

Research Article

Research on Low-Resolution Image Fusion Algorithm Based on Deep Learning

Yan Zhang 

Intelligent Science and Information Engineering College, Xi'an Peihua University, Xi'an 150761, China

Correspondence should be addressed to Yan Zhang; 150761@peihua.edu.cn

Received 14 January 2022; Accepted 16 April 2022; Published 17 June 2022

Academic Editor: Sheng Bin

Copyright © 2022 Yan 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.

The emergence of portable devices provides great convenience for image acquisition, but the image resolution is too low, which affects the identification and use of the image. Multiframe super-resolution algorithm makes a great contribution to extracting image features, but there is a problem with too much computation. Based on this, this paper proposes a low-resolution image fusion algorithm based on deep learning. Based on the introduction of extracting image features from low-resolution images, this paper proposes to fuse multiple low-resolution images, reduce the time of calculation error per frame, reduce the amount of calculation, extract feature information, and improve the alexnet network model to realize image extraction. The convolutional self-coding fusion network is used to realize image fusion and complete the reconstruction and fusion of low-resolution images. In the simulation analysis of the fusion algorithm, objective indexes such as network structure and loss function are selected to evaluate the effectiveness of the algorithm. This paper reduces the computational complexity of image acquisition and ensures that the amount of calculation can be reduced in each iteration. Analyze the effectiveness of the fusion algorithm, comprehensively evaluate the objective indicators, and verify the advantages of the algorithm through indicators such as gradient change and information entropy.

1. Introduction

Image sampling is to divide a spatially continuously distributed analog image into $m * n$ network. $M * n$ is called the spatial resolution of the image. According to the Shannon sampling theorem, as long as the sampling frequency is greater than twice the maximum frequency of signal sampling, the original signal can be completely restored by the sampling signal. The sampling is scanned in a straight line along the horizontal direction from top to bottom at a certain interval in the vertical direction and then sampled horizontally. The larger the sampling interval, the fewer the pixels (in line with the general law, the smaller the interval, the larger the pixels) i.e., the image quality is poor, and the mosaic effect will appear seriously.

With the development of equipment and technology, all kinds of low-cost and small equipment provide a lot of convenience for image acquisition, but it also brings the characteristics of low image resolution. In order to better

observe and analyze images, the reconstruction and fusion of low-resolution images has become a problem that must be faced [1]. In previous research and analysis, image restoration is often realized by super-resolution and image features are generated by noise blurred image sequences. This method is widely used in meteorology, medicine, and so on [2]. At present, there are many SR restoration algorithms, covering the frequency domain and spatial domain. Compared with the frequency domain, the amount of calculation in the frequency domain is low. It is only suitable for images with translational motion, unable to analyze video images, and the amount of calculation in the spatial domain method is large [3]. Based on this, this paper proposes a low-resolution image fusion algorithm based on deep learning.

This paper studies the low-resolution image fusion algorithm based on deep learning, which is mainly divided into four sections. Section 1 introduces the research background and overall framework of this research. Section 2 describes the research progress of image fusion algorithms at

home and abroad. On the basis of extracting super-resolution features from low-resolution images, Section 3 decomposes the images by using the improved alexnet network model and fuses the images by convolution self-coding fusion network. Section 4 simulates and analyzes the effectiveness of the constructed image fusion algorithm and obtains the simulation results and conclusions.

Compared with previous studies, the innovation of this paper lies in the selection of images. The selected images are some low-resolution images, which can quickly realize image restoration, reduce the estimated image sampling operation time, and fuse on the basis of feature extraction. In the simulation analysis, it is more reliable to evaluate with reference to objective indicators (such as gradient change, mean square deviation, and information entropy).

2. Related Work

In recent years, there are many research studies on image fusion, and good research results have been achieved [4]. Wang et al. summarized the existing DL based image fusion methods into several general frameworks by integrating the information contained in multiple images of the same scene into a composite image [5]. Lee et al. proposed a strict joint registration method of kompsat-3 MS and pan images based on physical sensor modeling for meteorological images and evaluated the influence of CCD line offset, ephemeris, and terrain elevation on image coordinate difference [6]. Asadi and Ezoji proposed to realize multiexposure image fusion based on intensity map pyramid integration in the research and analysis [7]. Gómez-sánchez et al. can process the image fusion structure through the alternating least squares (MCR-ALS) algorithm of multivariable curve resolution based on the mixed bilinear/trilinear model [8]. Geng and Liu proposed a multifocus image fusion method based on a guided filter [9]. Piqueras et al. proposed a special variant decomposition algorithm, multivariate curve resolution alternating least squares (MCR-ALS), which allows the simultaneous analysis of images collected by different spectral platforms without losing spatial resolution and ensures the spatial consistency between processed images [10]. In the research of Cabazos Marín et al., the auto focusing method selects the best focusing image (BFI) from the image stack captured from different distances of the object, calculates the spectrum of each vector using Fourier transform, and determines the BFI of the spectrum of each vector corresponding to the image in the stack [11]. Gupta et al. proposed a new high dynamic range (HDR) image generation method. The system uses a genetic algorithm to learn the relationship between different exposure images, generate long exposure and short exposure images and different exposure images to generate HDR like images [12].

To sum up, the purpose of the existing image fusion algorithms is to improve the information contained in the image, and different image fusion requirements are different. The fusion method based on image block size appeared earlier, and many improved versions appeared. The quadtree method also began to be introduced. The multiscale weighted gradient method and dynamic scene matting

method have higher computational complexity. However, the research object is relatively narrow. Although individual studies can take low-resolution images as the research object, the amount of calculation is large, and it is difficult to improve the image resolution.

3. Methodology

3.1. Super-Resolution Restoration of Low-Resolution Images. Medical, meteorological, and other images are affected by equipment and environment and have the characteristics of the continuous gray distribution. The image is blurred and discretized [13]. This process can be described as follows:

$$Y[m, n] = [Hcam(x, y) * F(hatm(x, y)) * X(x, y)] \downarrow + V[m, n], \quad (1)$$

where $*$ represents convolution operator, F represents set transformation operator, and V represents noise.

Affected by the traditional imaging system, the influence of camera lens shaking on the image is more obvious. Therefore, only considering the influence of the camera, the formula can be expressed as follows:

$$Y_k = D_k H_k F_k X + V_k, \quad (2)$$

where k represents a natural number, F represents set motion operator, and X represents rearrangement in order.

The LR image is registered, and then the pixel value of each grid point is copied to the image. Through the difference method, the image with the same resolution is obtained. For most SR problems, the number of sampled frames will be less than the number of LR frames, so it is impossible to determine all pixel values in data fusion. These pixels can be added with independent interpolation steps to fill in the 0 value and ignore the influence of position. If there are multiple LR frames, it can be determined directly using the average value of pixels. Obtain the minimum quasi side of X from Z and iteratively restore the original image. The formula can be expressed as follows:

$$X = \text{Arg min} [\|A(HX - z)\| + \lambda' r_r(X)], \quad (3)$$

where A represents the diagonal matrix, and the value is equal to the measured data of each element. Multiple measurement data will have a great impact on the estimated value of the HR frame. Considering that A is a diagonal matrix, the minimum value problem can be degraded by the following formula:

$$X_{n+1} = X_n - \beta \{H^T A^T (AHX_n - AZ) + \lambda' TX_n\}. \quad (4)$$

Therefore, in the iterative operation, there is no need to go through sampling and error analysis of each frame, and the operation speed is greatly increased.

3.2. Image Fusion Algorithm Based on Deep Learning. After getting a clearer image, image fusion is carried out. At present, there are many kinds of network models, among which lenet, alexnet, googlenet, etc. These kinds of network models mainly classify focused and defocused images [14].

Considering that the actual image is very small, if there are too many layers of the network model, it will easily lead to the loss of some image eigenvalues and the decline of image recognition ability. Although the commonly used model has higher accuracy, there are too many network layers [15], so this paper adopts the alexnet model in the selection of the network model and improves the model.

The alexnet network model contains 8 layers, including 3 full connection layers and the others are convolution layers. The loss function adopted at the end of the model is the softmax function. There are 96 convolution cores in the convolution layer, and the pixel size of the convolution core is only 11×11 . After the convolution is completed, the nonlinear mapping is carried out, the selected activation function is the ReLU correction function, the local normalization processing is carried out, and the obtained data is used as the input data into the convolution layer of the next layer. The convolution kernel pixel size of the second layer is only 5×5 . There is no pool layer between the 3rd to 5th convolution layers, but it is directly connected. There are 4096 nodes from layer 1 to layer 5.

The activation function adopted by the network model is ReLU function, and the formula is expressed as follows:

$$\text{ReLU}(x) = \max(0, x). \quad (5)$$

This function does not need normalization and will not be saturated. However, in order to improve universality, local normalization is implemented, and the formula can be expressed as follows:

$$b_{x,y}^i = a_{x,y}^i \left(k + a \sum_{\max(0, i-n/2)}^{\min(N-1, j+n/2)} (a_{x,y}^i)^{-2\beta} \right), \quad (6)$$

where a represents the calculated activation value, b represents the data result after normalization, and N represents the value of the convolution kernel. In this formula, n, k, a, β is a constant and belongs to a parameter. Generally, n value is set as 5, k value is 2, a value is 10^{-4} , and β value is 0.75. This setting can reduce the error rate to an ideal range.

Alexnet network model often adopts the maximum pooling method and average pooling method. Compared with the traditional CNN model, the adjacent areas will not overlap. The network model set in this paper adopts the overlapping pooling method, which belongs to the maximum pooling method. Compared with other methods, this pooling method can effectively reduce the error rate between the first category and the actual results, and avoid fitting. In order to avoid over fitting of the initial center, a dropout layer is added behind the full connection layer. With a certain probability, some neurons are activated first and other neurons are inhibited at the same time. The probability is set to 0.5 to reduce the complexity, no back propagation or forward propagation and reduce the network model parameters. Due to the large difference in image size, the network graphics need to be improved. Referring to the vggnet model, the minimum convolution kernel, replenishment up and down and central feature data are obtained. In this method, the network layer has only 5 layers, and the

size of all convolution cores in the improved 3-layer convolution layer is 3×3 . The pool layer image size is 16×16 .

The image with a clear texture is selected as the sample and smoothed by the Gaussian filter. Restricted, the Sobel operator is used for filtering, and the calculated scores of the obtained images are assigned, respectively. In consideration of possible misjudgment, correction shall be carried out according to the situation, and the formula is

$$\begin{cases} e(i, j) = 1, & \text{cot} > 4, \\ e(i, j) = 0, & \text{cot} \leq 4. \end{cases} \quad (7)$$

Here, cot represents the number near 1. After correction, the binary diagram is obtained. The boundary region can be calculated by the Brenner gradient function. The formula is

$$D(f) = \sum_j \sum_i |f(i+2, j) - f(i, j)|^2, \quad (8)$$

where f represents the pixel position.

3.3. Fusion Algorithm Based on Convolutional Neural Network. In order to further improve the image resolution, a convolutional neural network is used to optimize and improve. Among all kinds of deep feedforward neural network models, the convolutional neural network is special. If the hierarchical connection mode of full connection is adopted, there will be training redundancy, which is attributed to the excessive dependence on the amount of data. Using locally connected convolutional neural networks can greatly reduce the training parameters. In image processing, convolution, a linear operation, is used to obtain high robustness. At present, in several common signal processing, full, valid, and the same convolution are mainly used. If the input is a one-dimensional signal and the filter is also one bit, full convolution can be expressed as follows:

$$y(t) = \sum_{i=1}^m x(t-i+1)w(i), \quad (9)$$

where t is a natural number.

The SAMC convolution and valid convolution formulas are expressed as follows:

$$\begin{aligned} y &= \text{center}(\text{conv}(x, w, \text{full}), n) \in R^n, \\ y &= \text{conv}(x, w, \text{valid}) \in R^{n-m+1}. \end{aligned} \quad (10)$$

Valid is often used in convolution flow, and the above-given functions can be extended to two-dimensional scenes. In practical operation, it needs to be applied to two important parameters: zero filling and step size. Step size refers to the unit data required for the moving position of the filter. The number of weights can be through the convolution layer. Because the weights will be shared, the weights can be greatly reduced in the training process, the amount of data can be improved, and the fitting can be effectively avoided. In the deep learning fusion method, this paper combines a neural network with self-coding network to input supervision information and reduce training data. The self-coding

network used is a refined unsupervised network. The dimension of image data is reduced through network training. The working principle is shown in Figure 1. X represents input data, Z represents low latitude feature information, and Y represents decoded output data.

During the operation of the encoder, the data are mapped to Z through the hidden layer, and the formula is

$$Z = \delta(wx + b), \quad (11)$$

where Z represents the activation function, w represents the convolution weight, and b represents the offset value.

The decoding process is opposite to encoding, and the image is reconstructed by decoding. The self-coding fusion image is shown in Figure 2. The input image is numbered and expressed as I . In this figure, the subject results are divided into three layers, the middle is the fusion layer, and the coding layer covers the convolution layer. The results are similar, and the sizes are all 3×3 . Fill the edge image by mirror filling to avoid shadows on the edge. The convolution kernel is used to extract all the input image feature information, and the addition strategy is used to fuse the feature information. The feature information output from the fusion layer enters the decoding layer, which contains three convolution layers with the same result, all of which are 3×3 . Using this method to reconstruct the image can ensure the integrity of feature information.

4. Result Analysis and Discussion

4.1. Experimental Simulation. In measuring the effect of network training, the loss function is used to compare and analyze the difference between input data and output image data to calculate the error and update it continuously. In the evaluation gap, the detailed parameters of SSIM results are selected for judgment and the contrast, structure, and brightness are considered. These three indicators are fully combined with the characteristics of the human visual senses. SSIM calculation formula is as follows:

$$\text{SSIM}(x, y) = \frac{(2u_x u_y + c1)(2\delta_{xy} + c2)}{(u_x^2 u_y^2 + c1)(\delta_x^2 + \delta_y^2 + c2)}. \quad (12)$$

In the analysis of image similarity, the value is between 0 and 1. The greater the value, the higher the similarity. The loss function can be expressed as follows:

$$\begin{aligned} \text{LOSS} &= 1 - \text{SSIM} \\ &= 1 - \frac{(2u_x u_y + c1)(2\delta_{xy} + c2)}{(u_x^2 u_y^2 + c1)(\delta_x^2 + \delta_y^2 + c2)}. \end{aligned} \quad (13)$$

In image retrieval, the traditional KNN algorithm is combined with self-coding network and compared with a database. Firstly, the self-coding network is used to retrieve the data, then converge, extract the feature information, calculate the cosine distance between the image and the database, and take the closest image as the reference image.

In this analysis, K takes 5 to get the closest image. The cosine similarity is used to calculate the difference between different horizontal images. This calculation method pays more attention to the vector angle, not just the distance. The calculation formula is

$$\cos\left(\theta = \frac{ab}{\|a\| \times \|b\|}\right), \quad (14)$$

where a, b is the cosine value. When calculating, the cosine value close to 1 indicates high similarity.

In order to test the operation efficiency of the fusion method, different images are restored. Lena gray image with 256 pixels is selected as the test image $\times 256$, convolute the test image with the symmetric Gaussian low-pass filter and take factor 4 as the sampling to obtain different LR frames and increase Gaussian noise. In order to ensure the accuracy of the test, the motion error is increased. At the same time, the execution time and segmentation threshold of the algorithm are tested. Analyzing the random decline factor of the network, selecting a typical learning rate for evaluation and setting a more reasonable learning rate can get better analysis results and faster speed. However, if the value is set too high, the random factor will drop too fast and cross the global situation. If the value is set too small, the gradient changes slowly, and the convergence speed of the network decreases greatly, or even cannot converge. Therefore, different learning rates are set for evaluation in this analysis. The learning rate is set at $4e-5$, $4e-6$, and $4e-7$, respectively.

4.2. Result Analysis. The simulation results of super-resolution feature acquisition are shown in Figure 3. It can be seen from the figure that the algorithm proposed in this paper has a better effect. In terms of execution speed, the time is not prolonged compared with the traditional algorithm, which can meet the needs of real-time image processing, and the segmentation effect is better than the traditional algorithm.

The simulation analysis results of the network random descent factor are shown in Figure 4. It can be seen from the figure that when the learning rate is $4e-5$, the overall change of the function is relatively slow; when the learning rate is set to $4e-7$, the function decreases greatly and cannot converge; only when the learning rate is set to $4e-6$, the convergence speed and the decrease of loss function can get good results. Therefore, if the learning rate exceeds $4e-6$, it will cause the function oscillation, which is smaller than him, and the convergence speed will decrease. In the fusion network, the decoding and encoder adopt the learning rate, and the ReLU function is selected as the activation function to improve the convergence stability.

The method proposed in this paper preserves the image feature information in the fusion algorithm, removes noise, and is easier to observe. In the image selection, the dune image, camp image, and Marne image are selected for fusion. At the same time, the fusion algorithm proposed in this paper is compared with the pyramid fusion algorithm, NSST fusion algorithm, DWT fusion algorithm, and GFF fusion

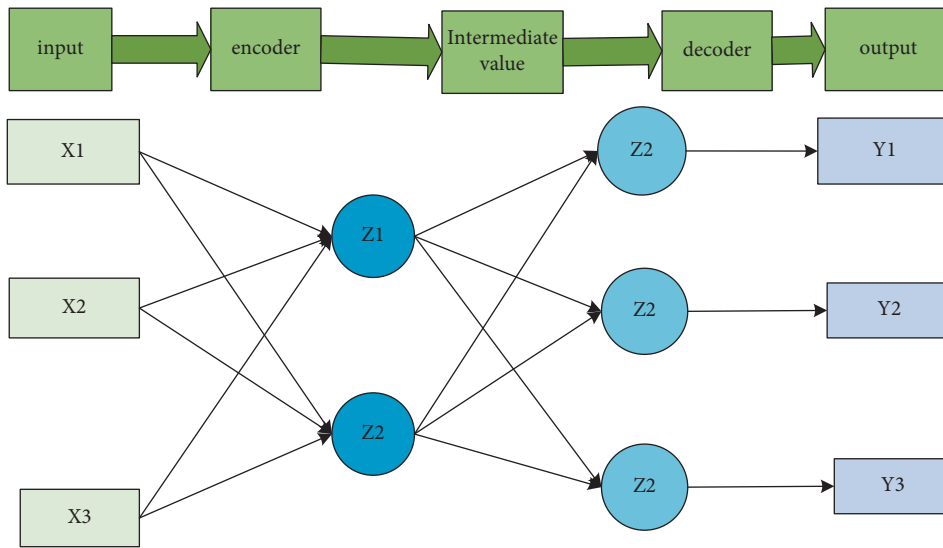


FIGURE 1: Working principle of self-encoder.

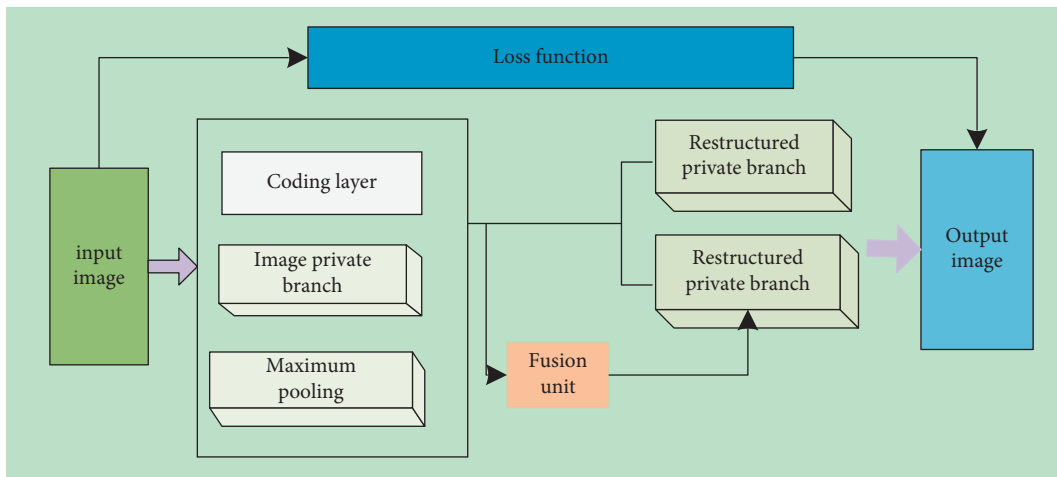


FIGURE 2: Self-coding fusion training network.

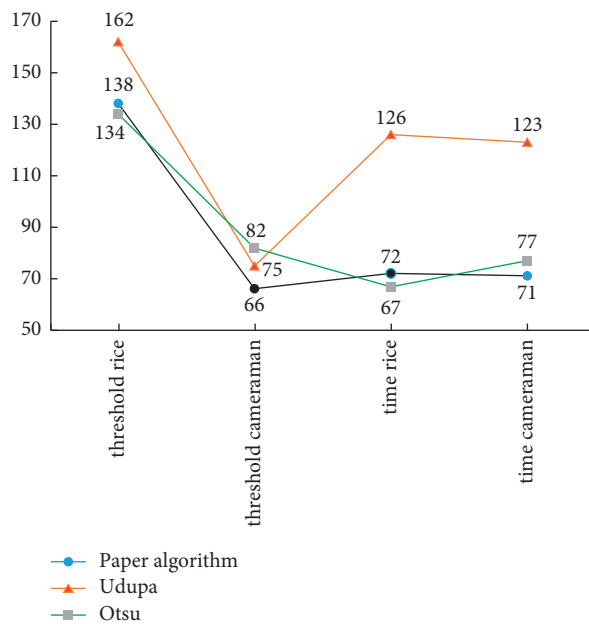


FIGURE 3: Comparison of threshold selection and execution time of different execution methods.

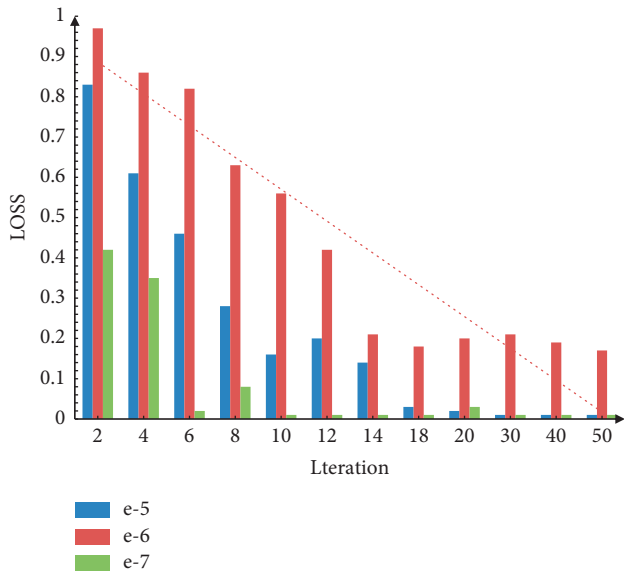


FIGURE 4: Influence of different learning rates on the loss function.

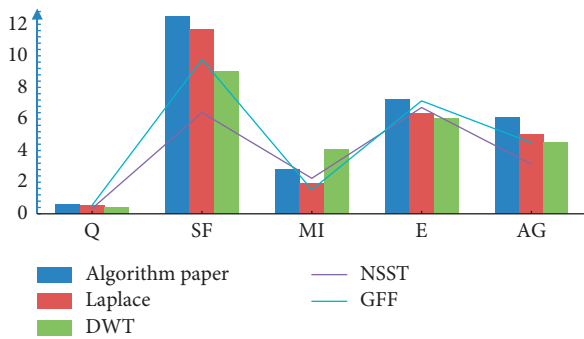


FIGURE 5: Simulation analysis and comparison results of image Camp images.

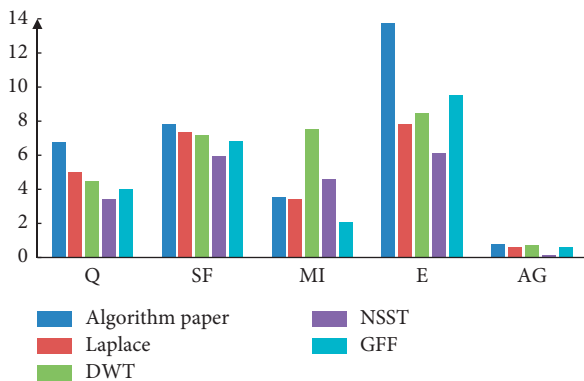


FIGURE 6: Simulation analysis and comparison results of image Kaptein images.

algorithm. The results of the camp image simulation analysis are shown in Figure 5. It can be seen from the figure that except Mi index, other indexes of the algorithm proposed in this paper are due to other algorithms.

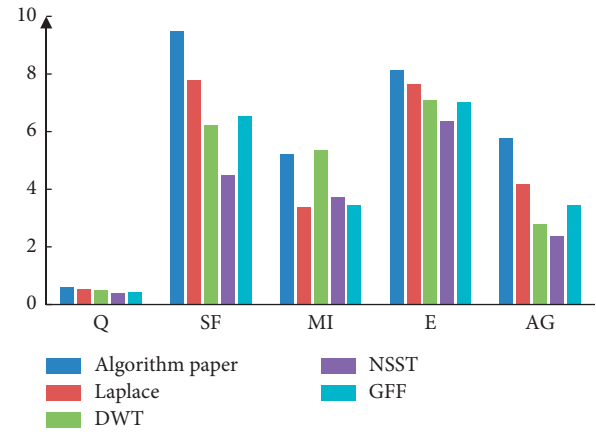


FIGURE 7: Simulation analysis and comparison results of image Marne images.

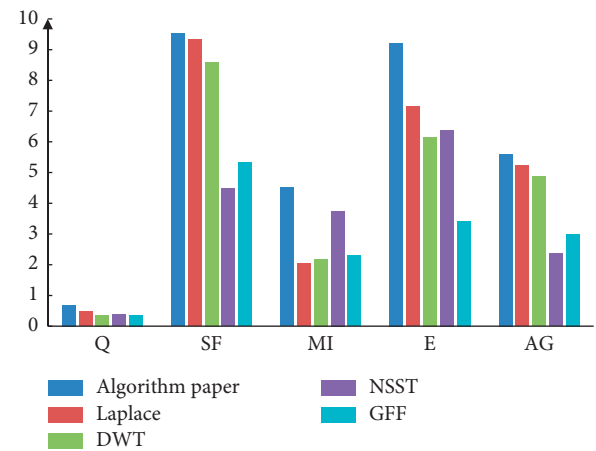


FIGURE 8: Simulation analysis and comparison results of image Dune images.

The kaptein image simulation results are shown in Figure 6. It can be seen from the figure that the MI index is basically the same as other algorithms, and other indexes are better than the other four algorithms.

The Marne image simulation analysis results are shown in Figure 7. It can be seen from the figure that the superior performance of index e is not obvious, and the other four objective indexes show obvious advantages.

The results of the dune image simulation analysis are shown in Figure 8. It can be seen from the figure that the superior performance of the SF index is not obvious, and the other four objective indexes show obvious advantages.

5. Conclusion

At present, remote sensing meteorological, medical images, and other professional images have high requirements for resolution, but due to subjective factors and equipment factors, the insufficient image resolution affects the normal use. Based on this, this paper studies the ground resolution image fusion algorithm based on depth learning. In order to

reduce the computational complexity, it is necessary to ensure that the amount of calculation can be reduced in each iteration. LR frames can not be calculated separately in each iteration, but all values can be found in the image through the difference, which can meet the relevant computational requirements. The improved alexnet model is used to realize the classification and acquisition of images, and the convolution neural network is combined to realize image fusion to obtain high robustness. In order to analyze the effectiveness of the fusion algorithm, the objective indexes are comprehensively evaluated, and different images such as the dune image, camp image, and Marne image are selected for comprehensive judgment. The indexes such as gradient change and information entropy show their advantages. It should be pointed out that some images, such as infrared and visible images, are difficult to fuse due to the lack of reasonable reference images, and image fusion involves multispectral images, so the adaptability of the algorithm needs to be further improved.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This study was supported by “Xi’an Peihua College-level subject,” Research on UAV visual tracking technology,” under Grant Number PHKT18104” and Xi’an Peihua College 2018 School-level Education and Teaching Reform Research Project “Research on the Reform of Computer Major Teaching System Based on Interdisciplinary Integration,” under Grant Number PHZ1806.

References

- [1] M. A. Herrera-Arellano, H. Peregrina-Barreto, and I. Terol-Villalobos, “Visible-NIR image fusion based on top-hat transform,” *IEEE Transactions on Image Processing*, vol. 10, no. 99, p. 1, 2021.
- [2] X. Wang, S. Bai, L. Zhi, Y. Sui, and J. Tao, “The PAN and MS image fusion algorithm based on adaptive guided filtering and gradient information regulation,” *Information Sciences*, vol. 9, no. 545, pp. 381–402, 2021.
- [3] B. Xiao, B. Xu, X. Bi, and W. Li, “Global-feature encoding U-net (GEU-Net) for multi-focus image fusion,” *IEEE Transactions on Image Processing*, vol. 30, pp. 163–175, 2021.
- [4] C. Chahine, C. Vachier-Lagorre, Y. Chenoune, R. El Berbari, Z. El Fawal, and E. Petit, “Information fusion for unsupervised image segmentation using stochastic watershed and Hessian matrix,” *IET Image Processing*, vol. 12, no. 4, pp. 525–531, 2018.
- [5] Y. Liu, X. Chen, Z. Wang, and R. K. Ward, “Deep learning for pixel-level image fusion: recent advances and future prospects,” *Information Fusion*, no. 42, pp. 158–173, 2018.
- [6] C. Lee and J. Oh, “Rigorous Co-registration of KOMPSAT-3 multispectral and panchromatic images for pan-sharpening image fusion,” *Sensors*, vol. 20, no. 7, pp. 152–156, 2020.
- [7] A. Asadi and M. Ezoji, “Multi-exposure image fusion via a pyramidal integration of the phase congruency of input images with the intensity-based maps,” *IET Image Processing*, vol. 14, no. 1, pp. 1147–1149, 2020.
- [8] A. Gómez-Sánchez, M. Marro, and M. Marsal, “3D and 4D image fusion: coping with differences in spectroscopic modes among hyperspectral images,” *Analytical Chemistry*, vol. 1, no. 10, p. 780, 2020.
- [9] P. Geng and J. Liu, “An effective multifocus image fusion method using guided filter,” *Industrial Robot: The International Journal of Robotics Research and Application*, vol. 46, no. 3, pp. 369–376, 2019.
- [10] C. B. S. Piqueras, C. B. Leites, C. Beleites et al., “Handling different spatial resolutions in image fusion by multivariate curve resolution-alternating least squares for incomplete image multisets,” *Analytical Chemistry*, vol. 90, no. 11, pp. 6757–6765, 2018.
- [11] A. R. Cabazos-Marín and J. Álvarez-Borrego, “Automatic focus and fusion image algorithm using nonlinear correlation: image quality evaluation,” *Optik*, vol. 164, no. 101, pp. 224–242, 2018.
- [12] S. S. Gupta, S. Hossain, and K. D. Kim, “HDR-like image from pseudo-exposure image fusion: a genetic algorithm approach,” *IEEE Transactions on Consumer Electronics*, vol. 10, no. 99, p. 1, 2021.
- [13] L. Ren, Z. Pan, J. Cao, J. Liao, and Y. Wang, “Infrared and visible image fusion based on weighted variance guided filter and image contrast enhancement,” *Infrared Physics & Technology*, vol. 114, no. 2, Article ID 103662, 2021.
- [14] X. Li, F. Zhou, and H. Tan, “Joint image fusion and denoising via three-layer decomposition and sparse representation,” *Knowledge-Based Systems*, vol. 224, no. 1, Article ID 107087, 2021.
- [15] S. Bhat and D. Koundal, “Multi-focus image fusion techniques: a survey,” *Artificial Intelligence Review*, vol. 54, no. 6, pp. 1–53, 2021.