Scale Input Adapted Attention for Image Denoising Using a Densely Connected U-Net: SADE-Net*

Vedat Acar[1][0000-0002-8321-2070] and Ender M. Eksioglu[2][0000-0002-7869-4159]

[1] Graduate School, Istanbul Technical University, Istanbul, Turkey
acarv19@itu.edu.tr

[2] Electronics and Communication Engineering Department, Istanbul Technical University, Istanbul, Turkey
eksioglue@itu.edu.tr

Abstract. In this work, we address the problem of image denoising using deep neural networks. Recent developments in convolutional neural networks provide a very potent alternative for image restoration applications and in particular for image denoising. A particularly popular deep network structure for image processing are the auto-encoders which include the U-Net as an important example. U-Nets contract and expand feature maps repeatedly, which leads to extraction of multi scale information as well as an increase in the effective receptive field when compared to conventional convolutional nets. In this paper, we propose the integration of a multi scale channel attention module through a U-Net structure as a novelty for the image denoising problem. The introduced network structure also utilizes multi scale inputs in the various substages of the encoder module in a novel manner. Simulation results demonstrate competitive and mostly superior performance when compared to some state of the art deep learning based image denoising methodologies. Qualitative results also indicate that the developed deep network framework has powerful detail preserving capability.

Keywords: Deep Learning · Convolutional Neural Networks · Image Denoising.

1 Introduction

Image denoising is one of the fundamental, low level tasks of computer vision. Noise is a commonly encountered distortion in digital images. There are several types of possible noise distributions encountered for vision data including additive white Gaussian noise (AWGN), Poisson noise, shot noise etc. A noisy digital image can be formulated as $y = x + v$, and the aim of the denoising process is to recover x from y. For this work, v is assumed to be AWGN.

* This work is supported by TUBITAK (The Scientific and Technological Research Council of Turkey) under project no. 119E248.

High-level computer vision tasks such as image classification, object detection, and segmentation have made significant advances with the introduction of deep Convolutional Neural Networks (CNNs). CNNs have attracted quite an interest due to their strong representation ability and wide applicability, leading to much improved results compared to conventional methods. The training of deep networks is a challenging issue due to the rise of the vanishing gradient problem with increasing depth. One structure which handles this issue is the ResNet which proposes residual connections to provide a better information flow [6]. Another example is the DenseNet which restrengthens connections through every layer [8]. Although these algorithms aid the vanishing gradient problem, there are more powerful modules to provide low loss feature transference. Channel attention is one of the popular such blocks, and it acts as an effective plug-and-play module. Another very recent approach is the use of smaller versions of the input which are named as scale inputs. These scale inputs can be used in the lower scales of the encoder to add more features to the encoder-decoder network. In this paper, we propose a U-Net architecture with attention layers, and the architecture also benefits from scale inputs in a novel manner. We will call this new structure as Scale input Attentive Network with Dense connections, namely SADE-Net.

2 Prior Art

2.1 Image Denoising

Image denoising is a fundamental task in image restoration, and its main aim is to preserve details while suppressing noise. There have been various methods tackling this problem. These have included transform domain methods and non local methods [2]. One particular algorithm which utilizes non-local similarities and transform methods together is the BM3D [4]. BM3D searches similar patterns in the image patches to process them together in a 3D transform. DnCNN was the earliest algorithm which combined denoising with CNNs [21]. DnCNN utilized a residual learning strategy to obtain better information flow and batch normalization to accelerate the training. The FFDNet algorithm on the other hand uses both the noise map and noisy image's subsamples [22] . Hence, FFDNet feeds both the noise map and noisy image's subsamples to the network to handle the problem of working in global noise level environment.

Recent deep network frameworks incorporate new structures and modules such as channel attention [1], non local blocks [19] or memory blocks [17]. Several state-of-the-art networks have benefited from the use of such novel blocks in the image denoising setting. We will give a short list of some of the well performing examples. MemNet structure proposed a memory network which incorporates short term and long term memory to cope with the long range dependency problem [17]. MWCNN adopted a wavelet transform strategy to upscale and downscale feature maps in a modified U-Net [11]. This structure demonstrated the efficiency of wavelet transform which avoids the detrimental gridding effects. RIDNet proposed a blind real image denoising network which utilized a residual-in-residual structure [1]. This network implemented feature attention inside Enhanced Attention Mechanism (EAM) blocks. PANET network on the other hand proposed pyramid attention blocks to better obtain long range correspondences of features and adopted a multi scale self-similarity prior [13].

2.2 Channel Attention

Attention mechanisms have become a quite popular ingredient which deep networks utilize for computer vision. Attention modules can model dependencies over longer distances, and their origins are motivated from human perception characteristics [3]. Treating all of the extracted feature maps in the same manner seems to hamper the discriminative power and the representation ability of the networks. An attention mechanism causes the deep neural network to focus its learning effort on more informative components of the input data by putting differing emphasis on different feature maps.

Attention modules are also rather lightweight, because they often utilize only two 1 × 1 convolutions. In this work, we have incorporated channel attention modules into a U-Net structure devised for gray level image denoising. Inspired by [14] and [9], we place the attention mechanism right next to the downsampling blocks, and the attention module outputs are transferred to the upsampling side by skip connections. We used the squeeze and excitation mechanism [7], which first applies global average pooling and extracts global spatial information. Afterwards, this mechanism uses two convolutions to capture the feature channel dependencies. Lastly, the input feature maps are rescaled by multiplying them with the obtained coefficients.

The mathematical description of the employed channel attention module is as follows. Let us consider f_c which carries features created by a convolutional layer having c feature maps of size $h \times w$. We first obtain global statistics of the feature maps.

$$g_p = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} f_c(i,j) \tag{1}$$

Here, $f_c(i,j)$ is the value of feature map at position (i,j). We implement an additional gating process to better exploit the channel dependencies.

$$s_g = \alpha(C_1(\delta(C_2(g_p)))) \tag{2}$$

C_1 and C_2 are the kernels to expand and contract the channels. δ is the ReLu operation, and α denotes the sigmoid function. Lastly, we rescale input f_c with s_g to obtain the final statistics.

$$\hat{f} = f_c \times s_g \tag{3}$$

The graphical description of the described channel attention structure is given in Fig. 1.

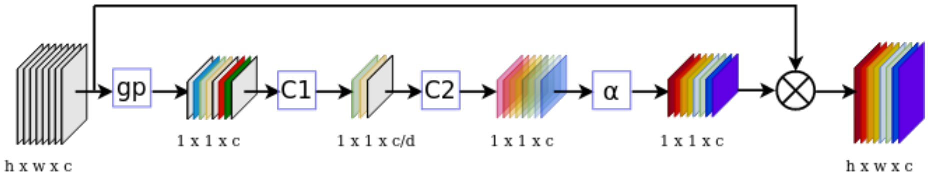


Fig. 1. Channel Attention Mechanism.

2.3 Scale Input for U-Nets

Traditional U-Net type networks only consider the overall input image, but not the subbands or scaled versions of this input. When one considers pyramid networks such as PANET [13], these networks also consider multi scale versions of the original input. These additional inputs give the network extra information and supervision to work from. These additional inputs do not burden the computational complexity of the network, while in general boosting the performance. In the new network here, we employ the rather recently introduced scale input approach [16]. In our strategy, the downsampled outputs at the encoder side get concatenated with subscaled versions of the noisy input image.

3 A Novel Network for Image Denoising: SADE-Net

In this section we detail the building blocks for the novel image denoising architecture as introduced here, namely "Scale input Attentive Network with Dense connections" (SADE-Net). Fig. 2 depicts the complete architecture of the novel SADE-Net framework for image denoising. As can be seen from Fig. 2, after the noisy image enters the network, a 1 × 1 convolutional layer followed by a parametric rectified linear unit (PReLu) extracts the initial features from the image. Then, two Densely Connected Residual (DCR) blocks transmit the information further. After the DCR blocks, the feature maps are downsampled by a ratio of two by using max pooling. The number of feature maps are also doubled at this stage. This operation is repeated three times for the encoder stage, leading to four distinct resolution levels. Every resolution scale has two DCR blocks both at the encoder and the decoder side.

At each resolution level of the encoder stage, downsampled versions of the noisy input image (scale inputs) get concatenated with the outputs of the downsampling stage. The scale inputs include the same number of maps as the outputs of downsampling. They are produced via a convolution kernel with 1 × 1 size and unit stride. After the concatenation of scale input and downsampled output, again a 1 × 1 convolution is used to halve the number of feature maps. In the decoder side, we utilized pixel shuffling to upsample the features by a ratio of two. We also transferred the input of the downsampling block to the encoder side via skip connections for all resolution levels. The channel attention block is applied inside this skip connections linking the encoder and decoder sides.

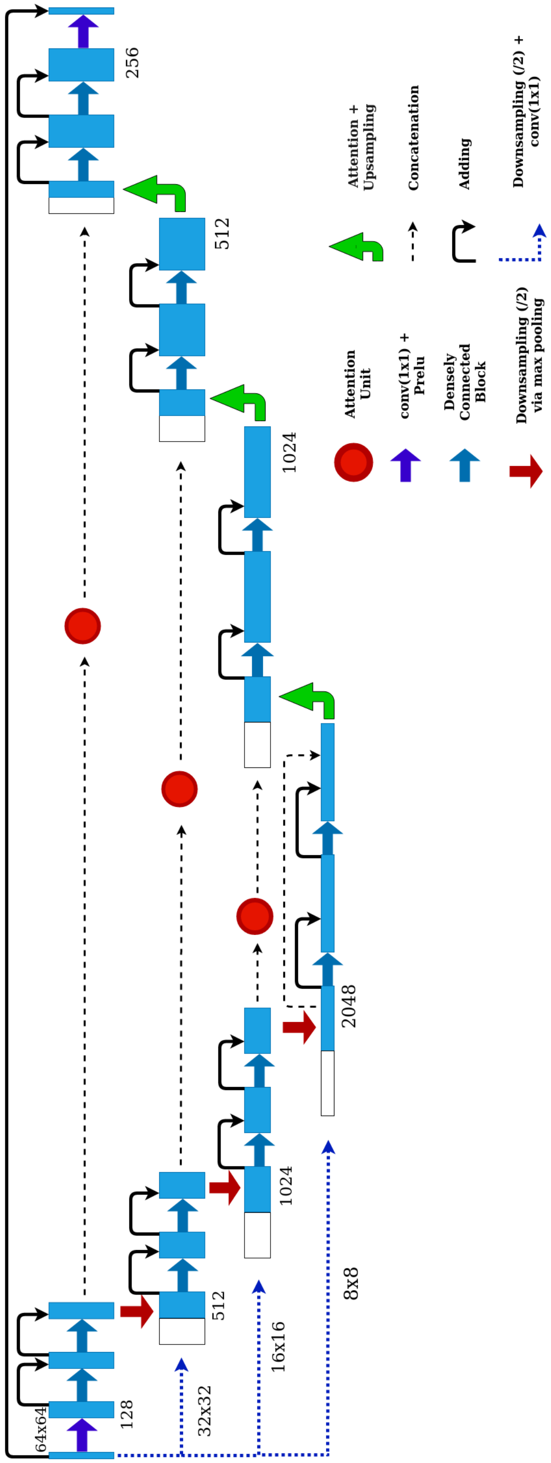


Fig. 2. Overall Architecture of the Proposed Network.

The overall structure of the employed channel attention block is again given in Fig. 1. The further stages at the decoder side reverse the actions of the decoder side. Finally, a global residual connection provides the final denoised output by adding the residual output of the network to the input noisy image as shown in the final stage of Fig. 2.

Table 1. Quantitative denoising results for BSD68 [12] dataset.

Method	Noisy	BM3D	DnCNN	FFDNet	IRCNN	DHDN	SADE-Net
Noise Level	PSNR(dB)						
$\sigma = 10$	28.26	33.32	33.88	33.76	33.74	33.42	33.89
$\sigma = 30$	18.97	27.75	28.36	28.39	28.26	28.55	28.50
$\sigma = 50$	14.92	25.60	26.23	26.29	26.19	26.44	26.41
Noise Level	SSIM						
$\sigma = 10$	0.7094	0.9158	0.9270	0.9266	0.9262	0.9213	0.9300
$\sigma = 30$	0.3348	0.7731	0.7999	0.8032	0.7989	0.8110	0.8090
$\sigma = 50$	0.1984	0.6838	0.7189	0.7245	0.7171	0.7296	0.7308

Table 2. Quantitative denoising results for Kodak24 [5] dataset.

Method	Noisy	BM3D	DnCNN	FFDNet	IRCNN	DHDN	SADE-Net
Noise Level	PSNR(dB)						
$\sigma = 10$	28.22	34.39	34.90	34.81	34.76	34.43	35.01
$\sigma = 30$	18.87	29.12	29.62	29.69	29.52	29.93	29.91
$\sigma = 50$	14.78	26.98	27.49	27.62	27.45	27.88	27.84
Noise Level	SSIM						
$\sigma = 10$	0.6573	0.9127	0.9223	0.9226	0.9215	0.9153	0.9273
$\sigma = 30$	0.2729	0.7877	0.8071	0.8123	0.8056	0.8211	0.8207
$\sigma = 50$	0.1998	0.7140	0.7368	0.7437	0.7342	0.7528	0.7545

4 Experimental Results

4.1 Implementation Details

The DIV2K dataset [18] includes a large number of high quality images, and currently it is a commonly used dataset in image denoising applications [15, 1]. We employ the DIV2K validation and training datasets for training the introduced denoising network. The training set constitutes 800 images with 1920 × 1080 resolution. Validation set has 100 images with the same resolution. For the testing, we use BSD68 [12] and Kodak24 [5] datasets which are also highly popular for image denoising [20, 15] . Kodak24 has 24 images at 768 × 512, and BSD68 consists of 68 images at 321 × 481 resolution.

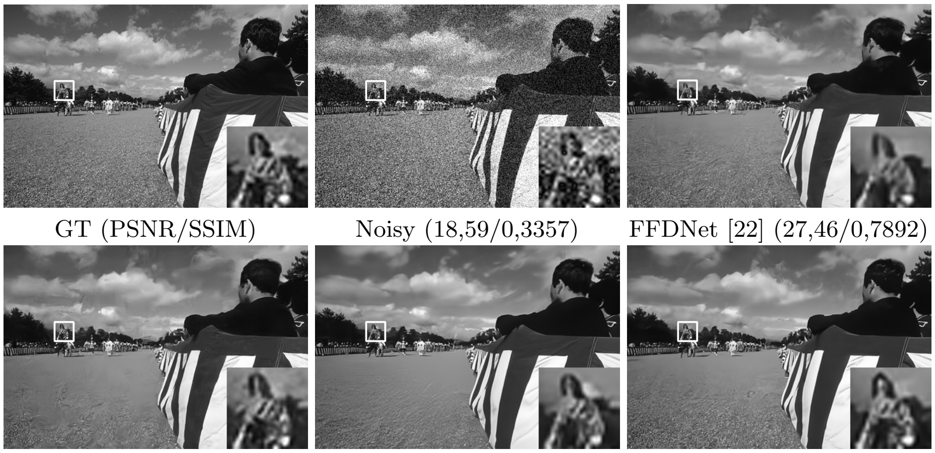
We firstly extract patches of size 64 × 64 from the training images, and we randomly flip these patches to augment the training data. One training batch has 16 randomly selected patches from the generated training data. We train the novel SADE-Net for grayscale image denoising with unknown noise level. We consider noise standard deviation levels which are between 5 and 55. For each training batch, we randomly sample the noise level σ from a uniform distribution defined on [5, 55]. Then, Gaussian noise realizations with the randomly chosen standard deviations are added to the images (patches) in the overall training batch.

For optimization, we utilize Adam optimizer [10] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. For the other hyperparameters of Adam, the default settings are used. For the competing methods, the hyperparameters were in general chosen as in their original papers, such as [15]. The initial learning rate is $1e^{-4}$, and it gets halved every three epochs. We use ℓ_1 mean absolute error as the loss function. All the training and testing procedures for the various deep networks experiments were conducted on two Nvidia RTX 2080 Ti GPUs. The training and testing of the deep networks are realized using CUDA version 10.2 in a PyTorch environment.

4.2 Performance Comparison

We compare the proposed network with several state of the art image denoising algorithms, including some recent and powerful image denoising networks. The competing methods are BM3D [4], DnCNN [21], FFDNet [22], IRCNN [23] and DHDN [15]. We used publicly available pretrained version of these networks for comparison purposes.

As quantitative performance metrics, we employed the peak-signal-to-noise-ratio (PSNR) and the structural similarity index (SSIM). Average PSNR and SSIM results for the image denoising experiments using the BSD68 test dataset are given in Table 1. The results for the the Kodak24 test dataset are listed in Table 2. In both tables, the highest result is marked with red, and the second best result is marked with blue. As can be inferred from Table 1 and Table 2, the proposed network performs better than the competing methods for a multitude of testing conditions. For all the simulated testing settings for both test datasets, the developed SADE-Net performed either best or second best in PSNR and SSIM among the realized approaches. To give an idea for the qualitative comparison of the various denoising results, we picked one sample test image from both BSD68 and Kodak24 datasets. We provide the denoised image results for our novel SADE-Net in addition to some high performance image denoising networks including DHDN [15], DnCNN [21] and FFDNet [22]. Figure 3 gives the denoised image results for a particular sample image from the BSD68 dataset, whereas Figure 4 includes the results for the sample image from the Kodak24 dataset. The denoised image results and the corresponding zoomed sections indicate that the introduced network is able to preserve details and texture while suppressing noise. The details in the zoomed sections showcase the improved detail preserving ability of the SADE-Net when compered with the competing deep networks.



GT (PSNR/SSIM) Noisy (18,59/0,3357) FFDNet [22] (27,46/0,7892)

DnCNN [21] (27,18/0,7703) DHDN [15] (27,33/0,7778) SADE-Net (27,57/0,7916)

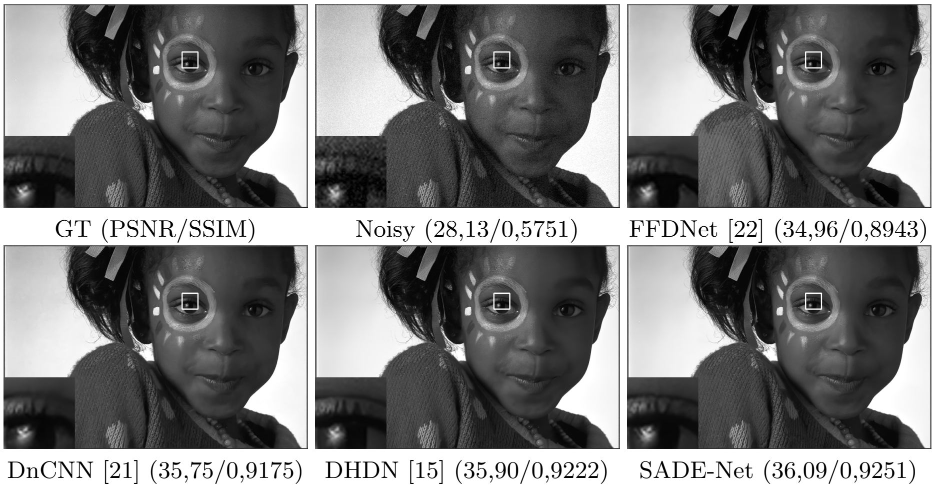
Fig. 3. Qualitative denoising results of our proposed network and other recent deep networks for the 'test019'image from BSD68 dataset, $\sigma = 10$.

5 Conclusion

In this paper, we propose a new image denoising deep network starting from a U-Net structure. In the novel network, we utilized a novel combination of channel attention blocks and the recently developed scale input idea. Another important feat is the use of densely connected DCR blocks to reduce the vanishing gradients problem and to facilitate robust transmission of information. The performance of the developed network is tested by using some of the most widely used test and training image dataset from the literature. The novel combination of scale inputs and channel attention blocks at the different resolution stages leads to improved denoising performance when compared to recent and effective deep methodologies for image denoising. The PSNR and SSIM results showcase the quantitative performance improvement. The proposed network is flexible and gives satisfactory denoising results for a wide range of noise levels. Denoised image samples on the other hand exhibit the qualitative performance enhancement in the preservation of details.

Acknowledgment

This work is supported by TUBITAK (The Scientific and Technological Research Council of Turkey) under project no. 119E248.



GT (PSNR/SSIM) Noisy (28,13/0,5751) FFDNet [22] (34,96/0,8943)

DnCNN [21] (35,75/0,9175) DHDN [15] (35,90/0,9222) SADE-Net (36,09/0,9251)

Fig. 4. Qualitative denoising results of our proposed network and other recent deep networks for the 'kodim15' image from Kodak24 dataset, $\sigma = 10$.

References

1. Anwar, S., Barnes, N.: Real Image Denoising With Feature Attention. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 3155–3164 (2019)
2. Colak, O., Eksioglu, E.M.: Image denoising using patch ordering and 3D transformation of patches. IET Image Processing **13**(13), 2636–2646 (2019)
3. Corbetta, M., Shulman, G.: Control of goal-directed and stimulus-driven attention in the brain. Nature reviews. Neuroscience **3**, 201–15 (04 2002). https://doi.org/10.1038/nrn755
4. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing **16**(8), 2080–2095 (2007). https://doi.org/10.1109/TIP.2007.901238
5. Franzén., R.: Kodak lossless true color image suite. source: http://r0k.us/graphics/kodak. vol. 4 (1999)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 770–778 (2016)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7132–7141 (2018)
8. Huang, G., Liu, Z., Weinberger, K.Q.: Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2261–2269 (2017)
9. Huang, Q., Yang, D., Wu, P., Qu, H., Yi, J., Metaxas, D.: MRI reconstruction via cascaded channel-wise attention network. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 1622–1626 (2019). https://doi.org/10.1109/ISBI.2019.8759423
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR **abs/1412.6980** (2015)

11. Liu, P., Zhang, H., Zhang, K., Lin, L., Zuo, W.: Multi-level Wavelet-CNN for Image Restoration pp. 886–88609 (2018)
12. 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: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 2, pp. 416–423 vol.2 (2001). https://doi.org/10.1109/ICCV.2001.937655
13. Mei, Y., Fan, Y., Zhang, Y., Yu, J., Zhou, Y., Liu, D., Fu, Y., Huang, T., Shi, H.: Pyramid attention networks for image restoration. ArXiv **abs/2004.13824** (2020)
14. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M.J., Heinrich, M., Misawa, K., Mori, K., McDonagh, S.G., Hammerla, N., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning where to look for the pancreas. ArXiv **abs/1804.03999** (2018)
15. Park, B., Yu, S., Jeong, J.: Densely Connected Hierarchical Network for Image Denoising. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 2104–2113 (2019)
16. Peng, Y., Cao, Y., Liu, S., Yang, J., Zuo, W.: Progressive training of multi-level wavelet residual networks for image denoising. ArXiv **abs/2010.12422** (2020)
17. Tai, Y., Yang, J., Liu, X., Xu, C.: MemNet: A Persistent Memory Network for Image Restoration (08 2017)
18. Timofte, R., Agustsson, E., Gool, L.V., Yang, M.H., Zhang, L., Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M., Wang, X., Tian, Y., Yu, K., Zhang, Y., Wu, S., Dong, C., Lin, L., Qiao, Y., Loy, C.C., Bae, W., Yoo, J., Han, Y., Ye, J.C., Choi, J.S., Kim, M., Fan, Y., Yu, J., Han, W., Liu, D., Yu, H., Wang, Z., Shi, H., Wang, X., Huang, T.S., Chen, Y., Zhang, K., Zuo, W., Tang, Z., Luo, L., Li, S., Fu, M., Cao, L., Heng, W., Bui, G., Le, T., Duan, Y., Tao, D., Wang, R., Lin, X., Pang, J., Xu, J., Zhao, Y., Xu, X., Pan, J., Sun, D., Zhang, Y., Song, X., Dai, Y., Qin, X., Huynh, X.P., Guo, T., Mousavi, H.S., Vu, T.H., Monga, V., Cruz, C., Egiazarian, K., Katkovnik, V., Mehta, R., Jain, A.K., Agarwalla, A., Praveen, C.V.S., Zhou, R., Wen, H., Zhu, C., Xia, Z., Wang, Z., Guo, Q.: NTIRE 2017 challenge on single image super-resolution: Methods and results. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 1110–1121 (2017). https://doi.org/10.1109/CVPRW.2017.149
19. Wang, X., Girshick, R.B., Gupta, A., He, K.: Non-local neural networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 7794–7803 (2018)
20. Yu, S., Park, B., Jeong, J.: Deep iterative down-up cnn for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2019)
21. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Transactions on Image Processing **26**, 3142–3155 (2017)
22. Zhang, K., Zuo, W., Zhang, L.: FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising. IEEE Transactions on Image Processing **27**(9), 4608–4622 (2018)
23. Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep CNN denoiser prior for image restoration. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2808–2817 (2017)