Denoising for Photon-limited Imaging via a Multi-level Pyramid Network

Shansi Zhang
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong, China
sszhang@eee.hku.hk

Edmund Y. Lam
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong, China
elam@eee.hku.hk

TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON) | 978-1-6654-5095-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/TENCON55691.2022.9977646

***Abstract*—Imaging under photon-scarce situations introduces challenges to many applications as the captured images are with low signal-to-noise ratio. Here, we target on the denoising under photon-limited imaging. We develop a multi-level pyramid denoising network (MPDNet), which employs the idea of Laplacian pyramid to learn the small-scale noise map and larger-scale high-frequency details at different levels. Feature extractions are conducted on the multi-scale input images to encode richer contextual information. The major component of MPDNet is the multi-skip attention residual block, which integrates multi-scale feature fusion and attention mechanism for better feature representation. Experimental results have demonstrated that our MPDNet can achieve superior denoising performance on the images with low photon counts.**

***Index Terms*—Photon-limited imaging, Multi-level pyramid denoising, multi-skip attention residual block**

I. Introduction

Imaging under photon-scarce situations bring challenges to many applications, such as surveillance, robotics and computational photography. As the photons become scarce, the signal-to-noise ratio of the captured images will be lower. The raw images could be severely corrupted by the noise under extremely low photon levels, which introduces large difficulties to image restoration. To alleviate the problem of photon-scarce imaging, some advanced sensors, including single-photon avalanche diode (SPAD) and quanta image sensor (QIS), which has single-photon detection sensitivity and photon-counting capability [1], [2], are developed. These sensors are good candidates for low-light imaging due to their higher sensitivity to photons than CMOS image sensors. However, noise still can not be avoided under photon-limited conditions. Here, we use 'photons per pixel' (ppp) representing the average number of photons that each pixel detects during the exposure period. As the average photon count decreases, the raw image is more severely corrupted. Our target is to develop a denoising framework that can deal with severe noise under photon-limited imaging.

The techniques for image denoising mainly include model-based methods and deep learning-based methods. The model-based methods utilize the priors of image or noise to obtain clean images through traditional optimization algorithm [3]–[5]. The concrete methods include non-local filter [6]–[9], sparse coding [10]–[12], external priors [13], [14] and low rank [15], [16]. However, these methods only assume additive white Gaussian noise and are not applicable to more complicated noise.

Nowadays, most researches on image denoising are based on deep learning. Some methods [17], [18] combine prior knowledge and deep neural network, which may not be robust enough due to the limitation of priors. Some methods adopt completely data-driven approach to learn the noisy-to-clean mapping. For example, Zhang et al. [19] proposed DnCNN, which adopts residual learning to remove Gaussian noise. Guo et al. [20] proposed CBDNet by considering realistic noise for blind denoising, with a noise estimation sub-network to avoid under-estimation for noise level by asymmetric learning. Anwar et al. [21] developed RIDNet for real image denoising, which uses residual on the residual structure to ease the low-frequency flow with feature attention for exploring the channel dependencies. Tian et al. [22] proposed ADNet, with global and local information integration and attention mechanism to achieve real and blind denoising. Zamir et al. [23] proposed MIRNet for image restoration, which contains multi-resolution branches with information exchange, spatial and channel attention mechanism and attention-based feature fusion to learn enriched features. Quan et al. [24] exploited the advantage of complex-valued operations and developed a complex-valued CNN for image denoising.

Most the existing methods are designed to learn a direct mapping from the noisy image to the clean image, which may not be effective to deal with images with high noise levels. Thus, we design a novel network, which is capable of suppressing severe noise under extremely photon-limited conditions. Our main contributions are as follows:

- We develop a multi-level pyramid denoising network (MPDNet), which adopts the idea of Laplacian pyramid to learn the small-scale noise map and larger-scale high-frequency maps at different levels. The learned high-frequency components are gradually added to the up-sampled denoised images to obtain the clean images with proper sharpness. We also introduce multi-scale image feature extraction to encode richer contextual information.
- We design a multi-skip attention residual block (MARB), which integrates multi-scale feature fusion and attention

mechanism to obtain better feature representation. Moreover, it has much fewer parameters than the common residual block.
- Our MPDNet is very lightweight with high efficiency. Experiments have been conducted to demonstrate its effectiveness for restoring images with severe noise.

II. Proposed method

A. Multi-level pyramid denoising network

To reduce the learning difficulty for severe noise suppression, we develop the MPDNet, which adopts multi-level learning approach to estimate the small-scale noise map and larger-scale high-frequency maps. We draw on the idea of Laplacian pyramid, which can decompose an image into a low-frequency component at the top level and a series of high-frequency components with different scales. Specifically, the lth high-frequency component $h_l \in \mathbb{R}^{\frac{h}{2^l} \times \frac{w}{2^l}}$ is obtained by $h_l = t_l - \mathrm{up_g}(\mathrm{down_g}(t_l))$, where t_l is the image at the lth level, $\mathrm{up_g}(\cdot)$ and $\mathrm{down_g}(\cdot)$ denote $2\times$ up-sampling and down-sampling with Gaussian smoothing, and the original image t_0 has a shape of $h \times w$. With Laplacian pyramid, the original image can be reconstructed without distortion.

Our MPDNet shown in Fig. 1 consists of three information flows, including the $\frac{1}{4}$-scale denoising flow f_{D} (purple lines), $\frac{1}{2}$-scale high-frequency flow $f_{\frac{1}{2}\mathrm{HF}}$ (blue lines) and full-scale high-frequency flow f_{HF} (red lines). These three flows share the same feature extraction process (orange lines).

The feature extractions are conducted on the multi-scale input images, i.e. I_1, $I_{\frac{1}{2}}$, $I_{\frac{1}{4}}$, to encode the contextual information with different scales. $I_{\frac{1}{2}}$ and $I_{\frac{1}{4}}$ are obtained by down-sampling the input image I_1. Gaussian smoothing is used before each down-sampling to avoid the aliasing effect. I_1 is first fed to a convolution layer and a multi-skip attention residual block (MARB) to obtain the full-scale features F_1. F_1 further passes through a convolution layer with $\mathrm{stride} = 2$ and an MARB, and the output is added with the features extracted from $I_{\frac{1}{2}}$, to obtain the $\frac{1}{2}$-scale features $F_{\frac{1}{2}}$. Similarly, $F_{\frac{1}{2}}$ is further down-sampled and then added to the features extracted from $I_{\frac{1}{4}}$ to obtain the $\frac{1}{4}$-scale features $F_{\frac{1}{4}}$. In this way, the features with different scales and levels are aggregated to better guide the subsequent learning of noise removal and high-frequency details.

In the denoising flow, $F_{\frac{1}{4}}$ passes through two MARBs and a convolution layer with Tanh activation to estimate the noise map, which is added to $I_{\frac{1}{4}}$ to yield the $\frac{1}{4}$-scale output image $O_{\frac{1}{4}}$. Usually, clean images can provide better features for the learning of high-frequency components. Thus, in the $\frac{1}{2}$-scale high-frequency flow, the features of the denoised image $O_{\frac{1}{4}}$ are first extracted, and then up-sampled and added to $F_{\frac{1}{2}}$ to supplement the high-resolution features. Next, these features are fed to the following two MARBs and a convolution layer with Tanh activation to obtain the $\frac{1}{2}$-scale high-frequency map $H_{\frac{1}{2}}$. As with the Laplacian pyramid, $\frac{1}{2}$-scale output image $O_{\frac{1}{2}}$ is obtained by up-sampling $O_{\frac{1}{4}}$ and adding $H_{\frac{1}{2}}$. In the full-scale high-frequency flow, the features before the last convolution layer in the $\frac{1}{2}$-scale high-frequency flow are up-sampled and then added to F_1, followed by the subsequent MARBs and convolution layer to yield the full-scale high-frequency map H_1. The final output image O_1 is obtained by up-sampling $O_{\frac{1}{2}}$ and adding H_1.

The information flows of MPDNet can be represented as

$$
\begin{aligned}
&I_{\frac{1}{2}} = \mathrm{down_g}(I_1),\ I_{\frac{1}{4}} = \mathrm{down_g}(I_{\frac{1}{2}}),\\
&O_{\frac{1}{4}} = f_{\mathrm{D}}(I_1, I_{\frac{1}{2}}, I_{\frac{1}{4}}) + I_{\frac{1}{4}},\\
&H_{\frac{1}{2}} = f_{\frac{1}{2}\mathrm{HF}}(I_1, I_{\frac{1}{2}}, I_{\frac{1}{4}}),\ H_1 = f_{\mathrm{HF}}(I_1, I_{\frac{1}{2}}, I_{\frac{1}{4}}),\\
&O_{\frac{1}{2}} = \mathrm{up}(O_{\frac{1}{4}}) + k \times H_{\frac{1}{2}},\ O_1 = \mathrm{up}(O_{\frac{1}{2}}) + k \times H_1,
\end{aligned}
$$

where $up(\cdot)$ and $down(\cdot)$ mean $2\times$ up-sampling and down-sampling operations, and k is a factor that can be manually designated during testing to control the image sharpness.

Since it is much easier to obtain small-scale clean and smooth images, denoising is only conducted on the $\frac{1}{4}$-scale images. However, the up-sampled small-scale images are always blurring. Thus, high-frequency components need to be learned to supplement the details. The high-frequency details are mainly reflected on the object boundaries and textures, and learning these sparse structures are definitely easier than learning the completed image features. Our MPDNet can be trained end-to-end to learn the noise removal and high-frequency details simultaneously.

B. Multi-skip attention residual block

The main component of our MPDNet is the MARB, as shown in Fig. 2. First, there are three 3×3 convolution layers to progressively extract features. Suppose the number of output channels is C, then we set the first convolution layer outputs $\frac{C}{2}$ channels, and the second and third convolution layers output $\frac{C}{4}$ channels. All these outputs are concatenated together, followed by 1×1 convolution layers for feature fusion. Here, we adopts two 1×1 convolution layers to reduce the number of parameters and increase the nonlinearity. The first one compresses the channels to $\frac{C}{4}$ and the second one expands the channels back to C, which constitutes a bottleneck structure. Then, the fused features are fed to another 3×3 convolution layer with sigmoid activation to determine the attention map, which is multiplied with the fused features to highlight more important ones and suppress less useful ones. Finally, the calibrated features are added to the input to form the residual connection.

This simple structure integrates both multi-scale feature fusion and attention mechanism. The outputs of the first three convolution layers have different receptive fields, which is equivalent to use convolutions with different kernel sizes to extract features. The concatenation and 1×1 convolutions aims to fuse the features with different scales. Then, spatial attention is adopted to adaptively determine the informative features for better feature representation. Our MARB is very lightweight compared with the common residual block (RB) with two 3×3 convolution layers, each of which outputs

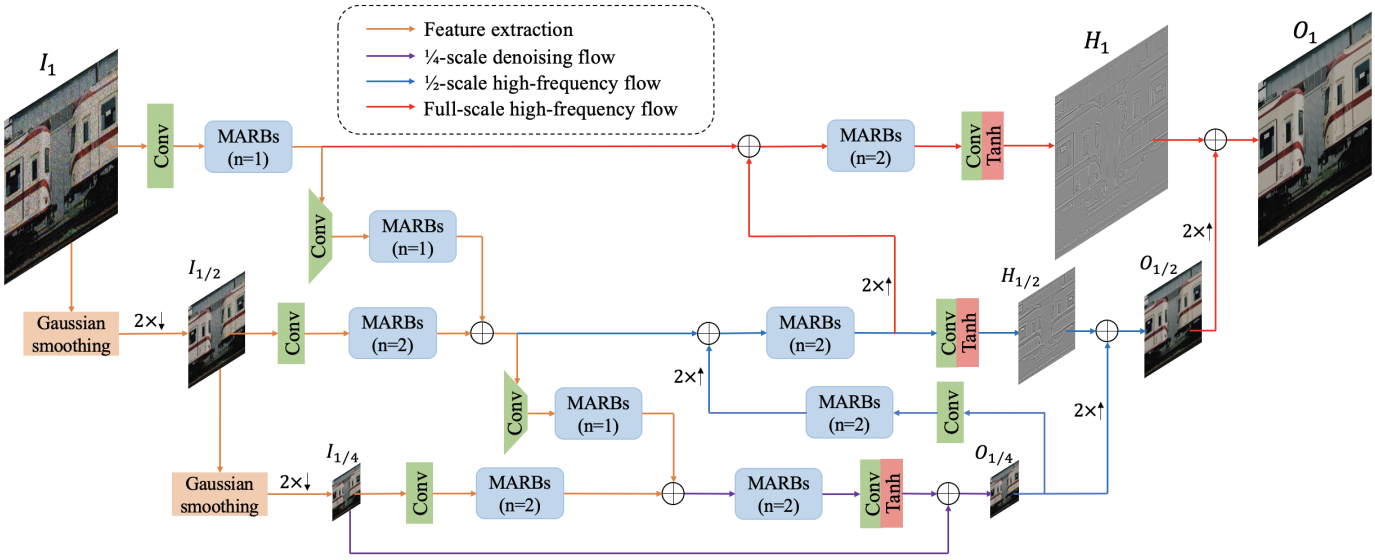


Fig. 1. The architecture of MPDNet, which consists of a $\frac{1}{4}$-scale denoising flow, a $\frac{1}{2}$-scale high-frequency flow and a full-scale high-frequency flow.

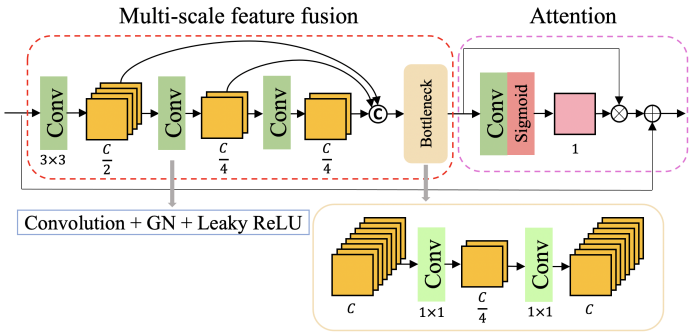


Fig. 2. The structure of MARB.

C channels. The Parameters of the two types of blocks are calculated as follows

$$\begin{aligned} P_{\text{MARB}} &= C\times 3\times 3\times \frac{C}{2}+\frac{C}{2}\times 3\times 3\times \frac{C}{4} \\ &\quad +\frac{C}{4}\times 3\times 3\times \frac{C}{4}+C\times 1\times 1\times \frac{C}{4}\times 2 \\ &\quad +C\times 3\times 3\times 1=\frac{107}{16}C^2+9C, \\ P_{\text{RB}} &= C\times 3\times 3\times C\times 2=18C^2. \end{aligned}$$

Then, we can obtain the ratio $r=\frac{P_{\text{MARB}}}{P_{\text{RB}}}=\frac{6.69C+9}{18C}$. As long as $C>0.796$, $r<1$. The channel numbers we set are between 16 and 256. Therefore, MARB has much fewer parameters than RB given the same input and output channel numbers.

C. Loss functions

The loss functions for training MPDNet include the ℓ_1 loss of high-frequency components

$$\ell_H=\sum_{s=1,\frac{1}{2}}\left\|H_s-\hat{H}_s\right\|_1, \tag{1}$$

and the ℓ_1 loss of output images with different scales

$$\ell_O=\sum_{s=1,\frac{1}{2},\frac{1}{4}}\left\|O_s-\hat{O}_s\right\|_1, \tag{2}$$

where $\hat{H}_{\frac{1}{2}}$, $\hat{H}_1$, $\hat{O}_{\frac{1}{4}}$ and $\hat{O}_{\frac{1}{2}}$ are the corresponding ground truths obtained by the Laplacian pyramid.

Following [25], we use SSIM [26] loss ℓ_{SSIM} and perceptual loss [27] ℓ_{vgg} for O_1 to further improve the visual quality of the output images. Then, the full loss is written as

$$\ell_{\text{MPD}}=\ell_H+\ell_O+\ell_{\text{SSIM}}(O_1,\hat{O}_1)+\ell_{\text{vgg}}(O_1,\hat{O}_1). \tag{3}$$

III. Experiments

A. Training data and implementation details

We adopt a similar method with [28], [29] to simulate the low-photon-count imaging, which is formulated by the following equation,

$$X_{\text{syn}}=K\Big(\underbrace{\text{Poisson}\big(X_{\text{Bayer}}/\text{mean}(X_{\text{Bayer}})\times p\big)}_{\text{photon count}}+n_r\Big), \tag{4}$$

where X_{Bayer} is the original Bayer image, p is a factor to control the photon level, and n_r is the readout noise following Gaussian distribution. K is to scale the pixel values to $0\sim 255$ to yield the raw Bayer image X_{syn} with Poisson-Gaussian mixed noise. Then, the interpolation demosaicing operation is employed to obtain the RGB images.

We synthesize the training data based on the Pascal VOC2012 dataset. The average photon count for each image is randomly sampled from the range $[1,10]$ during training. We show several examples of synthetic raw Bayer images and RGB images with different average photon counts, as shown in Fig. 3. It can be seen that the noise level becomes higher with the decrease of average photon count due to the lower signal-to-noise ratio.

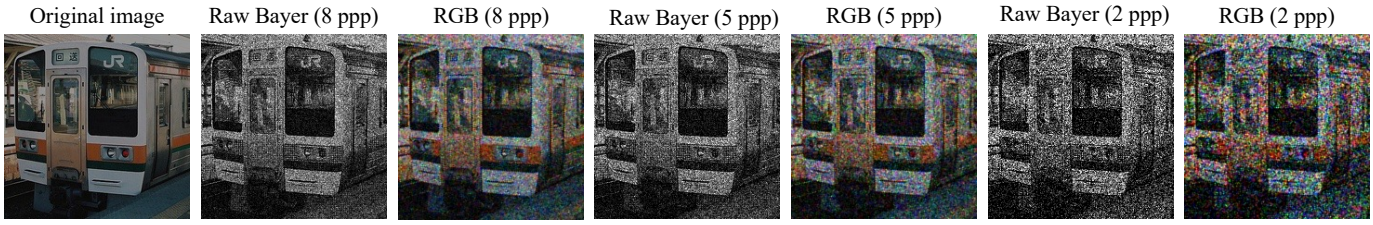


Fig. 3. Synthetic raw Bayer images and RGB images with different average photon counts.

During training, the input images were synthesized from the ground truths with random photon counts and cropped to 256×256 patches randomly. Adam optimizer with a initial learning rate of 1×10^{-3} and a decay rate of 0.8 was used. The MPDNet was trained for 200 epochs with batch size of 16. It was implemented by PyTorch on Nvidia Tesla P100 GPU.

B. Comparison

TABLE I
QUANTITATIVE RESULTS OF DIFFERENT METHODS

Methods	Params.	FLOPs	PSNR	SSIM	LPIPS↓
Our MPDNet	1.709M	15.431G	**28.36**	**0.8270**	**0.1726**
ADNet [22]	1.169M	76.631G	26.98	0.7765	0.2237
RIDNet [21]	1.499M	98.126G	27.24	0.7966	0.2081
MIRNet [23]	2.660M	66.565G	27.95	0.8089	0.1876

Here, we compare our MPDNet with several other blind denoising methods, including ADNet [22], RIDNet [21] and MIRNet [23]. These methods design different architectures to learn the residual map. We use the same datasets and training strategies to train them. The parameters, FLOPs and quantitative results (PSNR, SSIM, LPIPS [30]) of these methods are recorded in Table I. For this LPIPS, smaller value is better. It can be seen that the other methods have worse quantitative results but with much more FLOPs compared with ours. Therefore, our MPDNet can achieve better denoising performance with higher efficiency.

Table I lists the average results containing various photon levels. To compare the denoising performance at each photon level, we plot the PSNR, SSIM and LPIPS curves as a function of the average photon count ranging from 1 to 10, as shown in Fig. 6. The values of all the metrics become better with the increase of photon count, and our MPDNet outperforms other methods obviously at any photon level.

We present some visual results of different methods, as shown in Fig. 4 and 5. The input images are with low average photon counts, where some texture details are completely corrupted by the severe noise. The visual comparison indicates that our MPDNet can suppress noise more effectively, and restore clearer object details and edge information.

C. Ablation study

We first provide one example of our restoration (Fig. 7), as well as the intermediate outputs, including the $\frac{1}{4}$-scale denoised image, $\frac{1}{2}$-scale high-frequency image, $\frac{1}{2}$-scale denoised image and full-scale high-frequency image. As can be seen, our MPDNet can output smooth $\frac{1}{4}$-scale image and effectively predict the $\frac{1}{2}$-scale and full-scale high-frequency components to compose the high-quality denoised image.

TABLE II
ABLATION STUDY ON THE NETWORK ARCHITECTURE

Configurations	Params.	FLOPs	PSNR	SSIM
Default	1.709 M	15.431 G	**28.36**	**0.8270**
Replace MARB with RB	2.446 M	21.880 G	28.23	0.8257
w/o multi-scale input images	1.212 M	12.045 G	27.47	0.8081
w/o multi-level structure	1.543 M	14.543 G	27.26	0.7918

We conduct an ablation study on the network architecture by comparing the performance of several different configurations. To study the advantage of our MARB, we replace all the MARBs with RBs. As shown in Table II, this configuration ('Replace MARB with RB') achieves a little lower PSNR and SSIM than our default setting but with more parameters and computation, which reflects the effectiveness and efficiency of our MARB.

To investigate the benefit of feature extraction on multi-scale input images, we remove the corresponding convolution layers and MARBs after $I_{\frac{1}{2}}$ and $I_{\frac{1}{4}}$. Hence, only the features of full-scale image I_1 are extracted. It can be seen that the performance of this configuration ('w/o multi-scale input images') are degraded obviously even if the parameters and FLOPs are reduced. Therefore, the feature extraction on multi-scale images contributes to better denoising performance due to the encoding of richer contextual information.

Our MPDNet adopts an multi-level learning strategy with three information flows to learn the noise map and high-frequency maps, respectively. Thus, we investigate the benefit of this multi-level structure by building and testing another network, which has similar architecture with our MPDNet in general but directly outputs the full-scale denoised images without the multi-level learning. It can be seen that this configuration ('w/o multi-level structure') has the worst results, which suggests that the direct noisy-to-clean mapping can not suppress noise effectively when the noise level is high.

IV. CONCLUSION

In this paper, we study on the denoising under photon-limited imaging. We develop a MPDNet which adopts a

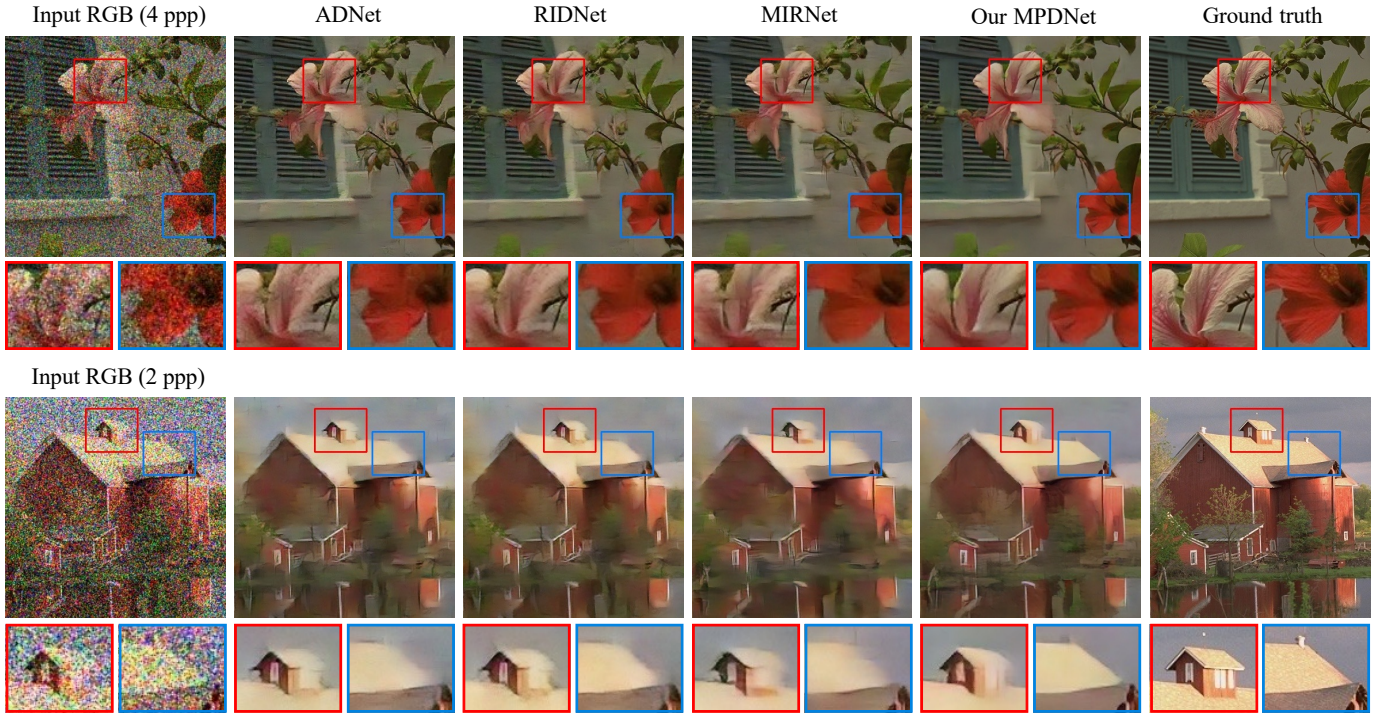


Fig. 4. Visual comparison among different methods on the Kodak dataset.

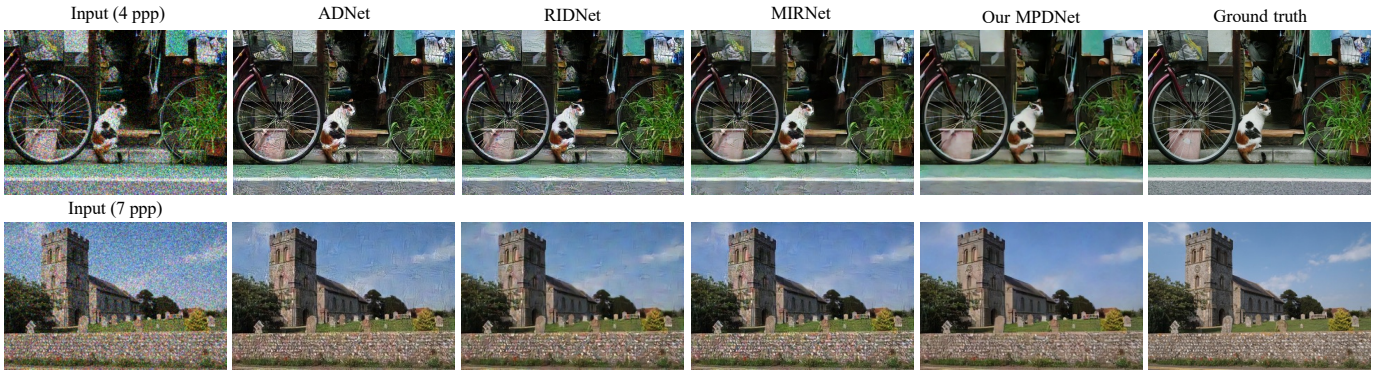
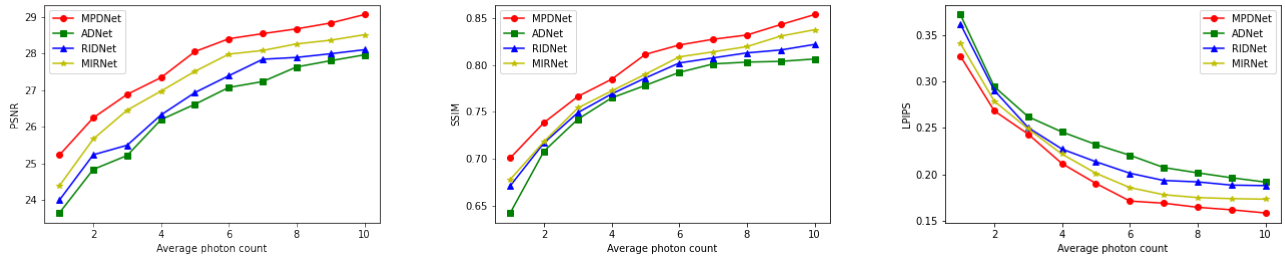


Fig. 5. Visual comparison among different methods on the Pascal VOC2012 dataset.



(a) PSNR (b) SSIM (c) LPIPS

Fig. 6. PSNR, SSIM and LPIPS of different methods at various photon levels.

multi-level structure and utilizes the idea of Laplacian pyramid to estimate the small-scale noise map and larger-scale high-frequency maps progressively. Feature extractions are performed on the multi-scale input images to encode richer

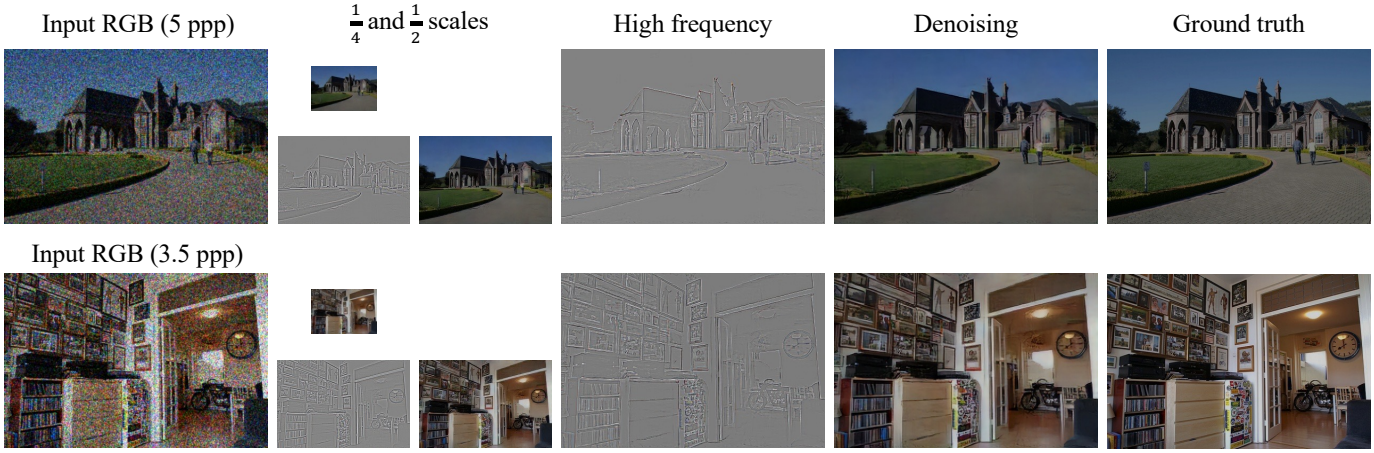


Fig. 7. One restoration example by our MPDNet, including the intermediate outputs, i.e., the $\frac{1}{4}$-scale denoised images, $\frac{1}{2}$-scale high-frequency images, $\frac{1}{2}$-scale denoised images and full-scale high-frequency images.

contextual information. The main component of our MPDNet is the MARB, which integrates multi-scale feature fusion and spatial attention for better feature representation, with much fewer parameters than the common residual block. Experimental results have verified the effectiveness of our method for image denoising.

ACKNOWLEDGMENT

The work is supported in part by the Research Grants Council of Hong Kong (GRF 17200019, 17201620, 17200321) and by ACCESS — AI Chip Center for Emerging Smart Systems, Hong Kong SAR.

REFERENCES

[1] S. H. Chan, O. A. Elgendy, and X. Wang, "Images from bits: Non-iterative image reconstruction for quanta image sensors," *Sensors*, vol. 16, no. 11, 2016.

[2] O. A. Elgendy and S. H. Chan, "Optimal threshold design for quanta image sensor," *IEEE Transactions on Computational Imaging*, vol. 4, no. 1, pp. 99–111, 2018.

[3] Y. Zhu, T. Zeng, K. Liu, Z. Ren, and E. Y. Lam, "Full scene underwater imaging with polarization and an untrained network," *Optics Express*, vol. 29, no. 25, pp. 41 865–41 881, 2021.

[4] L. Song, Z. Ge, and E. Y. Lam, "Dual alternating direction method of multipliers for inverse imaging," *IEEE Transactions on Image Processing*, vol. 31, pp. 3295–3308, 2022.

[5] L. Song and E. Y. Lam, "Fast and robust phase retrieval for masked coherent diffractive imaging," *Photonics Research*, vol. 10, no. 3, pp. 758–768, 2022.

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

[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[8] Z. Sun, S. Chen, and L. Qiao, "A general non-local denoising model using multi-kernel-induced measures," *Pattern Recognition*, vol. 47, no. 4, pp. 1751–1763, 2014.

[9] H. Li and C. Y. Suen, "A novel non-local means image denoising method based on grey theory," *Pattern Recognition*, vol. 49, pp. 237–248, 2016.

[10] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[11] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 2272–2279.

[12] J. Xu, L. Zhang, and D. Zhang, "A trilateral weighted sparse coding scheme for real-world image denoising," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 20–36.

[13] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 479–486.

[14] F. Chen, L. Zhang, and H. Yu, "External patch prior guided internal clustering for image denoising," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 603–611.

[15] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 700–711, 2013.

[16] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2862–2869.

[17] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1256–1272, 2017.

[18] S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3587–3596.

[19] 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.

[20] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1712–1722.

[21] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 3155–3164.

[22] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided cnn for image denoising," *Neural Networks*, vol. 124, pp. 117–129, 2020.

[23] 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 *European Conference on Computer Vision (ECCV)*, 2020, pp. 492–511.

[24] Y. Quan, Y. Chen, Y. Shao, H. Teng, Y. Xu, and H. Ji, "Image denoising using complex-valued deep cnn," *Pattern Recognition*, vol. 111, p. 107639, 2021.

[25] S. Zhang and E. Y. Lam, "Learning to restore light fields under low-light imaging," *Neurocomputing*, vol. 456, pp. 76–87, 2021.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[27] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4681–4690.

[28] O. A. Elgendy, A. Gnanasambandam, S. H. Chan, and J. Ma, "Low-light demosaicking and denoising for small pixels using learned frequency selection," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 137–150, 2021.

[29] S. Zhang and E. Y. Lam, "An effective decomposition-enhancement method to restore light field images captured in the dark," *Signal Processing*, vol. 189, 2021.

[30] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.