The Visual Computer (2024) 40:137–151
https://doi.org/10.1007/s00371-023-02771-8

ORIGINAL ARTICLE

Reference-based dual-task framework for motion deblurring

Cunzhe Liu[1] · Zhen Hua[1] · Jinjiang Li[2]

Accepted: 6 January 2023 / Published online: 4 February 2023

Abstract
Deep learning algorithms have made significant progress for deblurring in dynamic scenes. However, most of the existing image deblurring methods use a single blurry image as the input of the algorithm, which limits the acquisition of information and fails to preserve satisfactory structural texture. In contrast, we present a reference-based dual-task framework to recover a high-quality image by deblurring and enhancing a blurry image under the guidance of a reference image. Specifically, the framework includes two tasks: single image deblurring and reference feature transfer. The single image deblurring task deblurs the blurry image leveraging only the blurry image itself. The reference feature transfer task extracts and transfers abundant textures from the reference image to the coarsely result of the single image deblurring task. Benefiting from the reference image, our proposed method achieves more realistic visual effects with sharper texture details. Experimental results on GoPro, HIDE and RealBlur datasets demonstrate that our method outperforms state-of-the-art methods both quantitatively and qualitatively.

Keywords Motion deblurring · Reference image · Deep neural network · Structure and texture details

1 Introduction

Motion blur is usually caused by the tangled motion of objects in the captured scene or camera shake. Image deblurring has always been a challenging problem in the fields of computer vision and image processing, where the goal is to recover images with sharp details from a given blurred image. Blurry images not only affect the quality of people's visual perception, but also degrade the performance of vision tasks such as object detection [1] and face recognition [2]. Therefore, although image deblurring is a low-level computer vision task, it is of great significance to study an efficient deblurring algorithm to recover image structure texture.

Mathematically, the image degradation model due to blurring can be expressed as follows:

$$x = F(y, k) + n \tag{1}$$

where x and y denote the blurry image and the clear latent image, respectively. $F(y, k)$ usually stands for 2D convolution operator with kernel k. n represents the additive random noise. Since only the blurry image x is given and other quantities are unknown, there are infinitely many inconsistent solutions to solve Eq. (1), so image deblurring is a highly ill-posed problem. Traditional methods usually employ blur kernel estimation or natural image priors to deal with this ill-posed problem [3–8]. Thanks to the rapid development of deep learning techniques and the availability of large-scale datasets, a large number of learning-based methods employ end-to-end deep convolutional neural networks to learn the mapping relationship between blurry and clear images [9–17]. Despite the excellent performance of these learning-based methods, they are still insufficient in recovering the texture details of images. The method based on the generative adversarial network [18–21] can enhance perceptual quality, but they sometimes generate deblurred results with artifacts as shown in Fig. 1.

Recently, the field of image super-resolution [22–28] has introduced additional reference images to super-resolve

✉ Zhen Hua
huazhen66@foxmail.com

Cunzhe Liu
1365558766@qq.com

Jinjiang Li
lijinjiang@gmail.com

[1] School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264005, China

[2] School of Computer Science and Technology, Shandong Technology and Business University, Yantai 264005, China

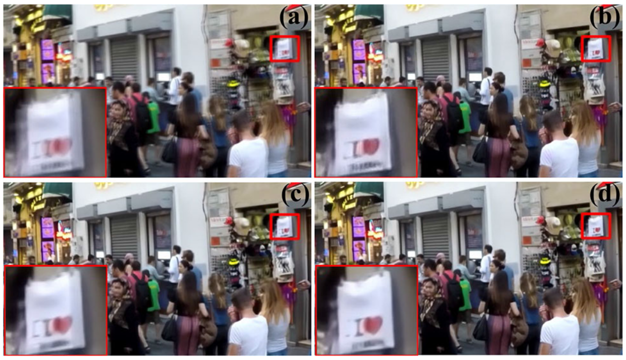


Fig. 1 One example. **a** Input blurry image. **b** Result of DeblurGAN-v2 [19]. **c** Result of MIMO-Unet+ [14]. **d** Our result

low-resolution(LR) images in order to compensate for the information lost in LR images and achieve excellent performance. This method, called reference-based super-resolution(RefSR), aims to transfer relevant texture details from a reference image to an LR image. Inspired by RefSR, we explore a new approach to utilize reference images to help deblur degraded images. Reference images and blurry images have similar content information and textures; they can be captured from different camera angles or obtained from video frames. Compared with using only one blurry image input to the neural network, the additional reference image can provide more complementary information, which alleviates the ill-posedness of the image deblurring problem to a certain extent. The state-of-the-art deblurring algorithm EFNet [29] also exploits additional information, achieving state-of-the-art deblurring performance. Reference-based methods have great potential for image restoration, but when applied to image deblurring, it is still very difficult to establish the corresponding matching relationship between the blurry image and the reference image, that is, how to transfer the relevant texture information to facilitate the reconstruction of the deblurred image. Specifically, there is a misalignment of complementary information between the reference image and the blurry image due to different camera angles or object movements. For efficient transfer of high-quality textures, alignment of the reference image and the blurry image is required. Most reference-based super-resolution adopts spatial alignment [23] and patch alignment [24,26] for alignment operations. However, unlike the downsampled LR image, the blurry images have a large degree of blur degradation, and it is difficult to directly establish the correspondence between the blurry images and the reference images. In addition, inaccurate alignment can transfer irrelevant textures to the deblurred image, resulting in severe degradation of deblurring performance. Hence, we need to explore a framework to achieve matching correspondence between blurry images and reference images, and adaptively transfer relevant texture features.

To address the issues mentioned above, we propose a novel reference-based image deblurring framework, which consists of two tasks: single image deblurring and reference feature transfer. First, we perform coarse image deblurring on the blurry input, which alleviates the difficulty of matching between blurry images and reference images. Given the results of the single image deblurring task, the reference feature transfer task further establishes the corresponding matching relationship and transfers the textures from reference image to deblurred image. In addition, to stably and efficiently transfer features from reference images, we propose the reference alignment module to extract high-quality features, which contains both patch alignment and deformable alignment. Finally, we fuse features and reconstruct the deblurred images using adaptive feature fusion. We conduct extensive experiments on synthetic and real-world datasets. Quantitative and visual experimental results demonstrate that our method achieves state-of-the-art performance.

The main contributions of this paper are summarized as follows:

- We propose a novel reference-based dual-task framework for image deblurring, which consists of a single image deblurring task and a reference feature transfer task.
- We propose the reference alignment module and adaptive feature fusion module, which effectively utilize the texture features of the reference image and refine the single image deblurring results.
- Extensive experiments on benchmark datasets show that our framework achieves excellent deblurring performance. Moreover, our framework also shows better performance when the reference image is dissimilar to the blurry input.

2 Related work

In this section, we briefly review some works related to our research, including learning-based deblurring methods and reference-based methods.

2.1 Learning-based deblurring methods

Learning-based methods have made significant progress in recent years. Sun et al. [30] utilized convolutional neural networks to estimate spatially varying motion blur kernels from local patches and then obtained deblurring results by deconvoluting the blurry images. Nah et al. [9] proposed an end-to-end multi-scale network to gradually restore clear images from coarse to fine. Similarly, Tao et al. [10] proposed a scale-recurrent structure on a multi-scale basis to reduce the amount of parameters. Meanwhile, generative adversarial networks [18,19] are also employed to improve the perceptual quality of deblurring results. Zhang et al. [12] divided the image into multiple patches as the input of the network and aggregated multiple image patches at different stages for better performance. Park et al. [31] proposed incremental temporal training, which uses temporal information to gradually restore blurred images. Li et al. [32] proposed a light global context refinement module into image deblurring for enriching global feature details. Cho et al. [14] proposed a fully convolutional deblurring network with multiple inputs and multiple outputs, and fused features of different scales to achieve excellent performance. Niu et al. [33] extracted the spatio-temporal information from the blurry input to assist deblurring. Although the methods mentioned above are beneficial for image deblurring, their algorithms are only based on a blurry image. The blurry image loses the low-frequency and high-frequency information of the image due to severe degradation. Low-frequency information and high-frequency information correspond to image structure and texture details, respectively. Due to the lack of sufficient information, it is difficult to recover results with structural and texture details when faced with images with large blur degradation, which limits the deblurring performance. In contrast, the reference images we introduce can provide more information and facilitate the restoration of image structure and texture.

2.2 Reference-based super-resolution

RefSR super-resolve LR images with the help of an additional reference image, the purpose of which is to extract and transfer texture information from the reference image after aligning the low-resolution image with the reference image. Zheng et al. [23] estimated the optical flow between the reference image and the low-resolution image, and then aligned them by the flow. Optical flow estimation is widely used in the field of computer vision. [34–36]. Inspired by video super-resolution [37,38], Shim et al. [39] further utilize the feature of deformable convolution(DCN) to extract relevant reference features. However, the above alignment methods usually lack to construct long-distance correlations between image pairs. Therefore, patch alignment based methods [24–26,40] are proposed. Zhang et al. [24] fuse reference features in a multi-scale feature space by computing the similarity between patches.Yang et al. [25] proposed texture transformer, in which hard and soft attention are used to extract and fuse textures. Lu et al. [26] proposed a coarse-to-fine patches correspondence matching pattern that significantly reduces the computational complexity. Wang et al. [40] first generalized RefSR to real-world dual cameras, super-resolve wide-angle images with telephoto images and obtained high-fidelity results. Huang et al. [28] proposed a reference-based dual-task [41] framework, which achieved state-of-the-art performance.

The above RefSR methods provide another idea for image deblurring. To this end, we explore new ways to utilize additional sharp reference images to assist in image deblurring. Inspired by previous methods [25,39], we adopt patch alignment and deformable alignment [42] strategies for different fields of view between blurry images and reference images. We also equip our framework with the adaptive fusion module, designed to effectively and efficiently aggregate reference feature and deblurred feature.

3 Method

In this work, to obtain deblurred images with fine textures, we propose a dual-task framework to fully utilize the additional reference images. We apply separate tasks to the blurry input and the reference image. To be specific, the blurry input provides content and structural information for the deblurred result, while the reference image is expected to provide texture details. To this end, we process the blurry input I_{Input} and the reference image I_{Ref} separately and then perform adaptive fusion. As shown in Fig. 2, our framework mainly consists of two parts:

For the *single image deblurring* task, we roughly deblur the blurry input to reduce its blur degradation degree and obtain the single image deblurring result I_{Deblur} and the deblurring feature F_{Deblur}:

$$F_{\text{Deblur}} = \mathcal{F}_{ID}(I_{\text{Input}}) \tag{2}$$

where $\mathcal{F}_{ID}$ represents the single image deblurring model, and here we use BAM [43] as the main building block of the model.

For the *reference feature transfer* task, we extract well-aligned features from the reference image and transfer to the deblurred feature F_{Deblur}. We first map I_{Ref} and I_{Deblur} into feature maps using a shared VGG19 pre-trained model. After that, we use the normalized inner product [24] in the feature space to calculate the cosine similarity matrix $M_{i,j}$ between I_{Ref} and I_{Deblur}. Then, we calculate the index map P and confidence map C based on the matrix $M_{i,j}$ for the following two modules:

$$P_i = \arg\max_j M_{i,j} \tag{3}$$

$$C_i = \max_j M_{i,j} \tag{4}$$

where the operation represented by Eq. (3) is to take the index of the maximum value of each row of the cosine similarity matrix M, and the operation represented by Eq. (4) is to take the maximum value of each row of M in the matrix. The index map P is used to warp the reference image to align the blurry input. The confidence map C is used to weight the

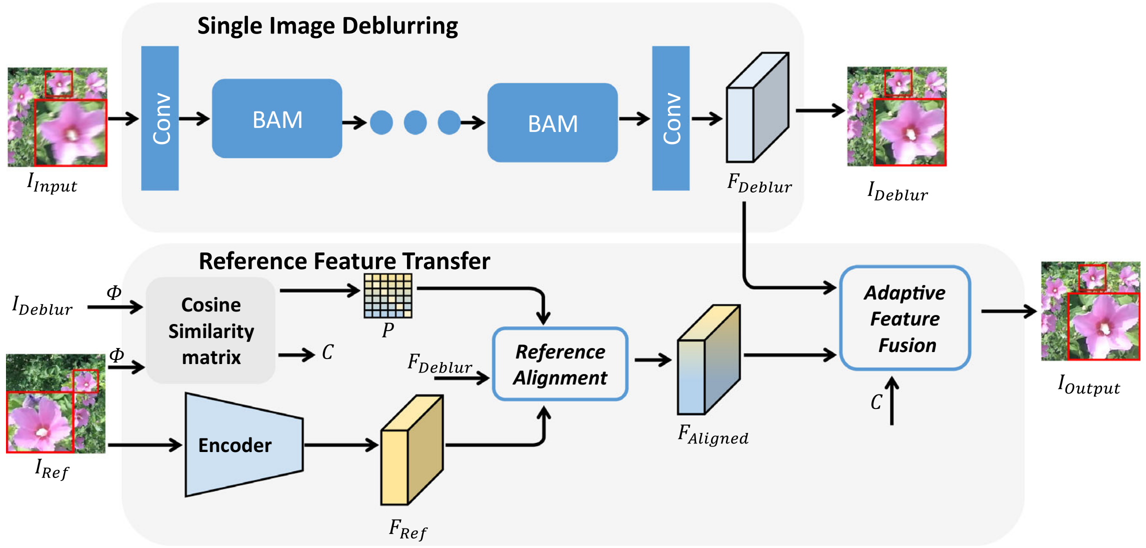


Fig. 2 The overview of the proposed deblurring framework, which contains a single image deblurring task and a reference feature transfer task. The single image deblurring network deblurs the blurry input I_{Input} to the deblurring image I_{Deblur} and obtains its feature F_{Deblur}. The reference feature transfer task first calculate the cosine similarity matrix between I_{Deblur} and the reference image I_{Ref}. The reference feature F_{Ref} is then warped by the reference alignment module. After the final adaptive feature fusion with F_{Deblur}, the framework yields a clear output I_{Output}

relevant reference features. Next we use the reference alignment module to align the images and extract well-aligned reference features F_{Aligned}, respectively:

$$F_{\text{Aligned}} = \mathcal{F}_{\text{RA}}(P, I_{\text{Ref}}, I_{\text{Deblur}}) \tag{5}$$

where $\mathcal{F}_{\text{RA}}$ represents the reference alignment module. Note that I_{Deblur} is only used to calculate the cosine similarity matrix with the reference image. Then, F_{Aligned} is aggregated with the deblurring feature F_{Deblur} through the adaptive feature fusion module, and the final deblurring result I_{Output} is obtained:

$$I_{\text{Output}} = \mathcal{F}_{\text{AFF}}(F_{\text{Aligned}}, F_{\text{Deblur}}, C) \tag{6}$$

where $\mathcal{F}_{\text{AFF}}$ represents the adaptive feature fusion module.

3.1 Feature extraction with reference alignment

The purpose of the reference alignment module is to obtain the features of the aligned blurry input, which will be used in the subsequent adaptive feature fusion. Inspired by reference-based super-resolution and video super-resolution, as shown in Fig. 3a, we adopt a combination of patch alignment and deformable alignment. The key to patch alignment is to use the index map P calculated by the cosine similarity matrix to select high-quality features in the reference feature F_{Ref}. So after getting the index map P, we need to unfold the reference feature F_{Ref} into patches. We use the Unfold operation in the Pytorch [44] framework to unfold it, where the sliding window is a 3×3 convolution kernel with a stride of 1. Then, we warp the reference feature F_{Ref} with the index map P followed by folding operation to obtain the aligned feature F_{Paligned}:

$$F_{\text{Paligned}} = \mathcal{W}(F_{\text{Ref}}, P) \tag{7}$$

where $\mathcal{W}$ represents the spatial warping operation. This warp step is used to transfer reference features according to index map P. In other words, we transfer feature patches according to the index of the most relevant position. The patch alignment method can stably find similar textures in the reference features. However, a simple patch-level alignment cannot fully exploit the similar features of the reference images [40]. Previous studies [38,42,45] have shown that deformable alignment has superior alignment performance; therefore, we introduce deformable alignment into the reference alignment module to align reference features and blurry features adaptively in the feature level [37]. Here, we are using DCNv2 [46]. Specifically, we first concatenate the reference feature F_{Ref} and the deblurred feature F_{Deblur} together to predict the offset o and modulation mask m:

$$o = \mathcal{E}_o([F_{\text{Ref}}, F_{\text{Deblur}}]) \tag{8}$$

$$m = \sigma(\mathcal{E}_m([F_{\text{Ref}}, F_{\text{Deblur}}])) \tag{9}$$

where $\mathcal{E}_o$ and $\mathcal{E}_m$ represent stacked convolutional layers, o represent activation functions, and [,] represent concatena-

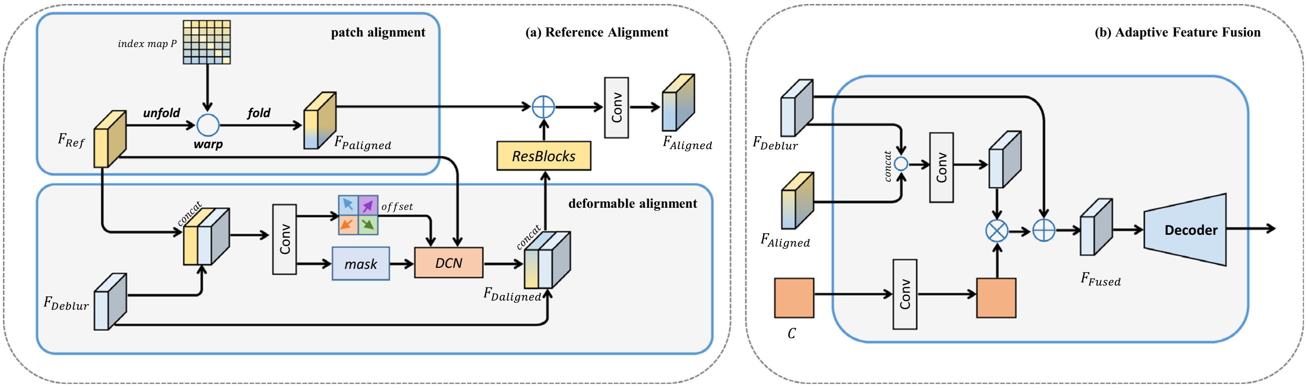


Fig. 3 Illustration of the reference alignment module and adaptive feature fusion module

tion operations. With the help of masks, we can perform the alignment operation adaptively even if the reference image and the blurry image are not the same scene. Then, we use deformable convolution $\mathcal{D}$ to compute aligned features F_{Daligned}:

$$F_{\text{Daligned}} = \mathcal{D}(F_{\text{Ref}}, o, m) \tag{10}$$

With the help of deformable convolutional offset diversity, the filter implicitly captures motion information by learning local samples. After obtaining the deformable aligned reference feature F_{Daligned}, we concatenate it with the deblurred feature F_{Deblur}, and then use the residual blocks R [47] for further feature aggregation. Finally, we fuse the aggregated feature and feature F_{Paligned} through a convolutional layer to obtain the final aligned feature F_{Aligned}:

$$F_{\text{Aligned}} = Conv(F_{\text{Paligned}} + R([F_{\text{Daligned}}, F_{\text{Deblur}}])) \tag{11}$$

where $Conv$ represents a convolutional layer. The alignment reference module not only uses stable explicit alignment of patch alignment, but also utilizes deformable convolution for implicit alignment. The combination of the two alignment methods facilitates the acquisition of better aligned features.

3.2 Adaptive feature fusion

Although the reference feature F_{Aligned} obtained by the reference alignment module has similar content to the deblurred feature F_{Deblur}, it is not optimal to simply concatenate or add them directly. This is because the alignment is not necessarily very precise and may introduce additional noise information. To effectively combine the deblurred feature F_{Deblur} and the aligned feature F_{Aligned}, as shown in Fig. 3b, we propose an adaptive feature fusion module to perform feature aggregation. We first concatenate the two together and then use the confidence map C to guide the fusion process adaptively. In order to suppress the features with inaccurate alignment and weight the high-quality aligned features, the confidence map C also uses a set of convolutional layers to aggregate the information of adjacent patches. Formally, we have:

$$F_c = Conv([F_{\text{Deblur}}, F_{\text{Aligned}}]) \otimes Conv(C) \tag{12}$$

Then, we use skip connections to synthesize the fused feature F_{fusion}:

$$F_{\text{fusion}} = F_c + F_{\text{Deblur}} \tag{13}$$

Finally, after the reconstruction of the decoder, we get the deblurred result I_{Output}.

$$I_{\text{Output}} = \text{Decoder}(F_{\text{fusion}}) \tag{14}$$

3.3 Implementation details

Our framework is mainly divided into two stages of training optimization. For the first stage, we trained the single image deblurring network separately, where the number of BAM basic blocks is 10. Our goal is to preserve the spatial structure and content information of the deblurring results. To this end, we only use the L1 loss to minimize the pixel minimum distance between the output I_{Deblur} of the single image deblurring task and the ground truth image I_{GT}:

$$\mathcal{L}_1 = \left\| I_{\text{GT}} - I_{\text{Output}} \right\|_1, \tag{15}$$

where $\|\cdot\|_1$ represents L1-norm. After training, we fixed the single image deblurring network for the second stage of training.

In the second stage training, the output result I_{Deblur} of the single image deblurring network is used to calculate the cosine similarity matrix between the single image deblurring network and the reference image I_{Ref}, and the deblurring feature F_{Deblur} is used for feature fusion to obtain the final deblurred output I_{Output}. We adopt the reconstruction loss proposed by [40]. This reconstruction loss computes the loss between the deblurred output I_{Output} and the ground truth image I_{GT} from both the low- and high-frequency domains. This loss can be defined as:

$$\mathcal{L}_{\text{rec}} = \left\| I_{\text{Output}}^{\text{filter}} - I_{\text{GT}}^{\text{filter}} \right\| + \sum_{i} \delta_i (I_{\text{Output}}, I_{\text{GT}}) \tag{16}$$

where the superscript $filter$ of the first item represents the filtering operation using a 3×3 Gaussian kernel. The second term $\delta_i(X, Y) = \min_j \mathbb{D}(x_i, y_j)$ is the contextual loss that minimizing the difference between the pixel x_i in I_{Output} and its most relevant pixel y_j in I_{GT} at the perceptual distance $\mathbb{D}$ [48–50]. The first term makes the deblurred output stably follow the low-frequency structure of the ground truth image, and the second term flexibly maximizes the similarity between I_{Output} and I_{GT}, improving the perceptual visual quality.

It took about 150 h to train the single image deblurring task. We trained the entire framework using the Adam optimizer [51], where = 0.9, = 0.999. It took about 80 h for our framework to converge. The batch size is 12. The initial learning rate was 0.0001 and gradually decreased using a cosine annealing strategy. The first four layers of the vgg19 pre-trained model were used for feature extraction. We built and trained the network framework using Pytorch [44] on an NVIDIA TITAN RTX GPU.

4 Experiments

4.1 Datasets and metrics

Datasets. In the experiments, we use the most popular GoPro [9] dataset. The training set and test set of the GoPro dataset are generated in the same way. Their ground truth images are all taken with a GoPro4 high frame rate camera, and the corresponding blurry images are generated by averaging consecutive ground truth frames. The dataset has a total of 3214 pairs of sharp and blurry images, of which 2103 pairs are used for training and 1111 pairs are used for testing. The resolution of the blurry and ground truth images is 1280 × 720. To evaluate the generalization ability of the model, we also tested on the HIDE [52] test set, which consists of 2025 images. To evaluate the ability of the model to handle real-world blur, we chose the high-quality RealBlur-J [53] test set, which contains 980 blurry images in low light, for our experiments. For the selection of sharp reference images, we choose the adjacent frames of ground truth images as reference images to facilitate the restoration of blurry images. The same is true for the HIDE and RealBlur-J datasets. All models in the experiments were trained on the GoPro training set.

Table 1 Quantitative evaluation results on the GoPro and HIDE test sets. The best scores are shown in bold. ∗ indicates that the author does not release source codes

	GoPro		HIDE	
Method	PSNR	SSIM	PSNR	SSIM
DeepDeblur [9]	27.83	0.915	25.73	0.874
SRN [10]	30.25	0.935	28.36	0.904
DeblurGAN-v2 [19]	29.55	0.934	27.40	0.882
MTRNN [31]	31.13	0.945	29.15	0.918
DSD [11]	30.96	0.942	29.01	0.913
Stack4-DMPHN [12]	31.39	0.948	29.10	0.918
DBGAN [55]	31.10	0.942	28.94	0.915
RADN∗ [16]	31.76	0.953	–	–
SAPHN∗ [56]	32.02	0.953	29.98	0.930
SimpleNet∗ [32]	31.52	0.950	–	–
MPRNet [15]	32.66	0.959	30.96	0.939
MIMO-UNet+ [14]	32.45	0.957	30.00	0.930
HINet [13]	32.71	0.959	30.33	0.934
Ours	**33.35**	**0.963**	**31.02**	**0.940**

Evaluation metrics. Like existing baseline deblurring methods, we use PSNR and structural similarity (SSIM) [54] to evaluate all experimental results. In general, larger PSNR and SSIM represent higher-quality restored images. All PSNR and SSIM in the experiments are calculated using built-in functions in MATLAB R2018a.

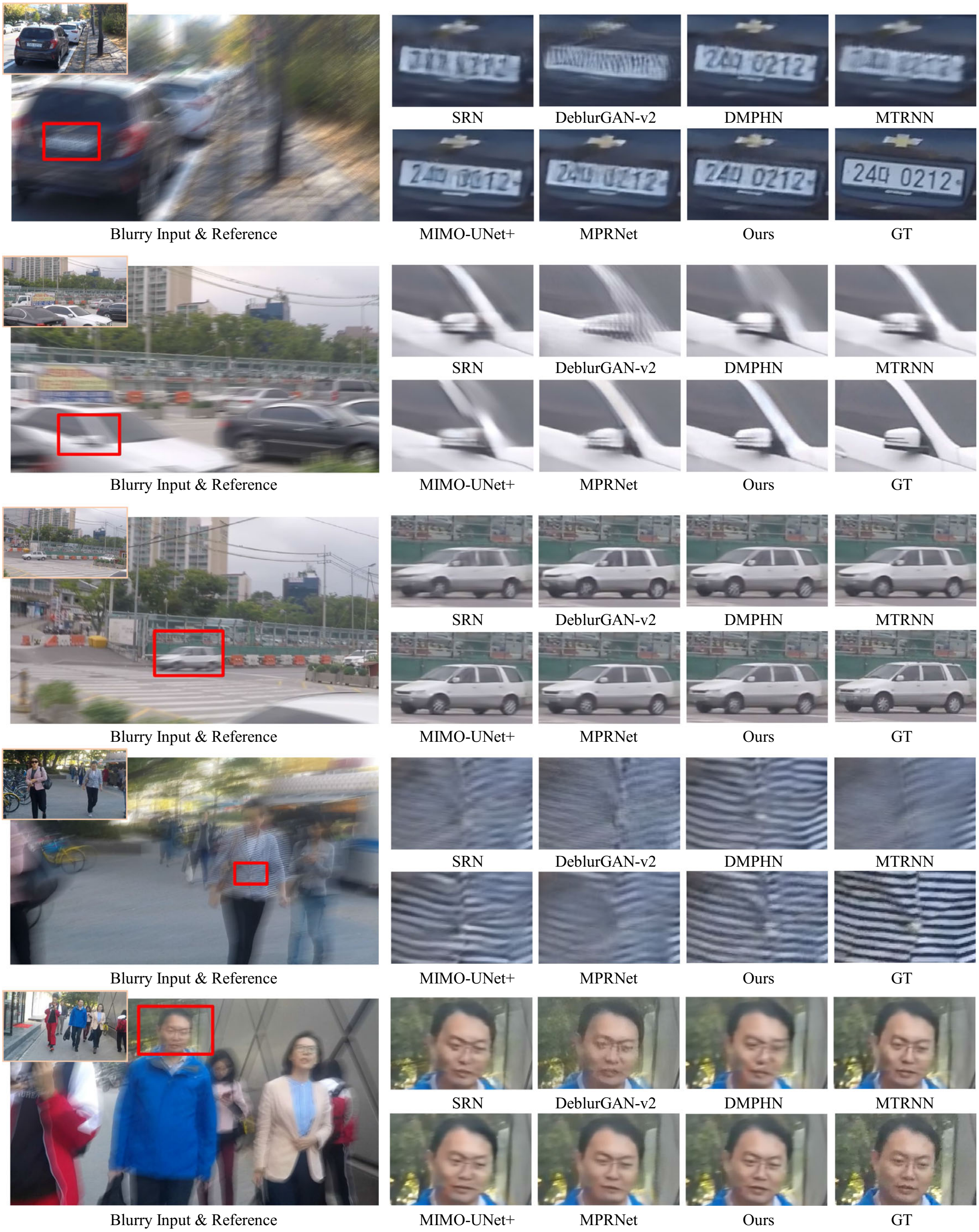
4.2 Comparisons with state-of-the-art methods

We first quantitatively compare the proposed method with several state-of-the-art methods, including DeepDeblur [9], SRN [10], DeblurGAN-v2 [19], DSD [11], MTRNN [31], DMPHN [12], DBGAN [55], RADN [16], SAPHN [56], SimpleNet [32], MPRNet [15], MIMO-UNet+ [14], HINet [13]. The test models of the above methods are all trained under the GoPro training set and tested under the GoPro, HIDE and RealBlur test sets. In addition to this, we also conduct qualitative evaluations and user study to measure the visual performance of different methods.

Quantitative evaluations. As shown in Table 1, we provide quantitative evaluation results on the GoPro and HIDE test set. As can be seen, our method achieves the highest PSNR and SSIM scores, surpassing all the other methods by at least 0.6 dB on the GoPro test dataset. This shows that with the help of reference images, our framework achieves state-of-the-art deblurring performance and generalization ability. Single image deblurring methods cannot utilize additional reference information. In contrast, our method can adaptively utilize useful information on the reference image. The quantitative

Table 2 Quantitative evaluation results on the RealBlur test set. The best scores are shown in bold

	SRN [10]	DeblurGAN-v2 [19]	DMPHN [12]	MPRNet [15]	MIMO-UNet+ [14]	Ours
PSNR	26.55	26.65	26.05	26.49	26.12	**26.69**
SSIM	0.864	0.862	0.856	0.865	0.853	**0.865**



SRN DeblurGAN-v2 DMPHN MTRNN

Blurry Input & Reference MIMO-UNet+ MPRNet Ours GT

SRN DeblurGAN-v2 DMPHN MTRNN

Blurry Input & Reference MIMO-UNet+ MPRNet Ours GT

SRN DeblurGAN-v2 DMPHN MTRNN

Blurry Input & Reference MIMO-UNet+ MPRNet Ours GT

SRN DeblurGAN-v2 DMPHN MTRNN

Blurry Input & Reference MIMO-UNet+ MPRNet Ours GT

SRN DeblurGAN-v2 DMPHN MTRNN

Blurry Input & Reference MIMO-UNet+ MPRNet Ours GT

Fig. 4 Visual comparison on the GoPro test dataset(top three examples) and HIDE test dataset(the forth and fifth examples). Our method recovers fine textures and major structures in text, textures, moving objects and human faces. Zoom-in for details

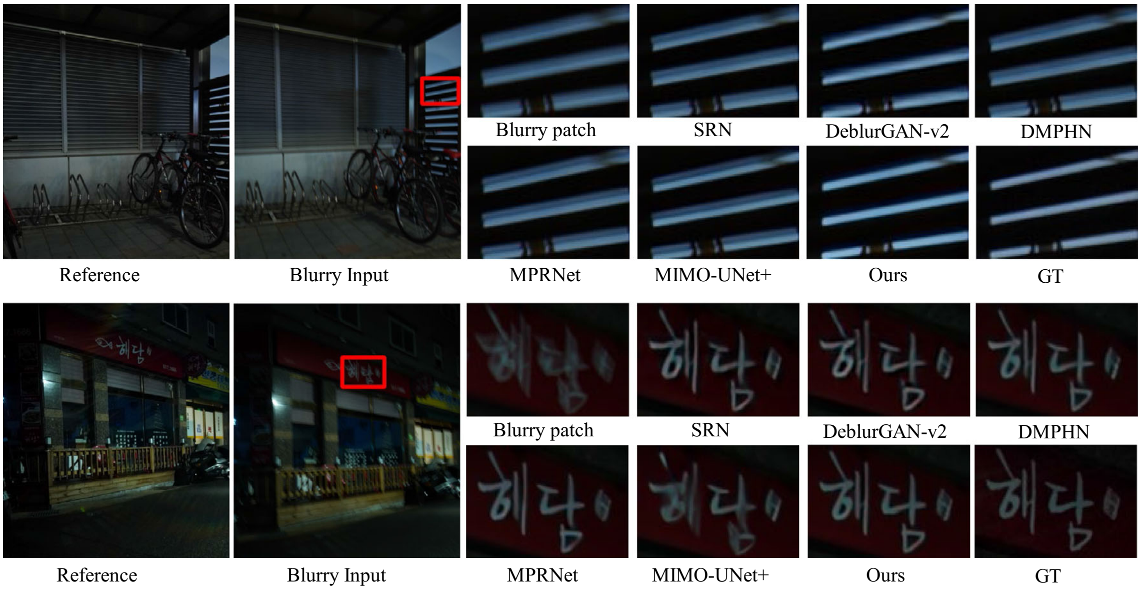


Fig. 5 Visual comparison on the RealBlur test dataset. Our method recovers more realistic details than other methods on blurry images in low light. Zoom-in for details

comparison results on RealBlur-J are shown in Table 2, due to the inconsistent way in which the training and test set data are generated, the metrics of the other state-of-the-art methods drop very low, but our method still achieves the highest performance. In addition, our framework does not require the use of complex multi-scale, multi-stage or multi-patch strategies. The above quantitative comparison results show that our approach achieves the best performance.

Qualitative evaluations. We further compare the proposed method with other state-of-the-art methods for visual quality. Among the best deblurring methods, we choose MPRNet [15], MIMO-UNet+ [14], SRN[10], DeblurGAN-v2 [19], MTRNN [31], DMPHN [12] as comparison methods. Thanks to the authors of these methods providing better quality source codes, we can make a fair comparison. The visual comparison results on the GoPro and HIDE test sets are shown in Fig. 4. All five examples have considerable restoration challenges, making the current single image deblurring algorithms intractable. Specifically, the top three examples were selected from the GoPro test set, and we selected texts with large motion blur, textures, and objects in high-speed motion for comparison. In the first example, the results of other methods produce severe artifacts, and some are even difficult to recognize, yet our method can still recover recognizable digital text. The second example contains texture details, but DeblurGAN-v2 [19] produces severe artifacts, MIMO-UNet+ [14] produces a certain scale distortion, and MPRNet [15], although the restoration effect is better than other methods, is still worse than our deblurring effect. Our method can effectively suppress blur diffusion and artifacts. The third example is high-speed motion blur, and our method still produces results that are closest to ground truth images. The last two examples are selected from the HIDE test set, and our method still achieves good visual performance. For example, in the last example, our method focuses more on reconstructing face shape and details. In contrast, due to severe motion shake, the high-frequency details of the image are lost, and it is difficult for other methods to recover clear facial information. Figure 5 shows the visual comparison results on the RealBlur test set; our method recovers more details in low-light blurry regions compared to other methods. Overall, with the help of reference images, our method recovers major structures and fine details on both synthetic and real-world datasets.

User study. To further evaluate the visual perceptual quality of the deblurring results, we also conduct a user study comparison of the deblurred results with three state-of-the-art methods. We choose DMPHN [12], MIMO [14] and MPRNet [15] as baseline comparison methods. The user study consisted of 30 users who had normal vision and were not aware of any experimental details, so it was objective. The selected images are derived from different scenes in the GoPro test set and are universal. We give users two images (ours and baseline) at a time and let users choose the image they think is the most realistic without a time limit. They do not know which algorithm has been used to recover the image. We collected 300 valid images for each set of comparisons. As shown in Fig. 7, more than 80% of people like our deblurring results, which again strongly proves that our method has better subjective quality.

Table 3 Quantitative evaluation for ablation study of reference feature transfer task. The PSNR is computed on GoPro

ID	Patch	DCN	AFF	PSNR
(1)				32.81
(2)	✓			32.99
(3)		✓		33.15
(4)	✓	✓		33.23
(5)	✓	✓	✓	33.35

4.3 Ablation study

4.3.1 Effect of reference feature transfer

Based on the single image deblurring task, we introduce a reference feature transfer task to improve the texture details of the deblurring results. Among them, in the reference feature transfer task, the reference module and the adaptive feature fusion module are essential, so we conduct ablation experiments on these two key modules. We retrained five network variants by adding important components of the model one by one: (1) single image deblurring task without reference image. (2) reference alignment module with patch alignment only. (3) reference alignment module with deformable alignment only. (4) reference alignment module based on patch alignment and deformable alignment. (5) entire framework with adaptive feature fusion. Each ablation experiment was trained for about 80 h.

Reference Alignment Module. Table 3 evaluates the effectiveness of the reference alignment module, compared with the baseline (1), (2) and (3) demonstrate the performance gain of utilizing only patch alignment or deformable alignment. (3) has better performance than (2), which indicates that deformable alignment can warp reference features better than patch alignment. As shown in Fig. 6, the patch alignment and the deformable alignment have different deblurring effects on the edges of license plate numbers, so we combine them in the reference alignment module. (4) achieves better performance, revealing that patch alignment and deformable alignment have a synergistic effect, resulting in better gain. Figure 6 also demonstrates that result (4) produces a clearer deblurred result. In summary, the reference alignment module combines the stability of patch alignment with the superior performance of deformable alignment to achieve better reference alignment.

The last row of Table 3 shows the performance of using adaptive feature fusion. For other control groups, we perform element-wise summation of deblurred features F_{Deblur} and aligned features F_{Aligned} without using confidence maps. As shown in Table 3, with adaptive feature fusion, the model achieves a gain of 0.12 dB in terms of PSNR, which indicates that the guidance of the confidence map is beneficial to the performance improvement. As shown in Fig. 6, adaptive feature fusion module further improves the deblurring effect, resulting in sharper structures and realistic textures (Fig. 7).

4.3.2 Robustness to different reference images

Previous experiments have demonstrated that sharp reference images have large gains in deblurring performance. We use the similar texture information of the reference image to help

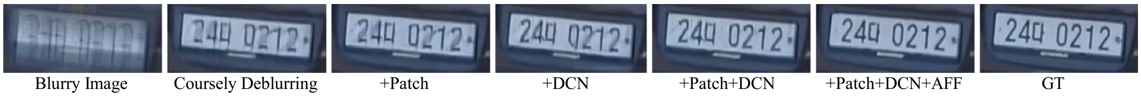
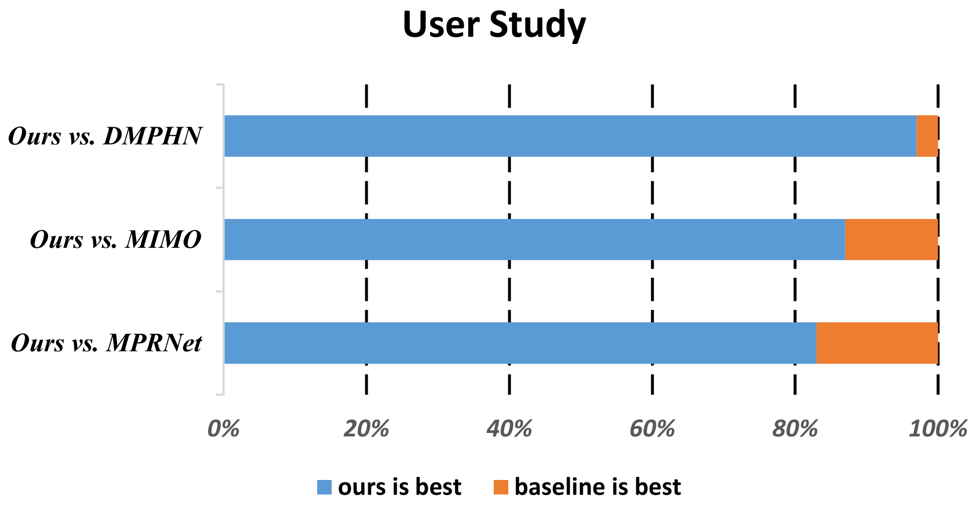


Fig. 6 Ablation study on reference feature transfer task. Note that " +" denote "with", "GT" stand for ground truth

Fig. 7 User study results on GoPro dataset



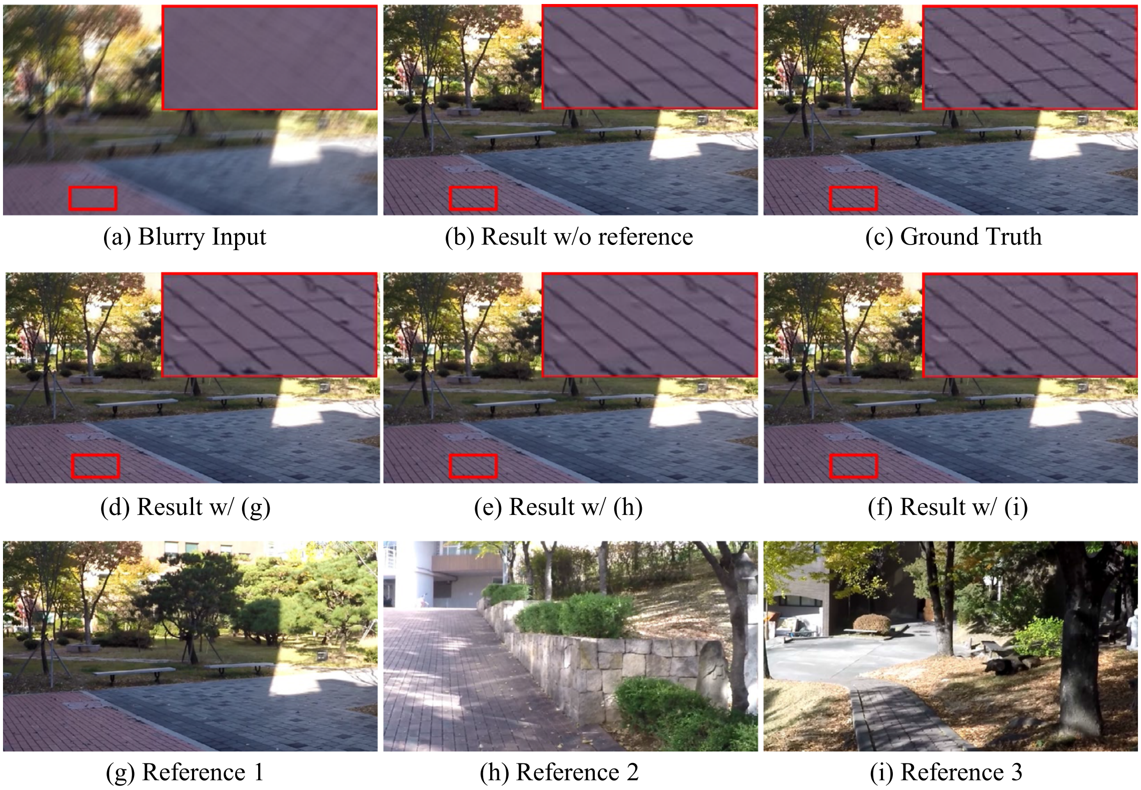


Fig. 8 Ablation result of using different reference images on the GoPro dataset. The proposed framework generates sharper results and fewer artifacts than the result without reference. Zoom-in for details

Table 4 Ablation study on different reference images

	Reference1	Reference2	Reference3	W/O Reference
PSNR	33.23	32.45	32.48	32.19
SSIM	0.963	0.957	0.957	0.955

image restoration, but do not analyze the influence of reference images in different scenes on the deblurring results. So to answer this question, we experimented with different reference images under the GoPro test set. As shown in Fig. 8, the Reference 1 is from the same scene as the blurry input and has a high degree of similarity, while the other reference images are randomly selected from other GoPro scenes. It can be seen that with the help of different reference images, our method achieves superior deblurring results. This is because patch alignment is a special attention mechanism, which can find and weight the most relevant texture features under the guidance of index map P and confidence map C.

We further conduct quantitative comparisons, as shown in Table 4, even though the reference image and the blurry image originate from different scenes; our method also has only a slight performance degradation. This means that our framework can robustly utilize reference images in different scenes to facilitate the restoration of deblurred images.

4.4 Object detection performance evaluation

As mentioned in the previous introduction, image blur can seriously affect other computer vision tasks, object detection is one of them. As one of the most basic and challenging problems in computer vision, object detection [57,58] has received extensive attention in recent years. With the development of deep learning, methods based on deep learning have been significantly improved in the field of object detection. Among them, the YOLO [59] series algorithms have gradually become the benchmark algorithms in the industry due to their better performance. However, most of the object detection algorithms assume that the input image is clear, so when blurred images are used as input, these methods all face severe performance degradation. At this point, the image deblurring technology can be applied to the object detection task to remove image blur and improve the accuracy of object detection. As shown in Fig. 9, we use the YOLO V5 object detector to detect the blurry and deblurred images, respectively. The first column is the result of detection using the

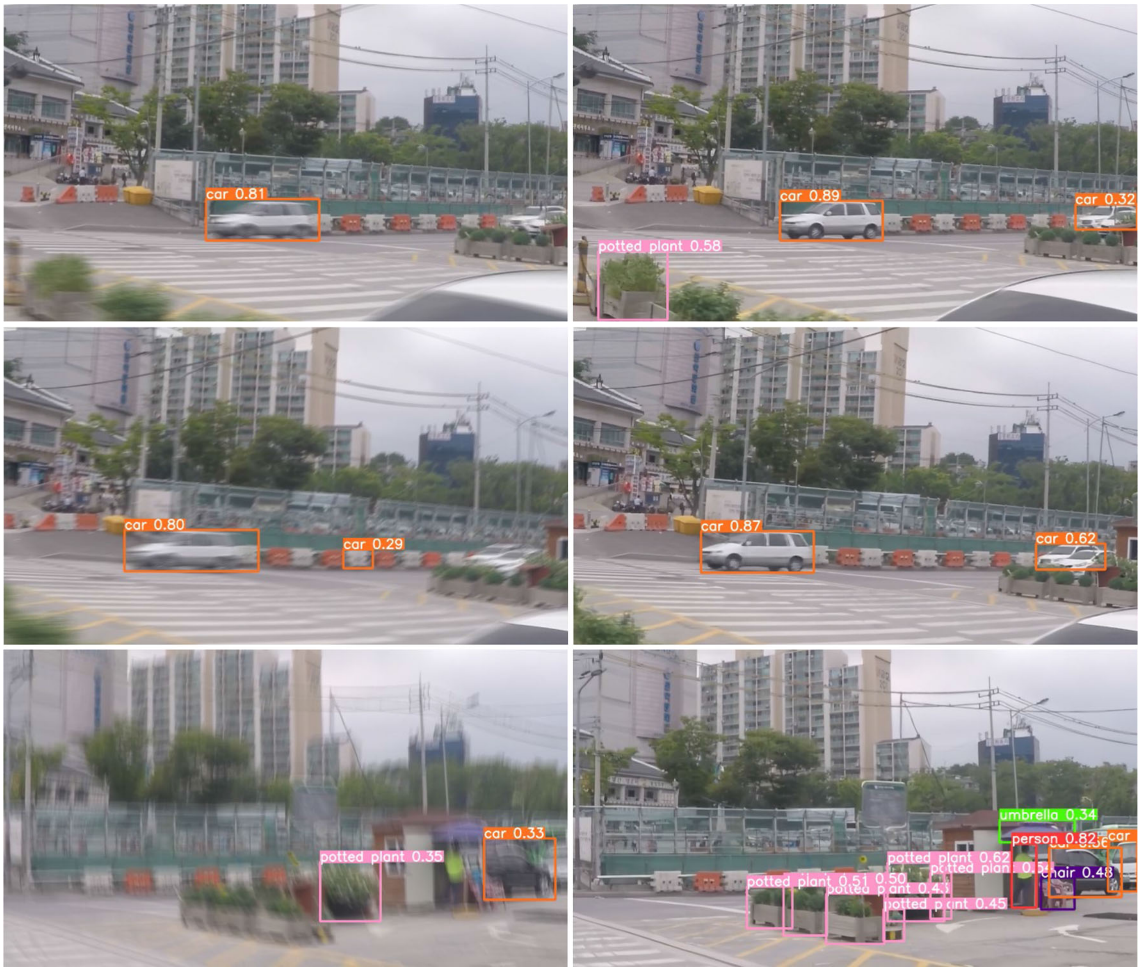


Fig. 9 The result of object detection. The first column is the detection result of the blurry input, and the second column is the detection result after deblurring by our method

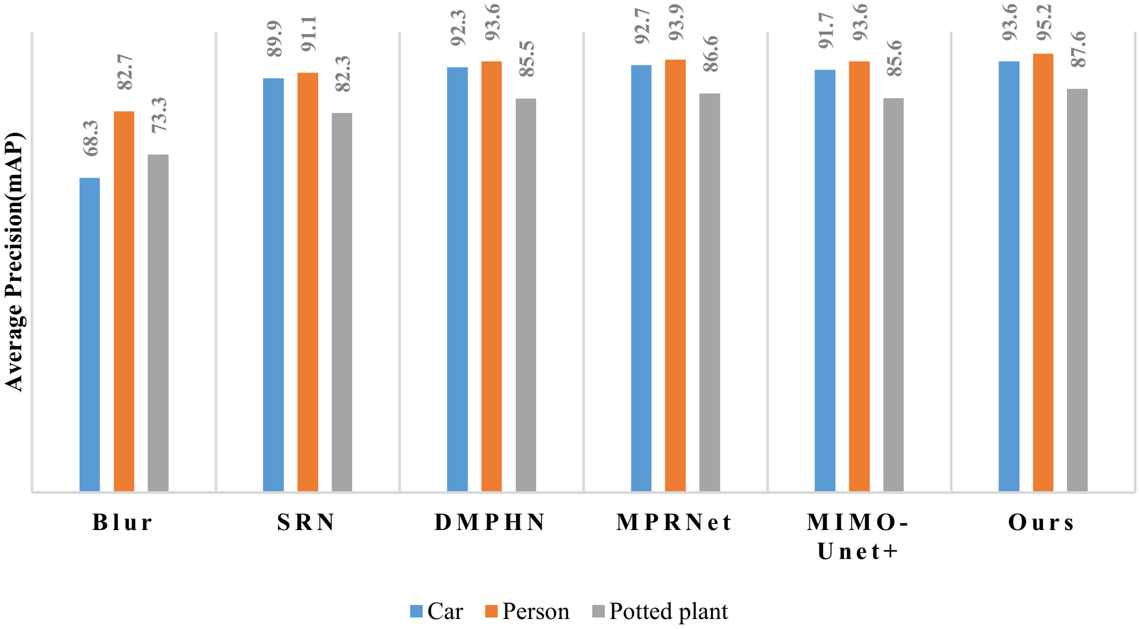


Fig. 10 Object detection performance evaluation

Table 5 Performance and efficiency comparison on the GoPro test dataset. We tested all the experimental metrics using an Nvidia Titan RTX GPU

Method	PSNR	FLOPs(G)	Runtimes(s)	Params(M)
DeepDeblur [9]	27.83	336	1.52	11.7
SRN [10]	30.25	167	0.83	6.8
MTRNN [31]	31.13	164	0.33	2.6
DSD [11]	30.96	471	1.31	2.84
Stack4-DMPHN [12]	31.39	235	0.49	21.7
MPRNet [15]	32.66	777	1.12	20.1
MIMO-UNet+ [14]	32.45	154	0.36	16.1
Ours	33.35	982	1.37	50.2

blurry image directly, and the second column is the detection result after image deblurring using our method. The first column clearly has many undetected objects and false detections; in contrast, the false negative examples are successfully detected in the second column.

Moreover, we further evaluate the proposed method with other deblurring methods in terms of performance improvement for object detection. Since the objects in the GoPro dataset are mostly people, cars, and potted plants, we only measure the average precision of these three classes for performance evaluation. As shown in Fig. 10, our method has the most obvious performance improvement for object detection, showing its superior performance.

4.5 Performance and efficiency comparison

In addition to quantitative and qualitative comparisons, we also compare the number of parameters, FLOPs and running time with state-of-the-art methods. Table 5 shows the experimental results. Compared with other single image deblurring methods, our method has relatively large FLOPs and parameter amount due to the integration of two different tasks. In addition, when testing, we divide the entire image into multiple patches for inference separately, and then splicing them into a whole image for output. Compared with other methods that input an entire image into the network for inference, this method of patch testing increases the inference speed to a certain extent. Compared with other methods, our method achieves the most excellent performance while maintaining an acceptable efficiency.

4.6 Influence of the batch size and initial learning rate

During the training phase of the whole framework, batch size and initial learning rate are important hyper-parameters that affect model performance. Specifically, increasing the batch size within an appropriate range can improve the stability of model convergence. If the initial learning rate is too large, the model will not converge, and if it is too small, the model will converge very slowly or cannot learn. Thus, we analyze the influence of these two hyper-parameters on the convergence of the model on the GoPro dataset. Figure 11 shows the ablation results.

Figure 11a shows the impact of batch size on the convergence speed of the model. It can be seen that as the batch size increases, the convergence speed of the model gradually increases. Therefore, we adjust the batch size to 12 to fully utilize the performance of the GPU.

Figure 11b shows the impact of the initial learning rate on model convergence. It can be seen that the model is more sensitive to the initial learning rate, too large or too small will seriously affect the performance, so we choose the initial learning rate as 1×10^{-4}.

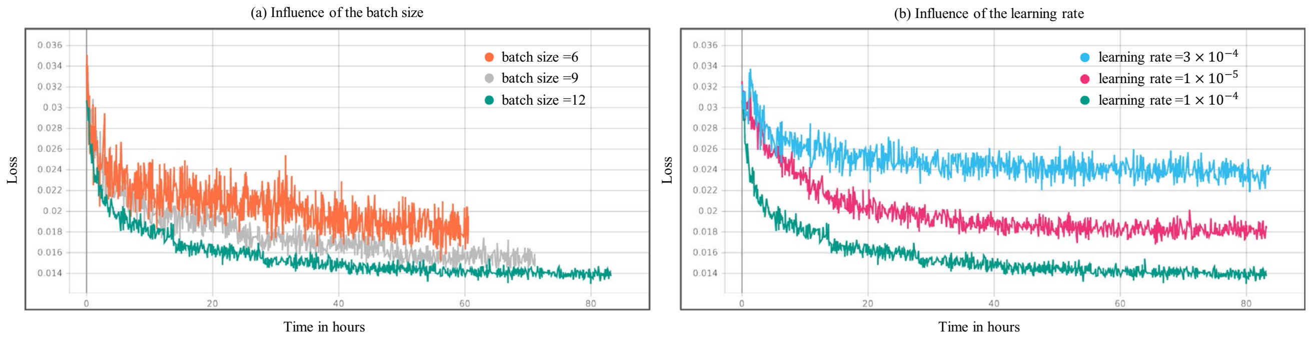


Fig. 11 Influence of different batch sizes and initial learning rates. **a** Influence of batch size. **b** Influence of learning rate

5 Conclusion

In this paper, we propose an effective framework for deblurring with reference images. The framework mainly includes a single image deblurring task and a reference feature transfer task. The single image deblurring task recovers a rough deblurred image from the blurry input, which will be used to compute the cosine similarity matrix with the reference image. The reference feature transfer task finally synthesizes high-quality deblurred results with the help of a well-designed reference alignment module and an adaptive feature fusion module. Quantitative and qualitative experimental results on synthetic and real-world datasets demonstrate that our framework achieves superior performance.

Limitations and Future Work. Although our framework achieves state-of-the-art performance, computing the cosine similarity matrix between the reference image and the single image deblurring results is memory-intensive. In addition, compared to the single image deblurring task, the framework yields expensive computation and inference speed due to the integration of the single image deblurring task and the reference feature transfer task. In the future, we are interested in exploring the use of lighter-weight single image deblurring models to trade off performance and computational speed.

Acknowledgements The authors acknowledge the National Natural Science Foundation of China (61772319, 62002200, 62272281 and 61972235), and Shandong Natural Science Foundation of China (ZR2021MF107, ZR2022MA076), Shandong Provincial Science and Technology Support Program of Youth Innovation Team in Colleges (2021KJ069, 2019KJN042), Yantai science and technology innovation development plan(2022JCYJ031).

Data Availability Data available on request from the authors.

Declarations

Conflict of interest Cunzhe Liu, Zhen Hua and Jinjiang Li declare that they have no conflict of interest.

References

1. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: Deblurgan: blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8183–8192 (2018)
2. Shen, Z., Lai, W.S., Xu, T., Kautz, J., Yang, M.H.: Deep semantic face deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8260–8269 (2018)
3. Pan, J., Sun, D., Pfister, H., Yang, M.H.: Blind image deblurring using dark channel prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1628–1636 (2016)
4. Sun, L., Cho, S., Wang, J., Hays, J.: Edge-based blur kernel estimation using patch priors. In: IEEE international conference on computational photography (ICCP), IEEE, pp. 1–8 (2013)
5. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: CVPR 2011, IEEE, pp. 233–240 (2011)
6. Whyte, O., Sivic, J., Zisserman, A.: Deblurring shaken and partially saturated images. Int. J. Comput. Vis. **110**(2), 185–201 (2014)
7. Xu, L., Zheng, S., Jia, J.: Unnatural l0 sparse representation for natural image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1107–1114 (2013)
8. Xu, X., Pan, J., Zhang, Y.J., Yang, M.H.: Motion blur kernel estimation via deep learning. IEEE Trans. Image Process. **27**(1), 194–205 (2018). https://doi.org/10.1109/TIP.2017.2753658
9. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 257–265 (2017)
10. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8174–8182 (2018)
11. Gao, H., Tao, X., Shen, X., Jia, J.: Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3843–3851 (2019)
12. Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 5971–5979 (2019)
13. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: Hinet: Half instance normalization network for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 182–192 (2021)
14. Cho, S.J., Ji, S.W., Hong, J.P., Jung, S.W., Ko, S.J.: Rethinking coarse-to-fine approach in single image deblurring. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 4641–4650 (2021)
15. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 14821–14831 (2021)
16. Purohit, K., Rajagopalan, A.: Region-adaptive dense network for efficient motion deblurring. Proc. AAAI Conf. Artif. Intell. **34**, 11882–11889 (2020)
17. Zhang, K., Ren, W., Luo, W., Lai, W.S., Stenger, B., Yang, M.H., Li, H.: Deep image deblurring: a survey. Int. J. Comput. Vis. **130**(9), 2103–2130 (2022)
18. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: Deblurgan: Blind motion deblurring using conditional adversarial networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8183–8192 (2018)
19. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp 8877–8886 (2019)
20. Zhang, K., Luo, W., Zhong, Y., Ma, L., Liu, W., Li, H.: Adversarial spatio-temporal learning for video deblurring. IEEE Trans. Image Process. **28**(1), 291–301 (2018)
21. Zhang, K., Luo, W., Stenger, B., Ren, W., Ma, L., Li, H.: Every moment matters: Detail-aware networks to bring a blurry image alive. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 384–392 (2020)
22. Zheng, H., Ji, M., Han, L., Xu, Z., Wang, H., Liu, Y., Fang, L. Learning cross-scale correspondence and patch-based synthesis for reference-based super-resolution. In: BMVC, vol 1, p 2 (2017)
23. Zheng, H., Ji, M., Wang, H., Liu, Y., Fang, L.: Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In: Proceedings of the European conference on computer vision (ECCV), pp. 88–104 (2018)

24. Zhang, Z., Wang, Z., Lin, Z., Qi, H.: Image super-resolution by neural texture transfer. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7974–7983 (2019)
25. Yang, F., Yang, H., Fu, J., Lu, H., Guo, B.: Learning texture transformer network for image super-resolution. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5790–5799 (2020)
26. Lu, L., Li, W., Tao, X., Lu, J., Jia, J.: Masa-sr: Matching acceleration and spatial adaptation for reference-based image super-resolution. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6364–6373 (2021)
27. Cao, J., Liang, J., Zhang, K., Li, Y., Zhang, Y., Wang, W., Van Gool, L.: Reference-based image super-resolution with deformable attention transformer. In: European conference on computer vision (2022)
28. Huang, Y., Zhang, X., Fu, Y., Chen, S., Zhang, Y., Wang, Y.F., He, D.: Task decoupled framework for reference-based super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5931–5940 (2022)
29. Sun, L., Sakaridis, C., Liang, J., Jiang, Q., Yang, K., Sun, P., Ye, Y., Wang, K., Gool, L.V.: Event-based fusion for motion deblurring with cross-modal attention. In: European Conference on Computer Vision, Springer, pp 412–428 (2022)
30. Sun, J., Cao, W., Xu, Z., Ponce, J.: Learning a convolutional neural network for non-uniform motion blur removal. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 769–777 (2015)
31. Park, D., Kang, D.U., Kim, J., Chun, S.Y.: Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In: European Conference on Computer Vision, Springer, pp 327–343 (2020)
32. Li, J., Tan, W., Yan, B.: Perceptual variousness motion deblurring with light global context refinement. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp 4096–4105 (2021)
33. Niu, W., Zhang, K., Luo, W., Zhong, Y.: Blind motion deblurring super-resolution: when dynamic spatio-temporal learning meets static image understanding. IEEE Trans. Image Process. **30**, 7101–7111 (2021)
34. Tu, Z., Xie, W., Cao, J., Van Gemeren, C., Poppe, R., Veltkamp, R.C.: Variational method for joint optical flow estimation and edge-aware image restoration. Pattern Recognit. **65**, 11–25 (2017)
35. Tu, Z., Poppe, R., Veltkamp, R.: Estimating accurate optical flow in the presence of motion blur. J. Electron. Imaging **24**(5), 053018 (2015)
36. Zhang, D., He, L., Tu, Z., Zhang, S., Han, F., Yang, B.: Learning motion representation for real-time spatio-temporal action localization. Pattern Recognit. **103**, 107312 (2020)
37. Tian, Y., Zhang, Y., Fu, Y., Xu, C.: Tdan: temporally-deformable alignment network for video super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3360–3369 (2020)
38. Wang, X., Chan, K.C., Yu, K., Dong, C., Change Loy, C.: Edvr: video restoration with enhanced deformable convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2019)
39. Shim, G., Park, J., Kweon, I.S.: Robust reference-based super-resolution with similarity-aware deformable convolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8425–8434 (2020)
40. Wang, T., Xie, J., Sun, W., Yan, Q., Chen, Q.: Dual-camera super-resolution with aligned attention modules. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1981–1990, (2021)
41. Abuolaim, A., Afifi, M., Brown, M.S.: Improving single-image defocus deblurring: how dual-pixel images help through multi-task learning. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1231–1239 (2022)
42. Chan, K.C., Wang, X., Yu, K., Dong, C., Loy, C.C.: Understanding deformable alignment in video super-resolution. Proc. AAAI Conf. Artif. Intell. **35**, 973–981 (2021)
43. Tsai, F.J., Peng, Y.T., Lin, Y.Y., Tsai, C.C., Lin, C.W.: Banet: Blur-aware attention networks for dynamic scene deblurring. In: arXiv preprint arXiv:2101.07518 (2021)
44. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. **32**, 8046 (2019)
45. Chan, K.C., Zhou, S., Xu, X., Loy, C.C.: Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5972–5981 (2022)
46. Zhu, X., Hu, H., Lin, S., Dai, J.: Deformable convnets v2: More deformable, better results. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9308–9316 (2019)
47. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778 (2016)
48. Mechrez, R., Talmi, I., Zelnik-Manor, L.: The contextual loss for image transformation with non-aligned data. In: Proceedings of the European conference on computer vision (ECCV), pp. 768–783 (2018a)
49. Mechrez, R., Talmi, I., Shama, F., Zelnik-Manor, L.: Maintaining natural image statistics with the contextual loss. In: Asian Conference on Computer Vision, Springer, pp. 427–443 (2018b)
50. Zhang, X., Chen, Q., Ng, R., Koltun, V.: Zoom to learn, learn to zoom. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3762–3770 (2019)
51. Kingma, D.P., Ba, J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
52. Shen, Z., Wang, W., Lu, X., Shen, J., Ling, H., Xu, T., Shao, L.: Human-aware motion deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5572–5581 (2019)
53. Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: European Conference on Computer Vision, Springer, pp. 184–201 (2020)
54. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
55. Zhang, K., Luo, W., Zhong, Y., Ma, L., Stenger, B., Liu, W., Li, H.: Deblurring by realistic blurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2737–2746 (2020)
56. Suin, M., Purohit, K., Rajagopalan, A.N.: Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 3603–3612 (2020)
57. Wei, L., Cui, W., Hu, Z., Sun, H., Hou, S.: A single-shot multi-level feature reused neural network for object detection. Vis. Comput. **37**(1), 133–142 (2021)
58. Zhang, H., Xu, M., Zhuo, L., Havyarimana, V.: A novel optimization framework for salient object detection. Vis. Comput. **32**(1), 31–41 (2016)
59. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788 (2016)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cunzhe Liu received his B.S. degree in Computer Science and Technology from Taishan University, Taian, China, in 2020. Currently, he is a M.S. degrees candidate in the School of Information and electronic engineering, Shandong Technology and Business University, Yantai, China. His research interests include image deblurring, image processing, and computer vision.

Zhen Hua received the B.S. and M.S. degrees in electrical automation from Taiyuan University of Technology, Taiyuan, China, in 1989 and 1992, respectively, and the Ph.D. degree in electronic information engineering from China University of Mining and Technology, Beijing, China, in 2008. She is currently a professor at Shandong Technology and Business University. Her research interests include computer aided geometric design, information visualization, virtual reality, and image processing.

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, and 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.