Bent Identity-based CNN for Image Denoising

Qiufeng Fan[1,*], Fanbo Hou[1,*], and Feng Shi[1]

[1] *School of Electronic Information Electrical Engineering, Anyang Institute of Technology Anyang 455000,China*
[*] *Corresponding author. E-mail: aqiufenga@163.com; 1135644495@qq.com*

Received: Nov. 20, 2019, Accepted: Mar. 31, 2020

In the process of image acquisition and transmission, the image will be polluted by noise. Therefore, we propose a bent identity-based convolutional neural network (BICNN) model. The model is a full convolu-tional network model with a depth of 30 layers, consisting of six feature extraction modules (FEM) and skip connection. Skip connection combines the output features of the first convolution layer with the output features of each FEM in series to guarantee the full extraction of image's features. Then we adopt the residual learning to alleviate the gradient disappearance and improve the convergence speed so as to ensure that the nonlinear mapping acquired by the trained denoising model is image noise. Bent identity is selected as the activation function, which has soft saturation and the output mean is close to zero, which can enhance the robustness of the model against input noise and accelerate the convergence of the model. Our extensive experiments demonstrate that our BICNN model can not only exhibit high effec-tiveness in several general image denoising tasks, but also make it highly attractive for practical denoising applications.

Keywords: Image denoising; Bent identity activation function; convolutional neural network; FEM

http://dx.doi.org/10.6180/jase.202009_23(3).0019

1. Introduction

Image is an important way for human to acquire, express and transmit information. People acquire about 75% of the external information coming from sight [1]. Images have rich information and it has become an important source for knowledge and data. The captured image can well convey the defect and emotional infor-mation of the image. In the process of image acquisition, formation and transmission, noise will be introduced to cause the degradation of image quality [2]. The complex natural environment will inevitably lead to noise in the acquired image.

In order to improve the quality of the image, it is necessary to de-noise the image. The noise is effectively removed and the important information such as image edges and details are better preserved [3, 4]. The rapid development of wavelet transform theory and its excellent time-frequency analysis characteristics make it widely used in many fields such as image compression and denoising, image fusion, digital watermarking, feature recognition and extraction. At present, wavelet image denoising method has become the mainstream of image denoising research and application [5].

A better denoising method should be universal and can deal with different background and different types of images with noise. Existing denoising methods mostly use standard images to test the denoising effect. The noise contained in the actual image is not single, so it is difficult to evaluate the noise level and type. And the denoising difficulty is much higher than the known noise added to the standard image. Therefore, it is essential to design an universal denoising method based on practical application.

Gaussian noise is the most common image noise, which obeys Gaussian distribution and seriously affects image quality. The widespread methods used for removing Gaussian noise, such as BM3D5, CSF and TRND, have achieved better denoising effects [6, 7], but manual modification for parameters is required. In order to solve the above problems, researchers began to use CNN method for image denoising. For example, Zhang [8] proposed DnCNN denoising model, which applied in-depth network architecture, residual learning algorithm and batch normalization to image denoising. On this basis, Zhang [9] proposed FFDNet model to realize blind Gaussian denoising. Isogawa [10] combined soft shrinkage with deep-CNN and proposed SCNN Gaussian de-noising model. Denoising method based on CNN can effectively make use of the global features of the image and significantly improve the denoising effect. However, with the deepening of the network depth, the features extracted by CNN method become more and more advanced, and a large number of features will be lost layer by layer. If the CNN layer is lower, the extracted low-level features are more primitivem, such as color and edge lines [11]. The high-level extracts texture details and other advanced features. In order to better extract the texture features in the image, this paper adopts skip connection to transfer the low-level features to the high-level and merges the high-level features. Therefore, we propose the BICNN denoising model.

The contributions for this paper are as follows.

1. We propose a bent identity-based convolutional neural network model.

2. Residual learning is adopted to alleviate the gradient disappearance and improve the convergence speed.

3. Bent identity is selected as the activation function, which can enhance the robustness of the model against input noise and accelerate the convergence of the model. The organization of this paper is as follows. In section 2, BICNN denoising model is detailed introduced. Section 3 gave the experiments and analysis. There is conclusion in section 4.

2. BICNN denoising model

2.1. Model structure

In order to effectively remove the noise contained in the image, six feature extraction modules (FEM), skip connection and residual learning are used. The BICNN denoising model shown in Figure 1 is designed. In Figure 1, X,Γ and $\hat{Y}$ represent the input noisy image, estimated noise image and estimated clean image, respectively. L_j is the output of j-th convolution layer, j=1,2,...,30. F_{KI} and F_{KO} and are the input and output of the k-th fea-ture extraction module, k=1,2,3,4,5,6. The layers of the BICNN model are thirty. The last convolution layer only uses a filter with size 3×3 and does not connect the activation layer. The remaining convolutional layers all use 64 filters with 3×3 size and connect the activation layer. Zero padding is performed before each convolution layer to ensure that the output image is the same size as the input. In the network model, skip connection is introduced to connect the output of the first convolution layer with the output of each feature extraction module. The lower level features extracted from the first layer are continuously fused with the higher level features extracted from the higher layer in series to ensure the full extraction of image features. The residual learning idea is introduced to make the denoising model acquire the noise information Γ in the noisy image through the end-to-end learning.

The key to realize denoising with BICNN model is to fully extract various feature information in the image and transmit it to the output end of the model. Inspired by the dense connection, the feature extraction module is designed, which can make full use of the different feature information extracted by different convolutional layers of the network. FEM module contains six convolution layers and four activation layers. The activation function of the activation layer is bent identity. The output of each FEM is fused by the output of uneven con-volution layer and the output of the first convolution layer in the BICNN model. Common active function has sigmoid function, tanh and ReLu, etc. Bent identity is perfect for tasks where the underlying behavior is linear (similar to linear regression). This activation function is not sufficient in the presence of non-linearity, but it can still be used as an activation function for regression tasks on the final output node. The convergence speed will be faster.

2.2. Mathematical model

Assume that the noise model in the image is $X = Y + n$, where X and Y are the noisy image and clean image, respectively. n is the Gaussian noise of standard deviation σ . The purpose of training BICNN denoising model by using $\{X_i, Y_i\}_{i=1}^{N}$ is to establish a non-linear mapping between X and Y, and to learn the noise feature of input noisy images. In order to minimize the difference between the actual clean image and the ideal clean image, the loss function is defined as:

$$L(w,b) = \arg\min_{w,b} \frac{1}{2N} \sum_{i=1}^{N} \left\| Y_i - \hat{Y}_i \right\|^2 \tag{1}$$

Where w and b are weight matrix and bias. X_i, Y_i and $\hat{Y}_i$ are the noise image, corresponding ideal clean image and the estimated clean image respectively.

The mathematical model of BICNN is shown in figure 1, where the output of the first convolutional layer is:

$$L_1 = \sum_{d=1}^{64} w_{3\times3}^{1} \times X + b^1 \tag{2}$$

Where $W_{3\times3}^{1}$ represents that the size of of the first convolution layer is 3×3, that is, the upper and lower indices of w represent the layer and the filter size of the layer. The upper index of b^1 represents its layer. d represents the number of input channels, namely the the number of filters. The input

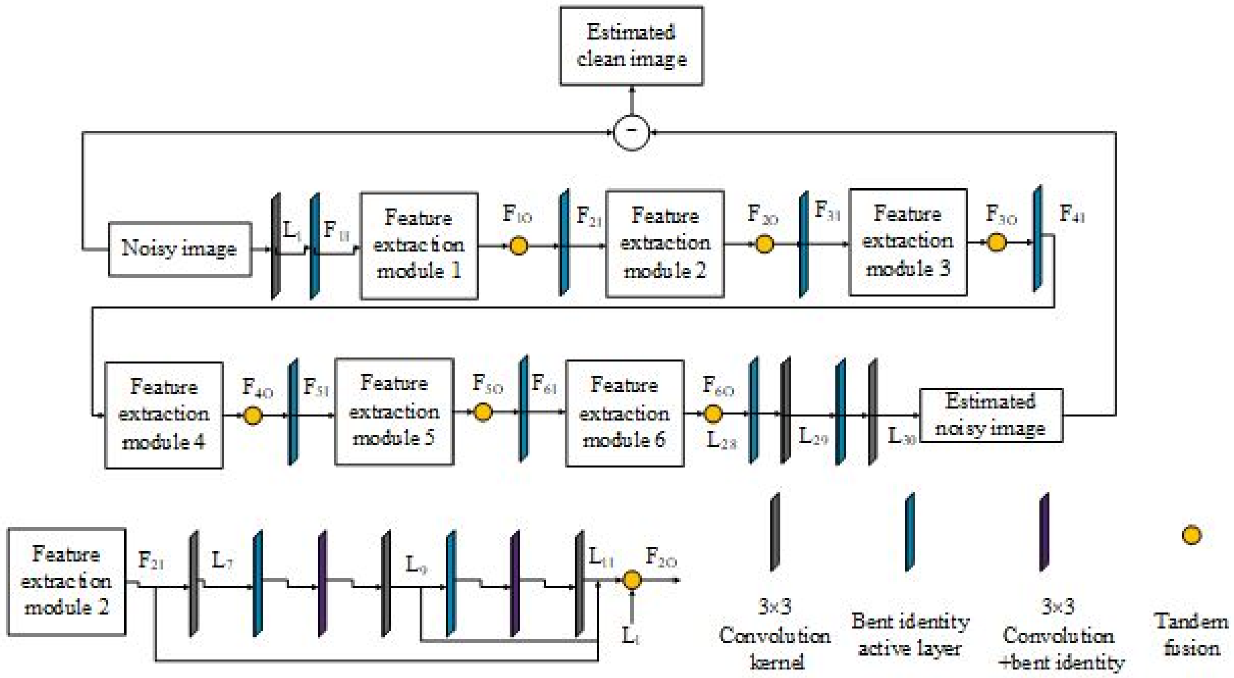


Fig. 1. BICNN model.

of first feature extraction module is:

$$F_{1I} = BI\left(L_1\right) = \begin{cases} L_1, & \sum_{j=1}^{9} w_j x_j > 0 \\ \alpha\left(e^{L_1} - 1\right), & \sum_{j=1}^{9} w_j x_j \le 0 \end{cases} \tag{3}$$

Where BI() is active function. Hyper-parameter $\alpha > 0$. x_j and w_j denote image pixel value and corre-sponding weight coefficient in local receptive field (filter window size is 3×3). The input of second feature extrac-tion module is:

$$F_{2I} = BI\left(F_{10}\right) \tag{4}$$

$$L_7 = \sum_{d=1}^{64} w_{3\times 3}^{7} \times F_{2I} + b^7 \tag{5}$$

$$L_9 = \sum_{d=1}^{64} w_{3\times 3}^{9} \times BI\left(\sum_{d=1}^{64} w_{3\times 3}^{8} \times BI\left(L_7\right) + b^8\right) + b^9 \tag{6}$$

$$L_{11} = \sum_{d=1}^{64} w_{3\times 3}^{11} \times BI\left(\sum_{d=1}^{64} w_{3\times 3}^{10} \times BI\left(L_9\right) + b^{10}\right) + b^{11} \tag{7}$$

$$F_{20} = \text{concat}\left[L_7, L_9, L_{11}, L_1\right] \tag{8}$$

Where *concat*[] denotes tandem fusion. The input of third feature extraction module is:

$$F_{3I} = BI\left(F_{20}\right) \tag{9}$$

$$L_{27} = \sum_{d=1}^{64} w_{3\times 3}^{27} \times BI\left(F_{50}\right) + b^{27} \tag{10}$$

$$\Gamma = L_{28} = w_{3\times 3}^{28} \times BI\left(L_{27}\right) b^{28} \tag{11}$$

We use BICNN model to get the noise information Γ , then we can obtain estimated clean image $\hat{Y} = X - \Gamma$.

3. Experiments and analysis

In order to verify the denoising effect of the proposed model, experiments are carried out on standard im-ages. 400 gray images with 180×180 pixels are selected to con-struct the training set. The training set of BICNN model is formed by randomly flipping 400 images and rotating them at any angle and then cutting them into 40×40 pixels images with step size 10. The experimental environment is Keras, which is carried out on a PC of CPU Intel (R) Core(TM) i5-8300h CPU2 30 GHz, using NVIDIA 1060 Ti GPU ac-celeration. Two widely used indicators are employed to indicate the effect of image denoising: the signal-to-noise-ratio (SNR), (normalized mean square error) NMSE and (Structural Similarity) SSIM [12]. The SNR, NMSE and SSIM are calculated as following:

$$SNR = 10 \ln \frac{\sum_{k=1}^{N} S(k)^2}{\sum_{k=1}^{N} [X(k) - S(k)]^2} \tag{12}$$

$$NMSE = 10 \ln \frac{\sum_{k=1}^{N} \left(S(k) - S'(k)\right)^2}{\sum_{k=1}^{N} |S(k)|^2} \tag{13}$$

$$SSIM = \frac{\left((2\mu_I \mu_I + C_1) \times (2\sigma_{II} + C_2)\right)}{\left((\mu_I^2 + \mu_I^2 + C_1) \times (\sigma_I^2 + \sigma_{I'}^2 + C_2)\right)} \tag{14}$$

We compare our method with state-of-the-art methods including DMS [13], NWT [14], PNM [15], EGL [16], NLG [17] to demonstrate the effectiveness of our new method. In the simulated study, four images are used for testing:

"Marilyn", "Chupa Chups", "David Hilbert" and "Old Tom Morris". We have extensively tested these values as the criteria for image denoising and find them succeed in virtually all cases.

In Marilyn test experiment, the denoised image using the new method in Figure 2 is compared with the de-noised image of other denoising methods. As observed, the new method successfully eliminates noise and ob-tains more accurate results than other methods. Note that there are some spots in soft threshold function. SDL method indicates that the denoising effect has better smoothness, but the image is fuzzy.

Similarly, in "Chupa Chups", "David Hilbert" and "Old Tom Morris" experiments (Figure 3, Figure 4 and Figure 5), the new method effectively removes the baseline noise and can better retrieve the images compared to other denoising methods.

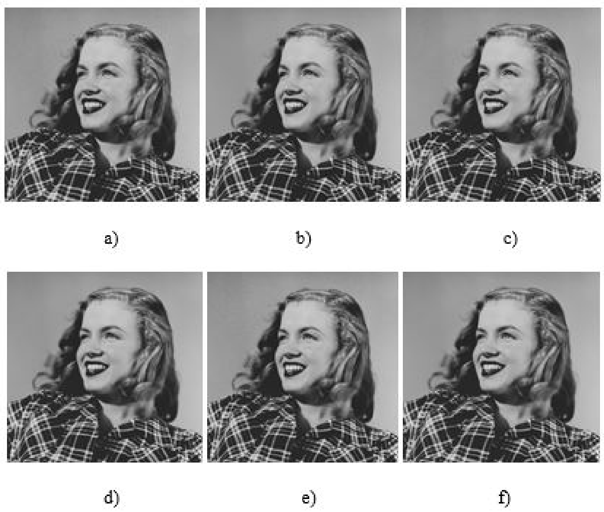


Fig. 2. Image Marilyn denoising results and experiment contrast. (a) DMS; (b) NWT; (c) PNM; (d) EGL; (e) NLG; (f) Proposed.

Similarly, in "Chupa Chups", "David Hilbert" and "Old Tom Morris" experiments (figure 3, figure 4 and figure 5), the new method effectively removes the baseline noise and can better retrieve the images compared to other denoising methods. Tables 1, 2, 3 and 4 show the SNR, NMSE, SSIM of the original noisy and denoised image with different denoising methods. According to table 1, SNR2, NMSE, SSIM with our new method are 19.9692, -8.7180 and 0.9946 which are the largest values among the methods; the proposed method has second largest SNR value, followed by NWT, PNM, EGL, NLG. Overall, the table shows that our new method can obtain better effect than the other methods.

Similarly, Tables 2, 3, and 4 imply that the new method

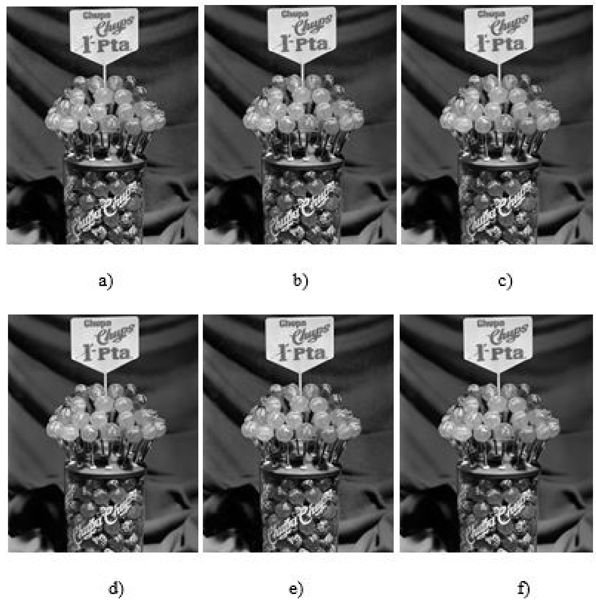


Fig. 3. Image Chupa Chups de-noising results and experiment contrast. (a) DMS; (b) NWT; (c) PNM; (d) EGL; (e) NLG; (f) Proposed.

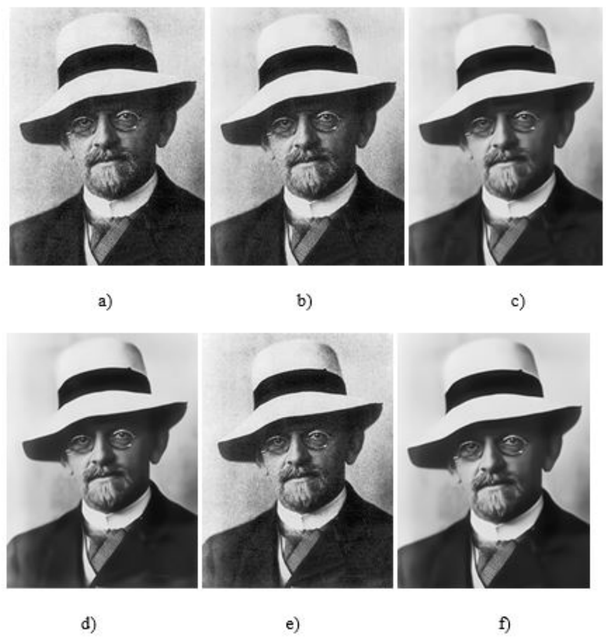


Fig. 4. Image David Hilbert denoising results and experiment contrast. (a) DMS; (b) NWT; (c) PNM; (d) EGL; (e) NLG; (f) Proposed.

outperforms the current denoising methods and successfully recovers the desired images. The values of SNR, NMSE, SSIM of the proposed adaptive denoising algorithm are slightly higher than that of the other compared state-of-

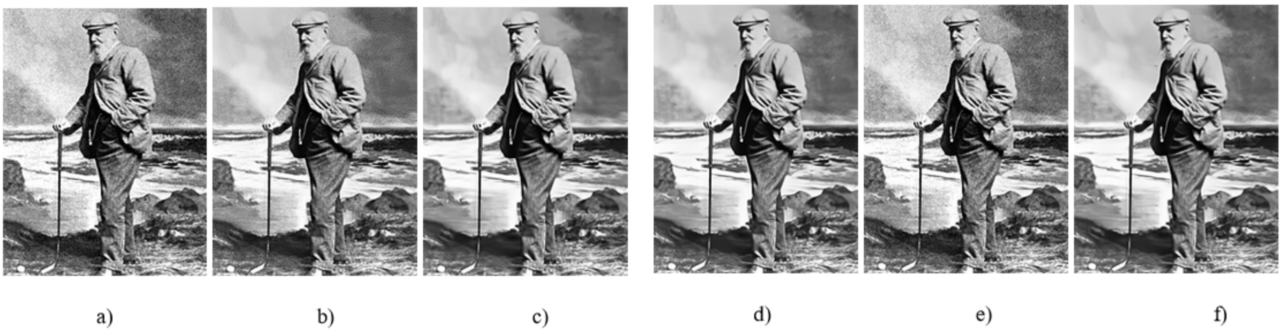


Fig. 5. Image Old Tom Morris de-noising results and experiment contrast. (a) DMS; (b) NWT; (c) PNM; (d) EGL; (e) NLG; (f) Proposed.

the-art methods. It shows that the pro-posed algorithm has a stronger ability to enhance the edges, the texture regions of the image, and preserve the smooth regions of the image while removing the noise.

Table 1. Comparison of Marilyn with different methods.

Method	SNR1	SNR2	NMSE	SSIM
DMS	8.8078	17.9387	-10.3537	0.7464
NWT	14.4077	17.1786	-10.1529	0.8122
PNM	8.3549	19.6941	-8.8761	0.8567
EGL	8.5837	19.7109	-8.8178	0.8924
NLG	8.6714	19.8827	-8.7596	0.9127
Proposed	8.7965	19.9692	-8.7180	0.9946

Table 2. Comparison of Chupa Chups with different methods.

Method	SNR1	SNR2	NMSE	SSIM
DMS	8.9129	15.2760	-10.3795	0.8693
NWT	8.5807	15.3085	-9.8547	0.8754
PNM	8.3253	15.3107	-9.7524	0.9237
EGL	8.5508	15.3193	-9.6688	0.9348
NLG	8.5674	15.2984	-9.5417	0.9717
Proposed	8.9038	15.3692	-7.9251	0.9935

In here, we emphatically introduce other two deep learning state-of-the-art methods (namely, FFDnet [8] and PCN [9]) based on DNN to conduct comparison experiments. The following is a brief review of FFDnet and PCN. The image size is 512×512pixel.

1) FFDnet. The proposed FFDNet works on downsampled subimages, achieving a good trade-off between inference speed and denoising performance. In contrast to the existing discriminative denoisers, FFDnet enjoys several desirable properties, including (i) the ability to handle a wide range of noise levels (i.e., [0, 75]) effectively with a single

Table 3. Comparison of Chupa Chups with different methods.

Method	SNR1	SNR2	NMSE	SSIM
DMS	8.5746	17.1832	-9.5517	0.8316
NWT	8.6472	17.8845	-7.1437	0.8763
PNM	8.2643	18.4126	-7.5526	0.8948
EGL	8.7549	18.4523	-7.7875	0.9124
NLG	8.7654	18.6721	-7.9147	0.9376
Proposed	8.9277	18.7836	-6.9872	0.9946

Table 4. Comparison of Chupa Chups with different methods.

Method	SNR1	SNR2	NMSE	SSIM
DMS	8.5746	17.1832	-9.5517	0.8316
NWT	8.6472	17.8845	-7.1437	0.8763
PNM	8.2643	18.4126	-7.5526	0.8948
EGL	8.7549	18.4523	-7.7875	0.9124
NLG	8.7654	18.6721	-7.9147	0.9376
Proposed	8.9277	18.7836	-6.9872	0.9946

network, (ii) the ability to remove spatially variant noise by specifying a non-uniform noise level map, and (iii) faster speed than benchmark BM3D even on CPU without sacrificing denoising performance. Figure 6 is the framework of FFDnet.

2) PCN. First, a convolutional neural network is proposed for image denoising which achieves the state-of-the-art performance. Second a deep neural network solution is presented that cascades two modules for image denoising and various high-level tasks, respectively, and it uses the joint loss for updating only the denoising network via back-propagation. It demonstrates that on one hand, the proposed denoiser has the generality to overcome the performance degradation of different high-level vision tasks. On the other hand, with the guid-ance of high-level vi-

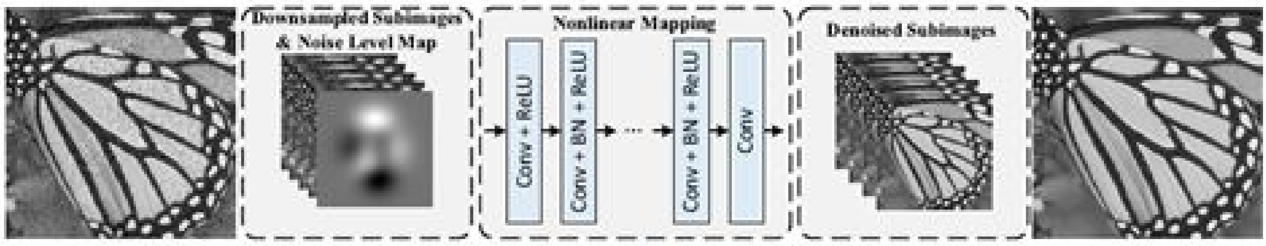


Fig. 6. The architecture of the proposed FFDNet for image denoising. The input image is reshaped to four sub-images, which are then input to the CNN together with a noise level map. The final output is reconstructed by the four denoised sub-images.

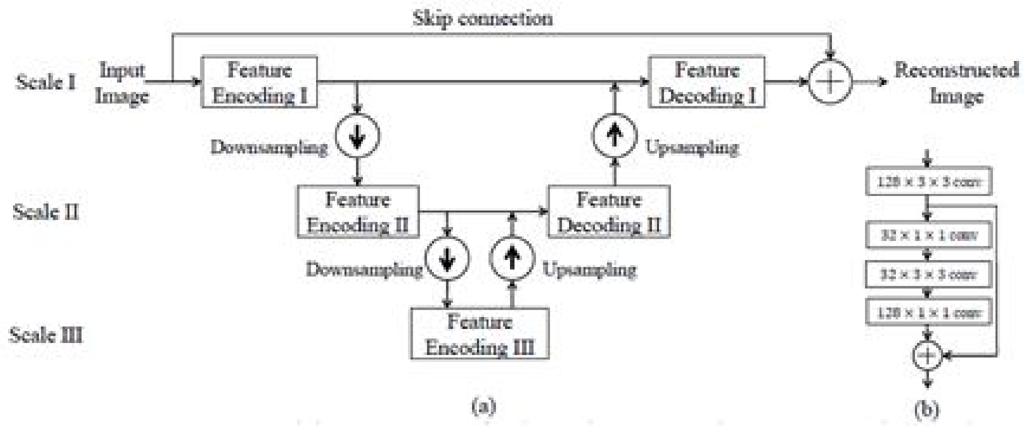


Fig. 7. (a) Overview of PCN denoising network. (b) Architecture of the feature encoding module. (c) Architecture of the feature decoding module.

sion information, the denoising network can generate more visually appealing results.

According to the experiments in the two papers, we set the added noise=40. The comparison result is shown in Figures 8-13. Also the color image denoising results are compared. It can be seen that proposed method and FFD-Net consistently outperforms PCN on different noise levels in terms of both quantitative and qualitative evaluation.

From the above results, we can see that the proposed method yields the better PSNR on most of the images. Because deep learning method needs to many training samples, the results change with the iteration number and sample number. Nevertheless, the proposed method has the similar results with deep learning method and even slightly is better than it in conclusion.

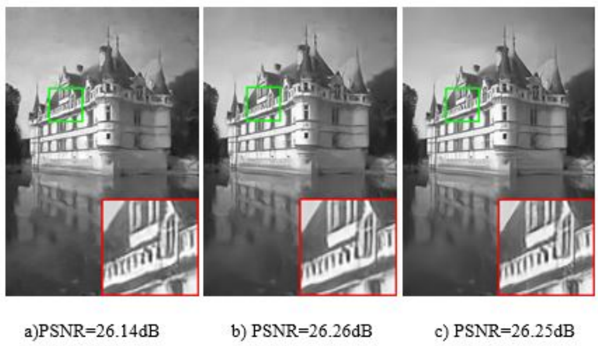
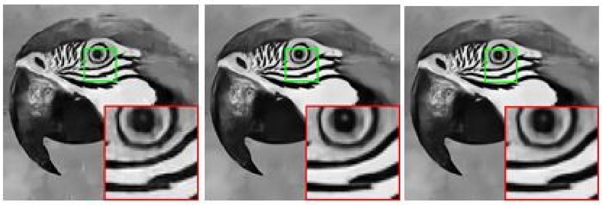


Fig. 8. Denoising results on image "House" from the BSD68 dataset with noise level 40 by different methods. a) FFDnet, b) PCN, c) Proposed.

4. Conclusions

The BICNN model proposed in this paper is a full convolutional network model. Each convolutional layer (not containing the last one) adopts 64 filters with 3×3 size to extract features of different types and levels. The combination of the six feature extraction modules FEM and jump connection in the model can transfer low-level features extracted from the shallow layer to the high level, so as to fully extract noise features in the input image and ensure that the model can effectively establish nonlinear mapping between input and output after end-to-end training. The



a)PSNR=31.22dB b) PSNR=32.18dB c) PSNR=31.56dB

Fig. 9. Denoising results on image "parrot" from the BSD68 dataset with noise level 40 by different methods. a) FFDnet, b) PCN, c) Proposed.

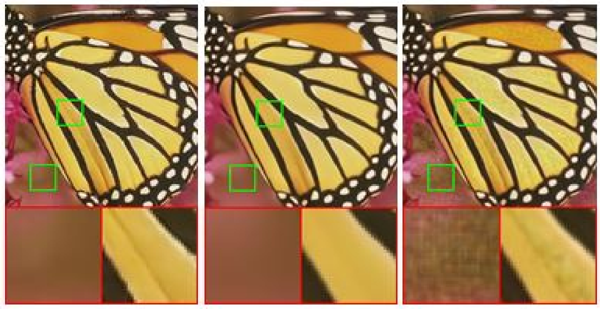
a) PSNR=28.54dB b) PSNR=28.72dB c) PSNR=28.68dB

Fig. 10. Denoising results on image "motorbike" from the BSD68 dataset with noise level 40 by different methods. a) FFDnet, b) PCN, c) Proposed.

a) PSNR=25.33dB b) PSNR=25.28dB c) PSNR=25.37dB

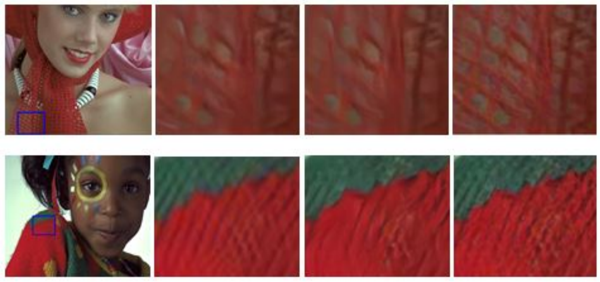
Fig. 11. Denoising results on colour image "zebra" from the BSD68 dataset with noise level 40 by different methods. a) FFDnet, b) PCN, c) Proposed.

information content of the noise in the noisy image is far less than that of the signal. The introduction of residual learning makes the trained model learn the noise, which can reduce the learning difficulty of the model. The soft saturation characteristic of the activation function and the output mean value is close to zero, which can improve the robustness of the model and accelerate the convergence of the model. Compared with other state-of-the-art methods, BICNN model has improved the effectiveness of the standard test set. Meanwhile, it can well retain the feature information of the original image. In the future, more deep learning methods will be adopted and combined to improve the accuracy of image denoising.



a) PSNR=27.84dB b) PSNR=27.91dB c) PSNR=27.95dB

Fig. 12. Denoising results on colour image "butterfly" from the BSD68 dataset with noise level 40 by different methods. a) FFDnet, b) PCN, c) Proposed.



a) b)PSNR=28.63dB c) PSNR=28.66dB d) PSNR=28.68dB

Fig. 13. Two image denoising examples from Kodak dataset. We show (a) the ground truth image and the zoom-in re-gions; (b) the denoised image by FFDnet; (c) the denoised image by PCN; (d) the denoising result of our pro-posed model.

Acknowledgement

This work was supported by the 2017 Anyang city science and technology breakthrough project (photovoltaic power station solar panel cleaning robot).

References

[1] Shoulin Yin, Ye Zhang, and Shahid Karim. Large Scale Remote Sensing Image Segmentation Based on Fuzzy Region Competition and Gaussian Mixture Model. *IEEE Access*, 6:26069–26080, may 2018.

[2] Lin Teng, Hang Li, and Shoulin Yin. Modified pyramid dual tree direction filter-based image denoising via curvature scale and nonlocal mean multigrade remnant filter. *International Journal of Communication Systems*, 31(16), nov 2018.

[3] Shoulin Yin and Jing Bi. Medical image annotation based on deep transfer learning. *Journal of Applied Science and Engineering*, 22(2):385–390, 2019.

[4] Ayesha Saadia and Adnan Rashdi. Incorporating fractional calculus in echo-cardiographic image denoising. *Computers and Electrical Engineering*, 67:134–144, apr 2018.

[5] Lin Teng and Hang Li. CSDK: A Chi-square distribution-Kernel method for image de-noising under the Internet of things big data environment. *International Journal of Distributed Sensor Networks*, 15(5), may 2019.

[6] Kostadin Dabov, Alessandro Foi, and Karen Egiazarian. Video denoising by sparse 3D transform-domain collaborative filtering. In *European Signal Processing Conference*, pages 145–149, 2007.

[7] Wensen Feng, Peng Qiao, and Yunjin Chen. Fast and Accurate Poisson Denoising with Trainable Nonlinear Diffusion. *IEEE Transactions on Cybernetics*, 48(6):1708–1719, jun 2018.

[8] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, jul 2017.

[9] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDNet: Toward a fast and flexible solution for CNN-Based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, sep 2018.

[10] Kenzo Isogawa, Takashi Ida, Taichiro Shiodera, and Tomoyuki Takeguchi. Deep shrinkage convolutional neural network for adaptive noise reduction. *IEEE Signal Processing Letters*, 25(2):224–228, feb 2018.

[11] Shoulin Yin, Ye Zhang, and Shahid Karim. Region search based on hybrid convolutional neural network in optical remote sensing images. *International Journal of Distributed Sensor Networks*, 15(5), may 2019.

[12] Boyang Chen, Xuan Feng, Ronghua Wu, Qiang Guo, Xi Wang, and Shiming Ge. Adaptive wavelet filter with edge compensation for remote sensing image denoising. *IEEE Access*, 7:91966–91979, 2019.

[13] Hongqiang Ma, Shiping Ma, Yuelei Xu, and Mingming Zhu. Deep Marginalized Sparse Denoising Auto-Encoder for Image Denoising. In *Journal of Physics: Conference Series*, volume 960, 2018.

[14] Lu Jing-Yi, Lin Hong, Ye Dong, and Zhang Yan-Sheng. A New Wavelet Threshold Function and Denoising Application. *Mathematical Problems in Engineering*, 2016, 2016.

[15] Fang Huang, Bo Lan, Jian Tao, Yinjie Chen, Xicheng Tan, Jie Feng, and Yan Ma. A Parallel Nonlocal Means Algorithm for Remote Sensing Image Denoising on an Intel Xeon Phi Platform. *IEEE Access*, 5:8559–8567, 2017.

[16] Yibin Tang, Ying Chen, Ning Xu, Aimin Jiang, and Lin Zhou. Image denoising via sparse coding using eigenvectors of graph Laplacian. *Digital Signal Processing: A Review Journal*, 50:114–122, mar 2016.

[17] Hongjun Li and Ching Y. Suen. A novel Non-local means image denoising method based on grey theory. *Pattern Recognition*, 49:237–248, jan 2016.