Towards Boosting Channel Attention for Real Image Denoising: Sub-band Pyramid Attention

Huayu Li[1], Haiyu Wu[2], Xiwen Chen[1], Hao Wang[1], and Abolfazl Razi[1(✉)]

[1] Northern Arizona University, Flagstaff, AZ 86011, USA
{hl459,xc53,hw328,Abolfazl.Razi}@nau.edu
[2] University of Notre Dame, Notre Dame, IN 46556, USA
hwu6@nd.edu

Abstract. Convolutional layers treat the Channel features equally with no prioritization. When Convolutional Neural Networks (CNNs) are used for image denoising in real-world applications with unknown noise distributions, particularly structured noise with learnable patterns, modeling informative features can substantially boost the denoising performance. Channel attentions in real-world image denoising tasks exploit dependencies between the feature channels; therefore, they can be viewed as a frequency-domain filtering mechanism. Existing channel attention modules typically use global statics as descriptors to learn inter-channel correlations. These methods deem inefficient in learning representative coefficients for re-scaling the channels at frequency level. This paper proposes a novel Sub-band Pyramid Attention (SPA) model based on wavelet transform to recalibrate the extracted features' frequency components in a more fine-grained fashion. Our method, in one sense, integrates the conventional frequency-domain filtering methods with deep learning architectures to achieve higher performance records. Experimental results show that ANNs equipped with the proposed attention module substantially improves upon the benchmark naive channel attention blocks. More specifically, we obtained a 3.97 dB gain compared to the best traditional algorithm, BM3D and a 1.87 dB to 0.18 dB gain over the DL-based methods in terms of denoising performance. Furthermore, our results show how the pyramid level affects the performance of the SPA blocks and exhibits favorable generalization capability for the SPA blocks.

Keywords: Deep learning · Attention model · Image denoising

This material is based upon work supported by the National Science Foundation under Grant 2008784.

H. Li and H. Wu—These authors contributed equally to this work.

Y. Peng et al. (Eds.): ICIG 2021, LNCS 12890, pp. 303–314, 2021.
https://doi.org/10.1007/978-3-030-87361-5_25

1 Introduction

Convolutional Neural Networks (CNN) have shown a remarkable performance in image denoising tasks compared to conventional filtering methods [4,22,32,33]. Mathematically, image denoising methods intend to recover a clean and high-quality target y from a corrupted low-quality observation x by eliminating noise, imaging artifacts, distortions, etc. In general, the overall impact of these undesired terms is modeled as a zero-mean Additive White Gaussian Noise (AWGN) with an arbitrary variance. Using the Gaussian noise model provides computational convenience. This model is backed by the central limit theorem that states that the normalized sum of independent and arbitrarily distributed random terms approach a normal distribution. Although appropriate for general and pure random noise modeling, the AWGN model is oversimplified for situations where the structured noise exhibits hidden patterns. For instance, image distortions due to cameras' loss of focus, lens scratch, dusty lens, camera shake, low illumination, and raindrops exhibit learnable patterns substantially different than a random noise [8].

Previous CNN-based methods [32–34] outperform the traditional methods [10,13] for image denoising due to the learning capacity of CNNs. However, they are mainly designed to deal with synthetic noise instead of real-world noise with more complicated and diverse compositions. Therefore, their performance might be suboptimal in real-world denoising tasks. Recently, new noise models [4,7,31,36] and new benchmark denoising methods [1,26] are proposed to tackle more realistic noise models. One popular way is using attention mechanisms to manage and quantify the interdependence between the extract feature maps. For instance, in RIDNet [4], a channel attention block [18] called feature attention was used for feature selection, which allows the network to focus on the feature channels of interest. Although feature attention methods obtain superior performance in real-world image denoising tasks, generating coarse feature descriptors based on the global pooling of the entire feature map is not optimal.

From the traditional image processing perspective, a natural image is composed of different frequency components, with high-frequency components representing the fine details and low-frequency components representing the global structure of the image; therefore, denoising is usually performed by smoothing and low-pass filtering. We re-think the DL-based denoising paradigm by leveraging traditional denoising concepts. Channel attention mechanisms can be regarded as an adaptive filter that suppresses the abundant frequency feature channels [35]. In most channel attention models, the feature maps' coarse statistics are generated by Global Average Pooling (GAP). Second-order channel attention was proposed in [14] for enriching the representation ability of the channel attention blocks. Nevertheless, these types of channel attention models are not flexible enough to deal with various frequency levels. In [5], Laplacian attention is proposed to integrate traditional methods and deep learning, adopting multiple convolutional layers with different receptive fields to model the frequency components of the input features. Despite the improvements obtained by this and similar methods, the following question remains open: Is there a

better way of obtaining a representation for the frequency characteristics of the input image?

In this paper, a CNN architecture with an efficient and plug-and-play Sub-band Pyramid Attention (SPA) is proposed as an alternative approach for the existing channel attention models (Fig. 1). Based on the wavelet decomposition, the SPA module performs a more fine-grained frequency selection that weighs the sub-bands at different levels. The SPA module exhibits a restoration performance while preserving the detailed textures with a negligible increase in computational complexity. Experimental results on real image denoising confirm the superiority of the proposed method. We also review how the utilized pyramid level affects the denoising performance and the generalization capability of the SPA blocks. The proposed SPA method with a proven superior performance can replace the existing attention mechanisms and is readily applicable to DL networks with arbitrary structures. Our study reminds the fact that leveraging fundamental knowledge in image processing can improve DL methods' performance.

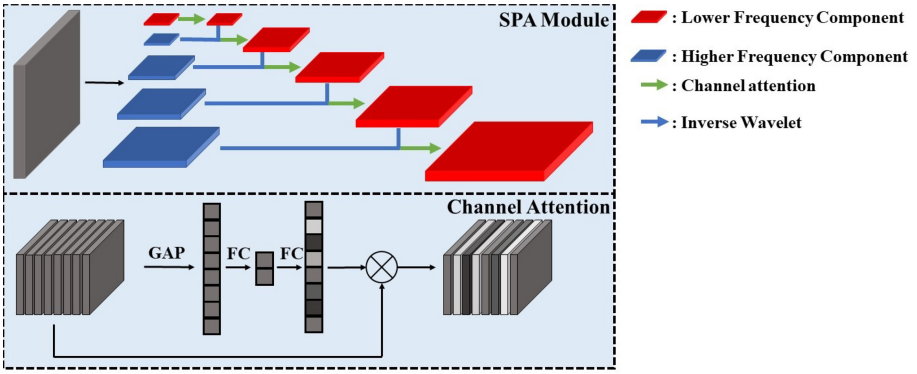

Fig. 1. The sub-band pyramid attention for frequency components selection.

2 Camera Noise vs Additive White Gaussian Noise

Most image denoising methods, such as [22,32,33] were trained based on image pairs $\{x_i, y_i\}_{i=1}^{N}$, where x_is are clean images easily available in large quantities, and $y_i = x_i + n$ are the artificial corrupted images by adding AWGN noise n with different standard deviations σ. While training a specific model for a certain noise level, the standard deviation σ for generating training and testing data is set to constant values *e.g.*, $\sigma = 15, 25, \ldots, 50$. For blind denoising, the models are trained under random noise levels and tested with unknown noise levels.

During camera imaging, a digital image goes through the sequential steps of photon-voltage conversion, analog amplification, and analog-to-digital conversion [19]. During these stages, multiple noise sources, including shot noise (photon noise, Gamma noise), read-out noise, and thermal noise, are involved [9].

We can roughly define the real noise model as a Poisson-Gaussian noise models [15,19,21] as $y = kP(\frac{x}{k}) + N(0, \sigma^2)$, where x and y are clean and noisy image pairs, k and σ are factors related to camera settings, and $P(\cdot)$ and $N(\cdot)$, respectively, denote the Poisson and Gaussian distributions. Unlike AWGN, real noise and image components are not independent that significantly limits the denoising performance. In addition to these noise terms, several other factors such as lens dirt and scratch, low light conditions, and natural air pollution can further distort the image.

3 Method

3.1 Frequency Sub-band Pyramid

The 2-D Discrete Wavelet Transform (DWT) is a powerful tool for analyzing image structures in spatial and frequency domains. At each layer, DWT decomposes the image (or the approximate coefficients of the previous layer) into four sub-bands using four orthogonal convolutional filters, including one low pass filter f_{LL}, and three high pass filters f_{LH}, f_{HL}, and f_{HH}. Haar wavelet, a popular mother wavelet which is also used in SPA, includes four orthogonal filters defined as: $f_{LL} = \left[\begin{smallmatrix} +1 & +1 \\ +1 & +1 \end{smallmatrix}\right]$, $f_{LH} = \left[\begin{smallmatrix} -1 & -1 \\ +1 & +1 \end{smallmatrix}\right]$, $f_{HL} = \left[\begin{smallmatrix} -1 & +1 \\ -1 & +1 \end{smallmatrix}\right]$, and $f_{HH} = \left[\begin{smallmatrix} +1 & -1 \\ -1 & +1 \end{smallmatrix}\right]$. Four sub-bands are generated by convolving the input image x with these filters to obtain: $x_{LL} = (f_{LL} \otimes x)$, $x_{LH} = (f_{LH} \otimes x)$, $x_{HL} = (f_{HL} \otimes x)$, and $x_{HH} = (f_{HH} \otimes x)$, where $\otimes$ is the convolution operator. Moreover, the bi-orthogonal property of DWT enables an easy and lossless reconstruction of the original image using the inverse transformation of Haar wavelet $IWT(x_{LL}, x_{LH}, x_{HL}, x_{HH})$.

In this work, we propose to build a Frequency Sub-band Pyramid using wavelet decomposition. A Frequency Sub-band Pyramid consists of multi-level frequency components of an image or its feature maps. Given an input x_0, DWT decomposes it into a set of detail coefficients $X_{1H} = [x_{1HH}, x_{1HL}, x_{1LH}]$ and approximate coefficients x_{1LL}. After further decomposition using multi-level DWT for n iterations, one low frequency component x_{nL} and n sets of high frequency components $X_{1H}, \ldots, X_{nH}$ are obtained. Stacking these components from the first to the last level forms a sub-band pyramid that represents the low-to-high frequency properties of the features to be modeled.

3.2 Sub-band Pyramid Attention

A channel attention [18] module is typically formulated as:

$$x^{'} = x * \sigma(f_2(ReLU(f_1(G_x)))) \tag{1}$$

for a 3D input $x \in C \times H \times W$, and a global descriptor $G_x \in C \times 1 \times 1$, which represents the statistics of each input map generated by the GAP. The functions f_1 and f_2 refer to two fully-connected layers activated by the Rectified Linear

Units ($ReLU$) [24] and sigmoid function (σ). The channel attention module captures the channel dependencies from the global descriptor of the entire input, which is too coarse and may lead to information loss. The proposed SPA module exploits a more fine-grained channel-wise correlation with a new strategy of selective amplification of different spectral layers. Overall, the decomposition results $[x_n L, x_n H, \ldots, x_2 H, x_1 H]$ of an input x obtained by the Frequency Sub-band Pyramid are re-calibrated by the channel attention from lower to higher frequency levels as shown in Fig. 1. After being processed by the channel attention module, each lower frequency component is concatenated with its corresponding higher frequency component. The Inverse Wavelet Transform (IWT) is used to build the lower-frequency components layer by layer, starting from the top layer to the base layer until the entire feature map with the original size (i.e., the size of the input image) is reconstructed. The SPA module performs a more precise frequency selection mechanism than the naive channel attention approach by this operation. The SPA explicitly calibrates the dependencies between the feature channels while selecting the desired frequency component inside each feature map.

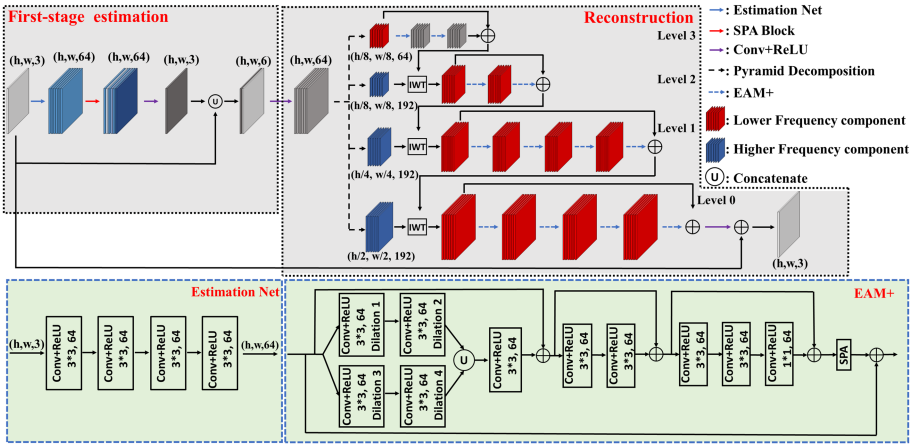

Fig. 2. The illustration of the entire network architecture: the network is divided into noise estimation and reconstruction stages. The noise estimation model is built by a plain CNN with SPA blocks. The reconstruction network is based on the sub-band pyramid, using a modified version of EAM [4], where we replaced the naive channel attention with SPA modules, and called it EAM+.

3.3 Network Architecture Overview

The network used in this work comprises two stages where the first stage performs the noise estimation, and the second stage performs the reconstruction, as shown in Fig. 2. For a noisy input $x \in C, H, W$, the first-stage F_e can be

regarded as a pixel-wise noise level estimation $x^{'} = F_e(x)$, where $x^{'} \in C, H, W$ is the estimation of the noisy map of the input channels. The first-stage estimation includes four 64-channel convolutional layers followed by a ReLU activation, a SPA block, and a 3-channel convolutional layer. The filter size is 3×3 for each convolutional layer in the first-stage estimation. The SPA block is considered a frequency estimator to selectively suppress the redundant information from the extracted features. The estimation results are stacked with the input along the channels as $[x, x^{'}]$ and are fed to the second stage.

Similar to the SPA block, the reconstruction stage is also designed based on the wavelet pyramid. This network consists of two convolutional layers (the first and last layers) and four sub-networks. The first convolutional layer extracts shallow features from the input image and the estimated noisy map. A 3-levels wavelet pyramid of the feature is constructed. The level-3 to level-1 sub-networks process the low-pass sub-bands $x_{3LL}, x_{2LL}, x_{1LL}$, and the level-0 sub-networks processes the basis features x_{0LL}. The sub-networks are built based on the Enhancement Attention Modules (EAM) proposed in RIDNet [4]. We replace the channel attention blocks in EAM with the proposed SPA blocks and name the modules EAM+. Each of the level-3 and level-2 sub-networks consists of two EAM+, and each of the level-1 and level-0 sub-networks consists of four EAM+. The sub-networks operate in a top-down manner, where each sub-network receives the lower-frequency map x_{iLL} of the wavelet's current layer as its input and extends it to the entire map of this current layer, which is equivalent to the low-frequency map of the wavelet's previous layer. This information is passed to the next subnetwork until the full-size map is recovered.

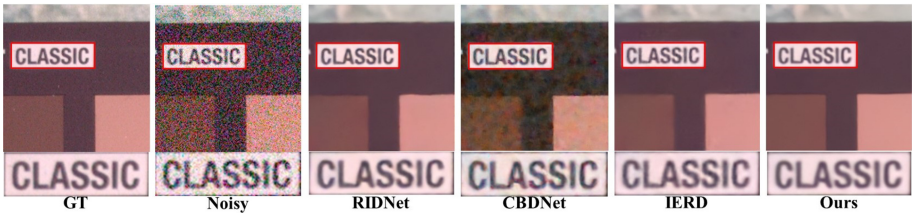

Fig. 3. A challenging example from SIDD dataset [1]. Our model exhibits a better color and edge preservation property.

4 Experiments

4.1 Experimental Setup

We use the Smartphone Image Denoising Dataset (SIDD) [1] and Darmstadt Noise Dataset (DnD) [26] for image denoising. The SIDD dataset provides 320 clean and noisy image pairs for training along with 1280 image pairs for validation. DnD dataset contains 50 pairs of real-world noisy and noise-free scenes. It

provides bounding boxes of size 512×512 of 1000 Regions of Interests (ROIs) for 50 scenes for generating testing data. Our model was implemented by the Pytorch Framework [25] and trained with a Tesla P40 GPU. We equipped the network with level-3 SPA blocks. The hyper-parameters of the network are defined in Fig. 2. We used 512×512 patches cropped from the SIDD [1] training set to train our model and used the DnD [26] and validation set of the SIDD dataset to evaluate the reconstruction performance. Data augmentation by applying random rotations at 90, 180, and 270° and horizontal flipping was used in the training phase. Peak Signal-to-Noise Ratio (PSNR) is used as the evaluation metric, while Mean Absolute Error (MAE) is used as the loss function. The model is trained by the Adam optimizer [20] with an initial learning rate of $1e{-}4$. We trained the model at $2.5e5$ iterations and halved the learning rate for each of the $1e5$ iterations.

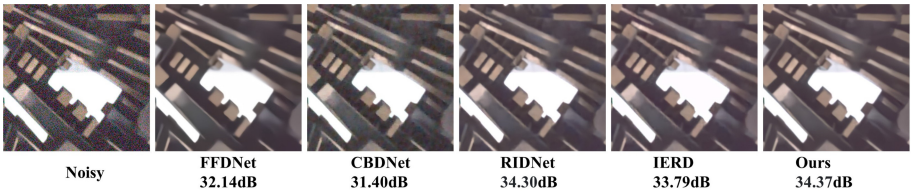

Fig. 4. Comparison of our method against some popular image denoising methods on DnD dataset.

4.2 Real Image Denoising

Table 2 represents the performances of the proposed method in terms of image denoising and reconstruction in comparison with several benchmark methods applied to the SIDD validation set. It can be seen that our method obtains remarkable results and outperforms the most commonly used DL-based denoising algorithms. Figure 3 presents an illustrative example, which shows that our method holds a competitive color and edge preservation property. Also, the results for the DnD dataset are summarized in Table 1. We compare the PSNR and SSIM [28] through the online evaluation system provided by the DnD [26] official website. The results show that the proposed method achieves a 3.97 dB gain compared to the best traditional algorithm, BM3D [13], while realizing a 1.87 to 0.18 dB gain over the DL-based methods [4,6,17,34]. Figure 4 shows the proposed method in restoring a noise-free image without over smoothing the details.

Table 1. PSNR and SSIM of the denoising methods evaluated on DnD [26] dataset.

Method	Blind/non-blind	PSNR	SSIM
CDnCNN-B [32]	Blind	32.43	0.7900
TNRD [12]	Non-blind	33.65	0.8306
LP [11]	Non-blind	34.23	0.8331
FFDNet [34]	Non-blind	34.40	0.8474
BM3D [13]	Non-blind	34.51	0.8507
WNNM [16]	Non-blind	34.67	0.8646
KSVD [3]	Non-blind	36.49	0.8978
MCWNNM [30]	Non-blind	37.38	0.9294
FFDNet+ [34]	Non-blind	37.61	0.9415
TWSC [29]	Non-blind	37.96	0.9416
CBDNet [17]	Blind	38.06	0.9421
RIDNet [4]	Blind	39.25	0.9528
IERD [6]	Blind	39.30	0.9531
Ours	Blind	**39.48**	**0.9580**

Table 2. The quantitative results (PSNR) for the SIDD dataset [1].

Method					
BM3D	FFDNet	CBDNet	RIDNet	IERD	Ours
30.88	29.20	30.78	38.71	38.82	**39.55**

4.3 Ablation Study

The major contribution of this work is proposing a novel implementation of the SPA attention module based on weighting wavelet layers. Although previous SPA-based attention mechanisms have already experienced remarkable performance, our approach offers a new perspective to this problem and answers new questions of "how does the pyramid level influence the denoising performance of the SPA blocks" and "if the SPA blocks can boost the performance of other networks with arbitrary architectures". Our results prove the superiority of the proposed SPA block compared to the naive channel attention block. At the same time, it also justifies the effect of different levels of SPA blocks. We also conducted an ablation study on the proposed network architecture. We test PSNR on the SIDD validation set for levels-0 to level-4. Notably, we define the naive channel attention as a level-0 pyramid (which means no pyramid). The second row of Table 3 shows that the higher pyramid level of the SPA blocks leads to a high PSNR and a better denoising performance. For evaluating the generalization of the SPA blocks, we conducted another ablation study by deploying SPA blocks at different levels in the RIDNet network as an alternative for its channel attention blocks. The third row of Table 3 shows the results of RIDNet when equipped with SPA blocks of different pyramid levels. The results show that the

SPA blocks generalize well in RIDNet while expecting to confirm that pyramids with more layers lead to better results.

Table 3. Investigation of effects of the pyramid level of SPA block and generalization on RIDNet.

Networks	Level	0	1	2	3	4
Ours	PSNR	39.24	39.33	39.47	39.55	39.57
RIDNet	PSNR	38.71	38.87	38.99	39.04	39.05

The proposed SPA attention is also generalized on the most commonly used AWGN model instead of only specific for real-world noise models. To investigate the AWGN noise, we use images with synthetic noise. We trained the blind denoising RIDNet and our network equipped with levels-0 to level-4 SPA attention on BSD500 [23] dataset with random AWGN (where the deviation σ varies between 5 to 55). The trained models were evaluated on BSD68 [27] dataset with noise level 15, 25, and 50. Table 4 compared the performance of each model. It consistently proves the generalizability.

Table 4. Comparisons on AWGN image denoising.

Noise levels	Method (RIDNet/Ours)			
	Level 1	Level 2	Level 3	Level 4
15	34.01/34.12	34.11/34.25	34.22/34.37	34.22/34.38
25	31.37/31.44	31.49/31.55	31.55/31.65	31.56/31.66
50	28.14/28.23	28.22/28.30	28.31/28.42	28.32/28.42

To further evaluate the performance of the proposed method, denoising is performed for Poisson-Gaussian noise model, which is defined as $y = P(x) + N(0, \sigma^2)$. This model includes Poisson-distribution $P(x)$ related to the clean image x, and followed by a Gaussian-distribution $N(0, \sigma^2)$. Here, we set the standard deviation σ to 15. We trained our network equipped with levels-0 to level-4 SPA attention on DIV2K [2] training set, and test them on DIV2K validation set. Table 5 presents the results of each level. The results show that the proposed SPA works better than naive channel attention (Level 0), and the performance gain still followed the obey the rules we observed in the previous experiments, that higher level pyramid brings more improvement.

Table 5. Comparisons on Poisson-Gaussian noise denoising.

Pyramid level	w/o CA	0	1	2	3	4
PSNR	31.11	31.52	32.21	32.57	33.04	33.06

Table 6. Comparisons on GPU runtime (seconds) and Model size (MB).

Level		w/o CA	0	1	2	3	4
Model size/GPU runtime	RIDNet	5.71/0.16	5.72/0.19	5.73/0.22	5.74/0.23	5.74/0.25	5.75/0.27
	Ours	17.11/0.22	17.13/0.24	17.16/0.29	17.18/0.32	17.21/0.35	17.23/0.41

4.4 Computational Overhead

As mentioned before, the SPA blocks improves the denoising performance while not adding considerable computational overhead. In this section, we compared the model sizes[1] when deployed different-level SPA blocks on both RIDNet and our network. It can be seen in Table 6 that the SPA blocks would not significantly increase the model size. Moreover, comparisons on average GPU runtime were conducted to show the proposed SPA module achieve a good trade-off between the calculation and performance. We fed images with a size of 512×512 for runtime evaluation. We tested each model 100 times and calculated the average runtime.

5 Conclusion

A novel channel attention module called Sub-band Pyramid Attention (SPA) is proposed in this work. The SPA blocks are built upon wavelet decomposition to realize a joint sub-band channel attention. The SPA is implemented as a plug-and-play module, hence can replace the naive channel attention in arbitrary deep learning networks. The proposed SPA block performs a more precise feature re-calibration that re-scales both the feature channels and the multi-level frequency components. The achieved gain for the proposed method over the conventional filtering methods and DL methods is considerable, ranging from 0.18 dB to 3.97 dB on the benchmark dataset of DnD. The generalization of the SPA blocks is verified by the ablation study, which suggests that the SPA module is compatible with other network architectures and can be widely used in other networks to boost image restoration performance.

References

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1692–1700 (2018)

[1] Model size was calculated by torchsummary package (https://github.com/sksq96/pytorch-summary).

2. Agustsson, E., Timofte, R.: NTIRE 2017 challenge on single image super-resolution: dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 126–135 (2017)
3. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Sig. Process. **54**(11), 4311–4322 (2006)
4. Anwar, S., Barnes, N.: Real image denoising with feature attention. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3155–3164 (2019)
5. Anwar, S., Barnes, N.: Densely residual Laplacian super-resolution. IEEE Trans. Pattern Anal. Mach. Intell. (2020)
6. Anwar, S., Phuoc Huynh, C., Porikli, F.: Identity enhanced residual image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 520–521 (2020)
7. Bao, L., Yang, Z., Wang, S., Bai, D., Lee, J.: Real image denoising based on multi-scale residual dense block and cascaded U-Net with block-connection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 448–449 (2020)
8. Boie, R.A., Cox, I.J.: An analysis of camera noise. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 671–674 (1992)
9. Boyat, A.K., Joshi, B.K.: A review paper: noise models in digital image processing. arXiv preprint arXiv:1505.03489 (2015)
10. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 2, pp. 60–65. IEEE (2005)
11. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: can plain neural networks compete with BM3D? In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2392–2399. IEEE (2012)
12. Chen, Y., Pock, T.: Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1256–1272 (2016)
13. 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)
14. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11065–11074 (2019)
15. Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. IEEE Trans. Image Process. **17**(10), 1737–1754 (2008)
16. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2862–2869 (2014)
17. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1712–1722 (2019)
18. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
19. Jain, U.: Characterization of CMOS Image Sensor. Ph.D. thesis, MS Thesis, Delft University of Technology (2016)

20. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Liu, X., Tanaka, M., Okutomi, M.: Practical signal-dependent noise parameter estimation from a single noisy image. IEEE Trans. Image Process. **23**(10), 4361–4371 (2014)
22. Mao, X., Shen, C., Yang, Y.B.: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Advances in Neural Information Processing Systems, pp. 2802–2810 (2016)
23. 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 8th IEEE International Conference on Computer Vision, ICCV 2001, vol. 2, pp. 416–423. IEEE (2001)
24. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: ICML (2010)
25. Paszke, A., et al.: Automatic differentiation in PyTorch (2017)
26. Plotz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1586–1595 (2017)
27. Roth, S., Black, M.J.: Fields of experts. Int. J. Comput. Vis. **82**(2), 205 (2009)
28. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
29. Xu, J., Zhang, L., Zhang, D.: A trilateral weighted sparse coding scheme for real-world image denoising. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11212, pp. 21–38. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01237-3_2
30. Xu, J., Zhang, L., Zhang, D., Feng, X.: Multi-channel weighted nuclear norm minimization for real color image denoising. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1096–1104 (2017)
31. Zamir, S.W., et al.: CycleISP: real image restoration via improved data synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2696–2705 (2020)
32. 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. **26**(7), 3142–3155 (2017)
33. Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep CNN denoiser prior for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3929–3938 (2017)
34. Zhang, K., Zuo, W., Zhang, L.: FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans. Image Process. **27**(9), 4608–4622 (2018)
35. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 294–310. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_18
36. Zhao, Y., Jiang, Z., Men, A., Ju, G.: Pyramid real image denoising network. In: 2019 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4. IEEE (2019)