IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 21, 2020, accepted March 28, 2020, date of publication April 2, 2020, date of current version April 20, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2985220*

# A Motion Deblur Method Based on Multi-Scale High Frequency Residual Image Learning

**KENG-HAO LIU[1], (Member, IEEE), CHIA-HUNG YEH[2,3], (Senior Member, IEEE), JUH-WEI CHUNG[1], AND CHUAN-YU CHANG[4], (Senior Member, IEEE)**

[1]Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-sen University, Kaohsiung 80424, Taiwan
[2]Department of Electrical Engineering, National Taiwan Normal University, Taipei 10610, Taiwan
[3]Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung 80424, Taiwan
[4]Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu 64002, Taiwan

Corresponding author: Chia-Hung Yeh (chyeh@ntnu.edu.tw)

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 108-2218-E-110-002-, Grant MOST 108-2218-E-003-002-, and Grant MOST 108-2221-E-110-018-.

**ABSTRACT** Non-uniform blind deblurring of dynamic scenes has always been a challenging problem in image processing because of the diverse of blurring sources. Traditional methods based on energy minimization cannot make accurate kernel estimation. It leads to that some high frequency details cannot be fully recovered. Recently, many methods based on convolution neural networks (CNNs) have been proposed to improve the overall performance. Followed by this trend, we first propose a two-stage deblurring module to recover the blur images of dynamic scenes based on high frequency residual image learning. The first stage performs initial deburring with the blur kernel estimated by the salient structure. The second stage calculates the difference of input image and initially deblurred image, referred to as residual image, and adopt an encoder-decoder network to refine the residual image. Finally, we can combine the refined residual image with the input blurred image to obtain the latent image. To increase deblurring performance, we further propose a coarse-to-fine framework based on the deblurring module. It performs the deblurring module many times in a multi-scale manner which can gradually restore the sharp edge details of different scales. Experiments conducted on three benchmark datasets demonstrate the proposed method achieves competitive performance of state-of-art methods.

**INDEX TERMS** Image deblurring, dynamic blur, non-uniform blind deblurring, deep learning.

## I. INTRODUCTION

Blind image deblurring (BID) has been received much attention and made a significant progress in the recent years [1]–[38]. The goal of image deblurring is to remove the blurring artifacts caused by object motions or camera shakes. More specifically, given an input blurry image, it aims to estimate unknown latent image when blur kernel is unknown. It implies that it is an ill-posed problem. The blur kernel $k$ must be estimated in advance some constraints or prior knowledge before running deconvolution process. Many deblurring methods have been proposed in the past. In the early stage, people devoted to estimate the blur kernel accurately with different types of priors [1]–[5]. Later, some people utilized salient structure to estimate the blur kernel [4]–[9].

Among the topics in BID, motion deblur is the one received wide attention. Motion deblur is an important topic in the problems of low-light and texture images. It is an image degradation phenomenon caused by various sources, such as camera shake and the movement of objects under a longer exposure. The blur kernel is usually spatially non-uniform, and difficult to be estimated. Ideally, to remove the motion blur, we must consider the all the blurring sources simultaneously. However, estimating an individual blur kernel for each pixel or region is infeasible because it would result in excessive computational burden. Therefore, many past researches only focused on single cases [10]–[13]. For instance, methods in [10], [13] assumed that the motion blur is caused by pure camera shake. They belong to global deblurring methods, and are not applicable for the image scenes

The associate editor coordinating the review of this manuscript and approving it for publication was L. Zhang.


This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

containing non-uniform blurring components. Kim *et al.* [14] combined segmentation and removal of non-uniform blurred images to restore a sharp image for dynamic scene. The same author [15] assumed that the blur kernel is roughly locally linear, and proposed a method that can estimate latent image and blur kernel at the same time. However, the performance of those methods depends on the quality of kernel estimation. In the cases of discontinuous motions and occlusions of objects, the blur kernel cannot be estimated correctly, so that the recovered image contains ringing artifacts.

Recently, as the rise of deep learning, convolutional neural networks (CNNs) were applied to resolve many tough problems in many areas such as image restoration [16], speech recognition [17], and fault diagnosis [18], [19]. The CNN methods were also applied to blur kernel estimation [20]–[23]. Those methods achieve satisfactory results but still have some drawbacks. For example, due to the lack of real-life blurred images and ground truths, they used synthetic data generated by self-defined blur kernels for training CNNs. It results in that the learned models can only handle the problems of spatially uniform or certain specific cases. The simple assumption they adopted also limits the capability of restoring the real-world blurred images. Later, [23]–[28] proposed other CNN-based methods which restore sharp images directly without estimating the blur kernel in advance. They integrate encoder and decoder networks into a single framework, and adopt end-to-end training strategy to make deblurring task simpler. Other deep neural structures, such as recurrent neural network (RNN) or generative adversarial network (GAN), were applied to deal with deblurring problem [29]–[31].

Those methods have become the most representative methods nowadays because of topnotch performance. However, most end-to-end methods require large amount of training samples to learn the relationship between shape and blurred images. In most cases, we cannot collect adequate training materials to well learn the model. Under such circumstances, introducing human-defined rules under a certain assumption into the deblurring framework can narrow down the depth of learning and reduce regression complexity so as to increase the accuracies of deblurring. For instance, it is known that the low-frequency components of a sharp image and the blurred one are similar. The main difference is the ''edge'' of object. It suggests that we only need to build a CNN to learn the rules of edge components instead of full image contents. There are several advantages about this way: 1. It can reduce the learning complexity of CNN, because the spatial contents involved in the training are limited to the edge components. 2. It can eliminate the artifacts caused by wrong edge estimation more effectively, such as ringing effects. 3. It is allowed to use relatively smaller amount of training samples.

Inspired by this idea, this paper proposes a multi-scale deblurring method based on residual image learning. This method adopts a two-stage deblurring module as the basic unit. The objective of the first stage is to recover low-frequency components while preserving high-frequency information and the second stage aims to refine high-frequency edge information. More specifically, the first stage performs initial deblurring with the blur kernel estimated by the salient structure of the input image. The salient structure is predicted by using a context aggregation network with dilated convolution units that can preserve contextual information of image. In the second stage, we define the difference of input image and the initially deblurred image as ''residual image''. Next, a residual image refinement process is performed by a pre-trained encoder-decoder network with residual image and high frequency image as inputs. Finally, the latent image can be reconstructed by combining the refined residual image with the input blur image. Since the residual image may contain edge details of different scales, it is impossible to recover all the details by using the deblurring module one time. To overcome this problem, we further design a multi-scale procedure to recover the details in coarse-to-fine manner by performing deblurring module multiple times from low to high resolution. Theoretically, this procedure can perform deblur for individual region/object so as to improve the ability of removing non-uniform blur.

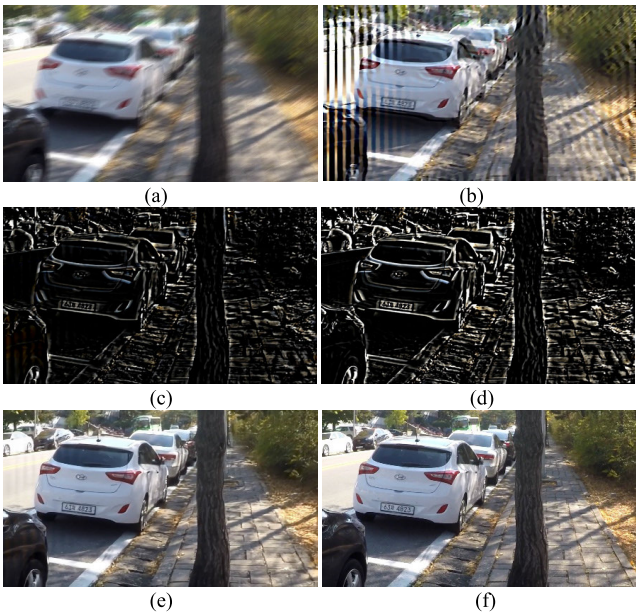


**FIGURE 1.** A deblurring example of the proposed method. (a) Input blurry image. (b) The initial deblurring result. (c) Residual image. (d) The modified residual image learned by a multi-scale refinement process. (e) The restored image obtained by combining (a) with (d). (f) Ground truth.

The experiments conducted on three benchmark datasets show that the proposed method is applicable for the various kinds of dynamic scene. Both quantitative and qualitative analysis shows that our method can achieve the state-of-art performance. Our contributions can be summarized as follows:

- We propose a novel concept for deblurring, which separates deblurring into low-frequency content recovery and high frequency information refinement.

- To fulfill the concept, we design a deblurring module based on deep learning. It consists of an initial deblurring method and a residual image refinement process.
- A multi-scale strategy is proposed to refine the edge details of different scales, and enhance the capability of non-uniform deblurring.

The remainder of this paper is organized as follows. The related work is introduced in Section II. Section III describes the proposed methodology. The dataset and experimental settings as well as the experimental results are demonstrated in Section IV. Finally, the conclusion is drawn in Section V.

## II. RELATED WORK

The statistical or image priors were first applied to blind deblurring. Fergus *et al.* [1] modeled the deblurring problem by assuming that a natural clear image has heavy-tailed distribution. They make a prior assumption that the gradient distributions of natural images are heavy-tailed, and estimate the blur kernel under a Bayesian framework. Xu *et al.* [2] used the $L_0$ constraint on image gradients to estimate the blur kernel. Krishnan *et al.* [3] proposed the normalized sparse prior to reduce the cost in image regularization, where the hyper-laplacian priors are used for image recovery. Pan *et al.* [4] utilized low rank prior to obtain the prior conditions of real data and estimate the blur kernel via Gaussian regularization. However, the design of those methods may cause the noise issue inside the blur kernel. To resolve it, the hysteresis threshold is used to remove the noise, but the details of blur kernel are also lost.

To suppress noise without losing the details, Xu and Jia [5] proposed an iteration-based approaches based on iterative support detection (ISD) and spatial priors to refine the kernel. Cho and Lee [6] proposed a nonparametric patch prior to the simulation of edges and angles. These work confirmed that the restoration quality will be poor if the blurry images lack of rich texture.

Cai *et al.* [32] removed complex motion blur by introducing a sparse regularization into original image and blur kernel. Machaeli and Irani [34] utilized the internal patch to deblur the images containing repetitive pattern. This method works well with repeated patches in the image, but fails in the case of lacking patterns in the image. Pan *et al.* [35] used dark channel priors for image deblurring, but the performance would degrade when the blur images contain noise.

On the other hands, some works used the edge distribution for blur kernel estimation [6], [7], [33], [36], [37]. For example, Cho and Lee [6] used bilateral and shock filters to find the sharp edges and used them to estimate blur kernel in a coarse-to-fine manner. However, the edge recovered from the blurry image is not necessarily effective to estimate the kernel. Xu and Jia [5] developed a method of selecting the amount of information for deblurring. Sun *et al.* [33] modeled image edge primitives using patch priors for blur kernel estimation. The main drawback of this category of approaches is that they heavily rely on the image filtering techniques and the ways to restore sharp edges.

Recently, the convolutional neural networks (CNNs) were applied to image deblurring [28]–[32], [38], [39]. For instance, Schulder *et al.* [40] proposed an integrated convolutional neural network for kernel estimation. However, this method needs to train different sizes of neural networks for the blur kernels with different sizes. Xu *et al.* [37] proposed a method for learning edge aware filters using a convolutional neural network. But this method cannot directly be applied to recover salient edges from blurred images to estimate the kernel.

There were some studies about motion blur removal. The studies [10]–[13] were designed to deal with the motion blur images caused by camera shake. Nah *et al.* [25] proposed an end-to-end convolutional neural network for real motion blur which combined with coarse-to-fine strategy, and [26], [27] also proposed an end-to-end architecture constructed by generative adversarial network. Those methods do not require kernel estimation but directly restore sharp image. In general, CNN-based methods achieved better results than traditional ones. However, the restored images still are not good enough in detail presentation. To further improve the performance, Kupyn *et al.* [47] and Miao *et al.* [48] proposed an end-to-end approach based on conditional generative adversarial network (cGAN).

## III. PROPOSED METHOD

In this study, we assume that the low-frequency components of blurred image $B$ and the latent image $I$ are similar. If we can accurately estimate their difference (called residual image, $R$), the latent image $I$ can be estimated by

$$I = R + B. \tag{1}$$

Based on this, we propose a two-stage deblurring module, as shown in Fig.2. It consists of two stages: Stage I performs initial deblurring, which roughly recovers low-frequency contents and obtain a temporary restored image $\tilde{I}$; Stage II performs residual image refinement, which refines the high-frequency details of residual image $\tilde{R}$ with the high-frequency image $H$. Finally, the latent image $I$ is obtained by combining refined residual image $R$ with $B$. In order to recover the image details of different scales, we further design a multi-scale strategy that performs the deblurring module in a successive way with different image scales.

The details of the deblurring module as well as multi-scale technique are introduced as follows.

### A. STAGE I: INITIAL DEBLURRING

We refer to [39] to develop our initial deblurring method. It consists of three steps: Salient structure prediction, blur kernel estimation, and image restoration. Fig.3 shows the flowchart of initial deblurring method.

#### 1) SALIENT STRUCTURE PREDICTION

We build a 9-layer CNN network to predict the salient structure $S$. The network architecture is shown in Table 1.

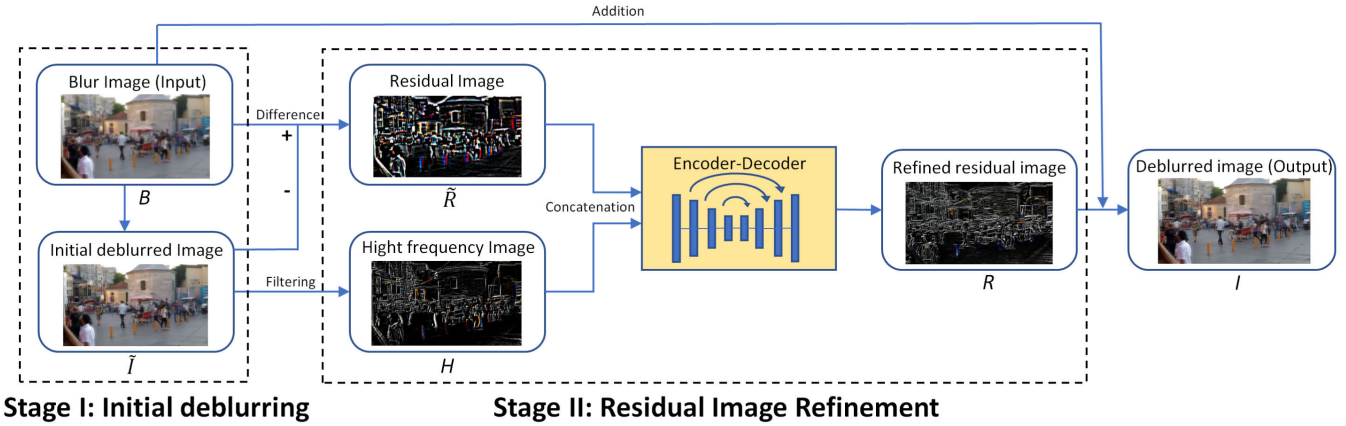


**FIGURE 2.** The architecture of the proposed two-stage deblurring module.

Differ from [39], we adopt multi-scale context aggregation network (CAN) [41] for better salient structure prediction.

The CAN aggregates multi-scale contextual information without losing resolution, so it can preserve more fine details of salient edges than general FCNs. Assume $\{L^0, L^1, L^2,.., L^l\}$ present each layer in the network. The dimensions of input layer $L^0$ and output layer $L^l$ are $h$ x $w$ x 3, the intermediate layers $L_i$ ($0< i < l$) are $h$ x $w$ x $d$, where $d$ denotes the number of feature maps. The response of each layer is given by:

$$L_n^i = \sigma\left(\Psi\left(\sum_{m=1}^{d}\left(L_m^{i-1} *_{df} K_{m,n}^i\right) + b_n^i\right)\right) \quad (2)$$

where $L_n^i$ and $L_m^{i-1}$ present the feature maps of $i$-th and ($i$-1)-th layers, $K_{m,n}^i$ is a 3 × 3 convolution kernel, $b_n^i$ is the bias, and $*_{df}$ is dilated convolution operator. $\sigma$ and $\Psi$ are leaky rectified linear unit (LReLU) and adaptive normalization, respectively.

Given the paired training data ($B_i$,$I_i$) (i=1,2,…,$N$), the network is trained by minimizing the reconstruction error of the gradient of salient structure

$$Loss\left(\nabla S_i\right) = \frac{1}{N}\sum_{i}^{N}\left\|\nabla S_i - \nabla S_i^*\right\|_2^2 \quad (3)$$

where $S_i$ is the salient output of $B_i$, $S_i^*$ is the target salient map simulated by applying $L_0$ smoothing filter on ground truth image $I_i$, and $\nabla$ is gradient operator.

### 2) BLUR KERNEL ESTIMATION

Given the salient structure $\nabla S$, we can use the minimization energy function to estimate blur kernel $k$

$$\min_k \|\nabla S * k - \nabla B\|_2^2 + \gamma \|k\|_2^2, \quad (4)$$

where $\gamma$ is regularization parameter. Then the latent image $I$ can be estimated by

$$\min_I \|I * k - B\|_2^2 + \lambda_1 \|\nabla I\|_0 + \lambda_2 \|\nabla I - \nabla S\|_2^2, \quad (5)$$

where $\lambda_1, \lambda_2$ are regularization parameters. Based on [39], (5) can be reformulated by using half-quadratic regulation and introducing an auxiliary variable $\mathbf{w}$ to replace $k$

$$\min_{I,\mathbf{w}} \|I * k - B\|_2^2 + \lambda_1 \|\mathbf{w}\|_0 + \lambda_2 \|\nabla L - \nabla S\|_2^2 + \lambda_3 \|\mathbf{w} - \nabla I\|_2^2, \quad (6)$$

where $\lambda_3$ is a scalar weight. The solution of (6) can be solved by minimizing $I$ and $\mathbf{w}$ in turn. Once $\mathbf{w}$ is fixed, $I$ can be obtained in the frequency domain

$$I = F^{-1}\left(\frac{F(L)\,\bar{F}(k) + \lambda_3 F(w)\,\bar{F}(\nabla) + \lambda_2 F_s}{F(k)\,\bar{F}(k) + (\lambda_2 + \lambda_3)\bar{F}(\nabla)F(\nabla)}\right). \quad (7)$$

where $F(\cdot)$ and $F^{-1}(\cdot)$ denote the Discrete Fourier Transform (DFT) and Inverse DFT (IDFT), $F_s = \bar{F}(\partial_x)F(\partial_x)F(S_x) + \bar{F}(\partial_y)F(\partial_y)F(S_y)$, $\partial_x$, $\partial_y$ are the vertical and horizontal derivative operators, and $\bar{(\cdot)}$ is complex conjugate operator. Similarly, when $I$ is fixed, the optimized $\mathbf{w}$ can be

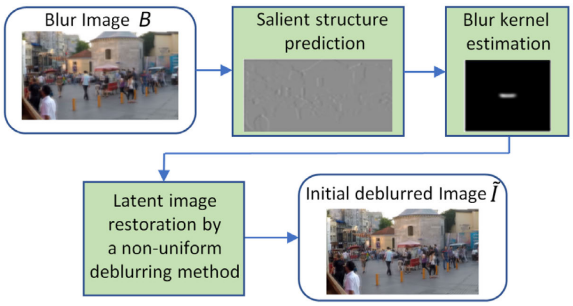


**FIGURE 3.** Overview of the initial deblurring method.

**TABLE 1.** The CNN network for salient structure prediction.

| Layer | Convolution | Dilation | Channel |
|---|---|---|---|
| 1 | 3x3 | 1 | 24 |
| 2 | 3x3 | 2 | 24 |
| 3 | 3x3 | 4 | 24 |
| 4 | 3x3 | 8 | 24 |
| 5 | 3x3 | 16 | 24 |
| 6 | 3x3 | 32 | 24 |
| 7 | 3x3 | 64 | 24 |
| 8 | 3x3 | 1 | 24 |
| 9 | 1x1 | 1 | 3 |

obtained by

$$\mathbf{w} = \begin{cases} \nabla L, & |\nabla L|^2 \geq \frac{\lambda_1}{\lambda_3}, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

#### 3) SHARP IMAGE RESTORATION

Once the blur kernel $k$ is estimated, we can utilize a common non-blind deblurring method to restore $I$, such as hyper-Laplacian prior method [3]. Although using hyper-Laplacian prior can recover a sharp latent image with main structure and fewer artifacts, but it would lose fine details. Another method, called isotropic total variation (TV) norm [4], can preserve fine details, but may preserve noise and ringing effect at the same time. To balance main structure and fine details, we adopt both hyper-Laplacian prior and isotropic TV norm to recover $I$. Firstly, the sharp image $I_1$ is acquired by solving

$$\min_{I_1} \|I_1 * k - B\|_2^2 + \rho \|\nabla I_1\|^{\alpha}, \tag{9}$$

where $\alpha$ is a smooth parameter set from 0.5 to 0.8, and $\rho$ is a weighting parameter. The smaller $\alpha$ is, the smoother $I_1$ will be obtained. Secondly, the latent image $I_2$ is obtained by TV-$L_2$ minimization

$$\min_{I_2} \|I_2 * k - B\|_2^2 + \mu \|I_2\|_{\mathrm{TV}},$$
$$\|I_2\|_{\mathrm{TV}} = \sqrt{(\partial_x I_2)^2 + (\partial_y I_2)^2}, \tag{10}$$

where $\mu$ is a regularization parameter and $\|I\|_{TV}$ is isotropic TV norm. Finally, the $\tilde{I}$ is obtained by averaging $I_1$ and $I_2$.

$$\tilde{I} = (I_1 + I_2)/2. \tag{11}$$

### B. STATG II: RESIDUAL IMAGE REFINEMENT

The resulted image $\tilde{I}$ obtained by initial deblurring method may contain fake edges or ringing artifacts, particularly when the blurring was caused by spatially non-uniform factors. In order to remove them, we propose a residual image refinement process. The details are introduced as follows.

#### 1) INPUT FOR REFINMENT

Based on Fig.2, after the temporal residual image $\tilde{R} = \tilde{I} - B$ is obtained, the next step is to refine $\tilde{R}$ ro $R$. However, $\tilde{R}$ may lack the high-frequency details necessary for refinement task. Thus, we further include the high-frequency image $H$ into the refinement process. The $H$ is generated by taking the high-frequency layer of the results of applying guided filter [43] to $\tilde{I}$. Finally, we concatenate $\tilde{R}$ with $H$ as the high-frequency materials for the input of an encoder-decoder network for refinement process.

#### 2) DESIGN OF ENCODER-DECODER NETWORK

We adopt U-Net [44] to build our encoder-decoder network. The U-Net's structure includes down-sampling (contraction) and up-sampling (expansion) processes. To increase the learning ability, we adopt modified ResBlock [25] as the basic unit to enhance the depth of network. The network architecture is shown in Fig.4.

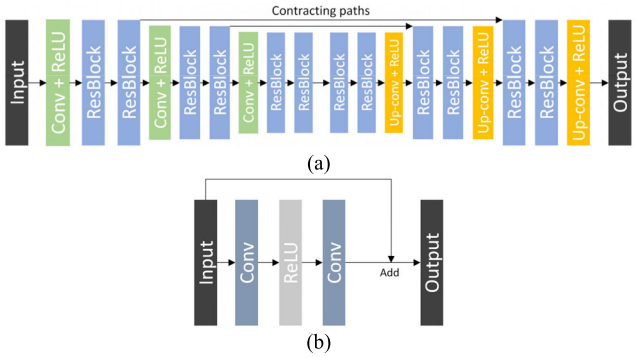


**FIGURE 4.** The encoder-decoder network for residual image refinement. (a) Network architecture. (b) The ResBlock.

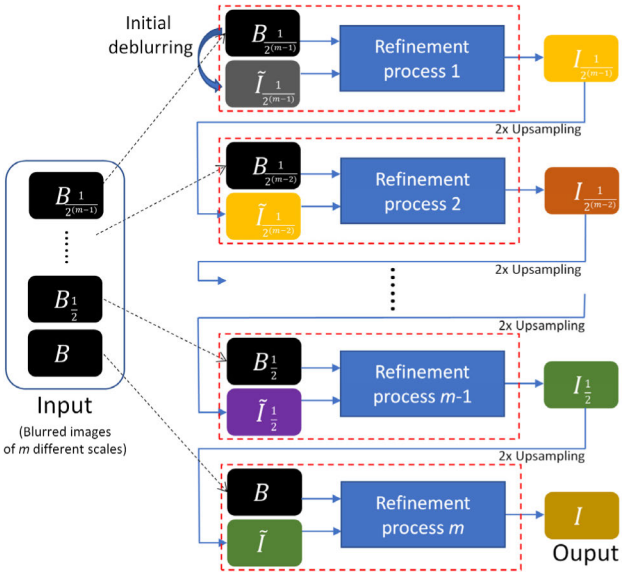


**FIGURE 5.** Flowchart of the proposed multi-scale refinement procedure. The red dotted frame presents the deblurring module.

### C. MULTI-SCALE DEBLURRING PROCESS AND TRAINING

#### 1) MULTI-SCALE DEBLURRING PROCESS

In order to restore the edge components of different scales, we design a coarse-to-fine deblurring procedure that uses the deblurring module multiple times. The concept is illustrated in Fig.5, where $m$ denotes the number of scales. According to the figure, there are $m$ blurred images of different scales used as input, and $m$ deblurring modules are operated. Each module performs deblurring in a specified resolution. From top to down, the resolution become twice successively, until original resolution is reached. Therefore, the top module performs deblurring in resolution of $\frac{1}{2^{m-1}}$, while the bottom one performs deblurring in original resolution. Through this design, the image details of different scales can be recovered progressively.

Take $m = 3$ for example, 1st module processes deblurring on quarter resolution, where $B_{\frac{1}{4}}$ presents the original blurred image with $\frac{1}{4}$ size and $\tilde{I}_{\frac{1}{4}}$ is the corresponding initially

deblurred image. After refinement, the deblurring output $I_{\frac{1}{4}}$ is then performed 2x upsampling to be the input of the second module, referred to as $\tilde{I}_{\frac{1}{2}}$. Similarly, the 2nd module operates deblurring task on 1/2 resolution, and the resulting deblurred image $I_{\frac{1}{2}}$ is up-converted for the input of 3rd module. It is noted that the initial deblurring only needs to be performed one time under this architecture.

### 2) TRAINING

We train the encoder-decoder network used in refinement process in a coarse-to-fine manner with multi-scale loss. The training is performed by using the prepared pair data of different scales.

Take $m = 3$ for example. Let $G$, $G_{\frac{1}{2}}$, and $G_{\frac{1}{4}}$ be the ground truth images of different scaled blurred image $B$, $B_{\frac{1}{2}}$, and $B_{\frac{1}{4}}$. For each training image, we run the multi-scale refinement with the same parameters of encoder-decoder, and obtain deblurred results $I$, $I_{\frac{1}{2}}$, and $I_{\frac{1}{4}}$,, and $I$. The we calculate the multi-scale loss defined by the average of loss values of $(I, G)$, $(I_{\frac{1}{2}}, G_{\frac{1}{2}})$ and $(I_{\frac{1}{4}}, G_{\frac{1}{4}})$. To preserve the details of edges, the edge-preserving loss function is adopted

$$Loss_{\text{total}} = Loss_{\text{ssim}} + Loss_{\text{E,g}}, \tag{12}$$

where $Loss_{\text{ssim}}$ is SSIM loss, and $Loss_{\text{E,g}}$ is two-directional gradient loss.

**SSIM loss** It adopts structural similarity to measure the difference of resulted image and ground truth. Assume $x$ and $y$ present the processed patch and the ground truth respectively. SSIM for pixel $p$ is defined as

$$\text{SSIM}\,(p) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{13}$$

where $\mu_x$ and $\mu_y$ are the mean of $x$ and $y$, $\sigma_x$ and $\sigma_y$ are the standard deviations, $\sigma_{xy}$ is the covariance, $C_1$ and $C_2$ are constants. The SSIM loss is then defined as

$$Loss_{\text{SSIM}}\,(P) = \frac{1}{N}\sum_{p\epsilon P} 1 - \text{SSIM}(p), \tag{14}$$

where $P$ is the input patch, $p$ is the pixel index, and $N$ is the number of pixels in $P$.

**Two-directional gradient loss** We also consider the loss of the gradients in horizontal and vertical directions

$$\begin{aligned} Loss_{E,g} = &\sum\nolimits_{w,h} \left\|(H_x(GT)_{w,h} - (H_x\,(I))_{w,h}\right\|_2 \\ &+ \left\|(H_y(GT)_{w,h} - (H_y(I))_{w,h}\right\|_2 \end{aligned} \tag{15}$$

where $H_x(\cdot)$ and $H_y(\cdot)$ are gradient operators in vertical and horizontal directions, $w$ and $h$ are spatial coordinates.

Finally, we summarize the proposed multi-scale deblurring method in Algorithm 1.

**Algorithm 1** Multi-scale Residual Image Learning-Based Deblurring

**Input:** Blur image $B$, Scale parameter $m$.
Find blur images $B_{\frac{1}{2}}, B_{\frac{1}{4}}, \ldots, B_{\frac{1}{2^{m-1}}}$
Find $\tilde{I}_{\frac{1}{2^{m-1}}}$ using initial deblurring method in Sec.III.A.
**for** $j = m\text{-}1 \rightarrow 0$ **do**
  Calculate $\tilde{R}_{\frac{1}{2^j}} = B_{\frac{1}{2^j}} - \tilde{I}_{\frac{1}{2^j}}$.
  Calculate $H_{\frac{1}{2^j}}$ by using guide filter [43].
  Refine $\tilde{R}_{\frac{1}{2^j}}$ to $R_{\frac{1}{2^j}}$ by using the encoder-decoder in Fig.4
  Calculate $I_{\frac{1}{2^j}} = R_{\frac{1}{2^j}} + B_{\frac{1}{2^j}}$.
  **if** j>0
    Up-sample $I_{\frac{1}{2^j}}$, then $\tilde{I}_{\frac{1}{2^{(j-1)}}} \leftarrow I_{\frac{1}{2^j}}$.
  **else**
    Set $I = I_{\frac{1}{2^j}}$.
  **end if**
**end for**

## IV. EXPERIMENTAL RESULTS

We conduct a comprehensive and comparative study as well as the performance evaluation for the proposed method. Three types of deblurring datasets are used in the experiments: GoPro dataset [25], Köhler *et al.* Dataset [45], and Lai *et al.* Dataset [46]. Several state-of-art methods, Xu *et al.* [2], Sun *et al.* [22], Whyte *et al.* [13], Gong *et al.* [38], Nah *et al.* [25], Ramakrishnan *et al.* [26], Kupyn *et al.* [47], and Miao *et al.* [48] are chosen for comparison. Three quantitative indices, PSNR, SSIM, and MS-SSIM are used for quantitative analysis. Our experiments are implemented on a PC with hardware setting: Intel i7-7700k CPU and NVIDIA 1080Ti GPU. The deblurring model is constructed and implemented on Tensorflow.

**TABLE 2. Quantitative comparison of four different settings performed on GoPro dataset.**

| Setting | PSNR | SSIM | MS-SSIM |
|---|---|---|---|
| Initial deblurring | 23.3121 | 0.7460 | 0.8652 |
| $m$=1 | 29.7921 | 0.8988 | 0.9561 |
| $m$=2 | 30.0014 | 0.9024 | 0.9578 |
| $m$=3 | **30.3125** | **0.9202** | **0.9763** |

**Training Data** Many related work utilize synthetic training data for training, that is, the blurred images are produced by using sharp image with self-defined blur kernels. Whether the blur kernel is defined as uniform or non-uniform, there is still a gap between the synthetic blur images and real-world ones. In this paper, we adopt GOPRO dataset [25] for training. Those training images were acquired by GoPro Hero4 operated in high speed 240fps mode at resolution $1280 \times 720$. For each sample, the blurred image is generated by the average of continuous frames in a specific video segment, and the center image of the segment is selected as sharp image (i.e., ground truth). The images produced in this way are closer to the real situation. GOPRO dataset provides 3214 pairs images, where 2103 pairs are used for training and 1111 pairs are used for test.


**FIGURE 6.** The deblurring results of using different settings.

The training of salient structure prediction network (Sec.III.A-1) is performed by end-to-end with the blur images in GOPRO dataset as the input and the corresponding L0-smoothing versions as the output (ground truth).

**Parameter setting** For the initial deblurring, the parameters $\lambda_1$, $\lambda_2$ and $\gamma$ are set to be 0.002, 0.002, and 1. For the residual image refinement, we adopt ADAM optimizer with the parameter $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The size of guided filter is set to be $15 \times 15$. For the training rate, we use polynomial decay rate to increase the stability of the model training, where the power is set to be 3, and training rate is set from $10^{-4}$ decreasing to 0. The size of guided filter is set to be $15 \times 15$. For training rate, we use polynomial decay to increase the stability of the model training, in which the power is set to be 3, and training rate is set from $10^{-4}$ decreasing to 0. The number of epochs is set at 4000. In each epoch, we set batch size = 15, and randomly select $256 \times 256$ residual and high-frequency patches as the inputs.

### A. ABLATION STUDY

In order to determine the parameter $m$ of the proposed method, we refer the [49], [50] to organize an extensive ablation study that compares the deblurring performance of the following four settings:

- Initial deblurring only
- Algorithm I with $m = 1$
- Algorithm I with $m = 2$
- Algorithm I with $m = 3$

**TABLE 3.** The quantitative performance of state-of-art methods and the proposed method in GoPro dataset experiment. (∗ denote the values declared in the corresponding paper).

| Method | PSNR | SSIM | MS-SSIM |
|---|---|---|---|
| Whyte [13] | 25.0933 | 0.8872 | 0.954 |
| Xu [2] | 25.1749 | 0.8911 | 0.9564 |
| Sun [22] | 24.709 | 0.856 | 0.9311 |
| Gong [38] | 27.2381 | 0.8972 | 0.9602 |
| Nah [25] | 28.4897/29.1* | 0.9012/0.9135 | 0.9646 |
| Ramakrishnan [26] | 28.241/28.9423* | 0.9131/**0.9220*** | 0.9679/0.9720* |
| Kupyn [47] | 29.013 | 0.9172 | 0.9728 |
| Miao [48] | 29.2168/29.221* | 0.9208/0.9184* | 0.9741 |
| OURS | **30.3125** | 0.9202 | **0.9763** |

**TABLE 4.** The quantitative performance of state-of-arts methods and the proposed method in Köhler dataset experiment. (∗ denote the values declared in the corresponding paper).

| Method | PSNR | SSIM | MS-SSIM |
|---|---|---|---|
| Whyte [13] | 24.4316 | 0.7426 | 0.8081 |
| Xu [2] | 26.8768 | 0.746516 | 0.8104 |
| Sun [22] | 25.7208 | 0.7328 | 0.7739 |
| Gong [38] | 26.1912 | 0.7335 | 0.7995 |
| Nah [25] | 26.1418/26.48* | 0.7450 | 0.8025/0.8079* |
| Ramakrishnan [26] | 26.0537/27.08* | 0.7468/0.7510* | 0.8005/0.8120* |
| Kupyn [47] | 26.2316 | 0.7482 | 0.8043 |
| Miao [48] | 25.91 | 0.7491 | 0.8113 |
| OURS | **27.1618** | **0.7515** | **0.8187** |

We conduct a quantitative comparison performed on GoPro dataset. The model of each setting is trained and tested individually. The testing results of PSNR, SSIM, and MS-SSIM are listed in Table 2. Based on the table, three indices increase as $m$ increase, and $m = 3$ setting produces the best performance. To illustrate the improvement provided by multi-scale technique in visual assessment, Fig. 6 demonstrates the deblurred results of four methods on a test image. In Fig. 6(b), the resulted image of initial deblurring contains ghosts in the surrounding area because the input blurred image has a non-uniform blur property. According to Fig. 6(c-e), those ghosts were gradually removed by the settings of $m = 1$, $m = 2$, and $m = 3$. It verifies that the multi-scale refinement process can remove the ghost phenomena and gradually restore the texture details.

In the following subsections, we choose $m = 3$ as the representative for the comparison with state-of-art methods.

### B. GOPRO DATASET

Table 3 tabulates the quantitative results of all the methods implemented on GOPRO dataset. It can be seen that our method almost outperforms all the state-of-art methods in PSNR, SSIM, and MS-SSIM. For the qualitative comparison, we choose five state-of-arts methods [13], [22], [25], [26], and [48] for comparison. Fig. 7 demonstrates the deblurring results of one test example. The method [13] is designed


**FIGURE 7.** Test sample 1 of GoPro dataset. (a) Input blur image. (b-f) State-of-art methods. (g) Our method. (h) Ground truth image.


**FIGURE 8.** Test sample 2 of GoPro dataset. (a) Input blur image. (b-f) State-of-art methods. (g) Our method. (h) Ground truth image.

for the scenario of simple camera shake. It can be seen in Fig. 7(b) that it produced small amount of ringing artifacts, and could not recover sharp content because it does not consider the relative motion and depth of different objects. Fig. 7(c) and Fig. 7(d) show the results of CNN-based end-to-end approaches [25] and [26]. We can find that method [25] which adopts multi-scale strategy, and method [26] which adopts GAN framework, significantly improved the image degradation caused by ringing effect and noise.

However, it is obvious that the end-to-end architecture seems to fail for the recovery of fine details. For example, the stitch in the bricks and the text on the license plate still lost some details. The method [44] is designed by conditional GANs. Their deblurring performance is shown in Fig. 7(e) and Fig. 7(f), respectively. Compared to [25] and [26], the issues of ringing effect and missing details are partially resolved. However, they could not restore the sharpness of object boundaries. Finally, the results of our proposed method are shown in Fig. 7(g). It not only provided almost zero ringing artifacts while retaining satisfactory sharp details. Fig. 8 shows another test image. We can simply rank the deblurring results by: Ours > Miao [48] > Ramakrishnan [26] > Nah [25] > Sun [22] > Whyte [13].

### C. KÖHLER DATASET

Köhler dataset [45] consists of 4 sharp images and 12 non-uniform blurry ones for each of them. So this dataset provides 48 pairs images. Table 4 tabulates the quantitative results of all the methods in PSNR, SSIM, and MS-SSIM. Similarly, our methods significantly outperformed all the compared methods. For visual quality assessment, Fig. 9 demonstrates the results of one selected examples. Compared to GOPRO

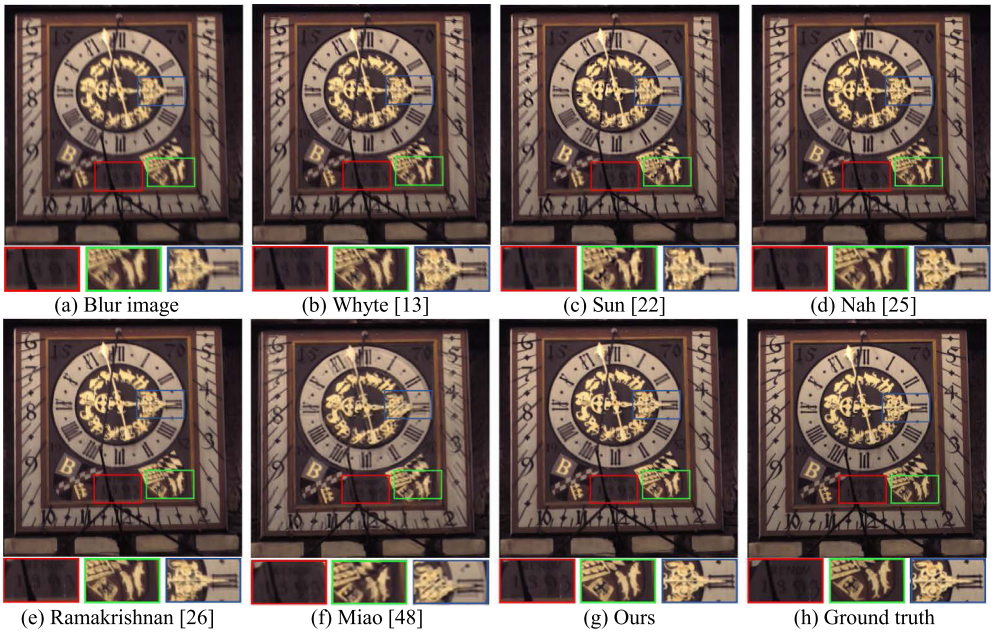


**FIGURE 9.** A test sample of *Köhler* of dataset. (a) Input blur image. (b-f) State-of-art methods. (g) Our method. (h) Ground truth image.

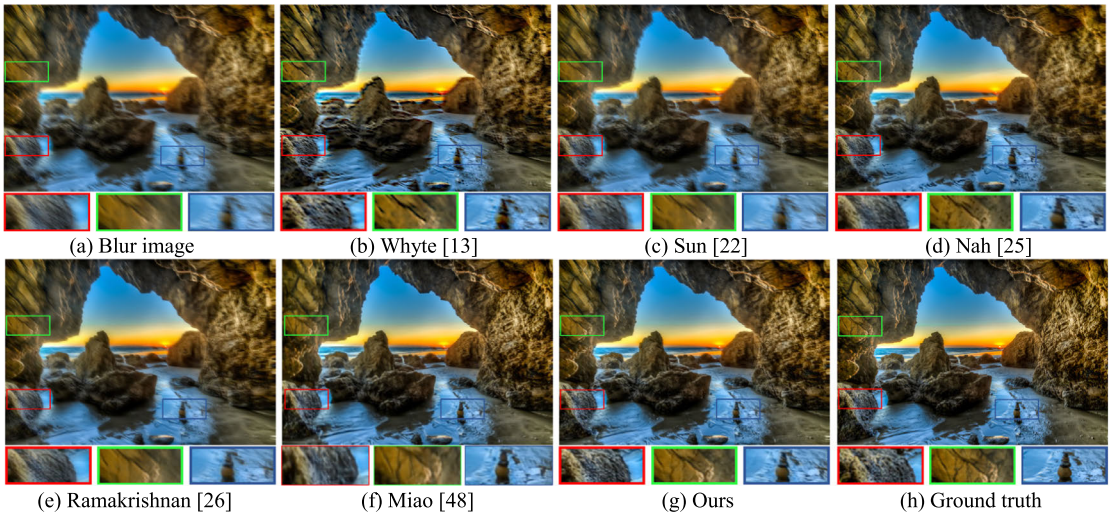


**FIGURE 10.** Test sample 1 of Lai dataset. (a) Input blur image. (b-f) State-of-art methods. (g) Our method. (h) Ground truth image.

dataset, the image content of Köhler dataset are more static. The overall deblurring performance of all selected methods is significantly better than that being applied on GOPRO dataset. However, we can still see the discrepancy between the different methods. In Fig. 9(b-e), it can be found that the results obtained by all methods contain certain degrees of sharpness and ghosting. For instance, the result of [48] provides sharp image while suffering from the most serious ringing effect, while the result of [26] has less ringing effect but loses sharpness. As shown in Fig. 9(g), our method not only eliminates the ringing artifacts but retains reasonable level of sharpness. It again confirms the superior restoration capability of our method.

### D. LAI DATASET

Lai dataset [46] is generated by convolving non-uniform blur kernels with several common degradations. It consists of 25 shape images and 100 non-uniformly blurred ones.

Since it was reported that the blurred images and the ground truth images are not aligned [25], we ignore the quantitative analysis in this section and only the visual qualitative comparison is presented. The main feature of Lai dataset is its diversity of subject. It includes many classes of static images, such as portrait, text, landscape, and building images. Therefore, using this dataset can reflect the pros and cons between our method and state-of-arts in different aspects. Figs. 10-12 shows the deblurring results of all the methods


**FIGURE 11.** Test sample 2 of Lai dataset. (a) Input blur image. (b-f) State-of-art methods. (g) Our method. (h) Ground truth image.


**FIGURE 12.** Test sample 3 of Lai dataset. (a) Input blur image. (b-f) State-of-art methods. (g) Our method. (h) Ground truth image.

implemented on three selected test images including landscape, portrait, and text, respectively.

Fig. 10 shows the superior deblurring ability of our method. We can simply rank the result by: Ours > Nah [25] > Ramakrishnan [26] > Miao [48] > Whyte [13] > Sun [22]. In addition, Fig. 11 show a qualitative comparison conducted on a portrait photo. It can be seen that the restored images of [13] contain significant ringing artifacts around the foreground region (i.e., the human face). This is probably due to the inaccurate kernel estimation caused by the interference of background. On the other hand, methods [25], [26], and [48] did not suffer this issue. Although end-to-end methods have the advantages of suppressing the artifacts and noise, the clarity of the restored images is still limited, particularly for the restoration of the edges of individual objects. Method [48] provided a non-ghost image, but it still loses the overall sharpness. In contrast to the state-of-arts, our method focuses on the restoration of high-frequency components. According to Fig. 11(g), our method successfully recovered the details of foreground in acceptable sharpness level while suppressing artifacts.

Finally, Fig. 12 shows a qualitative comparison of all the methods on a non-uniformly blurry text image. Obviously, our method produced the best results. All the text lines recovered by our method are readable, while the central lines recovered by most state-of-arts methods are not distinguishable. According to Fig. 12(a), the blurry degree of the texts in left side is higher than that in the right side.

It leads to some interesting results, for instance, the results of all state-of-arts have a characteristic that right-side texts were better recovered than left-side texts. Although Fig. 12(f) seems to provide sharp texts in the top half of the image, the texts below the image are not clear. Surprisingly, our method achieves consistency in the degree of deblurring on all the texts in different spatial locations. It again confirms the ability of non-uniform deblurring of our method.

## V. CONCLUSION

In this paper, a motion deblurring method based on high-frequency residual image learning is presented. We first propose a two-stage deblurring module. The first stage performs initial blurring which roughly recovers the low-frequency components of the latent image, while the second stage eliminates the ringing artifacts and retains shape details under a refinement framework. In order to recover the details of different scales, as well as enhance the capability of removing non-uniform blurring contents, we further design a coarse-to-fine framework in terms of the proposed deblurring module. It can perform the deblurring module in a multi-scale manner so that the details can be recovered gradually. The experiments conducted on three benchmark datasets demonstrate that the proposed method outperforms the state-of-arts in both qualitative and quantitative ways. The future work would be designing an end-to-end approach that combines initial deblurring and refinement process into a single framework, reducing the model complexity to save the computing time, and extending the design concept of our method to other topics of image restoration.

## REFERENCES


[1] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, Jul. 2006.

[2] L. Xu, S. Zheng, and J. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 1107–1114.

[3] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. CVPR*, Jun. 2011, pp. 233–240.

[4] J. Pan, R. Liu, Z. Su, and G. Liu, "Motion blur kernel estimation via salient edges and low rank prior," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.

[5] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proc. ECCV*, 2010, pp. 157–170.

[6] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–8, Dec. 2009.

[7] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring text images via L0-regularized intensity and gradient prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2901–2908.

[8] S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen, "An augmented Lagrangian method for total variation video restoration," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3097–3111, Nov. 2011.

[9] J. Pan, R. Liu, Z. Su, and X. Gu, "Kernel estimation from salient structure for robust motion deblurring," *Signal Process., Image Commun.*, vol. 28, no. 9, pp. 1156–1170, Oct. 2013.

[10] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 2901–2908.

[11] S. Harmeling, M. Hirsch, and B. Scholkopf, "Space-variant single-image blind deconvolution for removing camera shake," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 829–837.

[12] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf, "Fast removal of non-uniform camera shake," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 463–470.

[13] O. Whyte, J. Sivic, and A. Zisserman, "Deblurring shaken and partially saturated images," in *Proc. ICCV Workshops*, 2011, pp. 745–752.

[14] T. H. Kim, B. Ahn, and K. M. Lee, "Dynamic scene deblurring," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3160–3167.

[15] T. H. Kim and K. M. Lee, "Generalized video deblurring for dynamic scenes," in *Proc. CVPR*, Jun. 2015, pp. 5426–5434.

[16] M. Zhang, Y. Yuan, F. Zhang, S. Wang, S. Wang, and Q. Liu, "Multi-noise and multi-channel derived prior information for grayscale image restoration," *IEEE Access*, vol. 7, pp. 150082–150092, 2019.

[17] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019.

[18] Y. Wu, B. Jiang, and N. Lu, "A descriptor system approach for estimation of incipient faults with application to high-speed railway traction devices," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 10, pp. 2108–2118, Oct. 2019.

[19] Y. Wu, B. Jiang, and Y. Wang, "Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains," *ISA Trans.*, Sep. 2019, doi: 10.1016/j.isatra.2019.09.020.

[20] A. Chakrabarti, "A neural approach to blind motion deblurring," 2016, *arXiv:1603.04771*. [Online]. Available: http://arxiv.org/abs/1603.04771

[21] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1439–1451, Jul. 2016.

[22] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," 2015, *arXiv:1503.00593*. [Online]. Available: http://arxiv.org/abs/1503.00593

[23] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2014, pp. 1790–1798.

[24] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*. [Online]. Available: http://arxiv.org/abs/1606.08921

[25] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 257–265.

[26] S. Ramakrishnan, S. P. A. Gangopadhyay, and S. Raman, "Deep generative filter for motion deblurring," Sep. 2017, *arXiv:1709.03481*. [Online]. Available: https://arxiv.org/abs/1709.03481

[27] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," 2017, *arXiv:1711.07064*. [Online]. Available: http://arxiv.org/abs/1711.07064

[28] S. Zhang, P. Li, Y. Meng, L. Li, Q. Zhou, and X. Fu, "A video deblurring algorithm based on motion vector and an encorder-decoder network," *IEEE Access*, vol. 7, pp. 86778–86788, 2019.

[29] B. Zhao, W. Li, and W. Gong, "Deep pyramid generative adversarial network with local and nonlocal similarity features for natural motion image deblurring," *IEEE Access*, vol. 7, pp. 185893–185907, 2019.

[30] S. Xie, X. Zheng, W.-Z. Shao, Y.-D. Zhang, T. Lv, and H. Li, "Non-blind image deblurring method by the total variation deep network," *IEEE Access*, vol. 7, pp. 37536–37544, 2019.

[31] J. Liu, W. Sun, and M. Li, "Recurrent conditional generative adversarial network for image deblurring," *IEEE Access*, vol. 7, pp. 6186–6193, 2019.

[32] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Framelet-based blind motion deblurring from a single image," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 562–572, Feb. 2012.

[33] L. Sun, S. Cho, J. Wang, and J. Hays, "Edge-based blur kernel estimation using patch priors," in *Proc. IEEE Int. Conf. Comput. Photogr. (ICCP)*, Apr. 2013, pp. 1–8.

[34] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *Proc. ECCV*, 2014, pp. 783–798.

[35] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1628–1636.

[36] N. Joshi, R. Szeliski, and D. J. Kriegman, "PSF estimation using sharp edge prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[37] L. Xu, J. Ren, Q. Yan, R. Liao, and J. Jia, "Deep edge-aware filters," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 1669–1678.
[38] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. van den Hengel, and Q. Shi, "From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur," 2016, *arXiv:1612.02583*. [Online]. Available: http://arxiv.org/abs/1612.02583
[39] J. Pan, W. Ren, Z. Hu, and M.-H. Yang, "Learning to deblur images with exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1412–1425, Jun. 2019.
[40] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," Jun. 2014, *arXiv:1406.7444*. [Online]. Available: https://arxiv.org/abs/1406.7444
[41] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122
[42] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
[43] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," May 2015, *arXiv:1505.04597*. [Online]. Available: https://arxiv.org/abs/1505.04597
[45] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling, "Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database," in *Proc. Eur. Conf. Comput.*, 2012, pp. 27–40.
[46] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang, "A comparative study for single image blind deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1701–1709.
[47] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. M. Deblurgan, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 8183–8192.
[48] H. Miao, W. Zhang, and J. Bai, "Aggregated dilated convolutions for efficient motion deblurring," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, San Diego, CA, USA, Jul. 2018, pp. 1–6.
[49] Z. Wang, B. Du, and Y. Guo, "Domain adaptation with neural embedding matching," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 13, 2019, doi: 10.1109/tnnls.2019.2935608.
[50] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.

**KENG-HAO LIU** (Member, IEEE) received the B.S. degree in mathematical sciences from National Chengchi University, Taipei, Taiwan, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, Baltimore County, Baltimore, MD, USA, in 2009 and 2011, respectively.

He is currently an Assistant Professor with the Department of Mechanical and Electromechanical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan. His current research interests include remote sensing image processing, medical imaging, computer vision, and deep learning.

**CHIA-HUNG YEH** (Senior Member, IEEE) received the B.S. and Ph.D. degrees from the Department of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan, in 1997 and 2002, respectively. He was an Assistant Professor, from 2007 to 2010, an Associate Professor, from 2010 to 2013, and a Professor, from 2013 to 2017, with the Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan. He is currently a Distinguished Professor with National Taiwan Normal University, Taipei, Taiwan, and the Vice Dean of the College of Technology and Engineering. He has coauthored more than 250 technical international conferences and journal articles and held 47 patents in USA, Taiwan, and China. His research interests include multimedia, video communication, three-dimensional reconstruction, video coding, image/video processing, and big data. He was an Associate Editor of the *Journal of Visual Communication and Image Representation*, the *EURASIP Journal on Advances in Signal Processing*, and the *APSIPA Transactions on Signal and Information Processing*. He has been an active TC Member of the IEEE Communication Society on Multimedia Communication, APSIPA, and IWAIT. He is also one of the founding members of ACM SIGMM Taiwan Chapter. He has served as the Program Co-Chair of the IEEE Big Data Multimedia 2016, IWAIT&IFMIA 2015, ICS 2014, ICICS 2013, APSIPA 2013, ICS2012, and the IEEE ISIC 2012, and Co-Chaired the IEEE-TW 2016/2015, the IEEE ICME 2014, CVGIP2012, the IEEE PCM2012, VCIP 2012, APSIPA 2012, CVGIP2011, and VCIP 2010. He has also been on the Best Paper Award committee of JVCI and APSIPA. He was a recipient of the 2007 Young Researcher Award of NSYSU, the 2011 Distinguished Young Engineer Award from the Chinese Institute of Electrical Engineering, the 2013 Distinguished Young Researcher Award of NSYSU, the 2013 IEEE MMSP Top 10% Paper Award, the 2014 IEEE GCCE Outstanding Poster Award, the 2015 APSIPA Distinguished Lecture, the 2016 NARLabs Technical Achievement Award: Superior Achievement Award, the 2017 IEEE SPS Tainan Section Chair, the 2017 Distinguished Professor Award of NTNU, and the IEEE Outstanding Technical Achievement Award (the IEEE Tainan Section). He became a Fellow of IET in 2017.

**JUH-WEI CHUNG** received the B.S. degrees in mechanical and electro-mechanical engineering from National Sun Yat-sen University, Kaohsiung, Taiwan, in 2017, where he is currently pursuing the master's degree. His research interests include image processing and deep learning.

**CHUAN-YU CHANG** (Senior Member, IEEE) received the M.S. degree in electrical engineering from National Taiwan Ocean University, Keelung, Taiwan, in 1995, and the Ph.D. degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 2000. From 2001 to 2002, he was with the Department of Computer Science and Information Engineering, Shu-Te University, Kaohsiung, Taiwan. From 2002 to 2006, he was with the Department of Electronic Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan, where he has been with the Department of Computer and Communication Engineering (later the Department of Computer Science and Information Engineering), since 2007. He is currently a Full Professor and the Dean of the Research and Development, National Yunlin University of Science and Technology. He is also the Chair of the IEEE Signal Processing Society Tainan Chapter, and an Associate Editor of the *International Journal of Control Theory and Applications*. His current research interests include neural networks and their application to medical image processing, wafer defect inspection, digital watermarking, and pattern recognition. In the above areas, he has more than 150 publications in journals and conference proceedings.