A New Image Fusion Method Based on PCANet

Ying Zhu, Min Li*, Feng Chen, Xiaoxiang Qi and Xiaohua Qiu
Xi’an Research Institute of High-tech, Xi’an, China
*proflimin@163.com

Abstract. In recent years, a simpler principal component analysis network (PCA Net) has been proposed. The network needs fewer parameters to be adjusted, which can solve the problem of long training time. However, because of the infrared and visible images obtained under different environments, the gray distribution of the images is different. If only one base layer fusion strategy is used, the generality of the algorithm may be reduced. To solve this problem, this paper investigates two different base layer fusion strategies. First, the visible dataset is used to train PCANet; second, based on the alternating guided filter to single-scale decomposition of the image to obtain the base layer and the detail layer, then use the convolution kernel of the trained first layer network to extract the depth features of the source image, process it by kernel-norm to get the activity level map, and then construct the detail layer fusion weights; The base layer uses the two methods of maximum normalization of gray difference and weighted average to obtain the preliminary fusion map of the two base layers. By analyzing the methods in this chapter and other methods in simulation experiments, the QNCIE method in this chapter is 1.25% higher than the other methods. The QFMI is 150% higher than other methods. QP, QY and QCB are higher than other methods.

1. Introduction
Images play a very important role in people's lives, vivid image information is easier to remember than text information. Image fusion is an important part of image processing technology. It is a combination of useful information in pictures with certain rules to obtain images more suitable for visual information reception [1-3]. The fusion feature information based on deep learning construction fusion methods (such as VggML) [4] may be richer. However, classic deep learning networks, such as convolutional neural networks (CNN) [5-7] and residual networks, have problems such as long training time, difficulty in parameter adjustment, and high requirements on the experimental hardware environment. Recently, a simpler principal component analysis network (PCANet)[8-9] has been proposed.

This paper proposes an infrared and visible light fusion method based on improved PCANet. Through the simulation experiment analysis of the method of this paper and other methods, the performance of the method of this paper is evaluated.

2. Network Design

2.1 Framework Design
The network framework adopted in this article is a two-layer PCA Net work flow chart. First use the PCA filter of the first layer to convolve the image to extract features, the number of feature maps is the same as the number of filters. Then use the PCA filter of the second layer to convolve each feature map extracted from the first layer to obtain multiple more abstract feature maps.

2.2 Fusion Framework
Figure 1 shows the image fusion framework based on AGF[10] and PCA Net. It can be seen from the figure that the fusion of the detail layer extracts the feature image of the source image by using the trained PCA filter kernel. Use L1-norm to process it to get the activity level graph, and then use the average operator to get the final activity level graph, and then calculate the initial weight graph by soft-
max function. And calculate the initial activity level map to obtain the weight mapping matrix of detail layer fusion, and propose a new basic layer fusion strategy.

![Figure 1. The AGF and PCANet image fusion framework](image)

![Figure 2. the Detail layer fusion strategy diagram](image)

### 2.3 Convergence strategy

#### 2.3.1 Detail layer fusion. For the fusion method of the detail layer, at present, most of the saliency map detection algorithms are used to extract the salient features of the detail layer, and the weight mapping matrix is used to construct the fusion. The detail layer fusion strategy diagram of this paper is shown in Figure 2.

The number of PCA Net's first and second layer filters is 8. Suppose the source image is \( I_i \) \((i = 1, 2)\), the process of extracting the feature map using the convolution kernel of the trained first layer network is shown in equation (1)

\[
\phi_i = I_i \ast W_i^1
\]  

In the formula. \( W_i^1 \) represents the convolution kernel of the first layer of PCA Net. \( \phi_i \) represents the extracted \( i \)th feature map. \( L \) represents the number of extracted feature maps.

This paper introduces an improved L1 norm and average to process each feature map to obtain an activity level map, as shown in equation (2)

\[
\phi_i = \frac{1}{L} \sum_{j=1}^{L} \| \phi_j^i - {(x+t), (y-t)} \|, \quad (t=2)
\]

Construct the weight mapping matrix of detail layer fusion from \( L \), and fuse the detail layer to obtain the \( L \)th detail layer fusion Figure, as shown in equations (3) and (4):
\[ W_{IR}^j = \frac{\varphi_1^j}{\varphi_1 + \varphi_2} \] (3)

\[ F_D^j = W_{IR}^j D_{IR} + (1 - W_{IR}^j) D_{VIS} \] (4)

It is known from formula (4) that after the above steps, 1L detail layer fusion images can be obtained. The final detail layer fusion map as shown in (5)

\[ F_D = \max(F_D^j, F_D^j) \quad (i, j \in \{1, 2, ..., L\}) \] (5)

Finally, by adding the fused base layer BF and detail layer DF, the fused image F is obtained, as shown in equation (6):

\[ F = F_B + F_D \] (6)

2.3.2 Basic layer fusion. First, take absolute values of \( B_{IR} \) and \( B_{VIS} \) respectively. According to the principle that the greater the gray value, the more significant the target infrared characteristic. Find the difference between the two and remove the point where the pixel value is negative. The process is shown in formula (7).

\[ R = \max(|B_{IR}| - |B_{VIS}|, 0) \] (7)

In the pixel matrix R, the grayscale difference of different pixels is large, so the maximum normalization operation is introduced to obtain the weight mapping matrix of the infrared base layer fusion. As shown in formula (8).

\[ W_{IR} = W / \max(R) \] (8)

Use the obtained infrared weight map to make the first fusion image of the base layer, the process is as shown in formula (9)

\[ F_{B_1} = W_{IR} B_{IR} + (1 - W_{IR}) B_{VIS} \] (9)

This chapter uses a weighted average fusion strategy to obtain the second basic layer fusion graph, as shown in equation (10):

\[ F_{B_2} = 0.5 \times B_{IR} + 0.5 \times B_{VIS} \] (10)

Figure 3 gives the base layer images of infrared and visible light 0.5 weight and the base layer fusion image obtained by the method in this paper. It can be seen from the figure that the basic layer fusion method in this paper can effectively retain the infrared characteristics of the target, such as the infrared characteristics of pedestrians, and can retain the contrast of visible light, such as billboards.
3. Results and Discussion

3.1 Dataset and Parameters
The dataset used in this article is the Thermal World data set produced by Vladimir V. Kniaiz. The dataset contains registered visible and long-wave infrared images. There are 1568 images in the Thermal World dataset, 10 categories including people, cars, trucks, vans, buses, buildings, cats, dogs, trams, boats. 100 samples were randomly selected from the Thermal World data set as test data. All image sizes are reset to 375 × 500. The parameters of PCANet are set as follows: the number of network layers is set to 2, the number of filters of both layers is 8, and the filter size is set to \( k_1 \times k_c = 5 \).

3.2 Comparison Results

3.2.1 Experimental conditions. The operating environment of the experiment is: CPU Intel Core i7, main frequency 3.6GHz, memory 4GB, video memory 2GB, 64-bit Windows 7 operating system, Matlab2017a. In this chapter, the fusion method based on cross bilateral filter (CBF), the image fusion method based on NSCT, Two-scale image fusion method based on saliency detection (TSIFVS), Image fusion method (WLS) based on visual saliency map and least squares optimization [10], fusion method based on VGG19 (VggML) are selected.

3.2.2 Subjective visual analysis. To verify the fusion effect of the method in this paper, six groups of source images after strict registration were selected for fusion.
In the infrared image, the brightness of the lake surface is strong, but the vegetation near the water is blurred. In the visible light image, the lake surface is dark, but the vegetation and the ground of the lake are in good contrast, and the details are rich. Figure 4 (c) shows the fusion result of the CBF method. The fusion effect of the lake surface is poor, and this method generates more false information; Figure 4 (d) is the result of the NSCT method, which effectively integrates the infrared brightness characteristics of the lake surface, but the brightness of the lakeside vegetation is weak; Figure 4 (e) is the fusion result of TSIFVS method, the subjective visual fusion quality of this method is equivalent to that of NSCT method; Figure 4 (f) is the WLS method. Each target in the scene is fuzzy, which affects the recognition analysis; Figure 4 (g) is the result of VggML method, which effectively integrates the feature information of infrared and visible light; Figure 4 (h) is the fusion result of the method in this paper. There are many infrared brightness features on the lake surface, and the visible details of vegetation are abundant.

3.2.3 Analysis of objective evaluation indicators. The objective evaluation indicators used in the paper include: nonlinear related information entropy NCIEQ; feature mutual information FMIQ and phase consistency PQ, these two indicators can measure the edge, detail and other feature information contained in the fused image; Local structure similarity YQ; evaluation index CBQ based on human visual perception, which can measure the contrast of fusion results; The larger the value of these indicators, the better the fusion method.
From table 1, we can see the method of this chapter has higher indexes than other methods for all scene image fusion results. This chapter can effectively merge images of different scenes, highlight the target, and retain the detailed information of visible light.

**Table 1.** Comparison of objective evaluation indexes of the simulation experiments by different methods

| Images  | Metrics | CBF   | NSCT  | TSIFVS | WLS   | VggML | Mymethod |
|---------|---------|-------|-------|--------|-------|-------|----------|
| Camp    | $Q_{NCIE}$ | 0.8034 | 0.8033 | 0.8033 | 0.8034 | 0.8036 | 0.8128  |
|         | $Q_{FMI}$  | 0.2282 | 0.2213 | 0.2239 | 0.2311 | 0.2505 | 0.5896  |
|         | $Q_P$      | 0.1223 | 0.3148 | 0.2550 | 0.1965 | 0.2629 | 0.3726  |
|         | $Q_Y$      | 0.6442 | 0.8225 | 0.7858 | 0.6886 | 0.6832 | 0.8703  |
|         | $Q_{CB}$   | 0.5196 | 0.5794 | 0.5591 | 0.4896 | 0.5625 | 0.5597  |
| Smog    | $Q_{NCIE}$ | 0.8028 | 0.8029 | 0.8029 | 0.8032 | 0.8032 | 0.8124  |
|         | $Q_{FMI}$  | 0.2077 | 0.2219 | 0.2200 | 0.2260 | 0.2456 | 0.5966  |
|         | $Q_P$      | 0.1442 | 0.3050 | 0.2559 | 0.2089 | 0.2559 | 0.3843  |
|         | $Q_Y$      | 0.7115 | 0.8656 | 0.8242 | 0.7520 | 0.7311 | 0.8920  |
|         | $Q_{CB}$   | 0.5274 | 0.5996 | 0.5819 | 0.5011 | 0.5741 | 0.5925  |
| Lake1   | $Q_{NCIE}$ | 0.8126 | 0.8035 | 0.8034 | 0.8046 | 0.8044 | 0.8181  |
|         | $Q_{FMI}$  | 0.4961 | 0.2364 | 0.2261 | 0.2991 | 0.2934 | 0.6574  |
|         | $Q_P$      | 0.2682 | 0.2989 | 0.2560 | 0.1993 | 0.2606 | 0.3380  |
|         | $Q_Y$      | 0.8390 | 0.8907 | 0.8234 | 0.7858 | 0.7137 | 0.8956  |
|         | $Q_{CB}$   | 0.6004 | 0.5864 | 0.5764 | 0.3925 | 0.5131 | 0.6188  |
| Tank    | $Q_{NCIE}$ | 0.8063 | 0.8041 | 0.8042 | 0.8043 | 0.8082 | 0.8141  |
|         | $Q_{FMI}$  | 0.3698 | 0.2601 | 0.2631 | 0.2605 | 0.4453 | 0.5725  |
|         | $Q_P$      | 0.2177 | 0.3877 | 0.4014 | 0.0874 | 0.2932 | 0.4607  |
|         | $Q_Y$      | 0.7316 | 0.7507 | 0.7161 | 0.6014 | 0.5926 | 0.8753  |
|         | $Q_{CB}$   | 0.4291 | 0.3263 | 0.3407 | 0.2066 | 0.3710 | 0.4541  |

4. Conclusion
This chapter studies the application of PCANet in image fusion, including discuss the network structure of PCANet, the principle of feature extraction; select the visible light data set to pre-train PCANet. Experiments show that, compared with CBF, NSCT and other methods, The method in this paper can effectively merge images of different scenes. $Q_{NCIE}$, $Q_{FMI}$ and other objective evaluation index values are high. Compared with VggML, a fusion method based on VGG19, the CBQ value fused in this chapter has advantages, and the image contrast is better.

Reference
[1] Ma J, Ma Y, Li C. Infrared and visible image fusion methods and applications: A survey[J]. Information Fusion, 2018: S1566253517307972.
[2] Jin X, Jiang Q, Yao S, et al. A Survey of Infrared and Visual Image Fusion Methods[J]. Infrared Physics & Technology, 2017, 85.
[3] Zhang Y, Zhang L, Bai X, et al. Infrared and visual image fusion through infrared feature extraction and visual information preservation[J]. Infrared Physics & Technology, 2017, 83:227-237.
[4] Liu Y, Chen X, Peng H, et al. Multi-focus image fusion with a deep convolutional neural network[J]. Information Fusion, 2017, 36: 191-207.

[5] Liu Y, Chen X, Ward R K, et al. Image fusion with convolutional sparse representation[J]. IEEE signal processing letters, 2016, 23(12):1882-1886.

[6] Li Yong, Zeng Jiabei, Shan Shiguang, et al. Occlusion aware facial expression recognition using CNN with attention mechanism[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2439—2450.

[7] SHAKEEL M S, LAM K M. Deep—feature encoding—based discriminative model for age—inVariance face recognition[J]. Pattern Recognition,2019,93: 442—457.

[8] Chan T H, Jia K, Gao S, et al. PCANet: A Simple Deep Learning Baseline for Image Classification[J]. IEEE Transactions on Image Processing, 2015, 24(12):5017-5032.

[9] Ding Sijing, sun Zhongyang, sun Yankui, et al. Image classification of retinal optical coherence tomography based on pcnet and linear discriminant analysis [J]. Chinese Journal of image graphics, 2019, 24 (1): 115-123.

[10] Toet A. Alternating guided image filtering [J]. PeerJ Computer Science, 2016, 2(72):123-156.