Residual U-shaped Network for Image Denoising

김상민 O, 박범준, 정제창
한양대학교 전자컴퓨터통신공학과
ksmh1652@gmail.com, kkbbbj@gmail.com, jjeong@hanyang.ac.kr

Abstract

In this paper, we propose a new end-to-end convolutional neural network for image denoising. Conventional algorithms using CNN for image denoising did not fully utilize their network parameters, so the results of them can be more improved than before. The proposed network has a U-shaped hierarchical structure via residual learning and fully utilize network parameters. The proposed network uses ResBlocks with two convolution layers and one PReLU layer for extracting features in deeper network. In the U-shaped hierarchical structure, smaller feature maps make the network occupy the memory of GPU more efficiently. Maxpooling and sub-pixel interpolation layers also help the proposed network to have more important features while downscaling and upscaling. Experiments are executed for noisy images that have additive white Gaussian noise with specific noise levels: 15, 25, and 50. The experiments indicate that the proposed network reduces the Gaussian noise of noisy images better than conventional methods.

1. Introduction

There is a huge amount of information in images via televisions, the internet, and other media. However, in the real world, most of those are corrupted by some noises such as Gaussian noise and Poisson noise. Depending on the equipment which takes an image from the real world, the weather conditions and dust concentration, it is difficult to take a noise-free image in the real world. Therefore, image noise reduction or image denoising attracts many people to research. By reducing the noise of an image, recognizing some targets in the image more clearly is possible. Tasks using images such as object detection [1], image classification [2], and object segmentation [3] have more clear and accurate results by using the noise-free image.

Recently, image denoising is achieved using deep convolutional neural networks (CNN). CNN is a neural network making a new image by multiplying an input image and 2-d filters with weights. Before CNN, Dabov et al. [4] achieved image denoising by processing blocks in an image, matching similar blocks and filtering 3-d arrays that consists of similar blocks. Zhang et al. [5] proposed feed-forward denoising convolutional neural networks (DnCNN). DnCNN showed that reducing the noise of an image using deep learning algorithms has a good quality of results. Zhang et al. [6] also proposed a Fast and Flexible Solution for CNN-Based Image Denoising (FFDNet). FFDNet showed that the algorithm with the input of a noise level can reduce the noise and do not need to train the noisy image with a specific noise level every time. DnCNN and FFDNet have become benchmarks for many algorithms because of their performance and a few parameters. DnCNN and FFDNet have the number of network parameters between 10^5 and 10^6. Their primary difference is whether the algorithm should be trained to noise with a specific noise level.

In this paper, we also compare the proposed algorithm with a discrete wavelet denoising CNN (WDnCNN) by Zhao et al [7]. WDnCNN used discrete wavelet transform (DWT) and showed that balancing the low-frequency contents in the images is helpful for not only reducing additive white Gaussian noise (AWGN), but also real-world image denoising. WDnCNN also do not need to train the noisy image with a specific noise level. However, WDnCNN has less clear performance than DnCNN and FFDNet in lower noise level.

Conventional algorithms have good results for image denoising; however, DnCNN and FFDNet did not fully utilize their network parameters. Furthermore, WDnCNN did not solve the poor performance in noise level 15. Therefore, we propose a new end-to-end neural network named residual U-shaped network (RUN) for image denoising. The proposed RUN utilizes about 10 times more parameters and shows greater performance than DnCNN under the same memory of GPU. Residual learning [8] enables RUN to learn well in deeper condition. Thanks to the U-shaped structure [9] in RUN, the proposed network minifies the resolution of the feature maps and enlarge the number of feature maps; therefore, RUN can utilize the memory of GPU more efficiently. Finally, maxpooling and subpixel interpolation layers give the proposed network valuable feature maps helpful for training.

2. Proposed Algorithm

2.1 Residual learning

The proposed network is comprised of some residual blocks that enables the network to learn the difference between input image and output image by adding the input at the output of some convolutional layers, called residual learning. Residual learning is in the spotlight for those who train networks related to tasks for image restoration because designing deeper networks and improving performance become possible by training the difference between input image and output image and normalizing the differences to zero.

Recently, there are a lot of variations of residual blocks and we use residual blocks from Lim et al. [10, 11] and replace ReLU with PReLU [12] that gave good results on ImageNet classification. The residual blocks were also used by Yu et al. [13] and showed good results for JPEG artifacts removal. The structure of the residual blocks the proposed network uses is at Figure 1.

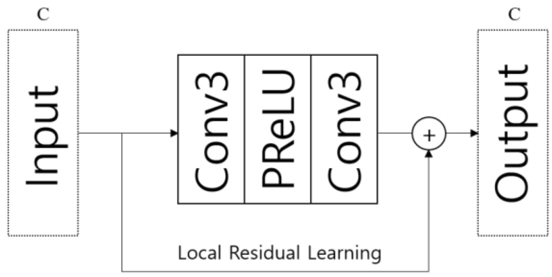

Figure 1. The structure of the residual block

2.2 U-shaped structure

The network that has U-shaped structure, also called U-net was first proposed for image segmentation [9]. Improving the results was possible by downscaling the feature maps and putting more convolutional layers because downscaling the feature maps reduces calculations of convolution provided that the kernel size of the convolutional layers the networks have is same. Recently, networks that has U-shaped structure also yield great results on image denoising and have been researched a lot [14, 15].

2.3 Scaling layers

The proposed network uses downscaling layer and upscaling layer for utilizing U-shaped structure. The scaling layers are used by Park et al. [15].

In downscaling layer, there are maxpooling layer, convolutional layer, and PReLU layer. Because the kernel size of maxpooling layer is 2, the resolution of the feature maps halves. To prevent the network from losing information, the number of output channels in convolutional layer in downscaling layer is double of that of input channels. By the composition of the downscaling layer, downscaling task can be trained well.

In upscaling layer, there are convolutional layer, PReLU layer, and sub-pixel interpolation layer first used by Shi et al. [16]. Because, the kernel size of maxpooling layer is 2, the resolution of the feature maps doubles and the number of channels of feature maps decreases four times less than before sub-pixel interpolation layer. By the composition of the upscaling layer, upscaling task can be also trained well.

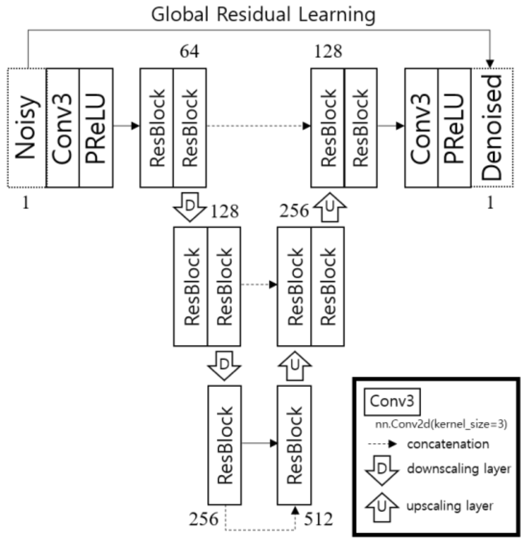

Figure 2. The structure of the proposed network

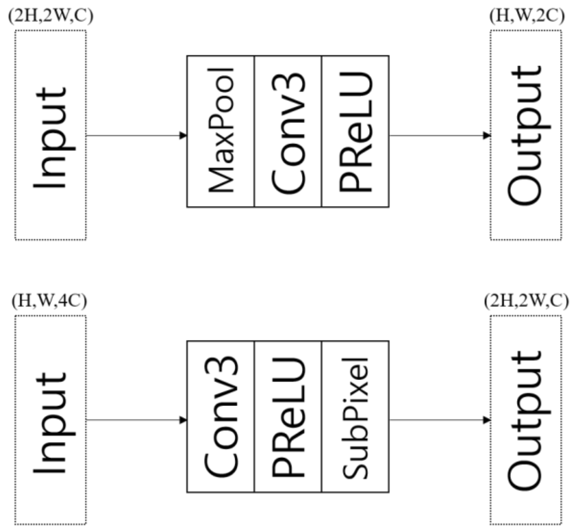

Figure 3. The structure of the scaling layers

3. Experiments and Analysis

To improve the results of image denoising by RUN, we used the DIV2K dataset [17] for training the proposed network. Furthermore, the datasets for testing the proposed network are BSD68 dataset and Set12 dataset that were used in DnCNN [5]. However, when testing the proposed network, we should set the height and width of an image to multiples of four because of two times of downscaling in the U-shaped structure. We didn't use flip & rotate on the training dataset and self-ensemble on the results of testing dataset to show the performance of the proposed RUN itself.

We trained the network using the datasets whose batch size is 32 and patch size is 96. The optimizer we used is ADAM optimizer [18] whose momentum is 0.9 and learning rate is 1e-4. The learning rate halves every 3 epochs, same as about 20k iterations. Finally, we set the loss to L1 loss because Zhao et al. [19] showed that L1 loss is best for image restoration tasks as a single loss.

We compare results from our proposed network with noisy images, those from BM3D [4], DnCNN [5], FFDNet [6], and WDnCNN [7] under the specific noise levels of AWGN: 15, 25, and 50.

Table 1. PSNR (dB) comparison of methods on grayscale image denoising

Methods	BSD68			Set12		
	Noise level			Noise level		
	15	25	50	15	25	50
Noisy	24.80	20.48	14.92	24.66	20.32	14.75
BM3D [4]	31.07	28.57	25.62	32.38	29.97	26.72
DnCNN [5]	31.72	29.21	26.21	32.84	30.36	27.11
FFDNet [6]	31.63	29.21	26.32	32.77	30.48	27.34
WDnCNN [7]	31.56	29.16	26.30	32.71	30.44	27.37
Proposed	**31.79**	**29.33**	**26.39**	**32.95**	**30.59**	**27.45**

Table 2. SSIM comparison of methods on grayscale image denoising in BSD68 datasets

Methods	BSD68		
	Noise level		
	15	25	50
Noisy	0.5808	0.4050	0.2099
BM3D [4]	0.8756	0.8071	0.6954
DnCNN [5]	0.8933	0.8327	0.7253
FFDNet [6]	0.8936	0.8346	0.7329
WDnCNN [7]	0.8928	0.8338	0.7329
Proposed	**0.8955**	**0.8375**	**0.7361**

Table 3. SSIM comparison of methods on grayscale image denoising in Set12 datasets

Methods	Set12		
	Noise level		
	15	25	50
Noisy	0.5553	0.3851	0.2055
BM3D [4]	0.8995	0.8564	0.7767
DnCNN [5]	0.9063	0.8656	0.7890
FFDNet [6]	0.9072	0.8700	0.7982
WDnCNN [7]	0.9063	0.8695	0.8002
Proposed	**0.9091**	**0.8720**	**0.8034**

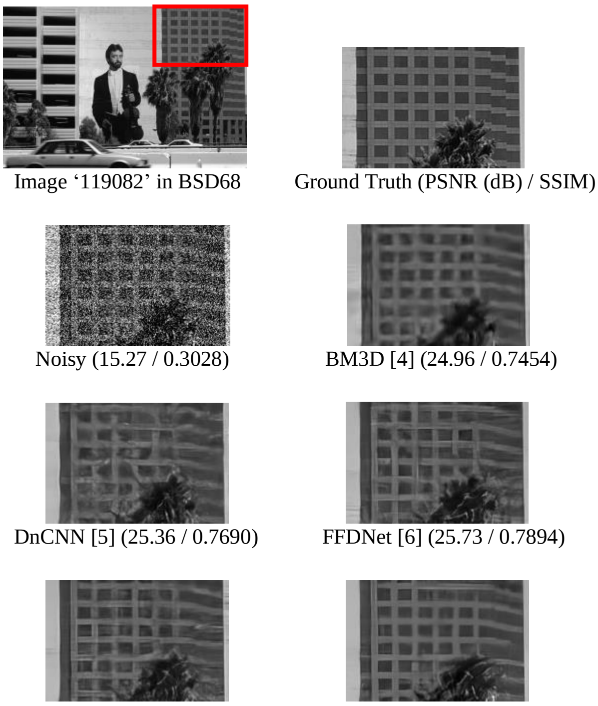

Image '119082' in BSD68 / Ground Truth (PSNR (dB) / SSIM)
Noisy (15.27 / 0.3028) / BM3D [4] (24.96 / 0.7454)
DnCNN [5] (25.36 / 0.7690) / FFDNet [6] (25.73 / 0.7894)
WDnCNN [7] (25.78 / 0.7899) / Proposed (25.82 / 0.7924)

Figure 4. Image denoising results at noise level 50

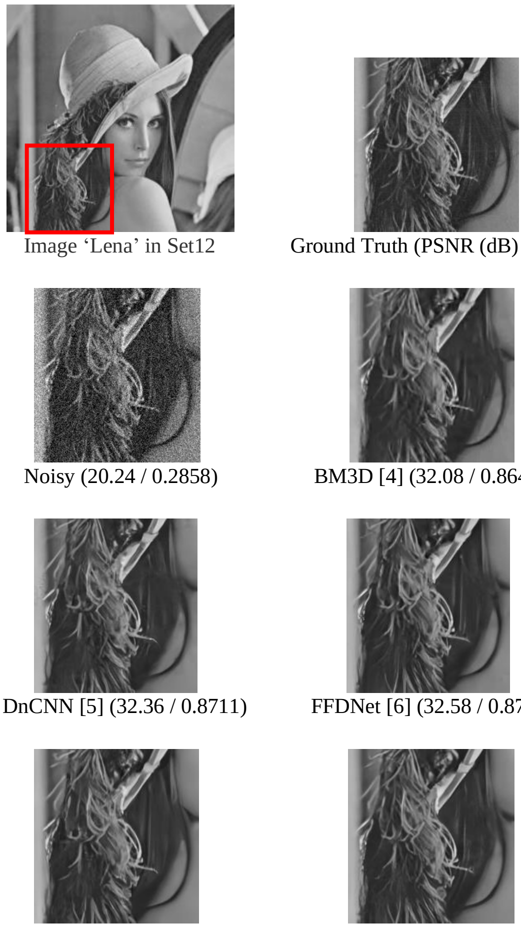

Image 'Lena' in Set12 / Ground Truth (PSNR (dB) / SSIM)
Noisy (20.24 / 0.2858) / BM3D [4] (32.08 / 0.8641)
DnCNN [5] (32.36 / 0.8711) / FFDNet [6] (32.58 / 0.8764)
WDnCNN [7] (32.64 / 0.8790) / Proposed (32.74 / 0.8807)

Figure 5. Image denoising results at noise level 25

The performance of them is measured in peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). Bolded results are the best results, and underlined results are the second-best results. The PSNR results between conventional algorithms and the proposed RUN are shown in Table 1. In addition, the SSIM results between conventional algorithms and the proposed RUN are shown in Table 2 and 3. DnCNN shows better results than FFDNet and WDnCNN at noise level 15 and vice versa at noise level 25 and 50; however, the proposed RUN produces a better not only quantitative results, but also qualitative results than any other conventional algorithms in this paper.

4. Conclusions

In this paper, we proposed a new end-to-end convolutional neural network for image denoising named RUN. Residual learning enables the proposed network to learn deeper without vanishing gradient problem. Furthermore, the proposed RUN fully utilized network parameters and the memory of GPU by adopting U-shaped structure. Scaling layers including maxpooling and sub-pixel layers helps RUN to extract information good for learning and keep the information from decreasing. In addition, we took a variation of patch size and learning rate considering the structure of the proposed network and the task of image denoising.

The proposed network showed greater results than any other previous algorithms in this paper. However, there is room for RUN to improve further. Because the performances of scaling layers depend on the tasks, there would be substitutes for maxpooling and sub-pixel. Furthermore, networks that have U-shaped structure have three or four of their scaling layers; however, RUN has two scaling layers to implement in the GPU, Quadro M4000.

Finally, as an end-to-end convolutional neural network for image restoration tasks, RUN can yield great results not only in grayscale image denoising, but also all image restoration tasks such as color image denoising, image demosaicing, single image super-resolution, and JPEG artifacts removal.

Acknowledgement

This work was supported by the Brain Korea 21 plus Project in 2014. (H0301-14-1018)

References

[1] X. Liu, Y. Lu, T. Wu, and T. Yuan, “An Improved Local Descriptor based Object Recognition in Cluttered 3D Point Clouds,” *International Journal of Computers Communications & Control*, vol. 13, no. 2, pp.221–234, 2018.

[2] H. Hosseini, B. Xiao, and R. Poovendran, “Google’s Cloud Vision API Is Not Robust To Noise,” International Conference on Machine Learning and Applications (ICMLA), 2017, pp.101-105.

[3] I. Despotović, V. Jelača, E. Vansteenkiste, and W. Philips, “Noise-Robust Method for Image Segmentation,” *International Conference on Advanced Concepts for Intelligent Vision Systems*, Berlin, Heidelberg, 2010, pp.153–162.

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising with block-matching and 3D filtering,” *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, International Society for Optics and Photonics, vol. 6064, p. 606414, 2006.

[5] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp.3929-3938.

[6] K. Zhang, W. Zuo, and L. Zhang, “FFDNet: Toward a fast and flexible solution for CNN-based image denoising,” *IEEE Transactions on Image Processing*, vol.27, no.9, pp.4608-4622, 2018.

[7] R. Zhao, K. M. Lam, and D. P. Lun, “Enhancement of a CNN-Based Denoiser Based on Spatial and Spectral Analysis,” *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp.1124-1128.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp.770-778.

[9] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp.234-241.

[10] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp.136-144.

[11] H. Kim, M. Choi, B. Lim, and K. Mu Lee, “Task-Aware Image Downscaling,” *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp.399-414.

[12] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” *Proceedings of the IEEE international conference on computer vision*, 2015, pp.1026-1034.

[13] S. Yu and J. Jeong, “Local Excitation Network for Restoring a JPEG-Compressed Image,” *IEEE Access*, vol.7, pp.138032-138042, 2019.

[14] S. Yu, B. Park, and J. Jeong, “Deep Iterative Down-Up CNN for Image Denoising,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp.0-0.

[15] B. Park, S. Yu, and J. Jeong, “Densely Connected Hierarchical Network for Image Denoising,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp.0-0.

[16] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp.1874-1883.
[17] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 126-135.
[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[19] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol.3, no.1, pp.37-47, 2016.