Deep Residual Network with Sparse Feedback for Image Restoration

Zhenyu Guo [1] **, Yujuan Sun** [1,]***, Muwei Jian** [2,3] **and Xiaofeng Zhang** [1]

[1] School of Information and Electrical Engineering, Ludong University, Yantai 264025, China; handlecoding@foxmail.com (Z.G.); iamzxf@126.com (X.Z.)

[2] School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China; jianmuweihk@163.com

[3] School of Information Science and Engineering, Linyi University, Linyi 276000, China

[*] Correspondence: syj_anne@163.com; Tel.: +86-158-5357-8596

check for updates

Abstract: A deep neural network is difficult to train due to a large number of unknown parameters. To increase trainable performance, we present a moderate depth residual network for the restoration of motion blurring and noisy images. The proposed network has only 10 layers, and the sparse feedbacks are added in the middle and the last layers, which are called FbResNet. FbResNet has fast convergence speed and effective denoising performance. In addition, it can also reduce the artificial Mosaic trace at the seam of patches, and visually pleasant output results can be produced from the blurred images or noisy images. Experimental results show the effectiveness of our designed model and method.

Keywords: image restoration; motion deburring; image denoising; sparse feedback

1. Introduction

In recent years, many hard problems in computer vision have been well solved, especially in the fields of image classification, object detection [1], and identification [2]. By using the deep learning method, the accuracy and robustness of these issues have been greatly improved.

Until now, many kinds of neural network structures have been proposed, such as five layers LeNet [3], eight layers AlexNet [4], 19 layers VGG [5], 22 layers GoogleNet [6], 152 layers ResNet [7], GAN [8], and so on. These networks are gradually deepened, and the training data set is also getting much larger. Although the technology of deep learning is especially effective for issues of classification, it still includes many problems that cannot be solved in image restoration. The main reason includes two aspects: First, to collect a large number of training data about image degradation is not easy; second, the degradation reason is various, and it is exceptionally difficult to enumerate all cases of image degradation. Therefore, when using a deep neural network to improve these problems, the training set is not usually enough to train a deep neural network model.

In this paper, we improve the network structure of ResNet [7] and propose a sparse feedback residual network, which is called FbResNet. It includes 10 layers with sparse "shortcut connections" (in this article we call the "shortcut connections" forward feedback). Figure 1 shows the structure of FbResNet. It can be seen that FbResNet only includes two forward feedbacks, which are derived from the input layer. One feedback is connected to the middle layer, while the other is connected to the last layer. The two feedbacks can provide an effective constraint of the loss function and help to train reasonable network parameters. Experimental results show FbResNet has a fast convergence speed and effective denoising performance.
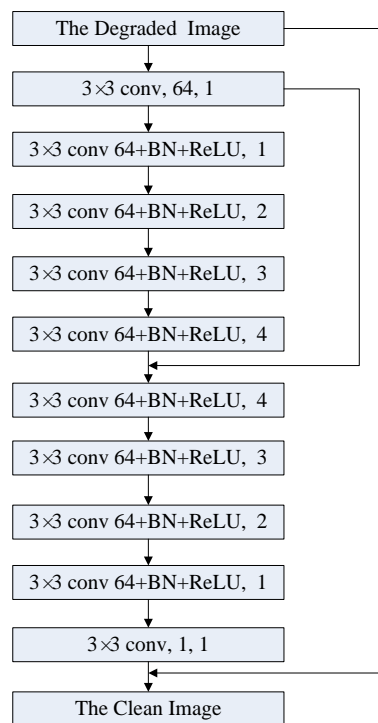
```
┌──────────────────────────────┐
│     The Degraded  Image       │────────┐
└──────────────────────────────┘        │
              ↓                          │
┌──────────────────────────────┐        │
│        3×3 conv, 64, 1        │───┐    │
└──────────────────────────────┘   │    │
              ↓                     │    │
┌──────────────────────────────┐   │    │
│  3×3 conv 64+BN+ReLU,  1      │   │    │
└──────────────────────────────┘   │    │
              ↓                     │    │
┌──────────────────────────────┐   │    │
│  3×3 conv 64+BN+ReLU,  2      │   │    │
└──────────────────────────────┘   │    │
              ↓                     │    │
┌──────────────────────────────┐   │    │
│  3×3 conv 64+BN+ReLU,  3      │   │    │
└──────────────────────────────┘   │    │
              ↓                     │    │
┌──────────────────────────────┐   │    │
│  3×3 conv 64+BN+ReLU,  4      │   │    │
└──────────────────────────────┘   │    │
              ↓←────────────────────┘    │
┌──────────────────────────────┐        │
│  3×3 conv 64+BN+ReLU,  4      │        │
└──────────────────────────────┘        │
              ↓                          │
┌──────────────────────────────┐        │
│  3×3 conv 64+BN+ReLU,  3      │        │
└──────────────────────────────┘        │
              ↓                          │
┌──────────────────────────────┐        │
│  3×3 conv 64+BN+ReLU,  2      │        │
└──────────────────────────────┘        │
              ↓                          │
┌──────────────────────────────┐        │
│  3×3 conv 64+BN+ReLU,  1      │        │
└──────────────────────────────┘        │
              ↓                          │
┌──────────────────────────────┐        │
│        3×3 conv, 1, 1         │        │
└──────────────────────────────┘        │
              ↓←───────────────────────── ┘
┌──────────────────────────────┐
│       The Clean Image         │
└──────────────────────────────┘
```

Figure 1. The architecture of proposed network FbResNet.

2. Related Works

Image restoration is a long-standing problem in low-level computer vision. In practice, the obtained images are usually degraded (such as images with noise, blurred images [9], sampled images [10], etc.). Image restoration is used to estimate the original image based on the degenerated image. Since this is an under-constraint problem, the solution is not unique. In order to reduce the solution space of the problem, the prior-based methods had been widely used, which had added the prior information or constraints and could recover the original image from the degraded image [10,11].

However, most of the prior-based methods build the objective function by simplifying the mechanism of the image degradation, and not considering the affection of noises or other factors. Hence, they cannot perfectly restore clean images for severely degraded images. In addition, the prior-based methods involve a complex optimization problem, and most of the prior-based methods can hardly achieve high performance without sacrificing computational efficiency. Furthermore, the prior-based models, in general, are nonconvex and involve several manually chosen parameters [12]. Therefore, there are many limitations in restoring images just using prior-based methods.

Deep convolution neural network (CNN) [13,14] has made a series of breakthroughs in many applications of computer vision [15,16], such as image classification, recognition and target detection, etc. The features of CNN are mainly exacted by increasing the depth of network model. Then, the lower, middle, and the advanced extracted features will be gradually obtained. In general, the advanced features will be used to connect with one or several fully connected layers. The reason for this remarkable achievement in computer vision task is mainly because many rich characteristics of different levels can be extracted by training the deep neural network.

Recent evidence reveals that the depth of neural network is very important. Many visual tasks [17–19], especially the low-level vision problems, have greatly benefited from very deep network. There are several references to perform the denoising problems using deep neural networks.

Reference [18] used a convolutional network as image denoising architecture and claimed that CNNs could provide comparable, and in some cases, superior performance to wavelet and Markov random field methods. Moreover, Reference [20] found that a convolutional network offered similar

performance in the blind denoising setting, as compared to other techniques in the non-blind setting. However, training the convolutional network architecture requires substantial computation and many thousands of updates to converge.

Reference [21] combined sparse coding and deep networks pretrained with denoising auto-encoder for the tasks of image denoising and blind inpainting and achieved comparable results to K-SVD [22]. This method could automatically remove complex patterns, such as superimposed text from an image, and improve the performance of unsupervised feature learning. However, the method in Reference [21] also strongly relied on supervised training and could remove only the noise patterns in the training data. In Reference [23], trainable nonlinear reaction diffusion had been proposed and could be used for a variety of image restoration tasks by incorporating appropriate reaction force. In [24], the multi-layer perceptron (MLP) had been used for image denoising.

The model in References [23,24] can achieve promising performance. Reference [24] claimed that training MLP with many hidden layers could lead to problems, such as vanishing gradients and over-fitting. Reference [24] also found back propagation will work well and concluded that deep learning techniques are not necessary.

In a deep network structure, can all these extracted features be fully used? There may be many useless layers or useless parameters, some high-level features that may be actually useless for the low-level applications of image processing. Therefore, a moderate depth neural network has been proposed in this paper, which is a 10-layer deep residual network with sparse feedback loops. The detail of the proposed network will be introduced as following.

3. Deep Residual Network with Sparse Feedback Loops

In the design and application of neural network, the researchers are only required to focus on the input and output, the number of hidden layers, and the initial parameters. As the network depth gradually increases, the parameters of neural network are also difficult to tune. There are also no relevant theories to be presented on how to tune these parameters. Moreover, the updating of neuron parameters depends on the gradient; the more far away from the output layer, the more difficult for updating of neuron parameters. It will be invariant, or will change dramatically, which is called gradient disappearance or gradient explosion problem. Although the dropout strategy, or batch normalization, was adopted to reduce explosion problems, it still often happened. The more layers of neural network, the more obvious the gradient disappearance or explosion problem is.

This question reminds me of the amplifier cascade problem in electronics. When connecting the circuit, the output signal is usually unstable. Single negative feedback or inter-stage negative feedback will generally be added to stabilize the output signal. In the amplifier circuit, the negative feedback is added to enhance the performance of the anti-noise and stability of the circuit, but the feedback will also reduce the amplification factor of the circuit. Usually, sparse feedback with a longer span is adopted to keep a tradeoff between the robustness and amplification factor. We believe that this situation is very similar to the shortcut in residual neural network. We try to adjust the shortcut in ResNet to a sparse longer connection. However, the idea of sparse longer feedback comes from the concept of negative feedback in the circuit; hence it is represented by "feedback."

However, the denoising results of ResNet are not better than that of the convolution neural network without feedback. For example, denoising convolutional neural network (DnCNN) [25] is good at removing Gaussian noise. Therefore, we combine the network structure of ResNet and DnCNN, and propose the deep residual network with sparse feedback for image restoration, which is called FbResNet. The proposed network structure is shown in Figure 1.

There are only two feedbacks in Figure 1; one is a short feedback; the other is a long feedback. The short feedback is connected from layer 1 to layer 5. Because the output dimension of layer 1 and layer 5 is equal, the outputs of layer 1 and layer 5 can be directly added. The dimension of the input layer is the same as that of the last convolutional layer. Hence, the long feedback adds the input image

to the output of the last convolutional layer, which can add a constraint to FbResNet to keep the most similarity between the input noisy image and output clean image.

The mean squared error between the clean images and the degraded images can be defined as the loss function to learn the trainable parameters Ξ of FbResNet as follows:

$$f(\Xi) = \frac{1}{N} \sum_{i=1}^{N} \|R(x_i, \Xi) - y_i\|^2 \tag{1}$$

where the input is a noisy image x_i, and FbResNet aims to learn a mapping function $R(x_i) = y_i$, to predict the clean image. $\{(x_i, y_i)\}$ represents N degraded-clean training image (patch) pairs; R represents the network structure, of which parameters Ξ require to be trained.

3.1. Network Structure

Inspired by the residual learning structure, we propose the deep residual network with sparse feedback loops for image restoration, and the structure of FbResNet is shown in Figure 1. It consists of ten layers. "Convolution" block is in the first layer. This layer has no "Batch Normal" and "ReLU," in other words, the information produced by this layer is the original information after filtering the input image, then it is used to estimate the residual information by feeding back to the middle and the last layer. Eight "Convolution + Batch Normalization + ReLU" blocks are in the middle layers. The number behind each middle layer is the dilation factors, which is set to 1, 2, 3, 4, 4, 3, 2 and 1, respectively. By using the increasing dilated factors, the first-half layers can learn the residual information using an enlarged receptive field, and the latter half layers can refine the residual information using the decreasing dilation factors. In order to ensure that the estimated residual information does not deviate greatly, two forward feedbacks from the first layer have been added. The first is connected to the middle of the dilation convolution. The second is connected to the last layer. The main task of FbResNet is to estimate the residual information between the input degraded image and the output clean image.

3.2. Implementation

In order to reduce the size and parameters of the neural network, we cut the input training image into small patches. But the restored image may exist annoying artifact boundary. There are two methods to deal with this problem: Symmetrical padding and zero padding (same padding). To verify the effectiveness of FbResNet in handling boundaries, we use the same padding strategy. Note the dilated convolution with dilation factor 4 pads 4 zeros pixels in the boundaries of each feature map. Batch normalization (BN) is adopted right after each convolution and before activation. We initialize the weights as in Reference [26] and Adam is used as minimizing function with a mini-batch size of 38. The learning rate starts from 0.1 and is divided by 10 when the error plateaus. We use a weight decay of 0.001 and a momentum of 0.9. The dropout is not used in the training phase.

3.3. Comparison

In order to verify the effectiveness of the proposed FbResNet, the comparison with the other network structures has been performed. In our opinion, very deep network architecture requires a huge training set, but in many computer vision tasks, a large number of training samples is not easy to be obtained. Nevertheless, small training samples can be easily constructed. Because our training set is small, for comparison on the same network scale, we reduce the depth of the ResNet and set it to 10. The reformed ResNet is shown in Figure 2. Figure 2 shows two kinds of network structures reformed from ResNet and the network structure of DnCNN. The network structure of (a) is same as that of ResNet except for the depth; besides of the first layer and the last layer, only 4 building blocks are used in the reformed ResNet. In order to compare the performance at the same configuration with Figure 1, Figure 2a is also improved to Figure 2b, which is called ResNet with dilated convolution. Figure 2c

shows the structure of DnCNN. The meaning of the parameters on each layer in Figure 2 is similar to that in Figure 1. The experiment setting of different models had been shown in Table 1.
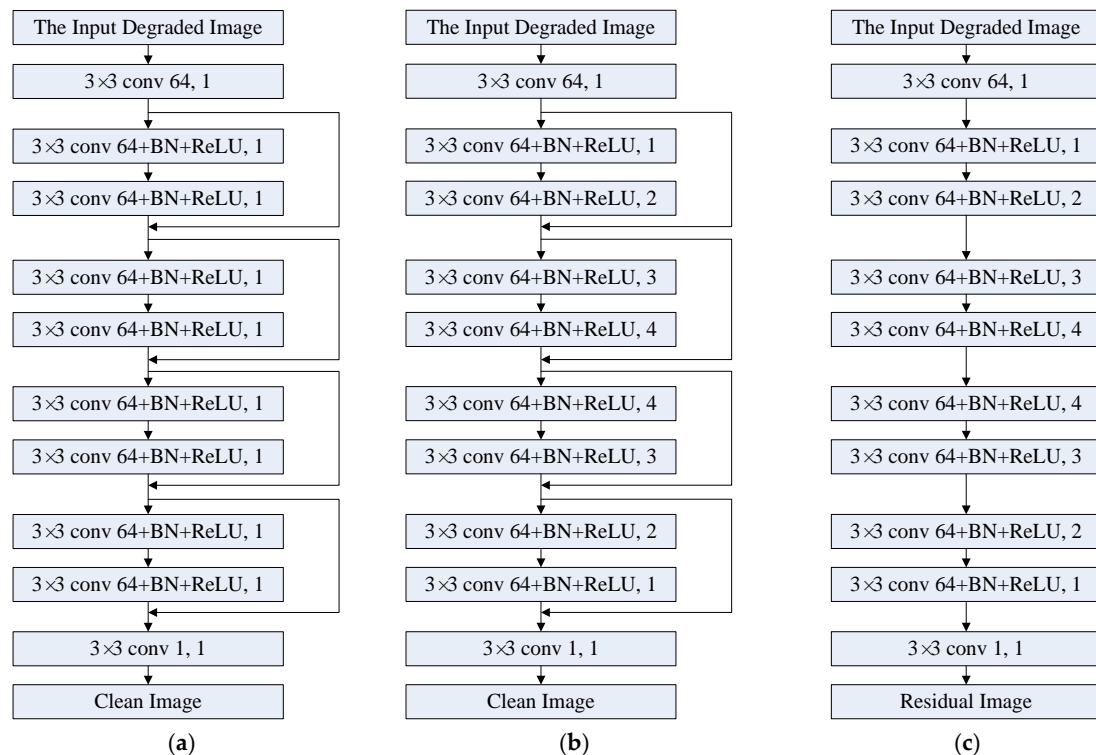
The Input Degraded Image	The Input Degraded Image	The Input Degraded Image
3×3 conv 64, 1	3×3 conv 64, 1	3×3 conv 64, 1
3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 1
3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 2	3×3 conv 64+BN+ReLU, 2
3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 3	3×3 conv 64+BN+ReLU, 3
3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 4	3×3 conv 64+BN+ReLU, 4
3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 4	3×3 conv 64+BN+ReLU, 4
3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 3	3×3 conv 64+BN+ReLU, 3
3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 2	3×3 conv 64+BN+ReLU, 2
3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 1	3×3 conv 64+BN+ReLU, 1
3×3 conv 1, 1	3×3 conv 1, 1	3×3 conv 1, 1
Clean Image	Clean Image	Residual Image
(a)	(b)	(c)

Figure 2. The architectures of the reformed ResNet and denoising convolutional neural network (DnCNN). (**a**) ResNet; (**b**) ResNet with dilated convolution; and (**c**) DnCNN.

Table 1. Experiment setting of different models.

Structure	ResNet	ResNet with Dilated Convolution	DnCNN	FbResNet
Input image	blur image	blur image	blur image	blur image
Number of layers	10	10	10	10
Padding strategy	Same	same	symmetric	same
Dilation convolution	No	yes	yes	yes
Output image	clear image	clear image	residual image	clear image

4. Experiments and Analysis

We evaluate our proposed model and method on two datasets. One is the human face database Facial Recognition Technology (FERET) [27], which includes 1403 human faces with size 64×64. The other is 660 images of size 180×180, which is a part of images presented in Reference [25], which includes the images of animals, humans, various landscapes, etc.

4.1. Human Face Database

To train FbResNet for motion deblurring with unknown motion direction, we consider four different motion directions: up, down, left, and right. The database used in this paper includes 1403 human faces. These faces are divided into two parts: 1044 images are used to train the parameters of FbResNet, and the remaining 359 images are used to test the network. The size of the human face image is 64×64. In this database, the technology of patch cutting is not adopted because the size of the input image is already small.

Figure 3 shows the restored images of human face. The first row shows the images with motion blur; the second row shows the deblurred images of DnCNN; the third row shows the deblurred images of ResNet; the forth row shows the deblurred images of ResNet with dilated convolution; the fifth row shows the deblurred images of FbResNet; and the last row shows the clean images. The restored results of FbResNet are clearer and more similar with the ground truth (clean images) than those of the other models.

Figure 3. Restored images of human face. The first row shows the images with motion blur; the images from the second row to the fifth row are restored images of DnCNN, ResNet, ResNet with dilated convolution and FbResNet; the last row shows the clean images.

In addition, we find that the restored images using DnCNN and FbResNet are darker than the other ones (the fifth column of Figure 3). The main reason is that there are a few feedbacks or no feedback in these two network structures. Hence, the network structure with a few feedbacks can maintain the average skin color of the restored face image and avoid overexposure, even if the input image has a small exposure effect.

Figure 4 shows the enlarged image of a human face. From left to right, it is the restored image of DnCNN, ResNet, ResNet with dilated convolution, FbResNet, and the ground truth, respectively. It is easier to see the advantages of FbResNet, which has less deformation and higher resolution than those of other models. Figure 5 shows the restored images with motion blur in different directions. The first row shows the blur images and the second row shows the deblured images; the below words describe the movement direction of the produced blur images. We find that the robust performance of FbResNet for motion blur is generally good, but the more mixture of the movement direction, the performance will gradually decline, and the deblurred robustness for various directions needs to be further improved.

Figure 4. Enlarged restored images. From left to right, it is the restored images of DnCNN, ResNet, ResNet with dilated convolution and FbResNet, and the last is the clean image.
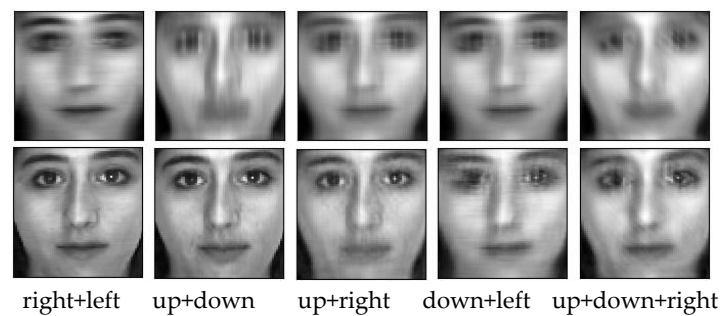
right+left   up+down   up+right   down+left   up+down+right

Figure 5. Restored images for input images with motion blur in different directions.

The left figure in Figure 6 shows the average peak signal-to-noise ratio (PSNR) improvement over the other models with respect to different motion direction by FbResNet model. It can be seen that the proposed FbResNet consistently outperforms the other models by a large margin. The right figure in Figure 6 shows the convergence of the loss function. The convergence speed of FbResNet model is faster than that of the other network models.
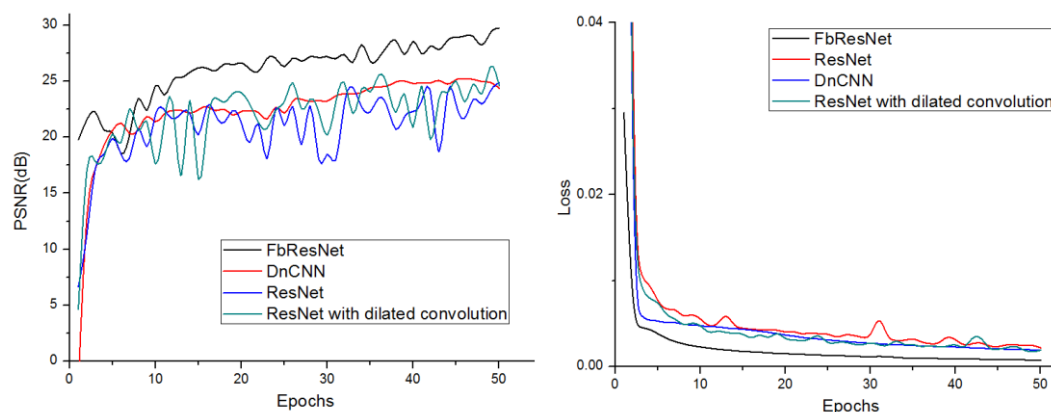
Figure 6. Peak signal-to-noise ratio (PSNR) and loss function for motion deblurring.

4.2. Image Denoising

To train FbResNet for image denoising with different noise level, 660 images of size 180×180 are used. The noise level 15, 35, 45 has been added in the image set. According to the method of Reference [25], we crop each image into small patches, and each patch size is set to 60×60. Then, we obtain 5940 training samples. These samples are divided into two parts, 80% of which is used to train the parameters of FbResNet; the rest is used as the validation set.

In order to validate the effectiveness of FbResNet, 192 images have been used to test the performance of the trained FbResNet and ResNet for image denoising. Figure 7 shows several denoised images with noise level 35. The first column shows the noisy input images; the second is the denoised images of ResNet model and the last column is the results of FbResNet model. Because each test image has been cropped into many small patches, the output patches of the network must be spliced in order to get a complete denoised image.

It can be seen that the images in second column of Figure 7 have obvious artificial stitching trace; nevertheless, it is almost impossible to find the presence of artificial traces from the images in last column. The main reason is that the sparse feedbacks have been added to FbResNet model and can be used to smooth the artificial traces at the seam of patches. Figure 8 shows the enlarged denoised images, and the images from left to right in the first row are input noisy image, restored image of ResNet, restored image of ResNet with "dilation convolution " + " symmetric padding," and restored image of FbResNet. The images in second row, from left to right, are the enlarged images of the red box corresponding to the position in the first row. The restored image of ResNet without symmetric

padding has the distinct artificial stitching traces, and the stitching traces have been improved in that of the ResNet with the symmetric padding. However, even if our algorithm is not added symmetric padding, it can also achieve the same effect as that of ResNet with symmetric padding.
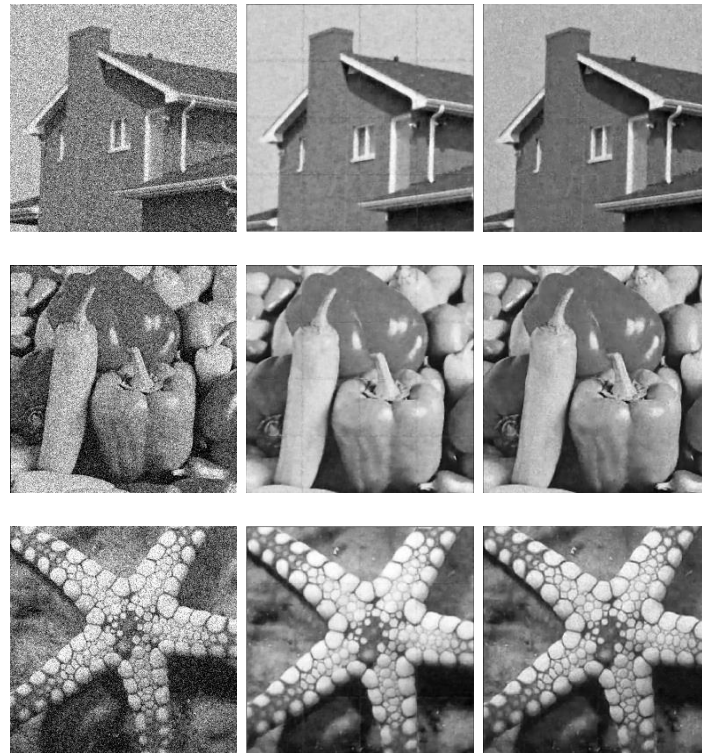

Figure 7. Several denoised image results with noise level 35. From left to right, the first column shows the noisy images, the second column shows the restored image of ResNet, and the last column shows the restored images of FbResNet.

The experiment results on these two datasets demonstrate the feasibility of training FbResNet, which can produce visually pleasant output result for the motion deblurring or image denoising.
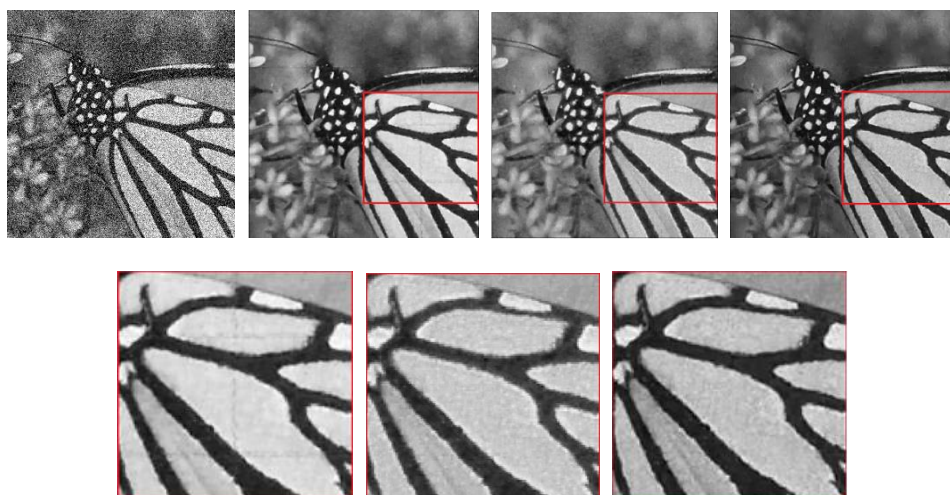
Figure 8. Enlarged denoised images. The images from left to right in the first row are input noisy image, restored image of ResNet, restored image of ResNet with "dilation convolution" + "symmetric padding," and restored image of FbResNet; the images in second row, from left to right, are the enlarged images of the red box corresponding to the position in the first row.

5. Conclusions

In this paper, we have designed and trained a deep residual network with sparse feedback loops for image restoration, especially for the restoration of motion deblurring and image denoising. The addition of sparse feedback improves the convergence speed and the training stability of network model. In addition, the proposed FbResNet is good at smoothing artificial stitching trace at the seam of patches, and visually pleasant output results can be produced from the deblurred images or denoised images.

References

1. Jian, M.; Lam, K. Simultaneous Hallucination and Recognition of Low-Resolution Faces Based on Singular Value Decomposition. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 1761–1772. [CrossRef]

2. Jian, M.; Lam, K.; Dong, J.; Shen, L. Visual-patch-attention-aware Saliency Detection. *IEEE Trans. Cybern.* **2015**, *45*, 1575–1586. [CrossRef] [PubMed]

3. Bottou, L.; Cortes, C.; Denker, J.S.; Drucker, H.; Guyon, I.; Jackel, L.D.; LeCun, Y.; Muller, U.A.; Sackinger, E.; Simard, P.; et al. Comparison of classifier methods: a case study in handwritten digit recognition. In Proceedings of the International Conference on Pattern Recognition. IEEE Computer Society, Jerusalem, Israel, Israel, 9–13 October 1994; pp. 77–82.

4. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Curran Associates Inc, Red Hook, NY, USA, 3–6 December 2012; pp. 1097–1105.

5. Ma, X.; Dai, Z.; He, Z.; Ma, J.; Wang, Y.; Wang, Y. Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. *Sensors* **2017**, *17*, 818. [CrossRef] [PubMed]

6. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

8. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Proceedings of the International Conference on Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2014; pp. 2672–2680.

9. Park, J.; Min, K.; Chang, S.K.; Lee, K.H. Estimation of motion blur parameters using cepstrum analysis. In Proceedings of the IEEE International Symposium on Consumer Electronics, Singapore, 14–17 June 2011; pp. 406–409.

10. Jian, M.; Lam, K.; Dong, J. A Novel Face-Hallucination Scheme Based on Singular Value Decomposition. *Pattern Recognit.* **2013**, *46*, 3091–3102. [CrossRef]

11. Jian, M.; Lam, K.; Dong, J. Facial-Feature Detection and Localization Based on a Hierarchical Scheme. *Inf. Sci.* **2014**, *262*, 1–14. [CrossRef]

12. Wang, X.; Gao, L.; Song, J.; Shen, H. Beyond Frame-level CNN: Saliency-Aware 3-D CNN with LSTM for Video Action Recognition. *IEEE Signal Process. Lett.* **2017**, *24*, 510–514. [CrossRef]

13. Chan, T.H.; Jia, K.; Gao, S.; Lu, J.; Zeng, Z.; Ma, Y. PCANet: A Simple Deep Learning Baseline for Image Classification. *IEEE Trans. Image Process.* **2015**, *24*, 5017. [CrossRef] [PubMed]

14. Lecun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **2014**, *1*, 541–551. [CrossRef]

15. Sun, Y.; Jian, M.; Zhang, X.; Dong, J.; Shen, L.; Chen, B. Reconstruction of normal and albedo of convex Lambertian objects by solving ambiguity matrices using SVD and optimization method. *Neurocomputing* **2016**, *207*, 95–104. [CrossRef]

16. Sun, Y.; Zhang, X.; Jian, M.; Wang, S.; Wu, Z.; Su, Q.; Chen, B. An improved genetic algorithm for three-dimensional reconstruction from a single uniform texture image. *Soft Comput.* **2016**, 1–10. [CrossRef]

17. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

18. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]

19. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]

20. Jain, V.; Seung, H.S. Natural image denoising with convolutional networks. In Proceedings of the International Conference on Neural Information Processing Systems; Curran Associates Inc.: Red Hook, NY, USA, 2008; pp. 769–776.

21. Xie, J.; Xu, L.; Chen, E. Image denoising and inpainting with deep neural networks. In Proceedings of the International Conference on Neural Information Processing Systems; Curran Associates Inc.: Red Hook, NY, USA, 2012; pp. 341–349.

22. Elad, M.; Aharon, M. Image Denoising via Sparse and Redundant Representations Over Learned Dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745. [CrossRef] [PubMed]

23. Chen, Y.; Pock, T. Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1256–1272. [CrossRef] [PubMed]

24. Burger, H.C.; Schuler, C.J.; Harmeling, S. Image denoising: Can plain neural networks compete with BM3D. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; IEEE Computer Society: Los Alamitos, CA, USA, 2012; pp. 2392–2399.

25. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]

26. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the ICCV, Las Condes, Chile, 11–18 December 2015; pp. 1026–1034.

27. Phillips, P.J.; Wechsler, H.; Huang, J.; Rauss, P.J. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* **1998**, *16*, 295–306. [CrossRef]