Selective Residual M-Net for Real Image Denoising

Chi-Mao Fan and Tsung-Jung Liu
Department of Electrical Engineering
National Chung Hsing University
Taichung 40227, Taiwan
Email: qaz5517359@gmail.com; tjliu@dragon.nchu.edu.tw

Kuan-Hsien Liu and Ching-Hsiang Chiu
Department of Computer Science and Information Engineering
National Taichung University of Science and Technology
Taichung 40401, Taiwan
Email: khliu@nutc.edu.tw; cjx860126@gmail.com

***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/FanChiMao/SRMNet.**

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

I. 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:

$$Y = D(X) + n, \tag{1}$$

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) [1], non-local means (NLM) [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 [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 [3, 6, 7, 8, 9], dense connection [10, 11, 12, 13, 14], residual dense block [15, 16, 17, 18], attention mechanisms [19, 20, 21], channel attention block [19, 22], and hierarchical architecture [23, 24, 25, 26]. 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 efficient structure named selective residual block (SRB). Moreover, we use the multi-scale feature fusion with two different sampling methods (pixel shuffle [27], 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) [28, 29] 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.

The rest of this paper is organized as follows. In Section II, we are going to introduce the related works on image denoising and some techniques applied at our proposed model. Then we will describe the proposed SRMNet model in Section III. In Section IV, we will report the visual quality [30, 31] results and numerical comparisons between our model and other state-of-the-art methods. In the end, we are going to make conclusion remarks in Section V.

II. Related Work

A. 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, 32], sparse coding [33, 34] and dictionary learning [33, 35]. Currently, CNN-based denoisers have demonstrated state-of-the-art results [3, 6, 23, 24, 25]. Moreover, the denoise models from [3, 4, 5, 21, 24, 25, 36] only deal with signal-independent noise (e.g., AWGN, read noise), while the model from [22, 37, 38, 39, 40] have the ability to process the real-world signal-dependent noise (e.g., shot noise, thermal noise).

B. Selective Kernel Network

Li *et al.* proposed the selective kernel convolution that has two branches. One of the path utilizes normal naive 3×3 convolution kernel to extract features, and the other path adopts different kernel size (e.g. 5×5, 7×7) to obtain the larger

This work is supported by the Ministry of Science and Technology, Taiwan, under Grant No. MOST 109-2221-E-005-059.

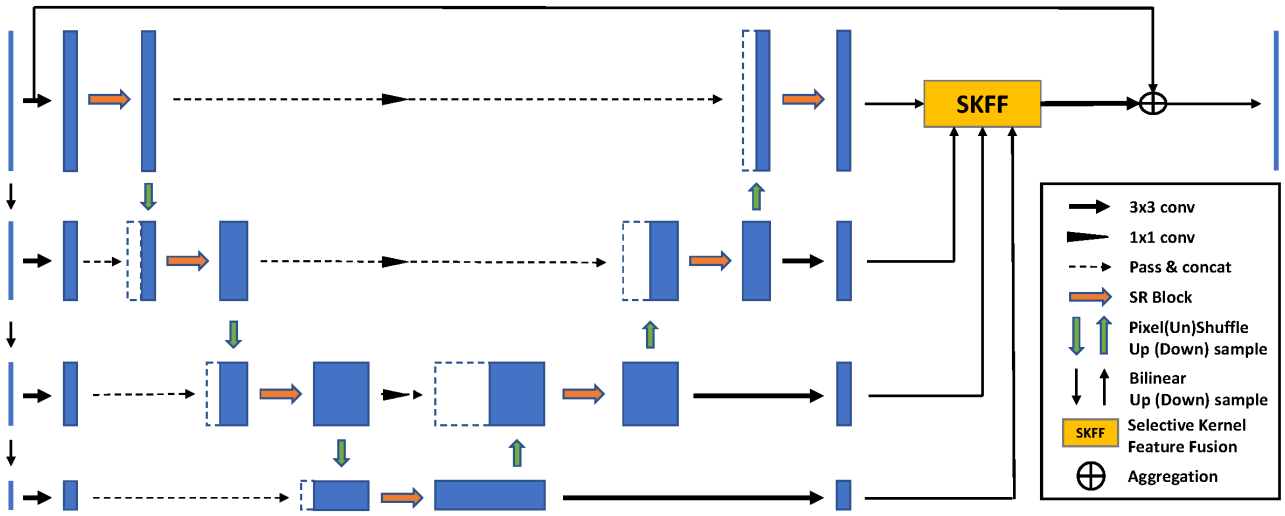

Fig. 1: Proposed Selective Residual M-Net (SRMNet) architecture. The source code and component structure of the model could be found in the provided URL indicated in the abstract. We set the initial channel in each resolution to 96 after 3×3 convolution, and totally we have 4 layers in the proposed M-Net.

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.* [29] were inspired by [28] and applied it to multi-scale feature fusion for image enhancement tasks, which also achieve good results.

C. Hierarchical architecture

Undoubtedly, the most representative hierarchical architecture is U-Net [23] which is first proposed to be used on image segmentation. Furthermore, due to the strong adaptive backbone, the U-Net can be easily applied with different extractive blocks to enhance the performance in lots of computer vision tasks. The hierarchical design has the advantages of learning the abundant representation with different feature sizes. However, the down-sample and up-sample in the encoder and decoder easily destroy the spatial information. The loss of pixel-localization has a big influence on pixel-wise vision tasks like instance segmentation and image restoration. Therefore, it is important to have trade-off between contextual information and spatial details.

III. 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.

A. M-Net

The M-Net architecture is first proposed for medical image segmentation [41]. Adiga *et al.* [42] used the same framework for fingerprint image denoising and also got 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 2×2 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 [29] which does not concatenate each feature map but aggregates the weighting features.

B. SRMNet

The proposed SRMNet for image denoising is shown in Fig. 1. We first use 3×3 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 [43] for image denoising as follows:

$$\mathcal{L}_{char} = \sqrt{||\hat{X} - X||^2 + \varepsilon^2}, \tag{2}$$

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 ε in Eq.(2) are empirically set to 10^{-3}.

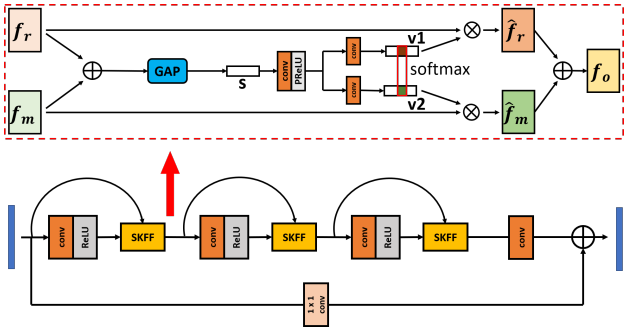

Fig. 2: Illustration of the Selective Residual Block (SRB).

C. Selective Residual Block

Fig. 2 shows the architecture of the proposed SRB which is improved from the residual dense block (RDB) [15]. 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 3×3 convolution and add the long skip connection with 1×1 convolution between the input and output.

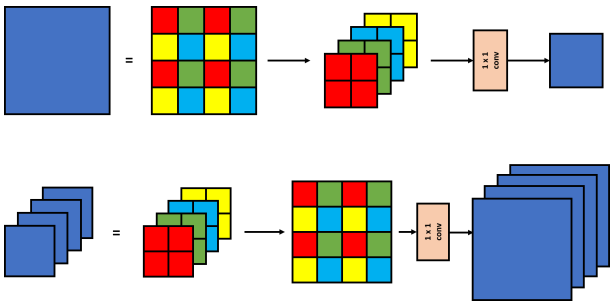

Fig. 3: Resizing module with pixel (un)shuffle.

D. Resizing Module

As for resizing module, we simply use bilinear downsampling 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 3×3 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 ingredients of bilinear features are obvious less than pixel unshuffled 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 increasing the weight of bilinear features.

IV. Experiments

A. 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 most commonly used objective evaluation indices [44, 45]: Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity (SSIM).

B. Experiment Datasets

Real-World Image Denoising. To train our SRMNet for real-world denoising, we follow [29, 46] to use 320 high-resolution images of SIDD dataset [47]. Evaluations also follow aforementioned methods, which is to perform the test on 1280 validation patches from the SIDD dataset [47] and 1000 patches from the DND benchmark dataset [48]. The resolution of all patches is 256×256 in both training and testing.

TABLE I: **Real-world image denoising**. Image denoising result on SIDD [47] and DND [48] 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 [47]		DND [48]	
	PSNR	SSIM	PSNR	SSIM
DnCNN [3]	23.66	0.583	32.43	0.790
BM3D [1]	25.65	0.685	34.51	0.851
CBDNet* [37]	30.78	0.801	38.06	0.942
RIDNet* [38]	38.71	0.951	39.26	0.953
DAGL [49]	38.94	0.953	39.77	0.956
AINDNet* [50]	38.95	0.952	39.37	0.951
VDN [39]	39.28	0.956	39.39	0.952
DeamNet* [51]	39.35	0.955	39.63	0.953
SADNet* [40]	39.46	0.957	39.59	0.952
CycleISP* [22]	39.52	0.957	39.56	0.956
DANet+* [52]	39.47	0.957	39.58	0.955
MPRNet [46]	39.71	0.958	39.80	0.954
MIRNet [29]	**39.72**	0.959	**39.88**	**0.956**
SRMNet (Ours)	39.72	**0.959**	39.44	0.951

Gaussian Color Image Denoising. For Gaussian denoising, we use the same experimental setup as image denoising [24, 25] and train our model on image super-resolution DIV2K [55] 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. Evaluation is conducted on noise levels 10, 30, 50 on CBSD68 [53] and Kodak24 [54]. It should be noted that our model does not know the noise level in the testing, which means SRMNet is the blind denoising model.

C. Image Denoising Performance

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 I. Compared to the previous state-of-the-art CNN-based method [29] in SIDD dataset, our model gains the same scores with

TABLE II: **Gaussian color image denoising**. Image denoising results on CBSD68 dataset [53] and Kodak24 dataset [54]. 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 256×256 color images.

Methods	CBSD68 [53]						Kodak24 [54]						FLOPs
	$\sigma = 10$		$\sigma = 30$		$\sigma = 50$		$\sigma = 10$		$\sigma = 30$		$\sigma = 50$		
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
BM3D [1]	35.89	0.951	29.71	0.843	27.36	0.763	33.32	0.943	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 [4]	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
DnCNN [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 [24]	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 [21]	36.43	-	30.63	-	26.83	-	37.24	-	31.86	-	29.58	-	-
DIDN [36]	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 [15]	36.47	-	30.67	-	28.31	-	37.31	-	31.94	-	29.66	-	1490G
RDUNet [25]	**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

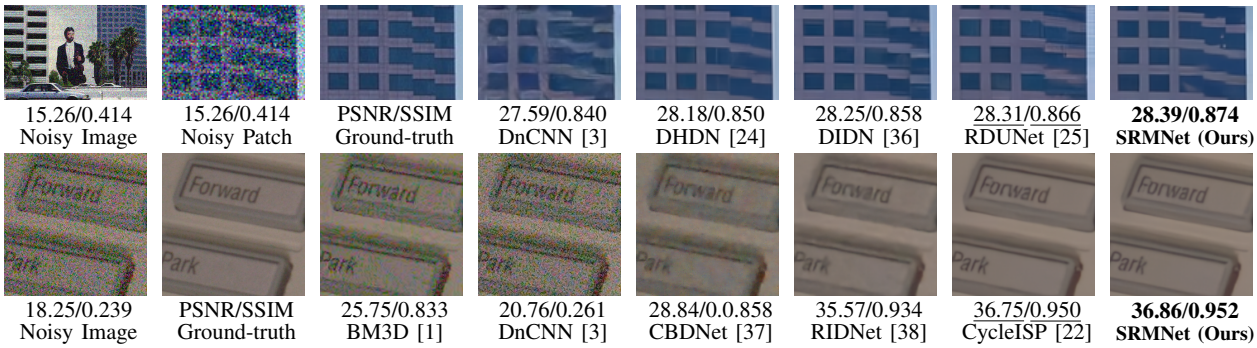

Fig. 4: Visual comparisons of image denoising on the CBSD68 [53] (upper row) and SIDD [47] (bottom row) datasets for color Gaussian and real image denoising, respectively. Due to the page limits, more visual results for different datasets could be found in our github page.

MIRNet but less computational complexity (e.g., FLOPs) and time cost. More specifically, Table III 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 removes noise and the denoised images are visually closer to the ground-truth.

Gaussian Color Image Denoising. In Table II and Fig. 4, we compare our SRMNet with the prior-based method (e.g., BM3D [1]), CNN-based methods (e.g., DnCNN [3], IrCNN [5], FFDNet [4]) and the models which are based on RDB (e.g., DHDN [24], DIDN [36], RDN [15], RDUNet [25]). According to Table II, we could observe three things: 1) The proposed 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 ($\downarrow 20\%$) 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.

TABLE III: Comparison of the PSNR and FLOPs for MIRNet [29], MPRNet [46] 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 $1 \times 3 \times 256 \times 256$. The inference times are measured on the computer equipped with NVIDIA GTX 1080Ti GPU.

Methods	PSNR	FLOPs (G)		Time (ms)	Speedup
MIRNet [29]	39.72	787.04	100%	212.94	1×
MPRNet [46]	39.71	573.88	72.9%	128.77	1.65×
SRMNet (Ours)	39.72	285.36	36.3%	71.29	2.98×

V. 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 unshuffle 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. In addition, we will also consider to combine the recent global feature extraction blocks (e.g., Transformer, MLP-Mixer) with our proposed modified M-Net to achieve better performances.

REFERENCES

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3d filtering," in *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, vol. 6064, 2006, p. 606414.

[2] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images," *IEEE transactions on medical imaging*, vol. 27, no. 4, pp. 425–441, 2008.

[3] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[4] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[5] K. Zhang, W. Zuo, S. Gu, and L. 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] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[7] K.-H. Liu, T.-J. Liu, C.-C. Wang, H.-H. Liu, and S.-C. Pei, "Modern architecture style transfer for ruin or old buildings," in *IEEE International Symposium on Circuits and Systems*, 2019, pp. 1–5.

[8] C.-C. Wang, H.-H. Liu, S.-C. Pei, K.-H. Liu, and T.-J. Liu, "Face aging on realistic photos by generative adversarial networks," in *IEEE International Symposium on Circuits and Systems*, 2019, pp. 1–5.

[9] K.-C. Chang, T.-J. Liu, K.-H. Liu, and D.-Y. Chao, "Locating waterfowl farms from satellite images with parallel residual u-net architecture," in *IEEE International Conference on SMC*, 2020, pp. 114–119.

[10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on CVPR*, 2017, pp. 4700–4708.

[11] G. Huang, S. Liu, L. Van der Maaten, and K. Q. Weinberger, "Condensenet: An efficient densenet using learned group convolutions," in *Proceedings of the IEEE conference on CVPR*, 2018, pp. 2752–2761.

[12] B.-X. Chen, T.-J. Liu, K.-H. Liu, H.-H. Liu, and S.-C. Pei, "Image super-resolution using complex dense block on generative adversarial networks," in *IEEE ICIP*, 2019, pp. 2866–2870.

[13] Y.-Z. Su, T.-J. Liu, K.-H. Liu, H.-H. Liu, and S.-C. Pei, "Image inpainting for random areas using dense context features," in *IEEE ICIP*, 2019, pp. 4679–4683.

[14] K.-H. Liu, T.-J. Liu, and F. Wang, "Cbl: A clothing brand logo dataset and a new method for clothing brand recognition," in *2020 28th European Signal Processing Conference*, 2021, pp. 655–659.

[15] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on CVPR*, 2018, pp. 2472–2481.

[16] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 517–532.

[17] K.-Y. Wen, T.-J. Liu, K.-H. Liu, and D.-Y. Chao, "Identifying poultry farms from satellite images with residual dense u-net," in *IEEE International Conference on SMC*, 2020, pp. 102–107.

[18] Y.-Z. Chen, T.-J. Liu, and K.-H. Liu, "Super-resolution of satellite images by two-dimensional rrdb and edge-enhancement generative adversarial network," in *IEEE ICASSP*, 2022, pp. 1825–1829.

[19] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the ECCV*, 2018, pp. 286–301.

[20] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proceedings of the IEEE/CVF Conference on CVPR*, 2019, pp. 11 065–11 074.

[21] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," *arXiv preprint arXiv:1903.10082*, 2019.

[22] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Cycleisp: Real image restoration via improved data synthesis," in *IEEE CVPR*, 2020, pp. 2696–2705.

[23] O. Ronneberger, P. Fischer, and T. 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.

[24] B. Park, S. Yu, and J. Jeong, "Densely connected hierarchical network for image denoising," in *IEEE/CVF CVPR Workshops*, 2019, pp. 0–0.

[25] J. Gurrola-Ramos, O. Dalmau, and T. E. Alarcón, "A residual dense u-net neural network for image denoising," *IEEE Access*, vol. 9, pp. 31 742–31 754, 2021.

[26] T.-J. Liu, C.-T. Lin, H.-H. Liu, and S.-C. Pei, "Blind stereoscopic image quality assessment based on hierarchical learning," *IEEE Access*, vol. 7, pp. 8058–8069, 2019.

[27] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE conference on CVPR*, 2016, pp. 1874–1883.

[28] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proceedings of the IEEE Conference on CVPR*, 2019, pp. 510–519.

[29] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *ECCV*. Springer, 2020, pp. 492–511.

[30] T.-J. Liu, H.-H. Liu, S.-C. Pei, and K.-H. Liu, "A high-definition diversity-scene database for image quality assessment," *IEEE Access*, vol. 6, pp. 45 427–45 438, 2018.

[31] T.-J. Liu, K.-H. Liu, and K.-H. Shen, "Learning based no-reference metric for assessing quality of experience of stereoscopic images," *Journal of Visual Communication and Image Representation*, vol. 61, pp. 272–283, 2019.

[32] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 60–65.

[33] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *IEEE CVPR*, 2011, pp. 457–464.

[34] L. Xu, S. Zheng, and J. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 1107–1114.

[35] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[36] S. Yu, B. Park, and J. 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.

[37] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE/CVF Conference on CVPR*, 2019, pp. 1712–1722.

[38] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3155–3164.

[39] Z. Yue, H. Yong, Q. Zhao, L. Zhang, and D. Meng, "Variational denoising network: Toward blind noise modeling and removal," *arXiv preprint arXiv:1908.11314*, 2019.

[40] M. Chang, Q. Li, H. Feng, and Z. Xu, "Spatial-adaptive network for single image denoising," in *ECCV*. Springer, 2020, pp. 171–187.

[41] R. Mehta and J. Sivaswamy, "M-net: A convolutional neural network for deep brain structure segmentation," in *IEEE 14th International Symposium on Biomedical Imaging*, 2017, pp. 437–440.

[42] S. Adiga V and J. Sivaswamy, "Fpd-m-net: Fingerprint image denoising and inpainting using m-net based convolutional neural networks," in *Inpainting and Denoising Challenges*. Springer, 2019, pp. 51–61.

[43] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proceedings of ICIP*. IEEE, 1994, pp. 168–172.

[44] T.-J. Liu, K.-H. Liu, J. Y. Lin, W. Lin, and C.-C. J. Kuo, "A paraboost method to image quality assessment," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 1, pp. 107–121, 2017.

[45] T.-J. Liu, "Study of visual quality assessment on pattern images: Subjective evaluation and visual saliency effects," *IEEE Access*, vol. 6, pp. 61 432–61 444, 2018.

[46] S. Waqas Zamir, A. Arora, S. Khan, M. Hayat, F. Shahbaz Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," *arXiv e-prints*, pp. arXiv–2102, 2021.

[47] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proceedings of the IEEE Conference on CVPR*, 2018, pp. 1692–1700.

[48] T. Plotz and S. Roth, "Benchmarking denoising algorithms with real photographs," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1586–1595.

[49] C. Mou, J. Zhang, and Z. Wu, "Dynamic attentive graph learning for image restoration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4328–4337.

[50] Y. Kim, J. W. Soh, G. Y. Park, and N. I. Cho, "Transfer learning from synthetic to real-noise denoising with adaptive instance normalization," in *IEEE Conference on CVPR*, 2020, pp. 3482–3492.

[51] C. Ren, X. He, C. Wang, and Z. Zhao, "Adaptive consistency prior based deep network for image denoising," in *Proceedings of the IEEE/CVF Conference on CVPR*, 2021, pp. 8596–8606.

[52] Z. Yue, Q. Zhao, L. Zhang, and D. Meng, "Dual adversarial network: Toward real-world noise removal and noise generation," in *European Conference on Computer Vision*. Springer, 2020, pp. 41–58.

[53] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of IEEE ICCV*, vol. 2, 2001, pp. 416–423.

[54] R. Franzen, "Kodak lossless true color image suite," *source: http://r0k. us/graphics/kodak*, vol. 4, no. 2, 1999.

[55] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE conference on CVPR workshops*, 2017, pp. 126–135.