Benchmarking Deep Deblurring Algorithms: 
A Large-Scale Multi-Cause Dataset and A New Baseline Model

Kaihao Zhang\(^1\) Wenhan Luo\(^2\) Boheng Chen\(^2\) Wenqi Ren\(^3\) 
Björn Stenger\(^4\) Wei Liu\(^2\) Hongdong Li\(^1\) Ming-Hsuan Yang\(^5,6,7\)
\(^1\) Australian National University \(^2\) Tencent \(^3\) IIE, CAS \(^4\) Rakuten Institute of Technology
\(^5\) Google Research \(^6\) University of California, Merced \(^7\) Yonsei University

Abstract

Blur artifacts can seriously degrade the visual quality of images, and numerous deblurring methods have been proposed for specific scenarios. However, in most real-world images, blur is caused by different factors, e.g., motion and defocus. In this paper, we address how different deblurring methods perform on general types of blur. For in-depth performance evaluation, we construct a new large-scale multi-cause image deblurring dataset called (MC-Blur) including real-world and synthesized blurry images with mixed factors of blurs. The images in the proposed MC-Blur dataset are collected using different techniques: convolving Ultra-High-Definition (UHD) sharp images with large kernels, averaging sharp images captured by a 1000 fps high-speed camera, adding defocus to images, and real-world blurred images captured by various camera models. These results provide a comprehensive overview of the advantages and limitations of current deblurring methods. Further, we propose a new baseline model, level-attention deblurring network, to adapt to multiple causes of blurs. By including different weights of attention to the different levels of features, the proposed network derives more powerful features with larger weights assigned to more important levels, thereby enhancing the feature representation. Extensive experimental results on the new dataset demonstrate the effectiveness of the proposed model for the multi-cause blur scenarios.

1. Introduction

Image deblurring is an important problem in computer vision and image processing, which aims to restore a sharp image from an observed blurry input. Deblurring has been widely used in applications such as medical image analysis, computational photography, and video enhancement. Conventional methods formulate the task as an inverse filtering problem, using the blur model

$$I_B = I_S * K + \sigma_N,$$  \hspace{1cm} (1)

where \(I_B\) is the observed blurry image, \(I_S\) is the latent sharp image, \(K\) is the unknown blur kernel, \(\sigma_N\) is the additive noise, and \(*\) is the convolution operation. Image deblurring is a well-known ill-posed problem, and numerous priors such as natural image statistics have been employed to constrain the solution space. However, estimating \(I_S\) based on this formulation typically involves iterative and time-consuming estimation processes.

Numerous deep models have recently been applied to image deblurring within the supervised learning framework. These models require a large number of paired sharp and blurry images to train networks in an end-to-end manner. A number of datasets have been created by averaging continuous frames, convolving with blur kernels, or directly taking photos with two cameras with different shutter durations to obtain pairs of images. Although these datasets have advanced the deep deblurring models, there remain unaddressed issues with these datasets: 1) As shown in [26], averaging sharp images of low frame rate to synthesize blurry images can cause unnatural blurs. For datasets that include motion blurs, images are usually generated by averaging continuous frames captured with a relatively slow shutter speed (e.g., the GoPro dataset (240 fps)), or from images in interpolated high fps videos (e.g., the REDS dataset [26]); 2) For datasets containing uniform blurs, e.g., the dataset by Köhler et al. [17], the number of images is insufficient for training deep networks, images are not high-definition, and the kernel size is relatively small. With an increasing number of devices being able to capture Ultra-High-Definition (UHD) images, existing datasets are not suitable for training models capable of handling such images; 3) Datasets of real-world blurry images typically require additional processing steps such as accurate alignment [37]; 4) While defocus is a common reason for blurry images, few datasets are developed specifically for this type of blurs, and existing ones [1] are not sufficient for deep deblurring methods.

To overcome these limitations, we construct a large-scale multi-cause dataset including blurry images caused by multiple factors, named MC-Blur dataset. This dataset is com-
posed of four subsets. The first subset includes images that are averaged from sharp frames to synthesize motion blurs. Different from existing datasets, sharp frames are captured with various ultra-high-speed cameras (iPhone, Samsung, Sony, etc.) at different frame rates (250, 500, and 1000 fps). With different types of devices and different frame rates, this subset mimics various motion blurs in the real world. The second subset contains motion blurs based on convolving sharp images with blur kernels. Due to the increasing number of high-definition cameras, we capture a large number of UHD images at 4K+ resolution. These UHD images are convolved with large blur kernels. The third subset is specific to defocus blurs. We capture a set of images with various defocus effects by manually changing the focus setting. The fourth subset comprises real-world blurry images captured by different types of devices, e.g., mobile cameras. While no sharp images are available as ground truth, this subset is included for qualitative performance evaluation on real-world scenarios.

In addition, we propose a new baseline model based on an attention mechanism, which we call Level-Attention Deblurring Network (LADN). This model integrates a Level Attention Module (LAM) to learn correlations among features from different levels. With different attention weights for different levels of features, important features are emphasized and redundant ones are ignored. Experimental results on the MC-Blur dataset validate its performance and demonstrate its effectiveness. The main contributions of this paper are summarized as follows:

- We introduce a large-scale dataset of images containing blur artifacts due to multiple causes: motion blurs by averaging high fps sharp images, motion blurs by convolving Ultra-High-Definition (UHD) images with large blur kernels, real-world defocus, and real-world blur artifacts from a diverse set of camera models.
- To the best of our knowledge, the proposed MC-Blur dataset is the largest single image deblurring dataset. Based on the MC-Blur dataset, we conduct an extensive benchmarking study, evaluating some state-of-the-art methods for a comprehensive understanding of how recent deblurring methods perform in this general multi-cause scenarios.
- We propose a level-attention deblurring network with a dedicated attention module. Both qualitative and quantitative results on the MC-Blur dataset demonstrate the effectiveness of the proposed baseline model.

2. Related Work

2.1. Image Deblurring Datasets

Several datasets have been developed for advances in image deblurring [10, 13, 17, 20, 21, 27, 28, 39–41, 44, 56]. For example, motion blurs are simulated by convolving images with uniform blur kernels by Levin et al. [21] and Sun et al. [44], or with non-uniform kernels by Köhler et al. [17]. In [20], Lai et al. introduce a dataset including images of both real-world and synthetic blurs (uniform blur kernels). However, the size of this dataset is still relatively small. Even when synthesizing blurry images via realistic blur kernels, the scale of the above-mentioned datasets is small, making them difficult to use for deep learning-based deblurring methods. Furthermore, the blur kernels are relatively small, which makes them less effective for deblurring Ultra-High-Definition (UHD) images.

Considering that images are captured for the duration of camera exposure, blurry images can be modeled by integration of neighboring frames [9],

\[ I_B = g \left( \frac{1}{T} \int_{t=0}^{T} S(t) \, dt \right), \]

where \( T \) is the exposure time period and \( g(\cdot) \) is the Camera Response Function (CRF). To model this process [9] and alleviate the problem of alignment [38], a number of deblurring datasets have been created based on the discrete formulation,

\[ I_B \simeq g \left( \frac{1}{M} \sum_{t=0}^{M-1} I_{S[t]} \right), \]

where \( M \) is the number of frames.

In particular, the GoPro dataset [29] has been widely used for training deep models. Its sharp images are captured by a GoPro Hero4Black camera with a shutter speed of 240 fps. Blurry images are generated by averaging continuous sharp frames over a time window. Similarly, the HIDE [40] and REDS [26] datasets are created based on this method.

Rim et al. [38] develop a dataset containing real blurry images and the corresponding sharp images. Image pairs are taken by two different cameras with different exposure times. While the blur is realistic, this work requires an additional image alignment step to generate image pairs, which causes the problem of imprecise alignment. In addition, this dataset also lacks defocus blurry images or UHR images, which is of great interest for real-world scenarios. On the other hand, Abuolaim and Brown [1] capture 500 images synthesized with defocus blurs, but this number is small compared to the recent large-scale deblurring datasets.

To address these problems, we propose a large-scale multi-cause image deblurring dataset called MC-Blur. We first create a Real High-fps based Motion-blurred set (RHM), which includes blurry images synthesized from high frame rate videos to balance the problem of alignment in [38] and frame rate [26, 29]. To address the issues with large blur kernels and UHD blurry images, we construct a large kernel UHD Motion-blurred (UHDM) dataset in this work. In addition, we introduce a Large-Scale real-world
### Table 1. Representative benchmark datasets for evaluating single image deblurring algorithms.

| Dataset                  | Sharp Images | Blurred Images | Motion | Defocus | Real | Aligned |
|--------------------------|--------------|----------------|--------|---------|------|---------|
| Levin et al. [21]        | 4            | 32             | ✓      | ×       | ×    | ✓       |
| Sun et al. [44]          | 80           | 640            | ✓      | ×       | ×    | ✓       |
| Köhler et al. [17]       | 4            | 48             | ✓      | ×       | ×    | ✓       |
| Lai et al. [20]          | 108          | 300            | ✓      | ×       | ✓    | ✓       |
| GoPro [28]               | 3,214        | 3,214          | ✓      | ×       | ×    | ✓       |
| HIDE [40]                | 8,422        | 8,422          | ✓      | ×       | ✓    | ✓       |
| Blur-DVS [13]            | 2,178        | 2,918          | ✓      | ×       | ✓    | ×       |
| Abuolaim et al. [1]      | 500          | 500            | ×      | ✓       | ✓    | ✓       |
| Rim et al. [38]          | 9476         | 9476           | ✓      | ×       | ✓    | ✓       |
| RHM-250fps               | 25,000       | 25,000         | ✓      | ×       | ×    | ✓       |
| RHM-500fps               | 25,000       | 25,000         | ✓      | ×       | ✓    | ✓       |
| RHM-1000fps              | 37,500       | 37,500         | ✓      | ×       | ×    | ✓       |
| UHDM                     | 2,000        | 10,000         | ✓      | ×       | ✓    | ✓       |
| LSD                      | 22,400       | 22,400         | ×      | ✓       | ✓    | ✓       |
| RMBQ                     |              | 10,000         | ✓      | ×       | ✓    | ✓       |
| MC-Blur dataset          | 111,900      | 129,900        | ✓      | ✓       | ✓    | ✓       |

Defocus blurry set (LSD) and a Real Mixed Blurry Qualitative set (RMBQ). The details of the representative datasets and the proposed MC-Blur dataset are listed in Table 1.

### 2.2. Deep Deblurring Methods

In recent years, numerous deep learning methods have been proposed to address single image deblurring [3, 5, 7, 8, 11, 13–15, 18, 19, 22–25, 28, 31, 34, 35, 42, 43, 45, 48, 51, 53], and video deblurring [2, 6, 12, 30, 32, 41, 46, 52, 55]. Deep deblurring methods typically train neural networks in an end-to-end manner, using blurry images as inputs and updating network parameters by comparing the outputs and the ground truth sharp images. The idea of multi-scale processing, together with an adversarial loss, is used in [29] for image deblurring. Kupyn et al. [18] adopt a conditional GAN for the deblurring task, resulting in the DeblurGAN model. An enhancement in terms of accuracy and speed was introduced with DeblurGAN-v2 [19]. In [5], patch-level deblurring is carried out to obtain an initial global estimation. The blur kernel is estimated and the final result is obtained via deconvolution. Motivated by Spatial Pyramid Matching, a multi-patch scheme is applied to learn hierarchical representations in [50]. A CNN is combined with an RNN in [51] to deblur images of spatial-variant blurs in dynamic scenes. Similarly, an LSTM with a CNN is employed in [45]. More recently, Gao et al. [8], Zhang et al. [53], and Park et al. [33] develop new architectures including nested skip connections or reblurring networks, which improve the image deblurring accuracy.

In the Sec. 5, we will choose six representative state-of-the-art image deblurring methods for evaluation. Apart from the image deblurring methods, there exist some related low-level image enhancement methods like image super-resolution networks [47, 54], image deraining networks [36] as well as image restoration networks [16, 49], which can also handle the deblurring task. Considering they are not specific for the image deblurring task, we do not evaluate them in this benchmark.

### 3. MC-Blur Dataset

The advances of learning-based methods for image deblurring relies heavily on the quality and quantity of datasets. To benchmark the current state-of-the-art image deblurring methods in various conditions, we construct the large-scale multi-cause (MC-Blur) dataset. It comprises four types of blurs: uniform blurs, motion blurs by averaging continuous frames, defocus blurs, and real-world blurs. In addition, the MC-Blur dataset includes a large amount of low-light images captured during both day and night time. Table 1 shows the differences between the MC-Blur dataset and existing ones in details. The four subsets are introduced in the following.

#### 3.1. RHM Subset

Averaging continuous frames within a time window to generate motion-blurred images is a common practice to synthesize images with motion blurs. For example, in the GoPro dataset, images captured at 240 fps were averaged to produce blurry images [29] at 30 fps. However, if the frame rate of the images to be averaged is not sufficiently high, the synthesized motion blur can be unnatural [26]. As such, Nah et al. [26] record videos at 120 fps and interpolate them to 1920 fps by CNNs. To remove this potential error source, we opt to capture sharp images using high frame-rate cameras to create the Real High-fps based Motion-blurred dataset, RHM. Images in this RHM dataset are captured in three settings. The first setting corresponds...
to the highest fps, up to 1,000 fps. The sharp videos are recorded using a Sony RX10 camera. This subset contains 30,000 images for training, and 7,500 images for testing. The sharp images in the second setting are captured with the same camera at 500 fps. The training and test sets contain 20,000 and 5,000 images, respectively. The third setting corresponds to images captured at 250 fps with mobile devices such as iPhone, Huawei, and Sony RX10 cameras. For training and testing, this set contains images of 20,000 and 5,000, respectively. All images are resized via bicubic downsampling to reduce noise. The image resolutions in this set are 960 × 540 and 640 × 360 pixels.

3.2. UHDM Subset

Another way to synthesize degraded images caused by motion blur is to convolve images with kernels. Existing datasets based on this approach either use low-resolution images or small blur kernels. For example, when the image resolution is lower than 720 × 720 pixels, the size of the blur kernels is usually set within the range from 15 × 15 to 27 × 27 pixels. We note that deblurring 4K+ images requires restoration with more details, which is challenging if the models are trained with low-resolution images. Thus, we capture sharp images of 4K-6K resolutions to create the Large-Kernel UHD Motion-blurred (UHDM) set. To convolve with the sharp images, we use blur kernel sizes of 111 × 111, 131 × 131, 151 × 151, 171 × 171, and 191 × 191, respectively. The training and testing sets contain 8,000 and 2,000 images, respectively. Blur kernels are generated via 3D camera trajectories [4].

3.3. LSD Subset

A few datasets on defocus image deblurring have recently been developed. To create the Dual-Pixel [1] dataset, Abuolaim et al. capture pairs of images of the same static scene at two aperture size via a Canon EOS 5D Mark IV DSLR camera. Focus distance and focal length differ across captured pairs to capture various defocus blurry images. However, this dataset is mainly designed for the dual-pixel problem. As a result, it provides only 500 pairs of blurry images and their corresponding all-in-focus images, and the scale is small for approaches based on deep learning models. Meanwhile, these blurry images contain large sharp patches. Given most of current deep deblurring networks take patches cropped from blurry images as input, the role of these sharp areas is limited during the training stage.

We collect a large-scale defocus blurred set, LSD, by capturing 18,000 image pairs of sharp image and blurry image with the defocus effect in the training set, and 4,400 image pairs for testing. The image resolution is at least 900 × 600 pixels. As in [1], we manually control the focus to obtain the blurry images and their corresponding sharp ones.

3.4. RMBQ Subset

The above three sets contain different kinds of real and synthesized blurry images. In the real world, blur artifacts can be more complex and difficult to approximate. For instance, real-world blurs are caused by multiple reasons, such as the blur caused by both camera shake and out-of-focus. Thus it is difficult to guarantee the generalization of models trained with images containing only a specific kind of blurs. We therefore capture another set of blurry images with various devices, including both high-end digital cameras and mobile phones (iPhone, Samsung, and Huawei). There are 10,000 images in this Real Mixed Blurry Qualitative (RMBQ) set. This set is designed only for qualitative evaluation, as there are no available sharp ground truth images.

4. Level-Attention Deblurring Network

In this work, we also present a new baseline model, Level-Attention Deblurring Network (LADN), for the image deblurring task. This model integrates an effective level attention module to enhance the representation capability of the features. In the following, we first introduce the LADN model architecture, and then represent the level attention module.

4.1. Network Architecture

An overview of the architecture of our proposed LADN model is shown in Fig. 1. Given a blurry image, features are extracted via a convolutional layer. A sequence of Residual-in-Residual Dense Blocks (RRDB) [47] is used to extract different levels of features. The level attention module is then used to form a 2D matrix that captures the correlation among different levels of features. Larger attention weights are given to more important feature levels and smaller weights to features at less important levels. Finally, the features are skip-connected to the first extracted features, and processed by several convolutional layers to produce the finally deblurred image.

Specifically, a blurry image $I_B$ is fed into the LADN, and a primary feature $F_P$ is extracted after the convolutional layers. A set of RRDB extract different levels of features, which are denoted as $F_{RRDB-1}, F_{RRDB-2}, \ldots, F_{RRDB-D}$. These features are concatenated and processed by the level attention module:

$$F_{LAM} = \Phi(\text{concat}(F_{RRDB-1}, F_{RRDB-2}, \ldots, F_{RRDB-D})),$$

where $D$ is the number of different levels, $\text{concat}(\cdot)$ is the concatenation operation, and $\Phi$ is a function of the level-aware attention mechanism approximated by a module introduced in Sec. 4.2.

A feature $F_{LAM}$ contains level-wise attentions, thereby emphasizing levels of features with larger weights and ne-
Figure 1. Model architecture of the proposed Level attention Deblurring Network. The LADN takes a blurry image as input, and uses Residual-in-Residual Dense Block (RRDB) and convolution layers to extract feature maps from different levels, which are further fed into a Level Attention Module to learn the correlations among different levels. Finally, a set of stacked layers generate the deblurred image.

Table 2. Performance evaluation of deep image deblurring methods on the proposed RHM set. PSNR and SSIM values are reported.

| Method         | DeepDeblur [29] | DeblurGAN [18] | SRN [45] | DeblurGAN-v2 [19] | DMPHN [50] | DBGAN [53] | LADN       |
|----------------|-----------------|----------------|----------|-------------------|------------|------------|------------|
| 250fps         | 30.38/0.8766    | 24.89/0.6364   | 30.57/0.8799 | 26.99/0.8061       | 30.42/0.8768 | 27.89/0.8191 | 31.31/0.8918 |
| 500fps         | 31.08/0.8974    | 24.66/0.6748   | 31.54/0.9051 | 27.67/0.8320       | 31.43/0.9018 | 28.36/0.8388 | 31.77/0.9104 |
| 1000fps        | 32.41/0.8966    | 25.20/0.6535   | 32.69/0.9016 | 29.81/0.8461       | 32.41/0.9096 | 29.66/0.8318 | 32.77/0.9031 |

4.2. Level Attention Module

Deep features from different levels are known to carry different importance weights for vision tasks including image deblurring. The layer attention module is designed to learn attention weights for the different feature maps for effective deblurring.

In this work, the feature maps from the sequential levels are concatenated first. As the lower portion of Fig. 1 shows (from the left to the right), each feature map is first reshaped as a 2D matrix with size $D \times HWC$, where $D$, $H$, $W$, and $C$ are the number of RRDBs, height, width and number of channels, respectively. This matrix is multiplied with its transpose to form a 2D matrix of size $D \times D$. Each element in this matrix represents the correlation between the two levels of features corresponding to the column and the row index. This correlation matrix is multiplied with the reshaped feature concatenation, and the formed features are reshaped into a $D \times H \times W \times C$ feature tensor. The feature tensor can be taken as feature residual, and added to the original feature concatenation in an element-wise manner. The resulting additive features are reshaped into an $H \times W \times DC$ tensor by absorbing the level number $D$. Subsequently, a $1 \times 1$ convolution is applied to decrease the channel number from $DC$ to $C$. The feature is again added with the primary features $F_P$ in an element-wise manner, which is processed by several convolutional layers to generate the latent sharp (deblurred) image $I_S$.
Table 3. Performance evaluation of deep deblurring methods on the proposed UHDM set.

| Method        | PSNR | SSIM  |
|---------------|------|-------|
| DeepDeblur [29] | 22.23 | 0.6322 |
| DeblurGAN [18] | 20.39 | 0.5568 |
| SRN [45]      | 22.28 | 0.6346 |
| DeblurGAN-v2 [19] | 21.03 | 0.5839 |
| DMPHN [50]    | 22.20 | 0.6378 |
| DBGAN [53]    | 21.52 | 0.6025 |
| LADN          | 22.49 | 0.6238 |

Table 4. Performance evaluation of deep deblurring methods on the proposed LSD set.

| Method        | PSNR | SSIM  |
|---------------|------|-------|
| DeepDeblur [29] | 20.73 | 0.7218 |
| DeblurGAN [18] | 20.04 | 0.6355 |
| SRN [45]      | 21.66 | 0.7664 |
| DeblurGAN-v2 [19] | 21.13 | 0.6964 |
| DMPHN [50]    | 21.23 | 0.7519 |
| DBGAN [53]    | 21.56 | 0.7536 |
| LADN          | 21.83 | 0.7658 |

5.1. Evaluated Deblurring Methods

We evaluate six state-of-the-art deblurring methods on the proposed MC-Blur dataset, including: DeepDeblur [29], DeblurGAN [18], SRN [45], DeblurGAN-v2 [19], DMPHN [50], and DBGAN [53]. These methods include a variety of different models, and include the multi-scale architectures, GAN and multi-patch networks.

Among these approaches, DeepDeblur and SRN are multi-scale networks. DeepDeblur takes blurry images with different resolution (e.g., 1/4 and 1/2) as input to generate their corresponding sharp versions with the original resolution. The SRN model introduces a scale-recurrent network that uses fewer parameters than DeepDeblur. DeblurGAN, DeblurGAN-v2, and DBGAN are GAN-based models which use generators to restore sharp images and apply discriminators to make the deblurred images more realistic. DeblurGAN is the first GAN-based deep deblurring network. To improve the deblurring performance, DeblurGAN-v2 introduces a feature pyramid network to replace the generator of DeblurGAN. DBGAN introduces a new model to generate realistic blurry images to help train the deblurring network. DMPHN is a multi-patch network that removes blur from small patches to help in the final deblurring operation. As the inputs at different levels have the same resolution, the DMPHN uses residual learning and thus requires only small filter sizes. Considering that the original authors have tried their best to find parameters to train the above models, we re-train six models on the MC-Blur dataset using the original parameter settings.

5.2. Implementation Details

The LADN model architecture shown in Fig. 1 includes four classical convolution layers and eight RRDB modules, except the model for the UHDM dataset, which uses three RRDB modules to save memory. The kernel size is $3 \times 3$. For the proposed LADN, we initialize weights using a Gaussian distribution with zero mean and a standard deviation of 0.01. All weights are updated after learning a mini-batch of size 8 in each iteration. During the training stage, we crop a $256 \times 256$ patch at any location and randomly flip frames to augment the data. Adam optimizer is used with a learning rate of $10^{-4}$.

5.3. Results on the RHM Dataset

To evaluate the performance of image deblurring methods, we first carry out experiments on the RHM dataset. Table 2 shows that the DeepDeblur, SRN, and DMPHN models perform well in terms of PSNR and SSIM. One reason is that these methods use a pixel-level loss function, which helps in obtaining high values of full-reference pixel-based metrics. The DeblurGAN, DeblurGAN-v2, and DBGAN models use discriminators to help synthesize more realistic deblurred images. These models are enforced to focus on pixel-wise measures and pay attention to the whole image. As the proposed LADN uses a level attention module to learn the correlation of features from different layers, it can better use feature maps and achieve better performance in terms of PSNR and SSIM. We also show a visual comparison of different methods on the RHM dataset in Fig. 3a.

5.4. Results on the UHDM Dataset

In this section, the same set of deblurring methods are evaluated on the UHDM images, which are synthesized by convolution. Table 3 shows that the PSNR and SSIM values of all methods are significantly lower than those in the Table 2. One reason is that we use large-size blur kernels to synthesize blurry images, which makes the deblurring task more difficult. Another reason is that, compared with HD (2K) image deblurring, deblurring UHD (4K+) images requires recovering more details. In addition, we show qualitative results corresponding to the large-kernel blur cause in Fig. 3b. The proposed LADN generates sharp text and structural information, which help people to understand the semantic information of images, such as text in Fig. 3b.

5.5. Results on the LSD Dataset

To investigate the performance of the SOTA deblurring methods along with our proposed LADN in the case of defocus blur, we conduct a comparison study on the LSD dataset. Quantitative results are shown in Table 4, and qualitative results in Fig. 3c. We note that defocus image deblurring is a more difficult problem compared with deblurring of motion-blurred images. While the SOTA deep deblurring methods can restore high-quality motion-deblurred images synthesized by averaging neighbouring frames, the performance of defocus deblurring is significantly lower.
Table 5. Ablation study and comparison on the GoPro dataset.

| Method               | PSNR  | SSIM  |
|----------------------|-------|-------|
| DeepDeblur [29]      | 29.08 | 0.914 |
| SRN [45]             | 30.26 | 0.934 |
| DeblurGAN-v2 [19]    | 29.55 | 0.935 |
| DMPHN [50]           | 30.25 | 0.937 |
| DBGAN [53]           | 30.43 | 0.937 |
| RNNDeblur [51]       | 29.19 | 0.931 |
| Shen et al. [40]     | 30.26 | 0.940 |
| AlJadannay et al. [3] | 30.35 | 0.961 |
| Hu et al. [11]       | 30.62 | 0.941 |
| Gao et al. [8]       | 30.92 | 0.942 |
| Park et al. [33]     | 31.15 | 0.945 |
| LADN (w/o LAM)       | 31.19 | 0.942 |
| LADN                 | 31.43 | 0.947 |

(a) Input  
(b) Tao et al. [45]  
(c) LADN without LAM  
(d) LADN

Figure 2. Deblurred results on the GoPro dataset. (a) input, (b) results of Tao et al. [45], (c) LADN without LAM, and (d) LADN with LAM. Best viewed in color.

Table 6. Run-time comparison of SOTA deep deblurring methods.

| Method               | Speed (in seconds) |
|----------------------|--------------------|
| DeepDeblur [29]      | 26.76              |
| DeblurGAN [18]       | 2.46               |
| SRN [45]             | 28.41              |
| DeblurGAN-v2 [19]    | 3.63               |
| DMPHN [50]           | 17.63              |
| DBGAN [53]           | 31.62              |
| LADN                 | 1.67               |

5.6. Results on the RMBQ Dataset

In addition, we evaluate the performance of deblurring methods in the real-world scenarios. The deblurred images by the evaluated methods are shown in Fig. 3d. Overall, the deblurring methods trained on the proposed MC-Blur dataset are able to generate relatively sharp images, recovering sharp text and image structure.

5.7. Ablation Study on the GoPro Dataset

We evaluate the LADN model against the SOTA methods on the widely-used GoPro dataset. To demonstrate the effective of the level attention module (LAM), we also design a variant of LADN without the LAM. The ablation study results are shown in Table 5. While the variant model without the LAM achieves competitive performance, the deblurred results are worse than those by the proposed LADN method. Qualitative results are shown in Fig. 2. The results by the SRN method [45] show improvement in terms of sharpness with regard to the input, while some letters and numbers remain blurry. In contrast, the proposed LADN model is able to restore sharper and clearer images.

5.8. Efficiency Analysis on UHD images

Efficiency should be taken into consideration when the image resolution is high, especially for large UHD images. Table 3 shows the evaluation results on UHD images. These experiments are carried out using a standard platform with a P40 GPU. Table 6 shows the run-time performance of the evaluated methods. The DeepDeblur, SRN, DMPHN and DBGAN models require more than ten seconds to process one UHD image. On the other hand, it takes a few seconds for the DeblurGAN or DeblurGAN-v2 models to process one image. In contrast, it takes 1.67 seconds for the proposed LADN model to deblur one UHD image.

5.9. Discussion

The evaluation results on different sets of the proposed MC-Blur dataset, lead to several interesting findings.

First, GAN based networks achieve lower values of PSNR & SSIM for motion-blurred images than methods without using the GAN framework. However, the two kinds of networks (with and without using the GAN framework) do not show significant differences for out-of-focus images. This indicates that paying attention to the whole images (e.g. the adversarial loss function), rather than just considering the pixel level (e.g., L1 and L2 loss functions), may be a direction for defocus deblurring. The results in Tables 2, 3, 4 support this finding.

Second, for motion-blurred images, current deep deblurring networks can generate high-quality images. However, it is difficult for them to achieve similar performance on large kernel based Ultra-High-Definition blurry images (Tables 2, 3). Considering that increasing modern mobile devices allow capturing UHD images, it may be a meaningful direction for researchers to study UHD image deblurring.

Third, current deep methods can generate a deblurred image in 2 seconds [50]. However, as shown in Table 6, it takes significantly longer time for them to handle a UHD image. Therefore, how to generate deblurred UHD images at a high speed while maintaining the performance of deblurring, is still an open task.

Limitations. In this paper, we propose a LADN for image deblurring. As we claim throughout this paper, the LADN is used as a baseline method and optimizing its performance is not our main goal. As a result, it does not
significantly outperform current SOTA methods, although it indeed achieves comparable performance in several settings.

6. Conclusion

We propose a new large-scale multi-cause dataset called MC-Blur to benchmark deblurring methods on images with blurs caused by various factors. The MC-Blur dataset includes a real high-fps based motion-blurred set, a large-kernel Ultra-High-Definition motion-blurred set, a large-scale real-world defocus blurry set and a real mixed blurry set. By decoupling the underlying factors for blur, the dataset can be used to develop effective deblurring methods for real-world scenarios, especially models based deep neural networks. We also propose a new baseline model, layer-attention deblurring network (LADN), which performs well on the MC-Blur dataset and GoPro dataset, in terms of PSNR, SSIM, and run-time.
References

[1] Abdullah Abuolaim and Michael S. Brown. Defocus deblurring using dual-pixel data. In *European Conference on Computer Vision*, 2020. 1, 2, 3, 4

[2] Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *European Conference on Computer Vision*, 2018. 3

[3] Raied Aljadaany, Dipan K Pal, and Marios Savvides. Douglas-rachford networks: Learning both the image prior and data fidelity terms for blind image deconvolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 3, 7

[4] Giacomo Boracchi and Alessandro Foi. Modeling the performance of image restoration from motion blur. *IEEE Transactions on Image Processing*, 2012. 4

[5] Ayan Chakrabarti. A neural approach to blind motion deblurring. In *European Conference on Computer Vision*, 2016. 3

[6] Huaqiu Chen, Jinwei Gu, Orazio Gallo, Ming-Yu Liu, Ashok Veeraraghavan, and Jan Kautz. Reblur2deblur: Deblurring videos via self-supervised learning. In *IEEE International Conference on Computational Photography*, 2018. 3

[7] 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 *IEEE International Conference on Computer Vision*, 2021. 3

[8] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 3, 7

[9] Michael Hirsch, Christian J Schuler, Stefan Harmeling, and Bernhard Schölkopf. Fast removal of non-uniform camera shake. In *IEEE International Conference on Computer Vision*, 2011. 2

[10] Michal Hradiš, Jan Kotera, Pavel Zemčík, and Filip Šroubek. Convolutional neural networks for direct text deblurring. In *British Machine Vision Conference*, 2015. 2

[11] Xiaobin Hu, Wenhui Ren, Kaicheng Yu, Kaihao Zhang, Xiaochun Cao, Wei Liu, and Bjoern Menze. Pyramid architecture search for real-time image deblurring. In *IEEE International Conference on Computer Vision*, 2021. 3, 7

[12] Tae Hyun Kim, Kyoung Mu Lee, Bernhard Scholkopf, and Michael Hirsch. Online video deblurring via dynamic temporal blending network. In *IEEE International Conference on Computer Vision*, 2017. 3

[13] Zhe Jiang, Yu Zhang, Dongqing Zou, Jimmy Ren, Jiancheng Lv, and Yebin Liu. Learning event-based motion deblurring. *arXiv preprint arXiv:2004.05794*, 2020. 2, 3

[14] Meiguang Jin, Givi Meishvili, and Paolo Favaro. Learning to extract a video sequence from a single motion-blurred image. In *CVPR*, 2018. 3

[15] Adam Kaufman and Raanan Fattal. Deblurring using analysis-synthesis networks pair. *arXiv preprint arXiv:2004.02956*, 2020. 3

[16] Heewon Kim, Sungyong Baik, Myungsung Choi, Janghoon Choi, and Kyoung Mu Lee. Searching for controllable image restoration networks. In *IEEE International Conference on Computer Vision*, 2021. 3

[17] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *European Conference on Computer Vision*, 2012. 1, 2, 3

[18] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 3, 5, 6, 7

[19] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *IEEE International Conference on Computer Vision*, 2019. 3, 5, 6, 7

[20] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2, 3

[21] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 2, 3

[22] Jichun Li, Weiming Tan, and Bo Yan. Perceptual various motion deblurring with light global context refinement. In *IEEE International Conference on Computer Vision*, 2021. 3

[23] Boyu Lu, Jun-Cheng Chen, and Rama Chellappa. Unsupervised domain-specific deblurring via disentangled representations. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 3

[24] Thekke Madam Nimisha, Kumar Sunil, and AN Rajagopalan. Unsupervised class-specific deblurring. In *European Conference on Computer Vision*, 2018. 3

[25] Janne Mustaniemi, Juho Kannala, Simo Särkkä, Jiří Matas, and Janne Heikkilä. Gyroscope-aided motion
[26] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sunghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2019. 3

[27] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sunghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2019. 1, 2, 3

[28] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In IEEE Conference on Computer Vision and Pattern Recognition, 2017. 2, 3

[29] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In CVPR, 2017. 2, 3, 5, 6, 7

[30] Seungjun Nah, Sunghyun Son, and Kyoung Mu Lee. Recurrent neural networks with intra-frame iterations for video deblurring. In IEEE Conference on Computer Vision and Pattern Recognition, 2019. 3

[31] Thekke Madam Nimisha, Akash Kumar Singh, and Ambasamudram N Rajagopalan. Blur-invariant deep learning for blind-deblurring. In IEEE International Conference on Computer Vision, 2017. 3

[32] Jinshan Pan, Haoran Bai, and Jinhui Tang. Cascaded deep video deblurring using temporal sharpness prior. In IEEE Conference on Computer Vision and Pattern Recognition, 2020. 3

[33] Park 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, 2020. 3, 6

[34] Kuldeep Purohit and AN Rajagopalan. Region-adaptive dense network for efficient motion deblurring. arXiv preprint arXiv:1903.11394, 2019. 3

[35] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. arXiv preprint arXiv:1908.02197, 2019. 3

[36] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In IEEE Conference on Computer Vision and Pattern Recognition, 2019. 3

[37] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In ECCV, Springer, 2020. 1

[38] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In European Conference on Computer Vision, 2020. 2, 3

[39] Ziyi Shen, Wei-Sheng Lai, Tingfa Xu, Jan Kautz, and Ming-Hsuan Yang. Deep semantic face deblurring. In IEEE Conference on Computer Vision and Pattern Recognition, 2018. 2

[40] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In IEEE International Conference on Computer Vision, 2019. 2, 3, 7

[41] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In IEEE Conference on Computer Vision and Pattern Recognition, 2017. 2, 3

[42] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. arXiv preprint arXiv:2004.05343, 2020. 3

[43] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In IEEE Conference on Computer Vision and Pattern Recognition, 2015. 3

[44] Libin Sun and James Hays. Super-resolution from internet-scale scene matching. In IEEE International Conference on Computational Photography, 2012. 2, 3

[45] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In IEEE Conference on Computer Vision and Pattern Recognition, 2018. 3, 5, 6, 7, 8

[46] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. EDVR: Video restoration with enhanced deformable convolutional networks. In IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2019. 3

[47] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRgcn: Enhanced super-resolution generative adversarial networks. In ECCV, 2018. 3, 4

[48] Xiangyu Xu, Jinshan Pan, Yu-Jin Zhang, and Ming-Hsuan Yang. Motion blur kernel estimation via deep learning. IEEE Transactions on Image Processing, 27(1):194–205, 2017. 3
[49] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 3

[50] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *CVPR*, 2019. 3, 5, 6, 7, 8

[51] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 3, 7

[52] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Wei Liu, and Hongdong Li. Adversarial spatiotemporal learning for video deblurring. *IEEE Transactions on Image Processing*, 28(1):291–301, 2018. 3

[53] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Bjorn Stenger, Lin Ma, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 3, 5, 6, 7

[54] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 3

[55] Shangchen Zhou, Jiawei Zhang, Jinshan Pan, Haozhe Xie, Wangmeng Zuo, and Jimmy Ren. Spatiotemporal filter adaptive network for video deblurring. In *IEEE International Conference on Computer Vision*, 2019. 3

[56] Shangchen Zhou, Jiawei Zhang, Wangmeng Zuo, Haozhe Xie, Jinshan Pan, and Jimmy S Ren. Davanet: Stereo deblurring with view aggregation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2