A Novel Adaptively Optimized PCNN Model for Hyperspectral Image Sharpening

Xinyu Xu 1, Xiaojun Li 1,2,3,* , Yikun Li 1,2,3, Lu Kang 4 and Junfei Ge 1

1 Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China; 11210900@stu.lzjtu.edu.cn (X.X.); liyikun@mail.lzjtu.cn (Y.L.); 12201844@stu.lzjtu.edu.cn (J.G.)
2 National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China
3 Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China
4 State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, China; kangl@lreis.ac.cn
* Correspondence: xjl@mail.lzjtu.cn

Abstract: Hyperspectral satellite imagery has developed rapidly over the last decade because of its high spectral resolution and strong material recognition capability. Nonetheless, the spatial resolution of available hyperspectral imagery is inferior, severely affecting the accuracy of ground object identification. In the paper, we propose an adaptively optimized pulse-coupled neural network (PCNN) model to sharpen the spatial resolution of the hyperspectral imagery to the scale of the multispectral imagery. Firstly, a SAM-CC strategy is designed to assign hyperspectral bands to the multispectral bands. Subsequently, an improved PCNN (IPCNN) is proposed, which considers the differences of the neighboring neurons. Furthermore, the Chameleon Swarm Optimization (CSA) optimization is adopted to generate the optimum fusion parameters for IPCNN. Hence, the injected spatial details are acquired in the irregular regions generated by the IPCNN. Extensive experiments are carried out to validate the superiority of the proposed model, which confirms that our method can realize hyperspectral imagery with high spatial resolution, yielding the best spatial details and spectral information among the state-of-the-art approaches. Several ablation studies further corroborate the efficiency of our method.

Keywords: hyperspectral sharpening; pulse-coupled neural network; multispectral image; remote sensing image fusion; high-resolution image

1. Introduction

Hyperspectral (HS) images are acquired by sampling the spectrum range into a large number of spectral channels, rendering them as enhanced multispectral (MS) images with multiple bands, a narrow spectral range, and abundant spectral information. Since the abundance of spectral bands enables HS imagery to identify ground cover types accurately, HS imagery is broadly applied in environmental monitoring [1], agricultural product assessment [2], geology [3], and mineralogical mapping [4]. However, because of the signal-to-noise ratio (SNR) constraints of satellite sensors, the spatial resolution and the spectral resolution of HS imagery would have to be inevitably compromised in a unique acquisition. Hence, HS images are characterized by high spectral resolution but low spatial resolution, which limits their applications in higher precision remote sensing (RS) interpretation. RS image sharpening is a cost-effective way to generate imagery with simultaneously rich spectral information and high spatial resolution by injecting spatial details into either HS or MS imagery, which could enhance the spatial resolution of the RS images. RS image sharpening has been extensively studied over the last four decades. Nonetheless, most of these methods aim to fuse the MS imagery and the panchromatic (PAN)
imagery, commonly known as pansharpening. As more HS satellites have been launched more recently (such as Earth Observing-1, ZY-1-02D, AVIRIS, and GF-5), HS sharpening has become increasingly prominent.

To solve the HS sharpening problem, many traditional pansharpening methods can still be applied. Pansharpening methods fall into two broad categories, namely the component substitution (CS) method and the multiresolution analysis (MRA) method [5]. The CS method involves projecting the original MS image onto the transform domain and substituting the spatial component with the PAN image [6–13]. Since the relatively large differences in spectral ranges between the PAN component and the replaced MS spatial component, the fusion result suffers from significant spectral distortion in the CS method. On the contrary, the MRA method preserves spectral information by injecting spatial details from the PAN imagery into the upsampled MS image, which is achieved by multiscale spatial filtering [14–20]. The MRA method is able to conserve the spectral information well, whereas it cannot get impressive spatial details.

HS sharpening aims to produce HS imagery with high spatial resolution by fusing low-resolution HS imagery and high-resolution MS imagery. In order to tackle the HS sharpening issue, traditional CS and MRA pansharpening methods are typically utilized by simply replacing both the PAN imagery and the MS imagery with the MS imagery and the HS imagery simultaneously [21]. Gomez et al. are the pioneers to apply the pansharpening approach to HS sharpening [22] and used the 2D-wavelet transform to render the MS and HS images fused in the same wavelength range. Chen et al. [23] present a generic HS sharpening framework based on the pansharpening method, which is the primary inspiration for later transmigration methods from pansharpening to HS sharpening. More recently, some methods for HS sharpening have also been proposed. Picone et al. addressed the band-assignment problem of the HS sharpening [24]. Lu et al. propose a spectral modulation hyper-sharpening framework [25], which mitigates the problem of large spectral distortion after fusion. Yokoya et al. [26] propose a coupled nonnegative matrix factorization method (CNMF), which generates the end members and abundance matrices by alternately unmixing HS and MS images with the NMF algorithm. However, the three-dimensional structure of the HS image is hardly reserved by the matrix factorization. Thus, tensor factorization is utilized in place of the matrix factorization [21]. Dian et al. present a nonlocal sparse tensor factorization method (NLSTF_SMBF) for the semi-blind fusion of HS and MS images [27], which primarily constructs full-band patches (FBPs), and similar FBPs can share the same dictionary.

In recent years, deep learning (DL) methods have been extensively applied in the field of HS sharpening [28,29]. Zhang et al. [30] proposed an unsupervised deep learning network architecture for the simultaneous optimization of the HS super-resolution and the degradation estimation. Qu et al. [31] proposed an unsupervised, unaligned Mutual HS super-resolution Dirichlet-Net, which effectively improves the robustness in the face of alignment errors.

Although promising fusion results are realized by various HS sharpening methods, it is expected to further improve the HS fusion accuracy by neural network algorithms. Nonetheless, due to the independent spectral features of the unique RS image, it is difficult to establish a universal training database for the RS fusion applications. Besides, it is time-consuming to train the network. A pulse-coupled neural network (PCNN) is a kind of biologically inspired neural network without training, which exhibits the characteristics of the pulse-synchronous phenomenon. The property of pulse-synchronous allows the synchronously stimulated pixels to generate the segmentation results, which turns out to be in accordance with the human visual mechanism [32]. In order to improve the accuracy of image fusion, it is useful to adopt different fusion strategies for different segmentation regions that are consistent with human perception. Traditional PCNN fusion models include simplified PCNN (SPCNN) [33], dual-channel PCNN (DCPCNN) [34], and shuffled frog-leaping PCNN [35]. Particularly, Panigrahy et al. propose adaptive DCPCNN for multi-aggregation and medical image fusion [36,37]. However, traditional PCNN fusion
approaches are applicable only to either the medical image fusion or the multi-focus image fusion, which cannot be applied to the RS image fusion. Recently, Li et al. presented a modified PCNN model for RS image fusion [32], but it only works on MS imagery.

In this paper, a novel adaptively optimized PCNN model for HS sharpening is proposed. Concretely, we summarize the main contributions of the paper as follows: (1) a SAM-CC band assignment method is proposed to group the HS bands with the MS band; (2) an improved PCNN (IPCNN) model is proposed to obtain irregular injection region of spatial details. (3) A Chameleon Swarm Optimization (CSA)-based IPCNN parameter optimization method is designed to achieve the optimal sharpening imagery. The comparative experiments were carried out on three datasets captured by the ZY-1-02D and GF-2 satellites, and the results substantiate the effectiveness of the proposed method.

The rest of the article is organized as follows. The description of the PCNN and CSA principles is given in Section II. Section III presents the details of our proposed fusion approach. Experimental results and discussions are provided in Section IV. Section V contains the conclusion of the paper.

2. Related Work

2.1. Standard PCNN Principle

PCNN belongs to the third generation of artificial neural networks, which is a kind of neural network model proposed by Johnson [38] on the basis of the observation of pulse delivery experiments in the cerebral cortex of cats and monkeys. The individual neuron in the PCNN model is partitioned into the receptive field, the modulation field, and the pulse delivery field according to its functions. Mathematical formulae of the receptive field are given in Equations (1) and (2). The mathematical description of the modulation field can be found in Equation (3), and the mathematical formulae of the pulse delivery field can be defined in Equations (4) and (5).

\[ F_y[n] = e^{-\alpha} F_y[n-1] + V_f \sum M_{y}^{W_{y}} Y_y[n-1] + I_y \]  

(1)\[ L_y[n] = e^{-\alpha} L_y[n-1] + V_l \sum W_{y}^{I_{y}} Y_y[n-1] \]  

(2)\[ U_y[n] = F_y[n] \left( 1 + \beta L_y[n] \right) \]  

(3)\[ Y_y[n] = \begin{cases} 1 & U_y[n] > E_y[n] \\ 0 & \text{otherwise} \end{cases} \]  

(4)\[ E_y[n+1] = e^{-\alpha} E_y[n] + V_e Y_y[n] \]  

(5)

where the subscripts \( ij \) and \( kl \) refer to the positions of the current neuron and the neighboring neurons, respectively. \( n \) denotes the \( n \)th iteration. \( I \) is the input image, whereas \( I_y \) corresponds to the external input of the neuron \( ij \). \( F \) and \( L \) indicate the feeding input and the linking input, respectively, while the difference between them is that the linking input \( L \) can only receive the local pulse stimulation from the neighboring neuron via the synaptic matrix \( W \). On the contrary, \( F \) receives not only the local stimulation but also the external stimulation from \( I \). \( \beta \) denotes the linking strength, which ranges from 0 to 1. A larger value of \( \beta \) indicates that the linking relationships are greater, i.e., the current neuron is more susceptible to neighborhood neurons. \( M \) and \( W \) represent the connection coefficient matrices, which are typically calculated by the Euclidean distance between the current neuron and the neighborhood. \( U \) refers to the internal activity, and \( E \) represents the dynamic threshold. \( U \) and \( E \) jointly determine whether or not the current neuron is stimulated in the current iteration, i.e., the boolean variable \( Y_y \) becomes equal to one whenever \( U_y \) is greater than \( E_y \). \( \alpha_l \), \( \alpha_e \), and \( \alpha_r \) are time decay constants, which are utilized to adjust the
decay rate of $F$, $L$, and $E$. $V_F$, $V_L$, and $V_E$ indicate the normalization constants. The matrix $Y[n]$ denotes the binary output of PCNN in each iteration $n$, whereas $Y_{ij}$ is 0 (when the neuron $ij$ is unstimulated) and 1 (when stimulated), respectively.

In addition, as illustrated in Figure 1, each PCNN neuron corresponds to a pixel in the image when applying PCNN to image processing applications. Therefore, the current neuron $ij$ will release a pulse as long as $U_{ij}$ is greater than $E_{ij}$. If the current pixel is stimulated in the current iteration, $E$ will automatically spike the value $V_E$; otherwise, $E$ will decrease gradually as the iteration number $n$ increases. After the $n$th iteration of PCNN, the neurons with either similar grayscale values or the adjacent positions will be synchronously stimulated, while the output $Y$ will eventually form a binary ignition map.

![Figure 1. Connection relationship diagram of PCNN.](image)

2.2. Chameleon Swarm Optimization Algorithm

The CSA algorithm is a meta-heuristic method proposed by Malik [39]. The model imitates the socially intelligent synergistic behavior of chameleons when foraging and capturing food near woods, swamps, and deserts. It is a bio-optimization algorithm for finding the global optimum of nonlinear, nonconvex, and other complex problems, which may prevent entrapment in the local optimum.

The principle of CSA is shown in Figure 2. The algorithm mathematizes the behavioral stages of a chameleon in looking for food, which include initializing the starting position, tracking the prey from a distance, locating the prey by eye rotation, and catching the prey with a high-velocity sticky tongue.
Start

Set parameters and initialize the population.

Set fitness function, calculated fitness values of the initial chameleon group according to $f$, and record the optimal value and its position.

Update the parameters: $\mu$, $\omega$, and $\alpha$.

$t=1$

Whether $r_i \geq P_p$?

Yes

Execute the part of $r_i \geq P_p$, and update the position.

No

Execute the part of $r_i < P_p$, and update the position.

Update the position of eyes after rotation.

Update the position of the attacked prey.

Calculated fitness values of the current chameleon group according to $f$, and record the optimal value and its position.

$t = t + 1$

Whether $t \leq T$?

Yes

No

Output global optimal value and its position.

End

Figure 2. Flowchart of CSA. See reference [39] for all variable definitions.
3. Proposed Method

In order to tackle the problem of spectral distortion and fuzzy spatial details in HS sharpening, we propose a novel adaptively optimized PCNN fusion algorithm. The flowchart of the proposed HS sharpening algorithm is shown in Figure 3, which contains the following modules: (1) the SAM-CC band assignment part; (2) the improved PCNN (IPCNN) model part; (3) the automatic parameter optimization of IPCNN by CSA part; (4) the extracting MS detail part; (5) the adaptive injected gains part; (6) the fusion output part.

Figure 3. Flowchart of the proposed fusion algorithm.

3.1. SAM-CC Band Assignment Block

In contrast to the multi-to-single band assignment in the pansharpening problem, HS sharpening is a multiband-to-multiband fusion. Thus, as shown in Figure 4, correctly assigning each HS band to the MS band is an essential procedure before the actual fusion operation, i.e., the choice of which band in the MS imagery can be used to sharpen the corresponding HS bands plays a crucial role. Classical band selection methods include the minimum spectral distance (MSD) assignment algorithm, the maximum cross-correlation (CC) assignment algorithm, and the minimum spectral distortion (SAM) assignment algorithm [24]. Since the optimal band selection is challenging for any single criterion, we propose a joint band selection algorithm using both the SAM and CC indices (SAM-CC) in order to group the HS bands more accurately.
Figure 4. Schematic diagram of HS and MS bands grouping.

Let $H = \{H_h\}_{h=1,\ldots,N_h}$ stands for the HS images and $M = \{M_m\}_{m=1,\ldots,N_m}$ represents the MS images, where $h$ and $m$ are the band numbers of HS and MS, respectively. $N_h$ and $N_m$ refer to the total number of bands of HS and MS. The proposed SAM-CC criterion is represented as Equation (6).

$$
SAM-CC\left(H_h, M'_m\right) = \arg\min_{\alpha} \left( E - CC\left(H_h, M'_m\right) \right) \times SAM\left(H_h, M'_m\right)
$$

(6)

$$
CC\left(H_h, M'_m\right) = \frac{\langle H_h, M'_m \rangle}{\sqrt{\langle H_h, H_h \rangle \langle M'_m, M'_m \rangle}}
$$

(7)

$$
SAM\left(H_h, M'_m\right) = \left( \frac{1}{N_h} \right) \sum_{n=1}^{N_h} \arccos\left( \frac{\langle H_h, M'_m \rangle}{\|H_h\| \cdot \|M'_m\|} \right)
$$

(8)

where $E$ is a unit matrix with size $N_M \times N_H$. The symbol $\langle \cdot \rangle$ indicates the inner product operation. The symbol $\| \cdot \|$ stands for the $l_2$ norm. $r$ refers to the spatial resolution ratio of HS to MS. $M'_m$ represents the low-pass filtered down-sampled image of $M_m$.

3.2. Improved PCNN Model

To make the PCNN model more suitable for HS sharpening, we propose an IPCNN model in the paper. In the standard PCNN model, the feeding input $F$ stands for the accumulative influence of the external stimuli. Nonetheless, the human eyes are more sensitive to the edges and orientations rather than individual pixels within an image. Therefore, a new feeding input $SF$ is designed in the IPCNN model, which considers the local neighborhood differences in both the horizontal and vertical directions. The $SF$ is calculated in Equations (9)–(11).

$$
SF_{ij} = \sqrt{RF_{ij}^2 + CF_{ij}^2}
$$

(9)
where \( RF_{ij} \) denotes the row difference of the two neighboring feeding inputs \( F \) via the local rectangular window \( M_1 \times M_2 \), and \( CF_{ij} \) indicates the column difference. After performing a number of experiments, we take the window size to be \( 5 \times 5 \). Furthermore, \( I_{ij}' \) is designed to strengthen the impact of the local neighborhood, which is given by Equation (12).

\[
I_{ij}' = \frac{1}{2} \left( H_{ij} + \frac{1}{M_1 \times M_2} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} H_{ij}' \right)
\]

where \( H_{ij} \) refers to the pixel value of the upsampled HS in row \( i \) and column \( j \).

After modeling the new \( SF \) and \( Y' \), the proposed IPCNN neuron model is illustrated in Figure 5. The IPCNN model has the following advantages compared to the standard PCNN: (1) the IPCNN model utilizes the spatial difference of the neighborhoods to stimulate the feeding input \( F_{ij} \) of the neuron, which can describe the local detail features; (2) the external stimuli \( I_{ij}' \) considers the influence of the surrounding pixels towards the central pixels.

---

**Figure 5.** IPCNN neuron model.

### 3.3. Automatic Parameter Optimization of IPCNN by CSA

Similar to the standard PCNN, the setting of appropriate parameters (i.e., \( \alpha_F, \alpha_L, \alpha_E, \beta \)) of the IPCNN model is most critical. As shown in Figure 6, different combinations of IPCNN parameters will lead to different segmentation results for the same input image. If there are too many segmentation pieces, the segmentation regions will be too small and will further affect the accuracy and complexity of the statistical computation of subsequent injection weights. Conversely, if the number of image segmentation pieces is too few, it is impossible to distinguish each region or to take advantage of the characteristics of each region. So, different parameters can lead to different segmentation results, and different segmentations will have corresponding impacts on the final fusion results. Most researchers choose to simplify the PCNN model or to use the manually set uniform parameters for the different images. However, fixed parameters for different input images do not lead to the optimal fusion results for all images. Thus, we propose a CSA-based optimization approach for setting IPCNN parameters \( \alpha_r, \alpha_a, \alpha_e, \beta, \) and \( W \), which can adaptively generate their own optimal parameters for various input images.
Figure 6. IPCNN segmentation results under pseudo-color maps with different parameters of IPCNN. (a) Original image of GF-2 satellite sensor. (b) Classical parameters combination, \([\alpha_F, \alpha_L, \alpha_E, \beta, w] = [0.1, 1, 0.62, 0.1, 0.707]\). (c) Change the parameters \(\alpha_E\) and \(w\), \([\alpha_F, \alpha_L, \alpha_E, \beta, w] = [0.1, 1, 1, 0.1, 0.9]\). (d) Change all parameters, \([\alpha_F, \alpha_L, \alpha_E, \beta, w] = [0.5, 2.5, 2, 0.3, 0.9]\).

The IPCNN parameters need to be jointly optimized together, so the CSA is employed to automatically optimize all five IPCNN parameters. For convenience, the connection weight \(W\) is denoted as \(W = [w, 1, w; 1, 0, 1; w, 1, w]\), where \(w\) is optimized rather than the entire matrix \(W\). The flowchart of the automatic IPCNN parameter optimization algorithm based on CSA is shown in Figure 7. Firstly, the chameleon position is initialized with the classical parameter values. Secondly, the fitness function is set as the weighted summation of spectral fidelity SAM [5] and spatial detail representation ERGAS [40]. The fitness function for the proposed optimization method is given as follows.

\[
f = \frac{SAM_h}{ERGAS_h + SAM_h}ERGAS + \frac{ERGAS_h}{ERGAS_h + SAM_h}SAM
\]

(13)

\[
ERGAS(\text{FH}, \text{F}) = 100 \frac{h}{l} \sqrt{\frac{1}{N_h} \sum_{h=1}^{N_h} \left( \frac{RMSE_h}{\mu_h} \right)^2}
\]

(14)

\[
SAM(\text{FH}, \text{F}) = \arccos \left( \frac{\langle f_{h_1}, f_{l_1} \rangle}{\| f_{h_1} \| \| f_{l_1} \|} \right)
\]

(15)

where \(\text{FH}\) denotes the reference imagery, and \(\text{F}\) refers to the fusion imagery. \(SAM_h\) and \(ERGAS_h\) stand for the ranges of SAM and ERGAS. \(h\) and \(l\) indicate the spatial resolutions of the MS and HS imagery, respectively. \(\mu_h\) indicates the mean value of the \(h\)th band of the reference imagery. <\> means the inner product. \(\| \cdot \|_2\) refers to the second-order norm operation.

Figure 7. Flowchart of CSA-based automatic IPCNN parameter optimization algorithm.
Since there are too many bands in the HS image, directly optimizing the parameters for all bands will be time-consuming. Now that the adjacent HS bands acquired by the SAM-CC band assignment method are usually strongly correlated, it would be more efficient to obtain optimized parameters by only using the representative bands of each group. Wang et al. proposed a HS band selection method with optimal domain reconstruction [41], which can adaptively achieve the optimal neighborhood reconstruction (ONR) to find the subsets of bands that can best represent the HS image. Thus, the ONR band selection method is used before IPCNN parameter optimization. In addition, the method can also reduce the impacts of noisy bands by taking advantage of the characteristics of neighborhood bands.

To be more efficient, the optimal IPCNN parameters are generated with the representative bands of HS images. Other HS bands could utilize these parameters as long as both bands are very similar to each other. Representative bands from each group are chosen by the ONR method. If there are more than five HS bands in each group, five of these are then selected using ONR to compose the optimal subset of bands. Otherwise, all of the bands from the original HS within the group are retained. Equation (16) is defined to express the procedure.

\[
H_{\hat{m}} = \left\{ H_j \right\}_{j \in \mathbb{N}} \xrightarrow{\text{ONR band selection}} H_{\hat{m}} = \begin{cases} \left\{ H_j \right\}_{j \in \mathbb{N}} & j > 5, j j = 5 \\ H_j & j \leq 5 \end{cases}
\]

3.4. Extracting MS Details

The multiscale analysis has demonstrated excellent capability in RS image fusion. The “atrous” wavelet transform is an undecimated multiscale analysis method, which is widely used to address image fusion issues due to its fast decomposition and reconstruction as well as the richer extracted details. For each histogram-matched MS band \( M_{hm} \) (matched by HS image), the “atrous” wavelet decomposition is performed to extract the spatial details from the multispectral imagery. After wavelet decomposition, the low-frequency component stands for the approximate image, and the high-frequency component indicates the noise and the local features. After setting the high-frequency components to zero, the low-frequency MS imagery \( M_i \) is then realized via wavelet reconstruction. Finally, the spatial detail information \( M_d \) is obtained through Equation (17).

\[
M_d = M_{hm} - M_i
\]

3.5. Adaptive Injected Gains

The HS imagery is divided into different irregular segmentation regions by the proposed CSA-based IPCNN segmentation algorithm. The injected gains are, therefore, calculated in each irregular region. The overall steps of the adaptive gain approach are described as follows:

1. Initialize IPCNN. Let \( V_r = 0.5, V_t = 0.2, V_L = 20, Y \{ \} = L \{ \} = U \{ \} = 0, E \{ \} = V_\epsilon \).
2. Optimize IPCNN parameters \( a_r, a_t, a_c, \beta, \) and \( W \) using the CSA-based IPCNN optimization algorithm.
3. Obtain the irregular segmentation region of the IPCNN model in the current iteration \( n \). And calculate the injected gain \( G[n] \) according to Equation (20).

\[
R_g[n] = \begin{cases} \frac{\text{cov}(H^*_r(i, j), M_L(i, j))}{\text{cov}(M_L(i, j), M_L(i, j))} & \text{if } Y[i][n] \neq 0 \\ 0 & \text{otherwise} \end{cases}
\]
\[
g^*_i = \text{corr} \left( H^*_i(i, j), M_L(i, j) \right) \frac{\text{std}(H_u)}{\text{std}(M_m)}
\]

\[
G^*[n] = \begin{cases} 
G^*_i \frac{\text{std}(H^*_i(i, j))}{\text{std}(M_L(i, j))} & \text{if } R_y[n] > 0 \text{ and } Y_y[n] \neq 0 \\
1 & \text{otherwise}
\end{cases}
\]

where \((i, j)\) stands for the pixel coordinates of the imagery. \(H_u\) denotes the upsampled image of HS, and \(M_l\) indicates the low-resolution version of the MS imagery. \(Y[n]\) refers to the activated neurons in the current iteration. \(R\) is the correlation ratio between the HS and MS imagery. \(\text{corr}(A, B)\) represents the correlation coefficient between matrix \(A\) and matrix \(B\. \text{cov}(A, B)\) denotes the covariance between matrix \(A\) and matrix \(B\. \text{std}(A)\) refers to the standard deviation of \(A\).

3.6. Fusion Output

The final fusion result can be calculated by Equation (21).

\[
\text{FH}_u = H^*_u + G^*M_m
\]

where \(\text{FH}_u\) denotes the high-resolution HS fusion image.

The fusion pseudo-code of the proposed HS sharpening algorithm is described as follows.

<table>
<thead>
<tr>
<th>Algorithm 1</th>
<th>AT-AIPCNN (“atrous” transform-adaptive IPCNN) method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong></td>
<td>(H, M, r%) HS imagery, MS imagery, and spatial resolution ratio between HS and MS</td>
</tr>
<tr>
<td><strong>Output:</strong></td>
<td>(\text{FH}%) HS sharpening result</td>
</tr>
<tr>
<td><strong>for</strong> (i = 1) to (N_h) do</td>
<td></td>
</tr>
<tr>
<td><strong>for</strong> (j = 1) to (N_m) do</td>
<td></td>
</tr>
<tr>
<td>( (H_{xy}, M_{xy}) \leftarrow \text{SAM-CC} (H, M) ) &amp; % Band assignment</td>
<td></td>
</tr>
<tr>
<td>( (H_{xy}, M_{xy}) \leftarrow \text{ONR} (H_{xy}, M) ) &amp; % Band selection</td>
<td></td>
</tr>
<tr>
<td><strong>for</strong> (t = 1) to (T_{\text{max}}) do</td>
<td></td>
</tr>
<tr>
<td>( (y = \alpha r, \alpha s, \beta, w) \leftarrow \text{CSA} (H_{xy}, M) ) &amp; % IPCNN parameters optimization</td>
<td></td>
</tr>
<tr>
<td>Reaching iteration stop conditions &amp;</td>
<td></td>
</tr>
<tr>
<td>( (y = \alpha r', \alpha s', \beta', w') ) &amp; % Optimal parameters combination</td>
<td></td>
</tr>
<tr>
<td><strong>for</strong> (j = 1) to (N_m) do</td>
<td></td>
</tr>
<tr>
<td>( (H) \leftarrow \text{upsampling} (H) ) &amp; % HS image upsampling</td>
<td></td>
</tr>
<tr>
<td>( (M_{hm}, H_{hm}) \leftarrow \text{histogram matching} (H_{xy}, M) ) &amp; % Histogram matching</td>
<td></td>
</tr>
<tr>
<td>( (M_t, M_o) \leftarrow ”\text{atrous”} (M) ) &amp; % Extracting MS Details</td>
<td></td>
</tr>
<tr>
<td>( H_{hm}'' \leftarrow \text{IPCNN} (H_{hm}') ) &amp; % IPCNN Segmentation</td>
<td></td>
</tr>
<tr>
<td>( G_t \leftarrow \text{correlation} (M_t, H_{hm}'', H_{hm}'') ) &amp; % Injection coefficient calculation</td>
<td></td>
</tr>
<tr>
<td>( \text{FH} \leftarrow \text{MRA} (M_t, H_{hm}'', H_{hm}'') ) &amp; % Fusion output</td>
<td></td>
</tr>
</tbody>
</table>

4. Experimental Results

4.1. Datasets

Three real datasets captured from the ZY1-02D and GF-2 satellites are utilized to test the validity of the proposed method. The HS sensor AHSI of the ZY1-02D satellite, which works in the spectral range from the visible to the short-wave infrared (with the wavelength from 395 nm to 2500 nm), provides 166 HS bands with a spatial resolution of 30 meters. The GF-2 satellite captures four multispectral bands with a spatial resolution of 3.4 meters, spanning the spectral range from the visible to the near-infrared (450–890 nm). For HS sharpening application, HS images from the ZY1-02D sensor would need to be
sharpened to the spatial resolution of MS imagery of the GF-2 sensor. Thus, dataset1 consists of imagery taken over the Liujiaxia reservoir in China from the ZY1-02D and GF-2 sensors, which mainly include the lake and the village. Dataset2 stands for the suburban area of Linxia City in China, mainly containing mountainous terrain. The dataset3 is the farmland area of Yongchang City in China, which primarily includes farmland, mountains, and buildings. The original images of the three datasets are shown in Figure 8.

![Figure 8. The original images of three datasets. (a) HS image of Liujiaxia dataset. (b) HS image of Linxia dataset. (c) HS image of Yongchang dataset. (d) MS image of Liujiaxia dataset. (e) MS image of Linxia dataset. (f) MS image of Yongchang dataset.](image)

### 4.2. Experimental Setup

In the experiments, most program code is executed with matlab2020a on the Intel CPU Core i7-13700K and NVIDIA GPU GeForce GTX 3090. The number of chameleon populations is set to 20 in the CSA-based IPCNN adaptive optimization, and the maximum number of iterations is set to 30.

Nine classical and competitive methods of different sharpening categories are compared with the proposed approach, i.e., the Gram-Schmidt adaptive (GSA) algorithm [12], the smoothing filtered-based intensity modulation (SFIM) algorithm [42], the generalized Laplacian pyramid (GLP) algorithm [43], the CNMF algorithm [26], the nonlocal sparse tensor factorization (NLSTF) algorithm [27], the NLSTF_SMBF algorithm [27], the HYSURE algorithm [44], the fast fusion based on Sylvester equation (FUSE) algorithm [45], and the UDALN algorithm [46]. Among which, GSA, SFIM, and GLP are owned by the classical pansharpening methods, GSA belongs to the CS methods, SFIM and GLP are the MRA method, CNMF belongs to the matrix decomposition methods, NLSTF and NLSTF_SMBF are owned by tensor decomposition categories, while the HYSURE and FUSE belong to Bayesian-based methods. In addition, GSA, SFIM, and GLP do not require any parameter setting. The main parameters for the NLSTF and NLSTF_SMBF methods are set according to Table 1, and the parameters of other comparative approaches are set to be the same as in the literature [47]. The UDALN is implemented in PyTorch, and the parameter settings are the same as in the literature [46].

| Table 1. Main parameter settings of NLSTF/NLSTF_SMBF methods. |
|-----------------|-----------------|-----------------|
| Parameter       | Value           | Description     |
| Selected         |                 |                 |
| Number of iterations | 30          |                 |
| Parameter       | Value           | Description     |
| Selected         |                 |                 |
| Number of iterations | 30          |                 |
Method | Main Parameters
--- | ---
NLSTF/NLSTF_SMBF | The atomic numbers for three different dictionaries: \( l_W = 10, \; l_H = 10, \; l_S = 14 \).
 | Parameters of sparse regularization: \( \lambda = 10^{-6}, \; \lambda_1 = 10^{-5}, \; \lambda_2 = 10^{-5}, \; \lambda_3 = 10^{-6} \).
 | Cluster scaling parameter: \( K = 151 \).
 | The spectral response matrix \( R \) was estimated by HYSURE [44].

The performance of the proposed HS sharpening approach is evaluated by eight complementary quantitative indices to verify the spectral and spatial qualities of the fusion imagery, such as the peak signal-to-noise ratio (PSNR) [47], the root mean square error (RMSE) [48], the error relative global adimensionnelle de synthèse (ERGAS) [40], the spectral angle mapper (SAM) [5], the structural similarity index measurement (SSIM) [49,50], the universal image quality index (UIQI) [48], the inter-correlation (CC) [51], and the degree of distortion (DD) [52]. The ideal values for RMSE, ERGAS, SAM, and DD are 0, whereas 1 for SSIM, UIQI, and CC. Moreover, the ideal value for PSNR is positive infinity.

PSNR measures the spatial similarity between the fusion imagery and the reference imagery. In general, the larger the PSNR, the more spatially similar the fusion imagery is to the reference imagery. The PSNR is defined as:

\[
PSNR(FH, F) = \frac{1}{N_h} \sum_{h=1}^N \log_{10} \left( \frac{\max(FH_h^2)}{\frac{1}{L} \sum_{j=1}^L (FH_{hj} - F_{hj})^2} \right)
\]

where \( FH \) denotes the reference imagery, and \( F \) refers to the fusion imagery. \( h \) stands for the band number, while \( N_h \) indicates the total number of all bands. \( \log_{10}(\cdot) \) represents the logarithm function with base 10. \( FH_h \) denotes the \( h \)th band of the reference imagery, \( \max(\cdot) \) denotes the maximum value. \( L \) denotes the total number of pixels per band. \( FH_{hj} \) denotes the \( j \)th pixel of the \( h \)th band of the reference imagery, and \( F_{hj} \) stands for the \( j \)th pixel of the \( h \)th band of the fusion image.

RMSE refers to the measure of deviation between two images. The RMSE is defined as:

\[
RMSE(FH, F) = \sqrt{\frac{1}{T} \sum_{h=1}^N \sum_{j=1}^L (FH_{hj} - F_{hj})^2}
\]

where \( T \) stands for the total pixel number of the reference imagery.

ERGAS is a comprehensive index that reflects both spectral distortions and spatial detail differences. The ERGAS can be computed through Equation (14). SAM calculates the similarity of the spectral vectors. The smaller one indicates that the spectrum is better maintained. The SAM index of two images \( FH \) and \( F \) is computed as Equation (15).

UIQI measures the similarity in terms of the brightness and contrast. For the window, \( a \times b \), \( Q \) is defined as:

\[
Q(a,b) = \frac{4\mu_x\mu_y}{\mu_x^2 + \mu_y^2 + \sigma_x^2 + \sigma_y^2}
\]

where \( \mu_x \) and \( \sigma_x^2 \) denote the expectation and variance of \( x \), respectively, and \( \sigma_{xy}^2 \) represents the covariance of \( x \) and \( y \).

Then, the UIQI can be computed through the following formula:

\[
UIQI(FH, F) = \frac{1}{N_h} \sum_{h=1}^N Q(FH, F)
\]
SSIM compares the structural similarities between two images, which indicate the luminance distortion, the contrast distortion, and the structural distortion. Its formula is described in Equations (26)–(27).

\[
\text{SSIM} (\mathbf{F}_h, \mathbf{F}) = \frac{1}{N_h} \sum_{k=1}^{N_h} \text{SSIMS} (\mathbf{F}_k, \mathbf{F}_h)
\]

\[
\text{SSIMS} (a, b) = \frac{\left(2\mu_a\mu_b + C_1\right) \left(\sigma_a^2 + \sigma_b^2 + C_2\right)}{\left(\mu_a^2 + \mu_b^2 + C_1\right) \left(\sigma_a^2 + \sigma_b^2 + C_2\right)}
\]

where \(C_1\) and \(C_2\) are infinitesimal small constants to ensure stability.

DD is an indicator to verify the quality of the fusion spectra. The smaller the value, the better the spectral retention. If the DD is equal to 0, there is no spectral distortion. Its calculation formula is shown in Equation (28).

\[
\text{DD} (\mathbf{F}_h, \mathbf{F}) = \frac{1}{T} \| \text{vec} (\mathbf{F}_h) - \text{vec} (\mathbf{F}) \|_1
\]

where \(T\) denotes the total pixel number of the reference imagery. \(\text{vec}(\mathbf{x})\) represents matrix \(\mathbf{x}\) as a vector, and \(\| \cdot \|_1\) is the one-order norm operation.

CC represents the spatial correlation between two images, which measures the geometric distortion of the fusion imagery. It is calculated as follows:

\[
\text{CC} (\mathbf{F}_h, \mathbf{F}) = \frac{1}{N_h} \sum_{k=1}^{N_h} \text{CCS} (\mathbf{F}_k, \mathbf{F}_h)
\]

\[
\text{CCS} (a, b) = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} (a_{ij} - \mu_a) (b_{ij} - \mu_b) \frac{1}{\sqrt{\sum_{i=1}^{H} \sum_{j=1}^{W} (a_{ij} - \mu_a)^2 \sum_{i=1}^{H} \sum_{j=1}^{W} (b_{ij} - \mu_b)^2}}
\]

where \(X_h\) refers to the \(h\)th band of \(X\), and \(\mu_x\) indicates the mean value of \(x\).

4.3. Experimental Results

The first experiment is conducted on the Liujiaxia dataset. The fusion results are shown in Figure 9, in which the local area is enlarged for convenient observation. As can be seen from Figure 9, the color of the GSA, SFIM, GLP, CNMF, HYSURE, UDALN, and FUSE methods appear to be dark. Our proposed method, NLSTF and NLSTF_SMBF methods, perform best in spectral preservation; nevertheless, the details of both the NLSTF and NLSTF_SMBF methods are blurred into chunks. Figure 10 presents the SAM error maps of the fusion images in all 166 bands with the Liujiaxia dataset. It is noted that the spectral distortion for all methods primarily occurs at the reservoir-mountain boundary, of which NLSTF and NLSTF_SMBF exhibit the largest aberration. Besides, UDALN has larger distortions in the reservoir area, and the proposed method, SFIM, and GLP perform better in the SAM error maps. Furthermore, quantitative assessments with the Liujiaxia dataset are shown in Table 2. It is clear from Table 2 that the proposed method obtains the best performance across all quantitative metrics, which demonstrates the superiority of the proposed method with respect to spectral preservation and spatial details.
Figure 9. Color composite image of fusion results with Liujiaxia dataset (bands 35,14,7 are input to RGB channels respectively). (a) Proposed method. (b) GSA. (c) SFIM. (d) GLP. (e) CNMF. (f) NLSTF. (g) NLSTF_SMBF. (h) HYSURE. (i) FUSE. (j) UDALN.

Figure 10. SMA error map of Liujiaxia dataset. (a) Proposed. (b) GSA. (c) SFIM. (d) GLP. (e) CNMF. (f) NLSTF. (g) NLSTF_SMBF. (h) HYSURE. (i) FUSE. (j) UDALN.

Table 2. Quantitative evaluation results of the Liujiaxia dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>SAM (°)</th>
<th>UIQI</th>
<th>SSIM</th>
<th>DD</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>37.6677</td>
<td>5.0560</td>
<td>1.2971</td>
<td>1.5961</td>
<td>0.7252</td>
<td>0.9296</td>
<td>2.3748</td>
<td>0.9863</td>
</tr>
<tr>
<td>GSA</td>
<td>27.6248</td>
<td>17.2337</td>
<td>5.1195</td>
<td>2.1357</td>
<td>0.5535</td>
<td>0.8875</td>
<td>11.6032</td>
<td>0.9767</td>
</tr>
<tr>
<td>SFIM</td>
<td>25.8724</td>
<td>21.2343</td>
<td>6.8882</td>
<td>1.7852</td>
<td>0.5902</td>
<td>0.8613</td>
<td>14.6026</td>
<td>0.9768</td>
</tr>
<tr>
<td>GLP</td>
<td>25.6146</td>
<td>21.8812</td>
<td>7.2056</td>
<td>1.7573</td>
<td>0.5920</td>
<td>0.8588</td>
<td>15.0508</td>
<td>0.9777</td>
</tr>
<tr>
<td>CNMF</td>
<td>26.7564</td>
<td>18.9197</td>
<td>5.8796</td>
<td>2.1720</td>
<td>0.4619</td>
<td>0.8707</td>
<td>12.8621</td>
<td>0.9653</td>
</tr>
<tr>
<td>NLSTF</td>
<td>26.7253</td>
<td>18.3908</td>
<td>4.4832</td>
<td>5.3202</td>
<td>0.3553</td>
<td>0.7393</td>
<td>11.5977</td>
<td>0.8935</td>
</tr>
<tr>
<td>NLSTF_SMBF</td>
<td>25.4524</td>
<td>19.5991</td>
<td>7.4406</td>
<td>9.6997</td>
<td>0.2621</td>
<td>0.6833</td>
<td>12.1336</td>
<td>0.8094</td>
</tr>
</tbody>
</table>
The fusion results with the Linxia dataset are shown in Figure 11. As can be seen, the large color difference and block blur are presented in the NLSTF and NLSTF_SMBF fusion methods. Moreover, the SFIM fusion image appears dark. GSA algorithm, as a kind of CS method, has greater spectral distortion, while MRA methods, such as SFIM and GLP, have less spatial detail. Furthermore, the proposed method, GSA, and UDALN, have fine spatial detail information. In addition, the spectra of the proposed method are the closest to the ground-truth image in subjective visualization. Figure 12 shows the SAM error maps of all the fusion results with the Linxia dataset. As can be seen from Figure 12, spectral distortion is more likely to occur in the ridge area and in the land-lake boundary. Although the SFIM method has less spectral distortion in most regions, it exhibits larger errors in some local areas (top left corner and bottom left of Figure 12d). NLSTF and NLSTF_SMBF appear to have more spectral distortion, whereas the proposed method performs well. Table 3 gives the quantitative evaluation results of the Linxia dataset, which demonstrates that our method achieves the best indicator results.

<table>
<thead>
<tr>
<th></th>
<th>HYSURE</th>
<th>FUSE</th>
<th>UDALN</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYSURE</td>
<td>27.9512</td>
<td>16.5909</td>
<td>4.8475</td>
</tr>
<tr>
<td>FUSE</td>
<td>23.1458</td>
<td>29.1700</td>
<td>11.5544</td>
</tr>
<tr>
<td>UDALN</td>
<td>30.1948</td>
<td>13.0902</td>
<td>3.1310</td>
</tr>
</tbody>
</table>

The best result is in bold, and the second best is underlined.

Figure 11. Color composite image of fusion results with Linxia dataset (bands 35,14,7 are input to RGB channel respectively). (a) Proposed method. (b) GSA. (c) SFIM. (d) GLP. (e) CNMF. (f) NLSTF. (g) NLSTF_SMBF. (h) HYSURE. (i) FUSE. (j) UDALN.
Figure 12. SMA error map of Linxia dataset. (a) Proposed method. (b) GSA. (c) SFIM. (d) GLP. (e) CNMF. (f) NLSTF. (g) NLSTF_SMBF. (h) HYSURE. (i) FUSE. (j) UDALN.

Table 3. Quantitative evaluation results of the Linxia dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>SAM (°)</th>
<th>UIQI</th>
<th>SSIM</th>
<th>DD</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>36.2331</td>
<td>5.7254</td>
<td>0.9960</td>
<td>1.9634</td>
<td>0.8498</td>
<td>0.9067</td>
<td>3.4608</td>
<td>0.9645</td>
</tr>
<tr>
<td>GSA</td>
<td>27.7164</td>
<td>16.0083</td>
<td>3.0327</td>
<td>2.6527</td>
<td>0.8158</td>
<td>0.8716</td>
<td>12.4741</td>
<td>0.9393</td>
</tr>
<tr>
<td>SFIM</td>
<td>17.7201</td>
<td>52.0659</td>
<td>22.7315</td>
<td>2.5678</td>
<td>0.3656</td>
<td>0.5387</td>
<td>42.0473</td>
<td>0.9452</td>
</tr>
<tr>
<td>GLP</td>
<td>29.1091</td>
<td>13.5721</td>
<td>2.4955</td>
<td>2.3762</td>
<td>0.8256</td>
<td>0.8826</td>
<td>10.5857</td>
<td>0.9526</td>
</tr>
<tr>
<td>CNMF</td>
<td>29.4453</td>
<td>12.4878</td>
<td>2.3389</td>
<td>2.8221</td>
<td>0.7854</td>
<td>0.8723</td>
<td>8.9418</td>
<td>0.8916</td>
</tr>
<tr>
<td>NLSTF</td>
<td>26.7873</td>
<td>16.4622</td>
<td>3.5301</td>
<td>4.5579</td>
<td>0.6711</td>
<td>0.7826</td>
<td>12.6480</td>
<td>0.8645</td>
</tr>
<tr>
<td>NLSTF_SMBF</td>
<td>22.4454</td>
<td>27.2926</td>
<td>9.9383</td>
<td>13.5531</td>
<td>0.4817</td>
<td>0.6763</td>
<td>20.5441</td>
<td>0.7075</td>
</tr>
<tr>
<td>HYSURE</td>
<td>29.1139</td>
<td>13.4809</td>
<td>2.4710</td>
<td>2.5469</td>
<td>0.8084</td>
<td>0.8665</td>
<td>10.3741</td>
<td>0.9299</td>
</tr>
<tr>
<td>FUSE</td>
<td>27.7103</td>
<td>16.0556</td>
<td>3.0428</td>
<td>2.8151</td>
<td>0.7959</td>
<td>0.8657</td>
<td>12.5806</td>
<td>0.9403</td>
</tr>
<tr>
<td>UDALN</td>
<td>28.3250</td>
<td>13.5625</td>
<td>2.5555</td>
<td>2.7875</td>
<td>0.7535</td>
<td>0.8709</td>
<td>10.3336</td>
<td>0.8834</td>
</tr>
</tbody>
</table>

The best result is in bold, and the second best is underlined.

Figure 13 presents the fusion results of the Yongchang dataset. From Figure 13, we can see that the color error occurred in the fusion result of the CNMF, NLSTF, and NLSTF_SMBF methods. In addition, The SAM error maps of the Yongchang dataset are shown in Figure 14. We note from Figure 14 that the proposed method has the least amount of spectral distortion at the edges, which is also verified by Table 4.
Figure 13. Color composite image of fusion results of with Yongchang dataset (bands 29, 19, 9 are input to RGB channel respectively). (a) Proposed method. (b) GSA. (c) SFIM. (d) GLP. (e) CNMF. (f) NLSTF. (g) NLSTF_SMBF. (h) HYSURE. (i) FUSE. (j) UDALN.

Figure 14. SMA error map of Yongchang dataset. (a) Proposed method. (b) GSA. (c) SFIM. (d) GLP. (e) CNMF. (f) NLSTF. (g) NLSTF_SMBF. (h) HYSURE. (i) FUSE. (j) UDALN.

Table 4. Quantitative evaluation results of the Yongchang farmland area data set.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>SAM (°)</th>
<th>UIQI</th>
<th>SSIM</th>
<th>DD</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>31.8816</td>
<td>11.0828</td>
<td>1.8716</td>
<td>4.4585</td>
<td>0.4979</td>
<td>0.6417</td>
<td>7.1603</td>
<td>0.7057</td>
</tr>
<tr>
<td>GSA</td>
<td>30.6948</td>
<td>12.9507</td>
<td>2.4569</td>
<td>5.1439</td>
<td>0.4744</td>
<td>0.6385</td>
<td>8.6661</td>
<td>0.6708</td>
</tr>
<tr>
<td>SFIM</td>
<td>30.4540</td>
<td>13.0830</td>
<td>2.5290</td>
<td>5.4665</td>
<td>0.4785</td>
<td>0.6243</td>
<td>8.5827</td>
<td>0.6538</td>
</tr>
<tr>
<td>GLP</td>
<td>31.4697</td>
<td>11.7546</td>
<td>2.2168</td>
<td>5.3264</td>
<td>0.4885</td>
<td>0.6284</td>
<td>7.2407</td>
<td>0.6606</td>
</tr>
<tr>
<td>CNMF</td>
<td>26.2153</td>
<td>21.5673</td>
<td>4.4503</td>
<td>6.3250</td>
<td>0.2844</td>
<td>0.5582</td>
<td>15.7905</td>
<td>0.4523</td>
</tr>
<tr>
<td>NLSTF</td>
<td>19.0226</td>
<td>51.3424</td>
<td>23.8635</td>
<td>12.8057</td>
<td>0.1418</td>
<td>0.3850</td>
<td>41.4662</td>
<td>0.3664</td>
</tr>
<tr>
<td>NLSTF_SMBF</td>
<td>19.5615</td>
<td>48.5621</td>
<td>21.1652</td>
<td>10.6630</td>
<td>0.1434</td>
<td>0.4451</td>
<td>38.9230</td>
<td>0.3652</td>
</tr>
</tbody>
</table>
### 4.4. Ablation Experiments

Three kinds of ablation experiments are carried out to test our three primary fusion modules, i.e., SAM-CC band assignment module, ONR band selection module, and CSA adaptive PCNN parameter module. In addition, different automatic parameter optimization strategies are also investigated.

The fusion results obtained on different datasets with and without ONR band selection strategies are presented in Table 5. For the Linxia dataset, our proposed SAM-CC assignment indicator performs best. Furthermore, the SAM-CC achieves better fusion results in most cases with the Linxia dataset and the Yongchang dataset, which illustrates the effectiveness of the SAM-CC band assignment.

#### Table 5. SAM-CC ablation.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>SAM (°)</th>
<th>UIQI</th>
<th>SSIM</th>
<th>DD</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM-CC (dataset1)</td>
<td>37.6677</td>
<td>5.0560</td>
<td>1.2971</td>
<td>1.5961</td>
<td>0.7252</td>
<td>0.9296</td>
<td>2.3748</td>
<td>0.9863</td>
</tr>
<tr>
<td>SAM (dataset1)</td>
<td>37.1726</td>
<td>5.2919</td>
<td>1.4142</td>
<td>1.7293</td>
<td>0.7139</td>
<td>0.9215</td>
<td>2.5195</td>
<td>0.9847</td>
</tr>
<tr>
<td>CC (dataset1)</td>
<td>37.2604</td>
<td>5.3198</td>
<td>1.3380</td>
<td>1.6597</td>
<td>0.7113</td>
<td>0.9237</td>
<td>2.5311</td>
<td>0.9852</td>
</tr>
<tr>
<td>SAM-CC (dataset2)</td>
<td>36.2331</td>
<td>5.7254</td>
<td>0.9960</td>
<td>1.9634</td>
<td>0.8498</td>
<td>0.9067</td>
<td>3.4608</td>
<td>0.9645</td>
</tr>
<tr>
<td>SAM (dataset2)</td>
<td>36.3003</td>
<td>5.6917</td>
<td>0.9988</td>
<td>1.9721</td>
<td>0.8493</td>
<td>0.9059</td>
<td>3.4370</td>
<td>0.9645</td>
</tr>
<tr>
<td>CC (dataset2)</td>
<td>35.0948</td>
<td>6.4739</td>
<td>1.1504</td>
<td>2.0009</td>
<td>0.8383</td>
<td>0.9017</td>
<td>4.2271</td>
<td>0.9634</td>
</tr>
<tr>
<td>SAM-CC (dataset3)</td>
<td>31.8816</td>
<td>11.0828</td>
<td>1.8716</td>
<td>4.4585</td>
<td>0.4979</td>
<td>0.6417</td>
<td>7.1603</td>
<td>0.7057</td>
</tr>
<tr>
<td>SAM (dataset3)</td>
<td>28.4326</td>
<td>16.8280</td>
<td>2.2748</td>
<td>4.1073</td>
<td>0.5638</td>
<td>0.7168</td>
<td>12.6765</td>
<td>0.7755</td>
</tr>
<tr>
<td>CC (dataset3)</td>
<td>31.2056</td>
<td>12.2060</td>
<td>1.9122</td>
<td>4.8604</td>
<td>0.5026</td>
<td>0.6967</td>
<td>8.1791</td>
<td>0.7334</td>
</tr>
</tbody>
</table>

The best result is in bold, and the second best is underlined.

To validate the efficiency of the ONR band selection module, we compared the fusion results of using ONR and without ONR. In order to verify the effectiveness of ONR band selection, we fused the data sets with and without ONR band selection. Comparisons of the fusion accuracy and the running time are shown in Table 6. As can be seen from Table 6, the running time has been reduced by more than 84% with the ONR band selection module, while the fusion accuracy remains almost unchanged. Overall, this indicates significant time cost savings for the ONR band selection module.

#### Table 6. ONR band selection for ablation.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>SAM (°)</th>
<th>UIQI</th>
<th>SSIM</th>
<th>DD</th>
<th>CC</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using ONR (dataset1)</td>
<td>37.6677</td>
<td>5.0560</td>
<td>1.2971</td>
<td>1.5961</td>
<td>0.7252</td>
<td>0.9296</td>
<td>2.3748</td>
<td>0.9863</td>
<td>221.1</td>
</tr>
<tr>
<td>Without ONR (dataset1)</td>
<td>37.3746</td>
<td>5.2096</td>
<td>1.3166</td>
<td>1.6199</td>
<td>0.7143</td>
<td>0.9272</td>
<td>2.5202</td>
<td>0.9859</td>
<td>1937.6</td>
</tr>
<tr>
<td>Using ONR (dataset2)</td>
<td>36.2331</td>
<td>5.7254</td>
<td>0.9960</td>
<td>1.9634</td>
<td>0.8498</td>
<td>0.9067</td>
<td>3.4608</td>
<td>0.9645</td>
<td>132.1</td>
</tr>
<tr>
<td>Without ONR (dataset2)</td>
<td>36.3467</td>
<td>5.6366</td>
<td>0.9918</td>
<td>1.9085</td>
<td>0.8533</td>
<td>0.9082</td>
<td>3.3977</td>
<td>0.9646</td>
<td>1794.2</td>
</tr>
<tr>
<td>Using ONR (dataset3)</td>
<td>31.8816</td>
<td>11.0828</td>
<td>1.8716</td>
<td>4.4585</td>
<td>0.4979</td>
<td>0.6417</td>
<td>7.1603</td>
<td>0.7057</td>
<td>179.2</td>
</tr>
<tr>
<td>Without ONR (dataset3)</td>
<td>30.9860</td>
<td>12.2864</td>
<td>1.8764</td>
<td>4.1296</td>
<td>0.5532</td>
<td>0.7081</td>
<td>8.6389</td>
<td>0.7665</td>
<td>1219.2</td>
</tr>
</tbody>
</table>

The best result is in bold, and the second best is underlined.

In order to examine the rationality of the CSA-based parameter optimization strategy in IPCNN, Table 7 shows the comparison of the CSA fusion metrics with other three parameter optimization strategies, i.e., the sparrow search strategy (SSA) [53], the improved
grey wolf optimizer (IGWO) strategy [54] and the enhanced whale optimization strategy (EWOA) [55]. As shown in Table 7, the CSA-based fusion algorithm achieves a better fusion performance as well as less time-consuming.

Table 7. Automatic parameter optimization ablation.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>SAM (°)</th>
<th>UIQI</th>
<th>SSIM</th>
<th>DD</th>
<th>CC</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSA (dataset1)</td>
<td>37.6677</td>
<td>5.0560</td>
<td>1.2971</td>
<td>1.5961</td>
<td>0.7252</td>
<td>0.9296</td>
<td>2.3748</td>
<td>0.9863</td>
<td>221.1</td>
</tr>
<tr>
<td>SSA (dataset1)</td>
<td>37.0112</td>
<td>5.4224</td>
<td>1.4200</td>
<td>1.6315</td>
<td>0.6999</td>
<td>0.9289</td>
<td>2.8151</td>
<td>0.9860</td>
<td>267.9</td>
</tr>
<tr>
<td>IGWO (dataset1)</td>
<td>37.1760</td>
<td>5.3224</td>
<td>1.3947</td>
<td>1.6253</td>
<td>0.7051</td>
<td>0.9291</td>
<td>2.7133</td>
<td>0.9861</td>
<td>394.6</td>
</tr>
<tr>
<td>EWOA (dataset1)</td>
<td>37.3093</td>
<td>5.2504</td>
<td>1.3730</td>
<td>1.6361</td>
<td>0.7099</td>
<td>0.9291</td>
<td>2.6281</td>
<td>0.9861</td>
<td>162.6</td>
</tr>
<tr>
<td>CSA (dataset2)</td>
<td>36.2331</td>
<td>5.7254</td>
<td>0.9960</td>
<td>1.9634</td>
<td>0.8498</td>
<td>0.9067</td>
<td>3.4608</td>
<td>0.9645</td>
<td>132.1</td>
</tr>
<tr>
<td>SSA (dataset2)</td>
<td>36.2738</td>
<td>5.7156</td>
<td>1.0086</td>
<td>1.9955</td>
<td>0.8466</td>
<td>0.9056</td>
<td>3.4595</td>
<td>0.9643</td>
<td>214.3</td>
</tr>
<tr>
<td>IGWO (dataset2)</td>
<td>36.2348</td>
<td>5.7139</td>
<td>1.0024</td>
<td>1.9743</td>
<td>0.8478</td>
<td>0.9059</td>
<td>3.4537</td>
<td>0.9640</td>
<td>246.4</td>
</tr>
<tr>
<td>EWOA (dataset2)</td>
<td>36.2304</td>
<td>5.7458</td>
<td>1.0128</td>
<td>2.0058</td>
<td>0.8473</td>
<td>0.9056</td>
<td>3.4952</td>
<td>0.9644</td>
<td>122.1</td>
</tr>
<tr>
<td>CSA (dataset3)</td>
<td>31.8816</td>
<td>11.0828</td>
<td>1.8716</td>
<td>4.4585</td>
<td>0.4979</td>
<td>0.6417</td>
<td>7.1603</td>
<td>0.7057</td>
<td>179.2</td>
</tr>
<tr>
<td>SSA (dataset3)</td>
<td>30.1932</td>
<td>13.5601</td>
<td>1.9849</td>
<td>4.1381</td>
<td>0.5554</td>
<td>0.7070</td>
<td>9.7518</td>
<td>0.7679</td>
<td>199.2</td>
</tr>
<tr>
<td>IGWO (dataset3)</td>
<td>30.7599</td>
<td>12.6262</td>
<td>1.9081</td>
<td>4.1242</td>
<td>0.5539</td>
<td>0.7080</td>
<td>8.9401</td>
<td>0.7672</td>
<td>280.4</td>
</tr>
<tr>
<td>EWOA (dataset3)</td>
<td>30.3997</td>
<td>13.2174</td>
<td>1.9643</td>
<td>4.1652</td>
<td>0.5525</td>
<td>0.7043</td>
<td>9.4456</td>
<td>0.7671</td>
<td>137.6</td>
</tr>
</tbody>
</table>

The best result is in bold, and the second best is underlined.

Table 8 exhibits the quantitative analysis of the fusion results with and without adaptive CSA optimization to verify the impact of the CSA algorithm. Besides, the traditional IPCNN parameters are set as classical values, i.e., $\alpha_r = 0.1$, $\alpha_s = 1$, $\alpha_c = 0.62$, $\beta = 0.1$, $w = 0.5$. In all three datasets, the adaptive IPCNN approach is significantly superior to the traditional parameters in most cases, which indicates that the adaptive parameter optimization effectively improves the quality of the fusion images.

Table 8. CSA adaptive IPCNN parameter ablation.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>SAM (°)</th>
<th>UIQI</th>
<th>SSIM</th>
<th>DD</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive (dataset1)</td>
<td>37.6677</td>
<td>5.0560</td>
<td>1.2971</td>
<td>1.5961</td>
<td>0.7252</td>
<td>0.9296</td>
<td>2.3748</td>
<td>0.9863</td>
</tr>
<tr>
<td>Traditional (dataset1)</td>
<td>29.4710</td>
<td>13.8064</td>
<td>3.8469</td>
<td>1.6911</td>
<td>0.6480</td>
<td>0.9020</td>
<td>9.1179</td>
<td>0.9818</td>
</tr>
<tr>
<td>Adaptive (dataset2)</td>
<td>36.2331</td>
<td>5.7254</td>
<td>0.9960</td>
<td>1.9634</td>
<td>0.8498</td>
<td>0.9067</td>
<td>3.4608</td>
<td>0.9645</td>
</tr>
<tr>
<td>Traditional (dataset2)</td>
<td>34.1057</td>
<td>7.2630</td>
<td>1.2945</td>
<td>1.9978</td>
<td>0.8346</td>
<td>0.9043</td>
<td>5.0036</td>
<td>0.9629</td>
</tr>
<tr>
<td>Adaptive (dataset3)</td>
<td>31.8816</td>
<td>11.0828</td>
<td>1.8716</td>
<td>4.4585</td>
<td>0.4979</td>
<td>0.6417</td>
<td>7.1603</td>
<td>0.7057</td>
</tr>
<tr>
<td>Traditional (dataset3)</td>
<td>29.7485</td>
<td>14.3252</td>
<td>2.0709</td>
<td>4.2138</td>
<td>0.5481</td>
<td>0.6916</td>
<td>10.3473</td>
<td>0.7591</td>
</tr>
</tbody>
</table>

The best result is in bold, and the second best is underlined.

5. Conclusions

In the paper, a novel algorithm for HS sharpening is proposed. Firstly, the band groups are acquired by performing joint SAM-CC band matching of HS and MS assignment method, the band grouping results are obtained, and then ONR is applied to the HS images of each group to improve the efficiency for subsequent adaptation. Besides, according to the characteristics of remote sensing image fusion application, an IPCNN model is proposed, which can obtain irregular injection regions of spatial details. In addition, a CSA-based IPCNN parameter optimization method is designed to achieve the optimal sharpening imagery. In summary, the proposed method is simple and easy to implement, which presents good fusion results on a variety of datasets, including reservoir, mountain, town, and river landscapes. Furthermore, several ablation experiments are also conducted to corroborate the efficiency of the proposed method.
**Author Contributions:** All the authors have contributed substantially to the manuscript. X.X. and X.L. proposed the methodology. X.X., X.L., and J.G. performed the experiments and software. X.X., X.L., and Y.L. wrote the paper. Y.L. and L.K. analyzed the data. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is funded by the National Natural Science Foundation of China (No. 41861055), China Postdoctoral Science Foundation (No. 2019M653795), and LZJTU EP Program (No. 201806).

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors are grateful to the editor and anonymous reviewers for their helpful and valuable suggestions. We also want to express our sincere gratitude to Jie Li for her support and help.

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

**References**


**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.