Two-Level Wavelet-Based Convolutional Neural Network for Image Deblurring

YEYUN WU, PAN QIAN, AND XIAOFENG ZHANG
Key Laboratory of Jiangxi Province for Image Processing and Pattern Recognition, School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China
Corresponding author: Xiaofeng Zhang (xfzhang1130@163.com)

ABSTRACT
Image deblurring aims to restore the latent sharp image from the blurred one. In recent years, some learning-based image deblurring methods have achieved significant advances. However, the tradeoff between the texture details and model parameters is still a crucial issue. In this paper, we propose a novel deblurring method based on two-level wavelet-based convolutional neural network (CNN), which embeds discrete wavelet transform (DWT) to separate the image context and texture information and reduces the complexity of calculation. Furthermore, we modify the Inception module by adding pixels-wise attention (PA) mechanism and channel scaling factor to make each convolution kernel have different weights, which increase the receptive field while significantly reduce the parameters of the module. Qualitative and quantitative evaluation on real-word and synthetic datasets shows that the deblurring performance of our method is comparable to the state-of-the-art algorithms. Moreover, compared to the traditional learning-based deblurring method, our model has fewer parameters.

INDEX TERMS
Image deblurring, two-level wavelet-based convolutional neural network (CNN), discrete wavelet transform (DWT), Inception module, pixel-wise attention (PA).

I. INTRODUCTION
Image blur, caused extrinsically or intrinsically by many factors (e.g., object motion, camera shake, defocusing), is one of the most common problems when we take photos. Blurred images not only seriously affect the quality of images, but also significantly reduce the performances of many computer vision applications, such as image recognition [1] and face detection [2]. Therefore, image deblurring has essential research significance. To solve this problem, researchers have proposed various methods [3]–[6], most of which are based on the following mathematical models:

\[ B = I \otimes K + n \]  

where \( B, I, K, \) and \( n \) denote blurry image, sharp image, blur kernel, and additive noise, respectively. And \( \otimes \) is the convolution operator. According to whether the blur kernel \( K \) is known, the family of image deblurring is divided into two types: non-blind deblurring and blind deblurring. However, this problem is highly ill-posed since there are always multiple pairs of sharp images \( I \) and blur kernel \( K \) corresponding to a single blurry image \( B \).

Most current single image deblurring methods [7]–[10] are performed on spatial domain data, to reconstruct pixel values as the output of the model. However, this way tends to produce under-deblurred (the deblurred images are still blurred) or over-deblurred (images contains many unpleasant artifacts) output, lacking some textural details. Besides, some recent works [1], [11]–[13] are based on deep learning to solve this problem. Jain [10] design a deep CNN to eliminate motion flow. Kupyn et al. [1] design a feature pyramid model based on the adversarial neural network to generate perceptually realistic results. However, since the weights of the CNN are spatially invariant, even though the blur is small, a larger image region is needed to increase the receptive field. Nah et al. [12] expand the receptive field by stacking \( 3 \times 3 \) convolution. The model has three scales and a total of 120 ResBlocks, which causes too many network parameters. Zhang et al. [13] use one CNN to learn pixel-wise weights for the recurrent neural networks (RNN) at every location to increase the receptive field. However, for complex motion blur, the deblurred image still lacks many details.
Due to various problems yet to be solved, image deblurring remains a challenging task. Wavelet transform (WT) has been proven to be an efficient tool, which can depict the contextual and textural information of an image at different levels [14]. This advantage motivates us to introduce WT into the image deblurring task. As shown in Figure 1: The approximate coefficients of different levels wavelet decomposition (i.e., the top-left patches in (b-c)) compress the context information of the image at different levels. The detail coefficients (i.e., the rest patches in (b-c)) reveal the structure and texture information of the image. Therefore, emphasizing high-frequency wavelet coefficients helps to restore texture details, while constrains low-frequency wavelet coefficient reconstruction to strengthen the consistency of context information. The combination of the two aspects can make the final deblurred image more photo-realistic.

To take full advantage of WT and CNN, we design a two-level WT-based CNN for image deblurring, which contains three subnetworks: embedding, two-level WT and reconstruction network. The embedding network takes the blurred image as input and uses one-level WT to represent it as a set of feature maps. Then, the modified Inception module with the large receptive field is used to extract potential features effectively. Moreover, we add a pixel-wise attention (PA) mechanism to select more important channels and features, which reduces the number of parameters. The two-level WT network again uses WT to reduce the scale of the original feature and the computation complexity. Besides, we use dense connections to reuse potential features, which improves the deblurring effect. The reconstruction network uses ResBlocks as the basic block to further deepen the model. Finally, we use the inverse wavelet transform (IWT) to restore the desired deblurred image. Experimental results show that our method can well capture the contextual structure information of the image while obtaining local texture details. The main contributions of our work can be summarized as follows:

- A novel WT-based approach is proposed for deep CNN-based single image deblurring. As far as we know, this is one of the few attempts to combine WT knowledge and deep learning in image deblurring domain.
- To make full use of WT, an effective and lightweight CNN (including the modified Inception module) is proposed, which can achieve a better tradeoff between deblurring effect and model parameters.
- Based on WT, the mixed wavelet-based loss and MSE loss are used to optimize the network in the sub-band image, which can make full use of the advantages of a wavelet function, and restore the texture details of the image.

The rest of this paper is organized as follows. In section II, a brief survey of the existing related work on image deblurring and WT is presented. Section III introduces the detailed framework of two-level wavelet CNN. To verify the deblurred performance of the proposed method, we conduct extensive fundamental experiments and ablation studies in Section IV. Finally, we conclude the paper in Section V.

II. RELATED WORKS

A. DEBLURRING RESEARCH

Over the years, people have proposed numerous deblurring methods, which can be categorized into three main strategies.
The earliest used strategy is the traditional method in digital image processing, which mainly adopt filtering [15], [16], WT [17], [18] and Fourier transform [19], [20] to recover the corrupted image. Tekalp and Pavlovic [15] propose a 2D space-variant Kalman filter method to restore linear, space-variant degraded images. Neelamani et al. [20] propose an effective hybrid Fourier-wavelet regularized deconvolution method that performs noise regularization in both the Fourier and wavelet domains. Most of these methods are aimed at digital images, and noise has a significant influence on the deblurring effect. These methods are no longer applicable when RGB images are prevalent, and blurring is exceptionally complicated.

The frequently used strategy is a mathematical model based on Eq (1). Pan et al. [21] find that the dark channel of a blurred image is less sparse. Enforcing the sparsity of the dark channel is helpful for blind deblurring on various scenarios. Besides, the sparsity of the dark channel introduces a non-linear non-convex optimization problem, and they introduce a linear approximation of the min operator to compute the dark channel. The principle of this method is the Bayesian framework: the true posterior probability \( P(I|B) \) is based on prior knowledge of the problem (i.e., \( P(I) \)) and is updated according to the compatibility of the given hypothesis and the evidence (i.e., the likelihood \( P(B|I) \)), and then use the maximum a posterior (MAP) to infer the statistical properties of \( I \) and \( K \). Similarly, Tikhonov [22] expect sharp image to have a small norm, so Tikhonov-miller regularizer is applied to stabilize the deblurring result. This type of method uses the Variational Bayesian methods to approximate the intractable integrals arising in the Bayesian framework and provides an analytical approximation of the posterior probability of sharp images \( I \). Since the proposed priors are only applicable to specific images, these methods can achieve good performance for certain specific blur, but the algorithms cannot be well promoted.

The third strategy is based on deep learning methods, and the current research on image deblurring is mainly concentrated in this area. Zhang et al. [23] present a basic multi-patch model, which deal with blurry images via a coarse-to-fine hierarchical representation and achieve an excellent deblurring effect. Lu et al. [24] present an unsupervised method for single-image deblurring based on disentangled representations. The disentanglement is achieved by splitting the content and blur features in a blurred image using content encoders and blur encoders, which achieves good performance on face, text and low-illumination images. Compared with other methods, these ideas are practical and novel, and the deblurring effect is significantly improved.

B. RELATED APPLICATION BASED ON WT

Wavelets are useful tools for analyzing image information because they can divide the image signal into multi-scale and directional sub-bands. Moreover, WT and IWT can replace the down-sampling and up-sampling of CNN. Therefore, WT has been widely applied in image processing tasks such as Super-Resolution [25], [26], image dehazing [28], and image deblurring [29], [30]. Min et al. [29] develop a novel recursive deep CNN improved by WT, WT is utilized to decompose and extract the frequency information of the blurred image, CNN eliminates or weakens the characteristics of high data redundancy and image smoothness caused by WT. Zhang and Hirakawa [30] propose double discrete wavelet transform, which dramatically enhances analyze, detect, and process capabilities of blur kernels and blurry images. Liu et al. [27] propose multi-level wavelet CNN, which embeds discrete wavelet transform to CNN, as a replacement for pooling operation.

III. PROPOSED METHOD

A. 2-D DISCRETE WAVELET TRANSFORM

Our model divides the image signal into multi-scale and oriented sub-bands based on WT. We choose Haar wavelet, for it is enough to get the context and texture information of different frequencies on the image. At the same time, the 2-D DWT is used to compute Haar wavelets.

The 2-D DWT operation includes filtering and horizontal downsampling by the 1-D filters \( \psi(x) \) and \( \varphi(x) \) to each column in the image and then using two filters to filter and vertically downsampling each row. Consequently, four sub-images \( I_{LL}, I_{LH}, I_{HL} \) and \( I_{HH} \) can be calculated. The 2-D DWT can be expressed as follows:

\[
\begin{align*}
\psi_{LL}(x, y) &= \phi(x) \varphi(y) \\
\psi_{LH}(x, y) &= \phi(x) \psi(y) \\
\psi_{HL}(x, y) &= \varphi(x) \psi(y) \\
\psi_{HH}(x, y) &= \varphi(x) \varphi(y)
\end{align*}
\]

where \( \phi(\cdot) \) and \( \varphi(\cdot) \) are 1-D scaling functions and wavelet function. In general, \( I_{LL} \) is the low-frequency information about horizontal, vertical and diagonal edges of the image, respectively. In contrast, IDWT uses the same filter and upsampling method to merge the four sub-images into the original image. Since DWT and IDWT involve down-sampling and up-sampling, once DWT or IDWT is used, the data size becomes a quarter or four times, and the number of channels becomes four times or a quarter. Figure 2 shows the illustration of DWT and IDWT. Similarly, for two-level DWT, each sub-band image \( I_{LL}, I_{LH}, I_{HL} \) and \( I_{HH} \) is further processed by WT to produce a decomposition result.

B. DETAIL STRUCTURE OF THE NETWORKS

As shown in Figure 3, our two-level wavelet CNN consists of 3 sub-networks: embedding, two-level WT and reconstruction network. The embedding network uses one-level DWT and modified Inception module to represent blurred images as a set of latent feature maps. The two-level WT network again utilizes WT to reduce the scale of features. Finally, the reconstruction network uses ResBlocks [12] and IDWT to restore the desired deblurred image.

The input of the embedding network is \( 3 \times h \times w \) blurred image. Through Eq. (2), we can get \( 12 \times \frac{h}{2} \times \frac{w}{2} \)
We use it as input for the modified Inception, as shown in Figure 4. In our Inception module, the size of the convolution kernel is $1 \times 1$, $3 \times 3$, $5 \times 5$, and $7 \times 7$, the corresponding padding number is $0$, $1$, $2$, $3$, and all the step size is $1$ so that each feature map has the same size as the input tensor. Besides, to reduce the number of parameters, we use channel scaling factor of $0.125$, $0.125$, $0.25$, $0.5$ make the large convolution kernel take up a more critical weight. Compared with the original Inception module, using this strategy will reduce the network parameters by $3.5$ times. Then we use PA attention adaptation to select more important channels and features.

The two-level WT network uses WT again, so the scale of the feature maps becomes $N_w \times \frac{4}{2} \times \frac{4}{2}$, which dramatically reduces the complexity of calculation and speeds up training. Moreover, since DWT is biorthogonal, it is guaranteed that this scheme can keep all information. This advantage is another reason why we use two-level DWT. Similarly, the base block is a modified Inception module. To alleviate the problem of vanishing gradients and reuse features, we use dense connections in this network. Finally, we use the IDWT to restore the scale of the feature maps to its original size.

The reconstruction network consists of a series of ResBlocks, and we do not use the Inception module. There are two reasons for this: First, the convolution kernel of the ResBlock module is $3 \times 3$, which significantly reduces the number of parameters. Second, the network does not need a large receptive field in the image reconstruction stage. We also add a skip connection between the embedding network and the reconstruction network, which is conducive to the transfer of image details and improves the deblurring effect.
C. LOSS FUNCTION

Image deblurring aims to learn the mapping function \( F_\theta(y) \) under the training dataset of fixed blurred/sharp image pairs. The common objective loss function for image deblurred is Mean Square Error loss (MSE loss) in the pixel-wise, which can be defined as:

\[
L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \| F_\theta(x_i) - y_i \|^2
\] (3)

where \( x_i \) and \( y_i \) denote the \( i \)-th input blurred/deblurred image and corresponding ground-truth sharp image, respectively. We use \( L_1 \) loss in the wavelet domain as the wavelet loss. Besides, we added SSIM loss to preserve edge details better. Therefore, our network consists of two loss functions:

\[
L_1(\theta) = \frac{1}{N} \sum_{i=1}^{N} \| LL(F_\theta(y_i), x_i) + LH(F_\theta(y_i), x_i) + HL(F_\theta(y_i), x_i) + HH(F_\theta(y_i), x_i) \|_1
\] (4)

where \( LL(\cdot), LH(\cdot), HL(\cdot), \) and \( HH(\cdot) \) represent the WT process at different frequencies, and

\[
L_{ssim}(\theta) = 1 - SSIM(F_\theta(y_i), x_i)
\] (5)

where SSIM(\cdot) can be calculated by structural similarity [37]. Finally, by combining the wavelet loss and SSIM loss, the network is jointly trained. Hence, our final loss term is:

\[
L_{total}(\theta) = L_1(\theta) + \lambda L_{ssim}(\theta)
\] (6)

where \( \lambda \) is a tradeoff parameter between \( L_1(\theta) \) and \( L_{ssim}(\theta) \).

IV. EXPERIMENTS

There are many existing deblurred datasets [12], [31], [32], we selected the real-word dataset GoPro and the synthetic dataset Köhler for qualitative and quantitative research of the model. And compared it with the state-of-the-art deblurring algorithm, including traditional algorithm [3], traditional algorithms combined with CNN [33], CNN-based deblurring [12], [34], and GAN-based method [35]. Besides, we also compared their model size and speed, which we will discuss in detail in Section C.

A. EXPERIMENT SETTING

1) TRAINING DATASETS

a: REAL-WORD DATASET

Nah et al. [12] provide a large-scale dataset with blurry/sharp pairs of realistic images, named GoPro, which is widely used in various deblurring task. Different from the earlier work [47]–[49], the blurry image is obtained by convolution of sharp image and blur kernel. In contrast, the blurry image corresponding to the sharp image in GoPro dataset is captured by Hero Black camera, so the whole dataset is closer to reality case. We select 2103 pairs for training, and the remaining 1111 pairs are used for the test.

b: SYNTHETIC DATASET

Köhler et al. [32] use different camera trajectories in the same scene to provide a dataset that contains 4 latent images and corresponding 12 different blurred images. This is a benchmark dataset widely used for single image deblurring algorithms. Our model is also tested on this dataset.

2) TRAINING DETAILS

In the three sub-networks, the number of channels is set to 64, 128, 256 pixels size. The tradeoff parameter \( \lambda = 1 \). Our experiments are conducted on a PC with Intel (R) Xeon (R) Gold 5120 CPU and two NVIDIA GTX 1080Ti GPU. The network is implemented on the Pytorch platform. To prevent our model from overfitting and augment the training data, the images are randomly cropped to 256 × 256 pixels size. The initial learning rate is set to \( 5 \times 10^{-5} \) and decreases with a decay rate of 0.3 every 400 epochs. Our network is optimized using ADAM [36]. The size of the training batch is 8, and the total training takes 4 × 10⁴ iterations to converge.

B. QUANTITATIVE AND QUALITATIVE EVALUATION ON DATASETS

1) EVALUATION ON REAL-WORD DATASETS

Figure 5 shows the deblurring visual quality of different state-of-the-art algorithms performance on real-word blurry dataset. For complex motion scenes, the traditional algorithm based on [3] has a poor deblurring effect, and the
deblurred image is still blurred, while the learning-based method has better deblurring effect. For simple motion scenes, the deblurring effect of traditional algorithms and learning-based algorithms is almost the same. In addition, CNN-based methods [47], [22] cannot eliminate large blur due to the limited receptive field and deblurring images...
TABLE 1. Quantitative deblurring performance comparison on the GoPro dataset, in terms of PSNR and SSIM.

| method            | Whyte et al. [3] | Nah et al. [12] | Sun et al. [33] | Zhao et al. [34] | Kupyn et al. [35] | ours       |
|-------------------|------------------|-----------------|-----------------|------------------|-------------------|------------|
| PSNR              | 24.53            | 28.49           | 25.31           | 30.90            | 29.55             | 30.75      |
| SSIM              | 0.846            | 0.917           | 0.851           | 0.910            | 0.934             | 0.913      |

FIGURE 6. Quantitative evaluations on the synthetic deblurring dataset [32].

will produce some tiny artifacts. Our method can restore the structure and details of the image very well. As shown in Table 1, for the real-word dataset GoPro. According to PSNR and SSIM, our method is superior to other traditional deblurring algorithms, for example, 6.22 PSNR and 6.7% SSIM higher than the traditional method [3], and 5.44 PSNR and 6.2% SSIM higher than the traditional algorithms combined with CNN [33]. Compared with other state-of-the-art learning-based methods, our method is comparable in PSRN or SSIM.

2) EVALUATION ON SYNTHETIC DATASETS

Our results are shown in Figure 6. In terms of visual quality, the effects of these deblurring algorithms are not much different for the blurred image synthesized by simple blur kernel, and the deblurring effect is significant. The visual effect of our method to deblur the image is almost the same as theirs, without much advantage.

The quantitative results are shown in Table 2. Compared to the GoPro dataset, both PSRN and SSIM have a certain degree of decline. Moreover, the learning-based deblurring
| method          | Whyte et al. [3] | Nah et al. [12] | Sun et al. [33] | Zhao et al. [34] | Kupyn et al. [35] | ours  |
|-----------------|------------------|-----------------|-----------------|-----------------|------------------|------|
| PSNR            | 24.43            | 26.48           | 25.22           | 26.54           | 26.72            | 26.83|
| SSIM            | 0.743            | 0.809           | 0.773           | 0.829           | 0.836            | 0.845|

| method          | Whyte et al. [3] | Nah et al. [12] | Sun et al. [33] | Zhao et al. [34] | Kupyn et al. [35] | ours  |
|-----------------|------------------|-----------------|-----------------|-----------------|------------------|------|
| time (sec)      | 700              | 15              | 1500            | 32              | 0.35             | 3.2  |
| size (MB)       | -                | 303.6           | 54.1            | 35.2            | 4                | 29.1 |

**FIGURE 7.** Deblurring qualitative evaluation of different Inception modules on GoPro dataset.

method is slightly better than the traditional method, which is similar to the GoPro, but the gap between traditional methods and CNN-based methods, our method has advantages in both PSRN and SSIM. For example: 2.4 PSNR and 10.2% SSIM higher than the traditional method [3], and 0.35 PSNR and 3.6% SSIM are higher than the CNN-based method [12].

**C. MODEL SIZE AND RUNNING TIME**

We also compare the model parameters and running time. As shown in Table 3, the traditional deblurring methods need to solve highly non-convex optimization problems [3], so the running time is relatively long. Although Sun et al. [33] develop CNN algorithms to solve motion blur, traditional
non-blind deblurring algorithms are required to generate the final sharp image, which increases the running time. The method in [12] use multi-scale CNN to estimate sharp images, and the calculation time is much less compared with traditional algorithms. However, the multi-scale strategy inevitably increases the model size and running time. The method proposed by Kupyn et al. [35] is not the best in qualitative and quantitative experiments. However, the model parameters are the smallest and the running time is the shortest, which has obvious advantages. Compared with the latest method [34], in addition to the deblurring effect, our method also has advantages in model size and running time.

D. ABLATION STUDY

1) ABLATION STUDY ON THE MODIFIED INCEPTION MODULE

To study the influence of the scale factor on the image deblurring effect, we conduct ablation studies on the modified Inception module. As shown in Table 4, the quantitative results of different scaling factor combinations are listed. With the increase of the proportion of large convolution kernels in models 1 to 4, the deblurring effect is also more significant. Model 4 and 5, compared with the original Inception module, our modified Inception module has greatly reduced parameters. In the case of sufficient parameters, the deblurring performance can be effectively improved by increasing the large convolution kernel’s weight. In our entire network, the scale factor has played a role in initially optimizing the weights of different convolution kernels, and the PA module adaptively selects more important channels and weights so that our network can achieve an excellent deblurring effect with a few parameters. Figure 7 shows the effect of different Inception modules on images with relatively simple blur condition.

2) ABLATION STUDY ON LOSS

To study the deblurring effect of different loss strategies, we conduct the ablation study on loss. Throughout the experiment, we use the MAE + SSIM loss function because the MSE loss may over penalize the pixel value error, resulting in an under-deblurred image, as shown in Figure 8. In order to

![FIGURE 8. Deblurring results by different losses. We chose two different images (one is complex blurred image and the other is simple blurred image) as the test.](image)
keep the structural details of the partial patch image the same between the deblurred image and the sharp image, we use MAE loss and SSIM loss. As shown in Table 5, the quantitative results of different loss combinations are listed. Through the deblurred performance, MAE + SSIM loss significantly improves the image quality indicators PSNR and SSIM. Qualitative evaluation shows that our baseline loss function can better deblur and retain more details.

3) ABLATION STUDY ON FREQUENCIES
To analyze the advantages of WT combined with CNN, we add the visualization contrast between four sub-bands of deblurred image and ground truth by using wavelet transform. Figure 9 shows the visual effect of the deblurred image in the frequency domain. We find that the structure and details of the deblurred image frequency domain are richer in some areas. The traditional method removes image blur from a mathematical point of view, and the learning-based method uses direct reconstruction of pixel values as output. Our method fully combines the advantages of WT and CNN. By using CNN to reconstruct wavelet coefficients, it avoids solving non-convex nonlinear problems. Wavelet transform can assist CNN to remove blur in the frequency domain and reduce model parameters. Our method has a good balance between deblurring performance and model parameters and provides a new idea for the study of image deblurring.

V. CONCLUSION
In this paper, we design a two-level wavelet deep CNN. Our network uses WT to decompose and extract the frequency information of the blurred image and uses CNN predict wavelet coefficients to help reconstruct deblurred images, making full use of WT and CNN’s advantages. In addition, our modified Inception module allows the large convolution kernel to occupy more channels through the channel scaling factor and uses the PA attention mechanism to select more important channels and features. The separation experiment proves that this method can optimize the network and significantly reduce the size of the model. Compared with the existing image deblurring methods, it achieves a right balance between the rich texture details of the deblurred image and the model parameters. We explored image deblurring in the wavelet domain and achieved satisfactory results.

REFERENCES
[1] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, “DeblurGAN: Blind motion deblurring using conditional adversarial networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, Jun. 2018, pp. 8183–8192.
[2] B. Lu, J.-C. Chen, C. D. Castillo, and R. Chellappa, “An experimental evaluation of covariates effects on unconstrained face verification,” IEEE Trans. Biometrics, Behav., Identity Sci., vol. 1, no. 1, pp. 42–55, Jan. 2019.
[3] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, “Non-uniform deblurring for shaken images,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., San Francisco, CA, USA, Jun. 2010, pp. 491–498.
[4] A. Gupta, N. Joshi, C. L. Zitnick, and M. F. B. Cohen Curless, “Single image deblurring using motion density functions,” in Proc. Eur. Conf. Comput. Vis., Berlin, Germany, 2010, pp. 171–184.
[5] L. Chen, F. Fang, T. Wang, and G. Zhang, “Blind image deblurring with local maximum gradient prior,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 1742–1750.

[6] L. Yang and H. Ji, “A variational EM framework with adaptive edge selection for blind motion deblurring,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 10159–10168.

[7] Y.-W. Tai, P. Tan, and M. S. Brown, “Richardson-lucy deblurring for scenes under a projective motion path,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1603–1618, Aug. 2011.

[8] N. Joshi, S. B. Kang, C. L. Zinnick, and R. Szelski, “Image deblurring using inertial measurement sensors,” *ACM Trans. Graph.*, vol. 29, no. 4, p. 30, Jul. 2010.

[9] Y. Yuan, W. Su, and D. Ma, “Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 3552–3561.

[10] E. S. Angel and A. K. Jain, “Restoration of images degraded by spatially varying point spread functions by a conjugate gradient method,” *Appl. Opt.*, vol. 17, no. 14, pp. 2186–2190, 1978.

[11] 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,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3806–3815.

[12] 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.

[13] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. W. H. Lau, and M.-H. Yang, “Dynamic scene deblurring using spatially variant recurrent neural networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2521–2529.

[14] S. Mallat, “Wavelets for a vision,” *Proc. IEEE*, vol. 84, no. 4, pp. 604–614, Apr. 1996.

[15] A. M. Tekalp and C. Pavlovic, “Space-variant and color image restoration using Kalman filtering,” in *Proc. IEEE Int. Symp. Circuits Syst.*, Pacific Grove, CA, USA, Sep. 1989, pp. 186–187.

[16] M. Hirsch, S. Sta, B. Scholkopf, and S. Harmeling, “Efficient filter flow for space-variant multiframe blind deconvolution,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 607–614.

[17] R. D. Nowak and M. A. T. Figueiredo, “Fast wavelet-based image deconvolution using the EM algorithm,” in *Proc. Conf. Rec. 35th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, Nov. 2001, pp. 371–375.

[18] J.-L. Starck, M. C. Nguyen, and F. Murtagh, “Image deblurring using a fast iterative method with local maximum gradient prior,” in *Comput. Vis. (ICCV)*, Venice, Italy, Sep. 2017, pp. 1698–1706.

[19] Y. Wu, Y. Pan, D. Sun, H. Pfister, and M.-H. Yang, “Deblurring images via dark channel prior,” *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 4797–4809.

[20] R. Neelamani, H. Choi, and R. Baraniuk, “ForWaRD: Fourier-wavelet regularization method,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 10217–10226.

[21] T. Guo, H. S. Mousavi, T. H. Vu, and V. Monga, “Deep wavelet prediction for image super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1100–1109.

[22] H. Huang, R. He, Z. Sun, and T. Tan, “Wavelet- SRNet: A wavelet-based CNN for multi-scale face super resolution,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 1698–1706.

[23] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, “Multi-level wavelet-CNN for image restoration,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Salt Lake City, UT, USA, Jun. 2018, pp. 8860–8869.

[24] Q. Yi, Y. Yang, and Y. Fu, “Wavelet U-Net and the chromatic adaptation transform for single image dehazing,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Taipei, Taiwan, Sep. 2019, pp. 2736–2740.

[25] C. Min, G. Wen, B. Li, and F. Fan, “Blind deblurring via a novel recursive deep CNN improved by wavelet transform,” *IEEE Access*, vol. 6, pp. 69242–69252, Nov. 2018.

[26] Y. Zhang and K. Hirakawa, “Blurring processing using double discrete wavelet transform,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 1091–1098.

[27] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang, “Deep video deblurring for hand-held cameras,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 237–246.

[28] 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,” *IEEE Trans. Image Process.*, pp. 12–27, 2010.

[29] J. Sun, W. Cao, Z. Xu, and J. Ponce, “Learning a convolutional neural network for non-uniform motion blur removal,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 769–777.

[30] Z. Zhao, B. Xiong, S. Gai, and L. Wang, “Improved deep multi-patch hierarchical network with nested module for dynamic scene deblurring,” *IEEE Access*, vol. 8, pp. 62116–62126, Mar. 2020.

[31] O. Kupyn, T. Martynik, J. Wu, and Z. Wang, “DeblurGAN-v2: Deblurring (Orders-of-Magnitude) faster and better,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 8877–8886.

[32] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980 [Online]. Available: http://arxiv.org/abs/1412.6980.

[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.