Image Denoising with Local Dense and Adaptive Global Residual Networks

Lulu Sun[1], Yongbing Zhang[1(✉)], Chenggang Yan[2], Xiangyang Ji[1,3],
Xinhong Hao[4], Yongdong Zhang[5], and Qionghai Dai[1,3]

[1] Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China
zhang.yongbing@sz.tsinghua.edu.cn
[2] Institute of Information and Control,
Hangzhou Dianzi University, Hangzhou 310018, China
[3] Department of Automation, Tsinghua University, Beijing 100084, China
[4] Science and Technology on Mechatronic Dynamic Control Laboratory,
Beijing Institute of Technology, Beijing 100081, China
[5] School of Information Science and Technology,
University of Science and Technology of China, Hefei 230026, China

Abstract. Residual Networks (ResNet) and Dense Convolutional Networks (DenseNet) have shown great success in lots of high-level computer vision applications. In this paper, we propose a novel network with Local Dense and Adaptive Global Residual (LD+AGR) frameworks for fast and accurate image denoising. More precisely, we combine local residual/dense with global residual/dense to investigate the best performance dealing with image denoising problem. In particular, local/global residual/dense means the connection way of inner/outer recursive blocks. And residual/dense represents combining layers by summation/concatenation. Furthermore, when combining skip connections, we add some adaptive and trainable scaling parameters, which could adjust automatically during training to balance the importance of different layers. Numerous experiments demonstrate that the proposed network performs favorably against the state-of-the-art methods in terms of quality and speed.

Keywords: Adaptive global residual · Local dense · Image denoising

1 Introduction

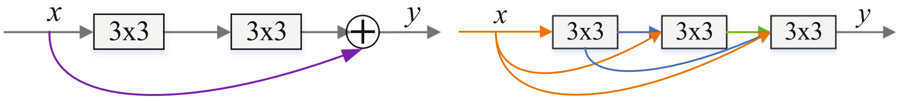
Image denoising, which aims to recover a clear image from its degraded observation caused by noise contamination, is a classic and fundamental problem in computer vision [12–14]. Since image denoising is highly ill-posed, it is very challenging to achieve satisfactory results.

This work was partially supported by the National Science Foundations of China under Grant 61571254, Guangdong Natural Science Foundation 2017A030313353, and Shenzhen Fundamental Research fund under Grant JCYJ20170817161409809.

R. Hong et al. (Eds.): PCM 2018, LNCS 11164, pp. 27–37, 2018.
https://doi.org/10.1007/978-3-030-00776-8_3

Numerous image denoising methods have been proposed [1–4, 9, 21–23, 26] in recent years with fantastic advancements. Most denoising methods are based on nonlocal self-similarity (NSS) priors [1, 3, 4, 18, 24]. NSS refers to the fact that a local patch often has many nonlocal similar patches across the image. Nonlocal means (NLM) [1] could be considered as a seminal work, bringing the new era of denoising by finding the NSS priors within a search window sliding across the image. It obtained a denoised patch by weighted averaging all other patches in the search window. Another famous benchmark, named block-matching and 3D filtering (BM3D) [4], remarkably combined NSS with an enhanced sparse representation in transform domain. It contained two general procedures: grouping and collaborative filtering. First, forming a 3D array by stacking together similar blocks. Second, obtaining 2D estimates of grouped blocks after performing collaborative filtering of the group. Instead of transforming images to other domains, low rank matrix approximation methods also attracted great attention in recent years. Representative and significant low-rank method was weighted nuclear norm minimization (WNNM) [9]. Based on the general prior knowledge that the larger singular values of the patch matrices of original image are more important than the smaller ones, WNNM achieved great success in image denoising.

Recently, methods based on neural networks [5–7, 11, 15, 17, 19] have shown significant success in many computer vision tasks, especially in image classification. Among these methods, Residual Networks (ResNet) [11, 25] and Dense Convolutional Networks (DenseNet) [15] are attracting the most attention. Inspired by such achievements, we try to investigate the properties of the two architectures: residual and dense. In this paper, we not only combine the two elements in terms of local/global way, but also adding adaptive parameters to keep a good balance when combining various skip connections (Fig. 1).



(a) Residual unit in [11]. (b) Dense unit in [15].

Fig. 1. Comparisons of residual/dense units. 3 × 3 represents a convolutional layer with kernel size of 3 × 3. ⊕ indicates the summation way of connecting. As shown in Fig. 3(b), many lines focusing on one point is the concatenation way for combination. Lines with the same color share the same value. The following symbols have the same meanings. (Color figure online)

2 Discussion of ResNet/DenseNet

ResNets are usually composed of lots of residual blocks, which only contains one skip connection and two convolutional layers. Such a simple residual architecture is easy to train. However, these units lack enough power to transmit sufficient information merely through cascading, leading to the lower ability of the whole network. Especially when dealing with image processing problems, these networks are not strong enough to extract features from massive data. In addition, it is likely to lose useful information during the process of deep-layers of delivery without any effective connection.

In contrast, DenseNets have plenty of skip connections in one dense block and the dense block is diverse to be able to simulate complex functions, which is beneficial to learn features. However, one big problem is that such powerful networks lack efficient contacts among outputs of each block. This will increase the time consumption of training. What is worse, no connections between blocks will cause some distortion when transmitting features.

Taking into account the shortcomings owned by single ResNet/DenseNet separately mentioned above, we are going to combine the two elements in two ways: local and global, which will be explained completely in the following sections.

3 Local Residual/Dense and Global Residual/Dense Networks

3.1 Local Residual and Global Residual Networks (LR+GR)

As shown in Fig. 2, both the local recursive block and global connecting way are the residual manner. So we name this style of framework as local residual and global residual networks (LR+GR). Normally, the first and last 3×3 convolutional layers are usually used for extracting features and reconstruction separately. In detail, this network is composed of three residual blocks, three inner and three outer identity skip connections, and two convolutional layers. In particular, we use parametric rectified linear unit (PReLU) [10] as activation function in all networks, which are omitted in the figures for simplicity.

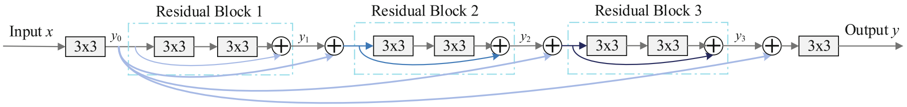


Fig. 2. Framework of local residual and global residual network.

3.2 Local Residual and Global Dense Networks (LR+GD)

From Fig. 3, we can see that the inner connecting way of each block is residual while the global manner is dense. Similarly, this kind of architecture is named as local residual and global dense networks (LR+GD). Particularly, there are three residual blocks and one summation skip connection in each unit. From the overall point of view, it uses dense style and there are six concatenating shortcuts.

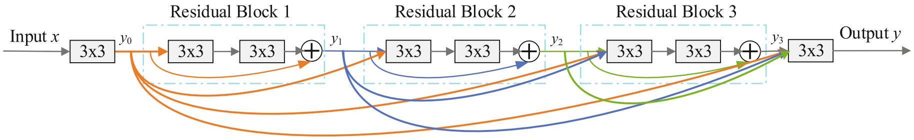


Fig. 3. Framework of local residual and global dense network.

3.3 Local Dense and Global Residual Networks (LD+GR)

If the recursive units are dense style while the global way is residual skip connection, we would call this framework as local dense with global residual network (LD+GR), as shown in Fig. 4. In particular, there are two dense blocks, two residual shortcuts and two convolutional layers in this network, and each block contains three convolutional layers and three dense skip connections.

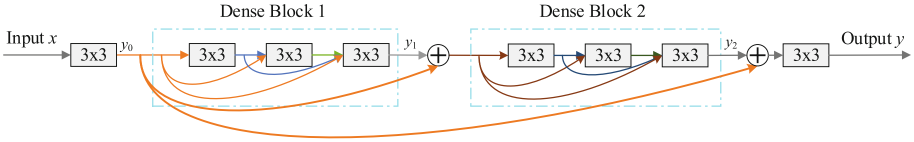


Fig. 4. Framework of local dense and global residual network.

3.4 Local Dense and Global Dense Networks (LD+GD)

Local dense with global dense networks (LD+GD) represent such frameworks that both inner and outer connections of blocks are dense, as shown in Fig. 5. To be specific, there are three concatenating lines in each dense unit and three skip connections in a global view.

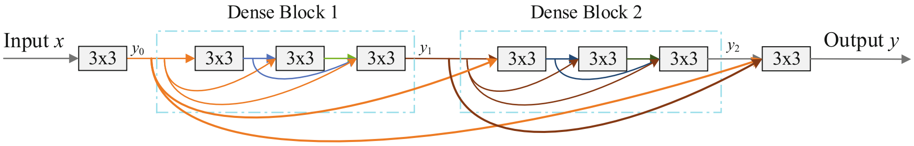


Fig. 5. Framework of local dense and global dense network.

4 Local Residual/Dense and Adaptive Global Residual Networks

4.1 Local Residual and Adaptive Global Residual Networks (LR+AGR)

Based on the framework of LR+GR, adding some trainable variables before summation, the network will become local residual and adaptive global residual network (LR+AGR). Seeing Fig. 6, there are three extra pairs of scaling parameters compared to the above LR+GR in Fig. 2.

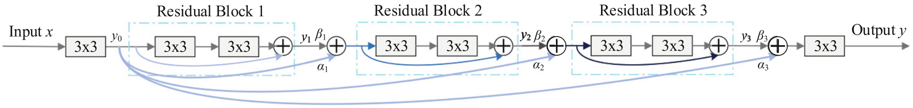


Fig. 6. Framework of local residual and adaptive global residual network.

4.2 Local Dense and Adaptive Global Residual Networks (LD+AGR)

Similarly, on the basis of LD+GR in Fig. 4, if we add some adaptive scaling parameters at the output of each dense block to balance the importance of each part automatically, the framework will become local dense and adaptive global residual network (LD+AGR), as shown in Fig. 7. We could see two pairs of scaling parameters after two dense blocks.

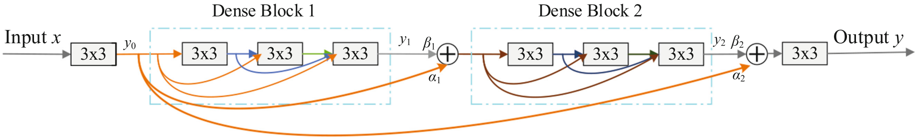


Fig. 7. Framework of local dense and adaptive global residual network.

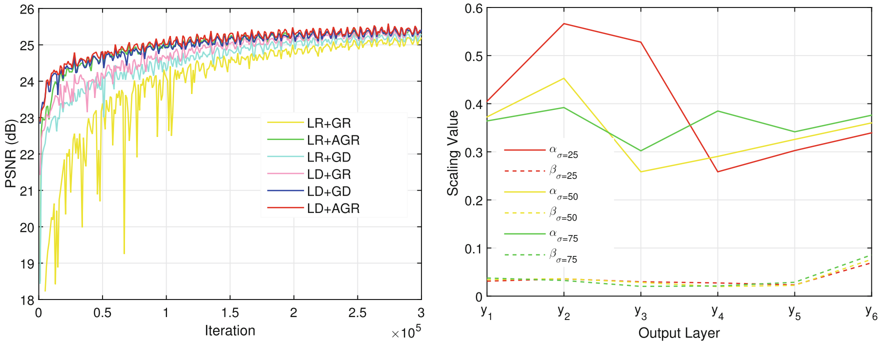
4.3 Analysis and Discussions

In order to investigate more properties of the four basic frameworks and two adaptive ones mentioned above, we conducted the image denoising experiments using these networks. The training process has been recorded in Fig. 8(a). We controlled all the variables the same except the frameworks. As iteration increases, they are going to converge. Clearly, LD+AGR has the fastest convergence speed and achieves the best value at last. The following are LD+GD, LR+AGR, LD+GR, LR+GD, and LR+GR. Compared to LD+GR, LD+AGR has superior performance, which fully demonstrates the importance of introducing the adaptive and trainable scaling parameters.

5 The Proposed LD+AGR Networks for Image Denoising

5.1 Architecture

Referring to the framework of LD+AGR, we build the improved network, as shown in Fig. 9(b). It is composed of six dense blocks and six adaptive residual



(a) (b)

Fig. 8. (a) PSNR(dB) comparisons of six frameworks during training. (b) Adaptive α and β of different output layers in our LD+AGR.

skip connections. Focusing on one dense block (See Fig. 9(a)), there are six 3×3 convolutional layers for learning features continuously, one 1×1 convolutional layer for decreasing the dimension of feature mappings, and fifteen dense lines for concatenating features together. The biggest difference is that we introduce two adaptive scaling parameters outside each dense block to adjust the importance of the first output y_0 and the latter output $y_i (i = 1, ..., 6)$. As for the number of convolutional layers in each dense block and total blocks, we choose seven (including the 1×1 convolutional layer) and six separately in this paper.

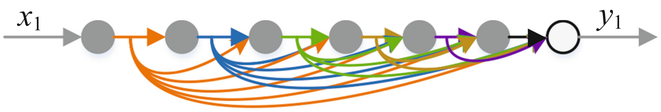
5.2 Adaptive Parameters

We trained three models for image denoising with noise level $\sigma = 25$, 50, and 75 using our LD+AGR framework. The learned parameters α and β of different layers can be observed in Fig. 8(b). Intuitively, all αs are much bigger than βs, which means the original output y_0 plays a more important role than latter output layers. Moreover, all αs change rapidly while all βs shake slowly and softly. But the last output y_6 seems to be more important than the other five ones. We also conducted such experiments on the condition that all αs and βs are 0.5, but the denoising performance is far worse than the adaptive ones.

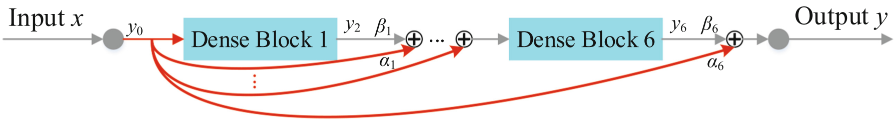
6 Experiments

In this section, we compare the proposed LD+AGR image denoising model with several state-of-the-art denoising methods, including BM3D [4], EPLL [26], WNNM [9], MLP [2], and PCLR [3]. The implementations are all from the publicly available codes provided by the authors.[1]

[1] The source code of the proposed method will be available after this paper is published.



(a)



(b)

Fig. 9. (a) Dense block of our network. (b) Architecture of our LD+AGR networks. Grey circles represent 3×3 convolutional kernels, and white circle is 1×1 convolutional kernel.

Fig. 10. The 14 test images (grey, 256×256). From left to right: Baboon, Barbara, Boat, Couple, Hill, Lena, Monarch, R.R.Hood, Pentagon, Starfish, Cameraman, Man, Paint-full, Parrots.

6.1 Training Details

We use Berkeley Segmentation Dataset BSD500 [20] as the training set and 14 widely used test images as the testing set (It can be found in Fig. 10). To increase the training set, we segment these images to overlapping patches of size 50×50 with stride of 10. We use the deep learning library Tensorflow on an NVIDIA GTX TITAN X GPU with 3072 CUDA cores and 12 GB of RAM to implement all operations in our network. The filter weights are initialized using the "Xavier" strategy [8] and biases are generated by tf.constant initializer using Tensorflow. We use Adam [16] algorithm to optimize the loss function of Mean Square Error (MSE).

6.2 Quantitative Results

We record PSNR comparisons to other state-of-the-art algorithms on noise level σ=25, 50, and 75 in Table 1. On the whole, our LD+AGR has the overwhelming superiority over the other methods on average, especially when σ= 25 and 50, the superiority can reach up to 0.33 dB and 0.36 dB over the second best methods on PSNR.

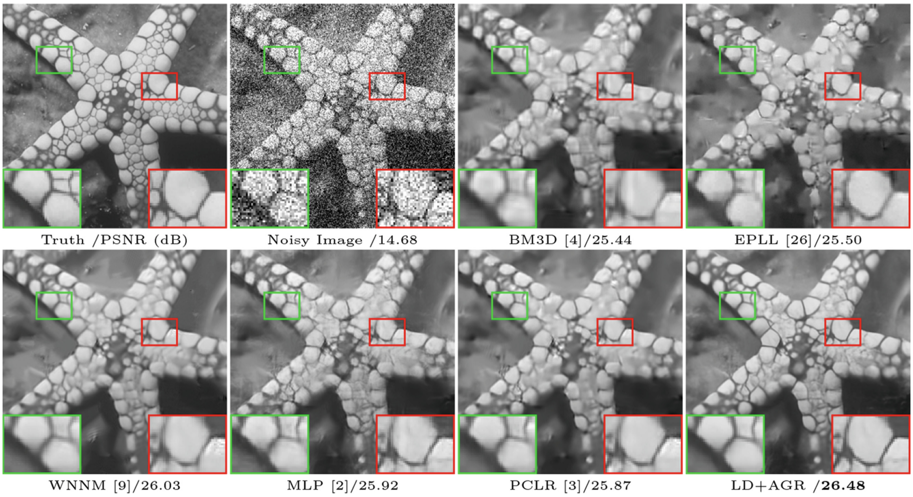
From Table 1, on average, we have the best results on three noise levels. Concretely, among 14 testing images, there are 13, 14, and 8 reconstructed images by our methods achieve the best performance. Hence, no matter on the whole or individuals, our LD+AGR shows tremendous advance over other methods in terms of PSNR.

Table 1. PSNR (dB) results with different σ over testing set (See Fig. 10). The best result for each image is **highlighted**.

	$\sigma = 25$														
Image / Method	Baboon	Barbara	Boat	Couple	Hill	Lena	Mona.	R.R.H.	Pent.	Star.	C.man	Man	Paint.	Parrots	Ave.
BM3D[4]	26.51	29.54	28.42	28.44	29.06	30.36	28.88	30.73	28.21	28.71	29.08	28.43	29.03	31.55	29.07
EPLL[26]	26.68	28.91	28.72	28.42	29.03	30.29	29.34	30.63	28.07	28.97	29.25	28.73	29.28	31.59	29.14
WNNM[9]	26.81	**30.20**	28.77	28.64	29.29	30.87	29.84	31.22	28.72	29.54	29.64	28.73	29.51	32.12	29.56
MLP[2]	26.87	29.14	28.85	28.75	29.31	30.79	29.62	30.93	28.23	29.26	29.30	28.80	29.65	32.10	29.40
PCLR[3]	26.75	29.73	28.88	28.73	29.30	30.78	29.75	30.98	28.42	29.35	29.68	28.74	29.58	32.01	29.48
LD+AGR	**27.06**	29.52	**29.27**	**29.22**	**29.53**	**31.22**	**30.16**	**31.55**	**28.77**	**30.09**	**29.95**	**29.10**	**30.18**	**32.64**	**29.88**
	$\sigma = 50$														
Image / Method	Baboon	Barbara	Boat	Couple	Hill	Lena	Mona.	R.R.H.	Pent.	Star.	C.man	Man	Paint.	Parrots	Ave.
BM3D[4]	24.07	26.33	25.25	25.17	26.22	27.05	25.62	27.75	25.35	25.44	25.98	25.55	25.66	28.32	25.98
EPLL[26]	24.13	26.02	25.48	25.09	26.14	26.93	25.76	27.58	25.02	25.50	26.04	25.59	25.88	28.00	25.94
WNNM[9]	24.21	26.64	25.54	25.27	26.35	27.47	26.32	28.13	25.72	26.03	26.45	25.71	25.97	28.71	26.32
MLP[2]	24.31	26.41	25.68	25.49	26.47	27.54	26.24	28.15	25.48	25.92	26.37	25.87	26.21	29.01	26.37
PCLR[3]	24.12	26.44	25.57	25.27	26.26	27.38	26.25	28.02	25.34	25.87	26.60	25.74	26.05	28.70	26.26
LD+AGR	**24.52**	**26.80**	**26.07**	**25.78**	**26.67**	**27.81**	**26.88**	**28.64**	**25.83**	**26.48**	**26.77**	**26.03**	**26.60**	**29.35**	**26.73**
	$\sigma = 75$														
Image / Method	Baboon	Barbara	Boat	Couple	Hill	Lena	Mona.	R.R.H.	Pent.	Star.	C.man	Man	Paint.	Parrots	Ave.
BM3D[4]	23.05	24.39	23.61	23.56	24.64	25.20	23.69	25.96	23.85	23.71	24.17	23.99	23.81	26.33	24.28
EPLL[26]	22.94	24.04	23.77	23.44	24.50	24.97	23.64	25.84	23.45	23.76	24.21	24.07	23.89	26.11	24.19
WNNM[9]	23.18	24.70	23.87	23.65	24.74	25.72	24.31	26.19	**24.21**	24.13	24.60	24.10	24.07	26.69	24.58
MLP[2]	23.30	**24.76**	24.11	**23.92**	**24.97**	**25.79**	24.40	26.41	24.13	24.16	24.67	**24.39**	**24.40**	**27.08**	24.75
PCLR[3]	22.99	24.58	23.87	23.64	24.64	25.52	24.28	26.28	23.79	24.01	24.78	24.12	24.15	26.74	24.53
LD+AGR	**23.33**	**24.76**	**24.30**	23.83	24.94	25.76	**24.83**	**26.69**	**24.21**	**24.32**	**24.98**	24.29	24.39	27.05	**24.83**

6.3 Visual Quality

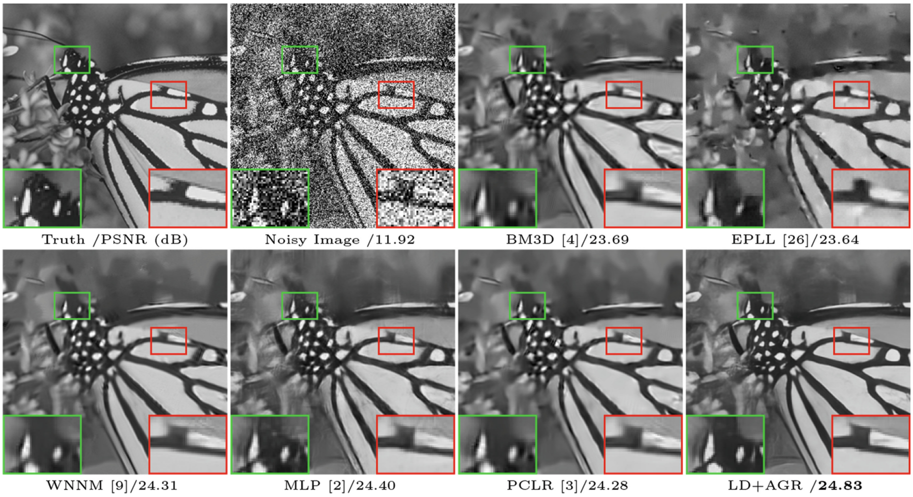
As shown in Fig. 11, similarly, our LD+AGR has the best visual quality compared to other methods. Especially, in the green and red windows, it is easy for us to recognize lines and shapes of the starfish in our result. Even with the noise level $\sigma = 75$, our method can still recover the most valuable information, which can be found in Fig. 12. In the green window, the head of butterfly in our



Truth /PSNR (dB) Noisy Image /14.68 BM3D [4]/25.44 EPLL [26]/25.50

WNNM [9]/26.03 MLP [2]/25.92 PCLR [3]/25.87 LD+AGR /**26.48**

Fig. 11. Sample image denoised results on *Starfish* with state-of-the-art methods ($\sigma = 50$). (Color figure online)



Truth /PSNR (dB) Noisy Image /11.92 BM3D [4]/23.69 EPLL [26]/23.64

WNNM [9]/24.31 MLP [2]/24.40 PCLR [3]/24.28 LD+AGR /**24.83**

Fig. 12. Sample image denoised results on *Monarch* with state-of-the-art methods ($\sigma = 75$). (Color figure online)

recovered image is distinct from others. Likewise, in the red block, our pattern is also much sharper than the others. In a word, from the view of visual quality, our LD+AGR performs better than other state-of-the-art image denoising methods.

6.4 Running Time

We profile the time consumption of all the methods in a Matlab 2015b environment using the same machine (an NVIDIA GTX TITAN X GPU with 3072 CUDA cores and 12 GB of RAM) in Table 2. Obviously, based on the adaptive networks, our method has enormous advantage than all the traditional algorithms.

Table 2. Average running time (s) for one image with different noise level σ over testing set (See Fig. 10). The best result for each dataset is **highlighted**.

σ	BM3D [4]	EPLL [26]	WNNM [9]	MLP [2]	PCLR [3]	LD+AGR
$\sigma = 25$	41.78	46.29	160.71	2.22	69.68	**1.25**
$\sigma = 50$	44.08	45.17	114.96	2.18	124.71	**1.25**
$\sigma = 75$	56.47	45.59	180.36	2.17	130.79	**1.25**
Average	47.44	45.68	152.01	2.19	108.39	**1.25**

7 Conclusions

In this paper, we address the image denoising problem via a local dense and adaptive global residual (LD+AGR) network which learns high effective features to reconstruct the latent clean images from the corresponding noisy ones. Moreover, we introduce adaptive scaling parameters to balance the importance of different outputs. Experimental results fully illustrate the effectiveness of the proposed method, which outperforms state-of-the-art methods by a considerable margin in terms of PSNR. Noticeable improvements can also visually be found in the reconstruction results.

References

1. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: CVPR, pp. 60–65 (2005)
2. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: can plain neural networks compete with BM3D? In: CVPR, pp. 2392–2399 (2012)
3. Chen, F., Zhang, L., Yu, H.: External patch prior guided internal clustering for image denoising. In: ICCV, pp. 603–611 (2015)
4. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007)
5. Dong, C., Deng, Y., Change Loy, C., Tang, X.: Compression artifacts reduction by a deep convolutional network. In: ICCV, pp. 576–584 (2015)
6. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10593-2_13
7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR, pp. 580–587 (2014)
8. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
9. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: CVPR, pp. 2862–2869 (2014)
10. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: ICCV, pp. 1026–1034 (2015)
11. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks (2016). arXiv preprint: arXiv:1603.05027
12. Hong, R., Hu, Z., Wang, R., Wang, M., Tao, D.: Multi-view object retrieval via multi-scale topic models. IEEE Trans. Image Process. **25**(12), 5814–5827 (2016)
13. Hong, R., Zhang, L., Tao, D.: Unified photo enhancement by discovering aesthetic communities from flickr. IEEE Trans. Image Process. **25**(3), 1124–1135 (2016)
14. Hong, R., Zhang, L., Zhang, C., Zimmermann, R.: Aesthetic tendency discovery by multi-view regularized topic modeling. IEEE Trans. Multimed. **18**(8), 1555–1567 (2016)
15. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks (2016). arXiv preprint: arXiv:1608.06993

16. Kingma, D., Ba, J.: Adam: A method for stochastic optimization (2014). arXiv preprint: arXiv:1412.6980
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
18. Liu, H., Xiong, R., Zhang, J., Gao, W.: Image denoising via adaptive soft-thresholding based on non-local samples. In: CVPR, pp. 484–492 (2015)
19. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431–3440 (2015)
20. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV, pp. 416–423 (2001)
21. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: ICML, pp. 1096–1103 (2008)
22. Mao, X.-J., Shen, C., Yang, Y.-B.: Image restoration using very deep fully convolutional encoder-decoder networks with symmetric skip connections. In: NIPS (2016)
23. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: NIPS, pp. 341–349 (2012)
24. Xu, J., Zhang, L., Zuo, W., Zhang, D., Feng, X.: Patch group based nonlocal self-similarity prior learning for image denoising. In: ICCV, pp. 244–252 (2015)
25. Zhang, Y., Sun, L., Yan, C., Ji, X., Dai, Q.: Adaptive residual networks for high-quality image restoration. IEEE Trans. Image Process. **27**(7), 3150–3163 (2018). https://doi.org/10.1109/TIP.2018.2812081
26. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: ICCV, pp. 479–486 (2011)