IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 16, 2020, accepted September 24, 2020, date of publication October 6, 2020, date of current version October 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028157

Semantic Information Supplementary Pyramid Network for Dynamic Scene Deblurring

YIMING LIU[1,2], YIFEI LUO[3], WENZHUO HUANG[4], YING QIAO[5], JUNHUI LI[4], DAHONG XU[4], AND DUQIANG LUO[1,2]

[1]Department of Bioinformatics, College of Life Science, Hebei University, Baoding 071000, China
[2]Institute of Life Science and Green Development, Hebei University, Baoding 071000, China
[3]Faculty of Arts and Social Science, University of New South Wales, Sydney, NSW 2052, Australia
[4]College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China
[5]School of Arts and Science, Rutgers University, New Brunswick, NJ 08901, USA

Corresponding author: Duqiang Luo (duqiangluo999@126.com)

ABSTRACT The algorithm in this paper is called semantic information supplementary pyramid network(SIS-net). We choose Generative Adversarial Network (GAN) as its fundamental model. SIS-net's generator imitates the feature pyramid network (FPN) structure to recycle features spanning across multiple receptive scales to restore a sharp image. However, to solve the problem caused by the phenomenon of semantic dilution in the FPN network, we have innovatively designed a semantic information supplement (SIS) mechanism. SIS mechanism contains two essential components: semantic information storage box (info-box) and feature fusion expanding. In the process of feature fusion expanding, the semantic information features coming from the info-box is supplemented to make greater use of detailed clues. In addition, SIS-net uses the intermediate layer path to extract image features in a single time to obtain a multi-scale effect. The running speed of SIS-net has obvious advantages over other algorithms, and can basically complete real-time deblurring tasks. Extensive experiments show that our SIS-net achieves both qualitative and quantitative improvements against state-of-the-art methods. The code is available at https://github.com/yimingliu123/SIS-NET

INDEX TERMS Generative adversarial network, feature pyramid network, semantic information.

I. INTRODUCTION

The goal of image deblurring is to recover a sharp latent image with the necessary edge and details when a single blurred image is given. Image deblurring has long been an essential task in computer vision and image processing. In [1], it is shown that standard network models, trained only on high-quality images, suffer significant degradation in performance when applied to the blur dataset. Thus, there is a serious need to tackle the issue of blur in images.

In practice, dynamic blur is more realistic and challenging since spatially variant blur is the combined effect of multiple factors, such as camera shake, defocusing, and object movement occurs during shooting. Also, there is currently no satisfactory compromise because of the choice of exposure time. To obtain images with rich colors and low noise, we selected a long exposure method with high sensitivity and low sensor. In a high dynamic scene, with the increase of exposure time, the blur caused by object movement is inevitable.

With the ''coarse-to-fine'' scheme has been extended to deep CNNs scenarios, the deblurring algorithm using a multi-scale network structure recently has achieved breakthroughs by exploiting the deblurring cues at different processing levels. Under the coarse-to-fine scheme, most networks import a heavy parameter weight, an suffering from a slow running speed. Such as multi-scale (Deepblur) [2] scale-recurrence Network(SRN) [3], DMPHN [4] and DeblurGAN [1]. To solve the problem of time-consuming, we choose Generative Adversarial Networks (GANs) [5] as its fundamental model and introduced a feature pyramid network (FPN) [6] to replace the multi-scale method in the generator. This FPN method constructs an intermediate path to actively collect the intermediate layer features obtained from the extraction path. This design can not only use multi-scale feature information but also reduce the network weight. The excellent results have been achieved in both angles of effect and speed, as shown in Figure 1.

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

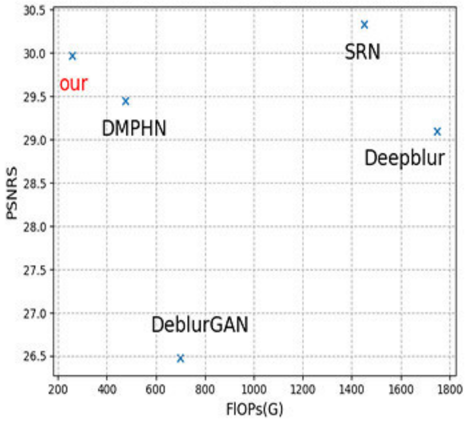


FIGURE 1. Compared to four state-of-the-art competitors (in blue): DeblurGAN [1], SRN [3] and Deepblur [2], our models (in red) is shown to achieve superior or comparable quality and is much efficient.

Though the superior performance of FPN [6] can handle both running time speed and multi-scale intermediate features, there is still lager room for improving it. Firstly, we find that the high-level semantic information in the FPN structure is gradually diluted in the process of gradual transmission to the shallower level. A similar drawback exists in [7]. This phenomenon will hinder the effect of repairing images. Therefore, we designed a semantic information supplement (SIS) mechanism in the generator [5] to solve the semantic dilution problem.

Secondly, the SIS mechanism can helps the network to repair the motion blur of more specifically. Since semantic information can distinguish the foreground object from the background, only the information contained in the foreground will be involved in the process of image repair. This is a bit similar to the non-uniform deblurring method based on segmentation [8]–[10]. The performance of these algorithms depends on whether or not the image can be segmented into different blur regions accurately. Furthermore, we use semantic information to replace segmentation, which can mostly avoid poor performance problems due to inaccurate segmentation. In addition to the highly concentrated semantic information, the receptive field of each feature pixel is special large, which can guide the network to use context information to restore blurred details in a larger range. The SIS mechanism contains two steps. First, the semantic information that needs to be supplemented is stored in the info-box in advance. Second, the feature fusion expansion mechanism gradually supplements the prepared semantic information to the merged feature map in the expanding path.

So, we have made three innovative improvements in this paper:

- To reduce the network parameters, we designed an intermediate layer path imitating the model of FPN [6] in our GAN-based [5] structure. This design makes the generator to actively preserve the intermediate layer features of different scales by extracting semantic information only once, rather than the multi-scale method.
- For different situations, we adopted two different upsampling methods according to specifical conditions to better restore the image details.
- This paper proposes the semantic information supplement mechanism (SIS) to solve the problem that high-level semantic information is diluted in its dissemination.

II. RELATED WORK

This section will briefly review some algorithms which better excellent deblurring effects, and these methods will appear in the section IV as comparative experiments. In recent years, dynamic scene deblurring became a hot topic for scenes that are not static, and the blur is caused by camera shake and complex object motion. Several deblurring methods [1], [9], [12]–[15], [17], [19], [20], [22], [23], [25] have been proposed to deal with dynamic blur. Sun *et al.* [11] employed a classification CNNs to predict blur direction and strength of a local patch. A dense motion field is obtained via Markov Random Fields (MRF) from the sparse blur kernel. The final latent image is generated by the non-blind deblurring method [12]. Gong *et al.* [13] utilized a fully convolutional network to estimate the dense heterogeneous motion flow from the blurred image and still used the method of [12] to recover the latent image. Nah *et al.* [2] is the pioneering application of multi-scale processing technology in image deblurring. A multi-scale loss function is employed to mimic the coarse-to-fine pipeline in conventional deblurring approaches. Tao *et al.* [3] improved the pipeline to model the scale-recurrent structure with shared parameters. In [3], the ConvLSTM module was used in the multi-scale network, so that the network parameters of different scales can be shared, and greatly reducing the number of parameters. Moreover, the ConvLSTM module can be used to aggregate feature maps from coarse-to-fine scales. Liu *et al.* [14] proposed a two-stage deblurring module to recover the blur images of dynamic scenes based on high-frequency residual image learning. EH-GAN [15] propose an edge heuristic multi-scale generative adversarial network (GANs) [5], which uses the coarse-to-fine scheme to restore clear images in an end-to-end manner. SR-Deblur GAN[16] solved the deblurring problem of local blurred natural images by the proposed self-reference deblurring GANs [5]. Kupyn *et al.* [1] also designed a GANs-based [5] deblurring algorithm, which is used to detect the object on the blurred image. Two better classification algorithms are proposed in [26], [27]. For the discriminator of the GANs-based network, the two methods are worthy of reference. Kupyn *et al.* [17] used the feature pyramid network[6] structure as the core building block to achieve a good deblurring effect. Recently, a blurred video frame interpolation method proposed by [18], which uses a low frame rate blur to synthesize the clear results of high frame rate from video enhancement. Yuan *et al.* [19] constructed a spatially variant deconvolution network using modulated deformable convolutions, which can adjust receptive fields adaptively according to the blur features. Chi *et al.* [20]

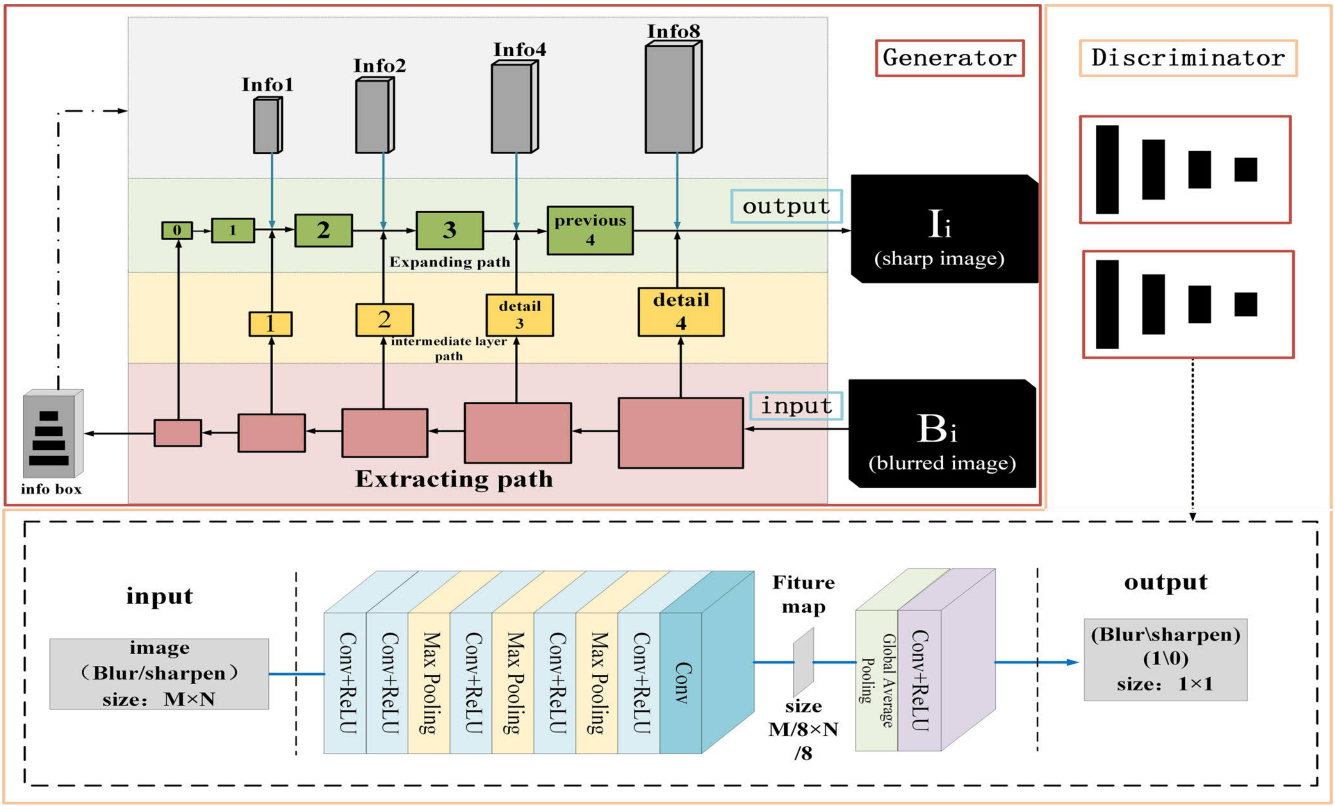


FIGURE 2. SIS-net network architecture. 1. Generator composed by extracting path (pink area), intermediate layer path (yellow area) and SIS mechanism. SIS mechanism consists of expanding path (green area) and semantic storage box (gray area). SIS mechanism consists expanding path (green area) and semantic storage box(gray area). 2. Two completely same discriminators are included in SIS-net. The detail of one of the discriminators' architecture is displayed in the bottom of the Figure 2. Ii represents a sharp image generated by SIS-net, and Bi represents the blurred image.

proposed a novel multi-scale deep convolutional neural network to solve demosaicking and deblurring jointly. Li *et al.* [21] proposed a brand-new convolution architecture named 'hole convolution', of which the kernel takes a rectangular ring of the neighbors to the center pixel into computation, and the reception field is greatly expanded. Moreover, they also presented a scale-aware convolutional neural network to recover the latent sharp image. Ye *et al.* [22] exploited a weights sharing method to restore sharp images in an end-to-end fashion, and use the super-resolution structure to replace the traditional upsampling layer.

III. PROPOSED METHOD

The GANs-based [5] network we proposed contains a generator, imitates the model of FPN [6], and two discriminators. In the generator part, the SIS mechanism is introduced as the main innovation.

A. GENERATOR

The generator of SIS-net is divided into three stages in the process of generating clear images. The first stage: the semantic information feature of the blurred image extracting process, its visual display corresponds to the pink background area of Figure 2; The second stage: in the intermediate layer path, the network actively collects the multi-scale intermediate layer features, whose visual display corresponds to the yellow background area of Figure 2; The third stage: Semantic Information Supplement (SIS) mechanism, which is used as the primary innovation point of this paper, solves the phenomenon of semantic dilution in FPN [6]. The visually displaying corresponds to the green and gray background area of Figure 2.

1) EXTRACTING PATH

As the first stage of image generation, the extracting path is mainly responsible for the semantic information feature extraction task of the blurred image. The blurred image passes through five inception blocks [23] of different receiving scales in turn, which serves as the basic network unit of the extracting path. The output of the previous inception block is used as the input of the next inception block. Since each block comprises several convolutional layers and pooling layers stacked in a specific order, its scale accepted by each inception block decreases by 1/2 in turn. For the specific inception block formation method, please refer to the literature [23]. Moreover, the feature map output by the last inception block is called basic semantic information(info$_0$). This highly concentrated semantic information makes each pixel have a larger receptive field. In the next feature fusion process, the semantic information can be combined with the higher-resolution detail features, that is, the multi-scale intermediate layer features. Semantic information is used to guide the network to use more comprehensive detail information to restore the image. The extracting path corresponds to the pink area in Figure 2, and each pink square corresponds to a network unit: inception block [23].

2) INTERMEDIATE LAYER PATH

In the second stage of image generation, the intermediate layer path actively collects intermediate products produced in the feature extraction process with a different scale. This design imitates the FPN [6] structure and also can handle both lower-resolution semantic information and higher-resolution detail features. Both the semantic information of the blurred image can be obtained from the lower-resolution deep feature maps to increase the receptive field of the network, and more detailed clues can be obtained from the higher-resolution shallow feature maps. The intermediate layer path extracting feature only one-time can avoid the problem of slow operation speed problem caused by multiple feature extraction in traditional multi-scale methods. For this issue, it can act as a lighter weight alternative to incorporate multi-scale features. We will explain this process in detail: when blur image passing through the extracting path, the inception block [23] performs pixel-level filtering on the input image or feature map. This process causes more important representative feature pixels are retained, with a certain number of feature pixels are discarded. Simultaneously, the details of the image are reduced. The intermediate layer path actively saves the feature map obtained by each inception block [23] in the feature extraction process as the intermediate layer features. The intermediate layer features can also be regarded as the detailed features of multi-scale. The larger the scale, the richer the detailed features can be provided, and we can recover a clear latent image by using these detailed clues reasonably. Since the extracting path contains five inception block [23], the corresponding intermediate layer path also covers a total of 5 different scale detailed feature maps, and they orderly decrease to half of the former. The smallest scale feature map is called info_0, and the remaining four features are called detailed features. These multi-scale detailed features will participate in the subsequent feature fusion process as the detailed information. In Figure 2, the yellow area represents the intermediate layer path; these yellow rectangles represent the actively collected intermediate layer features of different scales.

3) SEMANTIC INFORMATION SUPPLEMENT MECHANISM (SIS)

Semantic information as highly concentrated feature information has a super-wide receptive field of each feature pixel that can more effectively guide the network to use contextual relationships and restore the blurred images with a larger range of adjacent pixels. According to the semantic information, the network can distinguish the foreground and background well and avoid the background interference under the guidance of semantic information; it is quite suitable for removing the motion blur caused by the foreground target in the motion scene. What is more, to solve the problem that the semantic information in FPN [6] gradually dilutes with the increase of the number of fusions, we innovatively designed a semantic information supplement(SIS) mechanism.

In the third stage of image generation, SIS is mainly composed of the semantic information feature storage box(info-box) and expanding path. Info-box stores the semantic information features for the feature fusion to be performed next, corresponding to the gray rectangles in the gray area of Figure 2; The feature fusion and Fractional Strided Convolution upsampling occurs in the expanding path. In this process, three different forms of features are merged multiple times, and the clear image is gradually restored by Fractional Strided Convolution upsampling, corresponding to the green area in Figure 2.

a: BUILDING THE SEMANTIC INFORMATION STORAGE Box(info-BOX)

The semantic feature storage box is one of the critical components of the semantic information supplement mechanism, as shown in the gray area of Figure 2. There are four different scales of semantic information features. Its motivation is to use itself as a storage container, and to save the semantic information at the adopted scale of four in advance, so as to make sufficient preparations for the subsequent feature fusion operation. The basic semantic information Info_0 is derived from the feature map extracted by the last inception block in the extracting path. The semantic information of 4 different scales contained in the info-box: Info_1, Info_2, Info_3, and Info_4 are all obtained by using corresponding multiples on Info_0. It is worth noting that Info_0 is not stored in the info-box, because Info_0 will be directly transferred to the top of the symmetric expanding path. When selecting the upsampling method of semantic information storage box, considering the available feature pixels are scarce. For example, the scale of Info1 is only 22*40 if the network input image resolution is 720 * 1280. So in order to ensure the authenticity of the semantic features after up-sampling, unlike other mainstream up-sampling algorithms [1]–[3], we use pixel-shuffle convolution [24], [25].

Pixel-shuffle convolution [24] mainly includes two substeps: we assume that the size of the up-sampling sample needs to be r times. First, input the low-resolution feature info_0(w*h*c) into a convolutional layer. The number of channels in this convolutional layer is r^2*c, and the size of the convolution kernel is 1 * 1. Such a convolution design ensures that the scale of the feature map is not changed, and the number of channels is expanded by a factor of r^2 to obtain the info_0'(w*h*r^2c). The extended dimension of the number of channels is shown in Equation(1), where C_{nn} represents this convolution operation. The visualization flowchart corresponds to the channel generation part on the left side of the Figure 3.

$$Info_1'(w*h*r^2c) = C_{nn}(Info_1(w*h*c)). \tag{1}$$

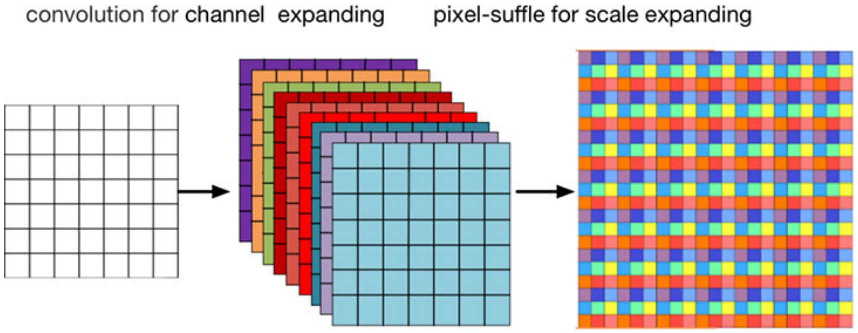

convolution for channel expanding
pixel-suffle for scale expanding

FIGURE 3. The upsampling pattern of pixel-shuffle convolution: the left part shows the process of the extending dimension of the channels; the right part represents the process of compressing channels to achieve high-resolution features.

The second step of pixel-shuffle convolution is to extrude the r^2c channels into c channels through the pixel-shuffle operation. As shown in Equation (2), the high-resolution feature is scaled (rw, rh, c), which completes the upsampling process. This corresponds to the compressed part on the right side of Figure 3.

$$Info_r(rw, rh, c) = pixel - shuffle(Info'_0(w * h * r^2c)). \quad (2)$$

Since increasing the number of feature channels is obtained by specific convolutional layers, these convolutions can be trained with other network parameters, so more realistic up-sampling results can be produced. It is not difficult to find that in the next feature fusion, the semantic information and the corresponding intermediate layer features are the same in scale, which makes feature fusion more convenient. By performing a pixel-shuffle convolution operation on the basic semantic information info_0, it sequentially upsamples two times, four times, eight times, and sixteen times to get Info_1, Info_2, Info_3 and Info_4 correspondingly, thus forming a semantic storage box.

b: FEATURE FUSION EXPANSION MECHANISM IN EXPANDING PATH

To make up for the shortcomings of the original FPN semantic information dilution, an actively incorporate detailed feature in each process of feature fusion process, we will also deliberately supplement the semantic information feature. In Algorithm 1, we visually show the that for the gradual recovery of a clear image through feature fusion expanding multiple times.

First, we introduce the preparation of the feature information required for feature fusion. These futures were divided into three types: semantic feature information, detail feature information and front layer feature information. "Info_k" represents the semantic feature information. The basic semantic information Info_0 is produced by extracting path, and the rest of Infok comes from the info-box. "Details" represents the intermediate feature map which are produced by the corresponding inception block [23] of intermediate layer path. We also treat them as detailed feature information: detail_1, detail_2, detail_3, detail_4. This detailed feature information provides detailed texture cues for the restoration of clear latent images. What is more, the previous fusion result is used as the front layer feature information of the next fusion, denoted as the previous k.This part corresponds to the top of the Algorithm 1.

Algorithm 1 Feature Fusion Mechanism Algorithm

```
Input: blurring image
Output: sharp
Features preparation:
    $\text{Info}_0$ = extracting path(blurring image)
    for i in range(1,4):
        $\text{detail}_i$ = intermediate feature $\text{map}_i$ = inception $\text{block}_i$ ($\text{detail}_{i-1}$)($\text{detail}_0$ = blurring image)
    for j in range(1,4):
        info j = info - box(j)
Feature fusion expansion:
    step1:
        $\text{merge}_0 = \text{Info}_0$
        $\text{previous}_1 = \text{FSconv}(\text{merged}_0)$
    step2:
        for k in range(1,3):
            $\text{Merged}_k = \text{previous}_k + \text{Info}_k + \text{detail}_k$
            $\text{previous}_{k+1} = \text{FSconv}(\text{merged}_k)$
    step3:
        for i in range(1,4):
            $\text{previous\_smooth} += \text{previous}_i$
        $\text{sharp} = \text{FSconv}(\text{info}_4 + \text{detail}_4 + \text{previous\_smooth})$
```

Second, we introduce the feature fusion expansion process, which totally has three different steps of feature fusion. Generally speaking, fusion requires splicing the three features prepared and then perform the upsampling process of Fractional-Strided Convolution (FSconv) [28] on the obtained splicing features. The specific three steps are as follows: In the first step, it is a bit special here. Since the semantic features of info_0 are relatively pure, there is no need to splice with other features, and info_0 is fractionally-strided convolutions directly, then the result is used as previous_1.

In the second step, identical feature fusion processes are completed three times. Merged$_k$ is spliced by the feature of previous$_k$, detail$_k$, and Info$_k$ with same corresponding scale and then the Fractional Strided Convolution upsampling is performed to generate the fusion result and it also will be treated as previous layer information previous$_{k+1}$ feature. In the third step, the first three fusion products act as the front layer information and splice with detail$_4$ and info$_4$. After upsampling the spliced features, a clear latent image is obtained. This part corresponds to the bottom of the Algorithm 1.

For the completeness of the paper, we briefly introduce the algorithmic principle of Fractional Strided Convolution upsampling [28](FSconv). When FSconv receives a feature map that needs upsampling, this feature map will be interpolated to expand its scale. The interpolation process involves inserting a certain number of digits 0 between each original feature pixel to achieve the purpose of upsampling. The number of zeros inserted between each original feature pixel depends on the size of the stride in formula A. The height or width of the feature map after FSconv upsampling is calculated as in Equation(3).

$$Height' = Height + (Stride - 1) * (Height - 1). \quad (3)$$

The height or width of the feature map after FSconv upsampling is calculated as in Equation(3). Where "Height′" is the height of the feature map after FSconv upsampling, "Height" is the height of the original input, and "Stride" is the number of 0s inserted between pixels. The value of "Stride" also directly determines the extension multiple of the feature after upsampling. The calculation method for weight is the same.

B. THE SPECIFIC STRUCTURE OF THE DISCRIMINATOR

Our network is equipped with two discriminators. It is worth noting that these two discriminators are completely consistent in network structure and parameters. However, the size of the pictures it accepts varies. One of the discriminators receives the entire generated image and controls the clarity of the generated image globally. Another discriminator receives the local area of the generated image to correct the clarity of the generated image in detail. At the bottom of Figure 2, we visually show the network structure of the discriminator, and the parameters of the network are elaborated in Table 1. Therefore, the design selected a lightweight network layout to improve the speed and performance of the discriminator. The discriminator looks at the generated image from different angles due to the setting method of the dual discriminators. Thus, it does not affect the discriminant performance as the network is not deep enough.

C. LOSS FUNCTION

Distinguish the differences between the repaired image and ground-truth, thereby guiding and correcting the network to produce a sharper image. An effective method is to calculate the difference in pixel level between the two images. This pixel-level operation can repair the details of the blurred image such as texture, color, and blurred information. Its effectiveness has been proved by many previous experiments [2], [11]. In our algorithm design, we choose MSE act as a pixel-level loss. It is also used as the most important part of the total-loss. For details of MSEloss, see Equation(4). The total-loss can be seen in Equation(6).

TABLE 1. Specific parameters of adversarial network layer.

Layers	Filter size	Stride	Padding
Conv 1+relu	3×3×1×64	1	1
Conv 2 + relu	3×3×64×64	1	1
Max pooling 3	2×2	2	0
Conv 4 + relu	2×2	2	0
Max pooling 5	2×2	2	0
Conv 6 + relu	3×3×64×64	1	1
Max pooling 7	2×2	2	0
Conv 8 + relu	3×3×64×64	1	1
Conv 9	3×3×64×64	1	1
Globle pooling 10	(M/8)×(N/8)	1	0
Sigmod 11	-	-	-

$$L_{MSE} = \frac{1}{wh}\sum_{x=1}^{w}\sum_{y=1}^{h}((I_i)_{x,y} - G_\theta(B_i)_{x,y})^2. \quad (4)$$

where B$_i$ represents the input blurred image, G$_\theta$ represents the generator, I$_i$ is the standard clear image, w and h are the length and width of the input/output image, respectively.

However, using the MSEloss alone will produce artifacts. For instance, considering two identical images offset by one pixel from each other; although they are similar in perception, the results will be very different. When processing the boundary contour of a moving object, the difference at the boundary will be very large because the movement has shifted. In this case, the network will use the average value of all possible solutions as the convergence value, resulting in artifacts, which is the reason of the contour edge still have a large degree of blur. So we incorporate the Perceptual loss [29] into the total-loss function to loosen the restrictions of using MSEloss alone. This Perceptual loss [29] calculates the least square distance between ground-truth and image generated by SIS-net based on the feature map, which is extracted by the VGG16 network [29]. The receptive field of each pixel in the feature map corresponds to the 8*8 adjacent fields of the original map. Therefore, in calculating of perception loss in the dynamic scene, the displacement of 8 pixels in the horizontal/vertical direction can be included. The process is shown in Equation5, where w and h are the length and width of the feature map,G$_\theta$ represents the generator, D$_\theta$ represents the discriminator, and the parameter of φ are obtained by the VGG-16 network in the ReLU 3_3 layer.

$$L_{percep} = \frac{1}{wh}\sum_{x=1}^{w}\sum_{y=1}^{h}(\varphi(I_i)_{x,y} - \varphi(G_\theta(B_i))_{x,y})^2. \quad (5)$$

The SIS-net is based on Generative Adversarial Networks [5]. This framework will input the blurred image and the real image into the discriminator so that the discriminator

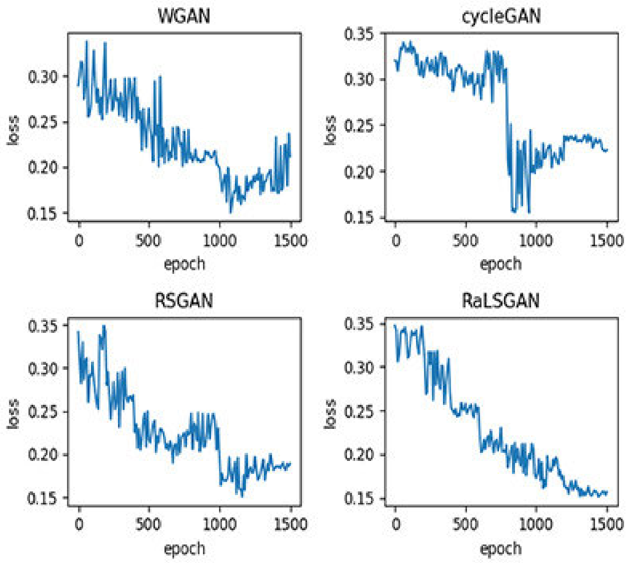


FIGURE 4. Network loss graph during training.

can distinguish blur and sharp image well. Therefore, the generator needs to generate images that are closer to clear to sharp to deceive the discriminator. Through the screening of experiments, it was found that the use of WGAN [30],cycleGAN [31],RSGAN [32] would cause problems such as model collapse, gradient explosion/disappearance during training, which makes it difficult to optimize the objective function. As shown in Figure 4, in the screening process, $\tau_1 = 0.5$, $\tau_2 = 0.006$, $\tau_3 = 0.01$. In comparison, using RaLSGAN [17] as a discriminate loss $L_D^{RaLSGAN}$ is more stable during the training process. As shown in Equation(6), where G_θ represents the generator and D_θ represents the discriminator.

$$\begin{aligned} L_D^{RaLSGAN} &= E_{x\sim Pdata(x)}[(D_\theta(x) - E_(x \sim P_z(z))D_\theta(G_\theta(z)) - 1)^2] \\ &\quad + E_{z\sim P_z(z)}[(D_\theta(G_\theta(z)) - E_{x\sim Pdata(x)}D_\theta(x) + 1)^2]. \end{aligned} \tag{6}$$

Our overall loss function is shown in formula(7):

$$L_{total} = \tau_1 L_{MSE} + \tau_2 L_{percep} + \tau_3 L_D^{RaLSGAN}. \tag{7}$$

For the selection of hyperparameters in Equation(7), we also conducted quantitative experiments showed in sectionIV of the paper. We subjectively set up several sets of candidate parameters for comparison based on the priority of L_{MSE},L_{percep},$L_D^{RaLSGAN}$. Through the continuous transformation of single variables, we find the optimal values of the three variables τ_1,τ_2,τ_3.

IV. EXPERIMENTAL DESIGN

A. INTRODUCTION TO DATASET

In this chapter, to prove the effectiveness of network and avoid the phenomenon, the network only has excellent performance on specific datasets due to overfitting. We will conduct comparative experiments on two different public datasets. In a real dynamic scene, it is challenging to obtain a completely consistent blur/sharp image pair at the same time. Because after obtaining the blurred image, it is impossible to return to the moment to get a sharp image corresponding to it. Even if we deliberately create certain conditions, slight deviations are inevitable. Therefore, in the process of data production, the Kohler dataset [33] uses a synthetic method. He artificially convolves the sharp image with the blur kernel. These blur images are not taken from the real world. It is determined by the blur kernel and cannot fully reflect the blur of dynamic scenes in the real world. The GoPro dataset [2] is obtained by superimposing multiple clear images through a particular process, and there is no fixed limit of the blur kernel compared with the synthetic method. So GoPro is relatively more able to reflect the vague state of the real world. Next, we give a detailed introduction to the production process of the two datasets.

1) THE KOHLER BENCHMARK DATASET

The Kohler benchmark dataset [33] was generated by recording and analyzing real camera motion, played on an automatic platform, recorded a series of sharp images, and sampled the motion trajectory of the 6D camera. Then it uses the artificially designed blur kernel to perform convolution on the sharp image to obtain the blur/sharp image pair. This chapter will compare and evaluate our dynamic scene down-blurring algorithm with other excellent algorithms on the Kohler dataset. Kohler contains 240 image groups, each of which consists of 4 blurred images, where each blurred image contains a blurred image generated by convolution of 12 different blur kernels and a corresponding sharp image.

2) GOPRP DATASET

The GoPro [2] dataset uses a GoPro Hero 4 camera to capture a video sequence at 240 frames per second (fps) and then averages a continuous variable number of latent (7-13) frames to produce blurs of varying intensities. For example, an average of 15 frames simulates an image taken at a shutter speed of 1/16, while the corresponding sharp image shutter speed is 1/240. It is worth noting that GoPro [2] uses multiple frames of sharp images to synthesize the blurred image. The sharp image corresponding to the blurred image is the middle frame in consecutive frames. Finally, the dataset consists of 3214 pairs of blurred and sharp images with a resolution of 1280 × 720. Among them, 1111 pairs are used as test sets instead of assuming a specific motion and designing a complex blur kernel. GoPro [2] follows the approximate camera imaging process during the blurry image generation process to integrate consecutive frames within an exposure time. Therefore, there are only pairs of clear/sharp image pairs in the dataset, but no blur kernels. Compared with the traditional synthetic blur dataset with uniform blur kernels, this blur dataset without kernel estimation has a better prospect in motion, and the static background exhibits more realistic spatial blur changes.

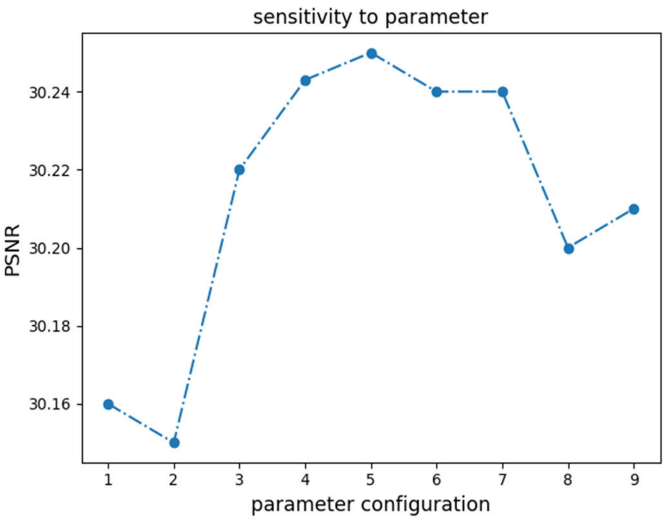


FIGURE 5. Fixed τ_2 fine-tuning the scope of the post-selection of τ_1.

B. EVALUATION STANDARD

Peak signal-to-noise ratio (PSNR) the most basic and objective image evaluation index, is widely used. Generally speaking, the larger the value means the less distortion. Its principle is to judge according to the error between pixels.

Structural similarity (SSIM) measures the similarity between two images in terms of brightness, structure, and contrast. A larger SSIM value corresponds to a higher image fidelity. Since the human eye easily obtains the structural information in an image, calculating the structural similarity between images is an excellent index for evaluating image quality.

C. HYPERPARAMETER SENSITIVE EXPERIMENT

From a subjective perspective, we first estimate the order of magnitude of each parameter. For τ_1: comparing the difference pixel by pixel on the two images is a very effective means of evaluating the loss of the latent image and the ground truth, which means L_{MSE} can be very accurate and meticulously correct the pixels that have been restored incorrectly. So τ_1 will occupy a large proportion. For τ_2: Our training is based on the GoPro dataset [2], considering that due to dynamic scenes, the blurred image will have a certain displacement during the camera exposure process. Therefore, forcibly using Lmse will cause certain artifacts, and then we use Lprecep for some relaxation. Based on the experience, we will choose τ_2 to be two orders of magnitude smaller than τ_1 because of the small proportion of moving objects in the global image; and too much relaxation will cause the network to be unable to perform delicate repairs.

For τ_3:$L_D^{RaLSGAN}$ is a loss for the binary judgment of the discriminator. The discriminator has certain filtering and culling ability for the more obvious blurred images after a better design and a lot of training, which could help the network to converge quickly in the early stage of training and jump out of the local optimal solution faster. However, for the relatively sharp or have only partial defect image, this binary classification method is difficult to remove them from the test dataset. Thus, we subjectively set τ_3 to an order of magnitude smaller than τ_1.

According to the idea of a single variable, we prefer to use τ_1 as the entry point since Lmse accounts for a large proportion of the total loss. Plenty of papers [1], [11], [17] prove that choosing $\tau_3 = 0.01$ based on experience can basically be applied to most situations. To simplify the problem, we also control $\tau_3 = 0.01$ consistent. We quantitatively use [0.1-0.9] as the candidate set of hyperparameter τ_1; Let [0.001-0.009] be the candidate set of τ_2. In the first experiment, we select the mediant of the candidate set 0.005 assigning to τ_2, and finally fine-tune τ_1. According to the results shown in Figure 5, it is obviously to see that τ_1 is excellent in the PSNR peak signal-to-noise ratio in the interval of 0.4, 0.5, and 0.6. After the above operations, we have reduced the range of τ_1 candidate set: $\tau_1 = 0.4, 0.5, 0.6$. Then we perform the second round of using τ_2 as a single variable to determine the value of τ_2 and at the same time find the maximum value of τ_1 optimal value at the same time. As shown in Figure 6, the blue line depicts a line chart obtained by fine-tuning τ_2 with $\tau_1 = 0.4$ and $\tau_3 = 0.01$; The red line depicts a line chart obtained by fine-tuning τ_2 with $\tau_1 = 0.5$ and $\tau_3 = 0.01$; The green line depicts a line graph obtained by fine-tuning with $\tau_1 = 0.6$ and $\tau_3 = 0.01$. The flatness of the line graph can represent the stability of the parameters. We found that the red line is the most stable and the leading result is also the best. Therefore, we determine the hyperparameters $\tau_1 = 0.5$, $\tau_2 = 0.006$, and $\tau_3 = 0.01$.

D. INNOVATION EFFECTIVENESS

In this paper, we added three innovative designs for traditional generators: first, add intermediate layer path to take

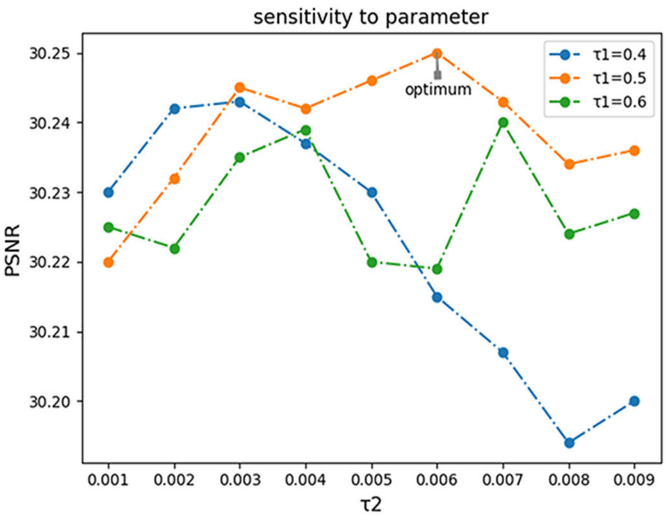


FIGURE 6. Enumerate τ_1 fine-tuning τ_2 to finalize the optimal solution.

TABLE 2. Comparison on the GoPro test dataset for Innovation effectiveness.

Model	Nolatent-net	NoPloss-net	Nosis-net	SIS-net(complete)
PSNR	28.21	29.1	28.44	30.28
SSIM	0.823	0.851	0.842	0.912

advantage of multi-scale intermediate layer features; second, the Semantic information supplementary mechanism supplements the semantic information multiple times in the expanding path; third, adding Perceptual loss [29] to relax the overall loss to reduce artifacts introduced in the process of recovering motion blur. Although these three innovative designs are logical in principle, we conducted detailed ablation experiments. According to the single variable principle, we removed one of the modules in the original network modules to compare quantitatively and stereotypedly. Among them, SIS-net(complete) represents the complete network. Nolatent-net means to remove the intermediate layer path on the basis of SIS-net. Noploss-net means to remove Perceptual loss [29] in the loss function of SIS-net. Nosis-net means to remove the semantic complement mechanism based on SIS-net. The quantitative results are shown in Table 2. It can be found that regardless of Nolatent-net, Noploss-net, NoSIS-net, the effect of image restoration is reduced to a certain extent. Further, it can be found that the lack of intermediate layer features has the most significant impact on the experimental results, which also verifies that the multi-scale approach can greatly improve the performance of deblurring. The intermediate layer features of different scales can provide different levels of detail clues; this is crucial for deblurring. In the case of removing the Semantic information supplementary mechanism, the network also performs poorly, which reflects the importance of semantic information. If we do not consider using the Perceptual loss function to relax the real penalty function, it will also have a certain impact on image repair quality because the displacement objects in the moving scene do deviate considerably greatly on a pixel-by-pixel comparison level. In addition, but we also provide a qualitative visual display, as shown in Figure 7. It can be found that whether it is in the green box shoe part or the yellow box letter part, the complete algorithm far exceeds the Nolatent-net and NoSIS-net in the recovery effect, thereby illustrating the effectiveness of the intermediate layer features and semantic complement mechanism. For NoPloss-net, since the shoes in the background are relatively stationary without movement displacement, the lack of Perceptual loss does not have much impact on the recovery effect. However, for a moving wheel, it can be found that the image restored by the NoPloss-net network will obviously have a certain degree of defects in the spoke position of the wheel. The result of NoPloss-net indeed shows that Perceptual loss can improve the deblurring effect in the moving scene.

E. QUANTITATIVE EVALUATION ON GoPro DATASET

We compare our method with recent state-of-art dynamic scene deblurring and non-uniform deblurring methods on the GoPro dataset: one is a traditional method by Xu *et al.* [34], while the rest are deep learning-based.We compare on standard performance metrics (PSNR, SSIM). The results are summarized in Table 3.

In terms of PSNR, SIS-net ranked first, slightly higher than SRN [3], and SIUN [22] ranked third. Unfortunately, our

TABLE 3. Performance and efficiency comparison on the GoPro test dataset.

Methds	Deep De-blur [2]	SRN [3]	Deblur GAN [1]	Sun et al [11]	Gong et al [13]	EG-GAN [15]	Deblur GAN -v2 [17]	Yuan et al [19]	Chi et al [20]	SIUN [22]	Xu et al [34]	Li et al [35]	SIS -net (ours)
PSNR	29.08	30.26	26.435	24.64	26.06	29.32	29.55	29.57	29.41	30.22	25.1	27.08	30.28
SSIM	0.841	0.934	0.892	0.842	0.863	0.933	0.901	0.933	0.918	0.904	0.894	0.857	0.912
Time	4.33s	1.31s	0.92s	20min	20min	0.64s	0.35s	0.418s	0.912s	3.67s	13.41s	1.7min	0.303s

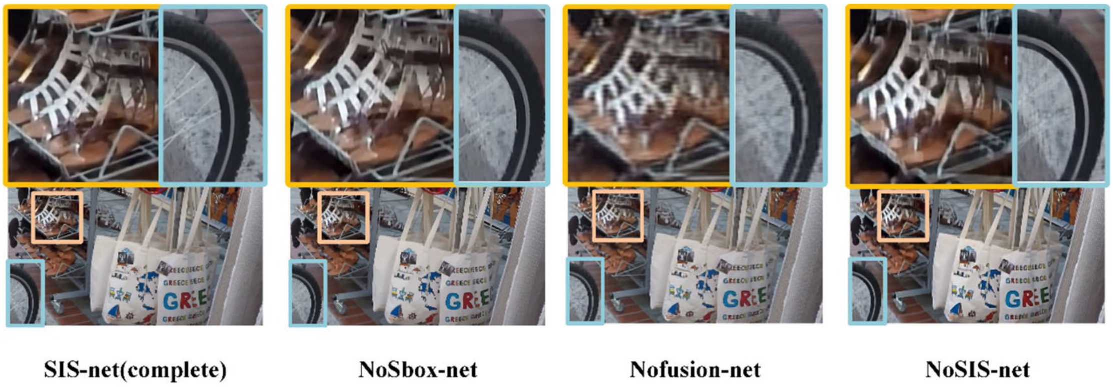

SIS-net(complete)
NoSbox-net
Nofusion-net
NoSIS-net

FIGURE 7. Visualization of the effectiveness of innovative components.

algorithm is at a disadvantage in the standard performance metrics of SSIM. Yuan *et al.* [19] and EG-GAN [15] tied for first place in SSIM. Nevertheless, our SIS-net is better both in PSNR and in time efficiency. This primarily shows that SIS-net(ours) still needs to improve the color fidelity in the process of repairing images. In the next work, I will make new improvements to the algorithm from this perspective. It has to be said that SRN [3] has achieved excellent results in image repair quality. However, we are very encouraged to observe that SIS-net (ours) takes 80% less inference time than SRN. This is a good illustration of the effectiveness of the design of the intermediate layer channel to take advantage of multi-scale features. Instead of extracting blur input image features of different scales multiple times like SRN [3] and Nah *et al.* [2], the speed of the network is greatly improved by reducing the weights. For pictures with a resolution of 720P, our average run-time reaches 0.303 seconds on our platform, which means it supports real-time image deblurring.

For other comparative experiments, the traditional methods [13], [34], [35] based on the prior assumption do not get good results,and they are not suitable for the task of real-time deblurring. Because only the convolutional neural network is used to estimate the blur kernel in Sun *et al.* [11], the error is massive in the process of inverse solution, So the high PSNR and SSIM coefficients are not obtained. Nah *et al.* [2] use a multi-scale scheme to remove motion blur. However, its [2] results in heavy network parameters and long running time due to the lack of sharing mechanism. Kupyn *et al.* [1] builds a generative adversarial model to remove motion blur. Because its generator does not design the intermediate layer channel, it cannot use multi-scale detailed clue information, and it also tends to be inferior in repair effect. DeblurGAN-v2 [17] also introduced the structure of the FPN [6], but they ignored the problem of semantic disappearance in the process of feature transmission. Therefore, it is inferior to our algorithm in each performance aspect. Yuan *et al.* [19] design an effective and real-time deblurring network by using a deconvolution operation. Despite [19] owned a high efficiency, it performs relatively poorly on the PSNR. Chi *et al.* [20] defines the deblurring as a two non-invertible tasks. But it did not get a remarkable result in the image repair effec. Please refer to table 3 for details.

The algorithms participating in the comparison of Kohler's benchmark dataset [33] remain unchanged, and the comparison results are shown in Table 4. It can be found that the algorithm in this paper still achieves good results, ranking first in the PSNR coefficient and second only to SRN-net in the SSIM coefficient. This shows that our network does not only perform well on specific datasets, but has a strong generalization.

F. SUBJECTIVE EVALUATION BASED ON PERCEPTION

In order to evaluate several models from the perspective of perception, canny edge detection is performed on the sharp image restored by each model:SRN [3], DeblurGAN [5], DeblurGAN-v2 [17], and SIS-net (ours). Figure 8 displays visual examples on the GoPro dataset. DeblurGAN [5], DeblurGAN-v2 [17], and SIS-net (ours) effectively restore the edges and textures, without noticeable artifacts. However, SRN for this specific example shows some color artifacts when zoomed in, and SRN also lost more edge pixels. DeblurGAN [25] and DeblurGAN-v2 [17] have better edge pixels

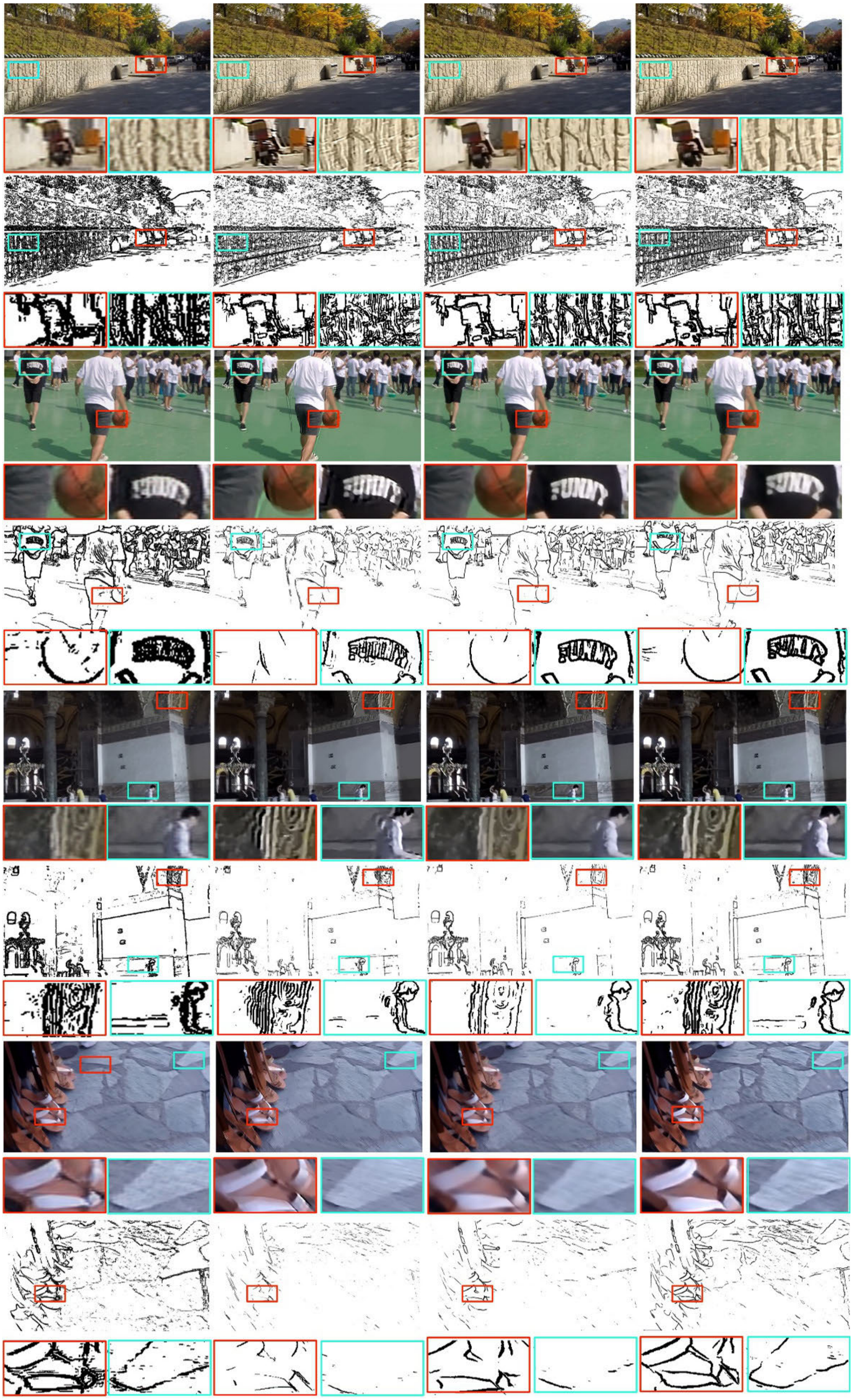


FIGURE 8. The above figure uses DeblurGAN [1], SRN [3], DeblurGAN-V2 [17] and SIS-net(Ours) methods to compare on the GOPRO dataset.

TABLE 4. Quantitative comparison experiment of Kohler dataset.

Methds	Deep Deblur [2]	SRN [3]	Deblur GAN [1]	Sun et al [11]	Gong et al [13]	Deblur GAN-v2 [17]	Chi et al [20]	Xu et al [34]	Li et al [35]	SIS-net (ours)
PSNR	26.48	26.75	26.10	25.245	25.96	26.36	24.90	24.61	25.88	26.79
SSIM	0.807	0.837	0.816	0.784	0.798	0.820	0.863	0.731	0.801	0.820

preserving ability than SRN, but it gets wider edge pixels. The results of SRN [3], DeblurGAN [5], DeblurGAN-v2 [17] indicate that the edge part of the object may have a higher degree of blur. From the perspective of edge detection results, SIS-net (ours) outperforms SRN, DeblurGAN [17], and DeblurGAN-v2 [5] in terms of edge pixel preserving ability and edge width refinement ability. Please refer to Figure 8 for details.

V. CONCLUSION

The algorithm in this paper can realize real-time dynamic scene deblurring. Due to the light-weight calculation of the algorithm, it can be transplanted to the smart terminal to realize the anti-shake function of the smart terminal. At the same time, the algorithm can be used as a preprocessing link in other visual tasks to improve the performance of visual tasks such as target detection, segmentation, classification, and behavior recognition. In terms of accuracy, the semantic supplement network solves the problem of semantic dilution in the feature pyramid network, and it is better than many algorithms proposed recently, and it can solve the blurred problem in dynamic scenarios well. However, the algorithm in this paper also has some defects. In the process of restoring the image, the color fidelity is not right, and the actual color of the image cannot be restored very well. In the other word, it means that our algorithm is at a disadvantage in the standard performance metrics of SSIM. It will be improved in the following work. What is more, this algorithm is only limited to deblurring a single image. With the advancement of image acquisition methods, the demand for various visual tasks will gradually increase at the video level. Video-based deblurring technology is also imminent. The video-based deblurring algorithm should not only consider the spatial information of a single frame image, but also take into account the time dimension. In principle, this spatiotemporal fusion network will have more detailed clues and deblurring clues than single-frame images. Furthermore, with the development of mobile phone camera hardware, but current smartphones also have already popularized two or more cameras. Whether or not binocular vision technology can be integrated into the deblurring process is also an irreversible trend for image deblurring in the future.

REFERENCES

[1] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.

[2] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3883–3891.

[3] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8174–8182.

[4] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5978–5986.

[5] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[6] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[7] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, "A simple pooling-based design for real-time salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3917–3926.

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

[9] M. Suin, K. Purohit, and A. N. Rajagopalan, "Spatially-attentive patch-hierarchical network for adaptive motion deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3606–3615.

[10] J. Pan, Z. Hu, Z. Su, H.-Y. Lee, and M.-H. Yang, "Soft-segmentation guided object motion deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 459–468.

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

[12] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 479–486.

[13] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. Van Den Hengel, and Q. Shi, "From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2319–2328.

[14] K.-H. Liu, C.-H. Yeh, J.-W. Chung, and C.-Y. Chang, "A motion deblur method based on multi-scale high frequency residual image learning," *IEEE Access*, vol. 8, pp. 66025–66036, 2020.

[15] S. Zheng, Z. Zhu, J. Cheng, Y. Guo, and Y. Zhao, "Edge heuristic GAN for non-uniform blind deblurring," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1546–1550, Oct. 2019.

[16] G. Gong and K. Zhang, "Local blurred natural image restoration based on self-reference deblurring generative adversarial networks," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2019, pp. 231–235.

[17] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8878–8887.

[18] W. Shen, W. Bao, G. Zhai, L. Chen, X. Min, and Z. Gao, "Blurry video frame interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5114–5123.

[19] Y. Yuan, W. Su, and D. Ma, "Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3555–3564.

[20] Z. Chi, X. Shu, and X. Wu, "Joint demosaicking and blind deblurring using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2169–2173.

[21] J. Li, K. Li, and B. Yan, "Scale-aware deep network with hole convolution for blind motion deblurring," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2019, pp. 658–663.

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

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[24] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[25] C. K. Huang and H. H. Nien, "Multi chaotic systems based pixel shuffle for image encryption," *Opt. Commun.*, vol. 282, no. 11, pp. 2123–2127, Jun. 2009.

[26] Z. Wang, B. Du, and Y. Guo, "Domain adaptation with neural embedding matching," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 7, pp. 2387–2397, Jul. 2020.

[27] B. Du, Z. Wang, L. Zhang, L. Zhang, W. Liu, J. Shen, and D. Tao, "Exploring representativeness and informativeness for active learning," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 14–26, Jan. 2017.

[28] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[29] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[30] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875

[31] A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. Courville, "Augmented CycleGAN: Learning Many-to-Many mappings from unpaired data," 2018, *arXiv:1802.10151*. [Online]. Available: http://arxiv.org/abs/1802.10151

[32] R. Natsume, T. Yatagawa, and S. Morishima, "RSGAN: Face swapping and editing using face and hair representation in latent spaces," 2018, *arXiv:1804.03447*. [Online]. Available: http://arxiv.org/abs/1804.03447

[33] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling, "Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 27–40.

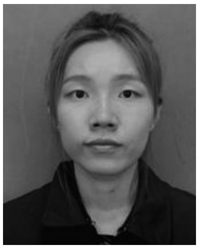
[34] L. Xu, S. Zheng, and J. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1107–1114.

[35] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, and M.-H. Yang, "Blind image deblurring via deep discriminative priors," *Int. J. Comput. Vis.*, vol. 127, no. 8, pp. 1025–1043, Aug. 2019.

YIMING LIU received the bachelor's degree in computer science from Liaoning University, China, in 2015, and the the M.S. degree from the College of Information Science and Engineering, Hunan Normal University, Changsha, China, in 2020. He is working as a Scientific Research Assistant with the College of Life Science, Hebei University. His research interests include bioinformatics and computer vision.

YIFEI LUO received the bachelor's degree in business administration from Wuhan University, China, in 2016, and the master's degree in public relations and advertising from the University of New South Wales, Sydney, Australia, in 2019. Her current research interests include intelligence media convergence and new media.

WENZHUO HUANG is currently pursuing the M.S. degree with the College of Information Science and Engineering, Hunan Normal University, Changsha, China. His research interest includes deep learning.



YING QIAO received the B.Sc. degree in computer science and mathematics (double major) from Rutgers University, New Brunswick, NJ, USA. Her research interests include image processing and computer vision.

JUNHUI LI is currently pursuing the M.S. degree with the College of Information Science and Engineering, Hunan Normal University, Changsha, China.

His research interests include deep learning, image processing, and computer vision.

DAHONG XU received the Ph.D. degree in electronic science from the National University of Defense Technology, in 2009. He is currently an Associate Professor with the College of Information Science and Engineering, Hunan Normal University, Changsha. His research interests include image processing, computer vision, and knowledge maps.

DUQIANG LUO received the Ph.D. degree from Northwest A&F University, in 2002. At present, he works as a Professor with the Department of Bioinformatics, College of Life Science and the Institute of Life Science and Green Development, Hebei University, Baoding, China, and serves as the Vice Director of the Key Laboratory of Pharmaceutical Chemistry and Molecular Diagnosis, Ministry of Education. During his research career, he has published over 70 scientific articles and holds patents. His research interests include AI accelerating drug development and chemical research.