IPOL IMAGE PROCESSING ON LINE

Published in Image Processing On Line on 2023–05–03.
Submitted on 2022–11–20, accepted on 2023–03–25.
ISSN 2105–1232 © 2023 IPOL & the authors CC–BY–NC–SA
This article is available online with supplementary materials,
software, datasets and online demo at
https://doi.org/10.5201/ipol.2023.446

An Analysis of Multi-stage Progressive Image Restoration Network (MPRNet)

Boshra Rajaei[1,2], Sara Rajaei[1,2], Hossein Damavandi[1,2]

[1]Saiwa Company, Ontario, Canada (info@saiwa.ai)
[2]Computer and Information Technology Department, Sadjad University, Mashhad, Iran
(brajaei@sadjad.ac.ir, hossein.damavandi@sadjad.ac.ir)

Communicated by Jean-Michel Morel *Demo edited by* Boshra Rajaei

Abstract

Multi-stage progressive image restoration network (MPRNet) is a three-stage CNN (convolutional neural network) for image restoration. MPRNet has been shown to provide high performance gains on several datasets for a range of image restoration problems including image denoising, deblurring, and deraining. The network is interesting because it manages to remove the three kinds of artifacts with a single architecture. Here, we provide an overview of the network and study its performance and computational complexity in comparison with other state-of-the-art methods.

Source Code

The source code and documentation for this algorithm are available from the web page of this article[1]. The source code is borrowed from the MPRNet original code and pre-trained models[2]. Usage instructions are included in the `README.txt` file of the archive.
This is an MLBriefs article, the source code has not been reviewed!

Keywords: image restoration; denoising; deblurring; deraining

1 Introduction

Image restoration is about restoring clean images by removing distortions such as blur, noise, and raining artifacts. There are several common causes of these artifacts in images. For instance blur artifacts may result from camera movement, subject movement, scattered light distortion, insufficient depth of field, and lens softness. Noise refers to the random appearance of undesired traces and variations in the brightness or color information. Images are inevitably contaminated by noise during acquisition and transmission. The level of noise typically varies with the length of exposure, physical

[1]https://doi.org/10.5201/ipol.2023.446
[2]https://github.com/swz30/MPRNet

temperature, and sensitivity setting of the camera. Depending on the noise source there are a few types of noises, such as: white Gaussian noise, impulse noise, periodic noise and banding noise. Here, we focus on Gaussian noise as a common sensor and electronic circuit noise that may also arise due to poor illumination or high temperature. Gaussian noise is a random statistical noise having a normal probability density function. Finally, in applications like video surveillance and self-driving cars, one has to process images and videos containing undesirable raining artifacts that may affect performance of the processing algorithm. Therefore, pre-processing steps to remove these artifacts are crucial.

Deblurring is about recovering a sharp image from a blurred input, and is inherently an inverse problem and hence does not have a unique solution. Image deblurring has been studied for decades. There are numerous image deblurring algorithms, from classic methods, e.g. [2, 4, 11], which mathematically estimate the blur kernel and then reverse its effect, to more recent machine learning based methods which benefit from recent advances in machine learning and deep learning, e.g. [17, 21, 23, 16, 5]. Anger et al [2] provide a blind kernel estimation and deblurring method based on the ℓ_0 gradient prior. The first step of this method is to estimate the blur kernel by alternating between a sharp image prediction using the ℓ_0 prior on the gradient image and a kernel estimation in a multi-scale manner. Once the kernel is estimated, a sharp image is predicted using a standard non-blind deconvolution method and the estimated kernel from the previous stage. For more mathematical details of the method, we refer to [2]. Since our main focus in this study is analyzing the MPRNet method [21] in deblurring mode, we compare the results experimentally with the other two deep networks, i.e. scale-recurrent network (SRN) [17] and multi-input multi-output U-net (MIMO-UNet) [5].

SRN [17] is one of the previous efforts besides MPRNet to bring the multistage design to image deblurring. Multi-scale networks followed the coarse-to-fine approaches in optimization based frameworks. SRN employs an encoder-decoder architecture which is effective in encoding broad contextual information but unreliable in preserving spatial image details. On the other hand, compared to other multi-stage methods SRN has a simpler network structure and a smaller number of parameters and hence, it is easier to train. Differently, MIMO-UNet [5] adopted the single-stage design of U-net and achieved competitive results. We also compare MPRNet results with MIMO-UNet as another coarse-to-fine network. MIMO-UNet proposed an architecture that facilitates information flow across different image resolutions in a multi-scale U-net, but it introduced complicated connections between various sized feature maps.

Additive white Gaussian noise is the classic noise model adapted to most digital cameras. The model assumes that noise at each pixel is an independent additive Gaussian normal random variable. Its variance is often assumed to be constant over all pixels, which is ensured if a previous Anscombe transform has been applied to the raw image. There are a wide range of classic image denoisers that try to model image noise mathematically, e.g. [6, 12, 8, 20]; and there are also many recent solutions motivated by recent advances in machine learning and deep learning, e.g. [21, 18, 7, 3]. Multi-scale discrete cosine transform (DCT) denoising is a classic denoising algorithm with low computational complexity [12]. Here, it is employed in our experimental results section for comparison. The original DCT denoising algorithm starts by thresholding a patch-wise DCT of the noisy input image and then aggregating the resulting patches. There are variants of DCT denoising. A two-step multi-scale version proposed in [12] enhances the performance of the original method significantly and also reduces halo artifacts in the denoised image.

We compare MPRNet denoising performance with non-linear activation free network (NAFNet) [3] that is trained on a similar dataset. The NAFNet is a deep learning model for image restoration based on the U-shaped network and contains a variable number of width and number of enclosing/decoding/middle blocks. The architecture has been shown to maintain a relatively small memory footprint while achieving exceptional output performance.

The last restoration mode of MPRNet is deraining. Beside MPRNet, image deraining has been studied in a few other efforts [22, 9, 13]. In this paper, we compare the performance of the density-aware image de-raining method with a multistream dense network (DID-MDN) [22] method. DID-MDN consists of two main stages architecture: a residual-aware rain-density classification (i.e. heavy, medium or light) and a multi-stream densely-connected network rain streak removal. For more technical details, we refer to [22].

In this paper, we first explain the MPRNet algorithm with complete detail in Section 2. Then, in Section 3 the MPRNet method is compared with three recent well-known restoration algorithms: SRN [17] and MIMO-UNet [5] deblurring, NAFNet [3] denoising and DID-MDN deraining [22]. Section 4 concludes the paper.

2 MPRNet Restoration Network Architecture

Over the past few years, data-driven CNN architectures have demonstrated superiority over conventional restoration approaches. In this paper, we are interested in a CNN model that can be applied to three restoration tasks: deblurring, denoising, and deraining. The multi-stage progressive image restoration network (MPRNet) [21] consists of three stages to progressively restore images, as shown in Figure 1. The first two stages have an encoder-decoder model based on the standard U-net [15] and learn the full contextual information of the input image. In the last stage, the original resolution subnetwork (ORSNet) operates at the original image resolution to generate spatially accurate outputs for a pixel-to-pixel correspondence between the input and output images.

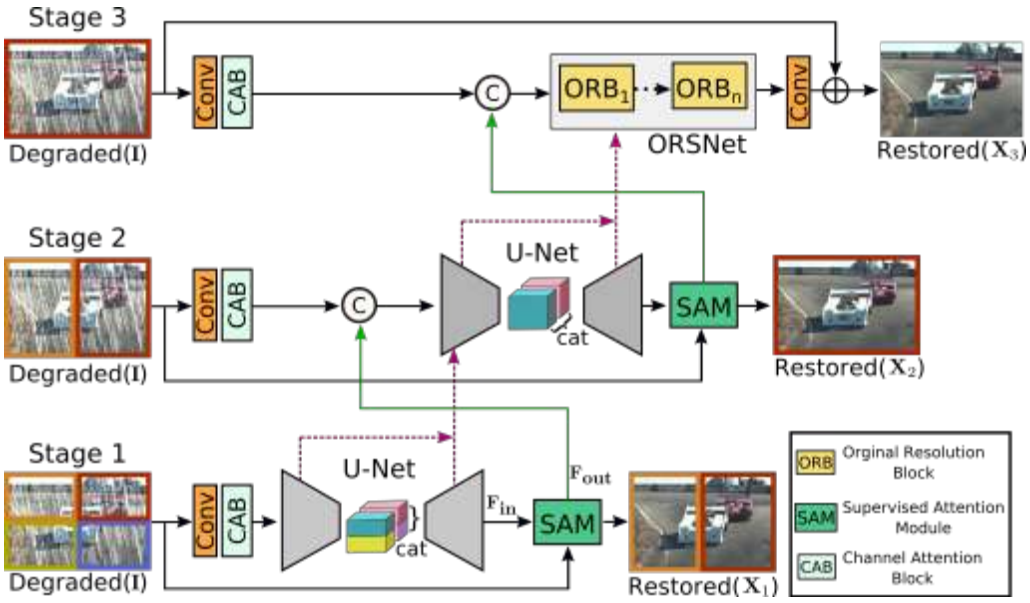


Figure 1: Multi-stage architecture for progressive image restoration of MPRNet (printed from [21]). The input image is divided into non-overlapping patches: four for stage 1, two for stage 2 and the original image for the final stage.

Among convolutional networks, encoder-decoder-based U-net architectures have been mainly studied for restoration due to their hierarchical multi-scale representation and computational efficiency. In order to selectively attend to relevant information, spatial and channel attention modules have also been incorporated. MPRNet presented an improved multi-layer architecture that is designed to take into account both high level global features as well as local details. In order to keep the same spatial resolution at different scales, and thus reduce the expensive runtime of the deconvolution/upsampling operation, MPRNet introduced a multi-patch hierarchy as an input. In addition, by

varying the patch count at each scale, the coarser scales concentrate on local information, producing residual information for the following finer scale.

In the encoder-decoder sub-network, features are extracted at each stage using channel attention blocks (CABs) and upsampled using bi-linear interpolation followed by a convolution layer. In the ORSNet stage of MPRNet we do not observe any down-sampling operation. High-resolution features are calculated using multiple original resolution blocks (ORBs), each of which further contains CABs.

Between every two stages, a cross-stage feature fusion (CSFF) module is introduced as shown in Figure 2. This makes the overall network more robust through several up- and down-sampling, and more stable as it eases the flow of information.

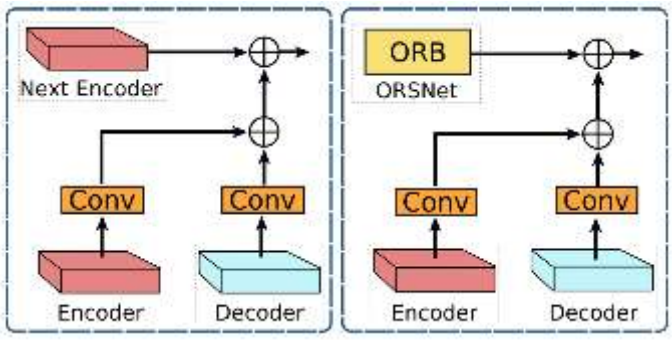


Figure 2: MPRNet cross-stage feature fusion (CSFF) modules. (Left) CSFF between stage 1 and stage 2. (Right) CSFF between stage 2 and ORSNet. (Printed from [21]).

Finally, a supervised attention module (SAM) is defined between every two stages, which results in significant performance gain. Figure 3 shows SAM components with the two main benefits. First, SAM provides ground-truth supervisory signals for restoration of the progressive image at each stage. Second, it suppresses features at the current stage through the attention maps and only passes useful features to the next stage.

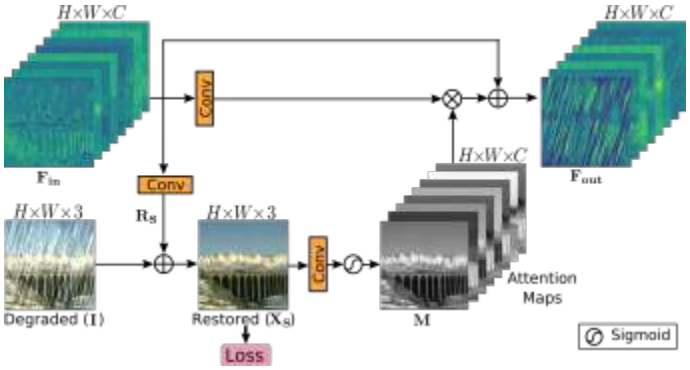


Figure 3: MPRNet supervised attention module (SAM) (printed from [21]).

In a nutshell, MPRNet multi-stage structure provides several key features:

1. An encoder-decoder for learning multi-scale contextual information in the first two stages.
2. Preservation of fine spatial details of the input image by operating on the original image resolution in the last stage.
3. A supervised attention module (SAM) that enables progressive learning.

4. Cross-stage feature fusion (CSFF) to propagate multi-scale contextualized features from early to late stage.

3 Experimental Results

MPRNet is trained for the three image restoration tasks: deblurring, denoising, and deraining, over different datasets. For deblurring, the network is trained using the GoPro dataset [10] with 2,103 pairs of clean-blur synthetically generated images. The network also is evaluated using 1,111 pairs from the same dataset. In the original paper, it has been experimentally proved that the trained model works well on images from the RealBlur dataset [14] that are captured in real-world conditions. For denoising, training is performed using 320 high-resolution images of the SIDD-Medium dataset [1] (gamma-corrected sRGB images are used) and is evaluated using 1,280 patches from the same dataset in addition to 1,000 patches from the DND dataset. Finally, deraining employed 13,712 clean-rain image pairs from multiple datasets for training like Rain100L dataset [19]. While in this paper, we evaluate the MPRNet performance using the pre-trained and available models from the original paper, the interested reader might re-train the network on other sets of images using the instructions that are provided in the public git that we introduced earlier in the source code section.

3.1 Deblurring

Deblurring is the first image restoration operation that we investigate using MPRNet. Here, we are interested in comparing the MPRNet deblurring algorithm with another deep learning based deblurring methods, i.e. SRN and MIMO-UNet. We introduced both methods briefly in Section 1. The output results of the SRN and MIMO-UNet methods were generated using their corresponding demos[3] or original source codes[4].

We compared the performance of MPRNet with SRN and MIMO-Unet over four images from the test category of the GoPro dataset as it is displayed in Figure 4. Note that the three methods have been trained over the same dataset and hence the models are completely comparable. MPRNet works subjectively better than the other two methods. In terms of PSNR, MPRNet outperforms MIMO-UNet on average by 0.51 dB and SRN by 1.39 dB. All the PSNRs are calculated over the large images and not the selected patches.

3.2 Denoising

The next restoration function of MPRNet is denoising. The MPRNet denoising method, as for its other functionalities, is fully blind. It works without any knowledge of the noise level. The only input of the network is a noisy image. In this section, we compare the denoising performance of MPRNet with another method based on deep learning, i.e. NAFNet [3]. Both MPRNet and NAFNet were trained over the SIDD dataset and hence their performance are comparable. For NAFNet, we perform the experiments using its original source code[5].

We examined the denoising power of the algorithms for three different noise levels: 10, 20, and 40 in terms of Gaussian standard deviation. Figures 5 to 7 show the accuracy comparison in terms of PSNR and visual quality using the four images from public datasets. While for low levels of noise we observe that NAFNet performs better than MPRNet, both visually and in PSNR, for higher levels, MPRNet outperforms NAFNet significantly by 0.31 to 1.51 dB for different images. We have to

[3] https://github.com/jiangsutx/SRN-Deblur
[4] https://github.com/chosj95/MIMO-UNet
[5] https://github.com/megvii-research/NAFNet

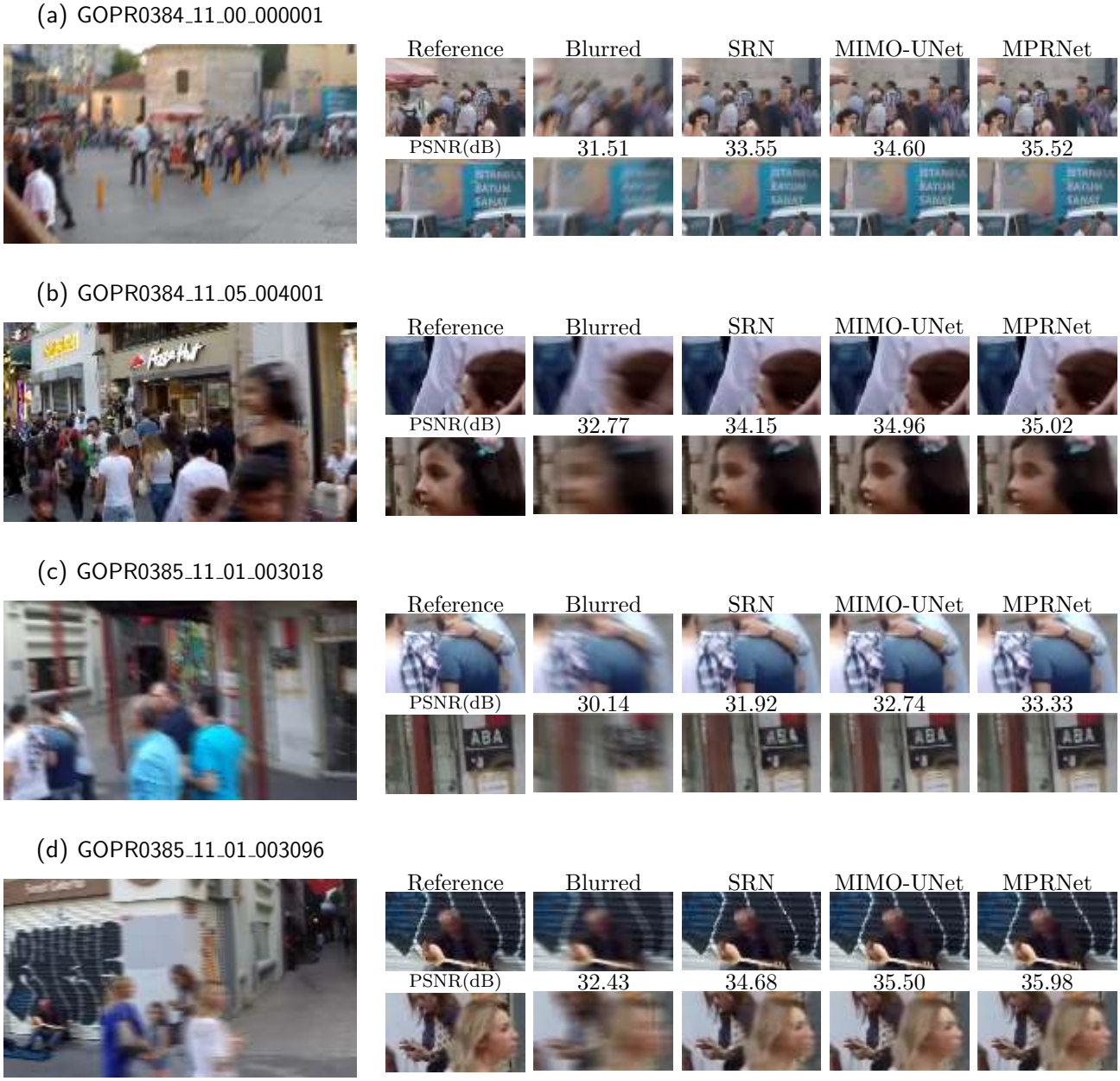


Figure 4: Visual and PSNR comparison of deblurring results using SRN, MIMO-UNet and MPRNet over four images from the GoPro dataset [10]. The size of the original images is 1280×720 and each patch is 256×144.

mention that the images are from public sources and not included in training and validation steps and are not even similar to the datasets that the network was trained on.

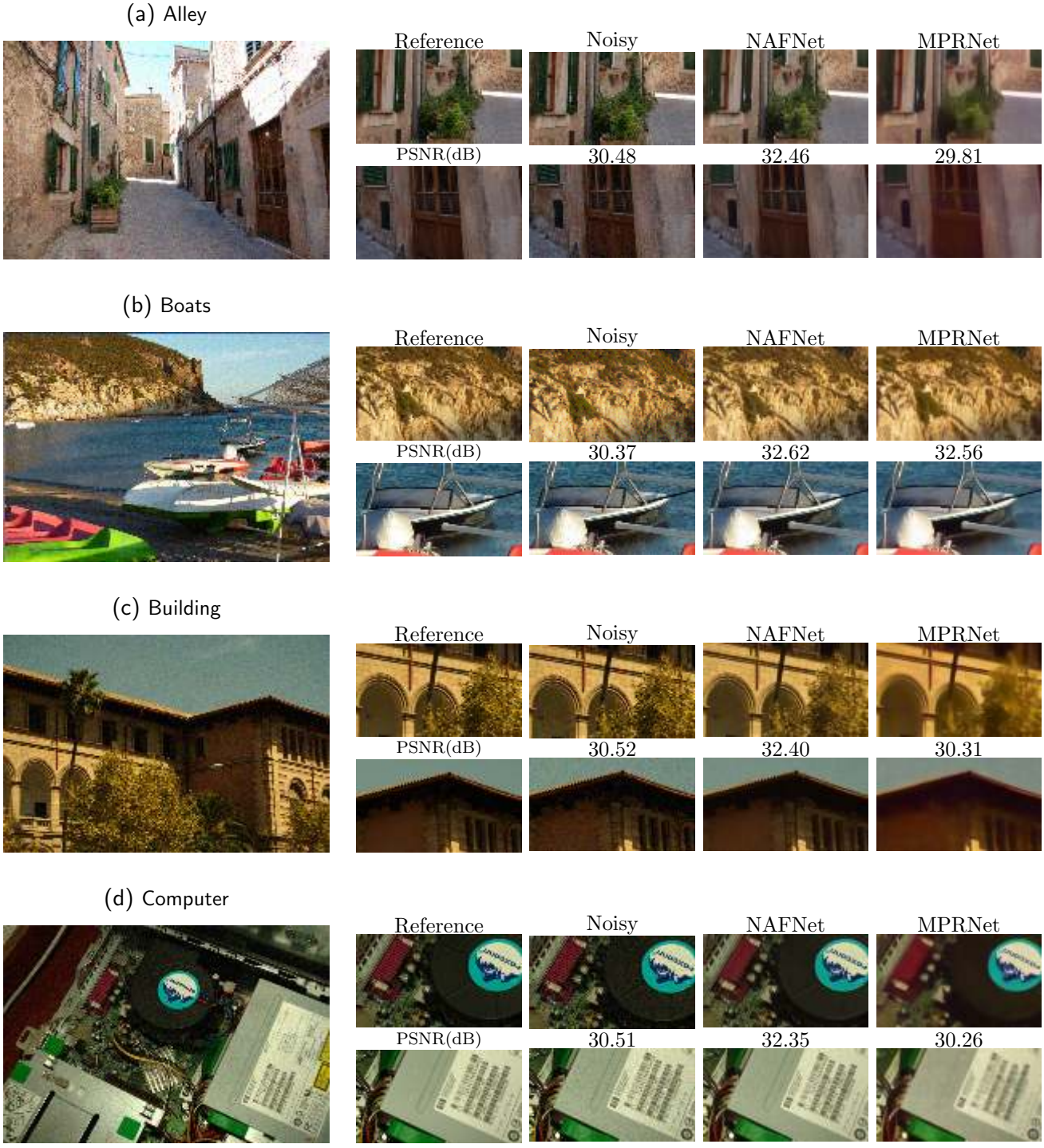


Figure 5: Visual and PSNR comparison of denoising results using NAFNet and MPRNet over four images from a public dataset with Gaussian noise level $\sigma = 10$. The size of each patch is 256×144.

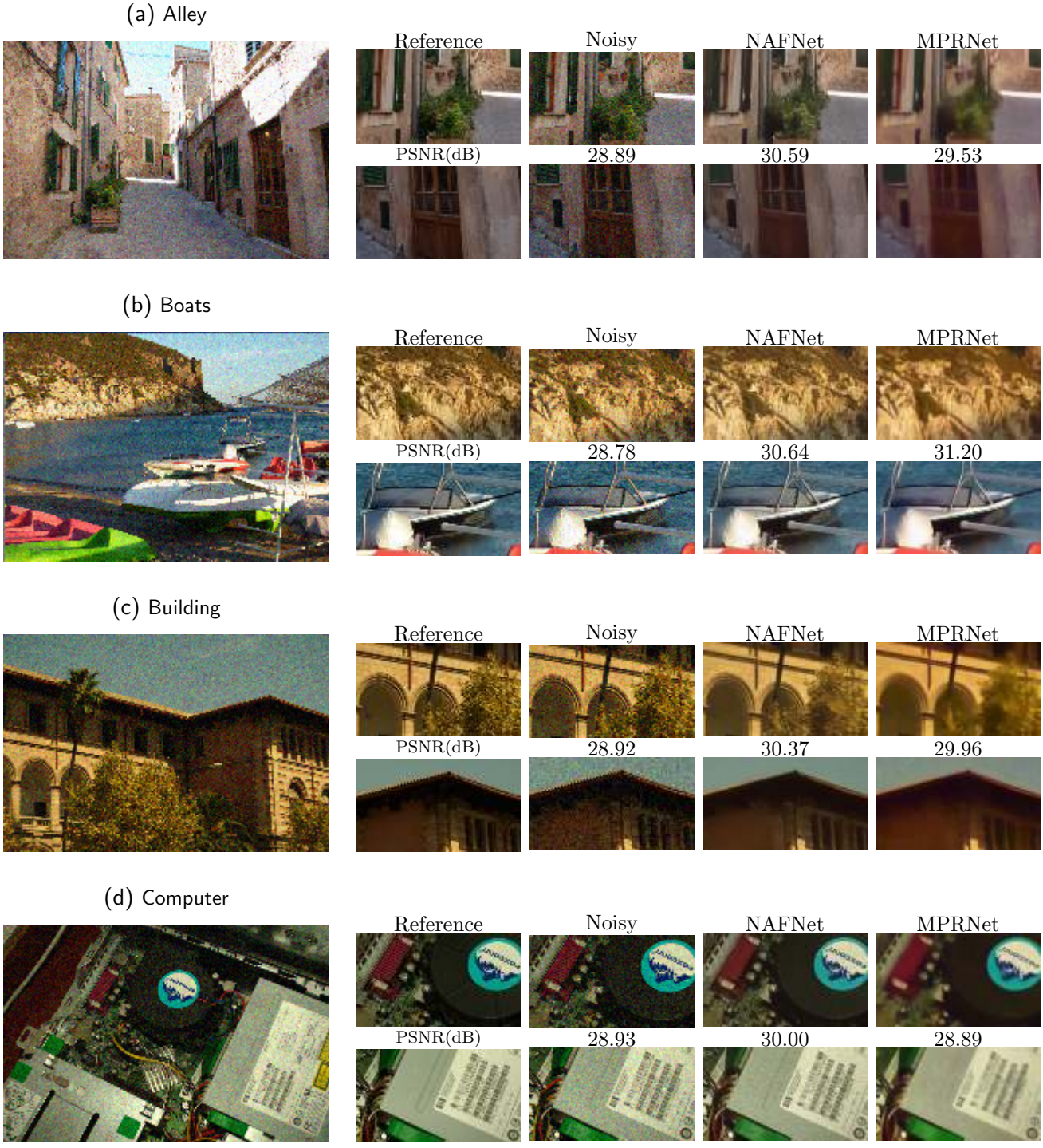


Figure 6: Visual and PSNR comparison of denoising results using NAFNet and MPRNet over four images from public dataset with Gaussian noise level $\sigma = 20$. The size of each patch is 256×144.

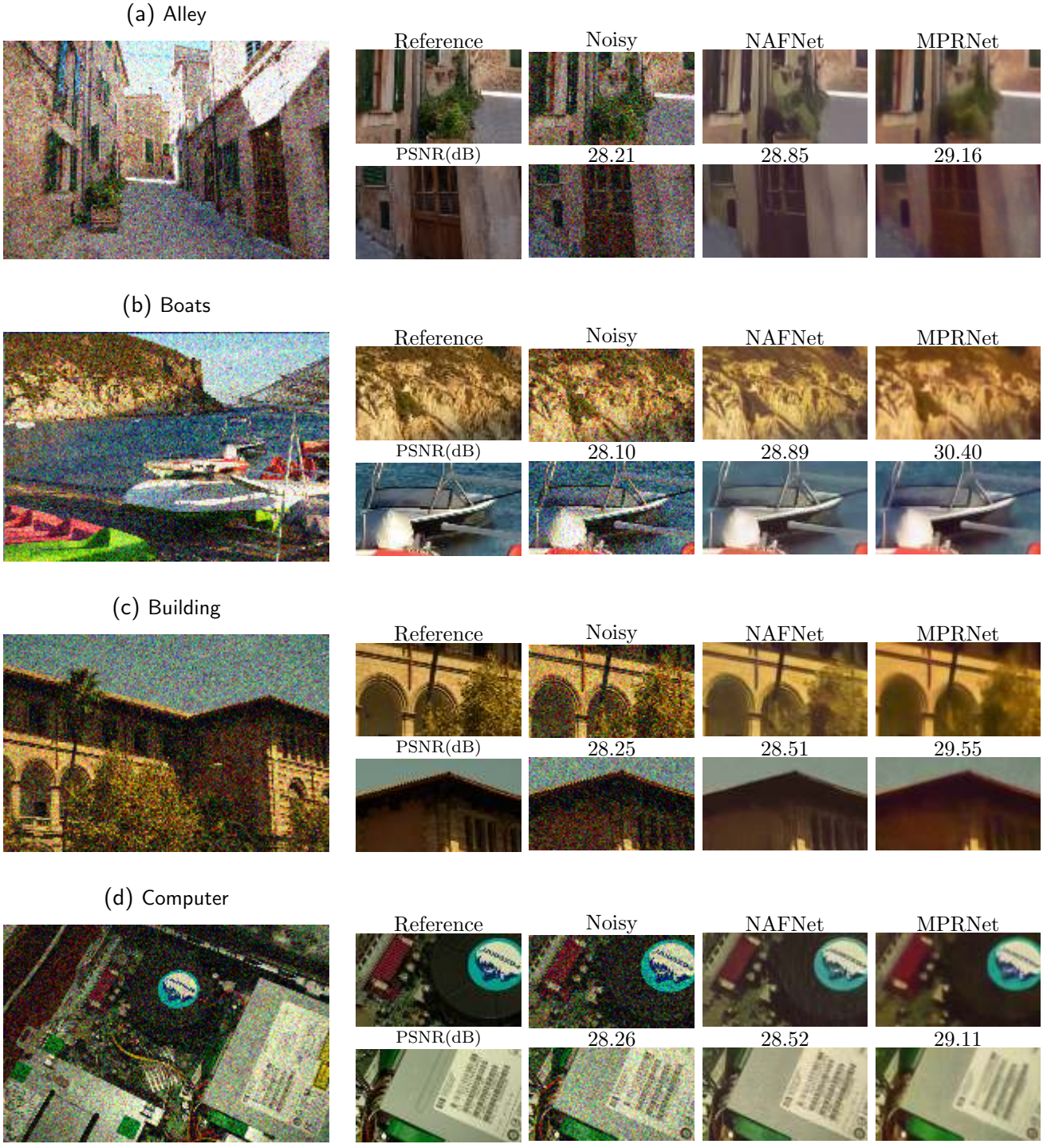
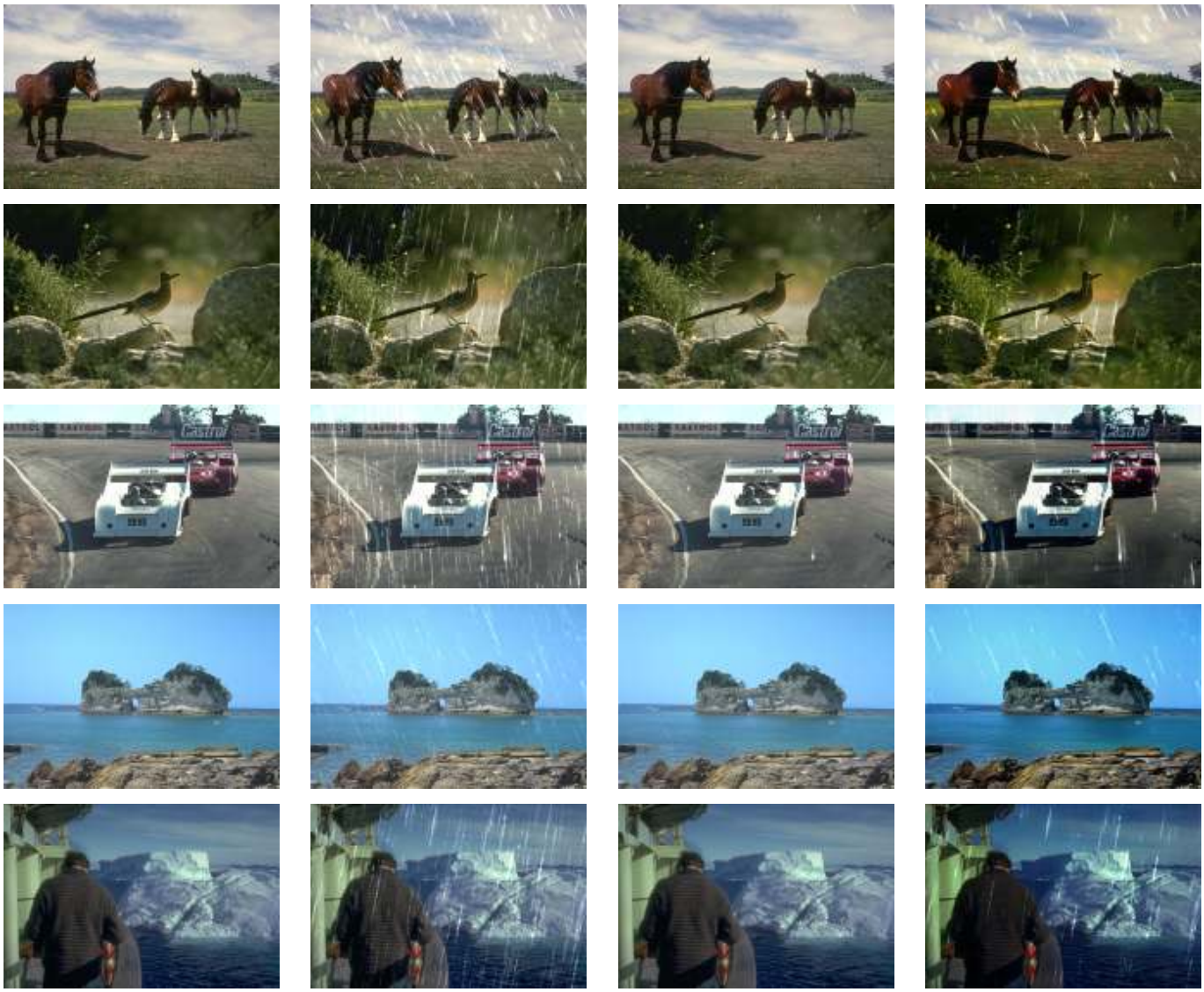


Figure 7: Visual and PSNR comparison of denoising results using NAFNet and MPRNet over four images from public dataset with Gaussian noise level $\sigma = 40$. The size of each patch is 256×144.

3.3 Deraining

For removing undesirable raining artifacts from input images, we compared the MPRNet performance with the density-aware image de-raining method using a multistream dense network (DID-MDN) method [22]. Figure 8 shows results over five synthetic rainy images from the Rainy100L dataset [19] and its test category. Both MPRNet and DID-MDN networks have been trained on the Rain100L dataset. As it is clear MPRNet removes rain artifacts better than DID-MDN in almost all images. Note that the results of MDN-DID were calculated using an online demo[6] of the original source code[7].



(a) Reference (b) Rainy (c) MPRNet (d) DID-MDN

Figure 8: Visual comparison of deraining results using MPRNet and DID-MDN method over five rainy images from the Rainy100L dataset [19]. The size of the images is 480×320.

[6] https://saiwa.ai/app/free/base-service/deraining
[7] https://github.com/hezhangsprinter/DID-MDN

4 Conclusion

In this paper, we investigated MPRNet method for image restoration with three functionalities, i.e. deblurring, denoising and deraining. The method is fully blind and works well without any knowledge of input artifact parameters like: blur kernel type and size or noise standard deviation. The MPRNet performance seems inferior to other restoration methods that we employed in this paper.

Acknowledgment

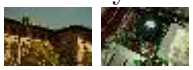
The authors would like to thank Prof. Jean-Michel Morel for his valuable comments on an early draft of this paper. This work has been fully supported by Saiwa company, Ontario, Canada.

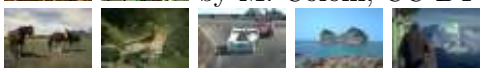
Image Credits

 from GoPro dataset [10]

 by A. Buades, CC-BY

 by J.-M. Morel CC-BY

 by M. Colom, CC-BY

 from Rain100L dataset [19]

References

[1] A. Abdelhamed, S. Lin, and M. S. Brown, *A High-Quality Denoising Dataset for Smartphone Cameras*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018, https://doi.org/10.1109/CVPR.2018.00182.

[2] J. Anger, G. Facciolo, and M. Delbracio, *Blind Image Deblurring Using the l0 Gradient Prior*, Image Processing On Line, 9 (2019), pp. 124–142, https://doi.org/10.5201/ipol.2019.243.

[3] L. Chen, X. Chu, X. Zhang, and J. Sun, *Simple Baselines for Image Restoration*, in European Conference on Computer Vision (ECCV), Springer, 2022, pp. 17–33, https://doi.org/10.1007/978-3-031-20071-7_2.

[4] S. Cho and S. Lee, *Fast Motion Deblurring*, in ACM SIGGRAPH Asia, 2009, pp. 1–8, https://doi.org/10.1145/1618452.1618491.

[5] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, *Rethinking Coarse-To-Fine Approach in Single Image Deblurring*, in IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 4641–4650, https://doi.org/10.1109/ICCV48922.2021.00460.

[6] F. Feschet, *Implementation of a Denoising Algorithm Based on High-Order Singular Value Decomposition of Tensors*, Image Processing On Line, 9 (2019), pp. 158–182, https://doi.org/10.5201/ipol.2019.226.

[7] B. Goyal, A. Dogra, S. Agrawal, B. S. Sohi, and A. Sharma, *Image Denoising Review: From Classical to State-Of-The-Art Approaches*, Information Fusion, 55 (2020), pp. 220–244, https://doi.org/10.1016/j.inffus.2019.09.003.

[8] S. Hurault, T. Ehret, and P. Arias, *EPLL: An Image Denoising Method Using a Gaussian Mixture Model Learned on a Large Set of Patches*, Image Processing On Line, 8 (2018), pp. 465–489, https://doi.org/10.5201/ipol.2018.242.

[9] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, *Recurrent Squeeze-And-Excitation Context Aggregation Net for Single Image Deraining*, in European Conference on Computer Vision (ECCV), 2018, pp. 254–269, https://doi.org/10.1007/978-3-030-01234-2_16.

[10] S. Nah, T. H. Kim, and K. M. Lee, *Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, https://doi.org/10.1109/CVPR.2017.35.

[11] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, *Deblurring Text Images Via L0-Regularized Intensity and Gradient Prior*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2901–2908, https://doi.org/10.1109/CVPR.2014.371.

[12] N. Pierazzo, J.-M. Morel, and G. Facciolo, *Multi-Scale DCT Denoising*, Image Processing On Line, 7 (2017), pp. 288–308, https://doi.org/10.5201/ipol.2017.201.

[13] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, *Progressive Image Deraining Networks: A Better and Simpler Baseline*, in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3937–3946, https://doi.org/10.1109/CVPR.2019.00406.

[14] J. Rim, H. Lee, J. Won, and S. Cho, *Real-World Blur Dataset for Learning and Benchmarking Deblurring Algorithms*, in European Conference on Computer Vision (ECCV), 2020, https://doi.org/10.1007/978-3-030-58595-2_12.

[15] 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, https://doi.org/10.1007/978-3-319-24574-4_28.

[16] M. Suin, K. Purohit, and A. Rajagopalan, *Spatially-Attentive Patch-Hierarchical Network for Adaptive Motion Deblurring*, in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3606–3615, https://doi.org/10.1109/CVPR42600.2020.00366.

[17] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, *Scale-Recurrent Network for Deep Image Deblurring*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8174–8182, https://doi.org/10.1109/CVPR.2018.00853.

[18] M. Tassano, J. Delon, and T. Veit, *An Analysis and Implementation of the FFDNet Image Denoising Method*, Image Processing On Line, 9 (2019), pp. 1–25, https://doi.org/10.5201/ipol.2019.231.

[19] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, *Deep Joint Rain Detection and Removal from a Single Image*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1357–1366, https://doi.org/10.1109/CVPR.2017.183.

[20] G. Yu, G. Sapiro, and S. Mallat, *Solving Inverse Problems with Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity*, IEEE Transactions on Image Processing, 21 (2011), pp. 2481–2499. https://doi.org/10.1109/TIP.2011.2176743.

[21] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, *Multi-Stage Progressive Image Restoration*, in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 14821–14831, https://doi.org/10.1109/CVPR46437.2021.01458.

[22] H. Zhang and V. M. Patel, *Density-Aware Single Image De-Raining Using a Multi-Stream Dense Network*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 695–704, https://doi.org/10.1109/CVPR.2018.00079.

[23] S. Zhang, A. Zhen, and R. L. Stevenson, *Deep Motion Blur Removal Using Noisy/blurry Image Pairs*, Journal of Electronic Imaging, 30 (2021), p. 033022, https://doi.org/10.1117/1.JEI.30.3.033022.