SGENet: Spatial Guided Enhancement Network for Image Motion Deblurring

Yu-Chieh Wang[1]
jacky.unbox@gmail.com
Chia-Hung Yeh[1,2]
chyeh@ntnu.edu.tw

[1] Department of Electrical Engineering, National Taiwan Normal University Taipei, Taiwan

[2] Department of Electrical Engineering, National Sun Yat-sen University Taipei, Taiwan

Abstract

Multi-stage architectures have been widely used for image motion deblurring and achieved significant performance. Previous methods restore the blurred image by obtaining the spatial details of the blurred input image. However, the blurred image cannot provide accurate high-frequency details, degrading the overall deblurring performance. To address this issue, we propose a novel dual-stage architecture that can fully extract the high-frequency information of the blurred images for reconstructing detailed textures. Specifically, we introduce a supervised guidance mechanism that provide precise spatial details to recalibrate the multi-scale features. Furthermore, an attention-based feature aggregator is proposed to adaptively fuse influential features from different stages in order to suppress redundant information from the earlier stage passing through to the next stage, allowing efficient multi-stage architecture design. Extensive experiments on GoPro and HIDE benchmark datasets show the proposed network has the state-of-the-art deblurring performance with low computational complexity when compared to the existing methods.

1 Introduction

Image deblurring aims at recover a sharp image from the blurred one with the necessary texture structure and high-frequency details [7], which is caused by camera shake, freely moving objects or defocus, resulting in visual discomfort and degraded image quality. Therefore, deblurring is an essential step and widely gets attention in the field of image processing, computer vision, pattern recognition, and etc.

Blind motion deblurring is a highly ill-posed problem with infinite feasible solutions. Most traditional methods [5, 10, 18, 26] use mathematical models or empirical observations and then manually design image priors to make images sharp. However, designing such priors does not generalize to real-world images of different scenes. To address this problem, recent deep-learning based methods have achieved great performance by directly learning the complex relationships between blurred and sharp images from large-scale data.

Single image deblurring is a position-sensitive task that requires pixel-to-pixel correspondence between blurred and sharp images. As a result, it is challenging to remove unwanted

degraded image content while preserving the natural edges and detailed texture. Existing CNN-based methods [1, 12, 20, 24, 31, 32] usually employ either high-resolution extraction pipelines or encoder-decoder sub-networks to increase the overall performance. In fact, previous researches employing high-resolution extraction pipelines [24, 31, 32] yield more accurate spatial details since there are no downsampling operations. However, such pipelines are ineffective in gathering contextual information due to the limited receptive fields. Moreover, the encoder-decoder architecture [1, 12, 20] utilizes a top-down and bottom-up manner to gradually map the input from high resolution to low resolution and then apply a reverse mapping to the original resolution. These methods can learn broad contextual information, but fine spatial details may also be lost, degrading the quality of restoration. Therefore, various variant architectures are introduced to restore the blurred image, and they can be roughly divided into three categories: multi-scale, multi-temporal and multi-patch architectures. Specifically, Nah *et al*. [15] proposed a multi-scale architecture to map blurry images to its sharp counterparts without estimating blur kernels. However, it is difficult to recover multi-scale information of blurry images due to the lack of receptive field. Park *et al*. [19] proposed an encoder-decoder network with the iterative strategy that can capture the non-uniform blur of the blurred input image repeatedly. However, the over iterations result in longer inference time, and the blurred input image cannot provide enough high-frequency information for reconstructing output with accurately detailed texture. Consequently, Zamir *et al*. [27] proposed a multi-stage architecture that leverages the advantages of the encoder-decoder and high-resolution extraction pipeline to learn spatially accurate and contextual-enriched features. Moreover, the method adopted a multi-patch strategy to obtain more details of the blurred input image. Despite its effectiveness, obtaining the image content of each patch through high-resolution extraction pipeline inevitably increases computational complexity and memory load, which makes it difficult to be applied to time-sensitive scenarios.

To address the aforementioned issues, we check again the bottleneck of multi-stage architectures and propose a novel deblurring architecture called Spatial Guided Enhancement Network (SGENet). We first explore the high-frequency information required for restoring the image, and found that the predicted images can provide high-frequency components, which is similar to the sharp images under the constraint of the loss function. Therefore, we focus on reconstructing detailed textures. Unlike the existing methods [4, 8, 28], which tried to obtain fine details by preserving the blurred image content, we present a practical contextual-aware enhancement module (CEM) with a supervised guidance mechanism that can calibrate semantic features through spatially-accurate predicted images to generate the detailed texture of output images. Also, other methods [8, 27] adopt a multi-stage architecture to decompose the recovery process into several subtasks, allowing each stage to learn only the valuable information for that stage. These methods try to aggregate useful features by simple summation or concatenation across stages; however, such kind of simple feature exchange causes redundant information to be erroneously fused, affecting the overall restoration performance. Inspired by [13], we develop a cross-stage selective aggregation module (CSAM), which can adjust the receptive field with consideration of incoming features from different stages, thereby suppressing the redundant information and only passing useful semantic features to enrich the features in the next stage.

The main contributions of the proposed method are in three aspects: First, we present a novel attention-based supervised guided mechanism that can fully exploit the high-frequency information from the predicted image for precisely generating outputs with detailed texture. Second, a new cross-stage weight adjustment strategy is introduced to adaptively select and

reuse features from different stages. This strategy only allows valuable features to pass to the next stage, ensuring a smooth information exchange for constructing an efficient multi-stage architecture design. Third, we provide extensive analysis and evaluations on dynamic scene deblurring benchmarks, demonstrating that our method produces state-of-the-art results while maintaining low time complexity. In addition, we provide detailed ablation studies, qualitative results, and generalization tests. The rest of this paper is organized as follows. The related work is reviewed in Sec. 2. Sec. 3 describes the details of the proposed method. Experimental results are demonstrated in Sec. 4. Finally, concluding remarks are made in Sec. 5.

2 Related Work

Motion blur is apparent streaking appeared in a single frame or a sequence of frames due to rapid movement or long exposure. Many deep learning-based approaches have been proposed with remarkable success, including single-stage and multi-stage architectures. Single-stage methods usually employ complex network structures to improve the capabilities of the model. Gao *et al*. [8] proposed the parameter selective sharing and nested skip-connection networks, which consist of encoder-decoder networks with shared parameters, thereby reducing memory consumption. However, the overall architecture remains being high time complexity since the recovery process is iterative. Zhang *et al*. [29] proposed a spatially-variant architecture to model the varying motion blur. Kupyn *et al*. [11] proposed a conditional generative adversarial network based on a feature pyramid with a Wasserstein loss to generate high-quality deblurred images. Shen *et al*. [21] introduced a human-aware to selectively remove the blurring of foreground and background. The multi-stage approaches stack multiple lightweight sub-networks to decompose the complex task into several solvable problems. Zhang *et al*. [28]proposed a deeply stacked hierarchical multi-patch network that leverages multiple local-to-coarse operations to focus on different scales of a blur. Zamir *et al*. [27] proposed a multi-stage progressively restoration network(MPRNet), which consists of two encoder-decoder subnetworks and an original resolution network. However, the original resolution network performs less efficiently and requires longer computation time. In addition, MPRNet also presented a supervised attention module (SAM) to improve recovering process at each stage. Inspired by [27], the proposed architecture consists of SAM module and two encoder-decoder sub-networks to facilitate performance. More discussions can be found in the NTIRE Challenge reports [16, 17].

3 Proposed Method

This section presents the details of the proposed Spatially Guided Enhancement Network (SGENet) for dynamic image deblurring. Then, we describe the proposed Contextual-Aware Enhancement Module (CEM) and finally illustrate our Cross-Stage Selective Aggregation Module (CSAM).

3.1 Network Architecture

A schematic of the proposed SGENet is shown in Fig. 1, which consists of two encoder-decoder sub-networks for restoring blurred images. The encoder-decoder subnetwork is

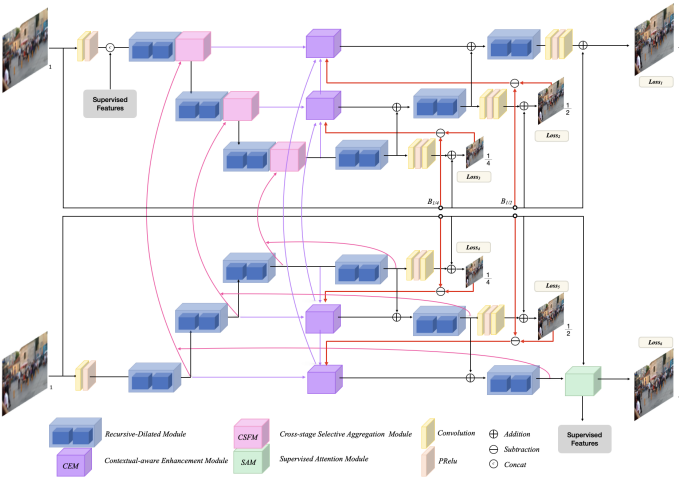

Figure 1: The framework of the proposed deblurring network.

based on U-Net [20] with the following modifications. First, a dilated convolution module [25] is added with a recursive residual design [6] to extract multi-scale features. Second, the feature maps at U-Net skip connections are processed with the proposed CEM, which is capable of synthesizing the detailed texture of the restored image. Finally, instead of simply stacking multiple stages, we incorporate a cross-stage selective aggregation module between the two stages, which can adjust the weights of different sub-networks adaptively to select useful feature representations.

Most previous methods [23, 28] attempt to restore images by extracting features from blurred input images. However, the high-frequency information of sharp images is quite diverse, complex, and difficult to learn. Therefore, more accurate reference images are required for the image reconstruction. Compared with the blurred input image, the predicted image contains precise high-frequency information and it would be more effective to fully exploit the spatial details of the predicted image to recover the image. Specifically, the proposed SGENet accesses the input image and predicts the image at each encoder-decoder scale. Given any stage t or scale s, the proposed model predicts the residual image $R_s^t \in R^{H\times W\times 3}$ and adds the degraded image $D_s^t \in R^{H\times W\times 3}$ up to obtain restored image $I_s^t \in R^{H\times W\times 3}$ defined as:

$$I_s^t = D_s^t + R_s^t \tag{1}$$

After that, we adopt the loss function shown in following equation to allow the predicted images to preserve accurate high-frequency information such as the detailed texture and edges. Then, we apply the predicted image as an attention map to precisely recalibrate the intermediate stage features to reconstruct outputs with detailed textures.

$$L = \sum_{t=1}^{2}\sum_{s=1}^{3}\left[L_{char}(I_s^t, Y) + \lambda L_{edge}(I_s^t, Y)\right] \tag{2}$$

where Y represents the ground truth image. L_{char} and L_{edge} are Charbonnier loss [2] and edge

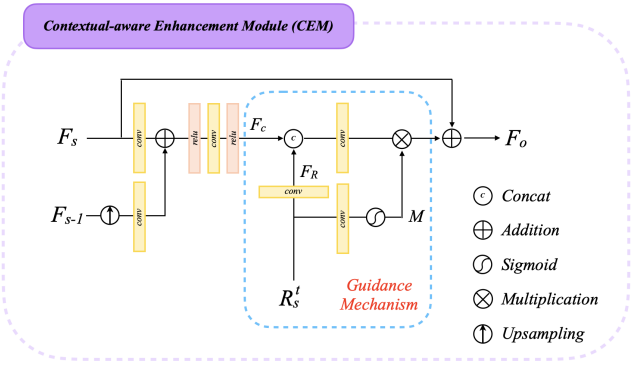

Figure 2: The architecture of contextual-aware enhancement module (CEM)

loss [9]:

$$L_{char} = \sqrt{\|I_s^t - Y\|^2 + \varepsilon^2} \tag{3}$$

$$L_{edge} = \sqrt{\|Lap(I_s^t) - Lap(Y)\|^2 + \varepsilon^2} \tag{4}$$

In addition, the constant ε is empirically set to 10^{-3}, where Lap denotes the Laplacian operator; The parameter λ in Eq. 2 is set to 0.05 to balance the loss terms.

3.2 Contextual-aware Enhancement Module (CEM)

Recent multi-stage networks for image deblurring [15, 27] typically adopt a single-scale feature pipeline to recover the spatial details and texture structure from input blurred images. However, such a pipeline attempts to extract high-resolution images leads to longer inference runtime. We introduce a contextual-aware enhancement module shown in Fig. 2, which can provide valuable supervised high-frequency information at each scale of the encoder-decoder to enhance multi-scale features progressively. Furthermore, with the help of the supervised guidance mechanism, we generate attention maps to suppress unfavorable features at the current scale, allowing only beneficial features to propagate to the next step, achieving significant performance gains.

Specifically, we first take the incoming features F_s and F_{s-1} from different scales to produce multi-scale features F_c by 3×3 convolution and ReLU activation, where $\{F_s, F_c\} \in R^{H \times W \times C}$, $F_{s-1} \in R^{\frac{H}{2} \times \frac{W}{2} \times C}$, $H \times W \times C$ denote the size of feature maps. Then, we obtain the predicted residual image R_s^t through Eq. 1 to generate residual features $F_R \in R^{H \times W \times C}$ and attention maps $M \in R^{H \times W \times C}$ using 3×3 convolution and sigmoid activation, respectively. Next, we fuse the residual features F_R with the multi-scale features F_c by concatenation and apply the attention maps M as a guided filter to obtain fine-detailed output features $F_{refined} \in R^{H \times W \times C}$. Finally, the output features would pass to the subsequent processing for further recovery.

3.3 Cross-stage Selective Aggregation Module (CSAM)

As shown in Fig. 3, the proposed CSAM is composed of two steps such as aggregate and select. The aggregate operation generates semantic features by fusing features of different

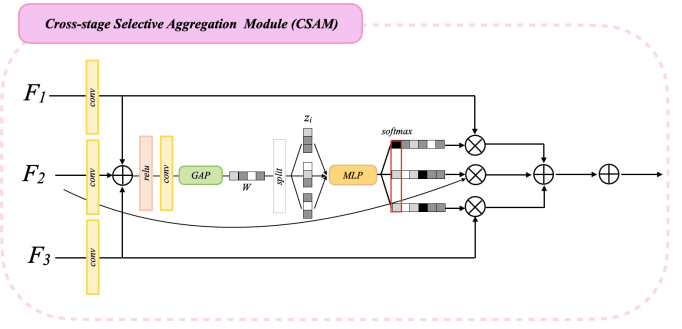

Figure 3: Schematic for the cross-stage selective aggregation module (CSAM)

stages; the select operation reweights the semantic features using an attention mechanism to select valuable information. CSAM takes the input features F_i from different stages, where $i \in \{1,2,3\}$, then combine these features to obtain semantic features $F_{semantic}$. After that, we employ the global average pooling (GAP) to generate channel-wise statistics as w:

$$F_{semantic} = \sum_{i=1}^{3} F_i \tag{5}$$

$$w = \frac{1}{H \times W} \sum_{p=1}^{H} \sum_{q=1}^{W} F_{semantic}(p,q) \tag{6}$$

where $F_{semantic} \in R^{H \times W \times C}$, and $w \in R^{1 \times 1 \times C}$. We split the feature vector w into 3 feature representations $z \in R^{1 \times 1 \times \frac{C}{3}}$, then pass through the multi-layer perceptron (MLP) module to learn the correlations in the latent space. Finally, we generate the attention feature vector using softmax activation to adaptively recalibrate the features of different stages as follow:

$$Y_i = F_i \times \sigma_{softmax}(\theta(z_i)) \tag{7}$$

where θ and Y_i denotes sets of MLP and output features, respectively.

4 Experiments

To demonstrate the advantages of our proposed deblurring framework, we evaluate the performance by comparing it with the state-of-the-art methods on two popular datasets, and conduct further ablation studies to analyze the contributions of individual components of our proposed network.

4.1 Dataset and implementation details

As shown in [4, 22, 27], we use the GoPro [15] dataset that contains 3,214 pairs of blurred and sharp images with the resolution of 720×1280, where 2,103 image pairs are for training and 1,111 pairs for evaluation. To demonstrate the generalization of our model, we take our GoPro [15] trained model and directly apply it to the test images of the HIDE [21] dataset, which consists of 2,025 images collected for human-aware motion deblurring.

The proposed framework is end-to-end trainable and has no pretraining process. Our SGENet uses 2 RDMs with 96, 120, 144 channels at each scale of the encoder-decoder. We

Figure 4: Qualitative comparisons among state-of-the-art method and our proposed SGENet on the GoPro test dataset.

first randomly crop the input image into 256×256 patches and train our model for 3,000 epochs using Adam optimizer with the initial learning rate of 10^{-4} steadily decreased to 10^{-7} using the cosine annealing strategy [14]. Horizontal and vertical flips are applied for data augmentation randomly. Our experiments are conducted on Intel i7-10700KF CPU and NVIDIA RTX 3090 GPU.

4.2 Performance comparisons

4.2.1 Quantitative Evaluation

We compare our method with the 9 latest approaches [4, 8, 15, 19, 22, 23, 27, 28, 30] on two popular datasets through the commonly-used metrics, i.e., PSNR and SSIM. The quantitative results on GoPro [15] and HIDE [21] datasets are listed in Table 1. For fair comparison, the runtime of models is measured using the released code with the image resolution of 720×1280 in the same environment. As shown in Table 1, the proposed SGENet outperforms other approaches on the GoPro dataset while achieving the fastest runtime. Specifically, the average PSNR and runtime of SGENet on the GoPro dataset are 32.96 dB and 0.017s, respectively. Our proposed model has 0.3 dB higher and 6.82× faster than the best model (MPRNet) among these approaches. Furthermore, our method has better performance than the second-best model (MIMO-UNet) on the GoPro dataset with almost the same inference time. We also evaluated our methods on the recent HIDE dataset [21] to verify the generalization ability of our model. As listed in Table 1, the proposed SGENet recorded the second-best performance in terms of PSNR and SSIM, which shows the robustness of our proposed method.

Figure 5: Qualitative comparisons among state-of-the-art method and our proposed SGENet on the HIDE test dataset.

Dataset	GoPro		HIDE		-	
Method	PSNR	SSIM	PSNR	SSIM	Runtime	Params.
DeepBlur [15]	29.23	0.916	25.73	0.873	-	11.7
SRN [23]	30.26	0.934	28.36	0.915	-	6.8
PNN+NSC [8]	30.92	0.942	29.11	0.913	-	<u>2.84</u>
DMPHN [28]	31.20	0.945	29.09	0.924	0.026	21.7
DBGAN [30]	31.10	0.942	28.94	0.915	0.225	11.5
MT-RNN [19]	31.15	0.945	29.15	0.918	0.050	**2.6**
SAPHNet [22]	31.85	0.948	29.98	0.930	-	-
MIMO-UNet [4]	32.45	0.957	29.99	0.930	<u>0.018</u>	16.1
MPRNet [27]	<u>32.66</u>	<u>0.959</u>	**30.96**	**0.939**	0.208	20.1
SGENet	**32.96**	**0.961**	<u>30.71</u>	<u>0.937</u>	**0.017**	19.5

Table 1: Evaluation results on the GoPro and HIDE dataset. The best score and second best are highlighted and underlined. The runtime and parameters are expressed in seconds and millions.

4.2.2 Qualitative Evaluation

We show the qualitative comparisons with the competing methods on the GoPro and the HIDE datasets in Figs. 4 and 5, respectively. It can be observed that the results of other methods still suffer from local region blurring and even producing ringing artifacts, which destroy the original image contents. In contrast, our method generates the most comparable results to the ground truth images, and our restored images can well recover the global structure and the sharper detailed textures. In addition, we visualize the output of the proposed CEM unit in Fig. 6 to analyze its contribution. The second and third columns of

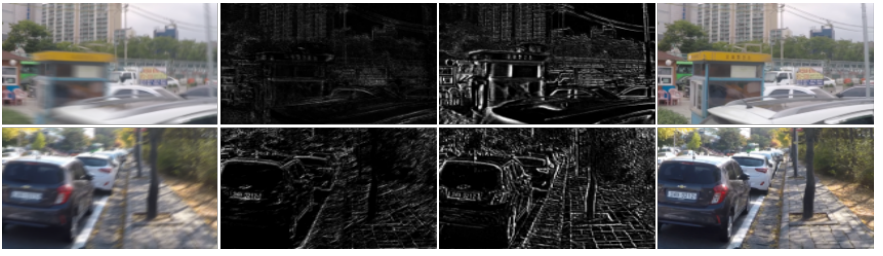

Figure 6: Two examples of visualization using our contextual-aware enhancement module. (a) Blurred Images. (b) Before enhancement processing. (c) After enhancement processing. (d) Ground-truth Image. It is showed that the features after CEM enhancement can highlight the detailed information such as edges and textures.

Fig. 6 shows the feature maps before and after the enhancement process, respectively. It is clear that the output of the proposed CEM contains detailed textures and sharp edge such as text and street. Therefore, our SGENet network has strong ability to generate sharper restored image. In conclusion, both the quantitative and qualitative results demonstrate that our method achieves superior performance.

Method	CSAM	CEM	PSNR
1-stage			30.05
		✓	30.24
2-stages			30.72
	✓		31.06
		✓	31.11
	✓	✓	**31.32**

Table 2: Effectiveness of individual components of the proposed SGENet on the GoPro test dataset.

Method	w/o guidance	w/ guidance
PSNR	31.15	**31.32**
Runtime	**0.017**	**0.017**

Table 3: Ablation studies on CEM for guidance mechanism.

4.3 Ablation Study

In this section, we conducted experiments to analyze the effectiveness of each component in our model on the GoPro dataset. We first train our model for 500 epochs and use a single-stage SGENet without any components as the baseline model. Table 2 demonstrates that removing the CEM causes a substantial performance drop from 31.32 dB to 31.06 dB, and from 31.32 dB to 31.11 dB when plugging out the CSAM. Note that the performance gain increases by a large margin from 30.72 dB to 31.32 dB when employing these two components. In addition, we also analyze the effectiveness of the proposed guidance mechanism. Fig. 2 illustrates the architecture of the proposed CEM with the guidance mechanism, and the experimental results are listed in Table 3. It shows that the proposed guidance mechanism can increase performance gain from 31.15 dB to 31.32 dB with nearly the same inference time. In other words, the proposed guidance mechanism can substantially increase the performance at almost no cost.

5 Conclusion

This paper proposes a novel spatial guidance enhancement network for single image deblurring, which aims at restoring blurred images with accurate spatial details. We develop the guidance mechanisms to progressively rebuild images by fully exploiting precise high-frequency information of predicted images. We require these high-level features with flexible information exchange across different stages. To this end, we propose a cross-stage selective aggregation strategy to adaptively utilize useful feature representations for an efficient multi-stage architecture. Experimental results show the proposed method achieves the state-of-the-art restoration performance with low time complexity when compared with the existing methods on two benchmark datasets.

References

[1] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11036–11045, 2019.

[2] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st International Conference on Image Processing*, volume 2, pages 168–172. IEEE, 1994.

[3] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 182–192, 2021.

[4] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4641–4650, 2021.

[5] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on image processing*, 20(7):1838–1857, 2011.

[6] Yixin Du and Xin Li. Recursive deep residual learning for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 730–737, 2018.

[7] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. Removing camera shake from a single photograph. In *ACM SIGGRAPH 2006 Papers*, pages 787–794. 2006.

[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*, pages 3848–3856, 2019.

[9] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8346–8355, 2020.

[10] Neel Joshi, C Lawrence Zitnick, Richard Szeliski, and David J Kriegman. Image deblurring and denoising using color priors. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1550–1557. IEEE, 2009.

[11] 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*, pages 8183–8192, 2018.

[12] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019.

[13] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. Selective kernel networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 510–519, 2019.

[14] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[15] 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*, pages 3883–3891, 2017.

[16] Seungjun Nah, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2020 challenge on image and video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 416–417, 2020.

[17] Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, and Kyoung Mu Lee. Ntire 2021 challenge on image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 149–165, 2021.

[18] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. Deblurring text images via l0-regularized intensity and gradient prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2901–2908, 2014.

[19] Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *European Conference on Computer Vision*, pages 327–343. Springer, 2020.

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[21] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5572–5581, 2019.

[22] 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*, pages 3606–3615, 2020.

[23] 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*, pages 8174–8182, 2018.

[24] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceedings of the IEEE international conference on computer vision*, pages 4799–4807, 2017.

[25] Boyan Xu and Hujun Yin. Dc-deblur: A dilated convolutional network for single image deblurring. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 234–245. Springer, 2021.

[26] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *European conference on computer vision*, pages 157–170. Springer, 2010.

[27] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.

[28] 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*, pages 5978–5986, 2019.

[29] Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson WH 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*, pages 2521–2529, 2018.

[30] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2737–2746, 2020.

[31] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.

[32] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2480–2495, 2020.