Research Article

Superresolution Reconstruction of Remote Sensing Image Based on Middle-Level Supervised Convolutional Neural Network

Xiu Zhang

School of Information Engineering, Shaanxi Xueqian Normal University, Xi'an, Shaanxi, China

Correspondence should be addressed to Xiu Zhang; zhangxiu@st.btbu.edu.cn

Received 17 September 2021; Revised 29 October 2021; Accepted 18 November 2021; Published 4 January 2022

Academic Editor: Mu Zhou

Copyright © 2022 Xiu Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image has become one of the important carriers of visual information because of its large amount of information, easy to spread and store, and strong sense of sense. At the same time, the quality of image is also related to the completeness and accuracy of information transmission. This research mainly discusses the superresolution reconstruction of remote sensing images based on the middle layer supervised convolutional neural network. This paper designs a convolutional neural network with middle layer supervision. There are 16 layers in total, and the seventh layer is designed as an intermediate supervision layer. At present, there are many researches on traditional superresolution reconstruction algorithms and convolutional neural networks, but there are few researches that combine the two together. Convolutional neural network can obtain the high-frequency features of the image and strengthen the detailed information; so, it is necessary to study its application in image reconstruction. This article will separately describe the current research status of image superresolution reconstruction and convolutional neural networks. The middle supervision layer defines the error function of the supervision layer, which is used to optimize the error back propagation mechanism of the convolutional neural network to improve the disappearance of the gradient of the deep convolutional neural network. The algorithm training is mainly divided into four stages: the original remote sensing image preprocessing, the remote sensing image temporal feature extraction stage, the remote sensing image spatial feature extraction stage, and the remote sensing image reconstruction output layer. The last layer of the network draws on the single-frame remote sensing image SRCNN algorithm. The output layer overlaps and adds the remote sensing images of the previous layer, averages the overlapped blocks, eliminates the block effect, and finally obtains high-resolution remote sensing images, which is also equivalent to filter operation. In order to allow users to compare the superresolution effect of remote sensing images more clearly, this paper uses the Qt5 interface library to implement the user interface of the remote sensing image superresolution software platform and uses the intermediate layer convolutional neural network and the remote sensing image superresolution reconstruction algorithm proposed in this paper. When the training epoch reaches 35 times, the network has converged. At this time, the loss function converges to 0.017, and the cumulative time is about 8 hours. This research helps to improve the visual effects of remote sensing images.

1. Introduction

Superresolution reconstruction (superresolution, SR) is the process of obtaining the highest quality image from one or more low-resolution images through signal processing and image processing methods. After the low-resolution (LR) small image is enlarged by interpolation (convolved with the interpolation function), it is enlarged to the required size and then reconstructed by the reconstruction algorithm. It can be seen that the selection of the interpolation function and the reconstruction algorithm determine the quality

and efficiency of the reconstruction. Because remote sensing images are acquired by satellites at high altitude, ground targets may only occupy dozens or even a few pixels in low-resolution remote sensing images; so, it is difficult to capture some smaller targets on the ground. The image superresolution technology can provide more effective information for the detection, recognition, and understanding of small targets in remote sensing images.

The image superresolution reconstruction technology not only overcomes the limitation of the inherent resolution of imaging equipment but also considers the influence of

down sampling, blurring, noise, and other factors in the image degradation process. It improves the resolution of the image and also improves the reconstructed quality of image. Therefore, it shows important application prospects in many fields. In satellite imagery, if the image resolution is high, you can see more detailed information or identify more objects in the same satellite remote sensing image. Compared with improving the performance of image capture equipment, using a more effective super-resolution reconstruction algorithm is undoubtedly a more cost-effective choice. The use of algorithms for high and low resolution conversion has low cost and has been successfully applied to video, face recognition, medical images, and satellite images. Therefore, in the field of image processing today, superresolution reconstruction algorithms are a hot issue that has attracted widespread attention.

Over the past few years, significant efforts have been made to develop various data sets or to come up with various methods for classifying scenes from remote sensing images. Gong believed that a systematic review of the literature on data sets and scene classification methods was still lacking. He began with a comprehensive review of recent developments. He then proposed a large-scale dataset called “NWPU-ResISC45” [1]. Zhang proposes an effective nonlinear approximation scheme that uses Saito and Remy’s multiharmonic local sine transform (PHLST) in conjunction with an algorithm to automatically and adaptively tile a given image based on its local smoothness and singularity. To measure this local smoothness, he introduced what is called the image local Besov index, which is based on the point-by-point modulus of image smoothness. This adaptive partitioning of the image is important for image approximations using PHLST, because PHLST stores angular and boundary information for each partition; so, it is wasteful to divide a smooth region of a given image into a set of smaller partitions. He used remote sensing images of the South Pole to demonstrate the superiority of the proposed algorithm over PHLST using uniform tiling. Due to global climate change, the analysis of these images, including their effective approximation and compression, is becoming increasingly important [2]. Alimjan believes that distance measurement and classification standards are equally important in remote sensing image classification. And the accuracy of any one will affect the classification accuracy. He believes this is based on the separability of classes using SVM and the spatial and spectral characteristics of remote sensing data. In addition, he proposed a distance formula as a metric to consider the brightness and direction of vectors. First, a SVM is trained, and a support vector (SV) is obtained for each class. In the test phase, enter a sample of the new test and use the distance formula to calculate the average distance between the test sample of each class and the SV. Finally, it was decided to classify the test samples into the category with the minimum mean distance. Repeat this process until all test samples have been classified [3]. Gong noted that while significant efforts have been made to develop a variety of remote sensing image scene classification methods, most of them rely on hand-crafted features. Extensive evaluation of publicly available remote sensing image scene classification benchmarks and comparison with the most advanced methods prove the

effectiveness of the proposed BoCF method for remote sensing image scene classification [4]. Wang proposed an effective deep neural network for remote sensing image registration. Unlike the traditional method of feature extraction and feature matching, he pairs patches in the perceptual image and reference image and then directly learns the mapping between these patch pairs and their matching tags for later registration. This end-to-end architecture makes it possible to optimize the entire processing (learning mapping function) through information feedback while training the network, which is lacking in traditional methods. In addition, to alleviate the small data problem of remote sensing images for training, his proposal introduces self-learning by using the image and its transformed copy learning mapping function. In addition, his application of transfer learning to reduce the training phase of the huge computing costs. It not only speeds up our framework but also gets an additional performance boost [5]. Guo believes that the explosive availability of remote sensing images poses challenges to supervised classification algorithms such as support vector machines (SVM), because training samples are often very limited due to the expensive and laborious tasks of ground reality. Temporal correlation and spectral similarity between multitemporal images provide an opportunity to alleviate this problem. In his research, he proposed a SVM-based sequence classifier training (SCT-SVM) method for multitemporal remote sensing image classification. His method uses classifiers of previous images to reduce the number of training samples required for classifier training of input images. For each input image, the crude classifier is first predicted based on the temporal trend of a set of previous classifiers. The current training sample is then used to fine-tune the predicted classifier to a more accurate position. This method can be gradually applied to sequential image data, and only a small number of training samples are required for each image [6]. Munoz-mari sees HyperLabelMe as a network platform that allows automated benchmarking of remote sensing image classifiers. To demonstrate the attributes of the platform, he collected and coordinated large datasets of labeled multispectral and hyper spectral images with different categories, dimensions, noise sources, and levels. He suggested that registered users could download training data pairs (spectrum and land cover/usage tags) and submit predictions for the invisible test spectrum. The system then evaluates the accuracy and robustness of the classifier and reports the different scores as well as a ranking list of the best methods and users. The system he studied was modular and extensible, and the data set and classifier results were growing [7]. The goal of artificial intelligence (AI) is to design a machine that can perceive, remember, and recognize like humans. Perceptron is a machine learning algorithm that can feel and learn first, but its learning ability is limited. Later, a neural network model with multiple hidden layers appeared, which can learn more complex functions, but it is not a learning algorithm that can meet the needs of artificial intelligence. Although deep learning research has brought impressive theoretical results, learning algorithms and theoretical experiments, due to its numerous parameters, high training data and computational requirements, and prone to overfitting, the future development of algorithms still faces challenges. Convolutional neural

network is a feed-forward neural network, including convolutional layer, pooling layer, and output layer. It is an efficient pattern recognition method.

The current superresolution reconstruction based on a single image can be divided into two categories: (1) interpolation-based methods without training samples and (2) learning-based methods with training samples. Among them, the method based on interpolation only upsampling the low-resolution image, recovering the high-frequency components of the image, and finally enhancing the edge of the image, but does not fundamentally increase the amount of information contained in the image, and the reconstructed image lacks deeper details. This research mainly discusses the superresolution reconstruction of remote sensing images based on the middle layer supervised convolutional neural network. This paper designs a convolutional neural network with middle layer supervision. There are 16 layers in total, and the seventh layer is designed as an intermediate supervision layer. The middle supervision layer defines the error function of the supervision layer, which is used to optimize the error back propagation mechanism of the convolutional neural network to improve the disappearance of the gradient of the deep convolutional neural network. The algorithm training is mainly divided into four stages: the original remote sensing image preprocessing, the remote sensing image temporal feature extraction stage, the remote sensing image spatial feature extraction stage, and the remote sensing image reconstruction output layer. The current research results of superresolution reconstruction algorithms based on single images are mostly devoted to finding new features that make it easier to obtain image detail information for training, and these methods usually have high dependence on pre-defined features and calculations. The complexity is high, the robustness is not strong, and the input image needs to be a fixed size problem. The three-layer reconstruction model based on the convolutional neural network solves the problem of the lack of robustness of these methods and the need for fixed-size input images, and the network can obtain features by itself, without manual settings, and only need to set the parameters of the network to be trained in advance.

2. Methods

2.1. Image Superresolution Reconstruction. In improving the accuracy of pattern recognition, high-resolution images also play a pivotal role. Therefore, how to obtain and analyze high-resolution images under limited conditions has become a research hotspot in image processing.

In the 1970s, image sensor technology developed, and image acquisition devices based on CCD and CMOS sensors were comprehensively developed and popularized. Although these digital image devices can basically meet the various applications in people's daily life, they have higher image resolution. In order to meet people's growing demand for images, it is also a hotspot of modern research. By increasing the density of the acquisition unit and increasing the area of the sensor, better quality images can be obtained intuitively and quickly. The cost of the equipment is too expensive to

be widely used. These conditions limit the improvement of the image spatial resolution from the hardware aspect. People began to study how to use software processing technology to obtain higher resolution images, and the image superresolution reconstruction technology (superresolution, SR) has thus been developed.

The current sequence image superresolution algorithm is mainly divided into two stages: motion estimation and image reconstruction. The motion estimation stage makes full use of the time-dimensional motion information between the sequence images through accurate motion estimation to obtain the registered image; the image reconstruction stage uses registration after the image is reconstructed, the image resolution is improved. It can be seen from this that motion estimation is essentially to extract the temporal motion information of the sequence image. Therefore, how to better extract the motion information has become the key content of the research of the sequence image superresolution algorithm. By designing multiple convolutional layers, the network can extract multiple different image features [8].

$$y_j = b_j + \sum_i w_{ij} \times x_i. \quad (1)$$

The feed forward operation of the convolutional neural network is based on the convolution operation, adding a bias B to each feature plane:

$$C_i^L = F(A_i^L + B_i^L). \quad (2)$$

Then, the training error of the network on the entire training sample is expressed as [9]

$$E = \frac{1}{2} \|t_n - y_n\|_2^2 = \frac{1}{2} \sum_{n=1}^N \sum_{t=1}^n \|t_n - y_n\|_2^2. \quad (3)$$

Then, the training error of the network on a training sample can be [10]

$$E = \frac{1}{2} \|t_n - y_n\|_2^2 = \frac{1}{2} \sum_{n=1}^N \|t_n - y_n\|_2^2. \quad (4)$$

At present, the input images of many single-frame image superresolution algorithms will be processed by cubic cubic interpolation [11, 12].

$$S(w) = |w|^3 - 2|w|^2 + 1, \quad (5)$$

$$S(w) = -|w|^3 + 5|w|^2 - 8|w| + 4. \quad (6)$$

The interpolation value of the corresponding pixel is [13, 14]

$$G(i + u, j + v) = A \times B \times C. \quad (7)$$

Among them,

$$A = [S(1 + u), S(u), S(1 - u), S(2 - u)]. \quad (8)$$

FIGURE 1: Convolutional neural network structure with intermediate layer supervision.

Assuming that $P(Y|X)$ is the maximum posterior probability of high-resolution image Y under low-resolution image X , then [15]

$$A = [S(1+u), S(u), S(1-u), S(2-u)]. \quad (9)$$

The error between the output image and the real high-resolution image is, namely, [16]

$$E = F(X_L) - X_H. \quad (10)$$

Among them, X_L is the network input [17, 18].

2.2. Intermediate Layer Supervised Convolutional Neural Network. Therefore, the method to improve the resolution of the image from the hardware point of view not only requires high cost, but also has limited room for improving the quality of the equipment. The research is difficult, the cycle is long, and it is not easy to make rapid progress in the short-term. In addition, even if the imaging device collects ultra-high-resolution images, there are new problems in image transmission. Ultra-high-resolution images take up a lot of space, and the pressure when transmitting images through the network is relatively large. Therefore, by improving the imaging equipment method of improving image resolution by quality has been greatly restricted in practical applications.

The structure of the convolutional neural network with intermediate supervision designed in this paper is shown in Figure 1. There are 16 layers, namely, Conv.1, Conv.2 ... Conv.16, and the seventh layer is designed as an intermediate supervision layer. The intermediate supervision layer defines the error function of the supervision layer, which is used to optimize the error back propagation mechanism of the convolutional neural network, and realize the iterative update of the weight parameters of the network layer to improve the disappearance of the gradient of the deep convolutional neural network [19].

2.3. Residual Activation Block. In this paper, the traditional residual structure is improved, and the residual activation block RAB is proposed. By adopting a new spatial and channel attention mechanism, the network can simultaneously analyze the spatial relationship between the features at different positions in the channel in the feature map. The

relationship between features located at the same position in different channels is modeled. The spatial channel attention mechanism enables the network to give higher response values to features that are rich in spatial information and are located in important channels, while suppressing features that are scarce in surrounding spatial information and are located on less important channels, so that the network can adaptively adjust features. The expression of the residual block is [20]

$$X_{l+1} = X_l + f(X_l). \quad (11)$$

In the formula, X_l is the input feature of the residual block. The ReLU function is a modified linear unit [21]:

$$\text{ReLU}(x) = \max(x, 0). \quad (12)$$

The improved residual network uses the L2 norm as the loss function, and its expression is [22]

$$L_2(y, f) = \frac{1}{2N} \sum_{i,j} [y(i, j) - f(i, j)]^2. \quad (13)$$

2.4. Superresolution Steps of the Sequence Image SRCNN (Image Superresolution Reconstruction) Algorithm Based on the Intermediate Layer Supervised Convolutional Neural Network. The main goal of the algorithm is to find the mapping function F between the input and the output, so that the high-resolution images output by the network and the real high-resolution remote sensing images are as close as possible. The algorithm training is mainly divided into four stages: the original remote sensing image preprocessing, the remote sensing image temporal feature extraction stage, the remote sensing image spatial feature extraction stage, and the remote sensing image reconstruction output layer.

- (1) Preprocessing the input layer: like the single-frame remote sensing image superresolution convolutional neural network algorithm, the original remote sensing image sequence X is first preprocessed and expanded into an interpolated picture of the same size as the output remote sensing image, as the input of the network remote sensing image sequence [23]

- (2) Remote sensing image temporal feature extraction stage: in this stage, a two-layer network can be designed. The two layers are equivalent to the use of filters and time sliding windows for sequence remote sensing image processing, and the middle layer supervised convolutional neural network is used to perform three-frame remote sensing. Image joint convolution, first three frames of remote sensing images are separately convolved, and then the corresponding positions obtained by the separate convolution operation are “and mapped” to obtain the feature remote sensing image that is jointly matched, and the value of each pixel of the feature remote sensing image is obtained. Schematic diagram of the supervised convolutional neural network in the middle layer is described above. The feature remote sensing image extracted in this network stage contains the information of all frames of remote sensing image, thereby enhancing the detailed information of the reference frame remote sensing image, which contains more remote sensing image information than the feature remote sensing image obtained from the reference frame remote sensing image alone. Schematic diagram of the middle layer supervised convolutional neural network. The feature remote sensing image extracted in this network stage contains the information of all frames of remote sensing image, thereby enhancing the detailed information of the reference frame remote sensing image, which contains more remote sensing image information than the feature remote sensing image obtained from the reference frame remote sensing image alone.

At the beginning of network training, the parameters of the middle layer supervised convolutional neural network are initialized. The result of spatiotemporal fusion feature remote sensing image may not be good, and mismatching and ghosting will occur, but after the iterative feedback of the network, error adjustment, and finally fusion feature, remote sensing image obtained at this stage will definitely contain more detailed information, which provides a good basic remote sensing image for the subsequent spatial feature extraction of remote sensing image.

- (3) Remote sensing image spatial feature extraction layer: the main purpose of this stage is to extract and abstractly combine a series of feature remote sensing images through a multilayer network to obtain the features needed to reconstruct high-resolution remote sensing images, which are essentially features process of extraction and aggregation. In this layer, the feature maps of spatiotemporal feature fusion have been obtained; so, there is no need to extract the time dimension information at this time, only the spatial dimension information needs to be extracted, using a series of 2D convolution kernels, that is, selecting a series of different filters to convolve the remote sensing image, and supervising

the selection and optimization of these filters through the iterative network [24]

$$F(X_{t-1}) = \max(0, W \times X_{t-1} + B_t). \quad (14)$$

Among them, X_{t-1} represents the feature remote sensing image of the previous network layer, including n_{i-1} feature planes. The input of the first layer network at this stage is the n feature remote sensing images obtained in the remote sensing image fusion stage, and the subsequent network layer input is the previous output of a layer network. W_i represents the filter bank of this layer, each layer of network contains n_i group filters, and the number of feature planes output through this layer of network is n_i . B_i is a n_i -dimensional vector, which represents the offset coefficient corresponding to each feature map of this hidden layer.

The spatial characteristics extraction stage of the remote sensing image here can set up a multilayer network to increase the ability to extract abstract features. As the number of network layers increases, the extracted features are more sufficient, providing more and better feature information for subsequent remote sensing image reconstruction. But the price that needs to be paid is that with the increase of network complexity and calculation, the efficiency of network training and remote sensing image reconstruction decreases.

- (4) Remote sensing image reconstruction output layer: the last layer of the network draws on the single-frame remote sensing image SRCNN algorithm, and the output layer overlaps and adds the remote sensing images of the previous layer, performs an average operation on the overlapped blocks, eliminates the effect of blackness, and finally gets high-resolution remote sensing image that is also equivalent to a filter operation; so, the operation expression of this layer is similar to the operation expression of the previous spatial feature extraction stage

For accuracy evaluation, the result of change detection can also be regarded as a special classification result, and the detection of change can be regarded as a multiclass classification problem. We use indicators including overall accuracy (OA), producer accuracy (PA), user accuracy (UA), and kappa coefficient to evaluate change detection accuracy. Overall accuracy (OA): the number of samples that correctly detect the type of change divided by the total number of samples [25].

$$OA = \frac{\sum_{i=1}^k N_{ii}}{N}. \quad (15)$$

Producer accuracy (PA): the ratio of the number of samples N_{ii} of the correct detection change type i and the number of samples of the change type $i(N_{+i})$ in the real data are also known as the inspection rate [26].

$$UA_i = \frac{N_{ii}}{N_{i+}}. \quad (16)$$

The calculation formula of Kappa coefficient is as follows [27]:

$$\text{Kappa} = \frac{N \sum_{i=1}^k N_{ii} - \sum_{i=1}^k (N_{+i} N_{+i})}{N^2 - \sum_{i=1}^k (N_{+i} N_{+i})}, \quad (17)$$

where N is the total number of samples evaluated for accuracy, and k is the number of categories of the change type. The accuracy of traditional CNN model change detection is shown in Table 1.

2.5. Implementation of Qt-Based Remote Sensing Image Superresolution Software Platform. In order to allow users to compare the superresolution effect of remote sensing images more clearly, this paper uses the Qt5 interface library to implement the user interface of the remote sensing image superresolution software platform and uses the intermediate layer convolutional neural network and the remote sensing image superresolution reconstruction algorithm proposed in this paper. This software allows users to see more intuitive superresolution results, also provides a platform for learners and researchers of remote sensing image superresolution, and allows them to easily run without understanding the superresolution reconstruction algorithm. The algorithm in this paper obtains superresolution images.

2.5.1. The Functional Framework of Remote Sensing Image Superresolution Reconstruction Software. The problem of image superresolution reconstruction is an ill-conditioned inverse problem. Because multiple images with different high resolutions may correspond to the same low-resolution image, there is no unique solution to solving a high-resolution image through a low-resolution image. In the imaging process, because there are many factors that can degrade the image, such as system noise, imaging noise, motion blur, or downsampling, the reasons for image degradation must be fully considered in the image reconstruction process to better deal with the degraded image. The main interface of the software contains the main panel, toolbar, and menu bar. The menu bar contains all the functional modules of the software, and the main panel and toolbar contain commonly used functional modules. The software is mainly divided into three modules, namely, file operation module, superresolution reconstruction module, and other functional modules. Among them, the superresolution reconstruction module implements a bicubic interpolation algorithm, an improved residual network remote sensing image superresolution algorithm, and a remote sensing image superresolution algorithm based on the middle layer supervised convolutional neural network. After the superresolution algorithm is executed, you can directly view the superresolution result image on the main panel and parameter indicators such as peak signal-to-noise ratio and structural similarity. Figure 2 is the module division diagram of this software, and the figure also contains shortcut keys for some functions.

2.5.2. Software Module. The software contains three modules:

TABLE 1: Change detection accuracy of the traditional CNN model.

Test	PA	UA	OA	Kappa
Unchanged	0.9692	0.8695		
Variation type 2	0.0082	0.0143		
Variation type 5	0.2330	0.2118	0.8172	0.3761
Variation type 6	0.1153	0.7426		
Variation type 7	0.3174	0.6918		

(1) *File Operation Module.* The function of this module includes opening the image, saving the image, and closing the software. Opening the image refers to opening the remote sensing image that needs superresolution on the disk, and saving the image refers to saving the generated superresolution image on the disk. And these three functions use shortcut keys consistent with other software.

(2) *Superresolution Reconstruction Module.* This module implements the improved convolutional neural network remote sensing image superresolution reconstruction algorithm in this article and also includes a faster bicubic interpolation algorithm. This module also uses the objective parameters PSNR (peak signal-to-noise ratio) and SSIM to evaluate the image quality of the superresolution reconstruction result image. The expression of peak signal-to-noise ratio (PSNR) is

$$\text{PSNR} = 10 \lg \frac{L^2}{\text{MSE}} = 10 \lg \frac{L^2 MN}{\sum_{i=1}^m \sum_{j=1}^n [y(i, j) - f(i, j)]^2}. \quad (18)$$

In the formula, MSE is the mean square error, L is the peak signal, and the value is 255 for the 8-bit gray scale image. The larger the PSNR, the better the image quality. Natural images have a specific structure, and each pixel in the image has a strong subordination relationship, which reflects the structure in the image. The structural similarity measures this structure and is used to evaluate the quality of the image [28].

$$\text{SSIM} = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (19)$$

In the formula, μ_x, μ_y is the mean value, σ_x, σ_y is the standard deviation, σ_{xy} is the covariance, and L is 255 for the 8-bit gray scale image. SSIM indicates the degree of similarity between the structure of the superresolution image and the actual high-resolution image, its value does not exceed 1, and the closer the SSIM is to 1, the more similar the structure and the better the reconstruction effect.

Mean pixel accuracy (MPA, mean pixel accuracy) is a simple improvement based on the PA index, which is to

FIGURE 2: Module division of the software.

average the proportion of all correctly classified pixels to the total pixels. The mathematical expression is as follows:

$$\text{MPA} = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ij}}{\sum_{j=0}^k P}. \quad (20)$$

The superresolution reconstruction module runs the SRR (image superresolution reconstruction) algorithm on the input image and displays the output image and various parameters after running. This software includes three algorithms for use, namely, the bicubic interpolation algorithm, the improved residual network superresolution algorithm in this article, and the improved densely connected network superresolution algorithm in this article. These three algorithms are placed in the main panel in the form of radio buttons. After selecting the buttons of different algorithms, the software will also run different superresolution algorithm programs.

(3) *Other Functional Modules.* This module provides users with more convenient functions, including viewing the gray-scale histogram, which is opened in the toolbar in the main interface, and the shortcut keys are Ctrl+1 and Ctrl+2, which are used to view the input image and output image, respectively, histogram of grayscale. Another function of this module is the help and about of the software, displayed in the menu. The configuration used to build this software: operating system Win10 Professional Edition, and CPU is Intel-Corei7-7700 K, memory 16GB, graphics card GTX10606GB.

3. Results

After the image degradation model is established, the reconstruction-based method mainly uses multiframe low-resolution (LR) images as a consistency constraint, combined with the prior knowledge of low-resolution images

for superresolution reconstruction, thereby obtaining super-resolution image. It makes full use of the information complementarity between multiple low-resolution images in the same scene and effectively fuses these complementary information to achieve the purpose of reconstructing high-resolution images.

The network performs a total of 40,000 training iterations, and the cumulative time is 40 minutes. At this time, the loss function value of the training set has converged to 1.44. A test is performed every 500 iterations, and the network with the largest PSNR in the test process is taken as the final network. The relationship between the network training loss function and the number of iterations is shown in Figure 3.

The generalization experiment results show that the superresolution effect of the network on the verification set is better than that of the network before the improvement. Here, 8 remote sensing images are selected for verification and display, as shown in Figure 4.

The advantages of the iterative back projection method are simple, intuitive, and fast calculation speed, but the selection of its back projection operator is more difficult, the solution is not unique, and it is very dependent on the initial value. This method cannot take advantage of the inherent characteristics of high-resolution images priori constraint knowledge. In the iterative process, the error of the back projection only accumulates evenly on the reconstructed image, resulting in a sawtooth effect on the edge of the reconstructed high-resolution image. The experiment chooses low-resolution images, bicubic interpolation, VDSR algorithm, and improved methods for comparison. For the rationality of the experiment comparison, the size scaling factor is 3 in the experiment. Only the comparison results of 4 groups of experiments are listed here as shown in Figure 5.

The image superresolution reconstruction algorithm based on learning is also called the image superresolution reconstruction based on example learning, and it is one of

FIGURE 3: The relationship between the network training loss function and the number of iterations.

FIGURE 4: Select 8 remote sensing images for result verification.

FIGURE 5: Four comparison results of groups of experiments.

the main research hotspots in the image superresolution reconstruction algorithm in recent years. This method finds a mapping relationship between low-resolution images and high-resolution images by learning to construct a sample library of low-resolution images and high-resolution images in advance and finally reconstructs high-resolution images through this mapping relationship.

Through the superresolution renderings, it can be seen that the subjective visual effect of the superresolution algorithm in this chapter is better than that of the VDSR network, and it can generate higher-quality superresolution images with more detailed information about the image texture. The second set of comparison results is shown in Figure 6.

FIGURE 6: The second set of comparison results.

TABLE 2: Eight PSNR comparison of the results of the three algorithms for remote sensing images.

Image	Bicubic interpolation	VDSR	Method of this article
Group 1	20.65	21.94	22.40
Group 2	16.31	17.37	17.82
Group 3	20.98	22.41	22.85
Group 4	23.99	24.96	25.41
Group 5	24.90	27.25	27.70
Group 6	19.85	21.95	22.40
Group 7	22.37	24.75	25.20
Group 8	22.80	23.82	24.28
Average value	21.48	23.06	23.51

Table 2 shows the PSNR comparison of the three algorithm results of 8 remote sensing images.

Most of the reconstruction algorithms based on single-frame image superresolution basically use the characteristics of the image detail information for training, but these methods rely on the pretrained image features relatively high, and the reconstruction speed is relatively slow. In addition, there are problems such as unclear edges of the reconstructed image, the need to fill the edges of the image, and the lack of robustness. The SSIM comparison is shown in Table 3. The PSNR of 8 sets of remote sensing images under different algorithms is compared with SSIM. It can be seen that compared with the VDSR algorithm, the objective parameter PSNR of the improved algorithm is increased by about 0.45 dB, and the SSIM is increased by about 0.023.

In order to improve the superresolution effect of remote sensing images, the experimental training samples use remote sensing images provided by the cooperative unit, the image size is 2000×2000 , and it is intercepted into 100×100 subimages as the data set. In order to increase the number of data sets, the previously captured subimage data sets were rotated by 90° , 180° , and 270° respectively, which increased the number of data sets by 3 times, and finally got about 25,000 100×100 small images. Then, use 80% of the data set as the training set, 10% as the test set, and 10% as the verification set. The training set is still used to train the network and update the parameter weights of the

TABLE 3: SSIM comparison.

Image	Bicubic interpolation	VDSR	Method of this article
Group 1	0.594	0.701	0.729
Group 2	0.456	0.590	0.630
Group 3	0.649	0.735	0.760
Group 4	0.654	0.738	0.763
Group 5	0.775	0.858	0.872
Group 6	0.712	0.828	0.845
Group 7	0.730	0.838	0.854
Group 8	0.779	0.834	0.850
Average value	0.669	0.765	0.788

network. The test set is only used to evaluate the superresolution ability that changes with the number of network iterations, not to update the weights. The training set image itself is used as the output of the network, and the training set image (2 times under sampling) is reduced twice to 50×50 , which is used as the low-resolution image of the network input. The network training performs an epoch for about 14 minutes. When the training epoch reaches 35 times, the network has converged. At this time, the loss function has converged to 0.017, and the cumulative time is about 8 hours. The relationship between the error and the number of iterations in the network training process is shown in Figure 7. The test carried out after each epoch training is close to stable at about 35 training sessions.

The four-layer convolutional network model based on multifeature map input proposed in this paper can input multiple images with different features, which provides more features of superresolution images to the greatest extent, and is more conducive to the reconstruction of superresolution images. Compared with the three-layer convolutional layer, the four-layer convolutional layer can extract more image features. This paper also discusses the influence of the number of convolution kernels and the parameters of the convolution layer on the network through experimental results. Because multiple images with different characteristics are obtained by interpolation from the same low-resolution image, the model proposed here is essentially a reconstruction model based on single-frame image superresolution. As shown in Figure 7, it can be seen that the training of

FIGURE 7: The relationship between the error and the number of iterations in the training process of the network.

TABLE 4: Accuracy evaluation of experimental area change detection results.

Model	5 × 5 neighborhood		5 × 5 neighborhood (CNN+)	
	Accuracy	Recall rate	Accuracy	Recall rate
CNN	0.8179	0.7336	0.7990	0.7518
LSTM	0.8007	0.7680	0.8597	0.7502
SRR	0.7228	0.8577	0.8256	0.8589
This research	0.8337	0.9252	0.9822	0.8981

the network has converged when the epoch reaches about 32 times, the training set error is still decreasing as the epoch increases, and the PSNR of the test set shows a slight downward trend, which is due to cause by over fitting. In order to solve the problem of network overfitting, this experiment takes the network when the network converges, and the PSNR of the test set reaches the maximum as the final network. At this time, the epoch is the 35th time.

The accuracy evaluation of the experimental area change detection results is shown in Table 4. The difference between different recurrent network models does not seem to be large. The 5 × 5 neighborhood (CNN+) input method combined with this research method achieves the highest accuracy rate (0.8337) and recall rate (0.9252).

We will divide into 5 groups according to the input method, each group contains 6 types of recurrent neural networks, we can see that the method of first facing the pixel neighborhood has higher accuracy, recall and accuracy, the input method of larger neighborhood can significantly improve the recall rate and accuracy rate, and the combination of CNN can further improve the recall rate and accuracy rate. Here, the average accuracy rate and average recall rate of the input method are in this study. It is obvious that they reach the highest level, and in this way, the accuracy of different cyclic neural networks is more different. The comparison between CNN and this research method is shown in Figure 8.

The change detection accuracy of the twin convolutional neural network model is higher, especially when the convolutional neural network model is simple and the number of

layers is small, and it is more obvious, and when the convolutional neural network model is more complex and the number of layers is large, the single channel accuracy difference between the change detection of the twin convolutional neural network model and the twin convolutional neural network model will be significantly reduced. The accuracy comparison of different convolutional neural network models is shown in Figure 9.

Because mask RCNN realizes the process of detection and segmentation through a network, the change detection method based on mask RCNN has higher detection accuracy than the change detection method based on fast RCNN. After adding FPN, the change detection accuracy based on mask RCNN can be further improved. Like the change detection method based on fast RCNN, the detection effect of this research model is better than mmrcnn. Remote sensing image change detection is shown in Table 5.

It can be seen from Figure 10 that the model test MPA of the middle layer supervised convolutional neural network model reached 0.86, MIOU reached 0.82, and the training loss dropped to 0.15. From the trend of each indicator curve of the model, it can be seen that the model obtained a good convergence. The change of MIOU of the middle layer supervised convolutional neural network model is shown in Figure 10.

4. Discussion

On the one hand, the single-frame image superresolution convolutional neural network is considered from a global

FIGURE 8: Comparison of CNN and this research method.

FIGURE 9: Accuracy comparison of different convolutional neural network models.

TABLE 5: Remote sensing image change detection.

Method	Variety	Unchanged	OA	Kappa
MMRCNN (NoFPN)	PA	0.6929	0.9954	0.9885
This research (NoFPN)	UA	0.7810	0.9928	0.7284
MMRCNN (FPN)	PA	0.7518	0.9971	0.9914
This research (FPN)	UA	0.8579	0.9942	0.7970
MMRCNN (FPN)	PA	0.7601	0.9984	0.9929
This research (FPN)	UA	0.9181	0.9944	0.8281
MMRCNN (FPN)	PA	0.8354	0.9980	0.9943
This research (FPN)	PA	0.9218	0.9961	0.8676

perspective. When the convolutional neural network is trained, the entire network parameters are adjusted synchronously, and finally, the optimization of the output is global. However, most of other superresolution algorithms based

on learning are optimized step by step; that is, the solution of each step is the current local optimal, but the final output is not necessarily the global optimal. On the other hand, when using a single-frame image superresolution

FIGURE 10: The change of MIOU in the middle layer supervised convolutional neural network model.

convolutional neural network to process images, the input image does not require complex preprocessing, and the hidden layer structure of the convolutional neural network is consistent, which can be distributed calculations, which greatly improves the computational efficiency, thereby improving the efficiency of the entire superresolution reconstruction. Convolutional neural networks are currently used for single-frame image superresolution processing and have achieved good results in image reconstruction efficiency and quality. However, in recent years, some people have used convolutional neural network for motion feature analysis, used the middle layer supervised neural network to extract the time characteristics of sequence images, and achieved certain results in the recognition of motion behavior. This discovery provides the possibility for convolutional neural networks to be used in the superresolution of sequence images [29].

In a convolutional layer of a convolutional neural network, it usually contains one or more feature planes composed of several neurons. All neurons on the same feature plane share weights, and the shared weights here are the convolution kernels. The convolution kernel generally generates a matrix in the form of Gaussian function for initialization and then obtains the weights in the convolution kernel through network training. The superresolution research of a single frame image is considered to be a typical ill-conditioned inverse problem. Most of its ideas are to combine some prior knowledge and use regularization methods to improve image quality. When there are more input images, that is, multiple continuous images are input, more structural similarity and information redundancy of the images are obtained. Using these additional information and previous image knowledge can better integrate and improve the image quality, so that the output image quality is better. Therefore, the research on the superresolution reconstruction algorithm of sequence images has gradually gained people's attention.

The last layer of the network is the image reconstruction layer. In traditional methods, it is often necessary to decompose the whole image into small image blocks and then restore the small image blocks, respectively. Finally, all image blocks are overlapped and fused, and the blocking effect is eliminated for the overlapped part of the image blocks. In srcnn algorithm, the operation of image reconstruction layer is to reconstruct the features extracted from the above network layer into a high-resolution image and smooth filter the image to obtain the final high-resolution image. This overlapping addition operation can also be regarded as a filtering operation in convolutional neural network, which is consistent with the previous network layer. In the stage of image superresolution reconstruction, image preprocessing is carried out first, and the image to be reconstructed is expanded by cubic convolution interpolation according to the superresolution factor. Then, the mapping coefficients are obtained, the interpolated three frame sequence images are input into the trained network, and the high and low resolution image mapping coefficients of the sampled frame images are solved through the network. Finally, the high-resolution image is solved, the mapping coefficient gray and the sampling frame interpolation input image are point multiplied, and the solution result is normalized to obtain the high-resolution output image [30].

The result of image registration directly affects the effect of subsequent image reconstruction. Only with accurate motion estimation can it be possible to accurately find the correspondence between high and low resolution subimage blocks and to obtain a better superresolution effect. However, in the case of complex motion, on the one hand, the number of images directly affects the computational efficiency of the motion estimation algorithm. To obtain accurate motion estimation, enough images are required, which will lead to a higher computational complexity of the algorithm. On the other hand, the current motion matching technology is difficult to ensure that its accuracy

requirements are met every time. If the accuracy of motion estimation is too poor, the registered image may have ghosts or mismatches, which will lead to poor subsequent superresolution reconstruction effects. It may even appear worse than the original image quality.

5. Conclusion

This research analyzes the difficult problems in the research of traditional serial remote sensing image superresolution, and realizes that the essence of motion estimation is to use the motion information between remote sensing images. Therefore, the middle layer convolutional neural network is studied, and the convolutional neural network is used to extract the sequence. The spatiotemporal characteristics of remote sensing images and the continuous optimization of parameters are improved on the basis of the superresolution convolutional neural network of single-frame remote sensing images. The superresolution algorithm of sequential remote sensing images based on the intermediate layer convolutional neural network is proposed, which avoids the traditional motion estimation problem in the superresolution algorithm of sequence remote sensing images has achieved good superresolution results. In addition, in order to improve the efficiency of the serial remote sensing image superresolution algorithm based on the intermediate layer convolutional neural network, the dictionary learning superresolution algorithm based on sparse representation is used for reference, and the convolutional neural network is used to obtain the coefficient matrix, which reduces the network scale. Improve the efficiency of superresolution calculation. There is a large amount of redundant information in the front and rear frames. If you directly use the single-frame remote sensing image superresolution processing, you may also achieve better results, but it is very wasteful, and the efficiency will not be improved. Sequence remote sensing image superresolution algorithms may be able to use redundant information and use a smaller-scale network for superresolution processing to achieve faster efficiency.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

The research was partially sponsored by the Scientific Research Program in Shaanxi Provincial Department of Education (Grant No. 20JK0583).

References

- [1] C. Gong, J. Han, and X. Lu, "Remote sensing image scene classification: benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.
- [2] Z. Zhang and N. Saito, "PhlSt with adaptive tiling and its application to antarctic remote sensing image approximation," *Inverse Problems & Imaging*, vol. 8, no. 1, pp. 321–337, 2014.
- [3] G. Alimjan, T. Sun, Y. Liang, H. Jumahun, and Y. Guan, "A new technique for remote sensing image classification based on combinatorial algorithm of SVM and KNN," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 7, p. 1859012, 2018.
- [4] G. Cheng, Z. Li, X. Yao, L. Guo, and Z. Wei, "Remote sensing image scene classification using bag of convolutional features," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1735–1739, 2017.
- [5] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 148–164, 2018.
- [6] Y. Guo, X. Jia, and D. Paull, "Effective sequential classifier training for SVM-based multitemporal remote sensing image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3036–3048, 2017.
- [7] J. Munoz-Mari, E. Izquierdo-Verdiguier, M. Campos-Taberner et al., "HyperLabelMe : a web platform for benchmarking remote-sensing image classifiers," *IEEE Geoscience & Remote Sensing Magazine*, vol. 5, no. 4, pp. 79–85, 2017.
- [8] M. Gohil, N. Patel, and R. D. Kamboj, "Investigation and analysis of land use/land cover (LULC) change using remote sensing image interpretation for urban area of Rajkot municipal corporation, Gujarat," *Disaster Advances*, vol. 14, no. 16, pp. 28–34, 2021.
- [9] C. Yu, L. Wang, J. Zhao, L. Hao, and Y. Shen, "Remote sensing image classification based on RBF neural network based on fuzzy C-means clustering algorithm," *Journal of Intelligent Fuzzy Systems*, vol. 38, no. 4, pp. 3567–3574, 2020.
- [10] F. Sun, H. Li, Z. Liu, X. Li, and Z. Wu, "Arbitrary-angle bounding box based location for object detection in remote sensing image," *European Journal of Remote Sensing*, vol. 54, no. 1, pp. 102–116, 2021.
- [11] D. Hou, Z. Miao, H. Xing, and H. Wu, "Two novel benchmark datasets from ArcGIS and Bing world imagery for remote sensing image retrieval," *International Journal of Remote Sensing*, vol. 42, no. 1, pp. 240–258, 2021.
- [12] J. Uthayakumar, M. Elhoseny, and K. Shankar, "Highly reliable and Low-Complexity image compression scheme using neighborhood correlation sequence algorithm in WSN," *IEEE Transactions on Reliability*, vol. 69, no. 4, pp. 1398–1423, 2020.
- [13] Z. Zhang, Y. Liu, T. Liu, Z. Lin, and S. Wang, "DAGN: a real-time UAV remote sensing image vehicle detection framework," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 11, pp. 1884–1888, 2020.
- [14] H. Chen, Y. Zhang, M. K. Kalra et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.
- [15] C. Zhang, P. Wang, K. Chen, and J. K. Kämäräinen, "Identity-aware convolutional neural networks for facial expression

- recognition,” *Journal of Systems Engineering and Electronics*, vol. 28, no. 4, pp. 784–792, 2017.
- [16] S. Kruthiventi, K. Ayush, and R. V. Babu, “DeepFix: a fully convolutional neural network for predicting human eye fixations,” *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4446–4456, 2017.
 - [17] E. Laxmi Lydia, J. S. Raj, R. P. Selvam, M. Elhoseny, and K. Shankar, “Application of discrete transforms with selective coefficients for blind image watermarking,” *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 2, article e377, 2021.
 - [18] W. Shen, M. Zhou, F. Yang et al., “Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification,” *Pattern Recognition*, vol. 61, no. 61, pp. 663–673, 2017.
 - [19] Y. Hou, Z. Li, P. Wang, and W. Li, “Skeleton optical spectra-based action recognition using convolutional neural networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 3, pp. 807–811, 2018.
 - [20] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, “Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network,” *Knowledge-Based Systems*, vol. 132, pp. 62–71, 2017.
 - [21] C. Yuan, X. Li, Q. Wu, J. Li, and X. Sun, “Fingerprint liveness detection from different fingerprint materials using convolutional neural network and principal component analysis,” *Computers, Materials and Continua*, vol. 53, no. 4, pp. 357–371, 2017.
 - [22] S. Biswas, D. Devi, and M. Chakraborty, “A hybrid case based reasoning model for classification in Internet of Things (IoT) environment,” *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 104–122, 2018.
 - [23] J. Zhu, S. Liao, Z. Lei, and S. Z. Li, “Multi-label convolutional neural network based pedestrian attribute classification,” *Image and Vision Computing*, vol. 58, pp. 224–229, 2017.
 - [24] S. Poria, H. Peng, A. Hussain, N. Howard, and E. Cambria, “Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis,” *Neurocomputing*, vol. 261, pp. 217–230, 2017.
 - [25] Q. Liu, X. Lu, Z. He, C. Zhang, and W. S. Chen, “Deep convolutional neural networks for thermal infrared object tracking,” *Knowledge-Based Systems*, vol. 134, pp. 189–198, 2017.
 - [26] P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento, and I. Hajirasouliha, “Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images,” *eBioMedicine*, vol. 27, pp. 317–328, 2018.
 - [27] J. C. Hou, S. S. Wang, Y. H. Lai, Y. Tsao, H. W. Chang, and H. M. Wang, “Audio-visual speech enhancement using multimodal deep convolutional neural networks,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 2, pp. 117–128, 2018.
 - [28] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, “Identification of Rice diseases using deep convolutional neural networks,” *Neurocomputing*, vol. 267, pp. 378–384, 2017.
 - [29] S. Pang, Z. Yu, and M. A. Orgun, “A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images,” *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 283–293, 2017.
 - [30] Y. Wang, Y. Li, and Q. H. Zhao, “Region-based multiscale segmentation of panchromatic remote sensing image,” *Kongzhi yu Juece/Control and Decision*, vol. 33, no. 3, pp. 535–541, 2018.