SELECTIVE RESIDUAL M-NET FOR REAL IMAGE DENOISING

Chi-Mao Fan⋆ Tsung-Jung Liu⋆ Kuan-Hsien Liu†

⋆Department of Electrical Engineering and Graduate Institute of Communication Engineering, National Chung Hsing University, Taiwan
†Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taiwan

ABSTRACT
Image restoration is a low-level vision task which is to restore degraded images to noise-free images. With the success of deep neural networks, the convolutional neural networks surpass the traditional restoration methods and become the mainstream in the computer vision area. To advance the performance of denoising algorithms, we propose a blind real image denoising network (SRMNet) by employing a hierarchical architecture improved from U-Net. Specifically, we use a selective kernel with residual block on the hierarchical structure called M-Net to enrich the multi-scale semantic information. Furthermore, our SRMNet has competitive performance results on two synthetic and two real-world noisy datasets in terms of quantitative metrics and visual quality. The source code and pretrained model are available at https://github.com/TentativeGitHub/SRMNet.

Index Terms— Image denoising, selective kernel, residual block, hierarchical architecture, M-Net

1. INTRODUCTION
Image denoising is a challenging ill-posed problem which also plays an important role in the pre-process of high-level vision task. In general, a corrupted image $Y$ could be represented as:

$$ Y = D(X) + n, $$

where $X$ is a clean image, $D(\cdot)$ denotes the degradation function and $n$ means the additive noise. Traditional model-based denoising methods, such as block-matching and 3D filtering (BM3D) \cite{1}, non-local means (NLM) \cite{2} are all based on the information of image priors. Although the conventional prior-based methods could handle most of denoising tasks and achieve acceptable performances, the key problems like computationally expensive and time-consuming hamper the efficiency of model-based methods.

In recent years, learning-based methods \cite{3,4,5} surpass traditional prior-based methods in terms of inference time and denoising performance. The performance gain of learning-based methods especially CNN over the others is mainly attributed to their elaborately designed model or block. For example, residual learning \cite{6,3}, dense connection \cite{7,8}, residual dense block \cite{9,19}, attention mechanisms \cite{11,12,13}, channel attention block \cite{11,14}, and hierarchical architecture \cite{15,16,17}. However, these complex architectures cause the restoration models to waste more computation and the improvement is only a little.

In this paper, we try to balance between the accuracy and computational efficiency of the model. First, we propose the hierarchical selective residual architecture which is based on the residual dense block with a more efficiency structure named selective residual block (SRB). Moreover, we use the multi-scale feature fusion with two different sampling methods (pixel shuffle \cite{18}, bilinear) based on the proposed M-Net to extract adequate and useful spatial feature information. For the reconstruction process, instead of using concatenation to fuse the feature maps with different resolutions, we adopt the selective kernel feature fusion (SKFF) \cite{19,20} to efficiently combine the features. Overall, the main contributions of this paper can be summarized as follows:

- We propose the novel hierarchical architecture (M-Net) to denoise for both synthesized additive white Gaussian noise (AWGN) and real-world noise.
- We propose the efficient feature extraction block called selective residual block which is improved from the residual dense block for image super-resolution.
- We experiment on two synthetic image datasets and two real-world noisy datasets to demonstrate that our proposed model achieves the state-of-the-art in image denoising quantitatively and qualitatively even with less computational complexity.

2. RELATED WORK

In this section, we first are going to discuss the development of denoising. Then, we will describe the hierarchical architecture and the selective kernel which we apply in our proposed selective residual block.

2.1. Image denoising
As aforementioned, traditional image denoising approaches are generally based on image priors or algorithms which
are also called model-based methods, such as self-similarity [1, 2, 21], sparse coding [22, 23] and dictionary learning [22, 24]. Currently, CNN-based denoisers have demonstrated state-of-the-art results [6, 3, 15, 16, 17]. Moreover, the denoise models from [3, 13, 16, 17, 4, 5, 25] only deal with signal-independent noise (e.g., AWGN, read noise), while the model from [14, 26, 27, 28, 29] have the ability to process the real-world signal-dependent noise (e.g., shot noise, thermal noise).

2.2. Hierarchical architecture

The hierarchical architecture is used to enrich the spatial information by extracting different sizes of features and make semantic information have more diversity. The most well-known hierarchical structure is the U-Net [15] which has the great contribution from high-level to low-level vision tasks including image segmentation [30, 31], image restoration [16, 17, 32], etc.

2.3. Selective Kernel Network

Li et al. proposed the selective kernel convolution that has two branches. One of the path utilizes normal naive $3 \times 3$ convolution kernel to extract features, and the other path adopts different kernel size (e.g., $5 \times 5$, $7 \times 7$) to obtain the larger receptive field. At the end of selective kernel convolution, they use softmax activation function to acquire the weights for two different features maps. Zamir et al. [20] were inspired by [19] and applied it to multi-scale feature fusion for image enhancement tasks, which also achieve good results.

3. PROPOSED METHOD

In this section, we mainly introduce the proposed Selective Residual M-Net (SRMNet), and provide detailed explanations for each component of the model in the following subsections.

3.1. M-Net

The M-Net architecture is first proposed for medical image segmentation [33]. Adiga et al. [34] use the same framework for fingerprint image denoising and also get good results. Compared with above two models, our proposed SRMNet has two improvements. 1) More diversity and plentiful multi-scale cascading features. The original M-Net used $2 \times 2$ max-pooling in both U-Net path and gatepost path, and then combined these two features together. Our SRMNet use pixel un-shuffle down-sampling in U-Net path and bilinear down-sampling for gatepost path, which makes the cascading features have more diversity. 2) Using different feature fusion methods to summarize the information in the decoder (reconstruction process). Actually, original M-Net has high-dimensional cascading features, especially the shallow layer in the model, which makes the M-Net have large number of parameters and high computational complexity. Therefore, they use some techniques such as batch normalization, reducing the size of input images and the dimension of input feature maps. In other words, the original M-Net is inappropriate to be directly applied to image denoising. To solve this problem, we use the selective kernel feature fusion (SKFF) method [20] which does not concatenate each feature map but aggregates the weighting features.
3.2. SRMNet

The proposed SRMNet for image denoising is shown in Fig. 1. We first use $3 \times 3$ weight sharing convolution in each resolution of corrupted input image acquired by doing the bilinear down-sampling from original-resolution input. Each layer has the proposed selective residual block (SRB) to extract high-level semantic information (more structure details will be illustrated in the next subsection). Then, we choose pixel unshuffled method as our down-sampling module at the end of SRB to obtain multi-scale feature maps. After that, the feature maps are concatenated with previous shallow features from bilinear down-sampling and keep going as normal U-Net process. The main purpose of choosing two different down-sample methods (pixel unshuffled and bilinear) is to make the cascaded features have more semantic information. Finally, using the SKFF (upper part of Fig. 2) to integrate features with different scales to reconstruct the denoised image.

We optimize our SRMNet end-to-end with the Charbonnier loss [35] for image denoising as follows:

$$L_{char} = \sqrt{||\hat{X} - X||^2 + \varepsilon^2},$$

(2)

where $\hat{X}, X \in \mathbb{R}^{B \times C \times H \times W}$ means the denoised and ground-truth images, respectively. $B$ is the batch size of training data, $C$ is the number of feature channels, $H$ and $W$ are the size of images. The constant $\varepsilon$ in Eq. (2) are empirically set to $10^{-3}$.

3.3. Selective Residual Block

Fig. 2 shows the architecture of the proposed SRB which is improved from the residual dense block (RDB) [9]. In the framework of SRB, each residual block has two input features ($f_r, f_m \in \mathbb{R}^{C \times H \times W}$ in Fig. 2) which denote the residual feature and mainstream feature, respectively. These two features will do the SKFF by multiplying the corresponding feature descriptor vectors ($v_1, v_2 \in \mathbb{R}^{C \times 1 \times 1}$ which are generated from channel-wise statistics $s \in \mathbb{R}^{C \times 1 \times 1}$) to get the weighted features ($\hat{f}_r, \hat{f}_m \in \mathbb{R}^{C \times 1 \times 1}$). Finally, we aggregate two channel-weighted features $\hat{f}_r, \hat{f}_m$ together to acquire the output feature $f_o$ of single residual block. After a few residual blocks (e.g., 3), we use a $3 \times 3$ convolution and add the long skip connection with $1 \times 1$ convolution between the input and output.

3.4. Resizing Module

As for resizing module, we simply use bilinear down-sampling for input images $Y \in \mathbb{R}^{3 \times H \times W}$ ($Y$ is the same as Eq. (1)), and use pixel unshuffled module shown in Fig. 3 for shallow features $\in \mathbb{R}^{C \times H \times W}$ after $3 \times 3$ convolution. Note that the size of the feature channel ($C$) before entering the resizing module is the same as output channel size but with smaller resolution (e.g. $W/2 \times H/2$).

It should be noticed that the the feature map of lower layer contains both bilinear and pixel unshuffled features but the in-
gredients of bilinear features are obvious less than pixel un-
shuffled features. It may cause the unbalanced problem. In
fact, the above unbalanced feature map will be passed through
the SRB which could solve the unbalanced problem by in-
creasing the weight of bilinear features.

4. EXPERIMENTS

4.1. Experiment Setup

Implementation Details. Our SRMNet is an end-to-end
model and trained from scratch. The experiments conducted
in this paper are implemented by PyTorch 1.8.0 with single
NVIDIA GTX 1080Ti GPU.

Evaluation Metrics. For the quantitative comparisons, we
consider the Peak Signal-to-Noise Ratio (PSNR) and Struc-
ture Similarity (SSIM) Index metrics.

4.2. Experiment Datasets

Gaussian Color Image Denoising. For Gaussian denois-
ing, we use the same experimental setup as image denois-
ing [16, 17] and train our model on image super-resolution
DIV2K [38] dataset which has 800 and 100 high-quality (the
average resolution is about 1920 × 1080) images for training
and validation, respectively. We randomly crop 100 patches
with size 256 × 256 for each training image and randomly add
AWGN to the patches with noise level from σ = 5 to 50. Eval-
uation is conducted on noise levels 10, 30, 50 on CBSD68
[50] and Kodak24 [37]. It should be noted that our model
does not know the noise level in the testing, which means
SRMNet is the blind denoising model.

Real-World Image Denoising. To train our SRMNet for
real-world denoising, we follow [20] to use 320 high-
resolution images of SIDD dataset [39]. Evaluations also fol-
low aforementioned methods, which is to perform the test on
1280 validation patches from the SIDD dataset [39] and 1000
patches from the DND benchmark dataset [40]. The resolu-
tion of all patches is 256 × 256 in both training and testing.

4.3. Image Denoising Performance

Gaussian Color Image Denoising. In Table 1 and Fig. 4
we compare our SRMNet with the prior-based method (e.g.,
BM3D [11]), CNN-based methods (e.g., DnCNN [3], IrCNN
[5], FFDNet [4]) and the models which are based on RDB
(e.g., DHDN [16], DIDN [25], RDN [9], RDUNet [17]). Ac-

Table 1 which we could observe three things: 1) The pro-
posed SRMNet achieves state-of-the-art quantitative scores,
especially for the difficult Gaussian noise levels (e.g., 30, 50).
2) Compared to RDB-based methods (DHDN, DIDN, RDN,
RDUNet), our SRMNet has the least FLOPs (↓20%) among
the five models, and still keeps the best scores because of the
efficient SRB design. 3) The SSIM scores of the SRMNet
are the best in both CBSD68 and Kodak24 datasets for each
noise level, which means that our denoised images are more
perceptually faithful. We think it is attributed to the M-Net
design which could gain more spatial details in the training
process.

Real-World Image Denoising. For real image denoising, we
evaluate the performance of 13 image denoising approaches
on real-world noise datasets (SIDD and DND) in Table 2.
Compared to the previous state-of-the-art CNN-based method
[20] in SIDD dataset, our model gains the same scores with
MIRNet but less computational complexity (e.g., FLOPs) and
time cost. More specifically, Table 3 shows our SRMNet
only use 36.3% of MIRNet’s FLOPs and about three times
faster than MIRNet. Fig. 4 also displays the visual results for
real image denoising on SIDD. Our SRMNet effectively re-
moves noise and the denoised images are visually closer to the
ground-truth.

5. CONCLUSION

In this paper, we present the SRMNet architecture and achieve
state-of-the-art performances on image denoising. The M-Net
design has the advantage of enriching features with different
resolutions by concatenating the results after pixel unshuf-

In this paper, we present the SRMNet architecture and achieve
state-of-the-art performances on image denoising. The M-Net
design has the advantage of enriching features with different
resolutions by concatenating the results after pixel unshuf-

bile and bilinear down-sampling. Moreover, we proposed the
SRB, which is an efficient block compared with the RDB. Our
future works are going to focus on different restoration tasks
such as image deblurring and image deraining.

6. REFERENCES

[1] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and
Karen Egiazarian, “Image denoising with block-matching and
3d filtering,” in Image Processing: Algorithms and Systems,
Neural Networks, and Machine Learning, International Society
for Optics and Photonics, 2006, vol. 6064, p. 606414.
[2] Pierrick Coupé, Pierre Yger, Sylvain Prima, Pierre Hellier,
Charles Kervrann, and Christian Barillot, “An optimized block-
wise nonlocal means denoising filter for 3-d magnetic reso-
nance images,” IEEE transactions on medical imaging, vol.
27, no. 4, pp. 425–441, 2008.
[3] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei
Zhang, “Beyond a gaussian denoiser: Residual learning of deep
cnn for image denoising,” IEEE transactions on image process-
ing, vol. 26, no. 7, pp. 3142–3155, 2017.
[4] Kai Zhang, Wangmeng Zuo, and Lei 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.
[5] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang,
“Learning deep cnn denoiser prior for image restoration,” in
Proceedings of the IEEE conference on computer vision and
pattern recognition, 2017, pp. 3929–3938.
[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun,
“Deep residual learning for image recognition,” in Proceed-
ings of the IEEE conference on computer vision and pattern
recognition, 2016, pp. 770–778.
[7] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kil-
lian Q Weinberger, “Densely connected convolutional net-
works,” in Proceedings of the IEEE conference on computer
vision and pattern recognition, 2017, pp. 4700–4708.
[8] Gao Huang, Shichen Liu, Laurens Van Der Maaten, and Kil-
lian Q Weinberger, “Condensenet: An efficient densenet using

Table 1: Gaussian color image denoising. Image denoising results on CBSD68 dataset [36] and Kodak24 dataset [37]. The best and second best scores are highlighted and underlined, respectively. All of scores are the average values of the whole dataset. The last column shows floating-point operations per second (FLOPs) which is conducted on $256 \times 256$ color images.

| Methods         | CBSD68 $\sigma=10$ | CBSD68 $\sigma=30$ | CBSD68 $\sigma=50$ | Kodak24 $\sigma=10$ | Kodak24 $\sigma=30$ | Kodak24 $\sigma=50$ | FLOPs  |
|-----------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|--------|
|                 | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | FLOPs  |
| BM3D [1]        | 35.89 | 0.951 | 29.71 | 0.843 | 27.36 | 0.763 | 33.32 | 0.956 | 27.75 | 0.773 | 25.60 | 0.686 | -    |      |      |      |        |
| IrCNN [5]       | 36.06 | 0.953 | 30.22 | 0.861 | 27.86 | 0.789 | 36.70 | 0.945 | 31.24 | 0.858 | 28.92 | 0.794 | 27G   |      |      |      |        |
| FFDNet [1]      | 36.14 | 0.954 | 30.31 | 0.860 | 27.96 | 0.788 | 36.80 | 0.946 | 31.39 | 0.860 | 29.10 | 0.795 | 18G   |      |      |      |        |
| DaCNN [3]       | 36.12 | 0.951 | 30.32 | 0.861 | 27.92 | 0.788 | 36.58 | 0.945 | 31.28 | 0.858 | 28.94 | 0.792 | 36G   |      |      |      |        |
| DHDN [16]       | 36.05 | 0.953 | 30.12 | 0.858 | 27.71 | 0.787 | 37.30 | 0.951 | 31.98 | 0.874 | 29.72 | 0.817 | 1019G |      |      |      |        |
| RNAN [13]       | 36.43 | -     | 30.63 | -     | 26.83 | -     | 37.24 | -     | 31.86 | -     | 29.58 | -     | -     |      |      |      |        |
| DIDN [25]       | 36.48 | 0.957 | 30.71 | 0.870 | 28.35 | 0.804 | 37.32 | 0.950 | 31.97 | 0.872 | 29.72 | 0.816 | 1121G |      |      |      |        |
| RDN [9]         | 36.47 | -     | 30.67 | -     | 28.31 | -     | 37.31 | -     | 31.94 | -     | 29.66 | -     | 1490G |      |      |      |        |
| RDUNet [17]     | 36.48 | 0.951 | 30.72 | 0.872 | 28.38 | 0.807 | 37.29 | 0.951 | 31.97 | 0.874 | 29.72 | 0.818 | 807G  |      |      |      |        |
| SRMNet (Ours)   | 36.46 | 0.961 | 30.72 | 0.878 | 28.38 | 0.814 | 37.29 | 0.957 | 31.97 | 0.882 | 29.72 | 0.826 | 285G  |      |      |      |        |

Fig. 4: Visual comparisons of image denoising on the CBSD68 [36] (upper row) and SIDD [39] (bottom row) datasets for color Gaussian and real image denoising, respectively. Due to the page limits, more visual results for different datasets could be found in our github page.

[1] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, “Residual dense network for image super-resolution,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2752–2761.

[2] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, “Cycleisp: Real image restoration via improved data synthesis,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2696–2705.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[4] Bumjun Park, Songhyun Yu, and Jechang Jeong, “Densely connected hierarchical network for image denoising,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0.

[5] Javier Gurrola-Ramos, Oscar Dalmau, and Teresa E Alarcón, “A residual dense u-net neural network for image denoising,” IEEE Access, vol. 9, pp. 31742–31754, 2021.

[6] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 517–532.

[7] Drew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 31742–31754, 2021.
Table 2: Real-world image denoising. Image denoising result on SIDD [39] and DND [40] datasets. * denotes the method used additional training data. The proposed SRMNet is only trained on the SIDD images and then tested on DND.

| Methods      | SIDD [39] | DND [40] |
|--------------|-----------|----------|
|              | PSNR | SSIM | PSNR | SSIM |
| DrCNN         | 23.66 | 0.583 | 32.43 | 0.790 |
| BM3D         | 25.65 | 0.685 | 34.51 | 0.851 |
| CBDNet*       | 30.78 | 0.801 | 38.06 | 0.942 |
| RIDNet*       | 38.71 | 0.951 | 39.26 | 0.953 |
| DAGL          | 38.94 | 0.953 | 39.77 | 0.956 |
| AINDNet*      | 38.95 | 0.952 | 39.37 | 0.951 |
| VDN           | 39.28 | 0.956 | 39.39 | 0.952 |
| DeamNet       | 39.35 | 0.955 | 39.63 | 0.953 |
| SADNet*       | 39.46 | 0.957 | 39.59 | 0.952 |
| CycleISP      | 39.52 | 0.957 | 39.56 | 0.956 |
| DAntNet*      | 39.47 | 0.957 | 39.58 | 0.955 |
| MPRNet        | 39.71 | 0.958 | 39.80 | 0.954 |
| SRMNet (Ours) | 39.72 | 0.959 | 39.88 | 0.956 |

Table 3: Comparison of the PSNR and FLOPs for MIRNet [20], MPRNet [32] and our method (SRMNet). We did the test on SIDD validation set which has 1,280 patches where FLOPs are estimated on the input with shape of 256 × 256. The inference times are measured on the computer equipped with NVIDIA GTX 1080Ti GPU.

| Methods      | PSNR | FLOPs (G) | Time (ms) | Speedup |
|--------------|------|-----------|-----------|---------|
| MIRNet       | 39.72 | 787.04    | 212.94    | 1×      |
| MPRNet       | 39.71 | 573.88    | 128.77    | 1.65×   |
| SRMNet (Ours)| 39.72 | 285.36    | 71.29     | 2.98×   |

of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1874–1883.

[19] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang, “Selective kernel networks,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 510–519.

[20] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, “Multi-stage progressive image restoration and enhancement,” in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII. 16, Springer, 2020, pp. 492–511.

[21] Antoni Buades, Bartomeu Coll, and J-M Morel, “A non-local algorithm for image denoising,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). IEEE, 2005, vol. 2, pp. 60–65.

[22] Weisheng Dong, Xin Li, Lei Zhang, and Guangming Shi, “Sparsity-based image denoising via dictionary learning and structural clustering,” in CVPR 2011. IEEE, 2011, pp. 457–464.

[23] Li Xu, Shicheng Zheng, and Jiaya Jia, “Unnatural 10 sparse representation for natural image deblurring,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 1107–1114.

[24] William T Freeman, Thouis R Jones, and Egon C Pasztor, “Example-based super-resolution,” IEEE Computer graphics and Applications, vol. 22, no. 2, pp. 56–65, 2002.

[25] Songhyun Yu, Bumjun Park, and Jechang Jeong, “Deep iterative down-up cnn for image denoising,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0.

[26] Shi Guo, Zifan Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang, “Toward convolutional blind denoising of real photographs,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1712–1722.

[27] Saeed Anwar and Nick Barnes, “Real image denoising with feature attention,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3155–3164.

[28] Zongsheng Yue, Hongwei Yong, Qian Zhao, Lei Zhang, and Deyu Meng, “Variational denoising network: Toward blind noise modeling and removal,” arXiv preprint arXiv:1908.11374, 2019.

[29] Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu, “Spatial-adaptive network for single image denoising,” in European Conference on Computer Vision, Springer, 2020, pp. 171–187.

[30] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.

[31] Xiaoming Li, Hao Chen, Xiaoyuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng, “H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes,” IEEE transactions on medical imaging, vol. 37, no. 12, pp. 2663–2674, 2018.

[32] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, “Multi-stage progressive image restoration,” arXiv e-prints, pp. arXiv–2102, 2021.

[33] Raghav Mehta and Jayanti Sivaswamy, “M-net: A convolutional neural network for deep brain structure segmentation,” in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017, pp. 437–440.

[34] Sukesh Adiga and Jayanti Sivaswamy, “Fpd-m-net: Fingerprint image denoising and inpainting using m-net based convolutional neural networks,” in Inpainting and Denoising Challenges, pp. 51–61. Springer, 2019.

[35] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud, “Two deterministic half-quadratic regularization algorithms for computed imaging,” in Proceedings of 1st International Conference on Image Processing. IEEE, 1994, vol. 2, pp. 168–172.

[36] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik, “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. IEEE, 2001, vol. 2, pp. 416–423.

[37] Rich Franken, “Kodak lossless color image suite,” source: http://riff.us/graphics/kodak, vol. 4, no. 2, 1999.

[38] Eirikur Agustsson and Radu Timofte, “Ntire 2017 challenge on single image super-resolution: Dataset and study,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 126–135.

[39] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown, “A high-quality denoising dataset for smartphone cameras,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1692–1700.

[40] Tobias Ploetz and Stefan Roth, “Benchmarking denoising algorithms with real photographs,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1586–1595.

[41] Chong Mou, Jian Zhang, and Zhuoyuan Wu, “Dynamic attentive graph learning for image restoration,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4328–4337.
[42] Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho, “Transfer learning from synthetic to real-noise denoising with adaptive instance normalization,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3482–3492.

[43] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao, “Adaptive consistency prior based deep network for image denoising,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8596–8606.

[44] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng, “Dual adversarial network: Toward real-world noise removal and noise generation,” in European Conference on Computer Vision. Springer, 2020, pp. 41–58.