Adaptive Weight Fusion Algorithm of Infrared and Visible Image Based on High-Frequency Domain CNN

Chuanyun Wang1*, Guowei Yang1, Dongdong Sun1, Jiankai Zuo2 and Ershen Wang3

1 College of Artificial Intelligence, Shenyang Aerospace University, Shenyang, China
2 Department of Computer Science, Tongji University, Shanghai 201804, China
3 College of Electronics and Information Engineering, Shenyang Aerospace University, Shenyang 110136, China
*Corresponding author’s e-mail: wangcy0301@sau.edu.cn

Abstract. Aiming at covering the shortage of single source sensor imaging and improving the contrast between the target and the background in image, this paper proposes an adaptive weight fusion algorithm of infrared and visible image based on a High-frequency Domain Convolutional Neural Network (HDCNN). Firstly, the high and low frequency components of the original image are obtained by using the Daubechies wavelet transform, and then a high-frequency domain convolutional neural network which can detect the frequency information ratio of infrared and visible light in the high-frequency subband is trained. Secondly, the network is used to perform adaptive weight fusion for the high frequency components and regional energy is used for fusion of the low frequency components. Finally, the fusion image is obtained by inverse wavelet transform. A large number of experiments have proved that the algorithm in this paper has a greater improvement over similar comparison algorithms in objective evaluation metrics such as standard deviation, spatial frequency and average gradient. The algorithm enhances the contrast between the target and the background in the fusion image, and enriches the characteristic information of the target itself.

1. Introduction
Infrared and visible image fusion technology can integrate the information of infrared and visible image into one image, making up for the deficiency of single source sensor imaging. Therefore, it has great application value in military reconnaissance, social security and other fields [1-4].

Image fusion methods mostly apply multi-scale transformation tools to complete fusion on the transformation domain [5-7]. An image fusion algorithm based on Laplace transform and sparse representation is proposed in [8]. The overall fusion effect is good, but the distinction between the target and the background of the fusion image is not great. In recent years, with the rise of neural network, infrared and visible image fusion through neural network has become a research hotspot [9-12]. Some scholars proposed to combine neural network with multi-scale transformation to achieve image fusion.[13,14] For example, [15] decomposed the source image by non-subsampled shearlet transform (NSST), and realized hierarchical fusion by convolutional neural network. In general, most of the image fusion algorithms based on CNN and multi-scale transform adopt the original image training neural network. The neural network trained by the original image dataset is difficult to extract and learn the
image target information, and can not effectively improve the target and background contrast of the fused image.

To solve the above problems, this paper proposes an adaptive weight fusion algorithm of infrared and visible image based on high-frequency domain convolutional neural network (HDCNN). The algorithm proposes HDCNN, which can extract and detect the characteristic information of high frequency target and improve the contrast between target and background.

2. Infrared and Visible Image Adaptive Weight Fusion Algorithm in Frequency Domain

2.1. Image Frequency Domain Transform Model
In the frequency domain transformation of images, this paper adopts the fourth-order Daubechies wavelet transform model [16]. Compared with haar wavelet transform, Daubechies wavelet transform considers the overlapping window, which is more conducive to the comprehensive acquisition of high frequency information. The original image can be decomposed into high frequency and low frequency components by Daubechies wavelet transform.

The fourth-order Daubechies wavelet basis is defined as \( (dwt) \), high-pass filter and low-pass filter are \( h_f \) and \( l_f \), respectively. The high-frequency components of the image decomposed by the fourth-order Daubechies wavelet transform have four sub-bands of different sizes. The high-frequency subbands corresponding to the infrared and visible light images are denoted as \( H_I^i, H_V^i \), \( i = 1,2,3,4 \), where \( H_I^i \) represents the first subband in the high-frequency component of the infrared light image. See Eq(1) for high frequency component acquisition.

\[
H_I = P_I(dwt)h_f, \ H_V = P_V(dwt)h_f
\]  
(1)

Where \( P_I \), \( P_V \) stands for infrared and visible image. The low frequency components \( L_I \) and \( L_V \) are obtained by the same method.

\[
L_I = P_I(dwt)l_f, \ L_V = P_V(dwt)l_f
\]  
(2)

2.2. High Frequency Component Fusion Based on High-Frequency CNN Model
The high frequency component mainly reflects the target characteristic information of the image. In order to achieve the adaptive weight fusion of the high-frequency components, a high-frequency domain convolutional neural network (HDCNN) is designed to detect the frequency information ratio between infrared and visible light in the high-frequency subbands.

As shown in Figure 1, the input data in the HDCNN is a high-frequency subband matrix with a size of 256\( \times \)256. The network uses two convolution and pooling operations to expand the features for full connection, and makes the data output between (0,1) through Softmax classifier. This data is the ratio of infrared to visible light frequency.

In the selection and generation of training set, this paper selects 100 groups of registered infrared visible images in TNO dataset, and expands the dataset to 1000 groups by rotation, translation and scaling. 4000 groups of high frequency subbands are obtained by fourth-order Daubechies wavelet transform as the final dataset. The high frequency subbands in the dataset are uniformly 256\( \times \)256, 10% of which are selected as test data and the remaining 90% as training data.
In this paper, the network model is built through Keras, and the cross entropy function is used as the loss function of the model, and the loss function is reduced by the way of random descent. The cross entropy function $C$ is shown in the Eq.(3).

$$C = \frac{1}{N} \sum_i C_i = \frac{1}{N} \sum_i [-y_i \times \log(p_i) + (1 - y_i) \times \log (1 - p_i)]$$  \hspace{1cm} (3)

Where, $y_i$ represents the label of sample $i$ of high frequency subband. The high frequency subband of infrared image is 1, and the high frequency subband of visible image is 0. Set the learning rate to $10^{-4}$. The data set is put into the network for training.

Taking the first group of high-frequency subbands in the high-frequency component as an example, the fusion rules are derived as follows.

**Step 1** Wait until the high-frequency subbands are fused by Eq (4). Where, $W_f$ is the weight of fusion, and the initial $W_f$ is a random number between (0.8,1).

$$H_f^1 = W_f \times H_{IR}^1 + (1 - W_f)H_{VIS}^1$$  \hspace{1cm} (4)

**Step 2** The frequency ratio $R_{S/V}$ of the infrared and visible subband is obtained by inputting $H_f^1$ into the model $HDCNN$.

$$R_{S/V} = HDCNN(H_f^1)$$  \hspace{1cm} (5)

**Step 3** Adjust $W_f$ with forward feedback to obtain a new fusion weight $W_{new}$.

$$W_{new} = W_f - W_f R_{S/V - T_0}^1, T_0 = \frac{T_{max} - T_{min}}{2}$$  \hspace{1cm} (6)

Where, the ratio interval of ideal fusion subbands is set as $T = [T_{min}, T_{max}]$, and $T_0$ is the optimal specific gravity.

**Step 4** Obtains $R_{S/V}$ under the condition of $W_{new}$ by Eq.(4) and Eq.(5).

**Step 5** Sets the maximum number of iterations $MAX_{iter}$, repeating steps 3 and 4. If $R_{S/V} \in T$, the iteration is stopped. The $W_{new}$ corresponding to the acquisition of $R_{S/V}$ is recorded as the actual optimal fusion weight $W_o$. Otherwise, $W_{new}$ in the last iteration is selected as the actual optimal fusion weight $W_o$.

**Step 6** Fusion of the first group of high-frequency subbands is completed by employing the actual optimal fusion weight $W_o$.

$$H_f^1 = W_o \times H_{IR}^1 + (1 - W_o)H_{VIS}^1$$  \hspace{1cm} (7)

The fusion Eq (8) of the high frequency component is obtained by generalizing step1 ~ step 6 to all the subbands of the high frequency component.

$$H_f = \{W_o^i \times H_{IR}^i + (1 - W_o^i)H_{VIS}^i | i \in Z, 1 \leq i \leq 4\}$$  \hspace{1cm} (8)

2.3. Low-Frequency Component Fusion Based on Regional Energy

The low frequency component mainly reflects the background texture information of the image. In order to retain the background texture information as much as possible, this paper proposes a fusion rule of low frequency component of infrared and visible image based on regional energy method.

**Step 1** Calculate the regional energy $L_{VIS}^{ARE(i,j)}$ of the low frequency component $(i,j)$ of the visible image.

$$L_{VIS}^{ARE(i,j)} = \sum_{m=1}^{m=n} w(m,n) \times |L_{VIS}(i + m, j + n)|$$  \hspace{1cm} (9)

Where, $m=n=1$, and $w$ is a region of $3 \times 3$ in size. $w(m,n)$ is the weight of the point $(m,n)$.

**Step 2** Confirm the fusion weight by regional energy.

$$W_{VIS}^{(i,j)} = \frac{1}{(2m+1)(2n+1)} L_{VIS}^{ARE(i,j)}, W_{IR}^{(i,j)} = 1 - W_{VIS}^{(i,j)}$$  \hspace{1cm} (10)

Among them, $W_{IR}^{(i,j)}$ and $W_{VIS}^{(i,j)}$ are the fusion weights of low frequency components of infrared and visible images at $(i,j)$ points, respectively. $W_{IR}$ and $W_{VIS}$ are the final fusion weight maps.

**Step 3** The fused low-frequency component $L_f$ is obtained through the weight map.

$$L_f = W_{IR}L_{IR} + W_{VIS}L_{VIS}$$  \hspace{1cm} (11)
2.4. Design and Implementation of Fusion Algorithm
In the algorithm proposed in this paper, the high and low frequency components of infrared and visible images are obtained through wavelet transform. The product HDCNN is applied to the adaptive fusion of high frequency components and the region energy is used to the weighted fusion of low frequency components. The fusion image is obtained by inverse wavelet transform of the fused high and low frequency components. See Figure 2 for the specific operation framework.

3. Experimental Results and Analysis

3.1. Parameter Setting
In order to verify the performance of the algorithm in this paper, 5 groups of representative infrared and visible images were selected from the TNO dataset as test cases. In order to unify the comparison, this paper unifies the image into a 256×256 size, 8bit single-channel grayscale image in the experiment. The experimental platform was PyCharm platform under 64-bit Windows10. In order to compare and analyze, the fusion algorithm based on Laplace transform and sparse representation in [8], the fusion algorithm based on convolutional neural network and image pyramid multi-scale transformation in [12] and the fusion algorithm based on convolutional neural network and NSST in [15] are replicated in this paper as the contrast algorithm. These algorithms are denoted as REF_8, REF_12, REF_15, and the algorithm in this paper is denoted as Proposed.

The training results of HDCNN model in this paper are shown in Figure 3. After 500 iterations of training, the accuracy and cross entropy of the model tended to be at 0.99 and 0.01 and remained stable. At this time, the model has high recognition accuracy and stability.

3.2. Evaluation Criterion
With the purpose of evaluating the performance of the algorithm more objectively, four commonly used image quality evaluation metrics were selected in this paper. The four evaluation indexes are as follows: Standard deviation (STD), Spatial Frequency (SF), Average Gradient (AG) and Information Entropy (IE) [17]. Among them, STD can macroscopically reflect the contrast between the image target and the
background. The larger the value of STD, the greater the contrast between the image target and the background is. SF can reflect the sharpness of the image in a positive correlation. AG can be positively correlated to reflect the clarity of image background details. IE is the information entropy of the image, which can reflect the amount of information carried by the image.

3.3. Subjective Analysis
The fusion results of the four groups of images are shown in Figure 4. The images fused by REF_8 algorithm basically reflect the target features of infrared images and the background information of visible images. But the gradient of the target against the background is not obvious. The REF_12 algorithm is better than REF_8 in the target feature information. However, it does not highlight the target feature information of the fusion image. REF_15 is a limited improvement over REF_12. The algorithm in this paper significantly improves the contrast between the target and the background of the fused image. For example, in the test of group (c), the image fused by the algorithm in this paper can clearly depict the characteristic information of the target figure, and the edge contour is relatively clear.

3.4. Objective Analysis
The four groups of test images were quantitatively evaluated, and the statistical results of specific performance indicators were shown in Table 1. In terms of evaluation metrics, the algorithm proposed in this paper surpasses the other three algorithms in STD, SF and AG, and also has certain advantages in
IE. The evaluation metrics objectively reflects that the proposed algorithm has a relatively good performance in enhancing the contrast between the target and the background. The fused image carries more information and has higher definition.

To further verify the robustness of the algorithm, this paper takes the group (c) of test images as an example to test the anti-noise performance of the algorithm. The original images tested in group (c) were successively added with Gaussian noise (0~10dB, with a step size of 2dB). The images interfered by noise were fused, and the performance of the fused images under different noise interference conditions was recorded and counted. The statistical results are shown in Figure 4. As shown in Fig. 4, Under different noise conditions, the performance of the proposed algorithm is better than that of comparison algorithms. This proves that the proposed algorithm has excellent robustness.

| Images | Evaluation metrics | REF_8 | REF_12 | REF_15 | Proposed |
|--------|--------------------|-------|--------|--------|----------|
| (a)    | STD                | 36.360| 37.140 | 43.210 | 45.230   |
|        | SF                 | 5.5575| 6.6754 | 6.8823 | 7.1995   |
|        | AG                 | 2.7152| 3.4678 | 3.5718 | 5.2094   |
|        | IE                 | 6.8889| 6.8632 | 6.9211 | 7.1397   |
|        | STD                | 16.970| 20.680 | 24.790 | 28.070   |
| (b)    | SF                 | 5.5410| 6.8142 | 6.9993 | 10.8994  |
|        | AG                 | 2.8863| 3.9051 | 4.0014 | 6.0469   |
|        | IE                 | 6.0993| 6.2629 | 6.4827 | 6.7185   |
|        | STD                | 28.580| 28.440 | 34.920 | 38.970   |
| (c)    | SF                 | 5.1749| 6.9406 | 7.3131 | 11.0546  |
|        | AG                 | 2.5193| 3.6890 | 3.8566 | 5.8205   |
|        | IE                 | **6.8974**| 6.4967 | 6.8084 | 6.8623   |
|        | STD                | 16.600| 14.530 | 17.130 | **18.280**|
| (d)    | SF                 | 1.9655| 3.1902 | 3.2400 | **6.0380**|
|        | AG                 | 0.9363| 1.8345 | 1.8529 | **3.4716**|
|        | IE                 | 5.1823| 5.4915 | 5.6400 | **5.9204**|

4. Conclusion
In this paper, an adaptive weight fusion algorithm of infrared and visible image based on a high-frequency domain convolutional neural network is realized. Firstly, the algorithm models the Daubechies wavelet transform to obtain the high and low frequency components of the original image. Then, the high-frequency components are fused with adaptive weights applying the trained high-frequency domain convolutional neural network. Low-frequency components are fused with weights based on regional energy. Finally, the image fusion is completed through Daubechies wavelet inverse transform. Experimental results show that the proposed algorithm has great advantages in STD, SF, AG and other indicators, enhances the contrast between the fusion image and the background, and has good anti-noise performance. In the future research, we will further improve the performance of the algorithm under the IE.

Acknowledgments
This research was supported by National Natural Science Foundation of China (Grant No. 61703287), 2021 Provincial College Student Innovation and Entrepreneurship Training Program (Grant No.D202104031404133082), Liaoning Provincial Key R&D Program of China (Grant No. 2020JH2/10100045), Talent Project of Revitalization Liaoning of China (Grant No. XLYC1907022) and High-Level Innovation Talent Project of Shenyang of China (Grant No. RC190030).
References

[1] Zhao Ye. Research on target detection method of UAV based on fusion of infrared and visible images. Harbin Institute of Technology, 2019.

[2] Bhavana D, Kumar K K, Tej D R. Infrared and visible image fusion using latent low rank technique for surveillance applications. International Journal of Speech Technology, 2021(2):1-10.

[3] Ren L, Pan Z, Cao J, et al. Infrared and visible image fusion based on weighted variance guided filter and image contrast enhancement. Infrared Physics & Technology, 2021, 114(2): 103662.

[4] J. Zuo, J. Chen, Z. Li, Z. Li, Z. Liu and Z. Han, "Research on Maritime Rescue UAV Based on Beidou CNSS and Extended Square Search Algorithm," 2020 International Conference on Communications, Information System and Computer Engineering (CISCE), 2020, pp. 102-106, doi: 10.1109/CISCE50729.2020.00027.

[5] Shen Ying, Huang Cunhong, Huang Feng, et al. Infrared and visible image fusion: review of key technologies [J/OL]. Infrared and Laser Engineering:1-16[2021-05-02].

[6] Nirmalraj S, Nagarajan G. An adaptive fusion of infrared and visible image based on learning of sparse fuzzy cognitive maps on compressive sensing. Journal of Ambient Intelligence and Humanized Computing, 2019(4).

[7] Ren L, Pan Z, Cao J, et al. Infrared and visible image fusion based on edge-preserving guided filter and infrared feature decomposition. Signal Processing, 2021(2): 108108.

[8] Li X, Zhang X, Ding M. A sum-modified-Laplacian and sparse representation based multimodal medical image fusion in Laplacian pyramid domain. Medical & Biological Engineering & Computing, 2019, 57(12 Suppl).

[9] Tang Chaoying, Pu Shiliang, Ye PengZhao, et al. Fusion of low-illuminance visible and near-infrared images based on convolutional neural network. Acta Optica Sinica, 2020, 40(16):37-45.

[10] Lou Xicheng, Feng Xin. Infrared and visible image fusion in latent low rank representation framework based on convolution neural network and guided filtering. Acta Photonica Sinica, 2021,50(03):188-201.

[11] Raza A, Liu J, Liu Y, et al. IR-MSDNet: Infrared and Visible Image Fusion based on Infrared Features & Multiscale Dense Network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, PP(99):1-1.

[12] Liu Y, Chen X, Cheng J, et al. Infrared and visible image fusion with convolutional neural networks. International Journal of Wavelets, Multiresolution and Information Processing, 2018,16(3).

[13] Ren Kan, Zhang Dawei, Wan Minjie, Miao Xin, Gu Guohua, Chen Qian. An Infrared and Visible Image Fusion Method Based on Improved DenseNet and mRMR-ZCA. Infrared Physics & Technology, 2021 (prepublish).

[14] Liu D, Zhou D, Nie R, et al. Infrared and visible image fusion based on convolutional neural network model and saliency detection via hybrid l0-l1 layer decomposition. Journal of Electronic Imaging, 2018, 27(6):1.

[15] Huan Kewei, Li Xiangyang, Chao Yutong, et al. Infrared and visible image fusion with convolutional neural network and NSST [J]. Infrared and Laser Engineering, 2021:1-7.

[16] Sun Shiyang. Quantum Image Watermarking Algorithm Based on Daubechies D(4) Wavelet Transform. Advances in Applied Mathematics, 2020, 09(12): 2292-2300.

[17] Evaluation of Spatial and Transform Fusion methods for Medical Images using Normalized Non-Reference Quality Metrics. International Journal of Computer Applications, 2016, 143(13):21-28.