Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105
Volume **9**, Number **2**, May **2024**

Remote Sensing Image Fusion based on Improved Super-Resolution Convolutional Neural Network

Xie Shan*

College of Information and Engineering
Chengdu Industrial Vocational and Technical College
Sichuan 600000, P. R. China
895293708@qq.com

Wang-Song Jin

Santa Clara University
New York 95053, USA
1468059565@qq.com

*Corresponding author: Xie Shan
Received February 21, 2023, revised March 24, 2023, accepted May 20, 2023.

Abstract. *By fusing multi-source remote sensing images, higher spatial resolution and richer detail information can be obtained to better serve the fields of environmental monitoring, crop estimation, and urban planning. In order to effectively improve the quality of low-resolution multispectral remote sensing images, this work proposes a remote sensing image fusion method based on improved super-resolution convolutional neural network. Firstly, the characteristics of super-resolution technique and convolutional neural network are investigated, and a novel three-layer convolutional neural network, SRCNN, is introduced. Then, the multispectral image is divided into four different channels for processing, and all the four different channel images are fed into the SRCNN for the enhancement of high-frequency detail information. The predicted multispectral and panchromatic images are sparsely represented before fusion. Secondly, the weights of SRCNN are generally initialised using two methods, namely Gaussian distribution as well as encoder assignment. However, these two algorithms have uncertainties that affect the reconstruction accuracy of the images. Therefore, the PSO algorithm is used to optimise the SRCNN weights, thus improving the resolution reconstruction accuracy. Finally, multiple sets of images from different areas of Landsat satellite data are used for simulation by both subjective and objective evaluation metrics. The experimental results show that the proposed methods all better maintain the rich information of remote sensing images and achieve better fusion results. The indicators such as source entropy, correlation coefficient, average absolute error and mean square error of the fused images are improved after the introduction of PSO algorithm.*
Keywords: remote sensing images; image fusion; convolutional neural net; low resolution images; PSO

1. **Introduction.** Remote sensing satellites are artificial satellites that allow remote sensing in outer space and continuous observation of the corresponding areas of the earth's surface on demand. Currently, remote sensing technology is widely used in various fields. Remote sensing data acquired by various sensors can be converted into visible images by simulating human visual system techniques [1,2].

The update and development of remote sensing technology bring us more and more remote sensing information data. Compared with single-source remote sensing images,

multi-source remote sensing image data can provide more data information [3,4], and there are complementary relationships and information redundancy between data. Remote sensing image data can provide more data information, and there are complementary relationships and information redundancy between data. Redundant information is the same description of the same area or carrying the same information. The system often exists for the same region of the repetition of data information, these data, although different forms of expression, but can be mapped to the data space through some kind of transformation [5,6]. The judicious use of redundant information can improve recognition rates and accuracy, and reduce the impact of noise and incomplete information. Complementary information can also be data information given by multiple sensors for the same specific area, and the information data provided by different sensors that are independent of each other have a complementary nature between the amount of data [7,8]. Currently, the development of image fusion techniques for remote sensing images has the following two main research focuses: selecting a particular region for the application of image fusion and image fusion for the image as a whole for the purpose of optimisation.

Remote sensing image fusion refers to the fusion of multiple remote sensing images from different sensors or different wavelength bands into a single image with richer and more comprehensive integrated information [9,10]. Enhancement of image details and spatial resolution: by fusing multi-source remote sensing images, higher spatial resolution and richer detail information can be obtained, which improves the quality and usability of the images. Fusing multi-scale and multi-band remote sensing images can provide more comprehensive feature information, making feature classification and target detection more accurate and reliable. Remote sensing image fusion can reduce the variability between remote sensing data acquired by different sensors and at different times, improve the spatial and temporal consistency of the data, and facilitate the analysis and comparison of the data. Remote sensing image fusion is a research hotspot that constantly drives the development of remote sensing technology [11]. The development of new fusion algorithms and methods can expand the application areas of remote sensing image processing and provide more possibilities for research in the fields of environment, agriculture and meteorology [12].

If a single sensor is used, the remote sensing image obtained may have defects such as blurred data or invisibility, but if different types of sensors are used for image fusion processing, the problem of unclear or invisibility in a single sensor can be significantly improved [13]. If the same type of sensors are used for fusion, the obtained image will get more obvious feature enhancement. Therefore, image fusion can significantly improve the readability of the image, and at the same time, it can enable a single image sensor to obtain richer and more diversified data information.

Any sensor has its advantages and disadvantages. Multi-spectral sensors may have cloud cover that prevents them from obtaining a complete image of the ground, leading to inaccurate interpretation of shadows in the ground image. Therefore, the fusion of image information from multiple types of sensors to a specific region can complement the lack of information from a single sensor to improve the completeness of the information.

The research significance of remote sensing image fusion is mainly in terms of improving image quality and comprehensiveness of information, improving feature classification and target detection, supporting remote sensing monitoring and resource management, enhancing spatial and temporal consistency of data, and promoting the development of remote sensing technology.

1.1. **Related Work.** Currently, the remote sensing image data used for fusion mainly come from different types of sensors. In the application of remote sensing image fusion,

low resolution multispectral images and high spatial resolution panchromatic images are used, which currently have the following drawbacks [14]: high-resolution images have more image details but poor spectral information, while multispectral images have sufficient spectral information but lower resolution. If we want to get both good spectral quality and high resolution, we need to improve the image fusion algorithm.

Image fusion technology is an emerging means of integrated information data processing in recent years, originally developed in the field of computer vision. Image fusion as an automatic information data integration processing means, can make full use of information from different data sources to complement each other and take advantage of the rapid processing speed of the computer. Through the fusion algorithm processing power and computer processing performance to improve the image quality to get more rich and multi-faceted data information, improve the reliability and stability of information. At present, image fusion methods are mainly divided into three categories:

(1) Spatial transform based image fusion methods. These methods are based on mathematical transforms or representations of transform domains, such as wavelet transforms, multiresolution analysis, etc. By transforming multiple images and fusing their low-frequency and high-frequency information appropriately, a clearer and more comprehensive image can be obtained. Common methods include wavelet transform fusion, singular value decomposition fusion, etc. Wan et al. [15] proposed a remote sensing image fusion method based on bootstrap filter and non-downsampled shear transform. By introducing a bootstrap filter to preserve image details and combining the non-downsampled shear transform for feature extraction, a high-quality remote sensing image fusion method is achieved. Wu et al. [16] proposed a remote sensing image fusion method based on the curvilinear wavelet transform and principal component analysis. By applying the curvilinear transform for decomposition and reconstruction, and combining with principal component analysis for feature fusion, the enhancement and information extraction of remote sensing images are achieved. However, after the spatial transformation, the edge parts in the image may have overlapping or artefact problems. This is due to the fact that the position of pixels may change during the transformation process, resulting in the loss of information or aliasing in the edge region.

(2) Image fusion methods based on multiscale analysis. Yan and Li [17] provide a comprehensive review of methods based on multiscale analysis in image fusion, including strategies at three levels. For each level, the article introduces common multiscale analysis algorithms and analyses their advantages and disadvantages. In addition, the applicability of these methods in different application scenarios is discussed and future research directions are proposed. However, in the process of multiscale analysis, image information of different scales is extracted and fused, but some unavoidable information loss or artefacts may be introduced in the fusion process. This may lead to the quality degradation of the final fused image.

(3) Neural network based image fusion method. Sun et al. [18] proposed an exposure fusion method based on deep unsupervised learning. By using extreme exposure image pairs, the authors proposed a self-encoder network to learn a mapping for extracting useful information from low- and high-exposure images, which produces better results in the exposure fusion task. Sun et al. [19] introduced a Generative Adversarial Network (GAN)-based fusion method for infrared and visible images. The authors proposed a multilevel GAN structure for generating fused images and introduced eye movement data for training. Experiments demonstrate that the method produces clearer and more visually perceptible results in infrared and visible light image fusion tasks.

Image super-resolution technology is also known as super-resolution image reconstruction. For the existing low-resolution images, the software algorithm can be used to obtain

a better high-resolution image. Convolutional neural network has the characteristics of displacement scaling and distortion invariance, which can better retain the spatial scale information of the image. Remote sensing image fusion, on the other hand, requires the extraction of features from multiple sources of image data, and then integrates a more comprehensive, accurate, and reliable information description of a scene or target.

In 2015, Dong et al. [20] proposed a Super Resolution Convolutional Neural Networks (SRCNN) based on deep learning framework. The SRCNN model is divided into three levels according to the three steps of image extraction and mapping as well as high-resolution image reconstruction. The SRCNN model can directly obtain the mapping criterion between the high and low of the image, which exhibits high image super-resolution performance.

1.2. **Motivation and contribution.** From the analysis above, it is clear that combining panchromatic and low-resolution multispectral images can preserve both the spatial resolution and spectral information of the panchromatic images while producing high-resolution multispectral images.

Therefore, in order to effectively improve the quality of low-resolution multispectral remote sensing images, this work adopts SRCNN convolutional neural network with certain prediction function under image super-resolution technique to process multispectral images and panchromatic images. The model of SRCNN convolutional neural network and the algorithmic steps for remote sensing image fusion are described in detail. Experiments are conducted on the proposed method and the corresponding experimental results are given for comparison.

The main innovations and contributions of this work include:

(1) Image super-resolution processing of multispectral images is performed using SRCNN model, and the up-sampled multispectral images are divided into four channels to be fed into SRCNN model respectively.

(2) The weights of SRCNN are generally initialised using two methods, i.e., Gaussian distribution and encoder assignment, which are uncertain and have an impact on the reconstruction accuracy of the image, but Particle Swarm Optimisation (PSO) algorithm has a better convergence, which can be used to set up a better initial parameter of the network, so in this paper, we use the PSO algorithm to optimise the initial weights of the SRCNN, which in turn achieve the improve the resolution reconstruction accuracy.

(3) The images of the four channels obtained after processing and the panchromatic image are sparsely represented with the overcomplete DCT dictionary. In accordance with the sparse coefficient absolute value of the larger of the multi-spectral image of each channel of the image and panchromatic image fusion, the fusion of the results of the reconstruction to obtain the final fusion image.

2. **Remote sensing image pre-processing.**

2.1. **Radiometric calibration.** Sensor measurements do not coincide with physical quantities such as the spectral reflectance or spectral radiant brightness of the target, so there is a discrepancy between the visual measurements of the target and physical quantities such as the spectral reflectance or spectral radiant brightness of the target [21,22].

The purpose of radiometric correction is to eliminate various distortions in the image data, thus radiometric correction lays the foundation for quantitative remote sensing inversion. In quantitative remote sensing analyses, values such as radiance, reflectance and

temperature are commonly used. These physical qualities are acquired by radiometric calibration, which is the transformation of voltage or digitally quantized quantities acquired by sensors into absolute radiance.

Radiometric calibrations are divided into absolute and relative calibrations. Absolute calibration is a quantitative relationship between the radiance value and the dimensionless DN value. For general sensors, absolute calibration requires a linear relationship to complete the conversion of the DN value to the radiant brightness value [23].

$$L = G \cdot DN + offset \tag{1}$$

where G denotes the gain value and offset denotes the offset.

The newest of the U.S. Landsat satellite datasets is the Landsat 9 satellite [24,25].The sensor types used in the Landsat 9 dataset are the Operational Land Imager 2 (OLI-2) and the Thermal Infrared Sensor 2 (TIRS-2).The bands of the Landsat 9 situation is shown in Table 1. In order to take advantage of the rich spectral information in the Landsat 9 dataset, radiometric calibration is required. The DN values on the raw images are converted to reflectance.

Table 1. Banding of Landsat 9

Band name	**Spectral range**
Coastal/Aerosol	0.433-0.453 μm
Blue	0.450-0.515 μm
Green	25-0.600 μm
Red	0.630-0.680 μm
Near Infrared (NIR)	0.845-0.885 μm
Short-Wave Infrared 1 (SWIR 1)	1.560-1.660 μm
Short-Wave Infrared 2 (SWIR 2)	2.100-2.300 μm
Panchromatic	0.500-0.680 μm
Cirrus	1.360-1.390 μm
Thermal Infrared (TIRS) 1	10.60-11.19 μm
Thermal Infrared (TIRS) 2	11.50-12.51 μm

2.2. **Geometric Correction of Remote Sensing Images.** The geometric distortion of remote sensing image data is related to the height, flight speed and rotation of the sensor, etc., and geometric correction is a common method to correct the geometric distortion of remote sensing images [26]. Before the image is geometrically corrected, we first find some typical features on the image as control points on the image, and then geometrically transform the selected control points using certain mathematical transformation relations. Among them, the selected control points need to be well distributed. In this work, polynomial correction method is used to perform geometric correction.

Firstly, the relationship between the coordinates is established. Let (x_i, y_i) and (x'_i, y'_i) be the coordinates of the corresponding pixel points of the remote sensing image before and after geometric correction, respectively.

$$\begin{cases} x_i = f_x(x'_i, y'_i) \\ y_i = f_y(x'_i, y'_i) \end{cases} \tag{2}$$

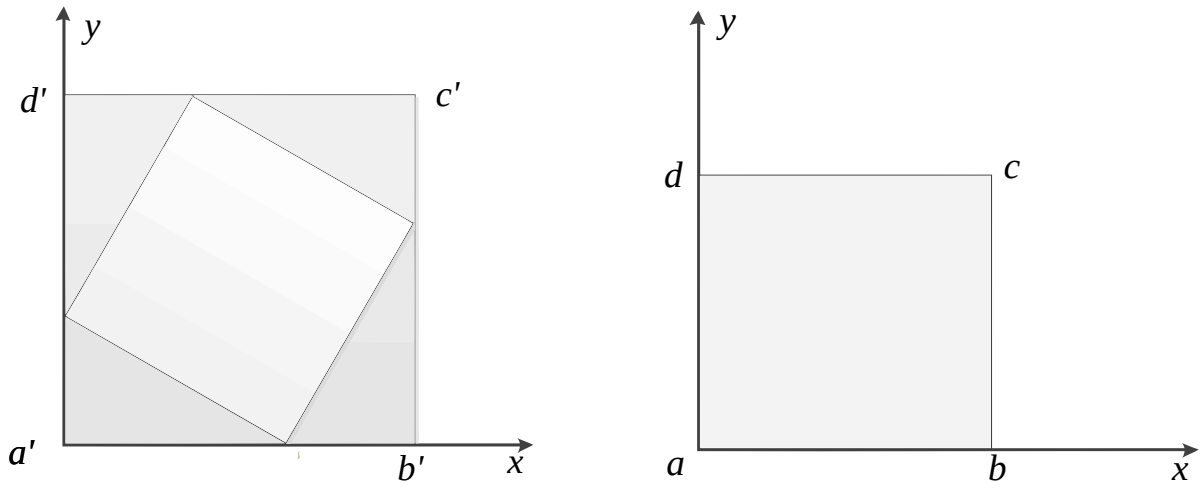
$$\begin{cases} x'_i = F_x(x_i, y_i) \\ y'_i = F_y(x_i, y_i) \end{cases} \tag{3}$$

where both the function *f()* and the function *F()* denote binary n-th degree polynomials

Equation (2) is the indirect method while Equation (3) is the direct method. The indirect correction method is to map each pixel point in the new image by the transform function *f()* to find out the position of that pixel point in the original image and assign the grey value of this pixel point to the corresponding pixel point in the new image [27,28]. The direct correction method is to obtain the position of the corresponding pixel point in the new image for each pixel point in the original image by the mapping relation of function *F()* and assign the grey value of this pixel point in the original image to the corresponding pixel point in the new image.

Then, the boundary of the image needs to be determined. We use the method of corresponding vertices to determine the boundary. Let the four vertices of the original image be denoted as *abcd*, then the four vertices of the corrected image be denoted as $a'b'c'd'$. The schematic diagram of geometric correction of remote sensing image is shown in Figure 1.

$$\begin{cases} x_1' = \max\left(x'a', x'b', x'c', x'd'\right) \\ x_2' = \min\left(x'a', x'b', x'c', x'd'\right) \\ y_1' = \max\left(y'a', y'b', y'c', y'd'\right) \\ y_2' = \min\left(y'a', y'b', y'c', y'd'\right) \end{cases} \tag{4}$$



(a) Image before geometric correction (b) Image after geometric correction

Figure 1. Schematic diagram of geometric correction of remote sensing image

Finally, the width and height of the new image needs to be determined in order for the boundary range of the image result to be transformed to the range required for accuracy. The width and height of the original image and the centre coordinate point are first determined, and then the width and height and centre point of the new image are adjusted.

$$\begin{cases} row = \frac{y_1' - y_2'}{\Delta y} + 1 \\ col = \frac{x_1' - x_2'}{\Delta x} + 1 \end{cases} \tag{5}$$

Where *row* denotes the number of rows of the new image, *col* denotes the number of columns of the new image, and Δx denotes the precision factor.

3. **Improved SRCNN.**

3.1. **SRCNN model and its principles.** The SRCNN model takes the relationship between deep learning and traditional sparse coding as a basis for super-resolution reconstruction of images using a three-layer network. The entire network structure can be viewed as three layers of feature extraction, nonlinear mapping, and high-resolution image reconstruction.The structure of the SRCNN model is shown in Figure 2.

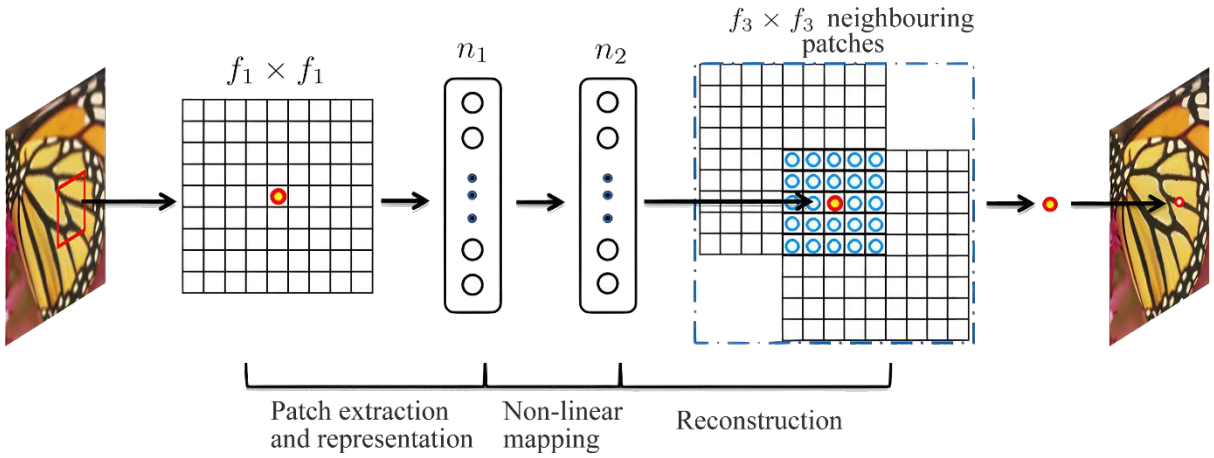


Figure 2. SRCNN model structure

Layer 1 network is used to extract and characterise the image blocks. Layer 1 network can be represented as:

$$F_1(Y) = \max(0, W_1 * Y + B_1) \tag{6}$$

where W_1 denotes the filter, B_1 denotes the bias, and $*$ denotes the convolution operation.

W_1 contains n_1 filters of size $c \times f_1 \times f_1$, c is the number of channels contained in the image, and f_1 is the filter null field size.

Layer 1 network extracts the n_1 dimensional features of the image block while layer 2 network maps this n_1 dimensional feature vector to n_2 dimensional feature vector. Layer 2 network can be represented as:

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \tag{7}$$

where W_2 contains n_2 filters of size $n_1 \times f_2 \times f_2$ and B_2 denotes the n_2-th dimensional vector.

Conventional approaches are usually predefined fusion methods, e.g., dealing with overlapping regions by arithmetic averaging. However, SRCNN merges overlapping image blocks using learning methods. A three-layer network is mainly used to complete the process of image reconstruction. The third layer can be represented as:

$$F(Y) = W_3 * F_2(Y) + B_3 \tag{8}$$

where W_3 contains c filters of size $n_2 \times f_3 \times f_3$ and B_3 denotes a c-th dimensional vector.

SRCNN uses the mean square error as the loss function.

$$L(W, h) = 1/n \sum_{i=1}^{n} \left\| F(Y_i; W, b) - X_i \right\|^2 \tag{9}$$

where n denotes the number of training samples, W denotes the weights, b is the bias term, X is the real image and $F(Y)$ is the mapped image.

The minimum mean square error between the real image and the fused image results in a high quality reconstructed image. SRCNN uses the standard gradient descent method

to minimise the randomness of the loss function. Compared with traditional learning algorithms, the SRCNN algorithm is able to automatically learn the hierarchical features, which facilitates the reconstruction of the image to extract the features, and enriches the edges of the reconstructed image with a good visual perception.

3.2. **PSO-SRCNN model.** The most commonly used methods of assigning or initialising the weights of Gaussian distributions are used to initialise the weights of the SRCNN network, but they have a certain degree of unqualifiedness that weakens the reconstruction accuracy. If PSO is used to initialise the parameters, it may be possible to exploit its search capability and convergence speed to achieve optimisation of the SRCNN network. Therefore, in this work, the PSO algorithm is used to initialise the SRCNN weights.

As a heuristic global optimisation algorithm, the PSO algorithm can simulate the foraging behaviour of birds. The PSO algorithm is often used to find the optimal solution during the training process of neural networks. Assuming that there are S particles in M-dimensional space, the spatial position of the k-th particle can be expressed as $x_k(x_{k1}, x_{k2}, ..., x_{km})$, the flight speed as $V_k = (V_{k1}, V_{k2}, ..., V_{kM})$, the best position in the history of the flight route as $P_k = (P_{k1}, P_{k2}, \ldots, P_{kM})$, and the best position of the population as $P_g = (P_{g1}, P_{g2}, \ldots, P_{gM})$. For each generation the velocity and position of the m-th dimensional $(1 < m < M)$ particle are updated.

$$V_{km}(l+1) = \omega V_{km}(t) + a_1 r_1[p_{km} - x_{km}(t)] + a_2 r_2[p_{gm} - x_{km}(t)] \tag{10}$$

$$x_{km}(m+1) = x_{km}(t) + V_{km}(t+1) \tag{11}$$

Where, t is the number of iterations, ω is the inertia weight, a is a constant called acceleration factor, and r is a constant between $[0, 1]$.

The flow of the proposed PSO-SRCNN model is as follows:

(1) The individual real images are cropped into sub-blocks of 32×32 size in the pre-training stage and these sub-blocks are taken as input images. Afterwards, Gaussian blurring is applied to the sampled images and combined with bicubic interpolation to select the appropriate factor for enlargement, thus obtaining low-resolution image samples;

(2) Subblocks of training images of SRCNN are used as input values. weights and thresholds of SRCNN are used as an initial population of PSO and the best fitness individual is selected as the initial value of SRCNN;

(3) According to the basic idea of BP network, the gradient descent method is used to optimise the parameters, and the residuals are used as a priori knowledge to find out the optimal weights and then execute the PSO framework for iterative optimisation. The best fitness of each particle is calculated, while the optimal solution in the population is selected;

(4) Repeat the above steps until the global best solution is found, thus obtaining the optimal weights and thresholds.

4. **Remote sensing image fusion based on PSO-SRCNN.**

4.1. **Multi-channel sparse representation.** In this work, the PSO-SRCNN model is used to process low-resolution multispectral images.

The four channel images of the up-sampled low-resolution multispectral image are fed into the convolutional neural network separately to obtain the enhanced images of the corresponding four channels. The corresponding multispectral images of these four channels are sparsely represented with the panchromatic image. Here, the DCT dictionary is used, which has a better representation of the texture information of the image. The sparse coefficients of the corresponding image blocks are derived using the OMP algorithm. Then the larger coefficients are selected as the sparse coefficients of the fused image block. The

principle of remote sensing image fusion method based on PSO-SRCNN is shown in Figure 3. Since the size ratio of the low-resolution multispectral image to the high-resolution

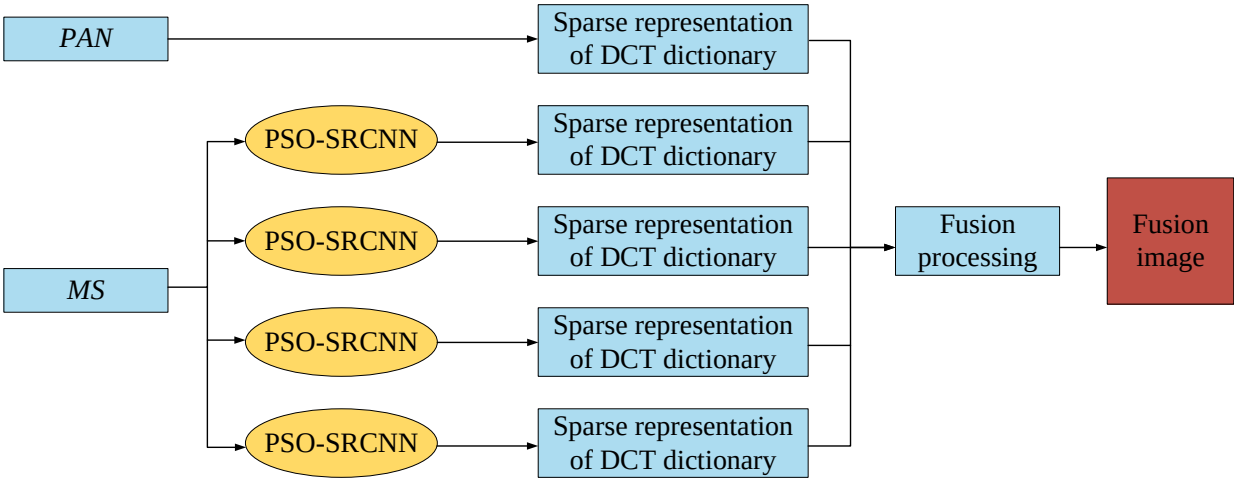
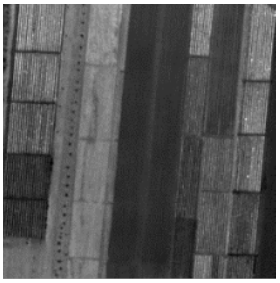


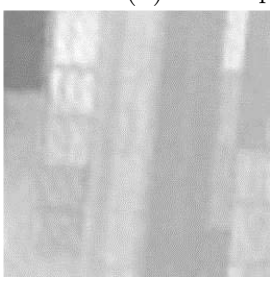
Figure 3. PSO-SRCNN-based remote sensing image fusion approach

panchromatic image of the data acquired by the Landsat 9 satellite sensor is 4, the first step is to up-sample the low-resolution multispectral image and feed the up-sampled image into the PSO-SRCNN neural network. The four-channel image after processing by PSO-SRCNN network is given in Figure 4.

(a) Multi-spectral images

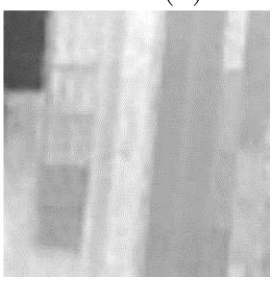


(b) Panchromatic images



(c) Channel 1

(d) Channel 2



(e) Channel 3

(f) Channel 4

Figure 4. Example of output image of PSO-SRCNN

The images of each of the four channels enhanced by the PSO-SRCNN model are fed into the overcomplete DCT dictionary for sparse representation and the corresponding sparse coefficients are obtained. Each small image block can be represented as the product of the overcomplete dictionary and the sparse coefficients.

$$\begin{cases} X'_{PAN} \approx D\alpha_{PAN} \\ X'_{MS} \approx D\alpha_{MS} \end{cases} \tag{12}$$

where α denotes sparse coefficients, D denotes DCT dictionary, X'_{PAN} denotes a small piece of panchromatic image corresponding to a particular channel, and X'_{MS} denotes a small piece of multispectral image corresponding to a particular channel.

After the image has been represented by the DCT dictionary, the item with the larger absolute value of the sparsity coefficient is selected as the corresponding fused image block X'_{Ft}.

$$X'_{Ft} = \begin{cases} D\alpha_{PAN} & \alpha_{PAN} > \alpha_{MS} \\ D\alpha_{MS} & \alpha_{PAN} < \alpha_{MS} \end{cases} \tag{13}$$

Then the whole image corresponding to the image block can be represented as:

$$X_{Ft} = \sum X'_{Ft} \tag{14}$$

The weighting method is adopted to fuse the images after sparse representation. $\mathrm{img}_{\mathrm{MS}}$ denotes the multispectral image output from the PSO-SRCNN model with a single channel, while $\mathrm{img}_{\mathrm{PAN}}$ denotes the panchromatic image, I denotes the unit matrix, α denotes the weighting coefficients, and X_F denotes the fused image under a particular channel. Before the image is subjected to DCT dictionary representation for solving the sparse coefficients, a unitary operation is performed on the image, and the obtained X_F is a block representation of the image under unitary, so it needs to be de-unitised while fusion.

$$X_F = a_1 \left(a_3 img_{PAN} + a_4 X_{Ft}\right) / \left(1 + a_4 I\right) + a_2 \left(a_3 img_{MS} + a_4 X_{Ft}\right) / \left(1 + a_4 I\right) \tag{15}$$

The fused image X_F obtained under the four channels is reconstructed to obtain the final fused image F.

4.2. **Over complete DCT dictionary.** In this work, the DCT dictionary size is set to 64×256, the redundancy factor is 4, the number of dictionary atoms is 256, the window sliding step is 1, and the window size is 8×8.

For a given sequence number $x(n)$, the DCT dictionary can be expressed as:

$$X_c(0) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \tag{16}$$

The corresponding matrix is of the form:

$$X_c = C_N x \tag{17}$$

where C_N is an N × N matrix whose row vectors are cosine based.

5. Experimental results and analyses.

5.1. **Experimental environment and experimental data set.** The experimental hardware environment is: Intel Core i5 2.2GHz processor, 6G RAM, 400G hard drive, GTX1060 discrete graphics card. The experimental software environment is: Windows 7 operating system, Matlab 2012 (R2012a) simulation software.

The simulation experiment selects Landsat 9 satellite data. The acquired satellite data is cropped in the experiment, and the block size is 64×64. In this experiment, the PSO-SRCNN model is trained on caffe, and the simulation test is carried out on the Matlab platform.The super-resolution factor is 2 in the SRCNN model, and the network structure includes 3 convolutional layers and 1 anti-convolutional layer, the learning rate is 0.001, the number of iterations is 10,000, and the loss function is the mean square error (MSE).The PSO algorithm has a particle number of 50, a dimension of 2, inertia weights of 0.6, and a maximum iterations number of 500. In order to validate the state-of-the-art of the proposed remote sensing image fusion method, compared with the wavelet transform [29], CNN [30], and the SRCNN [20] are compared.

5.2. **Image fusion quality evaluation metrics.** Subjective qualitative evaluation is also known as visual evaluation method. A direct observer evaluates the subjective effect of the fused image under strictly controlled conditions. This method is based on the subjective perception of a person and his or her own expertise to subjectively evaluate the quality of an image. Different people may evaluate the same image differently and the results are more subjective. Therefore, the international grading scale is divided into five levels [31], as shown in Table 2.

Table 2. Five Levels of Image Quality.

Mark	Visual criteria	Quality criteria
5	Image fusion at a glance	rare
4	Slight observation of differences In image fusion effects	(of an unmarried couple) be close
3	Image fusion effect is obvious	usual
2	Presence of visual obstruction	differ from
1	Unobservable	extremely poor

Objective quantitative evaluation of image fusion results is to evaluate the fusion effect using some quantitative indicators. Currently, there are many commonly used objective quantitative evaluation indexes [32]. In this work, mean square error, mean absolute error, correlation coefficient, and source entropy are chosen as the result quality evaluation indexes.

(1) Mean square error(MSE). The smaller the mean square error, the closer the fused image is to the reference image and the better the visual effect.

$$MSE = \frac{1}{M \times N} \sum_{i=0}^{M} \sum_{j=0}^{N} [I_F(i,j) - I_R(i,j)]^2 \tag{18}$$

where M is the number of rows of the image, N is the number of columns of the image, $I_F(i,j)$ is the grey value of the i-th row and j-th column in the standard reference image F, and $I_R(i,j)$ is the grey value of the i-th row and j-th column in the fusion image R.

(2) Mean absolute error (MAE).

$$MAE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |I_F(i,j) - I_R(i,j)| \tag{19}$$

(3) Correlation coefficient (CC).

$$CC(A,B) = \frac{\sum_{i=0}^{M} \sum_{j=0}^{N} \left[I_F(i,j) - \overline{I_F}\right] \times \left[I_R(i,j) - \overline{I_R}\right]}{\sqrt{\sum_{i=0}^{M} \sum_{j=0}^{N} \left[I_F(i,j) - \overline{I_F}\right]^2 \times \sum_{i=0}^{M} \sum_{j=0}^{N} \left[I_R(i,j) - \overline{I_R}\right]^2}} \tag{20}$$

where $\bar{I}_F$ and $\bar{I}_R$ are the mean values of the standard reference image F and the fused image R, respectively.

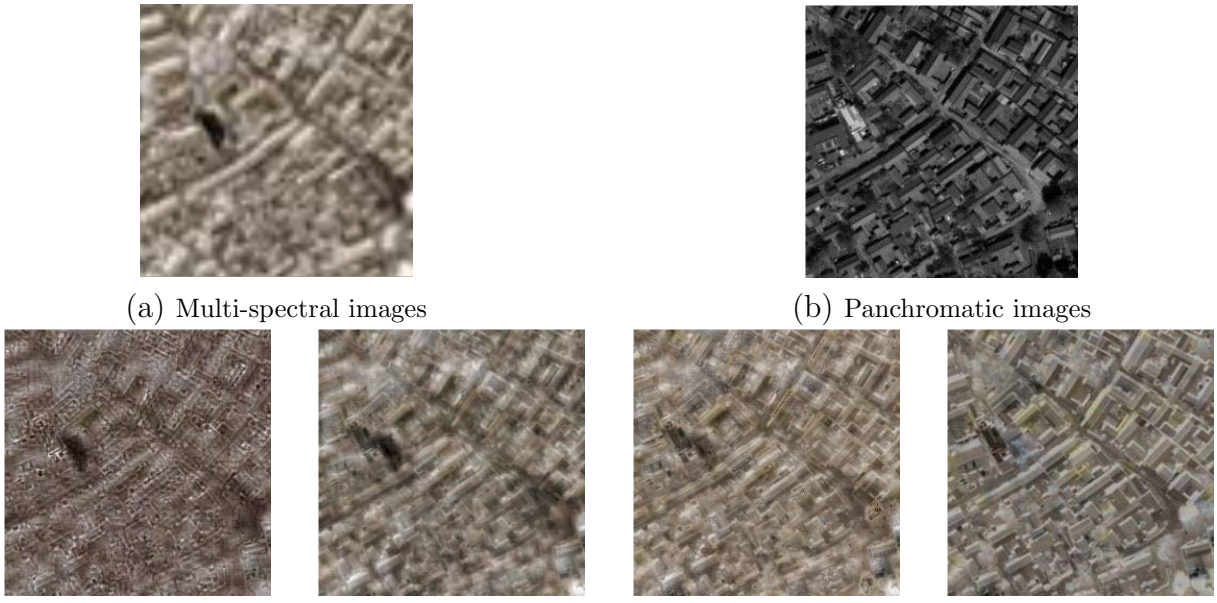
(4) Source Entropy. Improving the amount of visual information is one of the main goals of image fusion. An essential indication of the richness of visual information is the signal source entropy. The amount of information in the fused image increases with source

entropy value. The following is the definition of signal source entropy:

$$E = -\sum_{i=0}^{L-1} P_i \log_2(P_i) \tag{21}$$

where E is the source entropy of the image, L is the total number of grey levels of the image and P is the probability of occurrence of a pixel with grey value 1.

5.3. **Image fusion results analysis.** In this experiment, the PSO-SRCNN model is trained on Caffe and tested by simulation on Matlab platform. In order to verify the advancement of the proposed remote sensing image fusion method, it is compared with wavelet transform, CNN and SRCNN. The remote sensing image fusion simulation results are shown in Figure 5.



(a) Multi-spectral images (b) Panchromatic images

(c) Wavelet transform (d) CNN (e) SRCNN (f) PSO-SRCNN

Figure 5. Remote sensing image fusion simulation results

From the experimental results, it can be seen that the fused image of the wavelet transform method is darker, and the spectral distortion is more serious; although the CNN method can effectively improve the spatial resolution of the fused image, the spectral information is seriously impaired as in the case of the wavelet transform method; compared with the wavelet transform method and the CNN method, the SRCNN method achieves a better result, better maintains the spectral information, but the clarity is is not high enough; compared to other comparison methods, PSO-SRCNN method is slightly inferior to wavelet transform method in maintaining spatial resolution information, but it achieves better results in maintaining spectral information, and overall, PSO-SRCNN achieves better results.

Table 3 shows the numerical metrics after fusion of ten image pairs. Compared with the other three methods, the four objective evaluation indexes obtained from the fusion results of PSO-SRCNN are higher, so the fusion effect of PSO-SRCNN model is better compared with the original SRCNN model, which verifies its effectiveness and applicability.

Table 3. Accuracy analysis of experimental results

	Source Entropy	**CC**	**MAE**	**MSE**
Wavelet transform	1.0895	0.5681	0.8891	0.3215
CNN	1.1239	0.6408	0.8627	0.3004
SRCNN	1.2375	0.7981	0.8209	0.2967
PSO-SRCNN	1.4579	0.8754	0.7556	0.2375

6. **Conclusion.** In this work, a PSO-SRCNN model is proposed for processing the fusion of low-resolution multispectral images and panchromatic images. Firstly, the multispectral image is divided into four different channels for processing, and all four different channel images are fed into a convolutional neural network for the enhancement of high frequency detail information. Then, the predicted multispectral and panchromatic images are sparsely represented before fusion. A remote sensing image fusion method based on PSO-SRCNN is proposed by adding PSO algorithm to the original SRCNN model. Finally, several sets of images from Landsat 9 satellite data were tested by both subjective and objective evaluation indexes. The results show that all the proposed methods better maintain the rich information of remote sensing images and achieve better fusion results. The use of sparse algorithms leads to slower computing speed. Therefore, the direction of subsequent research is how to solve the problem of large data volume and time-consuming in the recovery process of the sparse algorithm under the premise of ensuring the performance of the algorithm, so as to improve the calculation speed.

REFERENCES

[1] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, and J. Wang, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sensing of Environment*, vol. 241, 111716, 2020.

[2] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121–2133, 2022.

[3] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10, 2339, 2023.

[4] T.-Y. Wu, H. Li, and S.-C. Chu, "CPPE: An Improved Phasmatodea Population Evolution Algorithm with Chaotic Maps," *Mathematics*, vol. 11, no. 9, 1977, 2023.

[5] J. Zhang, Y. Huang, R. Pu, P. Gonzalez-Moreno, L. Yuan, K. Wu, and W. Huang, "Monitoring plant diseases and pests through remote sensing technology: A review," *Computers and Electronics in Agriculture*, vol. 165, 104943, 2019.

[6] H. S. Munawar, A. W. Hammad, and S. T. Waller, "Remote sensing methods for flood prediction: A review," *Sensors*, vol. 22, no. 3, 960, 2022.

[7] A. Asokan, and J. Anitha, "Change detection techniques for remote sensing applications: A survey," *Earth Science Informatics*, vol. 12, pp. 143–160, 2019.

[8] M. D. Hossain, and D. Chen, "Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 115–134, 2019.

[9] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Systems with Applications*, vol. 169, 114417, 2021.

[10] E. K. Wang, C.-M. Chen, M. M. Hassan, and A. Almogren, "A deep learning based medical image segmentation technique in Internet-of-Medical-Things domain," *Future Generation Computer Systems*, vol. 108, pp. 135–144, 2020.

[11] K.-K. Tseng, J. Lin, C.-M. Chen, and M. M. Hassan, "A fast instance segmentation with one-stage multi-task deep neural network for autonomous driving," *Computers & Electrical Engineering*, vol. 93, 107194, 2021.

[12] Y. Li, J. Ma, and Y. Zhang, "Image retrieval from remote sensing big data: A survey," *Information Fusion*, vol. 67, pp. 94–115, 2021.
[13] Q. Xie, J. Dash, A. Huete, A. Jiang, G. Yin, Y. Ding, D. Peng, C. C. Hall, L. Brown, and Y. Shi, "Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 80, pp. 187–195, 2019.
[14] C. Kallimani, R. Heidarian, F. K. van Evert, B. Rijk, and L. Kooistra, "UAV-based Multispectral & Thermal dataset for exploring the diurnal variability, radiometric & geometric accuracy for precision agriculture," *Open Data Journal for Agricultural Research*, vol. 6, pp. 1–7, 2020.
[15] W. Wan, Y. Yang, and H. J. Lee, "Practical remote sensing image fusion method based on guided filter and improved SML in the NSST domain," *Signal, Image and Video Processing*, vol. 12, pp. 959–966, 2018.
[16] Z. Wu, Y. Huang, and K. Zhang, "Remote sensing image fusion method based on PCA and curvelet transform," *Journal of the Indian Society of Remote Sensing*, vol. 46, pp. 687–695, 2018.
[17] H. Yan, and Z. Li, "Infrared and visual image fusion based on multi-scale feature decomposition," *Optik*, vol. 203, 163900, 2020.
[18] C. Sun, K. Song, J. Su, Y. Yan, and T. Zhang, "A multi-exposure fusion method for reflection suppression of curved workpieces," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 1–4, 2022.
[19] C. Sun, C. Zhang, and N. Xiong, "Infrared and visible image fusion techniques based on deep learning: A review," *Electronics*, vol. 9, no. 12, 2162, 2020.
[20] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
[21] X. Zhang, F. Zhang, Y. Qi, L. Deng, X. Wang, and S. Yang, "New research methods for vegetation information extraction based on visible light remote sensing images from an unmanned aerial vehicle (UAV)," *International Journal of Applied Earth Observation and Geoinformation*, vol. 78, pp. 215–226, 2019.
[22] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, "Classification of remote sensing images using EfficientNet-B3 CNN model with attention," *IEEE Access*, vol. 9, pp. 14078–14094, 2021.
[23] L. P. Osco, J. M. Junior, A. P. M. Ramos, L. A. de Castro Jorge, S. N. Fatholahi, J. de Andrade Silva, E. T. Matsubara, H. Pistori, W. N. Gonçalves, and J. Li, "A review on deep learning in UAV remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, 102456, 2021.
[24] X. Li, B. Liu, G. Zheng, Y. Ren, S. Zhang, Y. Liu, L. Gao, Y. Liu, B. Zhang, and F. Wang, "Deep-learning-based information mining from ocean remote-sensing imagery," *National Science Review*, vol. 7, no. 10, pp. 1584–1605, 2020.
[25] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 14–34, 2021.
[26] W. Li, and C.-Y. Hsu, "Automated terrain feature identification from remote sensing imagery: A deep learning approach," *International Journal of Geographical Information Science*, vol. 34, no. 4, pp. 637–660, 2020.
[27] C. Zhang, P. Yue, D. Tapete, L. Jiang, B. Shangguan, L. Huang, and G. Liu, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 183–200, 2020.
[28] Y. Gu, Y. Wang, and Y. Li, "A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection," *Applied Sciences*, vol. 9, no. 10, 2110, 2019.
[29] W. Ma, Z. Pan, J. Guo, and B. Lei, "Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3512–3527, 2019.
[30] W. Zhou, D. Ming, X. Lv, K. Zhou, H. Bao, and Z. Hong, "SO–CNN based urban functional zone fine division with VHR remote sensing image," *Remote Sensing of Environment*, vol. 236, 111458, 2020.
[31] X. Nie, M. Duan, H. Ding, B. Hu, and E. K. Wong, "Attention mask R-CNN for ship detection and segmentation from remote sensing images," *IEEE Access*, vol. 8, pp. 9325–9334, 2020.
[32] D. Konstantinidis, V. Argyriou, T. Stathaki, and N. Grammalidis, "A modular CNN-based building detector for remote sensing images," *Computer Networks*, vol. 168, 107034, 2020.