# BURST DENOISING WITH GLOBAL-PATCH-RNN MODEL

[1] *Jyun-Ruei Wong* (翁浚瑞) *Chiou-Shann Fuh* (傅楸善)

[1] Graduate Institute of Networking and Multimedia,
National Taiwan University, Taipei, Taiwan,
E-mail: r07944028@ntu.edu.tw fuh@csie.ntu.edu.tw

**ABSTRACT**

Noise is an inevitable issue of nowadays photography technique, especially for mobile devices and low-light conditions. To address such problems, many computational photography techniques have been developed. While many of them focus on single-image denoising, recently the use of burst images is more and more popular. To exploit burst images, it means we should not only focus on the same aligned position information but also similar patches in the single image itself and across other patches in other images. We show that we can improve current burst denoising techniques by a global-patch-RNN model.

***Keywords:*** *Denoising, Burst Images, RNN*

## 1. INTRODUCTION

Denoising is an ongoing important topic. The noises come from many sources, such as hardware limitations, lacks of lights, and poor weather conditions. To deal with some of noise sources, we can try to address them from the point of view of hardware. For example, use a longer exposure time or more sensitive sensors. On the other hand, we can rely on computational photography such as state-of-the-art deep learning techniques to learn a denoising model.

Hardware and software solutions are all feasible. Burst image denoising is good mixture. By taking multiple images in a single shot, each image has better quality than a longer exposure time image, since the image is less influenced by the motion blur or handshake. After that, computational photography takes the role as mixing the images into a single denoised image.

In this paper, we exploit a supervised deep learning technique to learn a model by taking burst image with noise as input and a reference clean image as output. Traditional convolutional neural network has disadvantage of limited receptive fields. To deal with this problem, in each layer we feed all the patches into a RNN network, which we call global-patch-RNN, so that the model is able to learn from further-away patches. Global-patch-RNN acts like an attention model concept from natural image processing sequence-to-sequence model. If we want to learn every patch from all other patches, the whole model will look like a 2D fully-connected perceptron, which is not preferable for efficiency. As a result, we do not use the concept of attention model directly, while we use a shared RNN structure to encapsulate information from all same-sized patches.

Next, to better exploit the property of burst images, we use another RNN structure from [1] to learn relationships between multiple images. Thus our model comprises two different RNN structures, one for maintaining local information, both spatial and time domain neighborhood, and one for maintaining global information, both spatial and time domain whole image patches.

## 2. RELATED WORK

Traditional simple image denoising has remained as an ongoing studying topic for a long time, many methods such as Non-Local Means [2] and anisotropic diffusion [3] have been proposed. Moreover, Dabov, et al. bring up with a 3D filtering algorithm, BM3D [4], one of the most well-known denoising techniques.

Single image denoising methods based on machine learning then prevailed in the field. Methods including [5, 6, 8] have both found outstanding effect. Gharbi, et al. [7] found that denoising together with demosaicking can do a better job in both problems. Lehtinen, et al. [14] use a noise2noise model to learn without clean images, which is meaningful that we can directly train our model with noisy data.

Traditional multiple images variants focused on video denoising [9, 12, 13] or burst image denoising [10]. To work on limited devices such as mobiles, efficient algorithms [11] are proposed. Some variants of BM3D [12, 13] that exploit patches across and within images works quite well.

Learning-based multiple images methods often use RNN [1] to mix information across images. Or it is

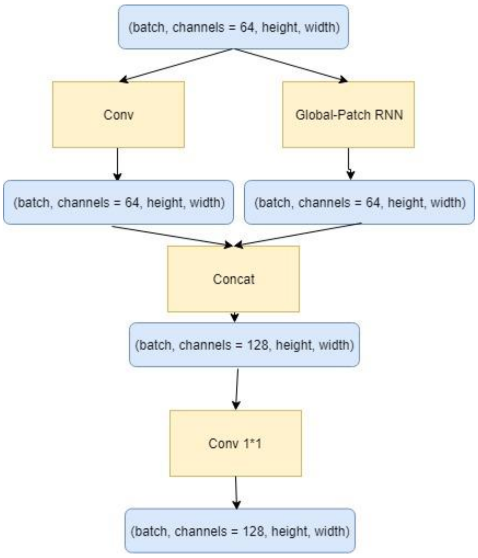


Fig. 1: The connection of our Global-Patch RNN with tradional CNN model. Note that it is just a rough structure of the architecture. The batch normalization and activation function are ignored in this figure.

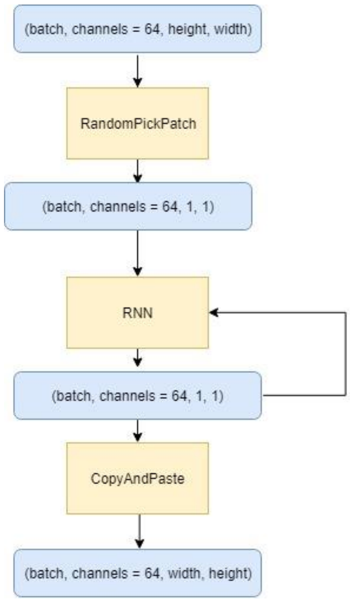


Fig. 2: Global-Patch RNN architecture. Note that the input size and the output size are the same so that it can be easily fit into CNN layers.

sometimes enough using a CNN [15, 16]. This area is still quite new that many researches are still needed.

## 3. METHOD

The convolution layers often have limited receptive fields. By adding RNN to learn patches, the model may try to learn the information from further-away patches.

Our method aims to learn a double RNN structure model to get both information from across images and inside images, both local and global information. The structure is derived from the work of Godard, et al. [1].

The course network goes through an RNN network which is fed in with the noisy burst images. Every image goes through a fine network to learn a single image denoising model. After the model is stable enough, another part of the model uses RNN to exploit the information learned from the single image denoising model to further learn across-image information.

The fine network tries to learn global patch information from the image itself. It is a spatial-RNN-fashioned network. While instead of learning in a directional way like bi-directional RNN, we try to eliminate the importance of distance between patches by feeding the patches into the network randomly. Thus RNN network cannot know which patch is close to another. In each layer, this kind of random patch RNN goes through once. Moreover, the information is then together mixed with CNN network to predict the pixel. Thus, we separate the job of CNN and RNN. CNN is used to learn local information; on the other hand, RNN is used to learn global information.

The difference between our Global-Patch RNN with other spatial-RNN networks is that we try to mix the information from all patches and use it to help CNN predict next layer. First, the RNN network only traverses the patches once, unlike spatial-RNN such as Bidirectional-RNN which traverses them twice. Next, RNN network even with the help of attention model still suffers from the pain of ignoring further-away patches, which is somehow like the limited receptive field problem we face in CNN network. Our method alleviates this problem by randomly feeding the patches so that the model has to learn to pay attention to every patch more evenly. Figure 1 shows the connection of our Global-Patch RNN and the traditional CNN architecture. The information from Global-Patch RNN and traditional CNN layers is further concatenated and conv1-1 back to tensors with the same channel size. Figure 2 shows the part of the global-patch-RNN layers in our model. RandomPick layers randomly but thoroughly select patches to feed into RNN layers. The outcome of the RNN layers is further duplicated to match the size of an image. Thus each patch can then gain this information by conv1-1 with it.

## 4. IMPLEMENTATION

We use subsets of High Dynamic Range (HDR+) dataset as our dataset for training, validation and testing, which contains 153 bursts. It is noted that each burst may contain different number of images, in a range of 6 to 12 images.

We first use an image pipeline to all raw DNG files to TIFF files format. To preprocess the image, we resize the images, so that the shorter edge is 512 pixels. Then we crop the images to 64*64 pixels for training. Moreover, to add noise, we use Gaussian distribution with standard deviation 0.1 for noise model.

We use Pytorch to build our network. To construct our network, first, we use an RNN network outside to feed multiple images. Second, inside the network is another RNN network for learning global patches. The patches are fed in randomly, so that it can learn the information from other positions from the image.

During training, we at first only train the part of single image denoising network. After the network is stable enough, we then train all the network with bursts, it is found that it is more stable to derive burst image denoising pipeline.

## 5. RESULT

We compare single image denoising with burst image denoising. The test is done on the full size of burst images. Table 1 shows the Peak Signal-to-Noise Ratio (PSNR) of different models. Figure 3 shows some result of the bursts denoising.

| method | PSNR |
|---|---|
| single image denoising | 22.97 |
| burst image denoising | 28.69 |

Table 1: PSNR of single image denoising and burst image denoising.

## 6. LIMITATION AND FUTURE WORKS

Our model can improve the detail of some images, but it is still not stable enough. We believe it is because of the nature of RNN still cannot encapsulate that much information from all across-image patches.

Moreover, feeding too much information into RNN network may cause the network to try to compromise sharpness for accuracy. We may try to give up on end-to-end pipeline by clustering the patches first and selectively feeding into the RNN structures. The idea of Point net from point cloud problems may also be an alternative for RNN structures.

## 7. CONCLUSION

Our Global-Patch RNN model is constructed of two domains of RNN model: one for across images in a burst, and one for across patches of all images in a bursts. We show that it brings more information for the traditional CNN model for burst denoising. Moreover, it is more efficient than most spatial-RNN-fashioned network.

## REFERENCES


[1] C. Godard, K. Matzen, and M. Uyttendaele, "Deep Burst Denoising," https://arxiv.org/abs/1712.05790, 2017.

[2] A. Buades, B. Coll, and J. M. Morel, "A Non-Local Algorithm for Image Denoising," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, Vol. 2, pp. 60–65, 2005.

[3] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, Issue 7, pp. 629–639, 1990.

[4] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Bm3d image denoising with shape-adaptive principal component analysis. In: SPARS'09-Signal Processing with Adaptive Sparse Structured Representations (2009)

[5] Roth, S., Black, M.J.: Fields of experts: A framework for learning image priors. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 2, pp. 860–867. IEEE (2005)

[6] Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep cnn denoiser prior for image restoration. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)

[7] Gharbi, M., Chaurasia, G., Paris, S., Durand, F.: Deep joint demosaicking and denoising. ACM Transactions on Graphics (TOG) 35(6), 191 (2016)

[8] Agostinelli, F., Anderson, M.R., Lee, H.: Adaptive multi-column deep neural networks with application to robust image denoising. In: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 26, pp. 1493–1501. Curran Associates, Inc. (2013)

[9] C. Liu and W. T. Freeman. A high-quality video denoising algorithm based on reliable motion estimation. ECCV, 2010.

[10] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. SIGGRAPH Asia, 2016.

[11] Z. Liu, L. Yuan, X. Tang, M. Uyttendaele, and J. Sun. Fast burst images denoising. SIGGRAPH Asia, 2014.

[12] Maggioni, M., Boracchi, G., Foi, A., Egiazarian, K.: Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms. IEEE Transactions on image processing 21(9), 3952–3966 (2012)

[13] Maggioni, M., Boracchi, G., Foi, A., Egiazarian, K.O.: Video denoising using separable 4d nonlocal spatiotemporal transforms. In: Image Processing: Algorithms and Systems. p. 787003 (2011)

[14] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, Timo Aila, Noise2Noise: Learning Image Restoration without Clean Data, ArXiv, 2018

[15] Miika Aittala, Frédo Durand, Burst Image Deblurring Using Permutation Invariant Convolutional Neural Networks, ECCV, 2018

[16] Ben Mildenhall, Jonathan T. Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, Robert Carroll, Burst Denoising with Kernel Prediction Networks, ArXiv, 2018

[17] X. Y. Xinyuan Chen, Li Song. Deep rnns for video denoising, 2016.

[18] 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, 2017.

[19] Wieschollek, P., Schölkopf, B., Lensch, H.P.A., Hirsch, M.: End-to-end learning for image burst deblurring. In: Lai, S.H., Lepetit, V., Nishino, K., Sato, Y. (eds.) Computer Vision – ACCV 2016. pp. 35–51. Springer International Publishing, Cham (2017)

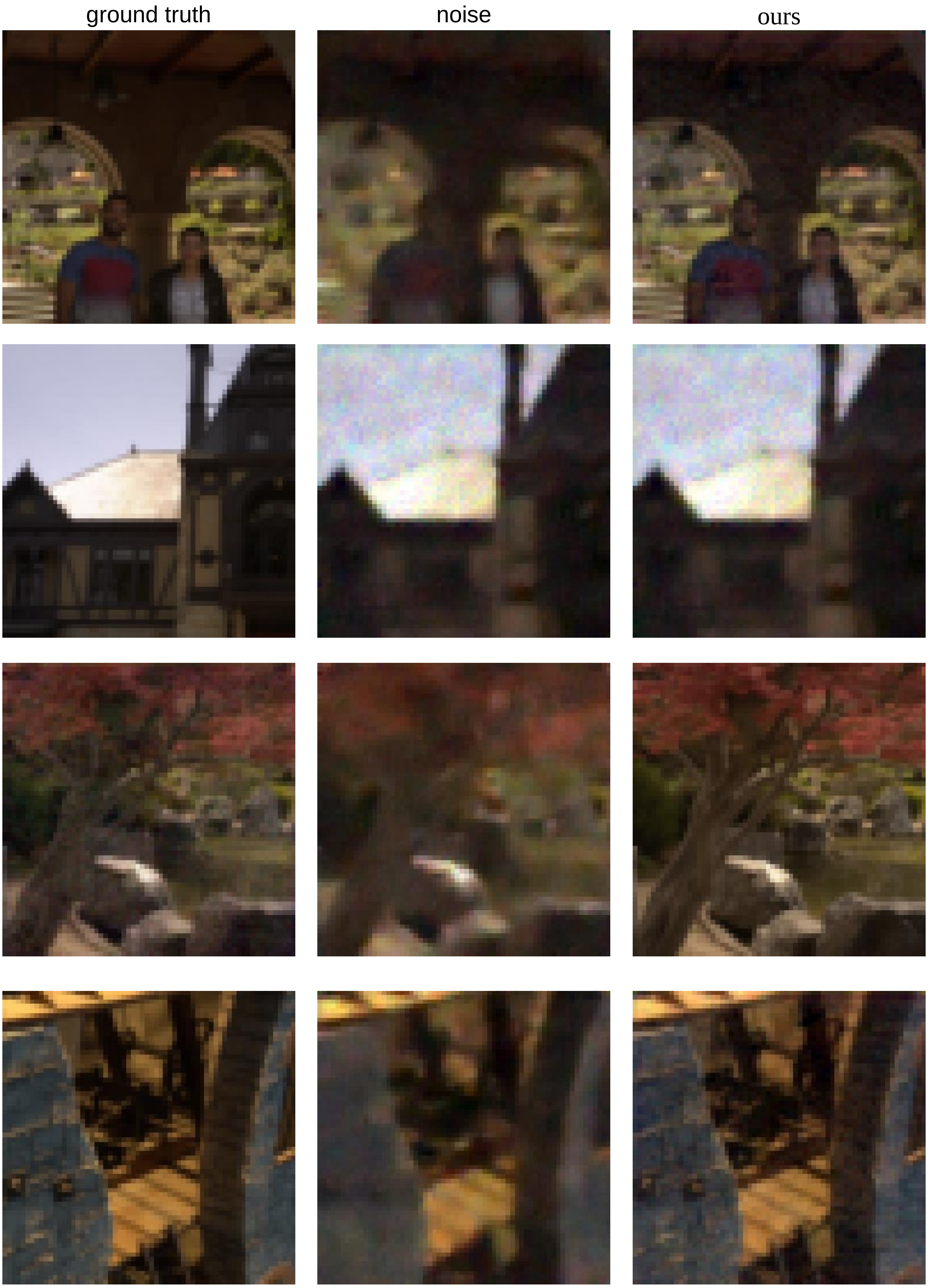

Figure 3: We choose a representative image in each burst for ground truth and noise image. Note that the model is trained on all images in bursts.