IMAGE DEBLURRING USING DEEP MULTI-SCALE DISTORTION PRIOR

Irina Kim, Dongpan Lim, Youngil Seo, Jeongguk Lee, Wooseok Choi and Seongwook Song

Samsung Electronics

ABSTRACT

Deep neural networks have recently advanced state-of-the-art in motion deblurring. However, non-uniform non-blind image deblurring has not been studied in depth. State-of-the-art methods shows improvement over conventional algorithms, but they are still not feasible for mobile deployment. Having informative prior information could improve performance of non-uniform deblurring. In this work, we propose a new deep framework that allows extracting spatially variant latent feature to Distortion Prior map from a pair of calibration sharp-blur images, without having to capture or model training dataset. We propose to use multi-scale Distortion Prior map that can fully utilize spatially variant information in the further restoration via multi-scale attention mechanism. Unlike prior art, we use image pyramid at decoder side, by fusing its fine level with coarse level of feature map via level attention and by injecting Distortion Prior at various resolution levels. Experiments show that proposed network outperforms state-of-the-art deblur networks both in terms of image quality and inference time. We demonstrate that proposed framework can successfully deblur non-uniform, non-blind applications, such as defocus blur removal. Being computationally efficient, it is feasible for mobile deployment.

Index Terms— Image deblurring, non-uniform image restoration, deep learning, defocus blur removal, under display camera

1. INTRODUCTION

Deep learning approach to image deblurring shows some improvement over conventional methods for uniform deblurring [1–4]. Xu *et al* proposed deconvolutional network to estimate blur and restore image [5]. Kupyn *et al* [6] used conditional adversarial networks. Encoder-decoder networks are quite popular in image deblurring [3, 7, 8]. Nah *et al* [7] and Tao *et al* [9] introduced a multi-scale cascade of networks using 'coarse-to-fine' approach that sequentially restores downscaled images.

Recently, Zhang *et al* advanced state-of-the-art surpassing 31.2dB on GoPro dataset by using a stack of encoder-decoder networks [10] and performing sequential restoration of small to large patches. Despite of being faster than other methods, it is not feasible for mobile deployment. Moreover, there is no significant improvement for non-uniform deblur.

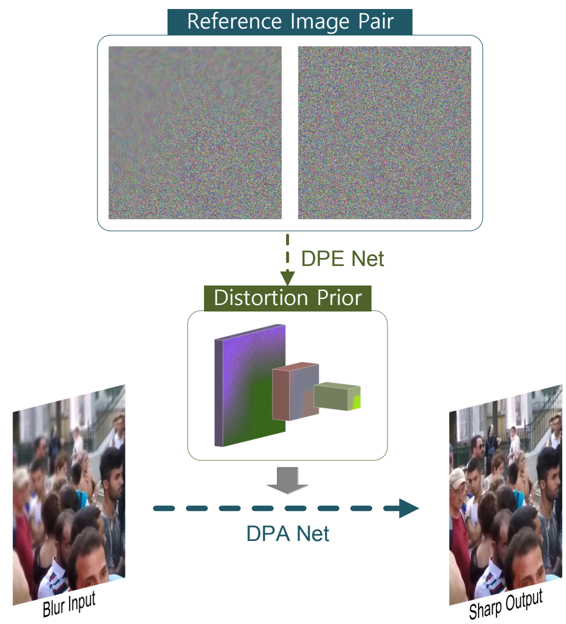

Fig. 1. Conceptual diagram of the proposed method. Based on a pair of blur and sharp reference images, we use a pair of networks to deblur input blur images: we extract spatially variant information to multi-scale Distortion Prior (DP) using DPE Net and then feed DP to main deblur network (DPA Net) via additive attention mechanism, without having spatially variant training data.

To address mentioned above issues, we introduce a deep Distortion Prior (DP) concept that allows to extract spatially varying blur information from a reference pair of sharp-blur images, without a need for a 'hard-to-get' training dataset.

Experiments show that proposed approach advances state-of-the-art exceeding 32dB on GoPro dataset, while being computationally efficient - 16 times faster than popular state-of-the-art deep deblurring method DMPHN [10], see Table 1. We also demonstrate superior results for non-blind deblurring: defocus blur removal.

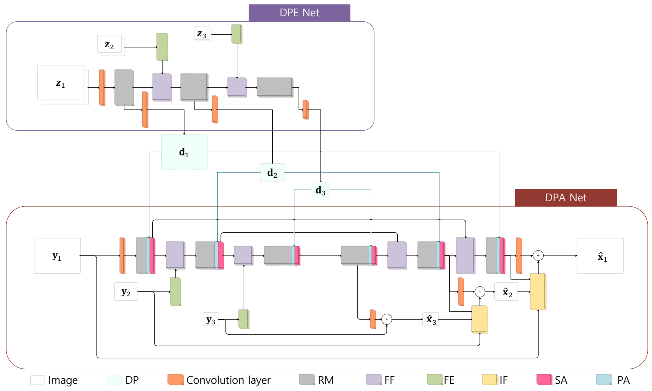

Fig. 2. An overall block diagram of the proposed architecture. Here DPE Net is Distortion Prior (DP) Estimator Network, DPA Net stands for DP Attentioned Network. Abbrevations for subBlocks are given in the caption of Fig. 3.

2. PROPOSED METHOD

In this work, we propose a novel method to solve non-uniform image deblurring. We can cover both blind and non-blind deblurring. We use linear image model below, where blur $\mathbf{K}$ varies spatially:

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mu \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^n$ - a blur observation, $\mathbf{x} \in \mathbb{R}^n$ - a reconstructed latent image, $\mathbf{K} \in \mathbb{R}^{n \times n}$ - a spatially varying blur matrix, $\mu \in \mathbb{R}^n$ - noise, $n \in \Omega$ - number of measurements.

Based on the model, we propose to estimate Distortion Prior from a pair of reference calibration images $\Psi = \{\mathbf{r}_{\text{sharp}}, \mathbf{r}_{\text{blur}}\}$:

$$\mathbf{d} = P(\mathbf{z}; \Theta_P) \;\; \forall \mathbf{z} \in \Psi \tag{2}$$

Next we propose to learn end-to-end mapping function from training pairs by using DP:

$$\hat{\mathbf{x}} = F(\mathbf{y}, \mathbf{d}; \Theta_F) \;\; \forall \mathbf{y} \in \Omega, \tag{3}$$

where Ω is training dataset.

2.1. Network Architecture

An overall architecture of the proposed framework is illustrated in Fig. 2. We have two networks in the framework: an auxiliary network that estimate DP as learnt feature map, namely DP Estimation (DPE) Network, and main restoration network, namely DPA Net. In this work we adopt encoder-decoder architecture for DPA Net and encoder architecture for DPE Net. Unlike previous state-of-the-art using a cascade of networks for main network, we use single hierarchical network inspired by Duplex Pyramid Network (DPN) [11].

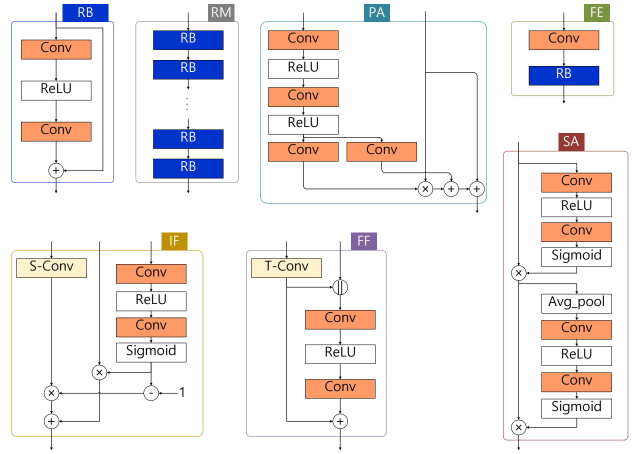

Fig. 3. SubBlocks diagram: RB - Residual Block, RM - Residual Module, PA - Prior Attention block, FE - Feature Extraction module, IF - Image Fusion, FF - Feature Fusion, SA- Self Attention block.

DPE Net generates a multi-scale distortion prior from a pair of sharp-blur image, then DPA Net restores sharp image $\mathbf{x}$ from a single blur observation $\mathbf{y}$ and $\mathbf{d}$. We use local residual learning in Residual Block (RB) depicted in Fig. 3. RB was inspired from SRResNet [12], with removed batch normalization.

We used RB as a building block of Residual Module (RM): in DPA Net, we used $K = 10$ RBs. Output of each Residual Module is connected to Prior Attention (PA) and Feature Fusion (FF) through Self-Attention (SA) block. The PA block merges the features received from DP and RM. In contrast to DPN [11], we build image pyramid by using bilinear interpolation and feed it to DPA Net via level attention Image Fusion (IF) block. The features of the input and downscaled images are extracted through Feature Extractor (FE) block. The FE block consists of one convolution and one RB as shown in the Fig. 3.

Distortion Prior Estimation (DPE) Net is generating multi-scale DP. We implemented DP estimator through three levels encoder, resembling encoder of DPA Net. For DP estimation, SA is removed and convolutional layer for extracting DP is added to the output of each RM. For blind deblurring, we use blur image for input; for non-blind - both sharp and blur images.

2.2. Attention Mechanism

Following transformer architecture [13] in natural language processing, attention is used in many image restoration tasks, including super-resolution, image deblurring, image dehazing [14,15]. Inspired by its sucess, we use three types of attention in our work: Self-Attention (SA), Level Attention (LA)

Fig. 4. Visual image quality evaluation on GoPro dataset: from the top to bottom we show blur input and deblur results of DeblurGAN [6], SRN [9], DMPHN [10] and proposed method. We can see that DeblurGAN results are very blur, SRN and DMPHN failed to restore faces, text or details (see billboard or car door).

and DP Attention (DPA). Self-attention module uses a cascade of channel attention and pixel attention and is depicted on Fig. 3 in SA block. SA can handle different blur information along pixel region and feature maps [16]. Channel attention models channel correlations and gives higher weights to important channels, to introduce global contextual information across channels. Accordingly, features can be emphasized in the encoding and decoding blocks.

$$\mathbf{G} = \text{Pool}(\mathbf{F}^*) \tag{4}$$

$$\mathbf{A} = \sigma(\text{Conv}(\delta(\text{Conv}(\mathbf{G}))) \tag{5}$$

$$\mathbf{F} = \mathbf{F}^* \odot \mathbf{A} \tag{6}$$

where, Pool - average pooling, σ - sigmoid, δ - ReLU, $A \in \mathbb{R}^C$ - channel attention, C - number of channel in the corresponding feature map $\mathbf{F}^*$.

Blur information and image characteristics are also spatially varying. To exploit this information and emphasize local pixels adaptively, we adopted pixel attention defined in Eq. 7.

$$\mathbf{A} = \sigma(\text{Conv}(\delta(\text{Conv}(\mathbf{F}^*))) \tag{7}$$

$$\mathbf{F} = \mathbf{F}^* \odot \mathbf{A} \tag{8}$$

where $\mathbf{A} \in \mathbb{R}^{H \times W}$ - pixel attention, W and H are image width and height, respectively. We fuse reconstructed image at coarse level $x^{\widehat{i+1}}$ with fine level of image pyramid y^i. We emphasized by attention block applied to corresponding feature map. The block diagram of Image Fusion (IF) block is provided in Fig. 3.

$$\mathbf{A} = \sigma(\text{Conv}(\delta(\text{Conv}(\mathbf{F}^i))) \tag{9}$$

Method	PSNR[dB]	Time[ms]
Sun [18]	24.64	20,000
DeepDeblur [7]	29.23	4,330
DeblurGAN [6]	28.7	850
AttentionNet [16]	28.81	1,300
SV-RNN [19]	29.19	930
SRN [9]	30.26	1,870
PSS-NSC [8]	30.92	1,600
DMPHN [10]	31.2	424
SA-PHN [20]	31.85	340
Proposed	**32.05**	**26**

Table 1. Performance on the GOPRO dataset [7].

$$\mathbf{x'^i} = \text{S-Conv}_{\uparrow}(\mathbf{x^{i+1}}) \tag{10}$$

$$\tilde{\mathbf{x}}^{\mathbf{i}} = \mathbf{x'^i} \odot \mathbf{A} + \mathbf{y^i} \odot (\mathbf{1} - \mathbf{A}) \tag{11}$$

where $\mathbf{A} \in \mathbb{R}^{H \times W}$ - level attention, $\mathbf{F}^i$ - feature map at i_{th} level, $\mathbf{y}^i$ - input image pyramid at i_{th} level, $\mathbf{x}^{(i+1)}$ - image pyramid at $(i+1)_{th}$ level, $\mathbf{x'}^i$ - upscaled $\mathbf{x}^{(i+1)}$ by transposed convolution with stride 2. We first apply attention to distortion prior $\mathbf{d}$ at each scale, block diagram is shown in PA block in Fig. 3. Inspired by Kauffman *et al* [17], we apply additive attention (15).

$$\mathbf{A} = \delta(\text{Conv}(\delta(\text{Conv}(\mathbf{d})))) \tag{12}$$

$$\mathbf{A}' = \text{Conv}(\mathbf{A}) \tag{13}$$

$$\mathbf{A}'' = \text{Conv}(\mathbf{A}) \tag{14}$$

$$\mathbf{F} = \mathbf{F}^* + \mathbf{F}^* \odot \mathbf{A}' + \mathbf{A}'' \tag{15}$$

where $\mathbf{F}^*$ and $\mathbf{F}$ - input and output feature map, respectively, $\mathbf{A} \in \mathbb{R}^{C \times H \times W}$ - prior attention.

3. EXPERIMENTS

In this work we used popular GoPro dataset [7] for training and tests. For non-blind deblurring, we used non-uniform lens distortion model:

$$\mathbf{x}' = \mathbf{R}^{-1}\mathbf{T}\mathbf{R}\mathbf{x} \tag{16}$$

where $\mathbf{R}$ is a rotation matrix with random angle θ, $\mathbf{T}$ is a random diagonal scaling matrix.

We provide objective image quality results in PSNR on the GoPro dataset and inference time required to process a single 720p HD image (1280x720) measured on NVidia TitanX GPU in the Table 1. From the Table 1, it is clear that proposed method outperforms both conventional and deep learning based state-of-the-art methods, while being computationally efficient. It is an order magnitude faster than previous fastest method [10] and three orders of magnitude faster than reference non-uniform deblur method [18]. In Fig. 4, proposed method shows better subjective image quality for GoPro dataset compared to other methods.

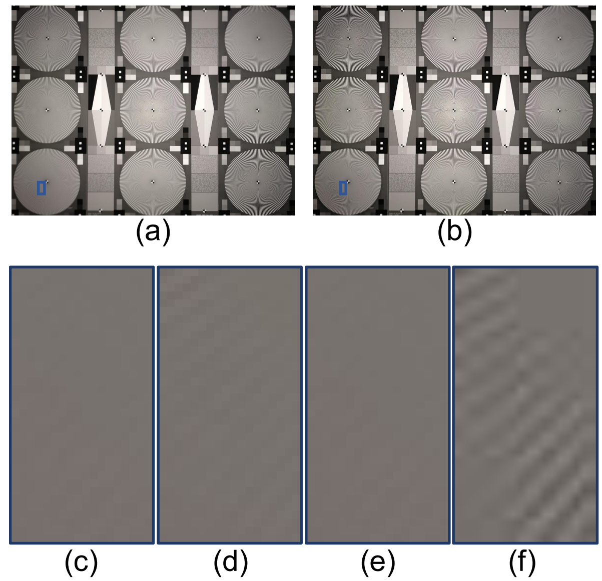

Fig. 5. Real Defocus example: (a) - input image; (b) - output of proposed method; (c) - magnified input patch; (d) - output of DPA Net using input concatenated with reference calibration image pair, without using Distortion Prior (DP); (e) - output using uniform DP; (f) - proposed output using non-uniform DP.

3.1. Defocus blur removal

We applied proposed method to defocus blur removal application in this section. Tests were done on real images captured by Samsung 64MP sensor. As you can see from the Fig. 5, peripheral part of the Star chart mage, especially on the right top and left bottom, did not contain any details. If we do not utilize spatially variant Deep Distortion Prior, we end up with the loss of resolution at peripheral parts of image, as shown in the Fig. 5 (d). However, once we apply proposed non-unform Distortion Prior, we can restore image at higher resolution at peripheral parts as you can see on the most right image (f), while keeping same quality for the central part of the image, see (c) and (e).

4. CONCLUSION

In this work, we proposed a new approach for spatially varying image deblurring using learnt multi-scale distortion prior, without implicit modeling of non-uniform blur. Experiments show that our approach outperforms state-of-the-art methods in both objective and subjective image quality, while being orders of magnitude faster. We also demonstrated superior image quality for non-blind non-uniform defocus blur removal. Proposed method can be applied to other non-uniform image restoration problems, such as under display camera.

5. REFERENCES

[1] Ruxin Wang and D. Tao, "Recent progress in image deblurring," *ArXiv*, vol. abs/1409.6838, 2014. 1

[2] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1964–1971. 1

[3] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang, "Deep video deblurring for hand-held cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1

[4] Ning Xu, Brian Price, Scott Cohen, and Thomas Huang, "Deep image matting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1

[5] Li Xu, Jimmy S. J. Ren, Ce Liu, and Jiaya Jia, "Deep convolutional neural network for image deconvolution," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1*, Cambridge, MA, USA, 2014, NIPS'14, p. 1790–1798, MIT Press. 1

[6] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1, 3, 4

[7] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1, 4

[8] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia, "Dynamic scene deblurring with parameter selective sharing and nested skip connections," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 4

[9] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia, "Scale-recurrent network for deep image deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1, 3, 4

[10] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 3, 4

[11] Irina Kim, Seongwook Song, Soonkeun Chang, Sukhwan Lim, and Kai Guo, "Deep image demosaicing for submicron image sensors," *Journal of Imaging Science and Technology (JIST)*, vol. 63(6), pp. 060410–1–060410–12, 11 2019. 2

[12] 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 *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 105–114. 2

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008. 2

[14] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang, "Non-local recurrent network for image restoration," in *Advances in Neural Information Processing Systems*, 2018, pp. 1673–1682. 2

[15] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran, "Image transformer," *arXiv preprint arXiv:1802.05751*, 2018. 2

[16] Qing Qi, Jichang Guo, and Weipei Jin, "Attention network for non-uniform deblurring," *IEEE Access*, 2020. 3, 4

[17] Adam Kaufman and Raanan Fattal, "Deblurring using analysis-synthesis networks pair," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 4

[18] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," *CoRR*, vol. abs/1503.00593, 2015. 4

[19] Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson W.H. Lau, and Ming-Hsuan Yang, "Dynamic scene deblurring using spatially variant recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 4

[20] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan, "Spatially-attentive patch-hierarchical network for adaptive motion deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3606–3615. 4