Computational Visual Media
DOI 10.1007/s41095-xxx-xxxx-x Vol. x, No. x, month year, xx–xx

Article Type (Research/Review)

Progressive edge sensing dynamic scene deblurring

Tianlin ZHANG[1], Jinjiang LI[2,3]✉, and Hui FAN[2,3]

Abstract Image deblurring in dynamic scene, since the blurring factors are formed by a combination of many reasons, it is a challenging task. In recent years, the use of multi-scale pyramid methods to recover potential high-resolution sharp images has been extensively studied. But because the cascade structure is insufficient in the recovery details, we have improved this. In this work, our network uses a progressive way to integrate data streams. We propose a new multi-scale structure and edge feature perception design. It is used to deal with blur changes in different spatial scales and enhance the sensitivity of network blur edges. The architecture is from coarse to fine in restoring the image structure, first performing global adjustment operations, and then performing local refinement operations. In this way, not only the global correlation is considered, but also the residual information is used to significantly improve the image restoration performance and enhance the texture details. Experimental results show that both quantitative and qualitative aspects can get better results than existing methods.

Keywords image deblurring, dynamic scene, multi-scale, edge feature.

1 Introduction

Due to the relative displacement of the object during the exposure process of the sensor, the image structure and texture are blurred and high-frequency details are degraded. It is not conducive to image processing tasks such as target detection and text recognition. Blind deblurring in dynamic scenes is a basic and low-level ill-posed inverse task in computer vision, and it is also a basic component of many practical applications. Its purpose is to recover the problem of potentially sharp images from blurred images with or without estimation of the unknown non-uniform blur kernel. To solve this problem, people propose methods based on traditional image processing and neural networks. One of the methods is to simplify the blur factor, by approximating the non-uniform blur to a uniform blur, and restoring the potential ground truth image and blur kernel. However, due to the irregular motion offset trajectory in space, it cannot be generalized to true blur. Therefore, people have done a lot of research on non-uniform blur[9–11, 22, 34], and extended the degree of freedom of the blur model from uniform to non-uniform. In order to limit the solution space of non-uniform blur, many natural image priors [2, 3, 21] are proposed for regularization, which promotes the research of deblurring. But it still stays at dealing with non-uniform blur caused by simple camera rotation and in-plane translation, which is certainly not enough to express sharp images and blur kernels. The image quality degradation caused by blur can be represented by a mathematical model.

$$B = KS + n, \tag{1}$$

where B and S are the blurred image and the potentially sharpened image respectively. K is an unknown or known blur kernel, that is, a blur matrix. Each row is a local blur kernel, which is combined with a sharp image to generate blurred pixels. n is additive noise. Since deblurring has a large solution space, K or B are generally constrained to simplify the solution of S.

1 College of Electronic and Communications Engineering, Shandong Technology and Business University, Yantai, 264005, China.

2 College of Computer Science and Technology, Shandong Technology and Business University, Yantai, 264005, China. E-mail: lijinjiang@gmail.com.

3 Co-innovation Center of Shandong Colleges and Universities: Future Intelligent Computing, Yantai, 264005, Cina.

Manuscript received: 2014-12-31; accepted: 2015-01-30.

In the past, traditional dynamic scene deblurring was generally done by using additional accurate image segmentation [4, 12] or motion estimation [12]. Among them, Kim et al. proposed joint segmentation of non-uniform blurred images based on energy model. Estimate the nonlinear blur kernel within the segment and realize parameter sharing. A groundbreaking addition of a non-static background, the dynamic scene blur is turned into a local deblurring problem. However, by introducing other additional data processing, the blur kernel that was originally estimated to be inaccurate will deviate further from the true blur kernel. Once the estimation of the blur kernel is in error, undesirable ringing artifacts will be produced. In order not to add redundant information, Kim and Lee approximated the blur kernel as locally linear, so that the motion flow and latent image can be estimated at the same time. Therefore, a non-segmented method is proposed to deal with this problem.

With the rapid development of deep learning, neural networks have been widely used in the field of computer vision. For the problem of image deblurring, they were first proposed for non-blind deconvolution [25]. Xu et al.[33] remove the blur by restoring the sharp image with a given blur kernel. They use separable kernels that can be decomposed into filters to form a deconvolution CNN. In [26], stack multiple CNNs in a coarse-to-fine manner to simulate iterative optimization. Because there is no pair of real blurred images and corresponding sharp images, a blurred image synthesized by a uniform blur core is used for training. In [28], a classification convolutional neural network is used to estimate the local linear blur kernel.

In recent years, people have proposed parameter models based on deep convolutional neural networks (CNN) to replace image formation models. In order to obtain a pair of blurred images and sharp images for network training, Schuler et al. [26] used Gaussian blurring to blur the sharp images collected in the ImageNet dataset [5], and proposed a blind deblurring algorithm based on CNN. The steps of feature extraction, kernel estimation and latent image estimation are carried out in a coarse-to-fine iterative manner. The method[1] predicts the deconvolution kernel in the frequency domain. In the mode where both blur kernel prediction and image prior are based on early learning methods. However, the models generated by these methods can be trained to simulate the nonlinear relationship between blurred images and ground truth, effectively overcoming the limited representative ability of traditional image processing methods to describe dynamic scenes.

Generative Adversarial Nets (GANs) are also used for deblurring due to their advantages in preserving detailed edges and generating approximate images. Kupyn et al. [15] used CNN as a generator and discriminator, and calculated the loss through content loss and adversarial loss. On this basis, they improved the network [16], and proposed a generative confrontation network based on a feature pyramid network and a relativistic discriminator with least square loss.

In order to use the fine image information as a feature aid, a multi-scale network structure is used for blind image deblurring, extending the ”coarse-to-fine” scheme to deep CNN scenes. This method first restores the potentially sharp image at the coarse scale, and then performs the coarse-scale output at the fine scale. In addition to the recent use of independent feature extraction layers on network structures of different scales [20], there have been some works on sharing network parameters in multi-scale pyramids [29] or other effective sharing parameters [7]. However, the use of deep network deblurring still has great challenges before it is promoted and applied. In order to inherit the traditional coarse-to-fine optimization methods, most multi-scale networks use a large number of training parameters. Even if the number of parameters is reduced by parameter sharing, skip connection and other methods, the multi-scale-based method has two main limitations:

1. In order to maintain the integrity of the blurred edge of the object and the sensitivity to large-scale blur, the network generally selects a larger size filter and excessively stacked convolutional layers. But this comes at the cost of the number of parameters and the speed of inference. Therefore, the calculation cost and memory consumption are greatly increased

2. Past experiments have shown that on multi-scale modules, the introduction of coarser or finer scale space inputs to further train model parameters cannot improve the overall deblurring performance of known models. Therefore, simply increasing the spatial scale cannot achieve better results.

According to the development status of deep learning deblurring technology, this paper proposes a progressive dynamic scene deblurring method. We first combine the multi-scale architecture and the encoder/decoder structure to construct a nonlinear function, and then use the residual information in the image to further optimize to break the above limitations. Multi-scale methods are widely used in image deblurring tasks because it is a task from coarse to fine to recover sharp images from blurred images. It is difficult for a single deep network to directly generate sharp images from severely blurred images. We speculate that it is much easier to recover a sharp latent image from a light blur than from a heavy blur. Recently, Park et al. [23] verified our guess from the benchmark dataset and iterative thinking. Therefore, the task is a gradual process, usually including two stages: the first step is to use a larger filter to generate a larger receptive field, restore the area with a larger blur range, and generate a rougher initial deblurred image. The second step refines the texture structure in the image as the final output.

Taking this feature into account, a progressive multi-scale edge-sensing residual network (PMERN) is specially designed. It consists of two corresponding units: Information Integration Unit (IIU) and Detail Optimization Unit (DOU). In the information integration unit, the entire network uses a modified encoder/decoder architecture. We spliced the saliency edge of the blurred image into the encoder as an auxiliary branch to help the network accurately locate the blurred area and degree. In the decoder, we try to change the single operation mode of deconvolution in the network, and then use multi-scale blurred images to be fused into the decoder to simulate the process of restoring potentially sharp images at different scales. The weight is automatically adjusted according to the blur features contained in the blurred image. Therefore, in the deep convolution, the shallow feature information will not be lost due to the excessive number of encoder layers. In order to significantly simplify the training process and bring obvious stability gains. In the detail optimization unit, the edge structure is further refined by learning the residual image. By using multiple blur features of different dimensions for residual learning, the network deblurring effect is improved.

There are two advantages of this network: 1) Since we merge images of different resolutions into the deconvolution network, the training time is much less than the scale cascade structure. 2) Compared with the scale recursive structure, no special parameter sharing method is used.

Our contributions are mainly:

1. Aiming at the limitation of the current depth deblurring model stacking depth and loop iteration, a new solution is proposed. Compared with the previous fixed-level architecture, our network is more flexible. 2. Change the traditional multi-scale method. There is no need to explicitly train the network with images of different scales as input to the scale cascade, and it perfectly combines the multi-scale architecture and the network structure. 3. Through the comprehensive evaluation of the benchmark dataset and the real dataset and the comparison with the most advanced deblurring methods, the performance of this method is better than the existing dynamic scene deblurring methods in both qualitative and quantitative evaluation, which proves our method It achieves better results with the fewest parameters.

2 Related work

In this section, we will briefly review the main applications of multi-scale concatenation and encoder-decoder methods in image deblurring. In the blind deblurring method, both the blur kernel and the image have certain a priori assumptions. However, these methods have little effect on large-scale blur nuclei. The method proposed by Fergus et al. [6] completely abandons the constraints on image prior hypotheses. The method in the article is mainly divided into two steps: estimation of the convolution kernel and deblurring. In-depth study of the gradient distribution of blurred images and non-blurred images, and proposed a deblurring algorithm based on the gradient distribution model. This method introduces a strategy from coarse to fine in the traditional deblurring. On this basis, almost all traditional methods based on energy optimization repeatedly deal with the problem of dynamic deblurring. Optimize from a low sampling scale and gradually expand upwards between iterations until it reaches the normal scale.

2.1 Multiscale

The multi-scale structure is designed to imitate the traditional coarse-to-fine optimization method. Because it does not participate in the estimation of the blur kernel, the artifacts caused by the kernel estimation error are avoided. Nah et al. [20] proposed a "from coarse to fine" neural network to eliminate ambiguity. After the Gaussian pyramid structure, the coarse-scale features are used to deblur the fine-scale images. At the same time, to speed up the model convergence, multi-scale loss and adversarial loss are added to each scale. This method establishes a deep neural network with independent parameters, which leads to the problem of excessive network parameters. To improve the network, simplify the network layer and parameters. Tao et al. [29] proposed a scale recursive network with long and short-term memory, using a codec structure with skip connections of different scales

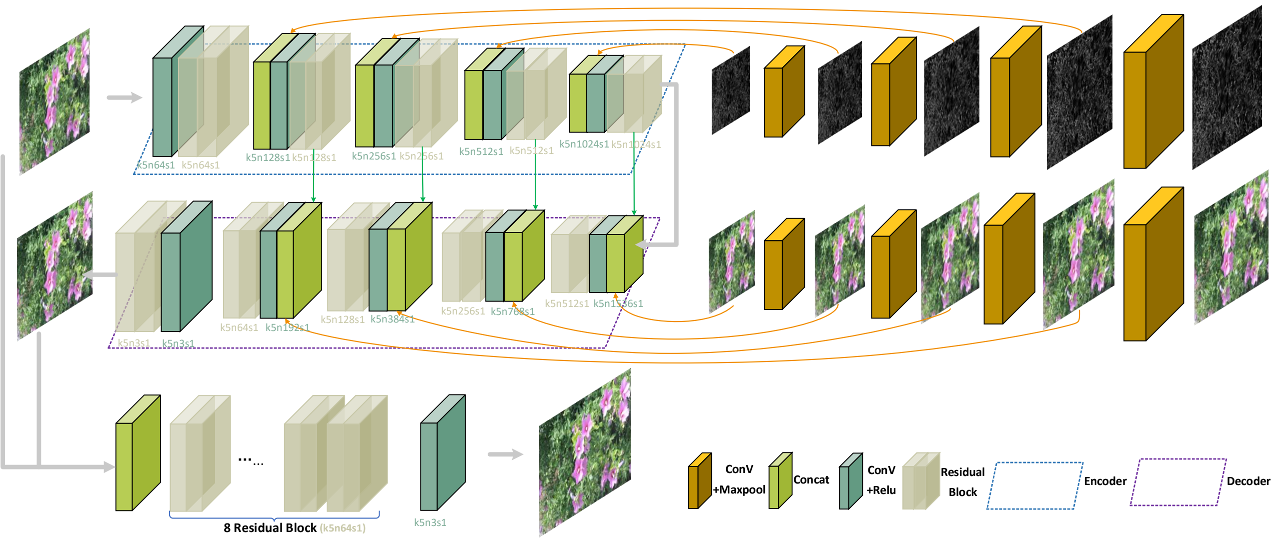

Fig. 1 Our proposed progressive multi-scale edge-sensing residual network.

and parameter sharing. Due to the large filter size, a large number of training parameters are used in the network, and adding low-resolution input to the multi-scale method does not improve the deblurring performance. Zhang et al. [36] proposed a deeply stacked multi-scale patch network, which takes a multi-scale patch structure as input and refines the entire image through a continuous upper layer. And it is proposed that the spatial recurrent neural network can be used to simulate the blur of spatial changes, in which the pixel-level weights of RNN are learned from CNN. In order to obtain a large receptive field and fuse features from different filtering directions, they used four RNNs and added a convolutional layer after each RNN. Then use the pre-trained VGG16 sub-network to predict the spatial variation weight of the RNN. However, because RNN cannot be calculated in parallel along the spatial dimension, the reasoning time of this method is still not reduced, and VGG16 is used as the weight generation network, which increases the network parameters and calculation amount. In order to further reduce the network parameters, [7] replaced the residual blocks in the subnets of each scale with the nested skip connection structure of the nonlinear transformation module. The network components are composed of three modules: feature extraction, nonlinear transformation and feature reconstruction.

2.2 Encoder-decoder

Codec structure [19] is a neural network design pattern. It is often used in natural language processing or other sequence-to-sequence prediction tasks. Specifically, the task of the encoder is to obtain a feature map of the input image through neural network learning after a given input image, and to classify and analyze the pixel values of the low-level regions of the image; The decoder then takes this feature map as input and maps it to the output image. The codec structure has also been successfully applied to various image processing problems. They use a symmetrical structure to first compile the input data into a small-size, multi-channel feature map, and then decode the feature map into an output with the same shape as the input. Among them, Ronneberger et al. [24] added skip connections between the corresponding feature maps in the encoder and decoder to improve their regression capabilities. First apply the structure[29] to image deblurring. Because the number of layers of the original codec is small, the perception field is small. If the number of layers is increased blindly, the size of the feature map will be too small to fully retain the spatial information, and parameters will be increased. Therefore, the author improved the codec, enlarged the field of perception, and had a better effect on recovering severe motion blur. Gao et al. [7] shared selective parameters on the basis of the codec for the change of blur image characteristics at coarse spatial scale. Ye et al. [35] proposed a scale-iterative upscaling network. The model is divided into two layers, and each layer uses a different U-Net as a sub-module. The first layer performs deblurring operations at a relatively small image scale to help the second layer perform large-scale deblurring operations. Similarly, we use an encoder-decoder network to restore the image structure.

3 Network Architecture

In this section, we first introduce the proposed deblur network architecture in detail. In addition, the encoder decoder and the salient edge training in the unit are described. Finally, the details of loss function, training and implementation are given. The purpose of the deep learning network proposed in this paper is to learn the end-to-end non-linear mapping between the blurred image and the corresponding potentially sharp image with the assistance of the sensitive information of the salient edge image, and use the progressive processing flow to achieve high efficiency and Precise blind deblurring. In the task of deblurring, compared with the real image, the blurred image has a huge deviation in both the image content and the image texture, so we cannot directly use the typical residual learning structure. The key idea of the progressive multi-scale residual network (PMRN) is to first reduce the degree of blur. This operation generates an initial deblurred image that is roughly the same structure as ground truth. Then extract the subtle information by learning the initial result and the residual image of the sharp image. In this way, richer details can be restored on the finally reconstructed potentially sharp image. Therefore, the entire network can be divided into two stages, as shown in Figure 1.

The first stage IIU consists of three parts, namely, the backbone, the prominent edge pyramid branch, and the blur image multi-layer guide branch. The backbone is composed of an optimized encoder/decoder, which takes a blurred image as input, extracts content features and blur features and maps them to the output image. The Significant Edge Pyramid Branch (SEPB) takes the significant edge map corresponding to the blurred image as input for convolutional down-sampling, and stitches the feature map to the encoder for feature extraction. The image multi-layer guided branch (IMGB) takes the blurred image as input for convolutional down-sampling, and stitches the feature map to the decoder for deconvolution. As the first stage of image deblurring, this output generates an initial set of sharp latent images.

The output of the IIU is fed back to the second stage DOU and used as the input of the DOU together with the blurred image. We extract structural details by stacking residual blocks and let the network learn the residual information between the numerically similar blur images and sharp images. At the same time, removing batch normalization after the convolutional layer of the classic residual block structure helps to improve the convergence speed and maintain the higher flexibility of CNN [18, 20]. Although the operation of the second stage is larger and more ambiguous, with the help of the first stage, the increase in complexity is not beneficial but gradual. Through the collaborative work of these two stages, a sharp image can be obtained.

3.1 Information Integration Unit

3.1.1 Encoder-decoder architecture

Recently, various computer vision tasks have also been inspired by the success of the encoder-decoder structure. For the classic encoder-decoder structure is not suitable for deblurring tasks. On the contrary, our fusion encoder-decoder network amplifies the advantages of various CNN structures and generates feasibility in training.

First of all, because of the formation of motion blur, a large receptive field is needed, and the depth of the network needs to be increased. However, in actual situations, this operation will introduce a large number of parameters. In addition, there are too many intermediate feature channels, the feature image will be convolved very small, it is difficult to extract the features. Secondly, the spatial scale of

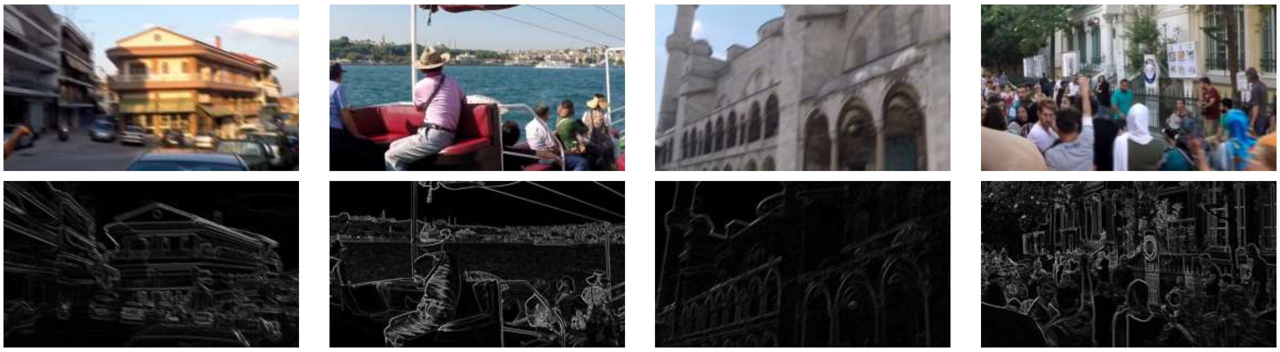

Fig. 2 Saliency edge.

the feature map in the decoder gradually increases, which will lose the original image detail information, which is not conducive to image reconstruction. So we have improved the encoder-decoder structure to adapt to the deblurring task. In order to prevent the disappearance of the gradient, we add a skip connection in the encoding part and the symmetrical decoding part, the purpose is to transfer the corresponding features from the encoder branch to the decoder branch. skip connection can solve the problem of gradient disappearance in the case of a deep network layer, and at the same time help the back propagation of the gradient and speed up the training process.

3.1.2 Significant edge pyramid branch

Motion blur in dynamic scenes often occurs at the edges of moving objects or at the edges of background objects due to camera shake. Blur edge features carry a lot of positioning information for the network to quickly extract feature maps, and better locate the blur area. Considering the fact that strong edge information is important for reliable deblurring, we use the Sobel operator to extract the salient edges in the image, and the result is shown in Figure 2.

We use a 2*2 convolution kernel and a maximum pooling layer with a step size of 2 to gradually obtain the down-sampled significant edge map. For the salient edge pyramid, we extract the hierarchical representation through the convolutional layer. Finally, the multi-scale spatial features are spliced to the encoder path.

$$E_{ed}^{1} = \max pool(E_{e}^{up}), \tag{2}$$

$$\mathrm{F}_{ed}^{1} = \sigma(W_{ed}^{1} * E_{ed}^{1} + b_{ed}^{1}), \tag{3}$$

$$E_{ed}^{i+1} = \mathrm{maxpool}(\mathrm{E}_{ed}^{i}), \tag{4}$$

$$\mathrm{F}_{ed}^{i+1} = \sigma(W_{ed}^{i+1} * E_{ed}^{i+1} + b_{ed}^{i+1}). \tag{5}$$

Where $i \in \{1,2,3\}$, $E_{ed}^{1}$is the salient edge map obtained by down-sampling E_{e}^{1} ,F_{ed}^{1} is the feature extracted from E_{ed}^{1} ,* represents the convolution operation, and maxpool is the maximum pooling operation. W_{ed}^{1} and b_{ed}^{1} represent the weight and bias of the first convolution operation in the significant edge pyramid branch, σ is the activation function of the modified linear unit, E_{ed}^{i+1} is the significant edge map obtained by downsampling E_{ed}^{i} ,F_{ed}^{i+1} is the feature extracted from E_{ed}^{i+1} , W_{ed}^{i+1} and b_{ed}^{i+1} represent the weight and bias in the i+1 th convolution operation.

Advantages of significant edge branching: (1) Strengthen the network's sensitivity to blurred areas of blurred images. (2) The encoder extracts more detailed information for easy analysis.

3.1.3 Image multi-layer guided branch

Different from the multi-scale in previous methods, we down-sample the original blurred images of different spatial scales and stitch them into the decoder to guide image reconstruction. The multi-layer branch of the blurred image can be expressed as:

$$F_{ms}^{1} = \sigma(W_{ms}^{1} * B + b_{ms}^{1}), \tag{6}$$

$$B_{ms}^{j} = \max pool(F_{ms}^{j}), \tag{7}$$

$$F_{ms}^{j+1} = \sigma(W_{ms}^{j+1} * B_{ms}^{j} + b_{ms}^{j+1}). \tag{8}$$

Where$j \in \{1,2,3\}$, F_{ms}^{1} is the feature extracted from the blurred image B, B_{ms}^{j} is the blur image in the coarse-scale space obtained by the down-sampling of F_{ms}^{j} , F_{ms}^{j+1} is the feature extracted from B_{ms}^{j} . Among them, the convolutional layer with a step size of 2 reduces the feature map size to half of the original size and doubles the number of channels. In contrast, the deconvolution layer with a step size of 2 halves the number of feature channels and doubles the size of the feature map.

The multi-scale guidance branch has the following advantages: (1) Multi-scale images are incorporated into the network structure, which reduces the number of training parameter updates and saves training time. (2) Increase the network width of the decoder branch.

3.2 Detail optimization unit

Due to the large gap between the blurred image and the corresponding potentially sharp image, if the normal network learning structure is used directly to deblur, the blurred edge of the image will have serious ringing artifacts. So like the method[20, 29, 36], the problem needs to be broken down into several sub-problems and completed step by step. Our key idea is to first generate an initial result with a sharp structure, and then concentrate on extracting subtle information by learning the initial result and the residual image of the sharp image.

4 Loss function

In deep learning, if it is a classification problem, you can use loss functions such as cross-entropy, softmax, or SVM. If it is a regression problem, the loss function generally adopts L1 or L2. With the development of network architecture, people have made many attempts to find a loss function to replace the widely used L1 loss and L2 loss. However, the trade-off of perceptual distortion has recently been demonstrated. Advanced loss functions (such as the adversarial loss of generative adversarial networks [8]) improve the perceptual image quality at the cost of distortion. Moreover, L2 is only very sensitive to large errors, but very tolerant of small errors. Most importantly, L2 loss does not consider human visual perception, which is different from the human visual system. Because in the subject of image deblurring, we not only need to restore the blur edges in the image to a potentially clear structure, but also need to retain the color information and detail information of the original image. In order to take human visual perception into consideration, the SSIM loss function [31] fully considers the brightness, contrast, and structural indicators, which can restore the texture details of the image well. At the same time, it also benefits from the progressive network structure, so the use of negative SSIM to train dynamic recursive fast multi-scale residual network will get good results.

Image illumination comparison part:

$$lx, y = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{9}$$

Image contrast comparison part:

$$cx, y = \frac{2\gamma_x\gamma_y + C_2}{\gamma_x^2 + \gamma_y^2 + C_2}, \tag{10}$$

Image structure comparison part:

$$\mathrm{s}(x, y) = \frac{\gamma_{xy} + C_3}{\gamma_x\gamma_y + C_3}, \tag{11}$$

among them:

$$\gamma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \gamma_x)(y_i - \gamma_y). \tag{12}$$

In the formula, C1 C2 C3 are constant terms to avoid instability when the denominator is close to zero. $x_i(y_i)$ and N are the image signal and the number of signals. μ acts on the average intensity of the discrete signals of x and y.

The SSIM loss function formula can be obtained by multiplying the product of the three comparison parts:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\gamma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\gamma_x^2 + \gamma_y^2 + C_2)}. \tag{13}$$

5 Experiment

5.1 Dataset

Artificially synthesized blurred images are not enough to express the complexity of real blurred images. The camera movement has 6 degrees of freedom (6D), including 3 translational freedoms and 3 rotational freedoms. The method of using the sharp image convolution blur kernel only considers two translational degrees of freedom on the two-dimensional plane [17, 27]. In addition, there are factors such as lens distortion, sensor saturation, camera nonlinear transformation, noise, compression, and depth of field that cannot be simulated by synthetic images.

GOPRO is a large-scale deblurring dataset proposed by Nah, taken by GOPRO Hero Black camera. Unlike the previous dataset that uses blur kernel and sharp image convolution to synthesize blurred images, it uses high-speed cameras to continuously short exposure sharp frames, and integrates and averages them to simulate long exposure blurred frames. The image formed in this way is closer to reality, and can simulate complex camera shake and non-uniform blur caused by multiple target movements in the scene. The GOPRO dataset contains a total of 3214 pairs of sharp and blurred images with an image size of 720*1280, of which 2103 pairs of images are used for training and the remaining 1111 pairs of images are used for testing.

The Kohler dataset [14] is a benchmark dataset for evaluating and comparing blind deblurring algorithms. The author records and analyzes the real camera movement over time, and then replays it with a robot carrier, and forms a dataset by leaving a series of sharp images on the movement track of the 6D camera. The Kohler dataset consists of 4 pictures, each picture is blurred with 12 different blur checks, and finally 48 blurred images are formed.

5.2 The experimental details

Our experiments are conducted on a PC equipped with eight TITAN RTX GPUs. Implement our framework on the pytorch platform. In addition, pixel filling is used to keep the output and input scales of the feature map unchanged. The adam[13] algorithm is used to train the initial learning rate at 0.0001 exponential decay to at 30 epoch using power 0.3, batchsize is set to 1 in the experiment. In experiments involving iterative model structure, the network shares the same training environment.

Our evaluation is comprehensive to verify different network structures and various network parameters. To

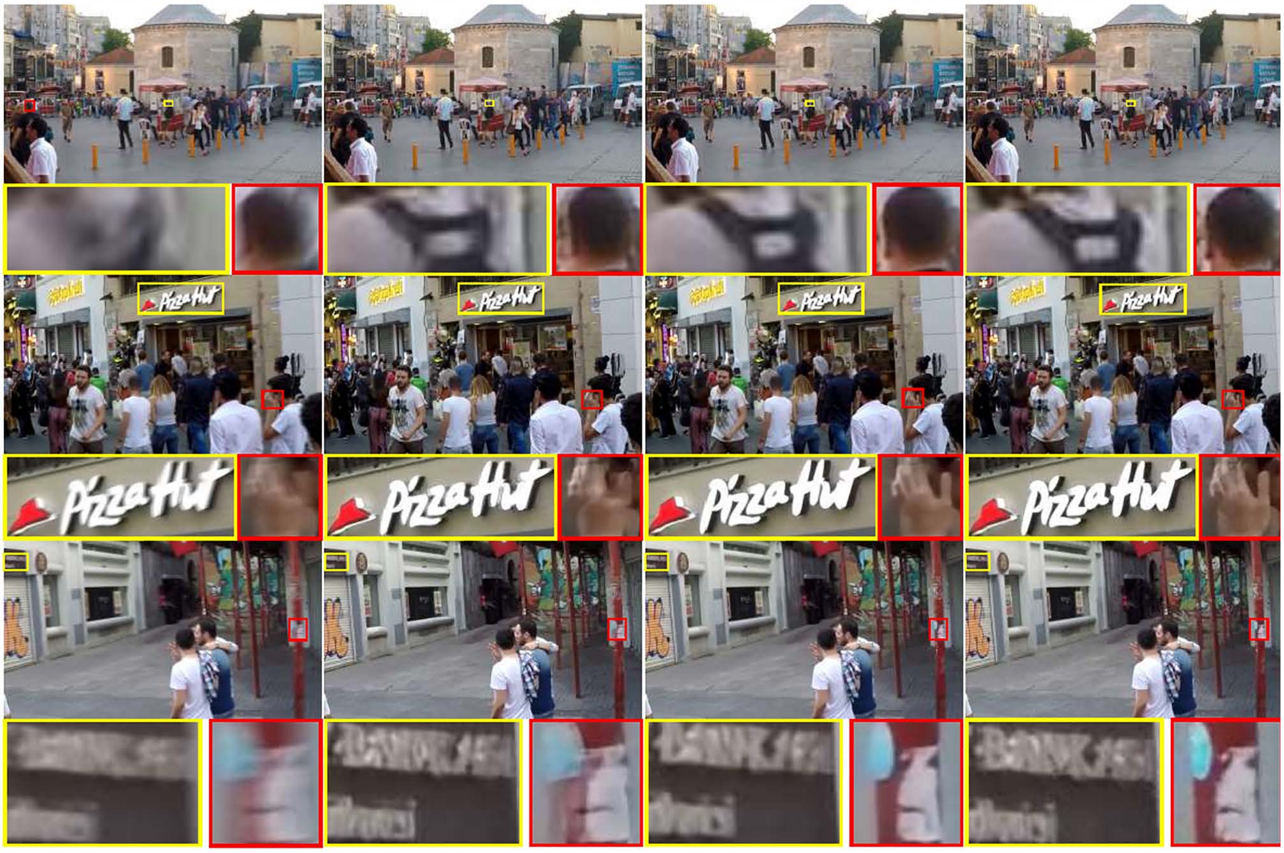

Fig. 3 From left to right are the blurred image, IIU output, complete network output and sharp image.

be fair, all experiments were performed on the same dataset with the same training configuration unless otherwise stated. In order to evaluate the performance of our proposed method, we use peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as our objective evaluation indicators(Our method also performs well on color-based evaluation indicators.[30]).

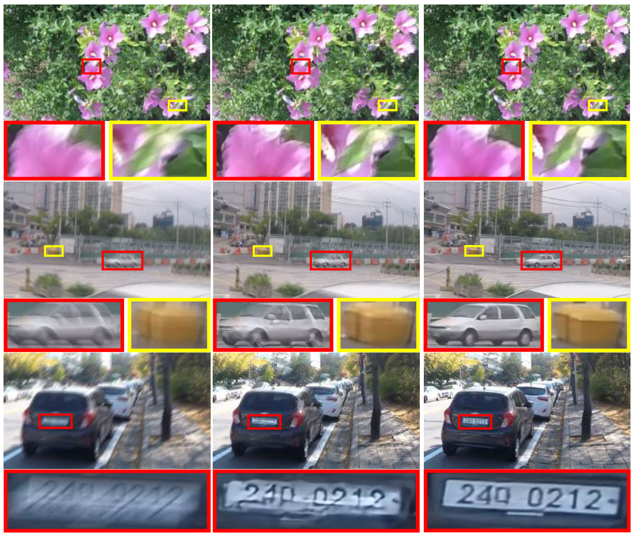

Fig. 4 From left to right, the blurred image is shown, the output of the edge branch is removed, and the complete network is output.

Method	blur	IIU	DOU + IIU
PSNR	20.54	29.32	32.65
SSIM	0.7998	0.9132	0.9512

Tab. 1 Quantitative results of progressive units.

5.3 Ablation study

5.3.1 Progressive deblurring

In order to verify our progressive defuzzification conjecture, we analyzed the influence of the DOU unit on the performance of the deblurring network, that is, the GoPro dataset test set is used in the IIU unit to perform the defuzzification operation. As shown in Figure 3, in the subjective sense, the output of the IIU is less blurred than the original blurred image, the blurred area gradually tends to be sharp, and some edges are restored. But there are still undesirable white spots and small areas of blur. Observing the detailed area of the image from left to right, the edges of the object gradually become sharper and closer to the GT. From the experimental results, it can be concluded that the image deblurring task gradually shifts from the input to the light blur, and finally approaches the sharp image. Although in terms of quantitative indicators(Tab 1), the output with fine details is only slightly better than the original deblurred image, but those further enhanced structure and texture details also play a very important role in achieving more realistic photo effects.

5.3.2 Edge perception

When predicting blur images, in order to verify the positive effect of edge information on the network, we remove the significant edge pyramid branch from the structure to observe the deblurring effect of the test set. It can be seen from the visual subjective that with the assistance of edge information, the image edge recovery is very obvious.

5.4 Compare with other methods

5.4.1 Results of the benchmark dataset

We compare our method with the existing image deblurring methods quantitatively on the benchmark evaluation GOPRO dataset. Then we put the Kohler dataset on our training model for testing. This dataset is widely used by traditional methods and learning-based methods for further performance evaluation. Finally, we use the blur images in the real scene in the Lai dataset to test the generalization ability of our network. Since our method deals with motion blur, it is unfair compared with the traditional uniform deblurring method. So we choose Whyte et al. [32] method as the representative traditional method of non-uniform blur processing. At the same time, we also choose the same de-motion blur method as ours for comparison. Both Nah et al. and Tao et al. use multi-scale architectures, but use parameter independence and parameter sharing to construct their deep networks. On this basis, Zhang et al. [36] used different depth stacking and layering methods to adapt to different images. We all use the author's official publicly available default parameters.

Figure 5 shows our deblurring results on the benchmark dataset(Unless otherwise stated, all images evaluated in our experiments are in RGB mode). For a fair experiment, we use the Nah method to test the code on the pytorch platform, which is different from the original lua code. For a fair experiment, we use the Nah method to test the code on the pytorch platform, which is different from the original lua code: The range of RGB has been changed to [0, 255]; the loss function only uses L1 loss, not adversarial loss; use Mixed-precision training; SSIM function is converted from MATLAB to python. This unifies the framework platform of each method and contributes to the objectivity of experimental results. The author provides 5 different models on the two datasets. Because we are all training and testing on the GOPRO dataset, we choose the best GOPRO_L1_amp as the test model for comparison. Three models are provided in the code provided by Tao. According to the author's description, the LSTM model works best, so it is also used as a test model for our comparison. It released models with different hierarchical structures, and we also chose their best DMPHN_1_2_4_8 model. Kupyn improved the deblurring effect and quality on the basis of the original GAN network implementation, and proposed a new version (DeblurGanv2). As a representative of the GAN network, we chose them to use Inception-ResNet-v2 to implement the best model for testing.

The PSNR and SSIM metrics of the deblurred image on the GOPRO dataset are quantitatively evaluated by the python code. The PSNR and MSSIM metrics on the Kohler dataset are calculated by the executable file provided by [14].

Synthesize the results in the chart. The structural similarity of PMERN in the gopro test is better than other methods. Especially in restoring the image edge texture details, the effect is particularly obvious. In the deblurring result of Nah et al.'s method, there are undesirable black patches; In the deblurring result of SRN, there are obvious artifacts and faults in the blur; DMPTH has achieved good results, but there is still room for improvement in detail and texture; GAN-v2 is too smooth, lacks texture details, and even erythema noise appears in some areas. In contrast, PMERN has a good effect on both the restoration of handwriting and the restoration of the edge structure of the image, and the edge texture and structural details of the object are retained to a large extent.

5.4.2 Results of the real blurred images

Although the GOPRO dataset simulates real blur by averaging continuous frame synthesis, it is synthesized by a high-speed camera, and the ground truth sharp image has severe noise and varying degrees of blur. Although the Kohler dataset is a real database, it only contains four different scenarios. Therefore, we further test our model on the real scene dataset of Lai et al.

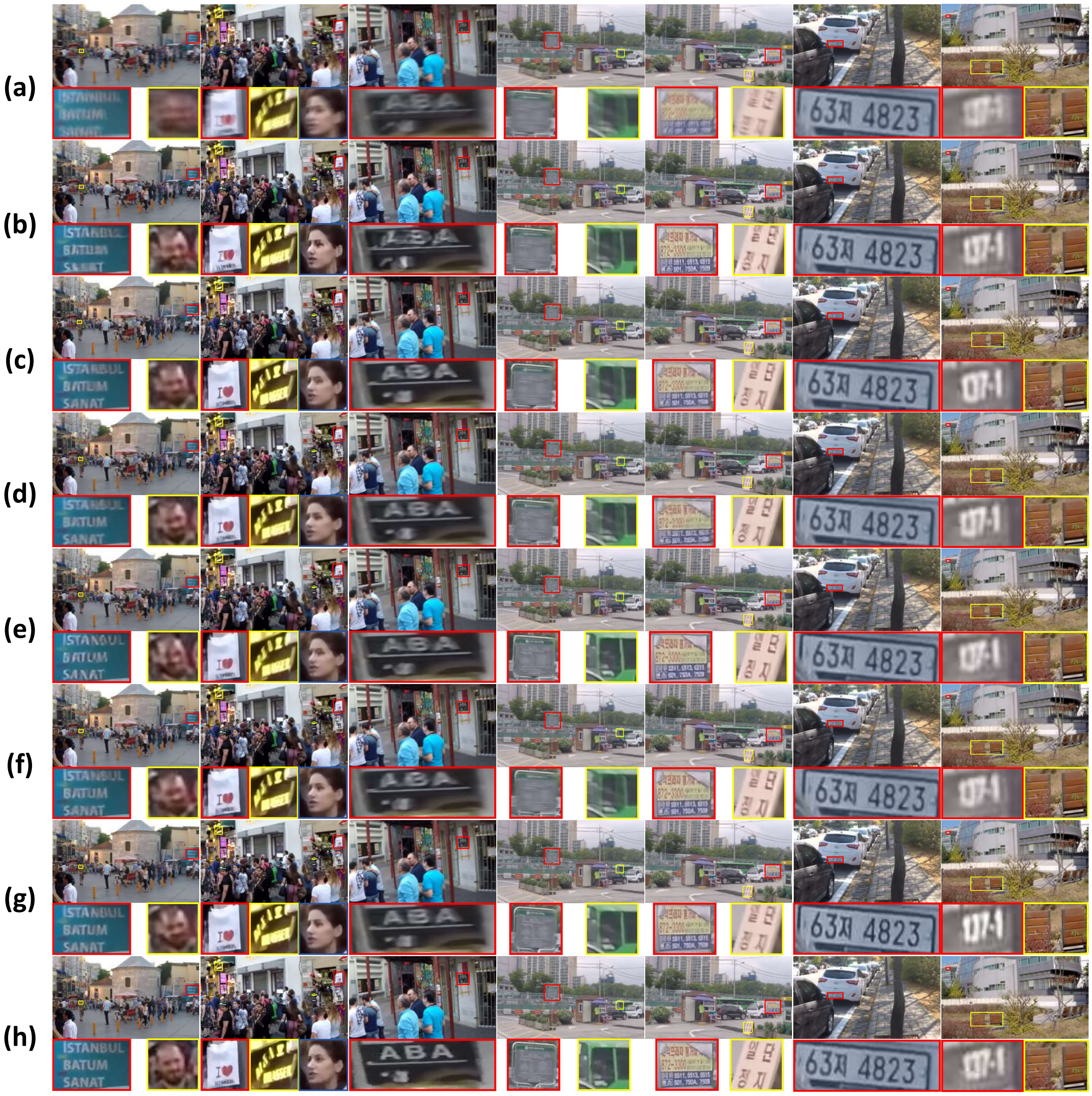

Fig. 5 Visual comparisons on testing dataset. In the top-down order, we show input, results of Whyte et al, Nah et al, Tao et al, Zhang et al, Kupyn et al,and our results, and sharp images.

Method	Whyte et al.	Nah et al.	Tao et al.	Zhang et al	Kupun et al.	Ours
PSNR	20.54	28.49	30.25	30.45	30.51	31.16
SSIM	0.7998	0.8543	0.9030	0.9057	0.9121	0.9225

Tab. 2 Quantitative results on test dataset (in terms of PSNR/SSIM).

The deblurring effect of the real scene can better reflect the adaptability of a network in the application field. As shown in Figure 8. Our method can generate sharp images for different scenes, and the texture details recovered by other methods are clearer. It has also made great progress in text processing. This

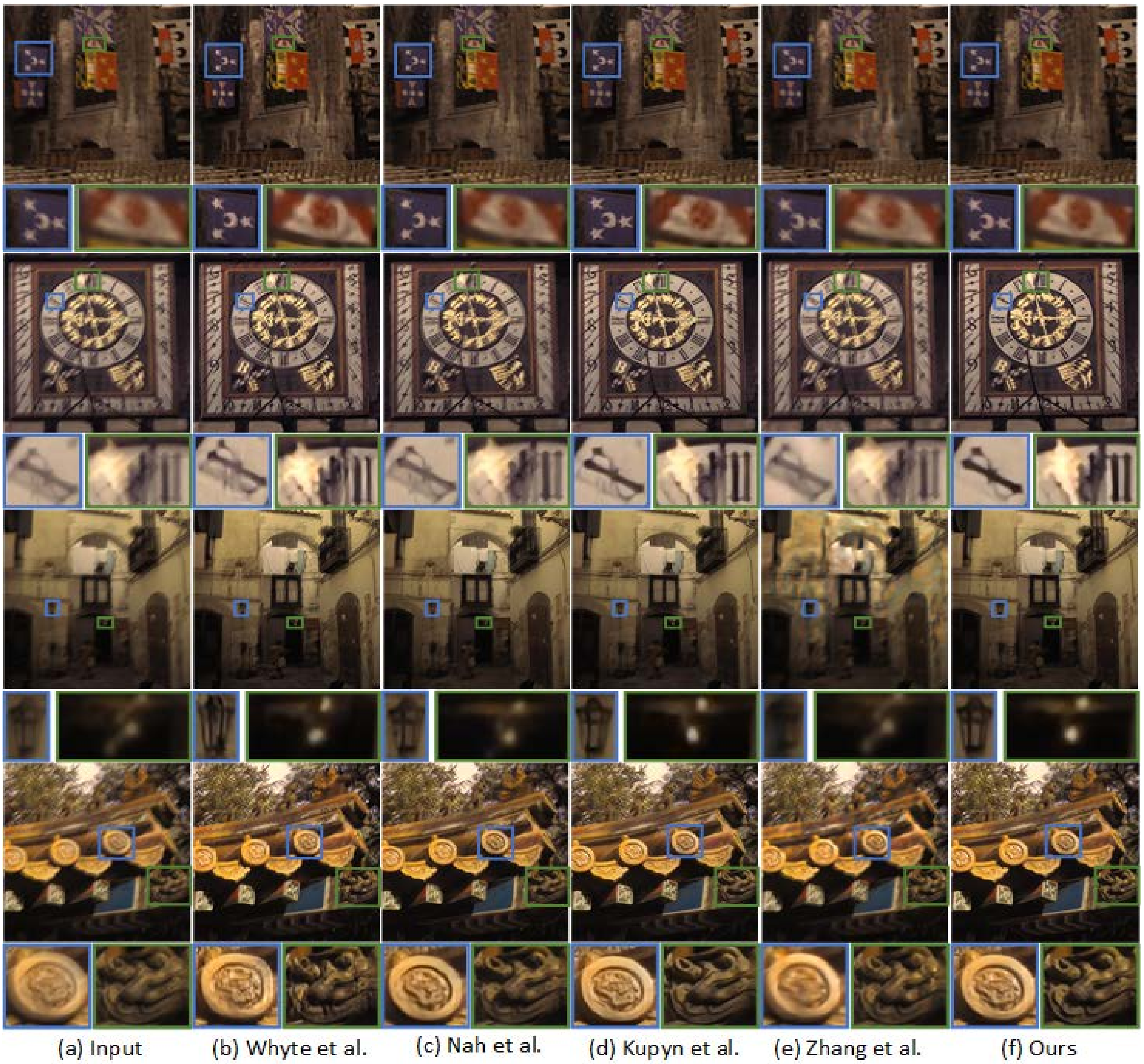

Fig. 6 Results on real data

Method	Whyte et al.	Nah et al.	Kupyn et al.	Zhang et al.	Ours
PSNR	24.68	26.48	27.76	25.56	28.28
MSSIM	0.7937	0.8079	0.8183	0.7867	0.8307

Tab. 3 Quality evaluations on Kohler dataset.

shows the wide compatibility of our network to different scenarios.

As shown in Figures 6, we show a qualitative comparison between Kohler datasets. Obviously, the restored images get high-quality visual effects, and our network can adapt to different scenarios.

6 Conclusion

In this article, we break through the limitations of current image deblurring tasks and describe a network structure of a multi-scale variant of blur edge perception. This structure effectively integrates edge features and scale information cues. We also propose a progressive network mode for single image deblurring in dynamic scenes. Outstanding performance in restoring image edge details. Our work provides new ideas for the follow-up of effective multi-scale deblurring deep networks. Experimental results show that, compared with traditional methods and learning-based methods, this method has better results on both benchmark datasets and real blurred images.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (61772319, 62002200, 61976125, 61976124), Shandong Natural Science Foundation of China (ZR2017MF049).

References

[1] A. Chakrabarti. A neural approach to blind motion deblurring. In *European conference on computer vision*, pages 221–235. Springer, 2016.

[2] T. F. Chan and C.-K. Wong. Total variation blind deconvolution. *IEEE transactions on Image Processing*, 7(3):370–375, 1998.

[3] S. Cho and S. Lee. Fast motion deblurring. In *ACM SIGGRAPH Asia 2009 papers*, pages 1–8. 2009.

[4] F. Couzinie-Devy, J. Sun, K. Alahari, and J. Ponce. Learning to estimate and remove non-uniform image blur. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1075–1082, 2013.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *ACM SIGGRAPH 2006 Papers*, pages 787–794. 2006.

[7] H. Gao, X. Tao, X. Shen, and J. Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3848–3856, 2019.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[9] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *European Conference on Computer Vision*, pages 171–184. Springer, 2010.

[10] S. Harmeling, H. Michael, and B. Schölkopf. Space-variant single-image blind deconvolution for removing camera shake. In *Advances in Neural Information Processing Systems*, pages 829–837, 2010.

[11] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf. Fast removal of non-uniform camera shake. In *2011 International Conference on Computer Vision*, pages 463–470. IEEE, 2011.

[12] T. Hyun Kim, B. Ahn, and K. Mu Lee. Dynamic scene deblurring. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3160–3167, 2013.

[13] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[14] 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 *European conference on computer vision*, pages 27–40. Springer, 2012.

[15] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018.

[16] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8878–8887, 2019.

[17] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1964–1971. IEEE, 2009.

[18] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[19] X. Mao, C. Shen, and Y.-B. Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in neural information processing systems*, pages 2802–2810, 2016.

[20] S. Nah, T. Hyun Kim, and K. Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3883–3891, 2017.

[21] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. Deblurring text images via l0-regularized intensity and gradient prior. In *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition, pages 2901–2908, 2014.

[22] J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016.

[23] D. Park, D. U. Kang, J. Kim, and S. Y. Chun. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. *arXiv preprint arXiv:1911.07410*, 2019.

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

[25] C. J. Schuler, H. Christopher Burger, S. Harmeling, and B. Scholkopf. A machine learning approach for non-blind image deconvolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1067–1074, 2013.

[26] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf. Learning to deblur. *IEEE transactions on pattern analysis and machine intelligence*, 38(7):1439–1451, 2015.

[27] L. Sun, S. Cho, J. Wang, and J. Hays. Edge-based blur kernel estimation using patch priors. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2013.

[28] Y.-W. Tai, X. Chen, S. Kim, S. J. Kim, F. Li, J. Yang, J. Yu, Y. Matsushita, and M. S. Brown. Nonlinear camera response functions and image deblurring: Theoretical analysis and practice. *IEEE transactions on pattern analysis and machine intelligence*, 35(10):2498–2512, 2013.

[29] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8174–8182, 2018.

[30] A. Valberg. *Light vision color.* John Wiley & Sons, 2005.

[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[32] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. *International journal of computer vision*, 98(2):168–186, 2012.

[33] L. Xu, J. S. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *Advances in neural information processing systems*, pages 1790–1798, 2014.

[34] L. Xu, S. Zheng, and J. Jia. Unnatural l0 sparse representation for natural image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1107–1114, 2013.

[35] M. Ye, D. Lyu, and G. Chen. Scale-iterative upscaling network for image deblurring. *IEEE Access*, 8:18316–18325, 2020.

[36] H. Zhang, Y. Dai, H. Li, and P. Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5978–5986, 2019.

TIANLIN ZHANG Tianlin Zhang received the B.S. degree in Software engineering from Shandong Jianzhu University, Shandong, China in 2018. Currently, he is a M.S. degrees candidate in the School of Computer Science and Technology, Shandong Technology and Business University, Yantai, China. His research interests include computer graphics, computer vision, and image processing.

JINJIANG LI Jinjiang Li received the B.S. and M.S. degrees in computer science from Taiyuan University of Technology, Taiyuan, China, in 2001 and 2004, respectively, the Ph.D. degree in computer science from Shandong University, Jinan, China, in 2010. From 2004 to 2006, he was an assistant research fellow at the institute of computer science and technology of Peking University, Beijing, China. From 2012 to 2014, he was a Post-Doctoral Fellow at Tsinghua University, Beijing, China. He is currently a Professor at the school of computer science and technology, Shandong Technology and Business University. His research interests include image processing, computer graphics, computer vision, and machine learning.

HUI FAN Hui FAN received the B.S. degrees in computer science from Shandong University, Jinan, China, in 1984. He received the Ph.D. degree in computer science from Taiyuan University of Technology, Taiyuan, China, in 2007. From 1984 to 2001, he was a Professor at the computer department of Taiyuan University Technology. He is currently a Professor at Shandong Technology and Business University. His research interests include computer aided geometric design, computer graphics, information visualization, virtual reality, and image processing.