remote sensing

MDPI

Article

Correlation Matrix-Based Fusion of Hyperspectral and Multispectral Images

Hong Lin [1], Jun Li [2,*], Yuanxi Peng [1], Tong Zhou [1], Jian Long [1] and Jialin Gui [1]

[1] State Key Laboratory of High-Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China; linhong11a@nudt.edu.cn (H.L.); pyx@nudt.edu.cn (Y.P.); zhoutong09@nudt.edu.cn (T.Z.); jian_long@nudt.edu.cn (J.L.); guijialin23@nudt.edu.cn (J.G.)

[2] College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China

* Correspondence: lijun14@nudt.edu.cn

Abstract: The fusion of the hyperspectral image (HSI) and the multispectral image (MSI) is commonly employed to obtain a high spatial resolution hyperspectral image (HR-HSI); however, existing methods often involve complex feature extraction and optimization steps, resulting in time-consuming fusion processes. Additionally, these methods typically require parameter adjustments for different datasets. Still, reliable references for parameter adjustment are often unavailable in practical scenarios, leading to subpar fusion results compared to simulated scenarios. To address these challenges, this paper proposes a fusion method based on a correlation matrix. Firstly, we assume the existence of a correlation matrix that effectively correlates the spectral and spatial information of HSI and MSI, enabling fast fusion. Subsequently, we derive a correlation matrix that satisfies the given assumption by deducing the generative relationship among HR-HSI, HSI, and MSI. Finally, we optimize the fused result using the Sylvester equation. We tested our proposed method on two simulated datasets and one real dataset. Experimental results demonstrate that our method outperforms existing state-of-the-art methods. Particularly, in terms of fusion time, our method achieves fusion in less than 0.1 seconds in some cases. This method provides a practical and feasible solution for the fusion of hyperspectral and multispectral images, overcoming the challenges of complex fusion processes and parameter adjustment while ensuring a quick fusion process.

Keywords: hyperspectral image; super-resolution; fusion; correlation matrix

check for updates

Citation: Lin, H.; Li, J.; Peng, Y.; Zhou, T.; Long, J.; Gui, J. Correlation Matrix-Based Fusion of Hyperspectral and Multispectral Images. *Remote Sens.* **2023**, *15*, 3643. https://doi.org/10.3390/rs15143643

Academic Editor: András Jung

Received: 16 June 2023
Revised: 16 July 2023
Accepted: 18 July 2023
Published: 21 July 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

Hyperspectral images have contributed significantly to various fields, including agriculture [1–4], environmental monitoring [5,6], image processing [7,8], mining [9–11], and urban planning [12–14]. Two commonly used imaging techniques in remote sensing are hyperspectral imaging and multispectral imaging. Hyperspectral imaging captures detailed information from hundreds of narrow and contiguous spectral bands, enabling a comprehensive material characterization based on spectral features [15]; however, it often suffers from lower spatial resolution [16]. On the other hand, multispectral imaging captures data in a broader range of spectral bands but with fewer spectral details, providing higher spatial resolution [17]. Therefore, widespread research has focused on improving the spatial resolution of hyperspectral images by fusing hyperspectral and multispectral images [18,19]. The fused HR-HSI obtained combines the rich spectral information of HSI with the spatial information of MSI. This fusion technique has been widely applied in various tasks, including change detection [20], anomaly detection [21], and target classification [22]. This indicates the significant application value of fusion techniques in remote sensing image processing.

However, the current methods used for hyperspectral and multispectral image fusion often prioritize fusion quality and overlook the requirement for fusion time, thereby limiting their feasibility in practical applications. Taking the work [23] as an example, it assumes that HR-HSI consists of a spectral base and a corresponding coefficient matrix. The spectral base can be approximated by performing singular value decomposition on the HSI. Solving the coefficient matrix requires dividing the MSI into patches, performing clustering learning to obtain the corresponding coefficient matrix, and further optimizing the coefficient matrix. Although this method can achieve a high-quality HR-HSI, the fusion process is relatively time-consuming. The situation is even more pronounced for deep learning-based methods. Although deep learning-based methods require less prior information, the training time of the models is often excessively long, and the trained models are usually only applicable to specific datasets. They cannot be directly used for new datasets. When facing new datasets, retraining the models is typically necessary. Additionally, most fusion methods involve non-fixed parameters that must be adjusted based on different datasets. However, we often cannot obtain effective reference images like experiments to adjust these parameters in practical scenarios. The non-fixed parameters can result in inconsistent fusion performance when the downsampling factor varies within the same dataset or when dealing with different datasets. These two drawbacks significantly impact the effectiveness of these methods in practical scenarios.

In current fusion methods, it is common practice to extract spectral and spatial information from HSI and MSI and reconstruct them to obtain HR-HSI. The work [24] above is a typical example of this approach. However, HSI and MSI already contain the spectral and spatial information required for reconstructing HR-HSI. Therefore, is it possible to use HSI and MSI directly without additional operations for the fusion step? In other words, can we combine HSI and MSI's spectral and spatial information using a single matrix instead of complex information extraction processes? This strategy can reduce computational complexity and shorten the fusion time. Inspired by this, we propose a fusion method based on a correlation matrix. We assume the existence of a correlation matrix $\mathbf{M}$ that connects the spectral information of the hyperspectral image (HSI) and the spatial information of the multispectral image (MSI), enabling fast fusion through simple matrix operations with HSI and MSI. Figure 1 provides an overview of this assumption. Since MSI and HSI are typically spectral and spatial downsampling of HR-HSI [23], we derive the correlation matrix $\mathbf{M}$ that satisfies our assumption. Detailed theoretical derivations will be provided in subsequent sections.

The fusion method proposed in this paper can be divided into two stages. We first utilize the obtained correlation matrix $\mathbf{M}$ to fusion the HSI and MSI via simple matrix operations, as shown in Figure 1. We name this stage CMF (correlation matrix-based fusion). CMF enables the rapid generation of high-quality HR-HSI without any parameters. As the fusion process only involves matrix operations, the fusion time required is short, addressing the issue of long fusion time. Additionally, no additional parameters are introduced in this stage.

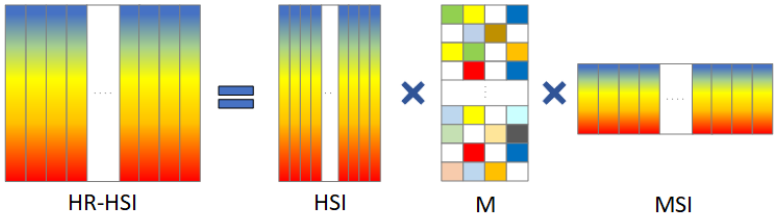


Figure 1. Correlation matrix assumption.

We name the second stage CMF+, where we optimize the fusion result from the first stage using the Sylvester equation in this stage. We only introduce a fixed parameter for this stage. Since the solution of the Sylvester equation will take little time, this stage does not significantly increase the time overhead. Our method utilizes simple computations

throughout the fusion process, and no parameters need to be adjusted. As a result, we can reconstruct HR-HSI quickly within a brief time.

The main contributions of this paper are as follows:

- We give an assumption of the correlation matrix, which establishes the correlation between the spectral information of the hyperspectral image (HSI) and the spatial information of the multispectral image (MSI). Through this correlation matrix, we avoid complex computational processes, greatly simplify the fusion process, and reduce the fusion time.
- The proposal of the correlation matrix assumption offers a novel method for future fusion processes, enabling efficient fusion through the solution of a matrix that fulfills the defined correlation matrix.
- Based on the generative relationship among HSI, MSI, and HR-HSI, we derive a method to solve the correlation matrix and construct a new fusion model using this correlation matrix. We achieve the initial fusion by performing simple matrix operations on HSI, MSI, and the correlation matrix. We further optimize the preliminary fusion result using the Sylvester equation. The entire fusion process is more straightforward compared to common fusion methods, without the need for complex operations or parameter adjustments.
- Experimental results on two simulated datasets and one real dataset validate the superiority of our proposed method over current state-of-the-art fusion methods. We can obtain high-quality fusion results while simplifying the fusion process and significantly improving fusion time.

We will organize the remaining sections of this paper according to the following structure. Section 2 will introduce some classical literature and methods. Section 3 will provide a detailed description of our proposed method. Section 4 will present the experimental details. Section 5 will showcase the experimental results and corresponding analysis. Finally, Section 6 will provide a conclusion of our research.

2. Related Work

Hyperspectral and multispectral image fusion is an important research area that has gained significant attention and interest. Over the past few decades, numerous methods and techniques have been proposed to improve the fusion performance between hyperspectral and multispectral images. These methods can be broadly categorized into four groups, Bayesian-based, matrix factorization-based, tensor factorization-based, and deep learning-based.

Bayesian-based methods utilize Bayesian theory to model the relationship between spectral and spatial information and achieve fusion through statistical inference. These methods typically rely on prior knowledge and probability models for image fusion. The fusion method proposed in work [25] utilizes a Bayesian non-parametric method with a Beta-Bernoulli process to learn dictionary elements and coefficient weights for achieving fusion between hyperspectral and multispectral images. Qi et al. [24] transformed the fusion problem into a Bayesian estimation framework. It introduces a prior distribution based on the linear mixing model and employs a Gibbs sampling algorithm with a Hamiltonian Monte Carlo step to obtain the posterior distribution required for fusion.

Matrix factorization-based methods represent hyperspectral and multispectral images in matrix form and employ techniques like principal component analysis (PCA) and non-negative matrix factorization (NMF) to extract underlying features, facilitating fusion. These methods effectively reduce data redundancy and preserve crucial information through dimensionality reduction and feature extraction. A novel adaptive non-negative sparse representation model is proposed for fusion in work [26]. Using a non-negative structured sparse representation model. The method begins with linear spectral unmixing to estimate the sparse coding of the spectral bases for HR-HSI from HSI and MSI. It generates a balance between sparsity and collaboration coefficients through adaptive sparse representation and performs alternating optimization on the spectral bases and coefficients. Ref. [27] considers HR-HSI as a composition of a hyperspectral dictionary and sparse codes. The method

proposes a non-negative dictionary learning approach based on block-coordinate descent optimization to learn the spectral dictionary from HSI. A clustering-based structured sparse coding method is introduced to solve the corresponding sparse codes. Huang et al. [28] present the SASFM fusion model. Based on reasonable assumptions, this model first learns a spectral dictionary from the HSI data. Then, the learned spectral dictionary and known MSI are used to predict HR-HSI.

The tensor factorization-based methods utilize tensors to represent the relationships between hyperspectral and multispectral images and achieve fusion through tensor operations. These methods can leverage the high-dimensional structural information of the images more effectively, thus improving the fusion performance. Li et al. [29] treat HR-HSI as composed of three sub-tensors and a sparse core tensor. This method proposes a fusion approach based on coupled sparse tensor representation, which achieves a high-quality fusion result through approximate alternating optimization. In ref. [30], a low tensor-train rank prior is utilized to learn the correlations among the HR-HSI cube's spatial, spectral, and non-local patterns. The HR-HSI is uniformly clustered based on the clustering structure in the MSI, and the LTTR (low tensor-train rank) constraint is imposed on each clustered block, effectively learning the spatial, spectral, and non-local similarities. The optimization is performed using alternating direction multiplication.

Methods based on deep learning utilize deep neural network models to learn the nonlinear mapping relationship between hyperspectral and multispectral images. By training the network models, they can automatically extract and learn feature representations of the images, achieving high-quality image fusion.

A novel coupled unmixing network (CUCaNet) with a cross-attention mechanism is proposed by Yao et al. [31]. The dual-stream convolutional auto-encoder framework serves as the backbone, and the MSI and HSI data are jointly decomposed into bases with spectral significance and corresponding coefficients for fusion. Work [32] presents the CNN-Fus network, which achieves unsupervised fusion without training. Based on the high correlation between spectral bands, the HR-HSI is obtained by multiplying a low-dimensional subspace by coefficients. Subspace is learned from HSI using singular value decomposition, and coefficients are estimated using a CNN denoiser inserted into the alternating multiplier method algorithm. Liu et al. [33] design a model-inspired deep network with an implicit auto-encoder network that treats each HR-HSI pixel as a separate sample. The non-negative matrix factorization (NMF) of the target HR-HSI is integrated into the auto-encoder network, consisting of spectral and spatial matrix NMF parts serving as decoder parameters and hidden outputs, respectively. Finally, a pixel-wise fusion model is proposed for fusion during the encoding stage.

In the above classification, methods based on Bayesian, matrix factorization, and tensor factorization are traditional fusion methods. This is primarily because they focus on constructing mathematical models and algorithms, utilizing traditional mathematical tools, such as statistics, matrix decomposition, and tensor decomposition to achieve image fusion. These methods typically emphasize modeling and analyzing the data to improve fusion performance and effectiveness.

In contrast, modern deep learning-based methods emphasize end-to-end feature learning and fusion using deep neural networks. They go beyond the limitations of traditional mathematical tools and techniques. These deep learning methods leverage the power of neural networks to learn hierarchical representations directly from the data and perform fusion in a more data-driven and automated manner.

3. Fusion Model

It is common to represent HSI, MSI, and HR-HSI using three-dimensional data. For clarity, in this paper, we define them using different symbols. We use the $\mathcal{Z} \in \mathbb{R}^{W \times H \times L}$ to denote the target HR-HSI, which has $W \times H$ pixels and L spectral bands. The $\mathcal{X} \in \mathbb{R}^{w \times h \times L}$ means the HSI, which comprises $w \times h$ pixels and L spectral bands. Finally, the $\mathcal{Y} \in \mathbb{R}^{W \times H \times l}$ represents the MSI, with $W \times H$ pixels and l bands. Since HSI contains more spectral

information than MSI, we can express this relationship as $L > l$. Similarly, $W > w$ and $H > h$ denote that the spatial resolution in MSI is larger than that in HSI.

The fusion method proposed in this paper is based on matrix operations. The basic principle is to use the hyperspectral image (HSI), multispectral image (MSI), and a correlation matrix to perform matrix multiplication for fusion. Therefore, we unfold the respective tensors along the spectral dimension and represent them as matrices, denoted as $\mathbf{Z} \in \mathbb{R}^{L \times WH}$, $\mathbf{X} \in \mathbb{R}^{L \times wh}$, and $\mathbf{Y} \in \mathbb{R}^{l \times WH}$, corresponding to the tensors. Since $\mathbf{X}$ and $\mathbf{Y}$ correspond to the spatial and spectral downsampling of $\mathbf{Z}$ [23], we can establish the following relationship

$$\mathbf{X} = \mathbf{ZBS} \tag{1}$$

$$\mathbf{Y} = \mathbf{RZ} \tag{2}$$

where the matrices $\mathbf{B} \in \mathbb{R}^{WH \times WH}$ represents the convolution blur operation, and $\mathbf{S} \in \mathbb{R}^{WH \times wh}$ denotes the spatial downsampling operation, both determined by the point spread function (PSF) of the hyperspectral camera sensor. $\mathbf{R} \in \mathbb{R}^{l \times L}$ represents the spectral downsampling operation, determined by the multispectral camera's spectral response function (SRF). In this paper, both PSF and SRF are known.

There is typically a high correlation among the spectral bands in HSI, and the spectral vectors often lie in a low-dimensional subspace [34]. Therefore, in common fusion models, it is commonly assumed that

$$\mathbf{Z} = \mathbf{DC} \tag{3}$$

where $\mathbf{D} \in \mathbb{R}^{L \times n}$ represents the spectral basis, and $\mathbf{C} \in \mathbb{R}^{n \times WH}$ represents the corresponding coefficient matrix. The reconstruction problem of the high spatial resolution hyperspectral image is thus transformed into a problem of solving for $\mathbf{D}$ and $\mathbf{C}$. The essential purpose of $\mathbf{D}$ and $\mathbf{C}$ is to extract the spectral information from HSI and the spatial information from MSI and reasonably combine them to achieve fusion; however, obtaining a high-quality fusion result typically requires complex computations and iterative optimization for solving $\mathbf{D}$ and $\mathbf{C}$. Therefore, we propose an assumption that there exists a matrix $\mathbf{M}$ satisfying the following expression

$$\mathbf{Z} = \mathbf{XMY} \tag{4}$$

where $\mathbf{M} \in \mathbb{R}^{wh \times l}$.

According to Equation (4), we can perform fusion by matrix operations when $\mathbf{M}$ exists. The role of $\mathbf{M}$ is to associate the spectral information of HSI with the spatial information of MSI, which we call this matrix as the correlation matrix. Due to the use of matrix operations without the need for complex feature extraction or iterative optimization, this fusion process is efficient.

From Equation (1), we can obtain that

$$\mathbf{Z} = \mathbf{X}(\mathbf{BS})^{-1} \tag{5}$$

where the $(\mathbf{BS})^{-1}$ means the inverse matrix of $\mathbf{BS}$. It means that we can derive $\mathbf{Z}$ once we have obtained $(\mathbf{BS})^{-1}$; however, $\mathbf{BS}$ is typically a non-square matrix, meaning it does not have a corresponding inverse matrix $(\mathbf{BS})^{-1}$. To address this, we introduce a generalized inverse matrix to approximate $(\mathbf{BS})^{-1}$. In this case, we can obtain the following equation.

$$\mathbf{Z} = \mathbf{X}(\mathbf{BS})^{+} \tag{6}$$

where the $(\mathbf{BS})^{+}$ is the generalized inverse matrix [35] of $\mathbf{BS}$.

Limited by the large dimensionality of $\mathbf{BS}$, directly calculating the generalized inverse of $\mathbf{BS}$ can be computationally demanding and require substantial hardware resources. To address this issue, we can explore Equation (2) as an indirect alternative for solving the problem.

We apply the same spatial convolution blur and downsampling operations to $\mathbf{Y}$ and obtain the $\mathbf{Y_d} \in \mathbb{R}^{l \times wh}$.

$$\mathbf{Y_d} = \mathbf{YBS} \tag{7}$$

According to Equation (7), we can indirectly obtain the generalized inverse matrix of $\mathbf{BS}$.

$$(\mathbf{BS})^{+} = \mathbf{Y_d}^{+}\mathbf{Y} \tag{8}$$

Since $\mathbf{Y_d}$ is obtained by spatially blurring and downsampling $\mathbf{Y}$, its data dimension is significantly reduced, much smaller than the dimension of $\mathbf{BS}$. Therefore, computing the generalized inverse matrix of $\mathbf{Y_d}$ does not require high hardware requirements and can be easily implemented.

Combining Equations (6) and (8), we can obtain the following relationship:

$$\mathbf{Z} = \mathbf{X}\mathbf{Y_d}^{+}\mathbf{Y} \tag{9}$$

By comparing Equations (4) and (9), we can find that matrix $\mathbf{Y_d^+}$ satisfies the requirements of matrix $\mathbf{M}$. Therefore, once we can solve for $\mathbf{Y_d^+}$, we can easily achieve the fusion process through simple matrix operations. By substituting Equation (1) into Equation (9), we can mathematically demonstrate why this approach works.

At this point, by using Equation (9), we can obtain a high-quality HR-HSI. To further enhance the fusion performance, we introduce a variable $\mathbf{V} \in \mathbb{R}^{L \times WH}$ and construct the optimization function as follows

$$\min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{ZBS}\|_F^2 + \|\mathbf{Y} - \mathbf{RZ}\|_F^2 + \rho\|\mathbf{Z} - \mathbf{V}\|_F^2 \tag{10}$$

where $\|\mathbf{A}\|_F^2$ means the Frobenius norm of $\mathbf{A}$ in this paper and ρ is a positive penalty parameter less than 1.

By minimizing Equation (10) to make the gradient of $\mathbf{Z}$ equal to zero, we obtain the following equation

$$(\mathbf{R^T R} + \rho\mathbf{I_z})\mathbf{Z} + \mathbf{Z}(\mathbf{BS})(\mathbf{BS})^T = \mathbf{X}(\mathbf{BS})^T + \mathbf{R^T Y} + \rho\mathbf{V} \tag{11}$$

where $\mathbf{I_z} \in \mathbb{R}^{L \times L}$ is the identity matrix. The detailed derivation process from Equation (10) to Equation (11) is provided in Appendix A.

We solve Equation (11) using the Sylvester Equation [36] and transform the problem as follows

$$\mathbf{C_1 Z} + \mathbf{Z C_2} = \mathbf{C_3} \tag{12}$$

where

$$\begin{cases} \mathbf{C_1} = \mathbf{R^T R} + \rho\mathbf{I_z} \\ \mathbf{C_2} = (\mathbf{BS})(\mathbf{BS})^T \\ \mathbf{C_3} = \mathbf{X}(\mathbf{BS})^T + \mathbf{R^T Y} + \rho\mathbf{V} \end{cases} \tag{13}$$

For computational convenience, similar to work [23], we assume that $\mathbf{B}$ satisfies a circulant–block–circulant (CBC) structure, which means that $\mathbf{B}$ can be diagonalized as $\mathbf{B} = \mathbf{FKF}^H$, where $\mathbf{F}$ represents the DFT matrix, and $\mathbf{F}^H$ represents the conjugate transpose of $\mathbf{F}$. The diagonal matrix $\mathbf{K} \in \mathbb{C}^{WH \times WH}$ contains the singular values of $\mathbf{B}$.

The method proposed in this paper can be divided into two stages. The first stage is the fusion achieved using Equation (9), where no constrained optimization conditions are introduced. We denote this stage as CMF. The second stage involves the fusion with the introduction of constrained optimization conditions, represented by Equation (10). We denote this stage as CMF+. This is performed to facilitate comparing the two stages in the subsequent analysis. The detailed process of fusion is presented in Algorithm 1.

Algorithm 1 The proposed method.

Require: $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$
Obtain the $\mathbf{Y_d}$ from Equation (7)
Compute the generalized inverse matrix of $\mathbf{Y_d}$ to obtain $\mathbf{Y_d^+}$.
Reconstruct the first stage HR-HSI $\mathbf{Z_{CMF}}$ using Equation (9).
Solve Equation (10) by substituting $\mathbf{Z_{CMF}}$ to obtain $\mathbf{C_1}$, $\mathbf{C_2}$, and $\mathbf{C_3}$ in Equation (12).
Eigen-decomposition of $\mathbf{B}$:$\mathbf{B} = \mathbf{FKF}^H$
$\widetilde{\mathbf{K}} = \mathbf{K}(1_s \otimes 1_{wh})$
Eigen-decomposition of $\mathbf{C_1}$:$\mathbf{C_1} = \mathbf{Q\Lambda Q^{-1}}$
$\widetilde{\mathbf{C_3}} = \mathbf{Q^{-1}C_3F}$
for i = 1 to L **do**
  $\widetilde{\mathbf{z}}_l = \lambda_l^{-1}(\widetilde{\mathbf{C_3}})_l - \lambda_l^{-1}(\widetilde{\mathbf{C_3}})_l\widetilde{\mathbf{K}}(\lambda_l s\mathbf{I}_{wh} + \sum_{t=1}^{s}\mathbf{K}_t^2)\widetilde{\mathbf{K}}^H$
end for
Set $\mathbf{Z_{CMF+}} = \mathbf{Q}\widetilde{\mathbf{Z}}\mathbf{F^H}$
return $\mathbf{Z} = \mathbf{Z_{CMF+}}$

In Algorithm 1, the $1_s \in \mathbb{R}^s$ and $1_{wh} \in \mathbb{R}^{wh}$ are vector, and their elements are all ones. $\widetilde{\mathbf{z}}_l$ denote l-th row of $\widetilde{\mathbf{Z}} = \mathbf{Q}^{-1}\mathbf{Z_{CMF+}}\mathbf{F}$, that is, $\widetilde{\mathbf{Z}} = [\widetilde{\mathbf{z}_1}^T, \widetilde{\mathbf{z}_2}^T, \ldots, \widetilde{\mathbf{z}_L}^T]^T$. λ_l is the elements in diagonal matrix $\mathbf{\Lambda}$ and the s means the spatial downsampling factor.

4. Experiments

4.1. Dataset

To demonstrate the universality and effectiveness of our proposed method, we conducted experiments on two simulated datasets, the Harvard [37] and Pavia University [38] datasets. In addition, we test our proposed method on a real dataset of remote sensing images generated by Sentinel 2 and Hyperion spectral cameras.

The Harvard dataset comprises 50 hyperspectral images, including indoor and outdoor scenes. Each image has a spatial resolution of 1392 × 1040 and 31 spectral bands, with a wavelength range of 400 to 700 nm and a band spacing of 10nm. We have selected one indoor image (imgh7) and one outdoor image (imgb2) and extracted a region with prominent features as the reference image. As a result, we obtained a high spatial resolution hyperspectral image of size 512 × 512 × 31 as a reference. The Pavia University dataset is an image with 610 × 340 pixels and 103 spectral bands, which shows the urban area of Pavia University, northern Italy. To better showcase the experimental results, we have removed the fuzzy bands and cropped a specific region from the reference image. As a result, we obtained a reference image of size 256 × 256 × 93 as HR-HSI.

To validate our method's effectiveness, we processed these two simulated datasets to obtain the corresponding HSI and MSI. For the Harvard dataset, we applied a Gaussian blur kernel of size 7 × 7 with a standard deviation of 2 to generate HSI with spatial downsampling factors of 16 and 32. As a result, we obtained HSI with dimensions of 32 × 32 × 31 and 16 × 16 × 31, respectively. Similarly, for the Pavia University dataset, we applied the same processing method and obtained HSI data of size 16 × 16 × 93 and 8 × 8 × 93, respectively. By performing such processing, we can evaluate the effectiveness and performance of our method under different downsampling factors. To obtain the MSIs for the Harvard dataset, we used the spectral response functions of the Nikon D700 camera [27] and obtained MSIs of size 512 × 512 × 3. As for the Pavia University dataset, we used the IKONOS-like reflectance spectral response filter [39] to generate MSI of size 256 × 256 × 4.

In addition, we used a real dataset acquired by the Hyperion sensor on the Earth Observation 1 satellite to test our proposed method. This dataset contains a hyperspectral image with 220 spectral bands. To simplify the experiment, we selected a spatial region and a subset of spectral bands, resulting in hyperspectral image data of size 100 × 100 × 89. The corresponding MSI was acquired by the Sentinel-2A satellite with a size of 300 × 300 × 4.

The real dataset will be used to evaluate the performance and effectiveness of our proposed method in real scenarios. Table 1 shows the settings of SRF and PSF on each dataset.

Table 1. The setting of SRF and PSF.

Dataset	SRF	PSF
Harvard	Nikon D700 [27]	7×7 Gaussian blur (standard deviation 2)
Pavia University	IKONOS [39]	7×7 Gaussian blur (standard deviation 2)
Sentinel 2 and Hyperion	PSF and SRF are Obtained by the Hysure [40] method	

4.2. Experimental Settings and Quantitative Metrics

The proposed method is compared with eight state-of-the-art HSI and MSI fusion methods, including low tensor multi-rank regularization (LTMR) [23], fast low tensor multi-rank regularization (FLTMR) [41], coupled sparse tensor factorization (CSTF) [29], non-negative structured sparse representation (NSSR) [27], clustering manifold structure (CMS) [42], image fusion by CNN denoiser (CNN-Fus) [32], model inspired autoencoder (MIAE) [33], and cross-attention in coupled unmixing nets (CUCaNet) [31]. Among the compared methods, CNN-Fus, MIAE, and CUCaNet are deep learning-based, and the others are traditional. In order to ensure the fairness of the experiments, we followed the parameter settings provided in the original literature for most methods. We did not perform targeted parameter adjustments for each method because it is often challenging to provide an effective reference image for parameter tuning in real-world applications. Therefore, in practical usage, we typically rely on the parameters provided by the original methods. However, in cases where certain methods were unable to run on the dataset, we made appropriate adjustments to the parameters. When the downsampling factors were 16 and 32 for the Pavia University dataset, we adjusted the parameter K in the FLTMR method to 25 and 4, respectively. Similarly, with downsampling factors of 16 and 32 for the Harvard dataset, we adjusted the parameter K in the FLTMR method to 100 and 25, respectively. We set the spectral basis L to 3 for all datasets for the CNN-Fus method. Additionally, for the NSSR method on the Pavia University dataset, when the downsampling factor of 32, we adjusted the parameter K to 64.

We use four objective evaluation metrics to measure the quality of the fusion results, including peak signal-to-noise ratio (PSNR) [23], the erreur relative globale adimensionnelle de synthese (ERGAS) [43], spectral angle mapper (SAM) [44], and universal image quality index (UIQI) [45]. PSNR calculates the mean square error between the original and processed images and compares it to the square of the maximum possible pixel value. PSNR is typically expressed in decibels (dB). ERGAS is a metric that measures the relative error between the reconstructed image and the original image, taking into account the spatial resolution of the image. SAM is a metric used to compare the similarity of multispectral images. It measures the difference between two-pixel vectors in the spectral space by calculating their angle. UIQI is a comprehensive metric used to assess image quality. It combines information from aspects such as luminance, contrast, and structure and calculates the similarity between two images.

5. Results and Analysis

5.1. Results on Simulated Dataset

Among the compared methods, MIAE and CUCaNet can perform image fusion without PSF and SRF information. In contrast, our proposed and other comparative methods perform fusion under the known PSF and SRF assumptions. The CUCaNet method can also be adapted to utilize PSF and SRF by modifying the source code. However, the MIAE method requires more extensive modifications to the source code to accommodate PSF and SRF. To minimize the impact of human factors on the experimental results, we retained the fusion results of the MIAE method without using PSF and SRF on the two

simulated datasets. For the CUCaNet method, we made corresponding adjustments to enable fusion using PSF and SRF. The CMF and CMF+ show the different fusion results of the two stages in our proposed method.

The Harvard dataset has two categories of images, indoor and outdoor scenes. We selected representative images from each category for our experiments, namely "imgb2" for the outdoor image and "imgh7" for the indoor image. The corresponding fusion results are listed in Tables 2 and 3, respectively. In the tables, the letter "s" denotes the downsampling factor. For each evaluation metric, we have highlighted the optimal results using **bold** font and indicated the second-best performance with an underline, enabling easy comparison and identification of the fusion result.

Table 2. Quantitative results on the Harvard dataset (imgb2).

Method	$s = 16$				$s = 32$			
	PSNR	**SAM**	**ERGAS**	**UIQI**	**PSNR**	**SAM**	**ERGAS**	**UIQI**
Best Values	$+\infty$	0	0	1	$+\infty$	0	0	1
CMF	42.503	1.940	1.150	0.972	42.422	1.954	0.578	**0.972**
CMF+	42.648	1.888	1.148	0.972	**42.454**	**1.942**	0.578	**0.972**
LTMR	43.519	1.827	1.054	0.972	39.957	3.247	0.648	0.957
FLTMR	**43.782**	1.775	1.042	**0.973**	38.792	4.003	0.705	0.946
CSTF	42.865	2.175	**0.960**	0.961	42.285	2.218	**0.556**	0.966
NSSR	28.997	3.891	1.841	0.905	28.431	3.428	0.896	0.926
CMS	43.350	**1.589**	1.032	0.970	42.015	2.223	0.624	0.965
CNN-Fus	42.036	1.980	1.168	0.970	41.948	2.005	0.588	0.969
MIAE	42.127	1.844	2.632	0.971	40.566	2.082	1.230	0.970
CUCaNet	41.447	1.725	1.500	0.972	39.929	2.033	1.530	0.970

Table 3. Quantitative results on the Harvard dataset (imgh7).

Method	$s = 16$				$s = 32$			
	PSNR	**SAM**	**ERGAS**	**UIQI**	**PSNR**	**SAM**	**ERGAS**	**UIQI**
Best Values	$+\infty$	0	0	1	$+\infty$	0	0	1
CMF	53.026	1.751	0.983	0.445	53.021	1.752	0.492	0.445
CMF+	**53.047**	**1.744**	0.982	**0.447**	**53.027**	**1.750**	0.492	0.445
LTMR	47.810	2.325	1.365	0.418	47.285	2.576	0.753	0.412
FLTMR	47.369	2.392	1.416	0.414	46.215	3.116	0.849	0.390
CSTF	52.799	1.775	0.989	0.414	52.672	1.807	0.502	0.419
NSSR	46.085	3.251	1.361	0.344	52.610	1.835	0.508	**0.447**
CMS	50.395	1.747	1.151	0.440	30.576	26.807	5.145	0.102
CNN-Fus	52.585	1.801	1.003	0.407	52.578	1.804	0.501	0.406
MIAE	52.963	1.763	**0.493**	0.443	52.932	1.771	**0.490**	0.445
CUCaNet	51.066	1.807	1.037	0.442	49.063	1.985	0.589	0.430

Table 2 shows that when the downsampling factor is 16, the FLTMR method achieves the best performance in terms of PSNR and UIQI metrics, outperforming our proposed method; however, it is worth noting that the fusion results obtained by our CMF+ method are very close to FLTMR. The PSNR difference is only 1.134 dB, and other metrics also exhibit a high level of similarity. When the downsampling factor increases to 32, our method achieves the best or second-best results in most metrics, with only a slightly lower ERGAS metric performance than the CSTF method. Notably, the impact of changing the downsampling factor on our method is relatively small compared to the downsampling factor of 16. Specifically, the CMF experiences a decrease of only 0.081 dB in PSNR, while CMF+ shows a decrease of only 0.194 dB. In contrast, the LTMR method experiences a significant drop of 4.99 dB in this metric.

When the downsampling factor increases, the spatial information in HSI decreases, reducing available prior information for fusion. Therefore, as the downsampling factor increases, the fusion accuracy tends to decrease. The proposed method does not require

additional feature extraction but operates based on the correlation matrix. When the downsampling factor changes, the correlation matrix is adjusted accordingly to adapt to the new HSI image, resulting in minimal changes. In contrast, the LTMR method assumes that the HR-HSI consists of a spectral base and corresponding coefficient matrix. The spectral base is obtained through singular value decomposition of the HSI, so changes in the downsampling factor directly affect the construction of the spectral base. Additionally, there are parameters in the fusion method that need to be adjusted to achieve optimal results; however, finding suitable references for parameter tuning in practical scenarios is challenging. As a result, the results of the LTMR method can vary significantly for different datasets and downsampling factors. FLTMR is an improvement upon LTMR, so it faces similar issues.

To provide a more intuitive display of the fusion results, we will show the false color images of the fusion results at a downsampling factor of 32. Figure 2 illustrates the fusion results of various methods mentioned in Table 2. The yellow box represents the image enlarged from the region marked by the red box, while the blue box represents the enlarged error image corresponding to that region. Figure 2 shows that the fusion results of CMF, CMF+, and CNN-Fus methods are superior to the other methods.

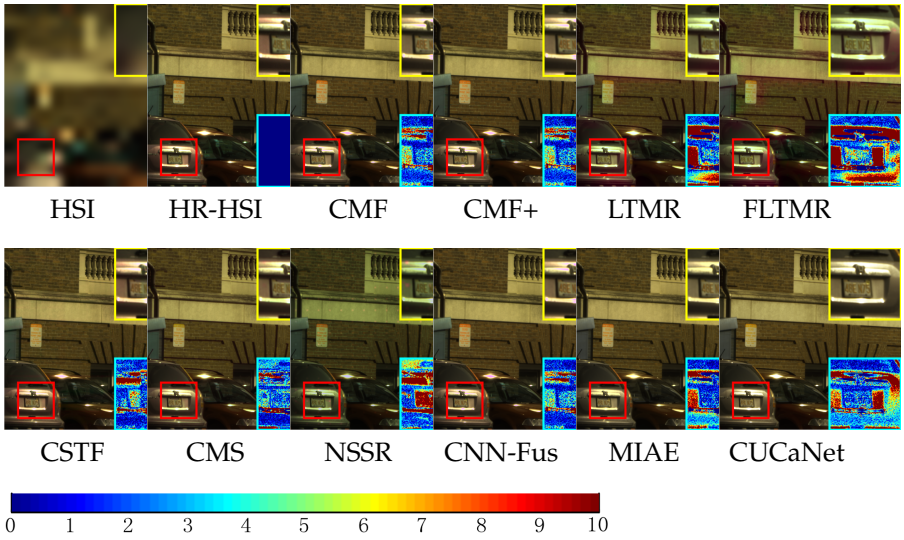


Figure 2. The false color images formed by bands 30, 20, and 10 of the fusion results (imgb2, $s = 32$). The yellow box region displays the image of the red box region magnified by two times. The blue box is the error image in the magnified portion at band 20. HR-HSI ($512 \times 512 \times 31$) is obtained by a commercial hyperspectral camera (Nuance FX, CRI Inc.).

According to the data in Table 3, our method achieves optimal results in most indicators for the indoor image of the Harvard dataset at downsampling factors of 16 and 32. Although it may slightly lag behind other methods in individual indicators, our method still performs at a second-best level. Overall, our method outperforms other methods; however, it is worth noting that the CMS method in the table shows significant changes in fusion performance at a downsampling factor of 32. This is mainly because the parameters of the CMS method were not adjusted according to the downsampling factor. Moreover, in the corresponding fusion result Figure 3, we can observe that CMF, CMF+, CNN-Fus, and MIAE methods exhibit significantly better fusion results than the others. These methods preserve image details and color accuracy, as evident from the enlarged regions and corresponding error images. The error images demonstrate their ability to reduce artifacts and distortions effectively.

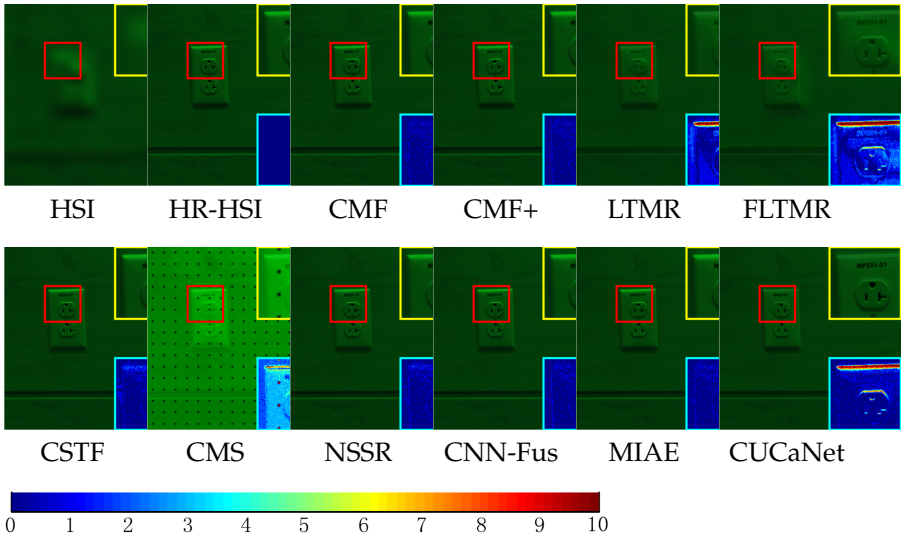


Figure 3. The false color images formed by bands 11, 21, and 31 of the fusion results (imgh7, $s = 32$). The yellow box region displays the image of the red box region magnified by two times. The blue box is the error image in the magnified portion at band 11. HR-HSI ($512 \times 512 \times 31$) is obtained by a commercial hyperspectral camera (Nuance FX, CRI Inc.).

According to the experimental results on the Pavia University dataset, Table 4 presents the performance of various methods. We observed that, at a downsampling factor of 16, the CUCaNet method achieved the best experimental results, while our method obtained the second-best results. However, when the downsampling factor increased to 32, our method achieved the best results in terms of PSNR and ERGAS while obtaining the second-best results in terms of SAM and UIQI evaluation metrics. CUCaNet, MIAE, and CNN-Fus are deep learning-based methods, where the CNN-Fus method does not require extensive training and has relatively low requirements on the training set. However, MIAE and CUCaNet methods require repeated training, and their performance can vary significantly across different datasets. The parameter settings in these three methods also impose limitations on fusion performance. For the CNN-Fus method, the fusion results may not be satisfactory if we do not modify the spectral basic L. Indeed, from the fusion results in Figure 4, it is evident that CMF, CMF+, and CSTF methods exhibit significant advantages in image fusion. They can better preserve image details and color consistency, producing more precise and natural fused images. In contrast, the fusion results of the NSSR method are relatively poorer, primarily because its parameters perform well only on specific datasets.

Table 4. Quantitative results on the Pavia University dataset.

Method	$s = 16$				$s = 32$			
	PSNR	**SAM**	**ERGAS**	**UIQI**	**PSNR**	**SAM**	**ERGAS**	**UIQI**
Best Values	$+\infty$	0	0	1	$+\infty$	0	0	1
CMF	43.258	1.997	0.281	0.989	43.219	1.998	0.141	0.989
CMF+	43.416	2.060	0.275	0.989	**43.253**	2.042	**0.140**	0.989
LTMR	33.655	6.802	0.958	0.935	31.292	9.074	0.656	0.908
FLTMR	32.108	8.047	1.161	0.908	30.340	10.145	0.733	0.892
CSTF	42.613	2.121	0.303	0.987	41.566	2.414	0.179	0.986
NSSR	26.883	5.591	1.705	0.901	26.554	4.156	0.859	0.923
CMS	36.507	5.168	0.746	0.946	34.313	6.555	0.492	0.930
CNN-Fus	41.517	2.477	0.342	0.987	41.447	2.476	0.173	0.987
MIAE	34.605	4.039	0.742	0.971	39.919	2.459	0.199	0.986
CUCaNet	**43.582**	**1.534**	**0.256**	**0.993**	42.064	**1.956**	0.153	**0.991**

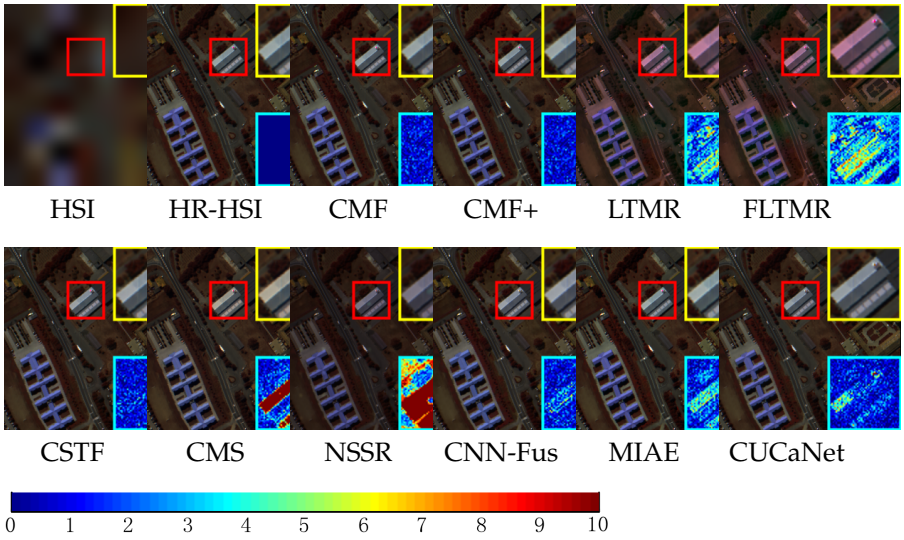


Figure 4. The false color images formed by bands 61, 36, 10 of the fusion results ($s = 32$). The yellow box region displays the image of the red box region magnified by two times. The blue box is the error image in the magnified portion at band 11. HR-HSI ($256 \times 256 \times 93$) was obtained by ROSIS (Reflective Optics Spectrographic Imaging System) on 2003.

5.2. Results on Real Dataset

This section will test our proposed method on a real dataset. To better simulate real-world usage scenarios, the parameter settings for each method were kept the same setting as the experiment on the Pavia University dataset ($s = 16$). The main reason for doing this is that it is often impossible to adjust the parameters based on reliable image pairs for optimal performance in real scenarios. As for our proposed and other methods, we employed the method in [40] to generate PSF and SRF during the fusion process. Figure 5 shows the fusion results of each method on the real dataset, where the yellow box highlights the effect of the region magnified three times compared to the red box region.

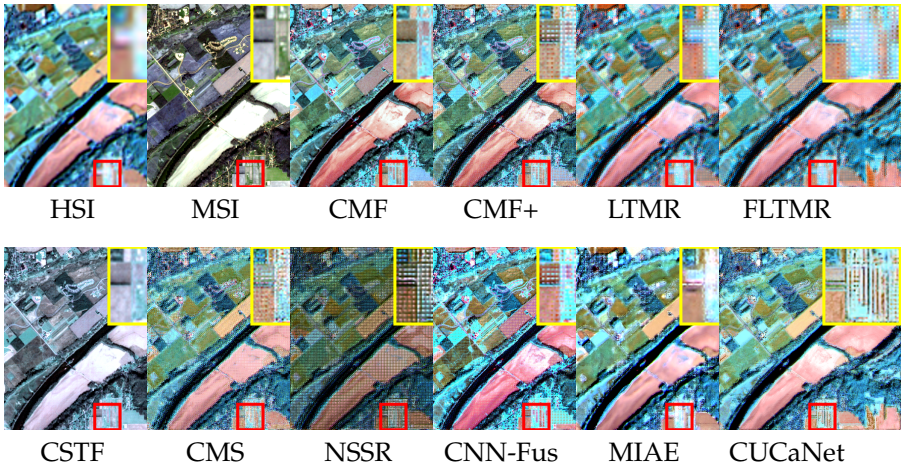


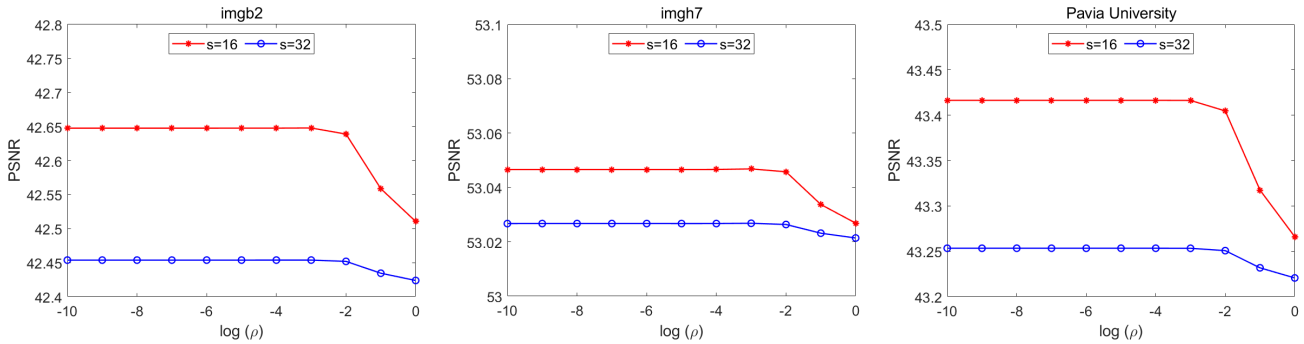
Figure 5. The false color images formed by bands 80, 60, and 29 of the fusion results on the real dataset. The yellow box region displays the image of the red box region magnified by three times. HSI ($100 \times 100 \times 89$) is obtained by Hyperion sensor on the Earth Observation 1 satellite, MSI ($300 \times 300 \times 4$) is obtained by the Sentinel-2A satellite.

Based on the observations from the false color images, we found that the CMF, CMF+, MIAE, and CUCaNet methods are closer to the false color images of the HSI. Furthermore, upon closer examination of the magnified regions, the CMF method exhibits better detail preservation than the MSI. The CUCaNet and CMF+ methods also achieve relatively good

results in preserving details. On the other hand, the MIAE method shows significant blurring. The fusion result of the CMF+ method is worse than CMF because CMF+ utilizes estimated PSF and SRF during the optimization process, which introduces additional errors. In contrast, the CMF method only employs PSF when generating $\mathbf{Y_d}$ and does not undergo further optimization, thus avoiding the introduction of additional errors.

5.3. Parameter Discussion

The proposed method in this paper involves only one parameter, ρ, in Equation (10). This section will discuss the impact of ρ on the fusion results. Figure 6 displays the variation of PSNR with different values of $\log(\rho)$ (log is base 10) on the Harvard and Pavia University datasets. According to the variation of PSNR in Figure 6, it can be observed that when the value of $\log(\rho)$ is below -3, it has a minimal impact on the experimental results. Therefore, we set ρ as a fixed parameter in our proposed method, specifically 0.001. Tables 2–4 show that the CMF stage alone can achieve decent fusion results even without introducing additional optimization processes. The optimization process further improves the fusion results based on the CMF stage. Although the variable ρ may influence the fusion results, it only improves the results obtained in the CMF stage when PSF and SRF are known. The parameter ρ constrains and optimizes the results without requiring iterative processes. Thus, its impact is limited. In contrast, parameters in other fusion methods are often related to selecting critical factors in the fusion process, directly affecting the fusion results. Moreover, these methods require iterative optimization, which may amplify the errors caused by parameters. Consequently, even in the case where both PSF and SRF are known, these methods often necessitate parameter adjustments for different datasets to achieve satisfactory fusion results, while our method does not have this concern. In the scenario where PSF and SRF are unknown, we can just use CMF for fusion, reducing the errors associated with the estimated PSF and SRF and eliminating the need for any parameter tuning.



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

Figure 6. The PSNR varies from parameter ρ. (**a**,**b**) are for Harvard dataset. (**c**) is for Pavia University dataset.

5.4. Comparison between CMF and CMF+

When discussing the differences between CMF and CMF+, we must consider two scenarios: PSF and SRF are known; PSF and SRF are unknown. In Section 3, we mentioned that CMF+ introduces the Sylvester equation to further optimize the results based on CMF. The downsampling factor and whether PSF and SRF are known largely influence the effect of the optimization.

When PSF and SRF are known, HSI contains more spatial information when the downsampling factor is low. In this case, the optimization equation can better adjust the fusion results, further improving the result of CMF. However, when the downsampling factor is high, the spatial information in HSI is significantly reduced, limiting the effect of the optimization equation. This observation is also evident in Figure 7, where the effectiveness of the optimization equation decreases with increasing downsampling factor.

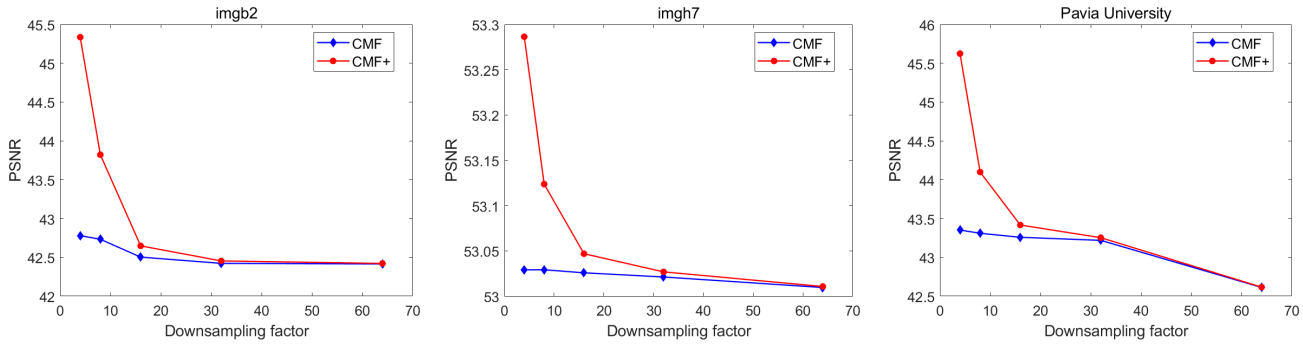


Figure 7. The PSNR varies from different downsampling factors (4, 8, 16, 32, 64). (**a**,**b**) are for Harvard dataset. (**c**) is for Pavia University dataset.

On the other hand, when PSF and SRF are unknown, as discussed in Section 5.2, the iterative optimization process introduces and amplifies errors in PSF and SRF, resulting in inferior performance of CMF+ compared to CMF. Therefore, in this scenario, CMF outperforms CMF+ in terms of overall performance.

In conclusion, CMF and CMF+ all achieve good fusion results when PSF and SRF are known, where CMF+ performs better at lower downsampling factors. When the PSF and RSF are unknown, the fusion results of CMF combined with related estimation methods will have better fusion results than CMF+. We can make corresponding choices according to the actual requirements.

5.5. Computational Cost

To showcase the computational efficiency of our method and other comparative methods, Table 5 displays the fusion time of each method on different datasets, and the average fusion time for all methods is calculated. Since MIAE and CUCaNet methods require an enormous amount of operating memory, we deployed them on a server configured with Intel(R) Xeon(R) Gold 5218 CPU @ 2.30 GHz and 125 GB RAM, along with NVIDIA TITAN RTX. The remaining methods were executed using Matlab 2021b on a local host with a 2.90 GHz 6-core Intel i5-9400F CPU and 8.0 GB DDR3 RAM.

Table 5. Fusion time (second).

Method	$s = 16$			$s = 32$				
	imgb2	imgh7	Pavia	imgb2	imgh7	Pavia	Real	Average
CMF	**0.053**	**0.053**	**0.517**	**0.067**	**0.059**	**0.048**	**0.082**	**0.126**
CMF+	1.227	1.228	0.903	1.154	1.174	0.862	1.808	1.194
LTMR	207.873	221.029	53.998	204.521	231.528	55.572	117.764	156.041
FLTMR	52.842	58.508	12.966	44.57	51.512	16.891	26.266	37.651
CSTF	130.619	129.765	102.492	105.265	135.655	125.971	233.009	137.539
NSSR	125.331	124.088	51.889	98.379	128.97	59.89	166.427	107.853
CMS	679.978	758.924	253.499	436.439	559.609	560.76	1100.2	621.344
CNN-Fus	167.128	171.829	63.745	132.762	167.446	73.035	26.529	114.639
MIAE	406.732	383.646	174.546	1678.538	1565.11	976.243	295.694	782.93
CUCaNet	58880	6469	5139	5767	5826	5097	5027	5600.714

Based on the data in Table 5, we can observe that both CMF and CMF+ methods require a very short time to complete the fusion, showcasing excellent fusion time regardless of the dataset. In comparison to other methods, their fusion time is significantly shorter. It is worth noting that the time for CNN-Fus, MIAE, and CUCaNet in Table 5 includes the training time, encompassing the entire process from training to fusion completion. Deep learning-based models often require retraining in practical scenarios to achieve satisfactory fusion results on new datasets.

Relative to the average fusion time of CMF, the average fusion time of CUCaNet is approximately 44,450 times longer. Even though the improved version of FLTMR shows significant improvement in average fusion time compared to LTMR, it is still around 298 times slower than CMF and nearly 32 times slower than CMF+. Additionally, we can infer from the objective metrics results in Tables 2–4 that these comparative methods take longer than CMF and CMF+, but their overall fusion results do not show equivalent improvements. In fact, the fusion results of most methods are inferior to our proposed methods.

Therefore, considering both fusion quality and efficiency, the method proposed in this paper is significantly superior to other comparative methods.

6. Conclusions

This paper proposes the assumption of the correlation matrix for the first time, and the matrix satisfying the assumption is derived. With this correlation matrix, we can realize the fusion process through simple matrix operations and quickly obtain high-quality HR-HSI. At the same time, we use the Sylvester equation to optimize the results further and improve the fusion quality. Experimental results on both simulated and real datasets demonstrate the superiority of our proposed method over other state-of-the-art fusion methods. The proposed method exhibits low fusion time cost, and the parameter settings do not significantly impact the fusion results, providing good generalization ability. The fusion approach through the correlation matrix provides a new direction for future research, in which we can find a more appropriate correlation matrix to achieve better fusion results. The correlation matrix-based method effectively reduces data complexity, better preserves the characteristics of hyperspectral and multispectral images, and resolves issues in common fusion methods, such as low fusion efficiency during the fusion process and the impact of non-fixed parameters. The proposed method enables the rapid reconstruction of HR-HSI, facilitating its integration with other applications faster. However, the proposed method still has some limitations, mainly reflected in the dependence on PSF and SRF. Specifically, when PSF and SRF are unknown, the estimation of SRF and PSF using other methods introduces additional errors, which can impact the optimization process of CMF+. This, in turn, leads to a decrease in fusion performance; therefore, appropriate prior information should be introduced in future research to reduce the dependence on PSF and SRF.

Author Contributions: Conceptualization, H.L. and J.L. (Jun Li); data curation, H.L.; formal analysis, H.L. and J.L. (Jun Li); funding acquisition, T.Z.; investigation, H.L. and J.L. (Jian Long); methodology, H.L.; resources, Y.P.; software, H.L.; validation, H.L. and J.L. (Jun Li); writing—original draft preparation, H.L.; writing—review and editing, H.L., J.L. (Jun Li), Y.P., T.Z., J.L. (Jian Long) and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Opening Foundation of State Key Laboratory of High Performance Computing, National University of Defense Technology, under Grant No 202201-05.

Data Availability Statement: The data presented in this study are available in the article.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Here, we will show the derivation process of Equations (10) to (11). Based on Equation (10), we can obtain the following function:

$$\mathbf{Fun}(\mathbf{Z}) = \|\mathbf{X} - \mathbf{ZBS}\|_F^2 + \|\mathbf{Y} - \mathbf{RZ}\|_F^2 + \rho\|\mathbf{Z} - \mathbf{V}\|_F^2 \tag{A1}$$

To convert a minimization problem into a derivative problem of $\mathbf{Fun}(\mathbf{Z})$. We can utilize the rules of Frobenius norm and trace operations. Based on these rules, we can establish the following relationship

$$\begin{aligned} \|\mathbf{X}-\mathbf{ZBS}\|_F^2 &= tr(\mathbf{X}-\mathbf{ZBS})^T(\mathbf{X}-\mathbf{ZBS}) \\ &= tr[\mathbf{X}^T\mathbf{X}-\mathbf{X}^T\mathbf{Z}(\mathbf{BS})-(\mathbf{ZBS})^T\mathbf{X}+(\mathbf{ZBS})^T(\mathbf{ZBS})] \\ &= tr[\mathbf{X}^T\mathbf{X}-2(\mathbf{BS})\mathbf{X}^T\mathbf{Z}+\mathbf{Z}(\mathbf{BS})(\mathbf{BS})^T\mathbf{Z}^T] \end{aligned} \tag{A2}$$

where tr represents the trace of a matrix. Taking the derivative of Equation (A2) with respect to $\mathbf{Z}$, we obtain:

$$\begin{aligned} \frac{\partial\|\mathbf{X}-\mathbf{ZBS}\|_F^2}{\partial\mathbf{Z}} &= \frac{\partial tr(\mathbf{X}-\mathbf{ZBS})^T(\mathbf{X}-\mathbf{ZBS})}{\partial\mathbf{Z}} \\ &= -2\mathbf{X}(\mathbf{BS})^T+2\mathbf{Z}(\mathbf{BS})(\mathbf{BS})^T \end{aligned} \tag{A3}$$

Similarly, we can obtain the results for Equations (A4) and (A5) using the same way.

$$\frac{\partial\|\mathbf{Y}-\mathbf{RZ}\|_F^2}{\partial\mathbf{Z}} = -2\mathbf{R}^T\mathbf{Y}+2\mathbf{R}^T\mathbf{RZ} \tag{A4}$$

$$\frac{\partial\rho\|\mathbf{Z}-\mathbf{V}\|_F^2}{\partial\mathbf{Z}} = 2\rho(\mathbf{Z}-\mathbf{V}) \tag{A5}$$

Therefore, we can obtain the derivative of $\mathbf{Fun}(\mathbf{Z})$ with respect to $\mathbf{Z}$.

$$\frac{\partial\mathbf{Fun}(\mathbf{Z})}{\partial\mathbf{Z}} = -2\mathbf{X}(\mathbf{BS})^T+2\mathbf{Z}(\mathbf{BS})(\mathbf{BS})^T-2\mathbf{R}^T\mathbf{Y}+2\mathbf{R}^T\mathbf{RZ}+2\rho(\mathbf{Z}-\mathbf{V}) \tag{A6}$$

By setting the derivative of Equation (A6) to zero:

$$-\mathbf{X}(\mathbf{BS})^T+\mathbf{Z}(\mathbf{BS})(\mathbf{BS})^T-\mathbf{R}^T\mathbf{Y}+\mathbf{R}^T\mathbf{RZ}+\rho(\mathbf{Z}-\mathbf{V})=0 \tag{A7}$$

Upon rearranging Equation (A7), we obtain

$$(\mathbf{R^TR}+\rho\mathbf{I_z})\mathbf{Z}+\mathbf{Z}(\mathbf{BS})(\mathbf{BS})^T=\mathbf{X}(\mathbf{BS})^T+\mathbf{R^TY}+\rho\mathbf{V} \tag{A8}$$

where $\mathbf{I_z}\in\mathbb{R}^{L\times L}$ is the identity matrix. Equation (A8) is indeed equivalent to Equation (11), completing the derivation.

References

1. Sethy, P.K.; Pandey, C.; Sahu, Y.K.; Behera, S.K. Hyperspectral imagery applications for precision agriculture-a systemic survey. *Multimed. Tools Appl.* **2022**, *81*, 3005–3038. [CrossRef]
2. Agilandeeswari, L.; Prabukumar, M.; Radhesyam, V.; Phaneendra, K.L.B.; Farhan, A. Crop classification for agricultural applications in hyperspectral remote sensing images. *Appl. Sci.* **2022**, *12*, 1670. [CrossRef]
3. Wang, B.; Sun, J.; Xia, L.; Liu, J.; Wang, Z.; Li, P.; Guo, Y.; Sun, X. The applications of hyperspectral imaging technology for agricultural products quality analysis: A review. *Food Rev. Int.* **2021**, *39*, 1043–1062. [CrossRef]
4. Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens.* **2020**, *12*, 2659. [CrossRef]
5. Ma, C.; Liu, X.; Li, S.; Li, C.; Zhang, R. Accuracy evaluation of hyperspectral inversion of environmental parameters of loess profile. *Environ. Earth Sci.* **2023**, *82*, 251. [CrossRef]
6. Stuart, M.B.; Davies, M.; Hobbs, M.J.; Pering, T.D.; McGonigle, A.J.; Willmott, J.R. High-resolution hyperspectral imaging using low-cost components: Application within environmental monitoring scenarios. *Sensors* **2022**, *22*, 4652. [CrossRef]
7. Cui, Y.; Zhang, B.; Yang, W.; Wang, Z.; Li, Y.; Yi, X.; Tang, Y. End-to-end visual target tracking in multi-robot systems based on deep convolutional neural network. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 1113–1121.

8. Cui, Y.; Zhang, B.; Yang, W.; Yi, X.; Tang, Y. Deep CNN-based visual target tracking system relying on monocular image sensing. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
9. Hirai, A.; Tonooka, H. Mineral discrimination by combination of multispectral image and surrounding hyperspectral image. *J. Appl. Remote Sens.* **2019**, *13*, 024517. [CrossRef]
10. He, J.; Barton, I. Hyperspectral remote sensing for detecting geotechnical problems at Ray mine. *Eng. Geol.* **2021**, *292*, 106261. [CrossRef]
11. Vignesh, K.M.; Kiran, Y. Comparative analysis of mineral mapping for hyperspectral and multispectral imagery. *Arab. J. Geosci.* **2020**, *13*, 160. [CrossRef]
12. Yuan, J.; Wang, S.; Wu, C.; Xu, Y. Fine-grained classification of urban functional zones and landscape pattern analysis using hyperspectral satellite imagery: A case study of wuhan. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3972–3991. [CrossRef]
13. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sens. Environ.* **2018**, *216*, 57–70. [CrossRef]
14. Weng, Q. A remote sensing? GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China. *Int. J. Remote Sens.* **2001**, *22*, 1999–2014.
15. Sun, W.; Du, Q. Graph-regularized fast and robust principal component analysis for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3185–3195. [CrossRef]
16. Akgun, T.; Altunbasak, Y.; Mersereau, R.M. Super-resolution reconstruction of hyperspectral images. *IEEE Trans. Image Process.* **2005**, *14*, 1860–1875. [CrossRef] [PubMed]
17. Hong, D.; Liu, W.; Su, J.; Pan, Z.; Wang, G. A novel hierarchical approach for multispectral palmprint recognition. *Neurocomputing* **2015**, *151*, 511–521. [CrossRef]
18. 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]
19. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. CoSpace: Common subspace learning from hyperspectral-multispectral correspondences. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4349–4359. [CrossRef]
20. Ferraris, V.; Dobigeon, N.; Wei, Q.; Chabert, M. Robust fusion of multiband images with different spatial and spectral resolutions for change detection. *IEEE Trans. Comput. Imaging* **2017**, *3*, 175–186. [CrossRef]
21. Qu, Y.; Qi, H.; Ayhan, B.; Kwan, C.; Kidd, R. DOES multispectral/hyperspectral pansharpening improve the performance of anomaly detection? In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 6130–6133.
22. Gómez-Chova, L.; Tuia, D.; Moser, G.; Camps-Valls, G. Multimodal classification of remote sensing images: A review and future directions. *Proc. IEEE* **2015**, *103*, 1560–1584. [CrossRef]
23. Dian, R.; Li, S. Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. *IEEE Trans. Image Process.* **2019**, *28*, 5135–5146. [CrossRef]
24. Wei, Q.; Dobigeon, N.; Tourneret, J.Y. Bayesian fusion of hyperspectral and multispectral images. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 3176–3180.
25. 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]
26. Li, X.; Zhang, Y.; Ge, Z.; Cao, G.; Shi, H.; Fu, P. Adaptive nonnegative sparse representation for hyperspectral image super-resolution. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4267–4283. [CrossRef]
27. Dong, W.; Fu, F.; Shi, G.; Cao, X.; Wu, J.; Li, G.; Li, X. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Trans. Image Process.* **2016**, *25*, 2337–2352. [CrossRef]
28. Huang, B.; Song, H.; Cui, H.; Peng, J.; Xu, Z. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 1693–1704. [CrossRef]
29. 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]
30. Dian, R.; Li, S.; Fang, L. Learning a low tensor-train rank representation for hyperspectral image super-resolution. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2672–2683. [CrossRef] [PubMed]
31. Yao, J.; Hong, D.; Chanussot, J.; Meng, D.; Zhu, X.; Xu, Z. Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; pp. 208–224.
32. Dian, R.; Li, S.; Kang, X. Regularizing hyperspectral and multispectral image fusion by CNN denoiser. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 1124–1135. [CrossRef] [PubMed]
33. Liu, J.; Wu, Z.; Xiao, L.; Wu, X.J. Model inspired autoencoder for unsupervised hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
34. Zhuang, L.; Bioucas-Dias, J.M. Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 730–742. [CrossRef]
35. Penrose, R. A generalized inverse for matrices. *Math. Proc. Camb. Philos. Soc.* **1955**, *51*, 406–413. [CrossRef]

36. Bartels, R.H.; Stewart, G.W. Solution of the matrix equation AX+ XB= C [F4]. *Commun. ACM* **1972**, *15*, 820–826. [CrossRef]
37. Chakrabarti, A.; Zickler, T. Statistics of real-world hyperspectral images. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 193–200.
38. Dell'Acqua, F.; Gamba, P.; Ferrari, A.; Palmason, J.A.; Benediktsson, J.A.; Árnason, K. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 322–326. [CrossRef]
39. 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]
40. 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]
41. Long, J.; Peng, Y.; Li, J.; Zhang, L.; Xu, Y. Hyperspectral image super-resolution via subspace-based fast low tensor multi-rank regularization. *Infrared Phys. Technol.* **2021**, *116*, 103631. [CrossRef]
42. Zhang, L.; Wei, W.; Bai, C.; Gao, Y.; Zhang, Y. Exploiting clustering manifold structure for hyperspectral imagery super-resolution. *IEEE Trans. Image Process.* **2018**, *27*, 5969–5982. [CrossRef]
43. Wald, L. Quality of high resolution synthesised images: Is there a simple criterion? In Proceedings of the Third International Conference Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images, Sophia Antipolis, France, 26–28 January 2000; pp. 99–103.
44. Yuhas, R.H.; Goetz, A.F.; Boardman, J.W. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In Proceedings of the JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop, Pasadena, CA, USA, 1–5 June 1992.
45. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [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.