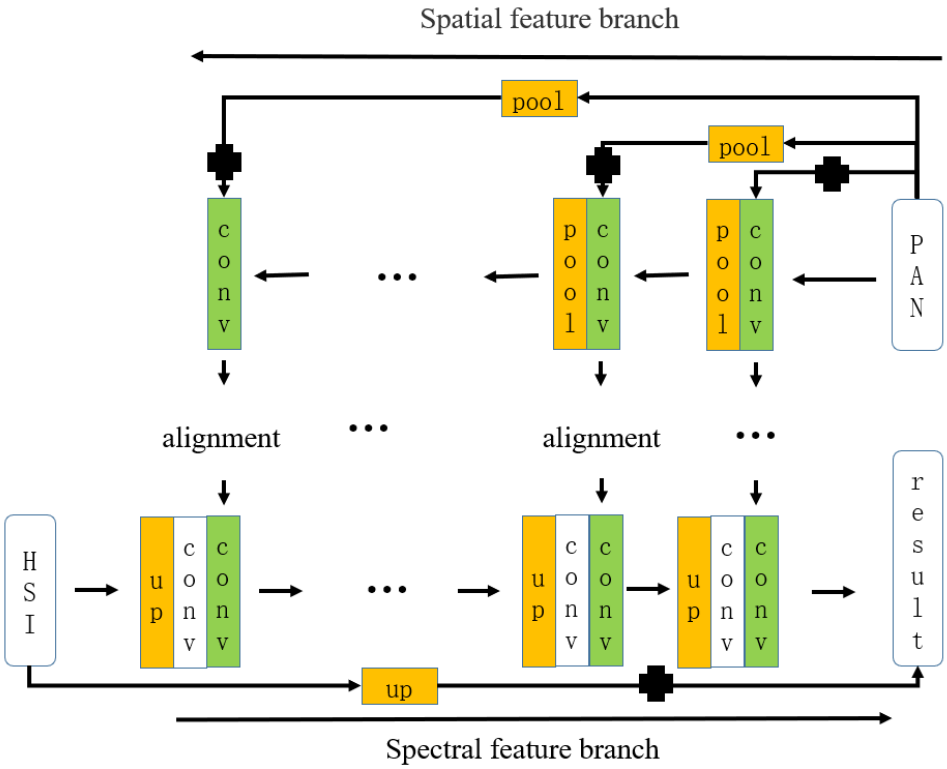


Figure 3. Spectral–spatial deep feature alignment.

3. Results

This section refers to a series of experiments undertaken to show the fusion results of the proposed SSRN method. The constituent parts are as follows.

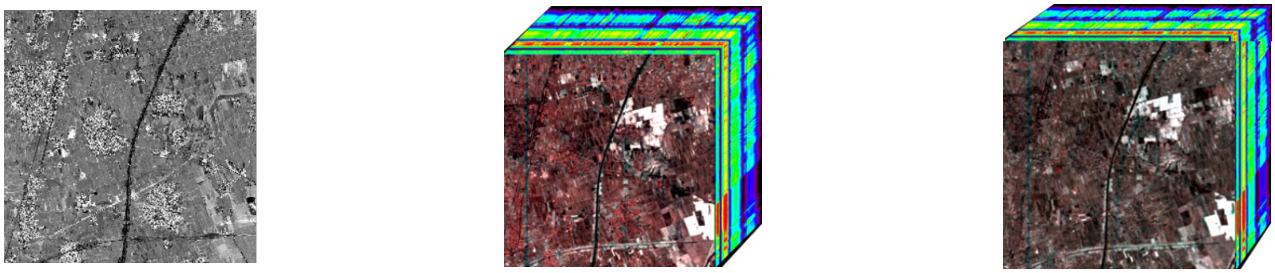
(1) Description of the hyperspectral, panchromatic and ground truth datasets used to investigate the effectiveness of the proposed SSRN method.
(2) Fusion results comparison of the proposed SSRN method with the state-of-the-art fusion methods on the experimental hyperspectral and panchromatic datasets.

3.1. Datasets Description

Three groups of real-world hyperspectral and panchromatic datasets were utilized in the experiments to investigate the effectiveness of the proposed SSRN method in the task of remote sensing image fusion. These three group datasets possess different properties, including the clutter and distribution of the ground land covers and the field of image coverage.

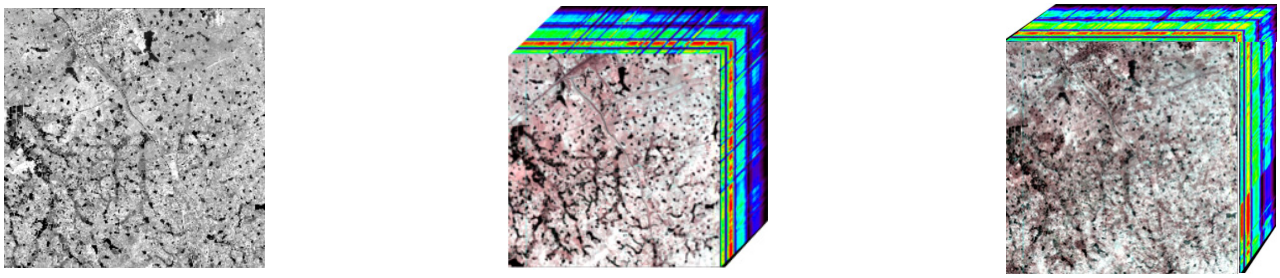
These three groups of real-world hyperspectral and panchromatic datasets were all collected by the ZY-1E hyperspectral and panchromatic remote sensors. For these three groups of datasets, there are three images in each dataset group. The hyperspectral remote sensor is with approximately 30-m spatial resolution and contains 90 spectral channels from 0.4 to 2.5 um. After removing the bands of the water absorption region, low signal-to-noise ratio and poor quality (22–29, 48–58, 88–90), 68 bands remained. The panchromatic remote sensor has approximately 2.5-m spatial resolution and with only one spectral band which is contained with relatively clear visual effect for the ground land cover. The sizes of these three hyperspectral datasets are all 300 $\times$ 300 pixels, and the size of the three panchromatic datasets are all 3600 $\times$ 3600 pixels.

The three group datasets possess different ground land covers and field of image coverage. The first hyperspectral and panchromatic dataset is obtained with the field of Baiyangdian region, which is located at the junction of Baoding City and Cangzhou City in Hebei, China. This hyperspectral and panchromatic dataset contains ground land covers of buildings, roads, croplands and shadows. The panchromatic, hyperspectral and the ground truth images of the Baiyangdian region dataset are shown in Figure 4. The second hyperspectral and panchromatic dataset is obtained with the field of Chaohu region, which is in the middle and lower reaches of the Yangtze River, one of the five major freshwater lakes in China. This hyperspectral and panchromatic dataset contains ground land covers of water, roads, mountains and croplands. The panchromatic, hyperspectral and the ground truth images in the Chaohu region are shown in Figure 5. The third hyperspectral and panchromatic dataset is obtained with the field of Dianchi region in the southwest of Kunming City. This dataset contains ground land covers of mountains, water, rivers and jungles. The panchromatic, hyperspectral and the ground truth images of the Dianchi region dataset are shown in Figure 6. For all of the three datasets, the ground truth images all have the same spatial resolution with the respective panchromatic image. In addition, the ground truth images have a size of 3600 $\times$ 3600 pixels. Meanwhile, the ground truth images all have the same spectral resolution with the respective hyperspectral image. All of the ground truth images have 68 spectral bands (after removing the bands of the low signal-to-noise, poor quality and the water absorption). All the ground truths of the three datasets were obtained by unmanned aerial vehicles with remote sensing image sensors. The sensor has the same retrievable spectral resolution and wave range with the hyperspectral images of the three datasets in our experiments. Meanwhile, with the low altitude flight of the unmanned aerial vehicle, the images obtained from this sensor have very high spatial resolution, which is same as the panchromatic images of the three datasets in our experiments. After the ground truths were obtained, geometric correction and radiometric calibration are carried out. Then, the ground truths of the three datasets in our experiments have the same spectral resolution with the hyperspectral images and the same spatial resolution with the panchromatic images in our experiments.



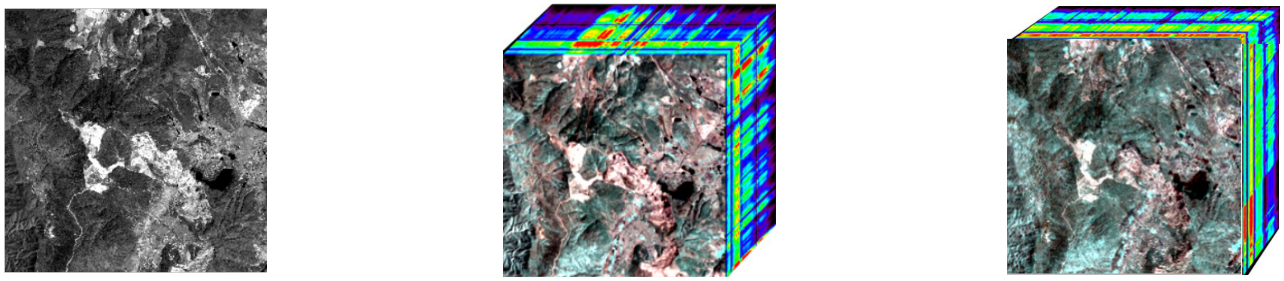
(**a**) (**b**) (**c**)

Figure 4. The Baiyangdian region dataset: (**a**) panchromatic image, (**b**) RGB 3D cube of hyperspectral image, (**c**) RGB 3D cube of ground truth image.



(**a**) (**b**) (**c**)

Figure 5. The Chaohu region dataset: (**a**) panchromatic image, (**b**) RGB 3D cube of hyperspectral image, (**c**) RGB 3D cube of ground truth image.



(**a**) (**b**) (**c**)

Figure 6. The Dianchi region dataset: (**a**) panchromatic image, (**b**) RGB 3D cube of hyperspectral image, (**c**) RGB 3D cube of ground truth image.

3.2. Experimental Setting

To verify the effectiveness of the proposed SSRN method, the experiments on the previously mentioned datasets are conducted. Four mainstream fusion methods are selected for the comparison, including Coupled Nonnegative Matrix Factorization (CNMF) [45], Modulation Transfer Function–Generalized Laplacian Pyramid (MTF_GLP) [46], General Intensity-Hue-Saturation (GIHS) [47], A-trous Wavelet Transform-based Pan-sharpening (AWLP) [48], High Pass Filtering (HPF) [24] and Smoothing Filter-based Intensity Modulation (SFIM) [24], which are implemented in MATLAB R2016b software. The proposed SSRN method is implemented in Python with the PyTorch frame.

The proposed SSRN method is implemented with three layers on spatial and spectral deep branches on all the three datasets. All the kernels in the convolution layers on both spatial and spectral deep branches have 3×3 kernel size. The kernels of the pooling layers on spatial deep branch have 2×2 kernel size. Then, on spatial deep branch, the output data

of one convolution layer and pooling layer will shrink to a quarter of its input data size. The first up-sampling layer on the spectral deep branch is set as three times up-sampled for the input data, and the following two up-sampling layers on the spectral deep branch are set as two times up-sampled for the input data. On the spatial deep branch, the input data is a 300 × 300 image patch; meanwhile, the input data is a 25 × 25 image patch on the spectral deep branch.

3.3. Experimental Results

In this section, the fusion performances of the proposed SSRN method and the state-of-the-art methods will be provided for result evaluation. Subjective nature of evaluation with different ground land covers fusion performance will be demonstrated in this section. In addition, the objective evaluation approach with performance indexes will be demonstrated later in Section 4.

The RGB images of the ground truth, the compared methods and the proposed SSRN method on the Baiyangdian region dataset are shown in Figure 7. The fusion result of the AWLP method roughly restores the ground land covers; nevertheless, some shadow regions are not as well fused as the other ground land covers. The fusion result of the CNMF method is relatively obscure. Most of the ground land covers are not well manifested in the fusion result of the CNMF method. The fusion result of the GIHS method has poor RGB contrast, that is, the GIHS method does not provide reasonable spectral revivification according to the ground truth image. The fusion result of the MTF_GLP method gives poor restoration for some croplands and shadows but gives good restoration for buildings and roads. The fusion results of the HPF and SFIM methods give good restoration for buildings and roads but poor restoration for croplands and shadows. The proposed SSRN method restores all of the ground land covers in the rough better than the compared fusion methods and gives better sharpness than the other fusion methods.

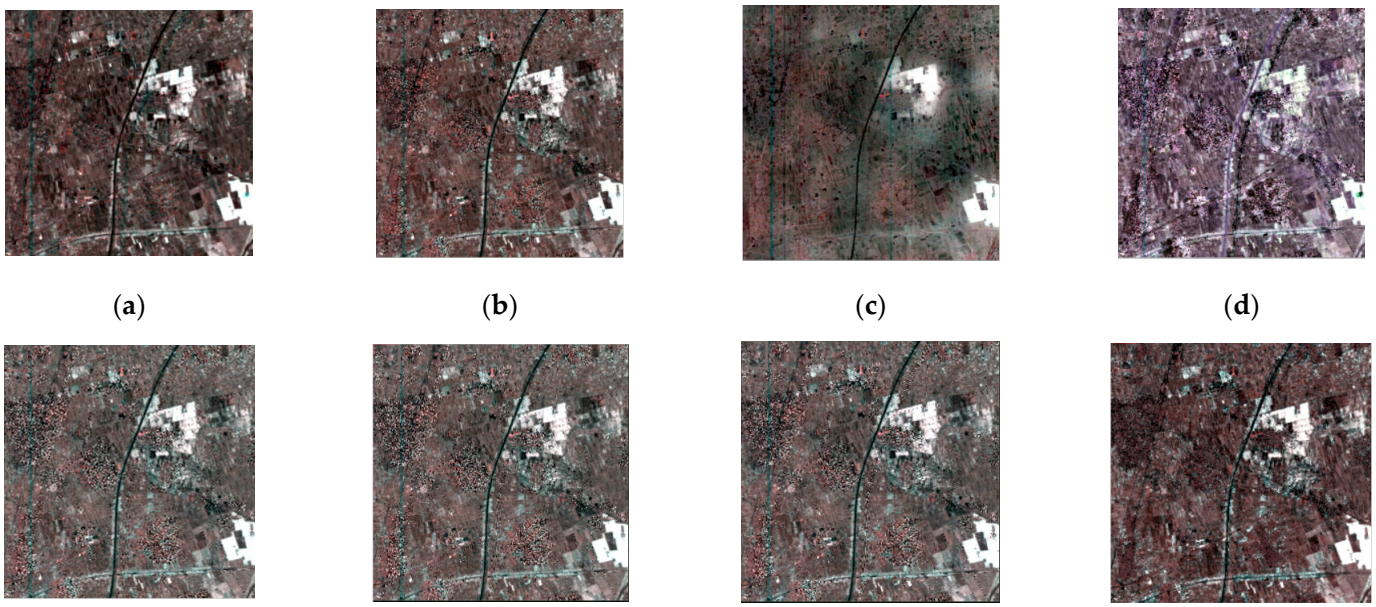


Figure 7. RGB images of the ground truth, the compared methods and the proposed SSRN method on Baiyangdian region dataset. (**a**) Ground truth, (**b**) AWLP, (**c**) CNMF, (**d**) GIHS, (**e**) MTF_GLP, (**f**) HPF, (**g**) SFIM and (**h**) SSRN.

The RGB images of the ground truth, the compared methods and the proposed SSRN method on the Chaohu region are shown in Figure 8. For the water land cover, the fusion results of the AWLP, CNMF and the proposed SSRN methods give better restoration

than the GIHS, MTF_GLP methods according to the ground truth image. However, the AWLP and CNMF methods give worse restoration than the proposed SSRN method for the mountain and cropland land covers. The fusion results of GIHS and MTF_GLP methods give worse restoration performance in all the ground land covers than the other compared fusion methods and the proposed SSRN method. For the road land cover, the AWLP method, CNMF method and the proposed SSRN method have better fusion performance than the GIHS and MTF_GLP methods. The fusion result of the GIHS method has poor RGB contrast and does not give reasonable spectral revivification according to the ground truth image. The MTF_GLP method has poor fusion performance on ground land covers of roads, croplands and mountains. The HPF and SFIM methods have poor fusion performance on ground land covers of shadows, roads and croplands but have good fusion performance for mountain land cover. In conclusion, the proposed SSRN method gives better fusion performance on all the ground land covers than the compared fusion methods in the rough and gives better sharpness than the other fusion methods.

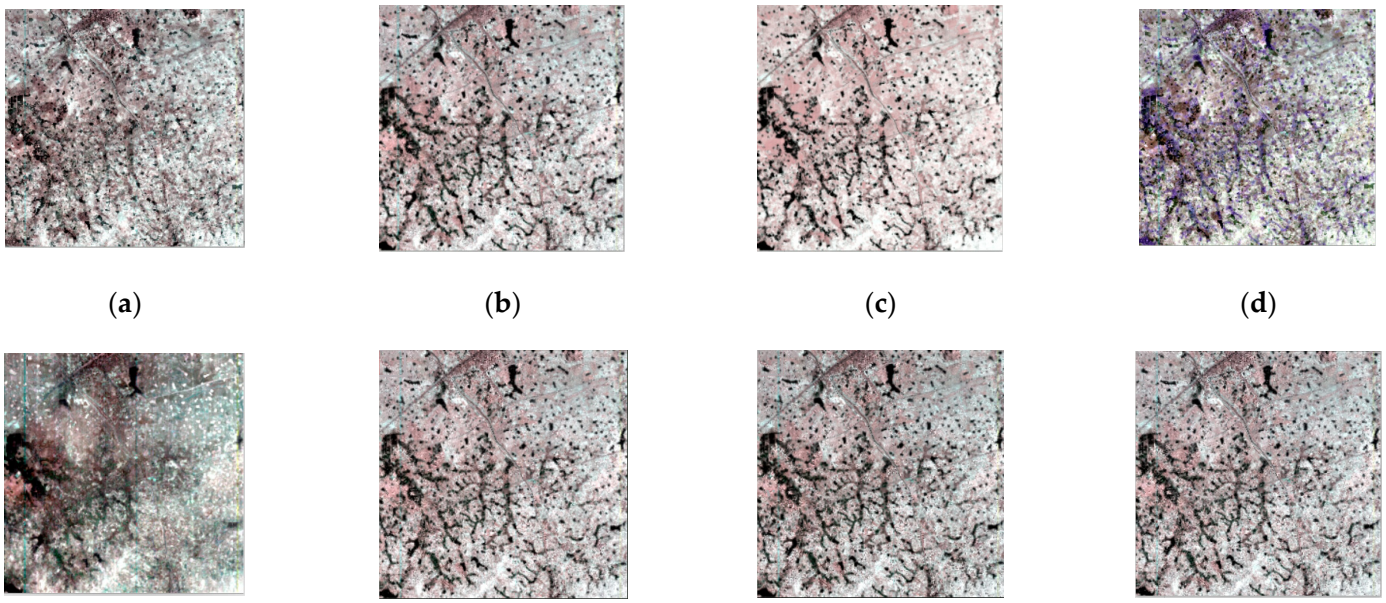


Figure 8. RGB images of the ground truth, the compared methods and the proposed SSRN method on Chaohu region dataset. (**a**) Ground truth, (**b**) AWLP, (**c**) CNMF, (**d**) GIHS, (**e**) MTF_GLP, (**f**) HPF, (**g**) SFIM and (**h**) SSRN.

The RGB images of the ground truth, the compared methods and the proposed SSRN method on the Dianchi region are shown in Figure 9. For the water land cover, the AWLP method, the GIHS method and the MTF_GLP method give poor restoration performance. Meanwhile, the CNMF and the proposed SSRN method have well fusion performance for the water land cover. By the way, the fusion result of the GIHS method has poor RGB contrast for the water land cover but weak spectral performance according to the ground truth images. For the jungle and mountain land covers, CNMF and MTF_GLP methods have poor fusion performance. Nevertheless, AWLP, GIHS and the proposed SSRN methods have better restoration performance than CNMF and MTF_GLP. The HPF and SFIM methods have good fusion performance on ground land covers of mountains and waters but poor performance for jungle land cover. In a word, the proposed SSRN method gives better fusion performance on all the ground land covers than the compared fusion methods.

In a word, with the experimental fusion results on the three datasets, in the compared fusion methods, the MTF_GLP method achieves better fusion results than the other com-

pared fusion methods, and GIHS achieves worse fusion results than the other compared fusion methods. Meanwhile, the proposed SSRN achieves better fusion results than all of the compared fusion methods.

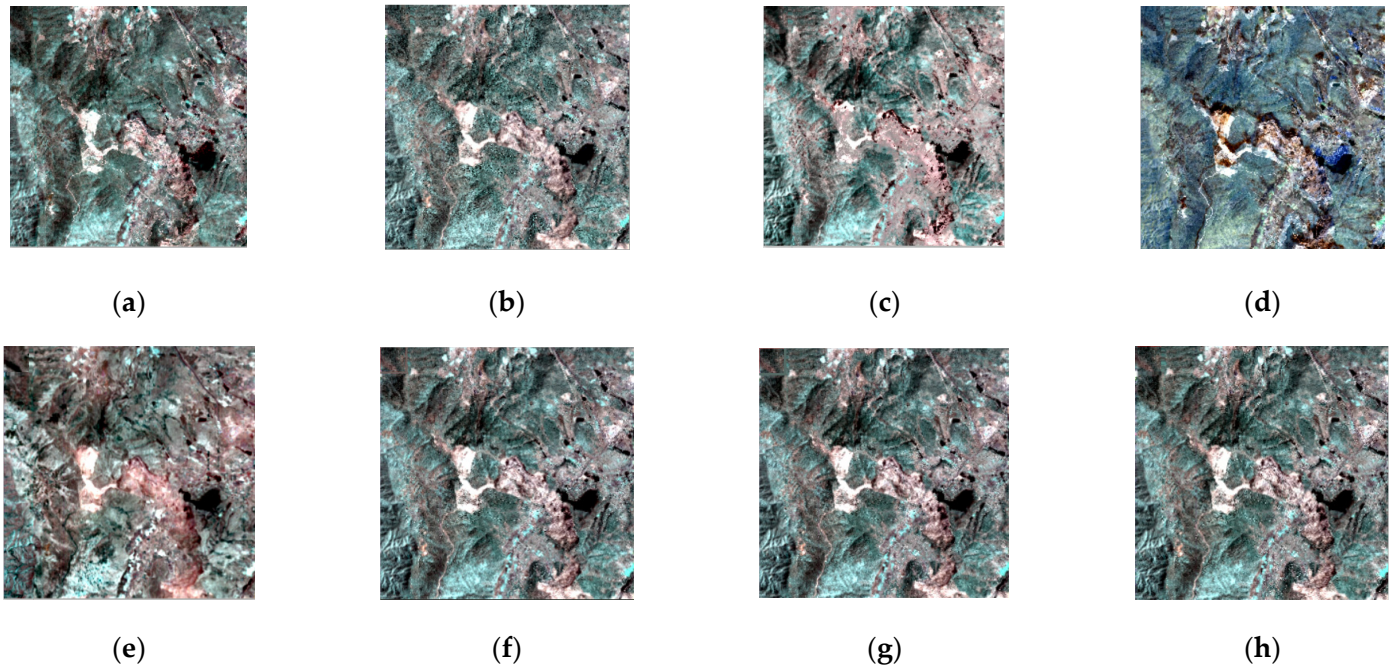


Figure 9. RGB images of the ground truth, the compared methods and the proposed SSRN method on Dianchi region dataset. (**a**) Ground truth, (**b**) AWLP, (**c**) CNMF, (**d**) GIHS, (**e**) MTF_GLP, (**f**) HPF, (**g**) SFIM and (**h**) SSRN.

4. Discussion

This section refers to the discussion of the experimental results, evaluating the fusion performance of the proposed SSRN method. A series of performance indexes are adopted to precisely evaluate the spectral and spatial performance of the fusion results. Eight representative performance indexes are utilized in this paper. The performance indexes are Root Mean Squared Error (RMSE), Spectral Angle Mapper (SAM), Spatial Correlation Coefficient (SCC), spectral curve comparison, Peak-Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), relative dimensionless global error in synthesis (ERGAS) and the Q metric [24]. The RMSE, SAM, PSNR, SSIM, ERGAS and Q metric performance indexes have the characteristics of value. Meanwhile, the SCC and spectral curve comparison have the characteristics of value on each band. Then, RMSE, SAM, PSNR, SSIM, ERGAS and Q metric are shown in a value graph, and the SCC and spectral curve comparison is shown in a line graph to connect the value in each band. Here, spectral curve comparison is comparing spectral curves of a pixel in the result of the fusion method with the spectral curve of the corresponding pixel in the original hyperspectral image. In our experiments, we compare the spectral curve on the (360, 360) pixel in the fusion result of all the compared and proposed methods with the spectral curve on the (30, 30) pixel in the corresponding hyperspectral image for all three datasets. Among the performance indexes, SAM and spectral curve are spectral quality metrics, SCC and SSIM are spatial quality metrics and RMSE, PSNR, ERGAS and Q metric are comprehensive spatial–spectral quality metrics. We utilize the SAM, spectral curve comparison, RMSE, PSNR, ERGAS and Q metric to prove the spectral information enhancement on the high spatial-resolution image. In addition, we utilize the SCC, SSIM, RMSE, PSNR, ERGAS and Q metric to prove the spatial information enhancement on the hyperspectral image. It is important to emphasize that SAM, SCC, PSNR, SSIM and Q metric are better when they are larger, and RMSE, SAM and ERGAS are better when they are smaller. For SCC, the performance is better when most SCC values

of bands are bigger than the other fusion methods. For spectral curve comparison, the performance is better when the spectral curve is near to the original spectral curve in the hyperspectral image.

Quality evolution for the compared and proposed fusion methods on the Baiyangdian region dataset is shown in Figure 10. For the RMSE index, the AWLP, MTF_GLP, HPF and the proposed SSRN methods achieve better performance than the CNMF, GIHS and SFIM methods to a great degree, while the proposed SSRN method achieves the best RMSE performance with RMSE lower than 100. For the SAM index, the AWLP, CNMF, MTF_GLP, HPF, SFIM and the proposed SSRN methods achieve better performance than the GIHS method to a great degree, while the SSRN method achieves the best performance. For the SCC index, GIHS and the proposed SSRN methods achieve better performance than the other compared methods, while the proposed SSRN method achieves the best SCC performance in most of the spectral bands. For the spectral curve comparison, the spectral curve of the proposed SSRN method is nearer than all of the compared fusion methods; then, the proposed SSRN method has better spectral curve comparison performance than the other compared fusion methods. For the PSNR index, AWLP, MTF_GLP, HFP and the proposed SSRN methods achieve better performance than the CNMF, GIHS and SFIM methods, and the proposed SSRN method achieves the best performance. For the SSIM index, GIHS, MTF_GLP, HPF and the proposed SSRN methods achieve better performance than the AWLP, CNMF and SFIM methods, and the proposed SSRN method achieves the best performance. For the ERGAS index, AWLP, MTF_GLP, HPF and the proposed SSRN methods achieve better performance than the CNMF, GIHS and SFIM methods, while the proposed SSRN method achieves the best performance. For the Q metric index, MTF_GLP and the proposed SSRN methods achieve better performance than the AWLP, CNMF, GIHS, HPF and SFIM methods, while the proposed SSRN method has the best performance.

Quality evaluation for the compared and proposed methods on the Chaohu region dataset is shown in Figure 11. For the RMSE index, the AWLP, MTF_GLP, HPF and the proposed SSRN methods achieve better performance than the CNMF, GIHS and SFIM methods to a great degree, while the proposed SSRN method achieves the best RMSE performance. For the SAM index, the AWLP, CNMF, MTF_GLP, HPF, SFIM and the proposed SSRN methods achieve better performance than the GIHS method to a great degree, while the SSRN method achieves the best performance. For the SCC index, the proposed SSRN method achieves better performance than all the compared methods in most of the spectral bands. For the spectral curve comparison, the spectral curve of the proposed SSRN method is nearer than all of the compared fusion methods; then, the proposed SSRN method has better spectral curve comparison performance than the other compared fusion methods. For the PSNR index, AWLP, MTF_GLP, HPF, SFIM and the proposed SSRN methods achieve better performance than CNMF and GIHS, and the proposed SSRN method achieves the best performance. For the SSIM index, GIHS, MTF_GLP, SFIM and the proposed SSRN methods achieve better performance than the AWLP, CNMF and HPF methods, while the proposed SSRN method achieves the best performance. For the ERGAS index, AWLP, CNMF, MTF_GLP, HPF and the proposed SSRN methods achieve better performance than the GIHS and SFIM methods, while the proposed SSRN method has the best performance. For the Q metric index, GIHS, MTF_GLP, HPF, SFIM and the proposed SSRN methods achieve better performance than the AWLP and CNMF methods, while the proposed SSRN method achieves the best performance.

Quality evaluation for the compared and proposed methods on the Dianchi region dataset is shown in Figure 12. For the RMSE index, the AWLP, MTF_GLP, HPF, SFIM and the proposed SSRN methods achieve better performance than the CNMF and GIHS methods to a great degree, and the proposed SSRN method achieves the best RMSE performance. For the SAM index, the AWLP, CNMF, MTF_GLP, HPF, SFIM and the proposed SSRN methods achieve better performance than the GIHS method to a great degree, and the SSRN method achieves the best performance. For the SCC index, the proposed SSRN method achieves better performance than all the compared methods in most of the spectral bands. For the

spectral curve comparison, AWLP, HPF, SFIM and the proposed SSRN methods achieve better performance than the CNMF, GIHS and MTF_GLP methods. The spectral curve of the proposed SSRN method is nearer than all of the compared fusion methods; then, the proposed SSRN method has better spectral curve comparison performance than the other compared fusion methods. For the PSNR index, AWLP, MTF_GLP, SFIM and the proposed SSRN methods achieve better performance than CNMF, GIHS and HPF methods, and the proposed SSRN method achieves the best performance. For the SSIM index, GIHS, MTF_GLP, HPF and the proposed SSRN methods achieve better performance than the AWLP, CNMF and SFIM methods, and the proposed SSRN method achieves the best performance. For the ERGAS index, MTF_GLP and the proposed SSRN methods achieve better performance than the AWLP, CNMF, GIHS, HPF, SFIM methods, while the proposed SSRN method achieves the best performance. For the Q metric index, GIHS, MTF_GLP, HPF, SFIM and the proposed SSRN methods achieve better performance than the AWLP and CNMF methods, while the proposed SSRN method achieves the best performance.

From discussing the performance indexes, some conclusions of the statistical reliability of the results can be concluded. With the spectral quality metrics of SAM, spectral curve comparison, RMSE, PSNR, ERGAS and Q metric indexes, the proposed SSRN method achieves the best performance; thus, the proposed SSRN method has strong reliability of spectral reconstruction ability. Meanwhile, with the spatial quality metrics of SCC, SSIM, RMSE, PSNR, ERGAS and Q metric indexes, the proposed SSRN method also achieves the best performance; thus, the proposed SSRN method also has strong reliability of spatial reconstruction ability. By the RMSE, ERGAS and Q metric indexes, the proposed SSRN method has strong reliability of holistic reconstruction similarity. By the SAM and spectral curve comparison indexes, the proposed SSRN method has strong reliability of spectral reconstruction similarity. By the SCC and SSIM indexes, the proposed SSRN method has strong reliability of spatial reconstruction similarity. By the PSNR index, the proposed SSRN method has reliability of more signal-to-noise-ratio reconstruction ability.

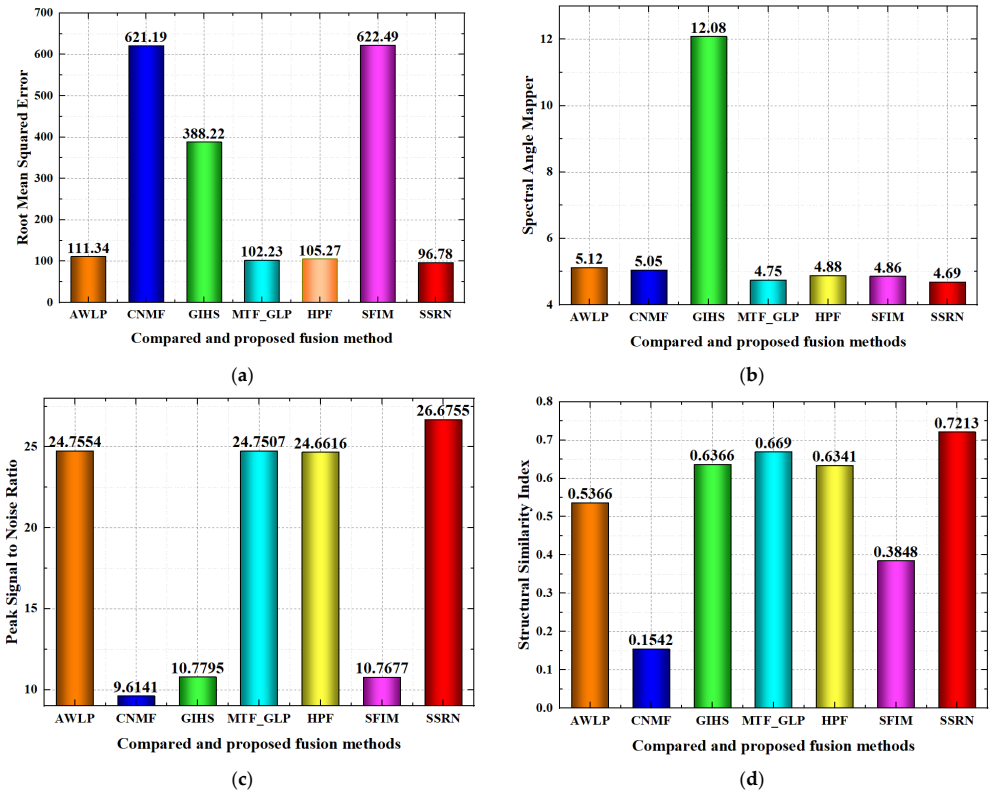


Figure 10. *Cont.*

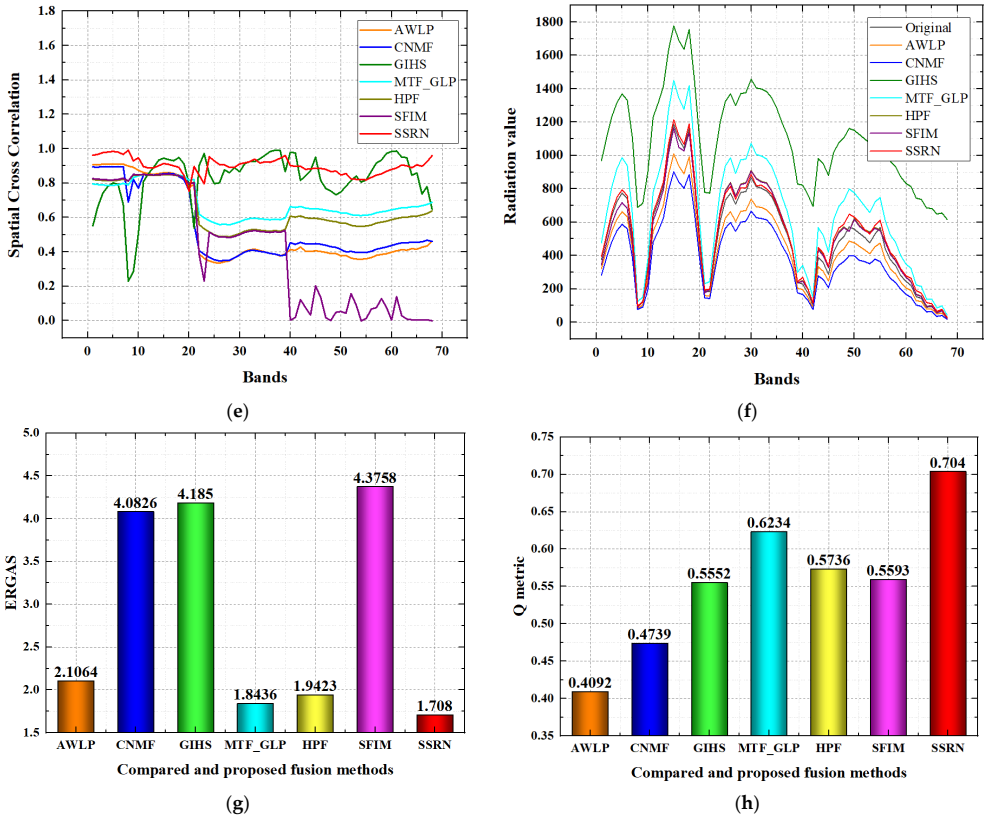


Figure 10. Quality evaluation for the compared and proposed fusion methods on the Baiyangdian region dataset. (**a**) RMSE, (**b**) SAM, (**c**) PSNR, (**d**) SSIM, (**e**) SCC, (**f**) spectral curve comparison, (**g**) ERGAS and (**h**) Q metric.

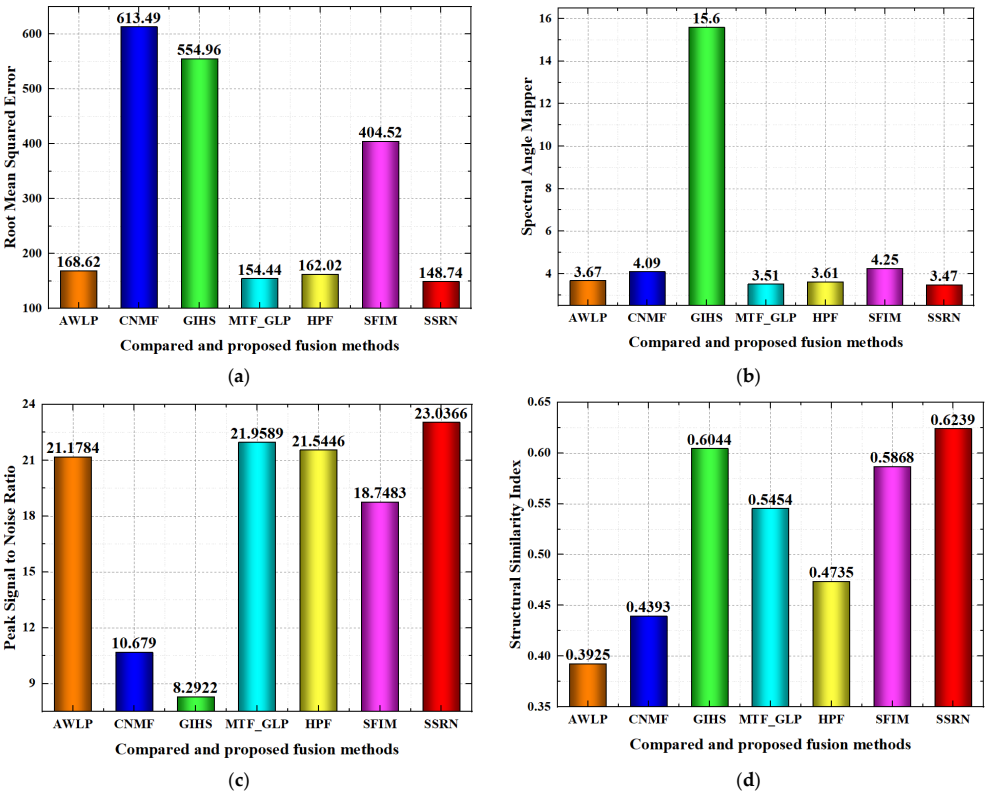


Figure 11. *Cont.*

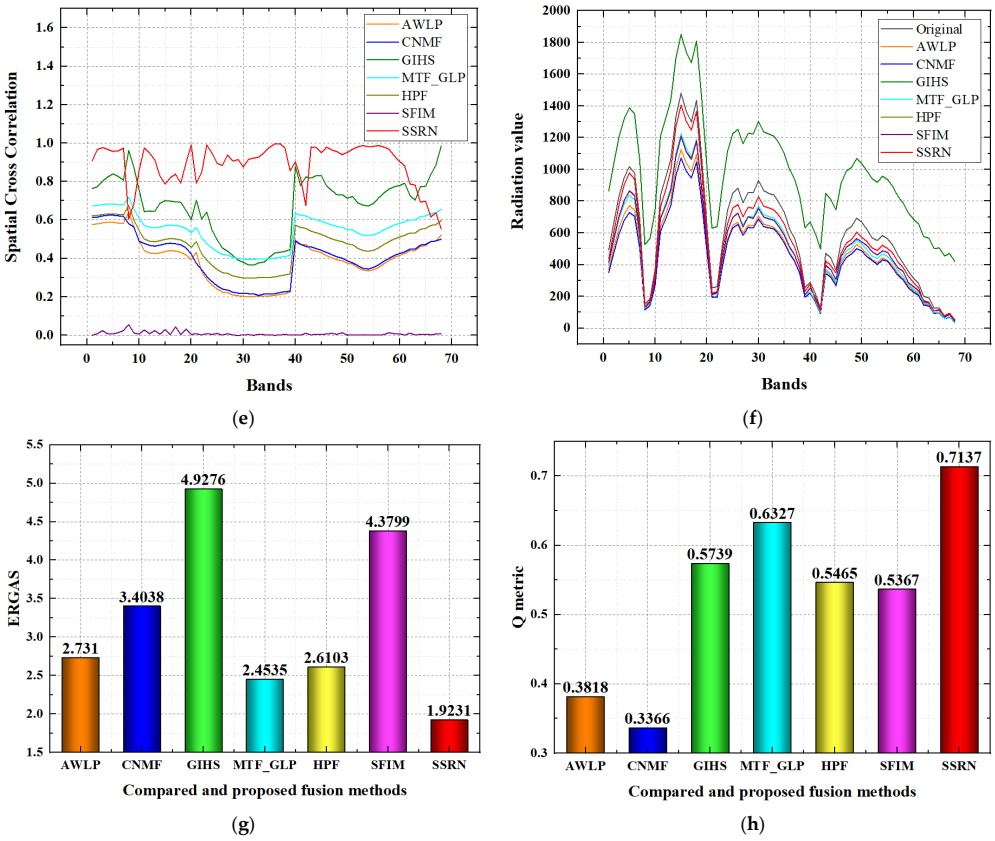


Figure 11. Quality evaluation for the compared and proposed fusion methods on the Chaohu region dataset. (**a**) RMSE, (**b**) SAM, (**c**) PSNR, (**d**) SSIM, (**e**) SCC, (**f**) spectral curve comparison, (**g**) ERGAS and (**h**) Q metric.

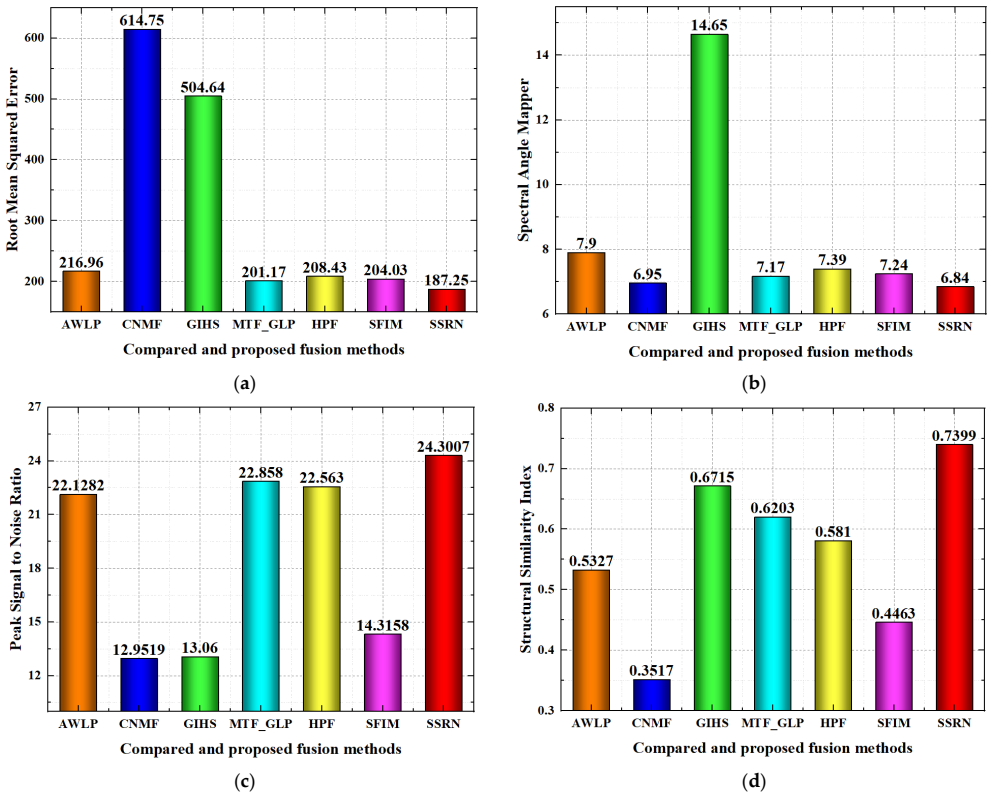


Figure 12. *Cont.*

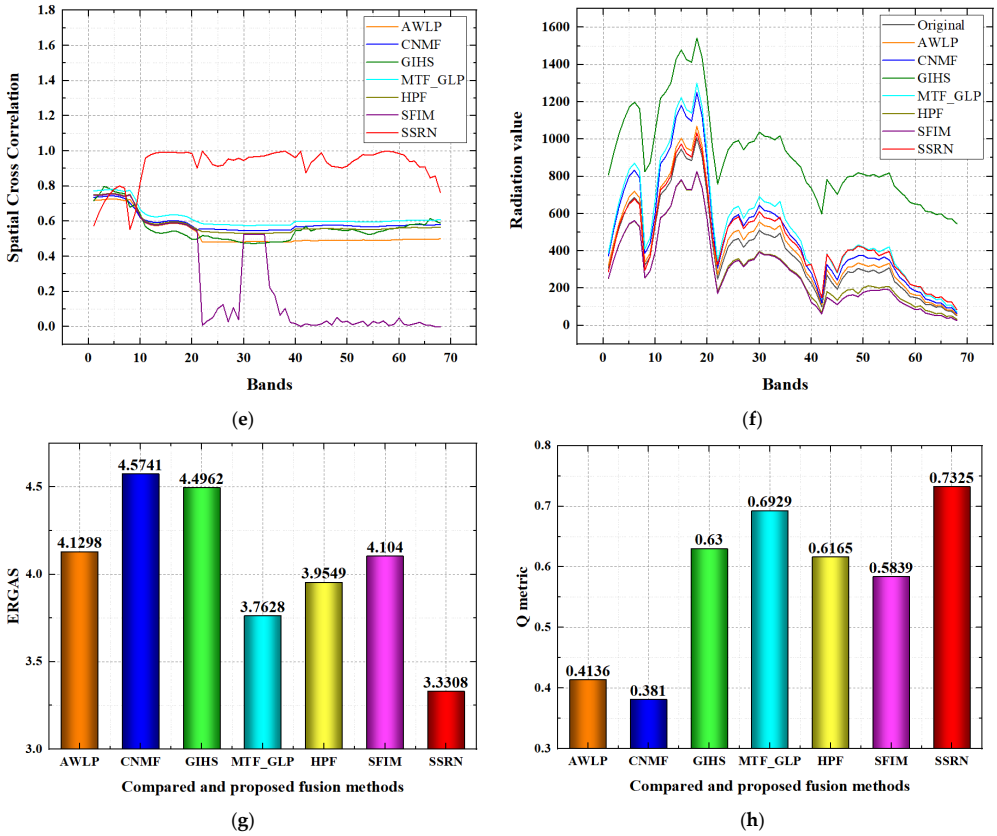


Figure 12. Quality evaluation for the compared and proposed fusion methods on the Dianchi region dataset. (**a**) RMSE, (**b**) SAM, (**c**) PSNR, (**d**) SSIM, (**e**) SCC, (**f**) spectral curve comparison, (**g**) ERGAS and (**h**) Q metric.

Here, we will provide the reflectance of the hyperspectral image and the SSRN fusion result on three experimental datasets. Reflectance of pixels (30, 30), (30, 270), (270, 30), (270, 270) in the hyperspectral image for each dataset is selected for the exhibition of reflectance. Accordingly, reflectance of pixels (360, 360), (360, 3240), (3240, 360), (3240, 3240) in the SSRN fusion result for each dataset is selected for the exhibition of reflectance.

Reflectance of four pixels in the Baiyangdian, Chaohu, Dianchi datasets are shown in Figures 13–15, respectively. From these figures, it can be seen that the fusion results of the proposed SSRN method have almost the same reflectance compared with the original hyperspectral images.

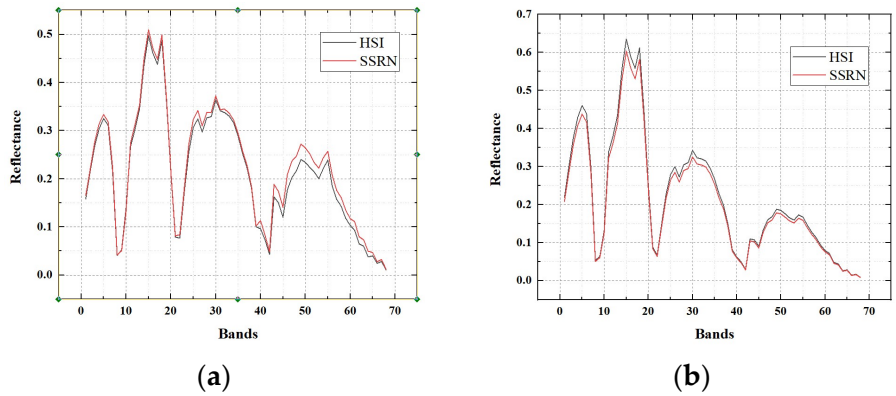


Figure 13. *Cont.*

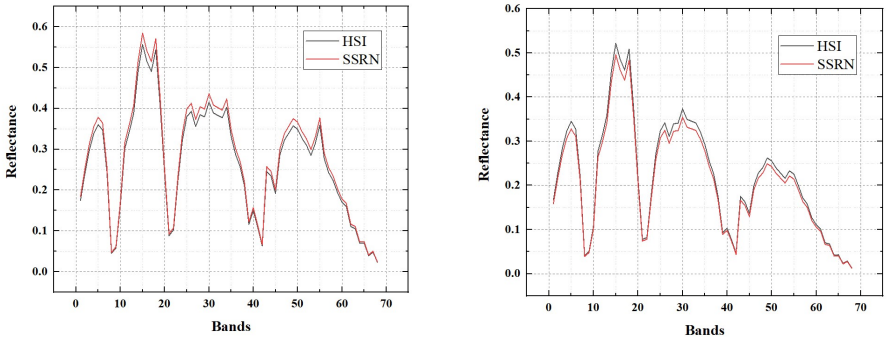


Figure 13. Reflectance of four pixels in Baiyangdian dataset for comparison of hyperspectral image and SSRN fusion result. (**a**) Pixel (30, 30) in hyperspectral image versus pixel (360, 360) in SSRN fusion result. (**b**) Pixel (30, 270) in hyperspectral image versus pixel (360, 3240) in SSRN fusion result. (**c**) Pixel (270, 30) in hyperspectral image versus pixel (3240, 360) in SSRN fusion result. (**d**) Pixel (270, 270) in hyperspectral image versus pixel (3240, 3240) in SSRN fusion result.

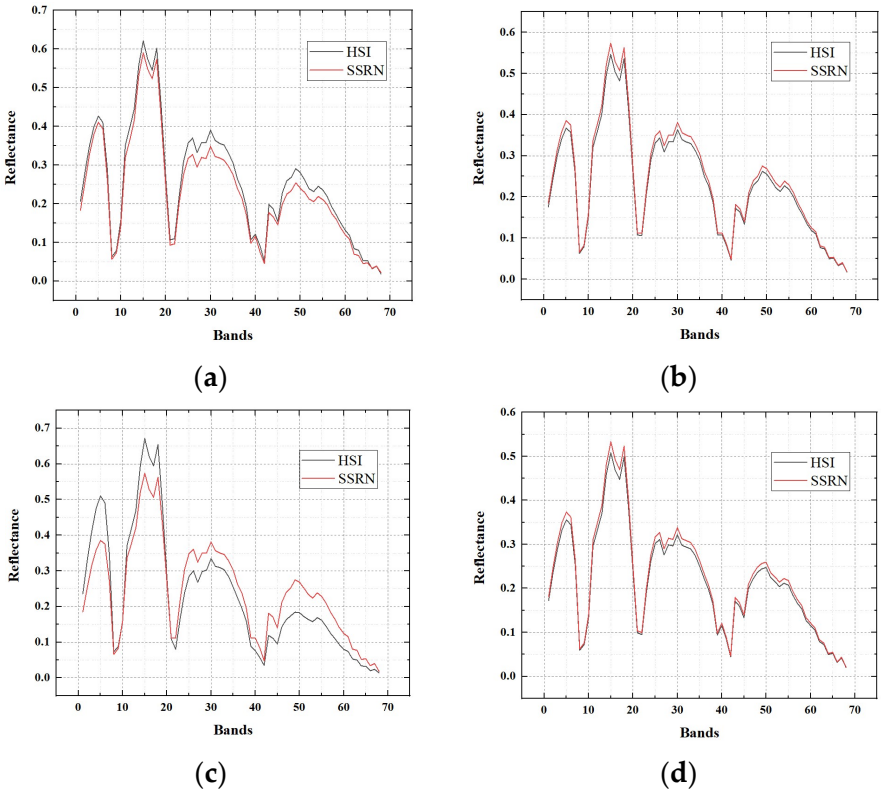


Figure 14. Reflectance of four pixels in Chaohu dataset for comparison of hyperspectral image and SSRN fusion result. (**a**) Pixel (30, 30) in hyperspectral image versus pixel (360, 360) in SSRN fusion result. (**b**) Pixel (30, 270) in hyperspectral image versus pixel (360, 3240) in SSRN fusion result. (**c**) Pixel (270, 30) in hyperspectral image versus pixel (3240, 360) in SSRN fusion result. (**d**) Pixel (270, 270) in hyperspectral image versus pixel (3240, 3240) in SSRN fusion result.

In terms of the discussion of the quality evaluation for the compared and proposed SSRN methods on the three datasets, it can be seen that the MTF_GLP method achieves the best evaluation performance in the compared fusion methods, while the GIHS method achieves worse fusion results than the other compared fusion methods. Meanwhile, the proposed SSRN method achieves better evaluation performance than all of the compared fusion methods.

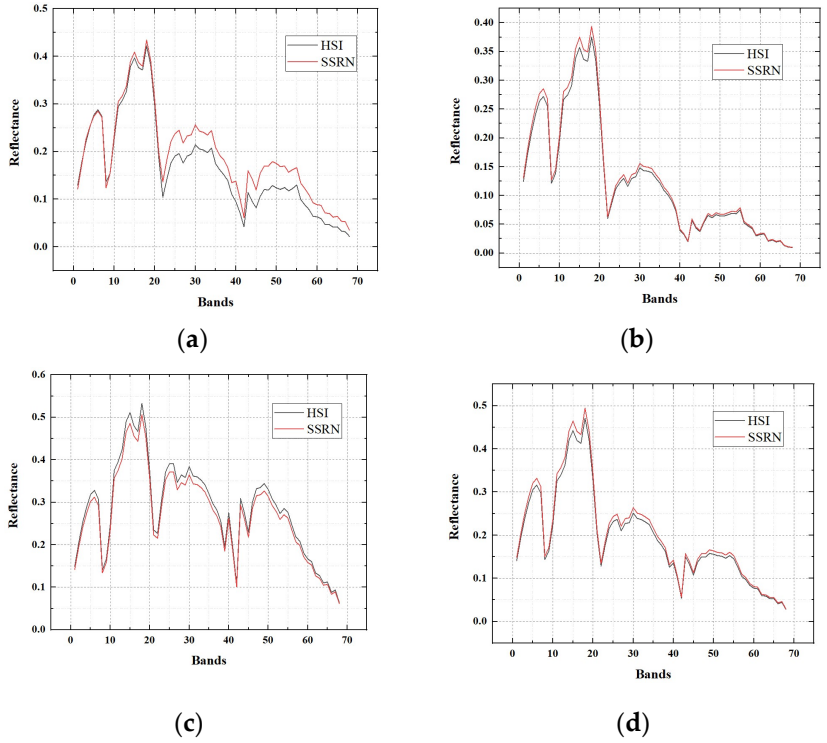


Figure 15. Reflectance of four pixels in Dianchi dataset for comparison of hyperspectral image and SSRN fusion result. (**a**) Pixel (30, 30) in hyperspectral image versus pixel (360, 360) in SSRN fusion result. (**b**) Pixel (30, 270) in hyperspectral image versus pixel (360, 3240) in SSRN fusion result. (**c**) Pixel (270, 30) in hyperspectral image versus pixel (3240, 360) in SSRN fusion result. (**d**) Pixel (270, 270) in hyperspectral image versus pixel (3240, 3240) in SSRN fusion result.

5. Conclusions

In this paper, a deep network fusion model is established for the intelligent fusion process of hyperspectral remote sensing images and panchromatic remote sensing images. This method firstly establishes the spectral–spatial deep feature branch. Secondly, this method establishes an enhanced multi-scale residual network of a spatial feature branch and residual network of a spectral feature branch. Finally, this method establishes spectral–spatial deep feature simultaneity. The latter two operations aim at adjusting the deep features learned from the deep learning network to make the deep features more representational and integrated. The proposed method is compared with the AWLP, CNMF, GIHS, MTF_GLP, HPF and SFIM methods. The experimental results suggest that the proposed method can achieve competitive spatial quality compared to existing methods and recover most of the spectral information that the corresponding sensor would observe with the highest spatial resolution. The method proposed in this paper treats hyperspectral and panchromatic remote sensing images as two independent units. The original deep features extracted from hyperspectral and panchromatic images are self-existent. In the future, we will try to further study uniformly existing hyperspectral and panchromatic images and extract deep features from the unity of hyperspectral and panchromatic images.

Author Contributions: Methodology, R.Z.; supervision, S.D. All authors have read and agreed to the published version of the manuscript.

Funding: National Key Research and Development Program of China: 2021YFE010519; National Key Research and Development Program of China: 2021YFE0117100.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data supported by Capital Normal University.

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

References

1. Ehlers, M. Multisensor image fusion techniques in remote sensing. *ISPRS J. Photogramm. Remote Sens.* **1991**, *46*, 19–30. [CrossRef]
2. Li, S.; Yin, H.; Fang, L. Remote Sensing Image Fusion via Sparse Representations Over Learned Dictionaries. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4779–4789. [CrossRef]
3. Zheng, S.; Shi, W.Z.; Liu, J.; Tian, J. Remote Sensing Image Fusion Using Multiscale Mapped LS-SVM. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1313–1322. [CrossRef]
4. Shaw, G.; Manolakis, D. Signal processing for hyperspectral image exploitation. *IEEE Signal Process. Mag.* **2002**, *19*, 12–16. [CrossRef]
5. Choi, M.; Kim, R.Y.; Nam, M.R.; Kim, H.O. Fusion of multispectral and panchromatic Satellite images using the curvelet transform. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 136–140. [CrossRef]
6. Churchill, S.; Randell, C.; Power, D.; Gill, E. Data fusion: Remote sensing for target detection and tracking. In Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2004), Anchorage, AK, USA, 20–24 September 2004.
7. Benediktsson, J.A.; Pesaresi, M.; Arnason, K. Classification and Feature Extraction for Remote Sensing Images from Urban Areas based on Morphological Transformations. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1940–1949. [CrossRef]
8. Huete, A.R.; Escadafal, R. Assessment of Biophysical Soil Properties through Spectral Decomposition Techniques. *Remote Sens. Environ.* **1991**, *35*, 149–159. [CrossRef]
9. Federico, C. A Review of Data Fusion Techniques. *Sci. World J.* **2013**, *2013*, 704504.
10. Carper, W.J.; Lillesand, T.M.; Kiefer, P.W. The Use of Intensity-Hue-Saturation Transformations for Merging SPOT Panchromatic and Multispectral Image Data. *Photogramm. Eng. Remote Sens.* **1990**, *56*, 459–467.
11. Tu, T.M.; Su, S.C.; Shyu, H.C.; Huang, P.S. A New Look at IHS-Like Image Fusion Methods. *Inf. Fusion* **2001**, *2*, 177–186. [CrossRef]
12. Chavez, P.S.; Sides, S.C.; Anderson, J.A. Comparison of Three Different Methods to Merge Multiresolution and Multispectral Data: Landsat TM and SPOT Panchromatic. *Photogramm. Eng. Remote Sens.* **1991**, *57*, 265–303.
13. Kwarteng, P.; Chavez, A. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens.* **1989**, *55*, 339–348.
14. Shettigara, V.K. A Generalized Component Substitution Technique for Spatial Enhancement of Multispectral Images Using a Higher Resolution Data Set. *Photogram Eng. Remote Sens.* **1992**, *58*, 561–567.
15. Shah, V.P.; Younan, N.H.; King, R.L. An Efficient Pan-Sharpening Method via a Combined Adaptive PCA Approach and Contourlets. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1323–1335. [CrossRef]
16. Pandey, P.C.; Tate, N.J.; Balzter, H. Mapping Tree Species in Coastal Portugal Using Statistically Segmented Principal Component Analysis and Other Methods. *Sens. J. IEEE* **2014**, *14*, 4434–4441. [CrossRef]
17. Aiazzi, B.; Baronti, S.; Selva, M. Improving Component Substitution Pansharpening Through Multivariate Regression of MS+Pan Data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239. [CrossRef]
18. Baronti, S.; Aiazzi, B.; Selva, M.; Garzelli, A.; Alparone, L. A Theoretical Analysis of the Effects of Aliasing and Misregistration on Pansharpened Imagery. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 446–453. [CrossRef]
19. Mallat, S.G. A Theory of Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693.
20. Sridhar, V.; Reddy, P.R.; Saiteja, B. Image Enhancement through Discrete and Stationary Wavelet Transforms. *Int. J. Eng. Sci. Res. Technol.* **2014**, *3*, 137–147.
21. Burt, P.J.; Adelson, E.H. The Laplacian Pyramid as a Compact Image Code. *IEEE Trans. Commun.* **2003**, *31*, 532–540. [CrossRef]
22. Starck, J.L.; Fadili, J.M.; Murtagh, F. The Undecimated Wavelet Decomposition and its Reconstruction. *IEEE Trans. Image Process.* **2007**, *16*, 297–309. [CrossRef]
23. Do, M.N.; Vetterli, M. The Contourlet Transform: An Efficient Directional Multiresolution Image Representation. *IEEE Trans. Image Process.* **2005**, *14*, 2091–2106. [CrossRef]
24. Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A Critical Comparison Among Pansharpening Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 2565–2586. [CrossRef]
25. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. A New Pansharpening Algorithm Based on Total Variation. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 318–322. [CrossRef]
26. Joshi, M.; Jalobeanu, A. MAP Estimation for Multiresolution Fusion in Remotely Sensed Images Using an IGMRF Prior Model. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1245–1255. [CrossRef]
27. Huang, B.; Song, H.; Cui, H.; Peng, J.; Xu, Z. Spatial and Spectral Image Fusion Using Sparse Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 1693–1704. [CrossRef]
28. Dong, W.; Fu, F.; Shi, G.; Cao, X.; Wu, J.; Li, G.; Li, X. Hyperspectral Image Super-Resolution via Non-Negative Structured Sparse Representation. *IEEE Trans. Image Process.* **2016**, *25*, 2337–2352. [CrossRef]
29. Wei, Q.; Bioucas-Dias, J.; Dobigeon, N.; Tourneret, J.Y. Hyperspectral and Multispectral Image Fusion based on a Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3658–3668. [CrossRef]

30. Simoes, M.; Bioucas-Dias, J.; Almeida, L.B.; Chanussot, J. A Convex Formulation for Hyperspectral Image Superresolution via Subspace-Based Regularization. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3373–3388. [CrossRef]
31. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [CrossRef]
32. Nascimento, J.M.P.; Dias, J.M.B. Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910. [CrossRef]
33. Bioucas-Dias, J.M. A Variable Splitting Augmented Lagrangian Approach to Linear Spectral Unmixing. In Proceedings of the 2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Grenoble, France, 26–28 August 2009.
34. Soysa, S.D.; Manawadu, S.; Sendanayake, S.; Athipola, U. Computer Vision, Deep Learning and IOT Based Enhanced Early Warning system for the safety of Rail Transportation. In Proceedings of the International Conference on Advances in Computing and Technology, Virtual Online, 28 November 2020.
35. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [CrossRef]
36. Li, H. Deep Learning for Natural Language Processing: Advantages and Challenges. *Natl. Sci. Rev.* **2018**, *5*, 22–24. [CrossRef]
37. Ahmad, F.; Abbasi, A.; Kitchens, B.; Adjeroh, D.A.; Zeng, D. Deep Learning for Adverse Event Detection from Web Search. *IEEE Trans. Knowl. Data Eng.* **2020**, *99*, 1. [CrossRef]
38. Silva, L.G.D.; Guedes, L.L.V.; Colcher, S. Using Deep Learning to Recognize People by Face and Voice. In *Anais Estendidos do XXV Simpósio Brasileiro de Sistemas Multimídia e Web*; SBC: Porto Alegre, Brazil, 2019.
39. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-adaptive CNN-based pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5443–5457. [CrossRef]
40. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by Convolutional Neural Networks. *Remote Sens.* **2016**, *8*, 594. [CrossRef]
41. Wei, Y.; Yuan, Q.; Shen, H.; Zhang, L. Boosting the Accuracy of Multispectral Image Pansharpening by Learning a Deep Residual Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1795–1799. [CrossRef]
42. Yuan, Q.; Wei, Y.; Meng, X. A Multiscale and Multidepth Convolutional Neural Network for Remote Sensing Imagery Pan-sharpening. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 978–989. [CrossRef]
43. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face Recognition: A Convolutional Neural Network Approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113. [CrossRef]
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
45. Yokoya, N.; Yairi, T.; Iwasaki, A. Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 528–537. [CrossRef]
46. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A. Context-Driven Fusion of High Spatial and Spectral Resolution Images based on Oversampled Multiresolution Analysis. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 300–312. [CrossRef]
47. Zhou, X.; Liu, J.; Liu, S.; Cao, L.; Zhou, Q.; Huang, H. A GIHS-based Spectral Preservation Fusion Method for Remote Sensing Images Using Edge Restored Spectral Modulation. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 16–27. [CrossRef]
48. Otazu, X.; González-Audícana, M.; Fors, O.; Núñez, J. Introduction of sensor spectral response into image fusion methods. Application of wavelet-based methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2376–2385. [CrossRef]