



Article

Multispectral and Hyperspectral Image Fusion Based on Joint-Structured Sparse Block-Term Tensor Decomposition

Hao Guo ¹, Wenxing Bao ^{1,*}, Wei Feng ², Shasha Sun ¹, Chunhui Mo ¹ and Kewen Qu ¹

¹ School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China; 20207169@stu.nun.edu.cn (H.G.); 20217392@stu.nun.edu.cn (S.S.); 20207173@stu.nun.edu.cn (C.M.); qukewen@nun.edu.cn (K.Q.)

² The School of Electronic Engineering, Xidian University, Xi'an 710071, China; wfeng@xidian.edu.cn

* Correspondence: baowenxing@nun.edu.cn

Abstract: Multispectral and hyperspectral image fusion (MHF) aims to reconstruct high-resolution hyperspectral images by fusing spatial and spectral information. Unlike the traditional canonical polyadic decomposition and Tucker decomposition models, the block-term tensor decomposition model is able to improve the quality of fused images using known endmember and abundance information. This paper presents an improved hyperspectral image fusion algorithm. Firstly, the two abundance matrices are combined into a single bulk matrix to promote structural sparsity by introducing the $L_{2,1}$ -norm to eliminate the scaling effects present in the model. Secondly, the counter-scaling effect is eliminated by adding the L_2 -norm to the endmember matrix. Finally, the chunk matrix and the endmember matrix are coupled together, and the matrix is reorganized by adding the $L_{2,1}$ -norm to the matrix to facilitate chunk elimination and solved using an extended iterative reweighted least squares (IRLS) method, focusing on the problem of the inability to accurately estimate the tensor rank in the chunk-term tensor decomposition model and the noise/artifact problem arising from overestimation of rank. Experiments are conducted on standard and local datasets, and the fusion results are compared and analyzed in four ways: visual result analysis, metric evaluation, time of the algorithm, and classification results, and the experimental results show that the performance of the proposed method is better than the existing methods. An extensive performance evaluation of the algorithms is performed by conducting experiments on different datasets. The experimental results show that the proposed algorithm achieves significant improvements in terms of reconstruction error, signal-to-noise ratio, and image quality compared with the existing methods. Especially in the case of a low signal-to-noise ratio, the proposed algorithm shows stronger robustness and accuracy. These results show that the proposed algorithm has significant advantages in dealing with multispectral high-resolution hyperspectral data.

Keywords: image fusion; tensor decomposition; structured sparse; iteratively reweighted least squares



Citation: Guo, H.; Bao, W.; Feng, W.; Sun, S.; Mo, C.; Qu, K. Multispectral and Hyperspectral Image Fusion Based on Joint-Structured Sparse Block-Term Tensor Decomposition. *Remote Sens.* **2023**, *15*, 4610. <https://doi.org/10.3390/rs15184610>

Academic Editor: Shih-Yu Chen

Received: 12 July 2023

Revised: 10 September 2023

Accepted: 15 September 2023

Published: 19 September 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the field of remote sensing, hyperspectral imaging, and multispectral imaging systems occupy an important position. Hyperspectral imaging has been widely used in mineral detection, environmental detection [1], climate monitoring [2], and classification [3]. Hyperspectral images are obtained by hyperspectral sensors capturing sunlight reflected from an object or scene. During the acquisition of the image, it is affected by the light, resulting in a reduced spatial resolution of the image [4], but with rich spectral information. In contrast, multispectral images have only a few bands, but the images are rich in spatial information. In the past two decades, a lot of efforts have been made to obtain hyperspectral images with high spatial resolution, and image fusion is the most economical, applicable, and effective means among all methods. The fusion mainly extracts effective high spatial resolution information from source images such as multispectral images and panchromatic

images (PAN) and incorporates it into hyperspectral images. Fusion into hyperspectral images to obtain hyperspectral images with high spatial resolution helps to accurately identify observation scenes with high spatial resolution, etc. The most common fusion method is to obtain high spatial resolution hyperspectral images (HRHS) by fusing high spatial resolution multispectral images (HR-MSI) and low spatial resolution hyperspectral images (LR-HSI). At the same time, when an image is reduced in size, the redistribution of pixel information and interpolation algorithms results in a loss of detail or distortion of the original image, a process known as the “scaling effect”. Conversely, when a low-resolution image is scaled up to a higher resolution, this is often accompanied by a loss of information, and the image may appear blurred, distorted, or show unrealistic details, a process known as the “counter-scaling effect”.

In the early studies of fusion algorithms, sharpening methods were mainly used to fuse PANs and MSIs possess high spatial resolutions [5]. In 2019, Meng et al. [6] reviewed the development history of pansharpening and divided it into two main categories: component replacement (CS) [7,8] and multi-resolution analysis (MRA) [9–11], among which the more famous methods are based on hue–intensity–saturation (HIS) [12], wavelet-based techniques [13,14], principal component analysis (PCA) [15], the Laplace pyramid [16], etc. Yokoya et al. [17] found that sharpening fusion methods does not improve the spatial resolution of each spectral band very well after conducting extensive experiments. There is also a growing interest in the study of multispectral and hyperspectral image fusion (MHF) algorithms. A method often referred to as the multi-sharpening process involves the most popular techniques, including matrix factorization [18–24] and tensor decomposition [25–32]. In addition, some new research results have appeared in recent years. For example, in 2020, Feng et al. [33] proposed a two-by-two matching based on a spatial–spectral graphical regularization operator for the hyperspectral band selection method, which encodes its geometric structure by constructing a spatial–spectral information map using spatial–spectral adjacencies. In 2021, Huang et al. [34] proposed a subspace clustering algorithm for hyperspectral images, which improves the model’s noise immunity by introducing an adaptive spatial regularization method. In 2022, Qin et al. [35] proposed a multi-focus image fusion method based on sparse representation (DWT-SR) by decomposing multiple frequency bands and shortening the algorithm’s running time by utilizing GPU parallel computing, which features high contrast and richer spatial details. In 2022, Zhang et al. [36] proposed a sparse demixing method in order to avoid inaccuracy in extracting the endmember information and to retain the original domain super-pixel inter-correlation and spectral variability in the approximated image domain. In 2023, Y et al. [37] designed a new attentional reinforcement learning network to accomplish the spatio-temporal prediction of wind power.

The matrix-based decomposition method is widely used in MHF problems because of its clear motivation and excellent performance. After matrix decomposition, the potential factors have a strong physical meaning, such as endmember and abundance are non-negative [22,38,39]. In 2011, Yokoya et al. [18] used a coupled non-negative matrix decomposition (CNMF) unmixing method to decompose hyperspectral and multispectral data into endmember matrices and abundance matrices alternatively; the cost function in the model is non-convex, resulting in an unsatisfactory algorithm fusion. In the same year, Kawakami et al. [40] decomposed the target image HRHS into a basis matrix and a sparse matrix and then recombined the spectral basis matrix of LR-HSIs and the coefficient matrix of HR-MSIs to obtain them. In 2014, Simoes et al. [21] introduced total variation (TV) regularization in a subspace-based HRHS fusion framework, which effectively eliminated the noise. In 2015, Lanaras et al. [41] completed the acquisition of HRHSs by jointly unmixing two input images into a pure reflection spectral matrix and a mixed sparse matrix. In 2018, Kahraman et al. [42] proposed a graph regularization $L_{1/2}$ -sparse constrained non-negative matrix decomposition-based image fusion method that enhances the fusion performance by using the $L_{1/2}$ and streamwise regularization to enhance the unmixing. In 2021, Mu et al. [43] introduced a spectral-constrained regularization term to avoid spectral distortion and maintain spectral integrity based on the coupled non-negative matrix

decomposition (CNMF) algorithm. In 2022, Khader et al. [44] proposed a new non-negative matrix factorization decomposition-inspired deep unfolding network (NMF-DuNet) to fuse LR-HSIs and HR-MSIs using a non-negative coefficient matrix and orthogonal transformation coefficient constraints. In 2022, Han et al. [45] proposed a fusion model based on spatial nonlocal similarity and low-rank prior learning to solve the problem of the existing methods that cannot simultaneously consider the spatial and spectral domain structure of hyperspectral image cubes; however, these problems still exist, such as the partial loss of spectral correlation, and the complexity of the algorithm is high. In 2023, Lin et al. [46] proposed a highly interpretable deep HSI-MSI fusion method based on a 2D image CNN-based Gaussian denoiser, which can fuse the LR-HSIs and HR-MSIs under a known imaging system and achieve good fusion performance.

While there are many other improved algorithms based on matrix decomposition, they all require the conversion of 3D data into a 2D data format, resulting in difficult exploitation of the spatial–spectral correlation in LR-HSIs. In contrast, the fusion method based on tensor decomposition can effectively avoid such defects and becomes a new research hotspot for the MHF problem. It regards LR-HSIs and HR-MSIs as a 3D data cube consisting of two spatial dimensions and one spectral dimension (i.e., spatial \times spatial \times spectral), where the spectral dimension is the key to identifying objects in the scene. Based on this fact, in 2018, Kanatsoulis et al. [47] proposed a coupled tensor decomposition-based (CTD) framework to solve the MHF problem, and, using canonical polyadic decomposition (CPD) uniqueness, they proved that HRHSs can be recovered under mild conditions. Xu et al. [48] constructed the nonlocally similar patch tensor of LR-HSIs based on HR-MSIs, calculated the smoothing order of all patches used for clustering, preserved the relationship between HRHSs and HR-MSIs by CPD, and proposed a multispectral and hyperspectral fusion algorithm based on the CPD decomposition of the nonlocally coupled tensor. In 2020, Prévost et al. [25] extended a similar idea to the Tucker decomposition for solving MHF problems with better results. In the same year, Xu et al. [27] proposed a one-way TV-based fusion method, which decomposes the HRHS into a core tensor and three modal dictionary forms, and imposes the L_1 -norm and the one-way TV characterizing sparsity and piecewise smoothness, respectively. In 2021, Peng et al. [49] introduced a global gradient sparse nonlocal low-rank tensor decomposition model to characterize the nonlocal and global properties of HRHSs, which can effectively capture the spatial and spectral similarity of the nonlocal cube, and introduced spatial–spectral total variation (SSTV) regularization to reconstruct HRHSs on the nonlocal domain low-rank cube, thus obtaining global spatial piecewise smoothness and spectral consistency. In 2022, Prévost et al. [50] proposed a fast coupling method based on the Tucker approximation into this work to take full advantage of the coupling information contained in hyperspectral and multispectral images of the same scene. After a large number of experiments, CPD and Tucker have been proven to have excellent recoverability on the MHF problem, and the frameworks of both have shown excellent performance; however, the failure to utilize the physical interpretation of the potential factors makes it difficult for both to utilize the a priori information to improve the algorithmic fusion effect. In addition, a new tensor decomposition, called tensor ring (TR) decomposition, is proposed in [51], which decomposes higher-rank tensors into cyclically contracted third-rank tensor sequences. This method inherits the representation capability of CPD and extends the low-rank exploration capability of Tucker decomposition. In 2022, Chen et al. [52] proposed a factorially smoothed TR decomposition model, which solved the problem that the spatial–spectral continuity of HRHSs could not be preserved in the existing TR decomposition model; however, the algorithm has high requirements on hardware equipment in the case of a large amount of spectral data. In summary, CPD and Tucker decomposition are not necessarily better than the methods for low-rank matrix estimation, and TR decomposition is not necessarily the best choice.

In contrast, linear mixed models (LMM) using matrixed LR-HSI and HR-MSI low-rank structures can circumvent such problems, and their endmembers and abundances possess a strong physical interpretation, which can be further improved by imposing structured

constraints or regularization on them, especially in the presence of noise, where this information is particularly important. In 2018, Li et al. [32] demonstrated, under a relatively restricted condition, that the coupled matrix decomposition model is able to identify recovered HRHSs. For example, if the abundance matrix of an HR-MSI is sparse, i.e., under tensor notation, the LMM can be rewritten to approximate the 3D spectral data as the sum of the outer products of the endmembers (vectors) and the abundance maps (matrices). This approach can fully match the matrix-vector third-order tensor decomposition, which decomposes the tensor into the sum of the component tensors, each of which in turn consists of matrix-vector outer products [53]. The matrix-vector third-order tensor factorization can also be regarded as the block-term tensor decomposition (BTD). Therefore, matrices and vectors in the BTD can be used as abundance maps and vectors, respectively, with non-negativity and coupling properties. In 2019, Zhang et al. [54] proposed a coupled non-negative BTD model [55–57]. The BTD was used to decompose the spectral images to take full advantage of the physical significance of their potential factors, but the use of regularization to improve the algorithm's performance and application in multiple scenarios was not considered. In 2021, Ding et al. [58] reformulated the recovery HRHS process as a special block-term tensor model, imposing constraints and regularization to improve the performance of the algorithm, but failed to take into account the effects between individual abundance matrices, and there were limitations in the experimental scenario. In 2022, Guo et al. [59] proposed a coupled non-negative block-term tensor decomposition algorithm based on regularization. Consider the sparsity and smoothness of the individual abundance and the endmember matrix; to these, add the L_1 -norm and bidirectional TV. However, the L_1 -norm has a non-smoothness property and is not the best choice for the regularizer. The algorithm takes too long to run. Therefore, a joint-structured sparse block-term tensor decomposition model (JSSL1) based on joint-structured sparsity is proposed by introducing the $L_{2,1}$ -norm to promote the structured sparsity of the block matrix composed of potential factors and eliminate the scaling effects present in the model, The L_2 -norm is introduced to the endmember to counteract the counter-scaling effects in the decomposition model. Finally, the block matrix and endmember are coupled together, and the $L_{2,1}$ -norm is imposed on them to promote the elimination of the block matrix. The contributions of the proposed algorithm are as follows.

- During the reconstruction of an HRHS, in order to promote the sparsity of the abundance matrix and counteract the scaling effect existing in the abundance matrix during image recovery, the two abundance matrices are superimposed to obtain the chunk matrix, and the blending parametrization $L_{2,1}$ -norm is added to it.
- In the process of reconstructing an HRHS, the counter-scaling effect of the endmember vector is considered and avoided by adding the L_2 -norm to the endmember vector.
- In the block-term tensor factorization model, the accuracy of the rank is a key factor that is unknown a priori. Inaccurate rank estimation may cause noise or artifacts to interfere with the experimental results. In order to solve this problem, this paper introduces the reweighted coupling of the block matrix and the endmember matrix and introduces the mixing norm. Through this method, the influence of the inaccuracy of the rank estimation can be reduced while maintaining flexibility, so as to improve the reliability of the experimental results.

Below is the remainder of this article. Section 2 describes the background and related works. Section 3 describes the MHF method and optimization scheme based on the joint-structured sparse LL1 model. Section 4 describes the experimental setup, visualization of experimental results, evaluation metrics, and classification results and develops the discussion. Section 5 summarizes the algorithm proposed in this paper and future improvement directions.

2. Background

Because LR-HSIs possess two spatial dimensions and one spectral dimension, they can be regarded as third-order tensors. In recent years, with the continuous improvement

of the MHF algorithm, the tensor decomposition method has been widely used because of its ability to avoid an excessive loss of spatial and spectral information. The two more common models are the CPD and Tucker models.

The CPD model [60] is a tensor rank decomposition model. The procedure is shown below, which allows the decomposing of the tensor into the total of multiple tensors of rank one. The mathematical model is as follows (note: see Appendix A for an explanation of the relevant symbols in the text).

$$\mathcal{Y} = \sum_{r=1}^R \mathbf{A}_r \circ \mathbf{B}_r \circ \mathbf{C}_r . \quad (1)$$

where $\mathbf{A}_r \in \mathbb{R}^{I \times 1}$, $\mathbf{B}_r \in \mathbb{R}^{J \times 1}$, and $\mathbf{C}_r \in \mathbb{R}^{K \times 1}$ are second-order factor matrices; $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$ is a third-order tensor. Figure 1 represents the flow of CPD.

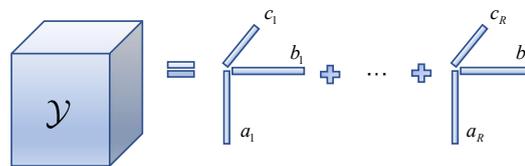


Figure 1. Canonical polyadic decomposition process.

The Tucker model [61] is also widely used in the field of imaging, and it is a higher-order form of principal component analysis (PCA), which focuses on decomposing the tensor into a core tensor and three-factor matrices. The Tucker decomposition process is shown in Figure 2. The mathematical model is expressed as [62]:

$$\mathcal{Y} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}. \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{I \times n_1}$, $\mathbf{B} \in \mathbb{R}^{J \times n_2}$, and $\mathbf{C} \in \mathbb{R}^{K \times n_3}$ are the factor matrices; $\mathcal{G} \in \mathbb{R}^{I \times J \times K}$ is a third-order core tensor.

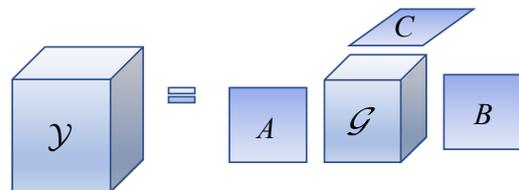


Figure 2. The Tucker decomposition process.

In addition to the two tensor models mentioned above, the BTM model is gradually being applied to the field of image fusion because the potential factors have a good physical interpretation, such as in the work performed previously [59]. In solving the MHF problem, compared with the CPD and Tucker models, potential factors in the BTM model, respectively, represent the abundance and the endmember, which have a physical sense of non-negativity and spatial smoothness, and contribute to the utilization of various prior knowledge in the image fusion process.

By combining the advantages of the CPD and Tucker models, the BTM of $rank - (L_r, M_r, N_r)$ (abbreviation: LMN model) in general form can be expressed as:

$$\mathcal{Y} = \sum_{r=1}^R \mathcal{G}_r \times_1 \mathbf{A}_r \times_2 \mathbf{B}_r \times_3 \mathbf{C}_r . \quad (3)$$

where $\mathbf{A}_r \in \mathbb{R}^{I \times L_r}$, $\mathbf{B}_r \in \mathbb{R}^{J \times M_r}$, $\mathbf{C}_r \in \mathbb{R}^{K \times N_r}$, and $\mathcal{G}_r \in \mathbb{R}^{L_r \times M_r \times N_r}$. The LMN decomposition process is shown in Figure 3. The above description shows that the BTM is approximating the tensor \mathcal{Y} by R component tensors, and each tensor is a Tucker of $rank - (L_r, M_r, N_r)$.

If each component tensor rank is one, it can be treated as a CPD. Similarly, if $R = 1$, then Equation (3) is the Tucker model of $rank - (L_r, M_r, N_r)$. In other words, the BTM can decompose a tensor into the sum of multiple component tensors, and this process can be viewed as the CPD, while each component tensor decomposition is viewed as a Tucker decomposition. This overcomes the problem that each component tensor rank must be one in the CPD and lifts the restriction that the Tucker decomposition has only one component tensor. It is a combination of the CPD and the Tucker decomposition.

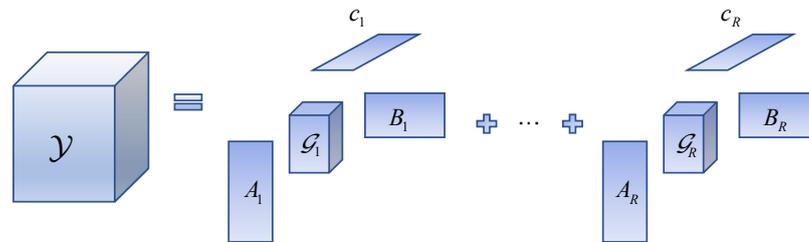


Figure 3. LMN decomposition process.

Usually, the special BTM model is adopted to resolve the MHF problem. The special BTM model of $rank - (L, L, 1)$ (abbreviation: LL1 model) is obtained when each member tensor degenerates to an $L_r \times L_r$ matrix, and each L_r takes the same value. The decomposition process of the LL1 model is shown in Figure 4. The mathematical model is as follows:

$$Y = \sum_{r=1}^R (A_r B_r^T) \circ c_r \tag{4}$$

where $A_r \in \mathbb{R}^{I \times L_r}$, $B_r \in \mathbb{R}^{J \times L_r}$, and $C = [c_1, \dots, c_R] \in \mathbb{R}^{K \times R}$. Among these, A_r and B_r are, respectively, the subblocks of the block matrices A and B , denoted as $A = [A_1, \dots, A_R] \in \mathbb{R}^{I \times \sum_{r=1}^R L_r}$, $B = [B_1, \dots, B_R] \in \mathbb{R}^{J \times \sum_{r=1}^R L_r}$. In general, the full column rank requirement for A and B is $\min(I, J) \geq \sum_{r=1}^R L_r$. Under mild conditions, the LL1 model possesses a nice property, i.e., $A_r B_r^T$ and c_r are identifiable until the alignment and scaling ambiguities arise, which is induced as follows (Theorem 4.7 of [56]). The potential factors $A_r B_r^T$ and c_r represent the abundance information; the former represents the abundance information of the endmember r , and the latter represents the r th endmember.

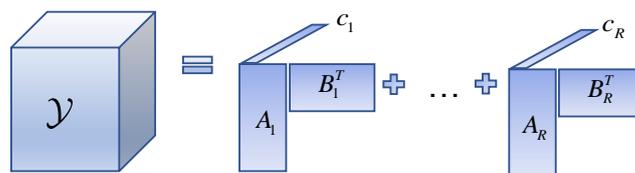


Figure 4. LL1 decomposition model.

In summary, the CPD and Tucker models are both low-rank models (see Figures 1 and 2), capable of “encoding” dependencies in different dimensions of the data and often used in spectral image modeling with good recoverability [25,47]. Although this is useful for recovering an HRHS, it is not necessarily better than the BTM-based MHF approach in practice. The BTM-based MHF framework can be used to introduce the structural information of potential factors using their physical interpretation, which helps to standardize the MHF criteria with a priori knowledge.

3. MHF Method Based on Joint-Structured Sparse LL1 Model

3.1. The MHF Method for the LL1 Model

The proof of how the model can be interpreted using physics is described in a separate work [59]. The LL1 model is used to model the spectral images. For all spectral images, they can be viewed as a third-order tensor of the form “space \times space \times

spectrum". $\mathcal{Y} \in \mathbb{R}^{I_M \times J_M \times K_H}$, $\mathcal{Y}_H \in \mathbb{R}^{I_H \times J_H \times K_H}$, and $\mathcal{Y}_M \in \mathbb{R}^{I_M \times J_M \times K_M}$ are used to denote the HRHS, LR-HSI, and HR-MSI, respectively. Further, I_H , J_H , and I_M , J_M denote the spatial dimensions of the LR-HSI and HR-MSI, respectively; K_H and K_M denote the spectral dimensions of the LR-HSI and HR-MSI, respectively. Due to the limitation of current optical sensors, multispectral images are rich in spatial information, and hyperspectral images are rich in spectral information; i.e.:

$$I_M \gg I_H, J_M \gg J_H, K_H \gg K_M. \quad (5)$$

The goal of the MHF problem is to obtain the HRHS by fusing an HR-MSI and LR-HSI in the same scene. The theoretical validation of the LMM model and the LL1 model was demonstrated in another work [59]. Thus, assuming that the abundance matrix is a low-rank matrix, the LL1 model formulation on the HRHS is:

$$\mathcal{Y}_S = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ \mathbf{c}_r. \quad (6)$$

where $\mathbf{A}_r \in \mathbb{R}^{I_M \times L_r}$, $\mathbf{B}_r \in \mathbb{R}^{J_M \times L_r}$, and $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R] \in \mathbb{R}^{K_H \times R}$, $r = 1, 2, \dots, R$. \mathcal{Y}_S denotes the fused image.

The expansion of the tensor \mathcal{Y} can be expressed as:

$$\mathbf{Y}_{(1)} = (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T, \quad (7)$$

$$\mathbf{Y}_{(2)} = (\mathbf{C} \odot \mathbf{A}) \mathbf{B}^T, \quad (8)$$

$$\mathbf{Y}_{(3)} = [(\mathbf{B}_1 \odot \mathbf{c}_1) \mathbf{1}_{L_1}, \dots, (\mathbf{B}_R \odot \mathbf{c}_R) \mathbf{1}_{L_R}] \mathbf{C}^T. \quad (9)$$

The above expressions are used in the subsequent alternating iterations to solve \mathbf{A} , \mathbf{B} , and \mathbf{C} . The LR-HSI can be obtained by spatially blurring and downsampling the HRHS. Thus, the fuzzy matrix and downsampling matrix are modeled by multiplying the rows and columns of each $\mathcal{Y}(:, :, k)$ component. The mathematical model is shown below:

$$\mathcal{Y}_H(:, :, k) = \mathbf{P}_1 \mathcal{Y}_S(:, :, k) \mathbf{P}_2^T, \quad (10)$$

$$k = 1, 2, \dots, K_H.$$

where $\mathbf{P}_1 \in \mathbb{R}^{I_H \times I_M}$ and $\mathbf{P}_2 \in \mathbb{R}^{J_H \times J_M}$, respectively, denote the spatial ambiguity and downsampling matrices along the two spatial dimensions.

If \mathcal{Y} conforms to the LL1 model and the spatial decay model of Equation (10) holds, there is:

$$\mathcal{Y}_H = \sum_{r=1}^R (\mathbf{P}_1 \mathbf{A}_r (\mathbf{P}_2 \mathbf{B}_r)^T) \circ \mathbf{c}_r. \quad (11)$$

The HR-MSI can be modeled as an HRHS spectral band aggregation form. The mathematical model is as follows:

$$\mathcal{Y}_M(i, j, :) = \mathbf{P}_3 \mathcal{Y}_S(i, j, :), \quad (12)$$

$$\forall i, j.$$

where $\mathbf{P}_3 \in \mathbb{R}^{K_M \times K_H}$ is the waveband aggregation matrix associated with the sensor specification. Similarly, if \mathcal{Y} conforms to the LL1 model and the spectral degradation model in Equation (12) holds, there is:

$$\mathcal{Y}_M = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ \mathbf{P}_3 \mathbf{c}_r. \quad (13)$$

Under this degenerate model, the task of reconstructing the HRHS is mainly to estimate the potential factors $\mathbf{A}_r \mathbf{B}_r^T$ and \mathbf{c}_r from \mathcal{Y}_H and \mathcal{Y}_M . The HRHS is then reconstructed using $\mathcal{Y} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ \mathbf{c}_r$. Thus, the MHF problem based on the LL1 model can be modeled as:

$$\min_{A,B,C} \frac{1}{2} \left\| \mathcal{Y}_H - \sum_{r=1}^R (\mathbf{P}_1 \mathbf{A}_r (\mathbf{P}_2 \mathbf{B}_r)^T) \circ \mathbf{c}_r \right\|_F^2 + \frac{1}{2} \left\| \mathcal{Y}_M - \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ (\mathbf{P}_3 \mathbf{c}_r) \right\|_F^2, \quad (14)$$

s.t. $A \geq 0, B \geq 0, C \geq 0$.

3.2. The MHF Modeling Based on Joint-Structured Sparse

The solution to the above model has an uncomfortable nature, so constraints need to be placed on the potential factors of the above objective function, and regularization methods need to be added to improve the performance of the algorithm. Consider that the operation time, computational complexity, and convergence speed of the BTM algorithm are affected by the rank R and L . In addition to the above, too high or too low estimates of R and L may allow noise/artifacts to interfere with the desired results [63]. The sparsity of the abundance matrix and the scaling/counter-scaling problems that occur in the fusion model can also affect the performance of the algorithm. Based on the above considerations, this paper introduces a regularization term that matches the structure of the BTM model, which improves the flexibility of the model by having hierarchical node blocks and column sparsity. Firstly, the potential factors \mathbf{A} and \mathbf{B} are formed into a chunk matrix \mathbf{U} , which introduces an $L_{2,1}$ -norm to promote structured sparsity and eliminate the scaling effects present in the model. Secondly, add a L_2 -norm value to the endmember \mathbf{c}_r , in order to counteract the counter-scaling effects that occur in the decomposition model. Finally, the chunk matrix \mathbf{U} and \mathbf{c}_r are coupled together to form a matrix $\varphi(\mathbf{A}, \mathbf{B}, \mathbf{C})$ of the size $2 \times R$. The introduction of the $L_{2,1}$ -norm promotes the elimination of the entire block of \mathbf{A} and \mathbf{B} , ensuring that the BTM algorithm can avoid noise/artifacts from interfering with the experimental results in the case of an overestimated rank. In summary, the modified scheme is expressed as follows:

$$\min_{A,B,C} \frac{1}{2} \left\| \mathcal{Y}_H - \sum_{r=1}^R (\mathbf{P}_1 \mathbf{A}_r \mathbf{B}_r^T \mathbf{P}_2^T) \circ \mathbf{c}_r \right\|_F^2 + \frac{1}{2} \left\| \mathcal{Y}_M - \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ (\mathbf{P}_3 \mathbf{c}_r) \right\|_F^2 + \lambda \|\varphi(\mathbf{A}, \mathbf{B}, \mathbf{C})\|_{2,1}, \quad (15)$$

s.t. $\mathbf{A} \geq 0, \mathbf{B} \geq 0, \mathbf{C} \geq 0$,

$$\varphi(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \begin{bmatrix} \|\mathbf{U}_1\|_{2,1} & \|\mathbf{U}_2\|_{2,1} & \cdots & \|\mathbf{U}_R\|_{2,1} \\ \|\mathbf{c}_1\|_2 & \|\mathbf{c}_2\|_2 & \cdots & \|\mathbf{c}_R\|_2 \end{bmatrix}. \quad (16)$$

where $\lambda > 0$ is the regularization parameter, matrix $\mathbf{A}_r = [a_{r1} \ a_{r2} \ \cdots \ a_{rL_r}] \in \mathbb{R}^{I \times L_r}$, $\mathbf{B}_r = [b_{r1} \ b_{r2} \ \cdots \ b_{rL_r}] \in \mathbb{R}^{J \times L_r}$, and $\mathbf{C} = [c_1 \ c_2 \ \cdots \ c_R] \in \mathbb{R}^{K \times R}$. R and L are the rank, and $r = 1, 2, \dots, R$. $\mathbf{U} = [\mathbf{A}^T \ \mathbf{B}^T]^T$ denotes the r th $(I + J) \times L_r$ block. Using the above modification scheme, while solving the MHF problem using the LL1 model, the ranks R and L can be overestimated, which can be reduced to real values by an appropriate selection of λ . In reference to the ideas in the literature [64,65], the well-known IRLS [66] method is extended to solve Model (15).

The flow chart of the algorithm is shown in Figure 5. The two images of the LR-HSI and HR-MSI are decomposed using the LL1 model to obtain the endmember matrix \mathbf{C} and the abundance matrix \mathbf{A}, \mathbf{B} . Then, \mathbf{A}, \mathbf{B} , and \mathbf{C} are brought into Model (15), and the extended IRLS method is used to iteratively solve \mathbf{A}, \mathbf{B} , and \mathbf{C} and bring them into Equation (6) to reconstruct the HRHS.

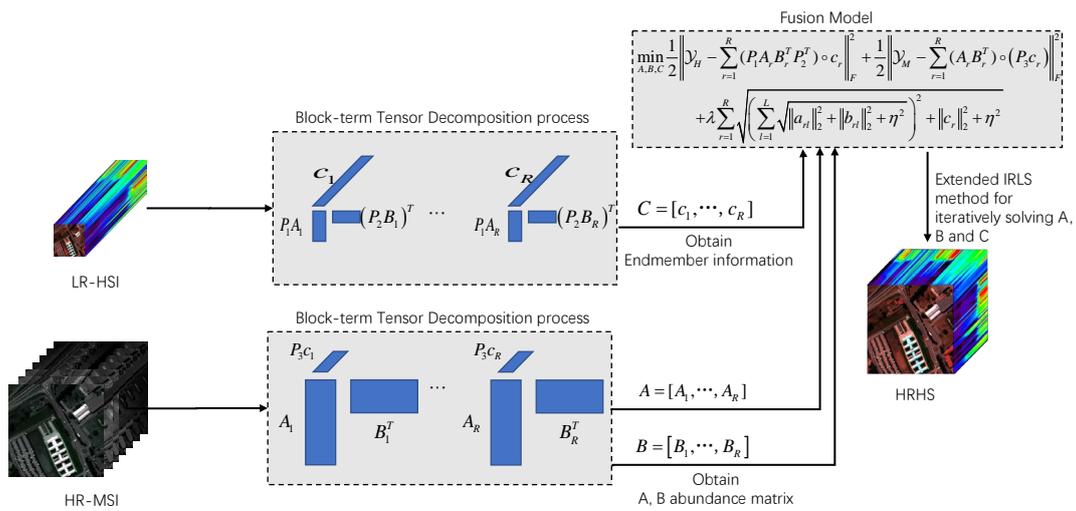


Figure 5. Algorithm flow chart.

The procedure for solving Model (15) is as follows. Express Model (15) in a more explicit form. The mathematical model is shown below:

$$\min_{A,B,C} \frac{1}{2} \left\| \mathcal{Y}_H - \sum_{r=1}^R (P_1 A_r B_r^T P_2^T) \circ c_r \right\|_F^2 + \frac{1}{2} \left\| \mathcal{Y}_M - \sum_{r=1}^R (A_r B_r^T) \circ (P_3 c_r) \right\|_F^2 + \lambda \sum_{r=1}^R \sqrt{\left(\sum_{l=1}^L \sqrt{\|a_{rl}\|_2^2 + \|b_{rl}\|_2^2 + \eta^2} \right)^2 + \|c_r\|_2^2 + \eta^2} \quad (17)$$

where η is a positive integer with guaranteed smoothness. For each potential factor **A**, **B**, and **C** in the model, they are convex with respect to the objective function. In addition, since each factor matrix in the regularization term is indistinguishable, the exact analytic formula cannot be obtained, and the extended IRLS method is used to solve the problem of Model (15). A hierarchical IRLS structure consisting of two independent reweighted least squares steps is identified from the regularization terms. Each step produces a separate reweighting matrix. The first matrix \mathbf{O}_1 is composed of the inverse of the external summation term of the regularization term in Model (15), which is the block of the joint weighted **A** and **B**. The second matrix \mathbf{O}_2 consists of the inverse of the terms summed inside the regularized terms in Model (15), which is used to balance the corresponding columns of A_r and B_r . The expressions \mathbf{O}_1 and \mathbf{O}_2 , respectively, are shown below:

$$\mathbf{O}_1^k(r, r) = \left(\left(\sum_{l=1}^L \sqrt{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2} \right)^2 + \|c_r^k\|_2^2 + \eta^2 \right)^{-1/2}, \quad (18)$$

$$\mathbf{O}_2^k((r-1)L + l, (r-1)L + l) = \left(\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2 \right)^{-1/2}. \quad (19)$$

where \mathbf{O}_1 is a diagonal matrix of order R , $\mathbf{O}_1(r, r)$ denotes the value of the r th diagonal element; \mathbf{O}_2 is a diagonal matrix of order RL , $\mathbf{O}_2((r-1)L + l, (r-1)L + l)$ denotes the value of the element in row $(r-1)L + l$, column $(r-1)L + l$.

Regarding the three potential factors, **A**, **B**, and **C** in Model (15) are estimated by iterative alternation.

3.2.1. Sub-Problem A

Estimating **A** at the $k + 1$ th iteration, the subproblem on **A** is formulated as follows:

$$\begin{aligned} \mathbf{A}^{k+1} = \arg \min_{\mathbf{A}} & \frac{1}{2} \left\| \mathbf{Y}_{\mathbf{H}(1)} - \mathbf{S}^k (\mathbf{P}_1 \mathbf{A})^T \right\|_{\mathbf{F}}^2 + \frac{1}{2} \left\| \mathbf{Y}_{\mathbf{M}(1)} - \mathbf{M}^k \mathbf{A}^T \right\|_{\mathbf{F}}^2 \\ & + \lambda \sum_{r=1}^R \frac{\left(\sum_{l=1}^L \frac{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2}{\sqrt{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2}} \right)^2 + \|c_r^k\|_2^2 + \eta^2}{\sqrt{\left(\sum_{l=1}^L \sqrt{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2} \right)^2 + \|c_r^k\|_2^2 + \eta^2}}. \end{aligned} \tag{20}$$

where $\mathbf{S} \triangleq \mathbf{C} \odot \mathbf{P}_2 \mathbf{B}$ and $\mathbf{M} \triangleq \mathbf{P}_3 \mathbf{C} \odot \mathbf{B}$. Equation (20) is obtained by the extended IRLS method. By deriving Equation (20) and making its derivative zero, the Sylvester equation with respect to **A** can be obtained as follows.

$$\mathbf{P}_1^T \mathbf{P}_1 \mathbf{A}^{k+1} \mathbf{S}^{kT} \mathbf{S}^k + \mathbf{A}^{k+1} (\mathbf{M}^{kT} \mathbf{M}^k + \lambda \mathbf{O}^k) = \mathbf{P}_1^T \mathbf{Y}_{\mathbf{H}(1)}^T \mathbf{S}^k + \mathbf{Y}_{\mathbf{M}(1)}^T \mathbf{M}^k. \tag{21}$$

where $\mathbf{O} \triangleq (\mathbf{O}_1 \otimes \mathbf{I}_L) \mathbf{O}_2$. Problem (21) is solved using the conjugate gradient (CG) algorithm [67].

3.2.2. Sub-Problem B

Similarly, the minimization problem with respect to **B** is as follows:

$$\begin{aligned} \mathbf{B}^{k+1} = \arg \min_{\mathbf{B}} & \frac{1}{2} \left\| \mathbf{Y}_{\mathbf{H}(2)} - \mathbf{V}^k (\mathbf{P}_2 \mathbf{B})^T \right\|_{\mathbf{F}}^2 + \frac{1}{2} \left\| \mathbf{Y}_{\mathbf{M}(2)} - \mathbf{Z}^k \mathbf{B}^T \right\|_{\mathbf{F}}^2 \\ & + \lambda \sum_{r=1}^R \frac{\left(\sum_{l=1}^L \frac{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2}{\sqrt{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2}} \right)^2 + \|c_r^k\|_2^2 + \eta^2}{\sqrt{\left(\sum_{l=1}^L \sqrt{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2} \right)^2 + \|c_r^k\|_2^2 + \eta^2}}. \end{aligned} \tag{22}$$

where $\mathbf{V} \triangleq \mathbf{C} \odot \mathbf{P}_1 \mathbf{A}$ and $\mathbf{Z} \triangleq \mathbf{P}_3 \mathbf{C} \odot \mathbf{A}$. Equation (23) is obtained by the extended IRLS method. By deriving Equation (22) and making its derivative zero, the Sylvester equation with respect to **B** can be obtained as follows:

$$\mathbf{P}_2^T \mathbf{P}_2 \mathbf{B}^{k+1} \mathbf{V}^{kT} \mathbf{V}^k + \mathbf{B}^{k+1} (\mathbf{Z}^{kT} \mathbf{Z}^k + \lambda \mathbf{O}^k) = \mathbf{P}_2^T \mathbf{Y}_{\mathbf{H}(2)}^T \mathbf{V}^k + \mathbf{Y}_{\mathbf{M}(2)}^T \mathbf{Z}^k. \tag{23}$$

Problem (23) is solved using the CG algorithm.

3.2.3. Sub-Problem C

Similarly, the minimization problem on **C** is as follows:

$$\begin{aligned} \mathbf{C}^{k+1} = \arg \min_{\mathbf{C}} & \frac{1}{2} \left\| \mathbf{Y}_{\mathbf{H}(3)} - \mathbf{Q}^k \mathbf{C}^T \right\|_{\mathbf{F}}^2 + \frac{1}{2} \left\| \mathbf{Y}_{\mathbf{M}(3)} - \mathbf{E}^k (\mathbf{P}_3 \mathbf{C})^T \right\|_{\mathbf{F}}^2 \\ & + \lambda \sum_{r=1}^R \frac{\left(\sum_{l=1}^L \frac{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2}{\sqrt{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2}} \right)^2 + \|c_r\|_2^2 + \eta^2}{\sqrt{\left(\sum_{l=1}^L \sqrt{\|a_{rl}^k\|_2^2 + \|b_{rl}^k\|_2^2 + \eta^2} \right)^2 + \|c_r^k\|_2^2 + \eta^2}}. \end{aligned} \tag{24}$$

where $\mathbf{Q} \triangleq (\mathbf{P}_2 \mathbf{B}_1 \odot_c \mathbf{P}_1 \mathbf{A}_1) \mathbf{1}_{L_1}, \dots, (\mathbf{P}_2 \mathbf{B}_R \odot_c \mathbf{P}_1 \mathbf{A}_R) \mathbf{1}_{L_R}$, and $\mathbf{E} \triangleq (\mathbf{B}_1 \odot_c \mathbf{A}_1) \mathbf{1}_{L_1}, \dots, (\mathbf{B}_R \odot_c \mathbf{A}_R) \mathbf{1}_{L_R}$. Equation (25) is obtained by the extended IRLS method. By deriving Equation (24) and making its derivative zero, the Sylvester equation with respect to **C** can be obtained as follows:

$$\mathbf{P}_3^T \mathbf{P}_3 \mathbf{C}^{k+1} \mathbf{E}^{kT} \mathbf{E}^k + \mathbf{C}^{k+1} (\mathbf{Q}^{kT} \mathbf{Q}^k + \lambda \mathbf{O}_1^k) = \mathbf{Y}_{\mathbf{H}(3)}^T \mathbf{Q}^k + \mathbf{P}_3^T \mathbf{Y}_{\mathbf{M}(3)}^T \mathbf{E}^k. \tag{25}$$

Then, the CG algorithm is used to solve Problem (25).

The steps regarding the proposed algorithm are shown in Algorithm 1.

Algorithm 1 JSLL1-IRLSCG.

Input $\mathcal{Y}, R, L, \lambda > 0, \eta > 0, k = 0$.

Output Target image HRHS.

While unconverged do

 Calculate \mathbf{O}_1 and \mathbf{O}_2 by Equations (18) and (19);

 Update \mathbf{O}^k by equation $(\mathbf{O}_1^k \otimes \mathbf{I}_L) \mathbf{O}_2^k$;

 Solve Problem (21) using the CG algorithm to update \mathbf{A}^{k+1} ;

 Solve Problem (23) using the CG algorithm to update \mathbf{B}^{k+1} ;

 Solve Problem (25) using the CG algorithm to update \mathbf{C}^{k+1} ;

$k \leftarrow k + 1$;

end While;

Reconstruction of images using type $\mathcal{Y} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ \mathbf{c}_r$ and acquisition of HRHS.

In summary, in the proposed algorithm, when the ranks R and L are overestimated in the model, reweighting by matrix \mathbf{O}_1 will impose joint block sparsity on \mathbf{A} and \mathbf{B} and column sparsity on \mathbf{C} , which helps to estimate R . Meanwhile, re-weighting the matrix \mathbf{O}_2 will jointly promote column sparsity on the corresponding blocks of \mathbf{A} and \mathbf{B} , thus estimating L .

4. Experiments and Analysis

4.1. Datasets

4.1.1. Standard Datasets

The Pavia University dataset [59] was acquired by the Reflection Optical System Imaging Spectrometer (ROSIS) over the city of Pavia. The size of the acquired image is 610×340 and consists of 115 spectral bands with an image resolution of 1.3 m. To enhance the experimental effect, the sub-image located in the upper left corner of the image was selected as the experimental image after removing the band containing water vapor. The size of the experimental image is $256 \times 256 \times 102$, and the image contains rich features such as houses, roads, and trees.

The Indian Pines dataset [59] was acquired by the Airborne Visible Infrared Spectroscopic Imager (AVIRIS) over Indian, U.S.A. The AVIRIS wavelength range is between $0.4 \mu\text{m}$ and $2.5 \mu\text{m}$ with a spatial resolution of 20 m. After removing 24 bands that cannot be reflected by water, the remaining bands were used for the study. Given the experimental requirements, the final image size selected was $144 \times 144 \times 200$.

4.1.2. Local Dataset

The dataset of Ningxia Sand Lake [59] was acquired by the WFV4 multispectral imager on the Gaofen-1 (GF1) satellite and the AHSI hyperspectral imager on the Gaofen-5 (GF5) satellite over Ningxia Sand Lake. The WFV4 wavelength range is $0.45\text{--}0.89 \mu\text{m}$, the spatial resolution is 16 m, and the size of the acquired multispectral image is $12,000 \times 13,400$. The wavelength range of AHSI is $0.4\text{--}2.5 \mu\text{m}$, the spatial resolution is 30 m, it consists of 330 bands, and the size of the acquired hyperspectral image is 2774×2554 . After removing 187 bands with high noise again, the remaining bands are used as the research object. The experimental images were intercepted at the same position as the acquired multispectral and hyperspectral images, and the final selected image sizes were $254 \times 254 \times 3$ for the HR-MSI and $254 \times 254 \times 143$ for the LR-HSI, respectively.

4.2. Experimental Setup

4.2.1. Experimental Environment

Eight algorithms, CNMF [18], Hysure [21], FUSE [22], SCOTT [25], IRSR [45], STEREO [47], BTM [54], and SCLL1 [58], were selected as comparison tests. Among them, SCOTT and STEREO are based on the Tucker and CPD models to solve the MHF problem, respectively; the BTM and SCLL1 are based on the BTM model to solve the MHF problem. Considering the different experimental environments of each algorithm, it is often difficult to achieve the effect in the original literature by directly adopting the parameter settings of the original literature under the current experimental environment, so this paper optimizes the parameters of other algorithms under the current experimental environment. Specifically, standard normally distributed pseudo-random numbers are used as initialization values for all algorithms, and the optimal parameter values are selected through ablation experiments; the hyperspectral decomposition threshold in the CNMF experiments is set to $\theta_H = 10^{-2}$, the multispectral decomposition threshold is set to $\theta_M = 10^{-2}$, and the outer recirculation decomposition threshold is set to $\theta_O = 10^{-4}$. The value of the weighting parameter in STEREO is set to $\lambda = 10^{-5}$. The regularization parameter is set to $\lambda = 10^{-3}$ and $\theta = 10^{-3}$ in SCLL1, and the parameters of the other algorithms that are not mentioned in the original text are used to set the parameters. The CPU in the experimental equipment is Intel(R) Xeon(R) Gold 6154; the RAM size is 256G. The operating system is 64-bit Windows 10. The experimental platform is Matlab R2017b.

4.2.2. Degradation Model

The local datasets are acquired using multispectral and hyperspectral sensors in the acquisition process to directly obtain the HR-MSI and LR-HSI, respectively. However, the standard dataset is different in that it is a hyperspectral dataset acquired by a hyperspectral sensor and lacks a corresponding multispectral dataset to perform fusion experiments. Therefore, the standard dataset is generally used as a reference image when fusion experiments are being performed, and the HR-MSI and LR-HSI used in the experiments are obtained from it by degradation, respectively. In this paper, the degradation model follows the literature [22,32,68] and strictly follows the Wald agreement. The process of acquiring the LR-HSI is such that the standard dataset is fuzzed using a 9×9 Gaussian kernel and sampled at 4-pixel intervals. The process of acquiring the HR-MSI is referred to in the literature [47]. At last, noise is added for the LR-HSI and HR-MSI.

4.3. Numerical Evaluation Indicators

In order to evaluate the differences between the algorithms proposed in this paper and the comparison algorithms in further detail, it is necessary to evaluate them with the help of numerical evaluation metrics, in addition to direct observation of the HRHS quality and difference images of each algorithm. In recent years, many evaluation metrics have emerged with the continuous improvement of the algorithms, but metrics such as the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [69], correlation coefficient (CC), universal image quality index (UIQI) [70], root mean squared error (RMSE), Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) [71], spectral angle mapper (SAM) [72], and TIME, which appear in the literature [4,17], are widely used, and they can comprehensively assess the superiority of the algorithms in terms of spatial information, spectral information, and algorithm efficiency. The specific details of these evaluation metrics are described in detail in our separate work [59]. Among these, the larger the PSNR value, the better the image quality; the closer the SSIM, CC, and UIQI values are to one, the better the quality of the HRHS fusion; the closer the values of the RMSE, ERGAS, SAM, and TIME are to 0, the better the fusion effect. Detailed definitions follow:

The peak signal-to-noise ratio (PSNR) evaluates the level of image distortion or noise. The general situation of PSNR is between 30 dB and 50 dB. In the ideal case, it is expected to be as high as possible, with ∞ representing the optimal value. It should be specifically

noted that when the *PSNR* is more than 30 dB, it is difficult for the human eye to see the difference between the fused image and the reference image. The definitions are as follows.

$$PSNR(\mathcal{Y}, \mathcal{Y}_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} PSNR(\mathcal{Y}_i, \mathcal{Y}_{S_i}) \quad (26)$$

The structural similarity (*SSIM*) is used to calculate the structural similarity between the HRHS and the reference image in the spatial domain [69]. The *SSIM* ranges from -1 to 1 . When the two images are identical, the *SSIM* value is equal to one. The definitions are as follows.

$$SSIM(\mathcal{Y}, \mathcal{Y}_S) = \frac{1}{M} \sum_{i=1}^M \frac{(2\bar{\mathcal{Y}}_i, \bar{\mathcal{Y}}_{S_i} + c_1)(2\sigma_{\bar{\mathcal{Y}}_i, \bar{\mathcal{Y}}_{S_i}} + c_2)}{[(\bar{\mathcal{Y}}_i)^2 + (\bar{\mathcal{Y}}_{S_i})^2 + c_1](\sigma_{\bar{\mathcal{Y}}_i}^2 + \sigma_{\bar{\mathcal{Y}}_{S_i}}^2 + c_2)} \quad (27)$$

The correlation coefficient (*CC*) describes the degrees of spectral likeness with respect to the objective image, as well as the fusion image. The *CC* is based on a scale from -1 to 1 . The closer the *CC* is to one, the higher their positive correlation. The definitions are as follows.

$$CC(\mathcal{Y}, \mathcal{Y}_S) = \frac{\sum_{i=1}^{N_W} \sum_{j=1}^{N_H} \{[\mathcal{Y}(i, j) - AVE_{\mathcal{Y}}] \cdot [\mathcal{Y}_S(i, j) - AVE_{\mathcal{Y}_S}]\}}{\sqrt{\sum_{i=1}^{N_W} \sum_{j=1}^{N_H} [\mathcal{Y}(i, j) - AVE_{\mathcal{Y}}]^2 \cdot \sum_{i=1}^{N_W} \sum_{j=1}^{N_H} [\mathcal{Y}_S(i, j) - AVE_{\mathcal{Y}_S}]^2}} \quad (28)$$

The universal image quality index (*UIQI*) [70] is specifically used to detect the average correlation between the reference image and the fused image. If the fused image is closer to the reference image, the *UIQI* value is closer to a value of one, which means a better reconstruction effect. The definitions are as follows.

$$UIQI(\mathcal{Y}, \mathcal{Y}_S) = \frac{1}{M} \sum_{i=1}^M \frac{4\sigma_{\bar{\mathcal{Y}}_i, \bar{\mathcal{Y}}_{S_i}} \cdot \bar{\mathcal{Y}}_i, \bar{\mathcal{Y}}_{S_i}}{(\sigma_{\bar{\mathcal{Y}}_i}^2 + \sigma_{\bar{\mathcal{Y}}_{S_i}}^2) [(\bar{\mathcal{Y}}_i)^2 + (\bar{\mathcal{Y}}_{S_i})^2]} \quad (29)$$

The root mean squared error (*RMSE*) is an important index describing the root mean squared error between each band of the reference image and the fused image. If the fused image is closer to the reference image, the *RMSE* value will be closer to a value of 0. The definitions are as follows.

$$RMSE(\mathcal{Y}, \mathcal{Y}_S) = \sqrt{\frac{\|\mathcal{Y}, \mathcal{Y}_S\|_F^2}{N_W N_H N_S}} \quad (30)$$

The Erreur Relative Globale Adimensionnelle de Synthèse (*ERGAS*) [71] is an important measure of the global quality and spectral variability of the reconstructed image. If *ERGAS* is closer to a value of 0, the reconstruction effect will be better. The definitions are as follows.

$$ERGAS(\mathcal{Y}, \mathcal{Y}_S) = \frac{100}{s} \sqrt{\frac{1}{N_S} \sum_{i=1}^{N_S} \frac{MSE(\mathcal{Y}, \mathcal{Y}_S)}{MEAN(\mathcal{Y}_S)}} \quad (31)$$

The spectral angle mapper (*SAM*) [72] is a metric used to measure the magnitude of the angle between the pixel spectra of the source image and the fused image. If the *SAM*

value is closer to a value of 0, the two images match more, and the similarity is higher. The definitions are as follows.

$$SAM(\mathcal{Y}, \mathcal{Y}_S) = \frac{1}{N_W N_H} \sum_{i=1}^{N_W N_H} \arccos \frac{(\mathcal{Y}, \mathcal{Y}_S)}{\|\mathcal{Y}_i\|_2 \bullet \|\mathcal{Y}_{S_i}\|_2} \quad (32)$$

where \mathcal{Y}_i and \mathcal{Y}_{S_i} are the i -th band of the initial and fused images, and $AVE_{\mathcal{Y}}$ and $AVE_{\mathcal{Y}_S}$ are the average pixel values of the initial and fusion images, respectively. $\bar{\mathcal{Y}}_i$ and $\bar{\mathcal{Y}}_{S_i}$ are the mean values of \mathcal{Y}_i and \mathcal{Y}_{S_i} , respectively. σ_i , σ_i^2 , and $\sigma_{i,j}^2$ are used to express the standard deviation, variance, and covariance, respectively. c_i is a constant. Here, the larger the value of $PSNR$, the closer the $SSIM$, CC , and $UIQI$ values are to 1, and the closer the $RMSE$, $ERGAS$, and SAM values are to 0, indicating a better fusion effect.

The signal-to-noise ratio (SNR) is the ratio of signal to noise in an image. A high signal-to-noise ratio is reflected in a clean and noise-free image, while a low signal-to-noise ratio makes the image rough and noisy, and the image quality is not translucent and has insufficient contrast. The definitions are as follows.

$$SNR = \frac{Signal}{Noise} \quad (33)$$

4.4. Classification Accuracy Evaluation

Here, we examine the effect of the HRHS on the pixel-level classification task. The fused images obtained by each fusion algorithm contain 11,781 labeled samples with the sample categories of Asphalt, Meadows, Trees, Painted Metal Sheets, Self-Blocking Bricks, and Shadows, and the amount for each sample categories is 2705, 5195, 1089, 1338, 1250, and 204, respectively. The fused images are used as the dataset for the classification algorithm, and 1% of the random points of each category of labeled samples is used as training samples, and the remaining 99% of the random points is used as test samples. Then, 1.5% of each category is taken as the validation set. The classification results of each HRHS were evaluated using three metrics: the overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ). Detailed definitions are as follows.

The overall classification accuracy (OA) refers to the ratio of the number of correctly categorized image elements to the total number of categories in an HRHS.

The kappa coefficient (κ) is an index to evaluate the degree of agreement or accuracy between the classification results and the actual features and is calculated as follows:

$$\kappa = \frac{N \sum_{i=1}^g w_{ii} - \sum_{i=1}^g (w_{i+} w_{+i})}{N^2 - \sum_{i=1}^g (w_{i+} w_{+i})} \quad (34)$$

where g is the total number of categories, i.e., the total amount of columns in the confusion matrix; w_{ii} is the total amount of correct classification, i.e., the count of the pixels on row i and column i of the confusion matrix; w_{i+} and w_{+i} are the total pixel counts of i rows and i columns, respectively; and N indicates the total amount of pixels to be used for accuracy estimation.

The average categorization accuracy (AA) indicates the average value of each accuracy for each category.

4.5. Model Parameter Selection

4.5.1. Selection of the Parameters λ and η

The parameters appearing in the model can affect the effectiveness of the fusion algorithm. For the regularization parameter λ and the smoothing parameter η that appear, a reasonable choice is made by ablation experiments. This is completed as follows: Firstly, λ and η are randomly selected for the experiments to roughly estimate the interval of better

results. Secondly, we fix the value of one of the parameters and assign a regular value to the other parameter, observing the changes in the evaluation indicators, such as the PSNR, and finding the better value. Finally, the optimal solution is found around the better value. With the above ideas, the parameter selection process for the Pavia University dataset is given, as shown in Figures 6 and 7. By fixing one parameter, the other parameter is scaled down 10 times each time; it was observed that the change in the PSNR and other indexes in the graph was found when $\lambda = 0.01$, $\eta = 0.001$, the evaluation indexes reached the optimum, and the algorithm fusion effect was the best. (Note: Other datasets selected the parameter values by this method).

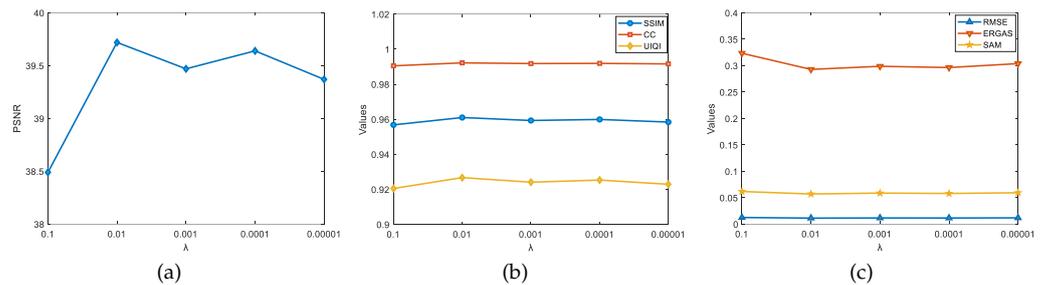


Figure 6. Different λ values of the indicators. (a) Different λ values of PSNR. (b) Different λ values of SSIM, CC, and UIQI. (c) Different λ values of RMSE, ERGAS, and SAM.

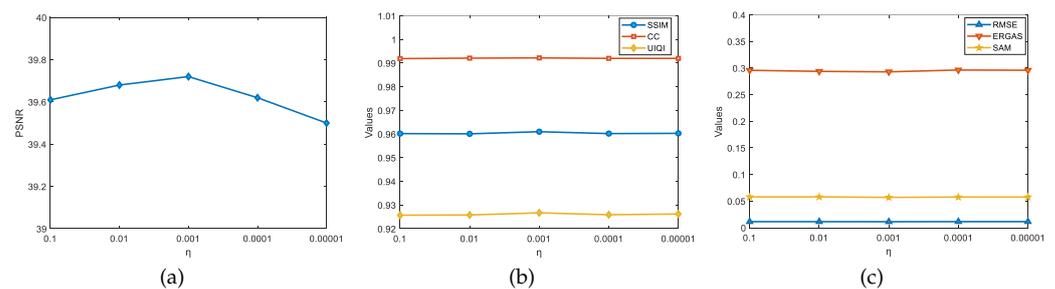


Figure 7. Different η values of the indicators. (a) Different η values of PSNR. (b) Different η values of SSIM, CC, and UIQI. (c) Different η values of RMSE, ERGAS, and SAM.

4.5.2. Rough Estimation of the Ranks L and R

For the LL1 fusion-based algorithm, the ranks L and R of the potential factors affect the overall performance of the algorithm. The selection process regarding the ranks L and R of the Pavia University dataset and the Indian Pines dataset are listed here, regarding the data in the experiments are the average of ten independent experiments conducted.

4.5.3. Estimation of L and R on Pavia University Dataset

Regarding the rough estimates of the ranks L and R in the experiments on the Pavia University dataset, as shown in Figure 8, the graphs show the changes in the evaluation metrics, such as the PSNR, SSIM, and CC, for different ranks of L . The L -value is estimated by fixing the rank R . Set $R = 25$ here. The L -value grows at a rate of 5 per increase from 10. It is clearly observed from Figure 8a that the PSNR values vary more significantly for L in the $[10, 30]$ range. The PSNR change remained essentially flat in the $[30, 45]$ range. In Figure 8b, L is in the $[10, 20]$ range where the three indicators SSIM, CC, and UIQI change more significantly. In summary, it can be roughly estimated that the optimal value of rank L is around $L = 30$.

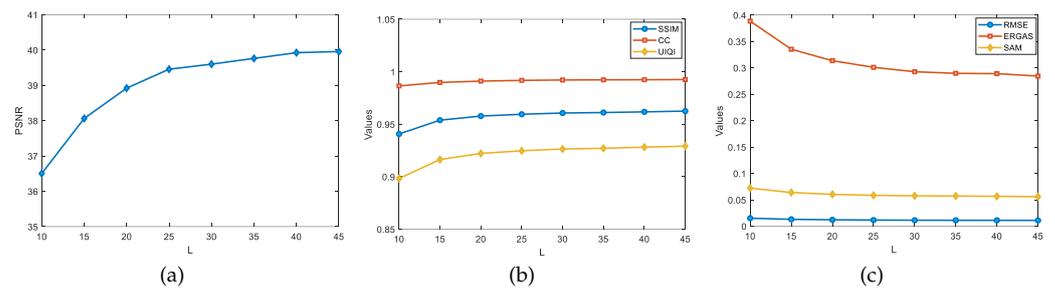


Figure 8. Evaluation metrics for different L values under the Pavia University dataset. (a) Different L values of PSNR. (b) Different L values of SSIM, CC, and UIQI. (c) Different L values of RMSE, ERGAS, and SAM.

Figure 9 shows the graphs of the variation of evaluation metrics, such as the PSNR, SSIM, and CC, for different ranks of R in the Pavia University dataset. Fix the rank L to estimate the R -value, set $L = 35$, and the R -value grows at a rate of 5 per increase from 10. As seen in Figure 9a, the PSNR values increase rapidly with R in the $[10, 15]$ range; there is a slow change in the PSNR during the $[15, 30]$ range; and the PSNR values remained essentially flat in the $[30, 45]$ range. In Figure 9b,c, when R in the $[10, 15]$ range, the three test metrics of the SSIM, CC, and UIQI improve more significantly, while the RMSE and SAM remain basically stable. In Figure 9c, there is a significant increase in the ERGAS values in the range $[10, 25]$. In the rest of the ranges, all of the tested metrics have basically leveled off without any major improvement. In summary, it can be roughly estimated that the optimal value of the rank R is around $R = 25$.

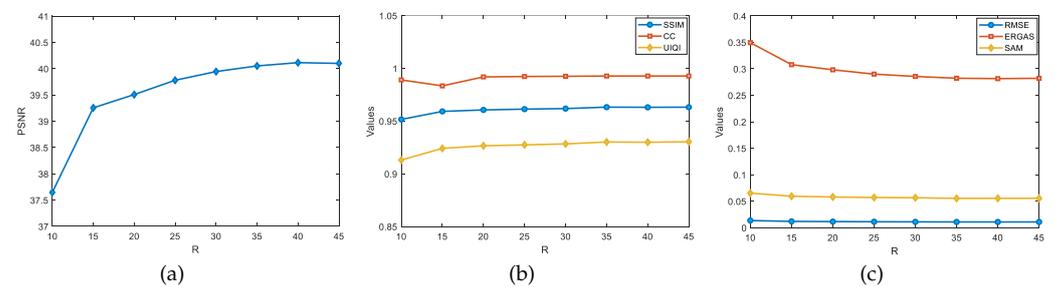


Figure 9. Evaluation metrics for different R values under the Pavia University dataset. (a) Different R values of PSNR. (b) Different R values of SSIM, CC, and UIQI. (c) Different R values of RMSE, ERGAS, and SAM.

4.5.4. Estimation of L and R on Indian Pines Dataset

Regarding the rough estimates of the ranks L and R in the Indian Pines dataset experiment, as shown in Figure 10, the graphs show the changes in the evaluation metrics, such as the PSNR, SSIM, and CC, for different ranks of L . Fix the rank R to estimate L . Initialize $R = 25$. The rank L grows from five at a rate of five per increase. The variation of PSNR values was observed from Figure 10a where the PSNR values are more significantly improved with L in the $[10, 30]$ range. The PSNR appears to decrease in the $[30, 35]$ range but not significantly. In the $[35, 40]$ range, it is more stable. In Figure 10b,c, L is in the interval of $[5, 20]$. The test indexes, such as the SSIM and CC, have more obvious improvements, and, in summary, the optimal value of the rank L can be roughly estimated to be around $L = 20$.

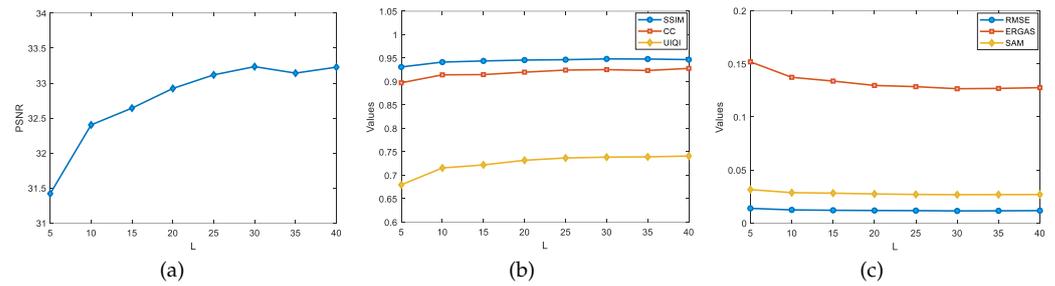


Figure 10. Evaluation metrics for different L values under the Indian Pine dataset. (a) Different L values of PSNR. (b) Different L values of SSIM, CC, and UIQI. (c) Different L values of RMSE, ERGAS, and SAM.

Figure 11 shows the graphs of the changes in the evaluation metrics, such as the PSNR, SSIM, and CC, for different ranks of R on the Indian Pine dataset. Fixing the rank L estimates R . By initializing $L = 30$, the rank R grows at a rate of five per increase starting from five. Observe the change in the PSNR values from Figure 11a where the PSNR values were most significantly enhanced by R in the $[5, 10]$ range. The PSNR shows more pronounced upward and downward fluctuations in the $[10, 30]$ range but largely maintains an upward trend. In Figure 11b,c, R has a more significant improvement in the test metrics, such as SSIM and CC, in the $[5, 10]$ range. In the rest of the ranges, the effect is not significantly improved. In summary, an inappropriate choice of the rank R suppresses the experimental effect of the fusion algorithm, so the optimal value of the rank R is roughly estimated to be around $R = 10$.

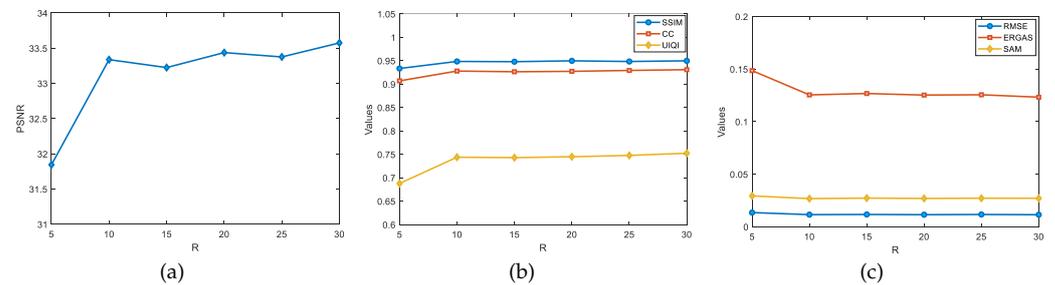


Figure 11. Evaluation metrics for different R values under the Indian Pine dataset. (a) Different R values of PSNR. (b) Different R values of SSIM, CC, and UIQI. (c) Different R values of RMSE, ERGAS, and SAM.

4.5.5. Signal-To-Noise Ratio Selection

The signal-to-noise ratio (SNR) is also one of the important factors affecting the fusion results of algorithms. The SNR of each algorithm is different, and it is necessary to select an SNR that all algorithms can “accept” to avoid the phenomenon of comparing the optimal and non-optimal results. The selected SNRs (unit: dB) range from $[15, 40]$ with a 5 dB interval. All experiments were conducted using the Pavia University dataset. The experimental results are shown in Figure 12. The values of the SNR range is $[15, 30]$. Evaluation indicators, such as the PSNR, SSIM, and CC, were significantly improved. In the range $[30, 40]$, all evaluation indicators tend to be stable. This shows that the performance of all algorithms is maximally released when $SNR = 30$ dB.

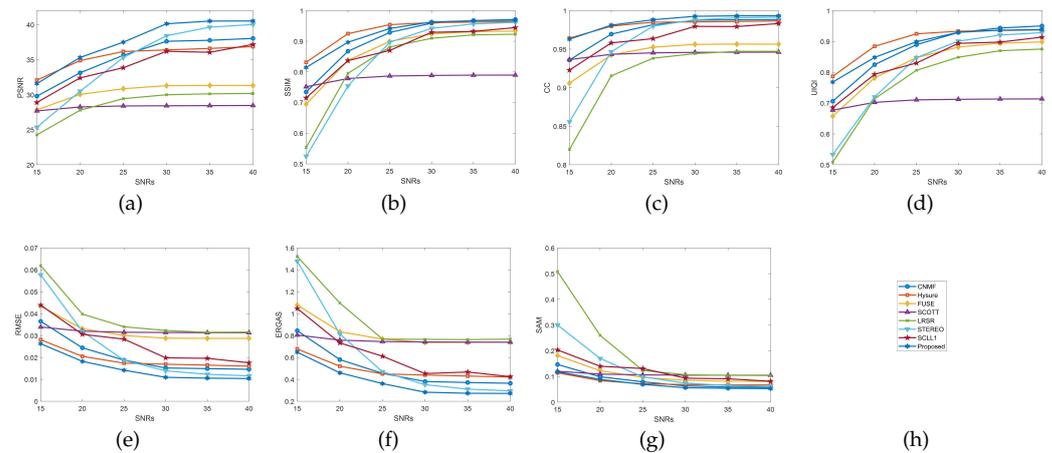


Figure 12. Performance comparison of all algorithms under different SNRs. (a) PSNR of different algorithms. (b) SSIM of different algorithms. (c) CC of different algorithms. (d) UIQI of different algorithms. (e) RMSE of different algorithms. (f) ERGAS of different algorithms. (g) SAM of different algorithms. (h) Icons.

4.6. Regularization Validity Discussion

A 2D-CNBTD was proposed in previous work [55], but the algorithm is not perfect and also suffers from low algorithm efficiency, long running time, and noise/artifacts. Therefore, based on the above considerations, a new algorithm (JSSL1) is proposed in this paper. As shown in Table 1, the numerical evaluation metrics of the three algorithms, BTD, 2D-CNBTD, and JSSL1, are given on the Pavia University dataset, Indian Pines dataset, and NingXia Sand Lake dataset, respectively. The better values are indicated in bold, and the time optimum is noted as 0 s. Other evaluation metrics, such as the PSNR and SSIM, are described in Section 4.3. Among these, the PSNR is used to evaluate the image distortion; the SSIM and UIQI evaluate the spatial features of the fused images; the CC, RMSE, and other metrics are used to evaluate the spectral features of the fused images. From Table 1, it can be seen that the spatial feature metrics SSIM and UIQI of 2D-CNBTD and JSSL1 are greatly improved compared with those before improvement. The CC, RMSE, and SAM metrics of the spectral features are also significantly improved. The PSNR is significantly improved as well, which can effectively remove noise from the images. The experimental results of 2D-CNBTD and JSSL1 on the Pavia University dataset and the NingXia Sand Lake dataset show that the difference between these in the spatial and spectral feature indexes is not obvious; on the NingXia Sand Lake dataset, the PSNR value of JSSL1 is better; on the Indian Pines dataset, each piece of data for JSSL1 is closer to the optimal value. In summary, the JSSL1 algorithm can effectively suppress noise, is suitable for more complex environments, has a lower computational complexity, and can achieve the goal in the least amount of time.

Table 1. Algorithm improvement before and after the value comparison table.

Algorithms	Pavia University			Indian Pines			NingXia Sand Lake		
	BTD	2D-CNBTD	JSSL1	BTD	2D-CNBTD	JSSL1	BTD	2D-CNBTD	JSSL1
PSNR(∞)	35.71	40.86	40.12	25.38	32.42	33.79	36.97	39.49	42.39
SSIM(1)	0.8764	0.9632	0.9652	0.7436	0.9328	0.9515	0.9753	0.9434	0.9261
CC(1)	0.9705	0.9932	0.9927	0.7091	0.8962	0.9338	0.9960	0.9976	0.9978
UIQI(1)	0.8354	0.9321	0.9341	0.4067	0.6949	0.7565	0.9390	0.9634	0.9518
RMSE(0)	0.0222	0.0104	0.0109	0.0306	0.0127	0.0112	50.57	50.362	34.84
ERGAS(0)	0.5454	0.2679	0.2819	0.4431	0.1530	0.1204	0.0882	0.0639	0.0647
SAM(0)	0.0957	0.0558	0.0551	0.0850	0.0336	0.0261	0.0436	0.0309	0.0338
Time(0/s)	80.94	89.68	44.07	7.26	11.62	5.53	64.79	73.44	43.12

4.7. Standard Dataset Experimental Results and Discussion

This section shows the experimental results of the eight fusion algorithms on the Pavia University dataset and the Indian Pines dataset. The experimental results include visual images of the fusion results, difference images, and the numerical results of evaluation metrics.

4.7.1. Pavia University Dataset Experimental Results and Discussion

Through the above experiments, the parameters of the Pavia University dataset are set as: $\lambda = 0.01$, $\eta = 0.001$, and $SNR = 30$ dB. Due to the specificity of the model in this paper, the estimation and selection of the rank can be “overestimated”, so the roughly estimated ranks R and L are slightly adjusted upwards and set to: $L = 35$, $R = 25$. The algorithm superiority was evaluated by observing the visualized images. For the Pavia University dataset, the three bands 61, 25, and 13 were selected as the red, green, and blue bands, respectively, and fused to generate pseudo-color images. The HRHS is shown in Figure 13. Figure 13a shows the original image of the Pavia University dataset. In order to see the differences in HRHSs more clearly, the red box part in each HRHS image is enlarged, and the enlarged part is shown in the upper right corner. The difference images of all algorithms are shown in Figure 14.

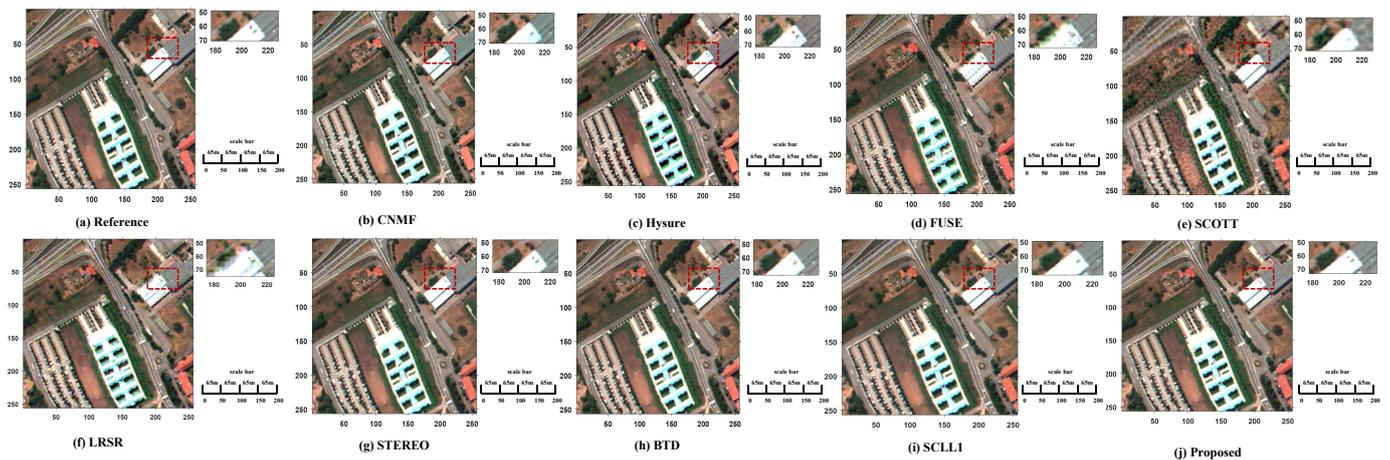


Figure 13. HRHSs of all algorithms on the Pavia University dataset.

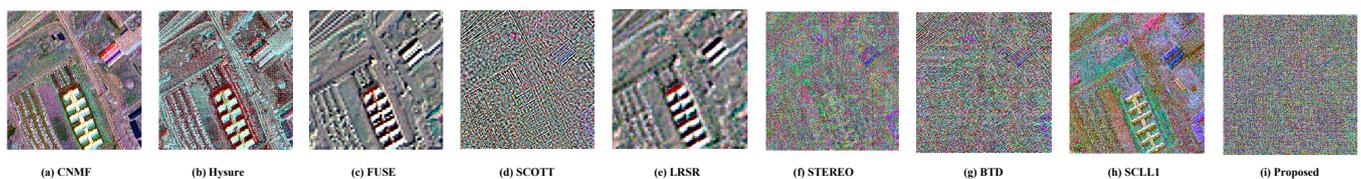


Figure 14. Pavia University dataset difference image.

As can be seen in the enlarged part shown in Figure 13, compared with the reference image, some spatial information is lost in the HRHS image of CNMF; a large number of blurred blocks and noise appear in the fusion result of SCOTT, and the spatial information is seriously missing; the fusion results of SCLL1 and the BTM are clearer; no blurred blocks and noise are seen, but some spatial information is lost. The HySure and STEREO fusion results in a low loss of spectral and spatial information; no noise and blurred blocks are seen in the fusion results of the algorithm proposed in this paper, and the spatial information is also preserved.

Figure 14 shows the difference images obtained after the difference between the target image and the fused image. Among these, the difference images of CNMF, Hysure, FUSE, LRSR, and SCLL1 have a lot of texture information; SCOTT has less texture information; the difference images of STEREO, the BTM, and the algorithm of this paper are more similar

and do not see more texture information, but, in comparison, STEREO and the BTM still have some building outline information.

When the PSNR is greater than 30 dB, it is difficult to observe the difference in the images with the naked eye, and the algorithm needs to be evaluated in terms of both the spectral and spatial features with the help of some evaluation metrics. The evaluation indexes of the Pavia dataset are shown in Table 2. As can be seen from the data in the table, the metrics describing the spectral and spatial information, the SSIM, CC, RMSE, ERGAS, and SAM of the algorithm in this paper, are closer to the optimal values, and only the UIQI is slightly inferior to the CNMF algorithm, which means that the algorithm in this paper performs the best in terms of spatial information, spectral information, image contrast, and brightness in the process of reconstructing images. In terms of the PSNR values, this paper's algorithm can effectively reject noise. Under the same number of iterations and in the same environment, the running time of this paper's algorithm, compared with the BTM and SCLL1, has been significantly improved.

Table 2. Performance for Pavia University Dataset.

Algorithms	CNMF	Hysure	FUSE	SOCTT	LRSR	STEREO	BTM	SCLL1	JSSLL1
PSNR(∞)	36.56	36.39	31.29	28.42	29.87	39.40	35.71	36.19	40.12
SSIM(1)	0.9627	0.9606	0.9240	0.7884	0.8929	0.9545	0.8764	0.9290	0.9652
CC(1)	0.9867	0.9853	0.9562	0.9458	0.9358	0.9913	0.9705	0.9795	0.9927
UIQI(1)	0.9343	0.9331	0.8818	0.7123	0.8299	0.9168	0.8354	0.8937	0.9341
RMSE(0)	0.0159	0.0169	0.0288	0.0314	0.0350	0.0120	0.0222	0.0198	0.0109
ERGAS(0)	0.4165	0.4395	0.7370	0.7415	0.7652	0.3086	0.5454	0.4528	0.2819
SAM(0)	0.0621	0.0684	0.0852	0.1051	0.1067	0.0643	0.0957	0.0932	0.0551
Time(0/s)	6.41	148.4	1.40	3.83	59.11	6.22	80.94	70.71	44.07

4.7.2. Pavia University Dataset Classification Experimental Results and Discussion

Image classification is one of the important application scenarios for high-resolution hyperspectral images. Acquiring high-resolution hyperspectral images is another important purpose of image fusion. A high-quality fused image will make the classification experiment "half the work". The experimental data contain information about buildings, vegetation, roads, bare soil, and other features. Therefore, a lightweight spectral-spatial convolutional network (LS^2CM) [73] with a low complexity and few network parameters was chosen to perform classification experiments on the fused images from Pavia University, as well as the original images for each algorithm. A 3×3 deep convolutional (Conv) filter was used for the experiments, and a filter with a convolution kernel of 1×1 was chosen for spectral feature extraction. A 3×3 filter was chosen for spatial feature extraction. All other parameter settings in this paper follow the literature [73]. Each experiment was repeated ten times. The results of the classification experiments on the Pavia dataset are shown in Table 3.

Table 3. Classification Performance for Pavia University Dataset.

Algorithms	CNMF	Hysure	FUSE	SOCTT	STEREO	BTM	SCLL1	JSSLL1	Original Image
KAPPA $\times 100$	96.52 \pm 1.66	96.27 \pm 0.88	94.40 \pm 7.07	94.90 \pm 5.08	97.78 \pm 1.00	97.71 \pm 2.13	94.99 \pm 5.00	97.30 \pm 2.36	97.93 \pm 1.51
OA(%)	97.50 \pm 1.21	97.32 \pm 0.62	95.89 \pm 5.31	96.24 \pm 3.87	98.40 \pm 0.71	98.36 \pm 1.51	96.40 \pm 3.54	98.08 \pm 1.65	98.51 \pm 1.09
AA(%)	95.52 \pm 1.91	95.66 \pm 1.67	95.43 \pm 3.64	93.66 \pm 5.50	97.09 \pm 1.69	97.52 \pm 1.35	94.56 \pm 4.86	97.03 \pm 1.90	97.20 \pm 3.08

As shown in Table 3, the HRHS image classification results of each algorithm are shown by the evaluation index. It should be specifically noted here that the HRHS of CNMF, Hysure, and the other algorithms are obtained by downsampling the original image, so the closer the classification result is to the original image, the better the classification result

will be. Therefore, compared with other algorithms, the HRHS classification result data of the algorithms in this paper, STEREO and the BTD, are closer to the reference image; the HRHS image classification effects of CNMF and Hysure are the second, and the HRHS image classification effects of FUSE, SCOTT and SCLL1 are the worst.

4.7.3. Indian Pines Dataset Fusion Experiment Results and Discussion

The parameters for the Indian Pines dataset are set to $\lambda = 0.01$, $\eta = 0.001$, and $SNR = 35$ dB. Due to the specificity of the model in this paper, the estimation and selection of the rank can be “overestimated”, so the roughly estimated ranks R and L are slightly adjusted upward and set to $L = 25$, $R = 10$. Here, the three bands 61, 25, and 13 are selected as the red, green, and blue bands, respectively, and fused to generate pseudo-color images. The HRHS is shown in Figure 15, and Figure 15a shows the original image of the Indian Pines dataset. To see the differences in the HRHS more clearly, the red-boxed part of each HRHS plot is enlarged, and the enlarged part is shown in the upper right corner. The difference images for all algorithms are shown in Figure 16.

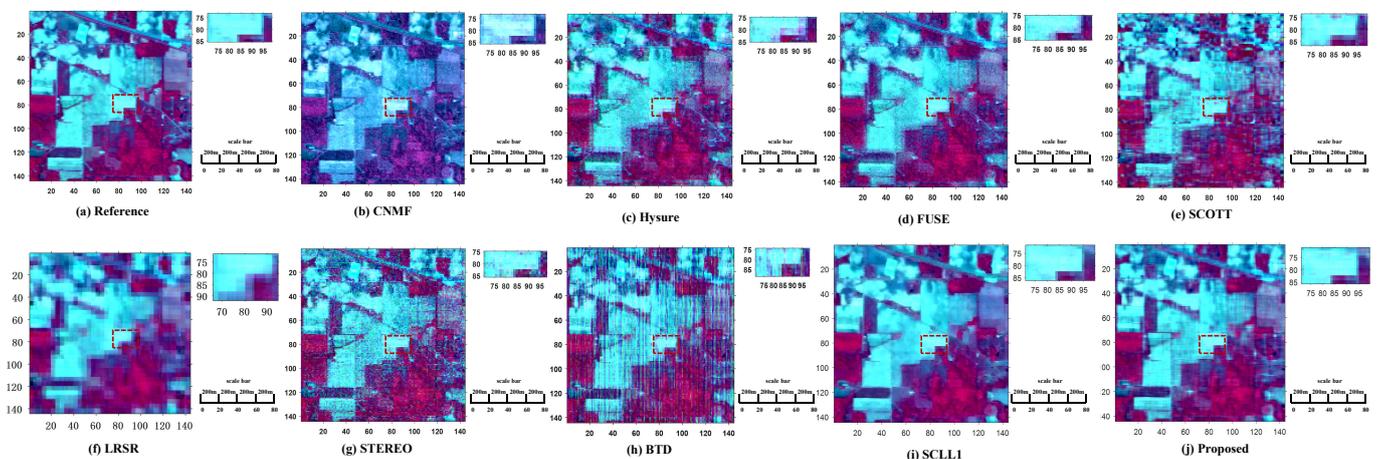


Figure 15. HRHSs of all algorithms on the Indian Pines dataset.

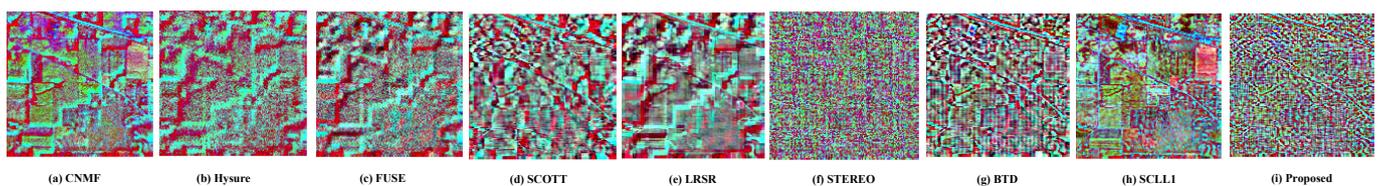


Figure 16. Indian Pines dataset difference image.

As can be seen in the enlarged part shown in Figure 15, compared with Figure 14a, the HRHS of CNMF appears distorted; the HRHS of Hysure, SCOTT, STEREO, and the BTD has a large fuzzy block and noise with serious loss of spatial information. The HRHS of FUSE, SCLL1, and this paper’s algorithm is clear; neither noise nor fuzzy block are seen, and the loss of spatial information is less.

Figure 16 shows the difference images of each algorithm on the Indian Pines dataset. Among these, the difference images of CNMF, Hysure, FUSE, SCOTT, LRSR, the BTD, and SCLL1 show a lot of spatial texture, indicating that there is a serious loss of spatial information in the HRHS obtained by the above algorithms; the difference images of STEREO and the algorithms in this paper do not show more spatial texture, indicating that the HRHS images of these two algorithms are closer to the reference image.

The data in Table 4 show that among the metrics describing the spectral and spatial information, the UIQI, CC, RMSE, ERGAS, and SAM of this algorithm are closer to the optimal values, and only the SSIM is slightly inferior to the SCLL1 algorithm, indicating that this algorithm loses less spatial and spectral information in the process of reconstructing

images, and the HRHS fused images are closer to the original dataset. In terms of the PSNR, this algorithm can better suppress noise; meanwhile, compared with the BTM and SCLL1, the running time of this algorithm is shorter.

Table 4. Performance for Indian Pines Dataset.

Algorithms	CNMF	Hysure	FUSE	SOCTT	LRSR	STEREO	BTD	SCLL1	JSSLL1
PSNR(∞)	27.55	30.44	31.24	27.76	26.90	26.27	25.38	32.08	33.79
SSIM(1)	0.9356	0.9078	0.9259	0.8586	0.8670	0.8166	0.7436	0.9551	0.9515
CC(1)	0.8175	0.8768	0.8793	0.7727	0.7754	0.8255	0.7091	0.8953	0.9338
UIQI(1)	0.6437	0.6438	0.6650	0.4804	0.4458	0.6800	0.4067	0.7298	0.7565
RMSE(0)	0.0241	0.0196	0.0167	0.0215	0.0264	0.4253	0.0306	0.0119	0.0112
ERGAS(0)	0.2736	0.2160	0.1873	0.2707	0.7754	0.0200	0.4431	0.1540	0.1204
SAM(0)	0.0664	0.0496	0.0401	0.0537	0.1793	0.3365	0.0850	0.0312	0.0261
Time(0/s)	2.73	47.44	0.75	0.58	31.54	4.05	7.26	23.85	5.72

4.8. Local Dataset Fusion Experimental Results and Discussion

The Ningxia Sand Lake dataset was acquired by two high-resolution satellites over the Sand Lake successively, and, here, the original dataset has been aligned and corrected by ENVI 5.3 software. By experimental verification, λ and η are not changed. R and L are set as $R = 20$, $L = 30$. Then, the three bands, 59, 38, and 20, are selected as the red, green, and blue bands, respectively, and fused to generate pseudo-color images. The HRHS is shown in Figure 17. Figure 17a,b show the HR-MSI and LR-HSI acquired by the satellites, respectively. In order to see the HRHS differences of each algorithm more clearly, each red boxed part of the HRHS plot is enlarged, and the enlarged part is shown in the upper right corner.

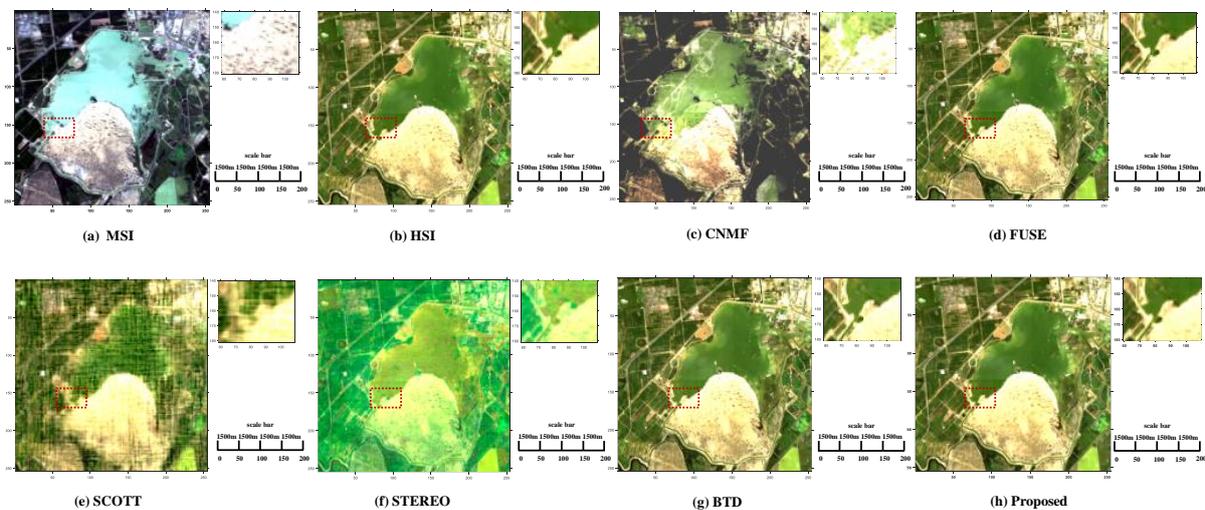


Figure 17. HRHSs of all algorithms on the Ningxia Sand Lake dataset.

As can be seen from the enlarged part in Figure 17, compared with the reference image, the HRHS of CNMF and STEREO showed distortion. The presence of a great number of fuzzy blocks and noise in the HRHS images of SCOTT and STEREO, shows that their spatial information was lost in the fusion process; the HRHS of FUSE and the BTD showed a small number of blurred blocks. In comparison, the HRHS of the algorithm in this paper was clearer; neither noise nor blurred blocks were seen, and less spatial information was lost.

The Hysure and SCLL1 algorithms show outliers when experimenting on the Ningxia Sand Lake dataset, and the LRSR algorithm also requires downsampling of the collected dataset for local dataset experiments, which is referred to as “pseudo-fusion”, while the algorithm in this paper directly fuses the two images. Therefore, these three algorithms are not compared. The numerical data of the other algorithms are shown in Table 5. Bold

means the best value. Among the indicators describing the spatial and spectral information, the UIQI, CC, RMSE, ERGAS, and SAM of the algorithms proposed in this paper are closer to the optimal values, and the SSIM is a little smaller than the BTD numerical indicators. This means that the algorithm in this paper can retain the spatial structure and spectral information in LR-HSI and HR-MSI images well during the process of reconstructing HRHS images; the PSNR value is also the highest, which shows that the proposed algorithm also performs well in suppressing noise.

Table 5. Performance for Ningxi Shahu Lake Dataset.

Algorithms	CNMF	FUSE	SOCTT	STEREO	BTD	JSSL1
PSNR(∞)	13.57	36.20	20.44	5.52	36.97	42.385
SSIM(1)	0.1649	0.9037	0.2383	0.2988	0.9753	0.92605
CC(1)	0.4650	0.9953	0.8690	0.8818	0.9960	0.9978
UIQI(1)	0.1773	0.9230	0.3617	0.3125	0.9390	0.9518
RMSE(0)	1067.9	114.44	374.84	922.06	50.57	34.8429
ERGAS(0)	0.8250	0.0780	0.3282	0.8229	0.0882	0.0648
SAM(0)	0.2505	0.0702	0.1746	0.3341	0.0436	0.0338
Time(0/s)	27.54	1.91	8.26	19.82	64.79	43.12

4.9. Analysis of Algorithm Complexity

The computational complexity of this model primarily involves matrix operations and element-wise computations. To calculate the complexity of the model, we can analyze the computational cost of each term and sum them up.

Each term in the model involves both matrix operations and element-wise computations. Specifically, the model consists of the following components:

$$\text{First term: } \frac{1}{2} \left\| \mathcal{Y}_H - \sum_{r=1}^R (\mathbf{P}_1 \mathbf{A}_r \mathbf{B}_r^T \mathbf{P}_2^T) \circ \mathbf{c}_r \right\|_F^2. \text{ Second term: } \frac{1}{2} \left\| \mathcal{Y}_M - \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ (\mathbf{P}_3 \mathbf{c}_r) \right\|_F^2.$$

$$\text{Regularization term: } \lambda \sum_{r=1}^R \sqrt{\left(\sum_{l=1}^L \sqrt{\|a_{rl}\|_2^2 + \|b_{rl}\|_2^2 + \eta^2} \right)^2 + \|\mathbf{c}_r\|_2^2 + \eta^2}.$$

For each term, we can separately consider the computational complexity of matrix operations and element-wise computations.

The first term involves matrix multiplication, element-wise multiplication, and summation. The most time-consuming part is the matrix multiplication. Assuming \mathcal{Y}_H has dimensions $H \times W$, \mathbf{A}_r and \mathbf{B}_r have dimensions $H \times K$ and $W \times K$, and \mathbf{P}_1 and \mathbf{P}_2 are dimension-adaptation matrices. The matrix multiplication $(\mathbf{P}_1 \mathbf{A}_r \mathbf{B}_r^T \mathbf{P}_2^T)$ has a complexity of approximately $O(HWK)$, followed by element-wise multiplication and summation with a complexity of $O(HW)$.

The second term is similar to the first term, involving matrix multiplication, element-wise multiplication, and summation. Assume \mathcal{Y}_M has dimensions $H \times W$, \mathbf{A}_r and \mathbf{B}_r have dimensions $H \times K$ and $W \times K$, and \mathbf{P}_3 is a dimension-adaptation matrix. The matrix multiplication $(\mathbf{A}_r \mathbf{B}_r^T)$ has a complexity of $O(HWK)$, followed by element-wise multiplication and summation with a complexity of $O(HW)$.

The regularization term involves element-wise operations such as squaring, square root, summation, and multiplication. Assuming each vector a_{rl} , b_{rl} , and c_r has a dimension of K , the computational complexity for each r is approximately $O(K)$, resulting in a total complexity of $O(RK)$.

Therefore, adding up the complexities of these three components, the overall computational complexity of the model is approximately $O(HWK + HW + RK)$.

5. Summary

In this paper, a coupled non-negative block-term tensor decomposition model based on joint-structured sparsity is proposed for estimating HRHSs. A hierarchically structured regularization scheme is adopted to fully consider the unknown nature of the tensor rank, the long running time of the algorithm, the sparsity of the abundance matrix, and the

scaling/counter-scaling effects that occur in the fusion model when the block-term tensor decomposition model is used to reconstruct the HRHS. Firstly, the sparsity of the abundance matrix is ensured, and the scaling phenomenon in the model is countered by superimposing the abundance matrix on the chunk matrix and adding the $L_{2,1}$ -norm; secondly, the counter-scaling effect in the model is suppressed by adding the L_2 -norm to each endmember vector; finally, the chunk matrix and the endmember matrix are reweighted and coupled, adding a $L_{2,1}$ -norm number, making it possible to select the appropriate regularization parameters to overestimate the tensor rank when the rank is unknown, thereby facilitating the elimination of the abundance matrix and the corresponding endmember matrix. To address the above facts, this paper conducts experiments based on standard and local datasets and solves the algorithmic model using the extended IRLS method. The experimental results show that the algorithm proposed in this paper performs well on both spatial and spectral information, in terms of both the visual sensory and numerical comparisons, and its fusion results are closer to the reference image. There are still some limitations in this paper, for example, the performance of the algorithm depends on the choice of multiple preset parameters and hyperparameters, such as the order of the matrix factorization, regularization parameters, etc. Different parameter settings may lead to different results and careful parameter tuning is required. Due to multiple matrix operations, element-wise operations, and iterative optimization, the computational complexity of the algorithm is high. This can be time-consuming when dealing with large amounts of data. In future work, we hope to further reduce the running time of the algorithm without compromising its overall performance.

Author Contributions: Funding acquisition, W.B.; Validation, K.Q. and W.F.; Writing—original draft, H.G.; Writing—review and editing, S.S. and C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study received support from multiple sources, including the National Natural Science Foundation of China (Project No. 62201438), the Ningxia Key R&D Program of China (Project No. 2021BEG03030), and the Innovation Projects for Graduate Students of North Minzu University (Project No.YCX23151).

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: https://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.

Acknowledgments: The authors would like to thank the Image and Intelligence Information Processing Innovation Team of the National Ethnic Affairs Commission of China for their support.

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

Abbreviations

The following abbreviations are used in this manuscript:

MHF	Multispectral and hyperspectral image fusion
HRHS	High spatial resolution hyperspectral image
LR-HSI	Low-resolution hyperspectral image
HR-MSI	High-resolution multispectral image
HSR	Hyperspectral super-resolution
PAN	Panchromatic image
CPD	Canonical polyadic decomposition
TV	Total variational
PAO	Proximal alternating optimization
ADMM	Alternating directional multiplier method
SNR	Signal-to-noise ratio
PCA	Principal component analysis
PSNR	Peak signal-to-noise ratio
SSIM	Structural similarity
CC	Correlation coefficient
UIQI	Universal image quality index

RMSE	Root mean squared error
ERGAS	Erreur Relative Globale Adimensionnelle De Synthèse
SAM	Spectral angle mapper

Appendix A. Symbol Definition

Upper- and lowercase bold letters are used to represent the vector and matrix, respectively. The higher-order tensor is denoted by \mathcal{Y} . For a tensor \mathcal{Y} , $\mathbf{Y}_{(n)}$ denotes its n th order expansion. \otimes denotes the Kronecker product. The Khatri–Rao product is denoted by \odot in its general version and by \odot_c in the columnar version. \circ denotes the outer product. The superscript T indicates a transpose. The unit matrix of order is denoted by I_N . The L_1 -norm, Euclidean, mixed $L_{2,1}$ -norm, and Frobenius norm numbers are, respectively, denoted by $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_{2,1}$, and $\|\cdot\|_F$. \mathbb{R} denotes the complex number field.

Appendix B. Algorithm Convergence Discussion

L_2 parametric regularization itself is a common method to prevent overfitting of the model, and the $L_{2,1}$ parametric is a hybrid parametric between the L_1 parametric and the L_2 parametric, which not only has the property of the L_2 parametric to prevent model overfitting but also has the sparse facilitation of the L_1 parametric to select the important features and discard the unimportant ones. Regarding this paper, the fusion model is shown in Equation (17). The three factors **A**, **B**, and **C** in this model with their associated minimization problems are Equations (20), (22) and (24).

Regarding the general IRLS [66], the expression is as follows:

$$x^{k+1} = \arg \min_x \frac{1}{2} \|b - Rx\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^n \frac{x_i^2}{\sqrt{x_i^{k+2} + \eta^2}} \quad (\text{A1})$$

The two reweighting matrices can be derived by comparing Equations (20), (22) and (24) with Equation (A1), respectively. The first matrix \mathbf{O}_1 is composed of the inverse of the summation terms external to the regularized terms in model Equation (17), which is the joint weighted **A**, **B** block; the second matrix \mathbf{O}_2 is composed of the inverse of the terms internal to the summation of the regularized terms in Equation (17), which is used to balance the corresponding columns of \mathbf{A}_r and \mathbf{B}_r . Here, the scheme of a hierarchical IRLS structure is generated. As the iterative process progresses, \mathbf{O}_1 and \mathbf{O}_2 gradually move closer to the direction with smaller weight parameters (i.e.: decaying weights), thus avoiding overfitting. The conjugate gradient algorithm is a method between the fastest descent method and the Newton method, which overcomes the disadvantage of the slow convergence of the fastest descent method and avoids the problem that the Newton method needs to store and calculate the Hessian matrix and find the inverse. Therefore, we choose the CG algorithm to complete the solution of the three factors **A**, **B**, and **C** to ensure that each factor is convergent.

References

1. Plowright, A.A.; Coops, N.C.; Chance, C.M.; Sheppard, S.R.; Aven, N.W. Multi-scale analysis of relationship between imperviousness and urban tree height using airborne remote sensing. *Remote Sens. Environ.* **2017**, *194*, 391–400. [\[CrossRef\]](#)
2. Posselt, R.; Mueller, R.; Stöckli, R.; Trentmann, J. Remote sensing of solar surface radiation for climate monitoring—The CM-SAF retrieval in international comparison. *Remote Sens. Environ.* **2012**, *118*, 186–198. [\[CrossRef\]](#)
3. Matikainen, L.; Karila, K.; Hyypä, J.; Litkey, P.; Puttonen, E.; Ahokas, E. Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating. *ISPRS J. Photogramm. Remote Sens.* **2017**, *128*, 298–313. [\[CrossRef\]](#)
4. Loncan, L.; De Almeida, L.B.; Bioucas-Dias, J.M.; Briottet, X.; Chanussot, J.; Dobigeon, N.; Fabre, S.; Liao, W.; Licciardi, G.A.; Simoes, M.; et al. Hyperspectral pansharpening: A review. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 27–46. [\[CrossRef\]](#)
5. Thomas, C.; Ranchin, T.; Wald, L.; Chanussot, J. Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1301–1312. [\[CrossRef\]](#)
6. Meng, X.; Shen, H.; Li, H.; Zhang, L.; Fu, R. Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges. *Inf. Fusion* **2019**, *46*, 102–113. [\[CrossRef\]](#)

7. Carper, W.; Lillesand, T.; Kiefer, R. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogramm. Eng. Remote Sens.* **1990**, *56*, 459–467.
8. 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](#)]
9. Liu, J. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *Int. J. Remote Sens.* **2010**, *21*, 3461–3472. [[CrossRef](#)]
10. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. An MTF-based spectral distortion minimizing model for pansharpening of very high resolution multispectral images of urban areas. In Proceedings of the 2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, Berlin, Germany, 22–23 May 2003; IEEE: Piscataway, NJ, USA, 2003; pp. 90–94.
11. 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](#)]
12. Leung, Y.; Liu, J.; Zhang, J. An improved adaptive intensity–hue–saturation method for the fusion of remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *11*, 985–989. [[CrossRef](#)]
13. Gomez, R.B.; Jazaeri, A.; Kafatos, M. Wavelet-based hyperspectral and multispectral image fusion. In *Geo-Spatial Image and Data Exploitation II*; SPIE: St Bellingham, WA, USA, 2001; Volume 4383, pp. 36–42.
14. Zhang, Y.; He, M. Multi-spectral and hyperspectral image fusion using 3-D wavelet transform. *J. Electron. (China)* **2007**, *24*, 218–224. [[CrossRef](#)]
15. 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.
16. Li, H.; Manjunath, B.; Mitra, S.K. Multisensor image fusion using the wavelet transform. *Graph. Model. Image Process.* **1995**, *57*, 235–245. [[CrossRef](#)]
17. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. [[CrossRef](#)]
18. Yokoya, N.; Yairi, T.; Iwasaki, A. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 528–537. [[CrossRef](#)]
19. Bendoumi, M.A.; He, M.; Mei, S. Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6574–6583. [[CrossRef](#)]
20. Berné, O.; Helens, A.; Pilleri, P.; Joblin, C. Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy. In Proceedings of the 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Reykjavik, Iceland, 14–16 June 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1–4.
21. 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.* **2014**, *53*, 3373–3388. [[CrossRef](#)]
22. Wei, Q.; Dobigeon, N.; Tourneret, J.Y. Fast fusion of multi-band images based on solving a Sylvester equation. *IEEE Trans. Image Process.* **2015**, *24*, 4109–4121. [[CrossRef](#)]
23. 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](#)]
24. Veganzones, M.A.; Simoes, M.; Licciardi, G.; Yokoya, N.; Bioucas-Dias, J.M.; Chanussot, J. Hyperspectral super-resolution of locally low rank images from complementary multisource data. *IEEE Trans. Image Process.* **2015**, *25*, 274–288. [[CrossRef](#)] [[PubMed](#)]
25. Prévost, C.; Usevich, K.; Comon, P.; Brie, D. Hyperspectral super-resolution with coupled tucker approximation: Recoverability and SVD-based algorithms. *IEEE Trans. Signal Process.* **2020**, *68*, 931–946. [[CrossRef](#)]
26. Dian, R.; Fang, L.; Li, S. Hyperspectral image super-resolution via non-local sparse tensor factorization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5344–5353.
27. Xu, T.; Huang, T.Z.; Deng, L.J.; Zhao, X.L.; Huang, J. Hyperspectral image superresolution using unidirectional total variation with tucker decomposition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4381–4398. [[CrossRef](#)]
28. Li, S.; Dian, R.; Fang, L.; Bioucas-Dias, J.M. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. *IEEE Trans. Image Process.* **2018**, *27*, 4118–4130. [[CrossRef](#)]
29. Zhang, K.; Wang, M.; Yang, S. Multispectral and hyperspectral image fusion based on group spectral embedding and low-rank factorization. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 1363–1371. [[CrossRef](#)]
30. Li, W.; Du, Q. A survey on representation-based classification and detection in hyperspectral remote sensing imagery. *Pattern Recognit. Lett.* **2016**, *83*, 115–123. [[CrossRef](#)]
31. Zhang, K.; Wang, M.; Yang, S.; Jiao, L. Spatial–spectral–graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1030–1040. [[CrossRef](#)]
32. Li, Q.; Ma, W.K.; Wu, Q. Hyperspectral super-resolution: Exact recovery in polynomial time. In Proceedings of the 2018 IEEE Statistical Signal Processing Workshop (SSP), Breisgau, Germany, 10–13 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 378–382.
33. Feng, Z.; Yang, S.; Wei, X.; Gao, Q.; Jiao, L. Regularized Sparse Band Selection via Learned Pairwise Agreement. *IEEE Access* **2020**, *8*, 40096–40105. [[CrossRef](#)]
34. Huang, S.; Zhang, H.; Pižurica, A. Subspace clustering for hyperspectral images via dictionary learning with adaptive regularization. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–17. [[CrossRef](#)]

35. Qin, X.; Ban, Y.; Wu, P.; Yang, B.; Liu, S.; Yin, L.; Liu, M.; Zheng, W. Improved image fusion method based on sparse decomposition. *Electronics* **2022**, *11*, 2321. [[CrossRef](#)]
36. Zhang, D.; Wang, T.; Yang, S.; Jia, Y.; Li, F. Spectral reweighting and spectral similarity weighting for sparse hyperspectral unmixing. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
37. Chengqing, Y.; Guangxi, Y.; Chengming, Y.; Yu, Z.; Xiwei, M. A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks. *Energy* **2023**, *263*, 126034. [[CrossRef](#)]
38. Wu, R.; Chan, C.H.; Wai, H.T.; Ma, W.K.; Fu, X. Hi, BCD! hybrid inexact block coordinate descent for hyperspectral super-resolution. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2426–2430.
39. Wu, R.; Ma, W.K.; Fu, X.; Li, Q. Hyperspectral super-resolution via global-local low-rank matrix estimation. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7125–7140. [[CrossRef](#)]
40. Kawakami, R.; Matsushita, Y.; Wright, J.; Ben-Ezra, M.; Tai, Y.W.; Ikeuchi, K. High-resolution hyperspectral imaging via matrix factorization. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2329–2336.
41. Lanaras, C.; Baltasvias, E.; Schindler, K. Hyperspectral super-resolution by coupled spectral unmixing. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3586–3594.
42. Kahraman, S.; Ertürk, A.; Ertürk, S. Graph Regularized L_{1/2}-Sparsity Constrained Non-Negative Matrix Factorization for Hyperspectral and Multispectral Image Fusion. In Proceedings of the 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 23–26 September 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–4.
43. Mu, T.; Nie, R.; Ma, C.; Liu, J. Hyperspectral and panchromatic image fusion based on CNMF. In Proceedings of the 2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC), Shanghai, China, 23–25 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 293–297.
44. Khader, A.; Yang, J.; Xiao, L. NMF-DuNet: Nonnegative matrix factorization inspired deep unrolling networks for hyperspectral and multispectral image fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5704–5720. [[CrossRef](#)]
45. Han, Y.; Bao, W.; Zhang, X.; Ma, X.; Cao, M. Hyperspectral and multispectral data fusion via nonlocal low-rank learning. *J. Appl. Remote Sens.* **2022**, *16*, 016508. [[CrossRef](#)]
46. Lin, B.; Guo, Y. Deep Hyperspectral and Multispectral Image Fusion via Probabilistic Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–14. [[CrossRef](#)]
47. Kanatsoulis, C.I.; Fu, X.; Sidiropoulos, N.D.; Ma, W.K. Hyperspectral super-resolution: A coupled tensor factorization approach. *IEEE Trans. Signal Process.* **2018**, *66*, 6503–6517. [[CrossRef](#)]
48. Xu, Y.; Wu, Z.; Chanussot, J.; Comon, P.; Wei, Z. Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 348–362. [[CrossRef](#)]
49. Peng, Y.; Li, W.; Luo, X.; Du, J. Hyperspectral image superresolution using global gradient sparse and nonlocal low-rank tensor decomposition with hyper-laplacian prior. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5453–5469. [[CrossRef](#)]
50. Prévost, C.; Chainais, P.; Boyer, R. Fast fusion of hyperspectral and multispectral images: A tucker approximation approach. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 2076–2080.
51. Zhao, Q.; Zhou, G.; Xie, S.; Zhang, L.; Cichocki, A. Tensor ring decomposition. *arXiv* **2016**, arXiv:1606.05535.
52. Chen, Y.; Zeng, J.; He, W.; Zhao, X.L.; Huang, T.Z. Hyperspectral and Multispectral Image Fusion Using Factor Smoothed Tensor Ring Decomposition. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [[CrossRef](#)]
53. Qian, Y.; Xiong, F.; Zeng, S.; Zhou, J.; Tang, Y.Y. Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 1776–1792. [[CrossRef](#)]
54. Zhang, G.; Fu, X.; Huang, K.; Wang, J. Hyperspectral super-resolution: A coupled nonnegative block-term tensor decomposition approach. In Proceedings of the 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Guadeloupe, West Indies, 15–18 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 470–474.
55. De Lathauwer, L. Decompositions of a higher-order tensor in block terms—Part I: Lemmas for partitioned matrices. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 1022–1032. [[CrossRef](#)]
56. De Lathauwer, L. Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 1033–1066. [[CrossRef](#)]
57. De Lathauwer, L.; Nion, D. Decompositions of a higher-order tensor in block terms—Part III: Alternating least squares algorithms. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 1067–1083. [[CrossRef](#)]
58. Ding, M.; Fu, X.; Huang, T.Z.; Wang, J.; Zhao, X.L. Hyperspectral super-resolution via interpretable block-term tensor modeling. *IEEE J. Sel. Top. Signal Process.* **2020**, *15*, 641–656. [[CrossRef](#)]
59. Guo, H.; Bao, W.; Qu, K.; Ma, X.; Cao, M. Multispectral and Hyperspectral Image Fusion Based on Regularized Coupled Non-Negative Block-Term Tensor Decomposition. *Remote Sens.* **2022**, *14*, 5306. [[CrossRef](#)]
60. Hitchcock, F.L. The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.* **1927**, *6*, 164–189. [[CrossRef](#)]
61. Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, *31*, 279–311. [[CrossRef](#)]

62. Liu, N.; Li, W.; Wang, Y.; Tao, R.; Du, Q.; Chanussot, J. A survey on hyperspectral image restoration: From the view of low-rank tensor approximation. *Inf. Sci.* **2023**, *66*, 140302. [[CrossRef](#)]
63. Hunyadi, B.; Camps, D.; Sorber, L.; Paesschen, W.V.; Vos, M.D.; Huffel, S.V.; Lathauwer, L.D. Block term decomposition for modelling epileptic seizures. *EURASIP J. Adv. Signal Process.* **2014**, *2014*, 139. [[CrossRef](#)]
64. Beck, A. On the convergence of alternating minimization with applications to iteratively reweighted least squares and decomposition schemes. *Optim. Online* **2013**, *25*, 185–209. [[CrossRef](#)]
65. Rontogiannis, A.A.; Kofidis, E.; Giampouras, P.V. Block-term tensor decomposition: Model selection and computation. *IEEE J. Sel. Top. Signal Process.* **2021**, *15*, 464–475. [[CrossRef](#)]
66. Daubechies, I.; DeVore, R.; Fornasier, M.; Güntürk, C.S. Iteratively reweighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math. J. Issued Courant Inst. Math. Sci.* **2010**, *63*, 1–38. [[CrossRef](#)]
67. Liu, J.S.; Chen, R. Sequential Monte Carlo methods for dynamic systems. *J. Am. Stat. Assoc.* **1998**, *93*, 1032–1044. [[CrossRef](#)]
68. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379. [[CrossRef](#)]
69. Sui, L.; Li, L.; Li, J.; Chen, N.; Jiao, Y. Fusion of hyperspectral and multispectral images based on a Bayesian nonparametric approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1205–1218. [[CrossRef](#)]
70. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [[CrossRef](#)]
71. Wald, L. Quality of high resolution synthesised images: Is there a simple criterion? In Proceedings of the Third Conference “Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images”, SEE/URISCA, Sophia Antipolis, France, 26–28 January 2000; pp. 99–103.
72. 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*; NASA: Washington, DC, USA, 1992; Volume 1, pp. 147–149.
73. Meng, Z.; Jiao, L.; Liang, M.; Zhao, F. A lightweight spectral-spatial convolution module for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [[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.