Domain-adaptive Video Deblurring via Test-time Blurring

Jin-Ting He*,[1], Fu-Jen Tsai*,[2], Jia-Hao Wu[1], Yan-Tsung Peng[3], Chung-Chi Tsai[4], Chia-Wen Lin[2], and Yen-Yu Lin[1]

[1] National Yang Ming Chiao Tung University, Taiwan
jinting.cs12@nycu.edu.tw, jiahao.11@nycu.edu.tw, lin@cs.nycu.edu.tw
[2] National Tsing Hua University, Taiwan
fjtsai@gapp.nthu.edu.tw, cwlin@ee.nthu.edu.tw
[3] National Chengchi University, Taiwan
ytpeng@cs.nccu.edu.tw
[4] Qualcomm Technologies, Inc., San Diego
chuntsai@qti.qualcomm.com

Abstract. Dynamic scene video deblurring aims to remove undesirable blurry artifacts captured during the exposure process. Although previous video deblurring methods have achieved impressive results, they suffer from significant performance drops due to the domain gap between training and testing videos, especially for those captured in real-world scenarios. To address this issue, we propose a domain adaptation scheme based on a blurring model to achieve test-time fine-tuning for deblurring models in unseen domains. Since blurred and sharp pairs are unavailable for fine-tuning during inference, our scheme can generate domain-adaptive training pairs to calibrate a deblurring model for the target domain. First, a Relative Sharpness Detection Module is proposed to identify relatively sharp regions from the blurry input images and regard them as pseudo-sharp images. Next, we utilize a blurring model to produce blurred images based on the pseudo-sharp images extracted during testing. To synthesize blurred images in compliance with the target data distribution, we propose a Domain-adaptive Blur Condition Generation Module to create domain-specific blur conditions for the blurring model. Finally, the generated pseudo-sharp and blurred pairs are used to fine-tune a deblurring model for better performance. Extensive experimental results demonstrate that our approach can significantly improve state-of-the-art video deblurring methods, providing performance gains of up to 7.54dB on various real-world video deblurring datasets. The source code is available at https://github.com/Jin-Ting-He/DADeblur.

Keywords: Video deblurring · Domain adaptation · Diffusion model

1 Introduction

Videos captured in dynamic scenes often appear blurred and have blurry artifacts due to camera shaking or moving objects captured during the exposure process.

* equal contribution

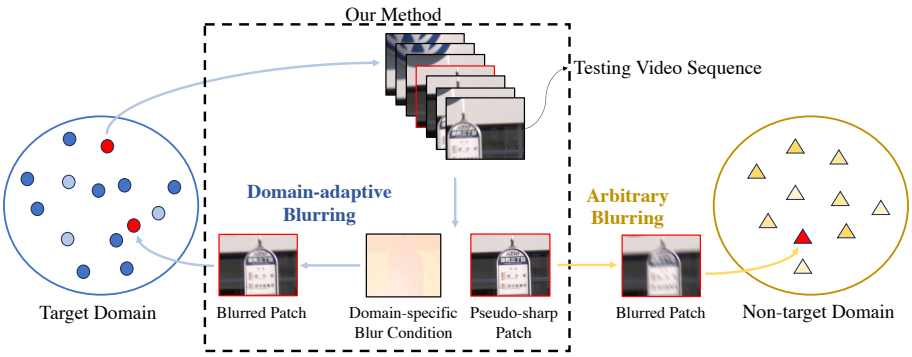


Fig. 1: Illustration of the proposed domain adaptation method. It can generate domain-specific blur conditions for a blurring model to produce blurred video frames from the chosen pseudo-sharp patch. These blurred frames can be used to fine-tune video deblurring models and improve their performance in the target domain.

The unwanted blur severely degrades visual qualities and reduces the accuracy of subsequent computer-vision applications. Dynamic scene video deblurring aims to restore such blurred videos, considered a highly ill-posed problem due to the unknown varying blurs.

Video deblurring has made remarkable progress with the development of deep learning. Numerous methods [3,16,17,22,25,34,36,48,50,56,60,63,64] have applied Convolutional Neural Networks (CNNs) for video deblurring. Among these methods, Recurrent Neural Network-based (RNN-based) methods [3, 16, 56, 63] are commonly employed to extract spatio-temporal features. To better capture motion cues in videos, several methods [22,36,48] compute optical flows to get additional motion information for alignment. In addition, motivated by the success of Transformer in computer vision [2,4,8,11,43,58,65], Transformer-based methods [25, 26, 28] have been proposed to model long-range, spatio-temporal dependency. Since these methods have been dedicated to improving deblurring through architectural designs, their performance often drops substantially when a domain gap exists between training and testing videos.

In real-world cases, cameras may capture blurry videos due to different settings and scenarios, such as shutter speed, aperture, or light sources, leading to different blurry patterns with various orientations and magnitudes. Previous deblurring models often face challenges when dealing with blurry patterns not seen during training. However, few methods discuss the domain gap issue in video deblurring. Liu *et al.* [29] attempted to address it via a blurring network that generates blurred images to enable meta-learning [12] for adaptation. Even though it could generate blurred data for adaptation, it did not consider helpful temporal information that exists in consecutive frames for domain adaptation. That is, continuous motion across consecutive frames reveals motion blur trajectories during capturing, and the degree of blurriness implies the blur intensities

during exposure. The information implicitly conveys domain-specific cues regarding blur orientations and magnitudes in unseen blurry videos. Therefore, motivated by this observation, we further explore the cues to enhance video deblurring models in unseen domains.

In this paper, we propose a domain adaptation method for video deblurring, which relies on a blurring model to generate domain-adaptive training pairs for adaptation. The generated data can be used to calibrate deblurring models in unseen domains. The proposed framework is illustrated in Figure 1. Since blurred and sharp pairs are unavailable for fine-tuning during inference, our scheme can extract relatively sharp regions from blurry videos using the proposed Relative Sharpness Detection Module (RSDM). These relatively sharp regions can be considered pseudo-sharp images. Inspired by the recent diffusion-based blurring method, ID-Blau [57], which can generate blurred images based on a sharp image and arbitrarily specified blur conditions, we adopt ID-Blau as the blurring backbone. Nevertheless, randomly generated blur conditions are not consistent with blur patterns present in test videos. Directly applying ID-Blau with such conditions for blurring and fine-tuning a deblurring model may not help. Considering blurry videos implicitly provide coherent motion blur cues, we propose a Domain-adaptive Blur Condition Generation Module (DBCGM) containing a Blur Orientation Estimator and Blur Magnitude Estimator to create domain-specific blur conditions tailored for blurring those pseudo-sharp images using ID-Blau. It turns out that the generated domain-specific blur pairs can be used to fine-tune deblurring models, thereby achieving domain adaptation during inference. Our contributions can be summarized as follows:

- We propose a test-time domain adaptation method for video deblurring based on ID-Blau, which can generate domain-specific blur conditions to achieve test-time fine-tuning for deblurring models.
- We propose a Relative Sharpness Detection Module to detect relatively sharp regions as pseudo-sharp images and a Domain-adaptive Blur Condition Generation Module to generate domain-specific blur conditions for blurring those pseudo-sharp images.
- Experimental results demonstrate that our method can significantly enhance state-of-the-art video deblurring models [21,37,56,63] on five real-world video deblurring datasets, including RealBlur [44], RBVD [3] and three versions of BSD [63].

2 Related Work

Video Deblurring Convolutional neural networks have significantly advanced the video deblurring task. Compared to image deblurring [13, 19, 32, 38, 39, 49, 51, 53, 54, 59], video deblurring [3, 16, 17, 22, 25, 34, 36, 48, 50, 56, 60, 63, 64] can utilize spatial and temporal information from videos to achieve better performance. Recently, several methods adopted RNN-based models [3, 21, 26, 34, 37, 56, 63] to leverage spatio-temporal information in videos. Zhong *et al.* [63] proposed

an efficient spatio-temporal RNN architecture to catch spatially and temporally varying blurs with RNN cells. Wang *et al*. [56] utilized a motion magnitude prior as guidance to improve deblurring performance. Pan *et al*. [37] developed wavelet-based feature propagation to transfer features in frequency space recurrently. Li *et al*. [21] proposed a grouped spatial-temporal shift operation to aggregate spatio-temporal features efficiently.

In addition to progressively propagating features through RNNs, several studies have explored Transformer-based architectures [25,26,28] to garner long-range information for deblurring. Liang *et al*. [26] proposed a recurrent video restoration transformer with guided deformable attention. Lin *et al*. [28] introduced a flow-guided sparse transformer that leverages optical flows to guide the transformer's attention mechanism. Although these architectural designs for RNNs and Transformers were used to improve video deblurring, they often do not perform up to par when dealing with blurry videos on unseen domains, especially for videos captured in the real world.

Domain Adaptation Domain adaptation aims to bridge the domain gap between the training set and testing set, which has been widely discussed in vision tasks, such as object detection [23, 24, 62] and semantic segmentation [5, 15, 20]. Nevertheless, little work has been done [6,29,33] to address the domain gap issue in deblurring, especially for video deblurring [29]. Since we solely have blurred inputs during inference, no ground truth can be used for test-time deblurring calibration. Some methods apply self-supervised strategies to enable fine-tuning during testing. Chi *et al*. [6] utilized a reconstruction branch to reset its deblurred result to the original blurry input. Nah *et al*. [33] employed a fixed reblurring model as a reblurring loss to supervise the deblurring result. Although these methods could update and improve deblurring models through self-supervised learning, they ignore domain-specific blur information in the test domain, which is beneficial to deblurring performance in unseen domains. To generate blurred images with blur characteristics on unseen domains, Liu *et al*. [29] utilized a GAN to supervise a blurring model to generate blurred images from sharp ones for fine-tuning. In contrast, our method explores domain-specific blur cues from consecutive testing video frames for domain adaptation.

Diffusion Models Diffusion models have demonstrated a strong ability to synthesize realistic images in image generation [10, 14, 47]. Based on the diffusion model, several studies [27,30,31,35,42,45,46,55,57,61] have utilized it with various conditions to generate controllable results. Some methods [35,46] used text prompts [9, 40, 41] in text-to-image generation. Although text embedding provides additional information to generate controllable results, it often has a limitation in spatially guiding the generated images. Recently, several studies proposed to spatially guide diffusion models with various prompts [27,30,31,45,55,57,61]. Zhang *et al*. [61] proposed Control-Net to fine-tune a pre-trained diffusion model with various prompts by zero convolution layers, making the diffusion model spatially controllable without requiring high computational costs. Liang *et al*. [27]

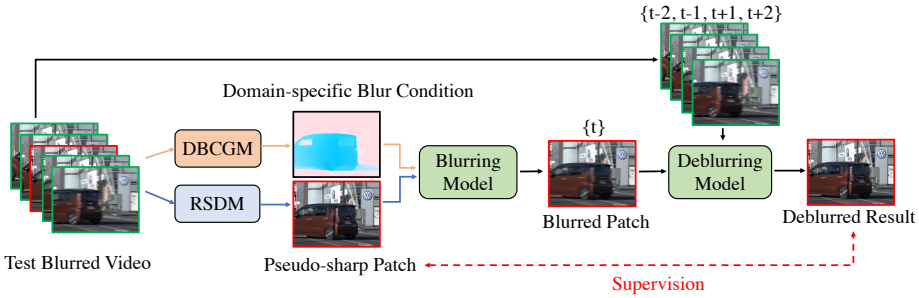


Fig. 2: Pipeline of the proposed domain adaptation scheme. Given a blurred video in the test domain, we use the Relative Sharpness Detection Module (RSDM) to extract relatively sharp patches to be pseudo-sharp patches and the Domain-adaptive Blur Condition Generation Module (DBCGM) to generate domain-specific blur conditions for blurring the pseudo-sharp patches using ID-Blau [57]. Finally, the pseudo-sharp and blurred pairs are used to update a deblurring model for domain adaptation.

proposed a controllable multi-modal image coloring method based on a content-guided deformable autoencoder. Wu *et al.* [57] proposed ID-Blau, a diffusion-based blurring model that can generate diverse blurred images based on a sharp image and arbitrarily-specified pixel-wise blur conditions. Motivated by the success of ID-Blau, we adopt ID-Blau as a blurring model to generate blurred images during testing. However, simply applying ID-Blau with randomly generated blur conditions would not improve a deblurring model for blurred images coming from different domains. Therefore, we aim to leverage the inherent motion blur cues in testing videos to generate domain-specific blur conditions for adaptation, thereby significantly boosting deblurring performance on unseen domains.

3 Proposed Method

We proposed a domain adaptation scheme for video deblurring based on ID-Blau [57], a diffusion-based blurring model that utilizes controllable blur conditions to generate corresponding blurred images. As shown in Figure 2, considering blurred and sharp pairs are unavailable during inference, we generate domain-adaptive training pairs from blurred videos to calibrate a deblurring model for the target domain. First, we propose a Relative Sharpness Detection Module (RSDM) to extract relatively sharp patches from blurred videos. These patches are treated as pseudo-sharp images, which are then blurred using ID-Blau to create pseudo-training pairs for updating deblurring models. In ID-Blau, a blur pattern is represented using pixel-wise blur orientations and magnitudes as a blur condition in a continuous blur condition field. Simply using ID-Blau with randomly sampled blurred conditions to produce blurred images may not help a deblurring model in unseen domains. Therefore, we proposed a Domain-adaptive Blur Condition Generation Module (DBCGM) to create domain-specific blur

conditions for ID-Blau. It allows ID-Blau to generate blurred images adaptive to the target domain. Finally, the generated pseudo-sharp and blurred pairs are used to fine-tune a deblurring model for better performance.

3.1 Relative Sharpness Detection Module (RSDM)

The proposed RSDM aims to search for relatively sharp patches in blurred videos. These patches are regarded as pseudo-sharp patches for domain adaptation. To achieve this, we propose a Blur Magnitude Estimator (BME) to predict a blur magnitude map for a blurred image, where the blurriness degree for each pixel is estimated.

As shown in Figure 3, the BME is a five-stage encoder-decoder network combined with Multi-Scale Feature Fusion (MSFF) proposed in [7], as

$$\tilde{E}_k = \begin{cases} MSFF(E_{k-1}, E_k, E_{k+1}) & \text{if } k = 1, 2, 3, \\ MSFF(E_k, E_{k+1}, E_{k+2}) & \text{if } k = 0, \\ MSFF(E_k, E_{k-1}, E_{k-2}) & \text{if } k = 4, \end{cases} \tag{1}$$

where E_k is the output of the k-th encoder layer and $\tilde{E}_k$ is its fusion result. MSFF includes a 3×3 convolutional layer plus resizing all the input features to the size of E_k. After resizing, the multi-scale features are concatenated, followed by a 1×1 and a 3×3 convolutional layers.

To optimize the BME, we choose the GoPro [32] dataset, which synthesizes blurred images by accumulating consecutive sharp frames from a high-speed camera as

$$B = g(\frac{1}{T}\int_{t=1}^{T} H(t)dt) \simeq g(\frac{1}{N_s}\sum_{n=1}^{N_s} H\left[n\right]), \tag{2}$$

where B, H, T, N_s, and g respectively denote the generated blurred image, the sharp images captured by a high-speed camera, the exposure time, the number of sampled sharp images, and the camera response function. Here, the center sharp image $H[\frac{N+1}{2}]$ is chosen to be the ground-truth sharp image S corresponding to the generated blurred image B.

Considering continuous motion during the exposure process causes various degrees of blurriness, we characterize the continuous exposure as motion trajectories by accumulating optical flows from the sharp image sequence $\mathbf{H} = \{H\left[1\right], ..., H\left[N\right]\}$ and aggregating them as

$$\mathcal{F} = \sum_{n=1}^{N-1} \frac{f(H\left[n\right], H\left[n+1\right]) - f(H\left[n\right], H\left[n-1\right])}{2}, \tag{3}$$

where f is a pre-trained optical flow network. Here, we adopt the method in [52] for obtaining the optical flows. The calculated motion trajectory map has two components: u and v, representing the average horizontal and vertical motion

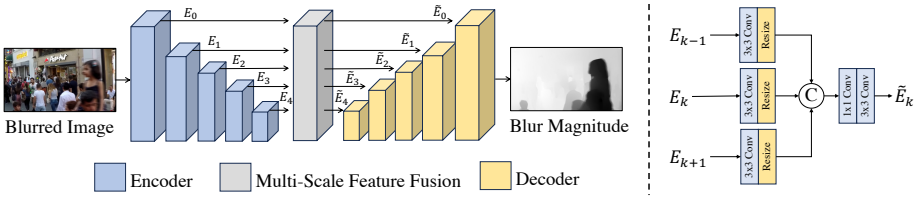


Fig. 3: The left figure shows the architecture of the Blur Magnitude Estimator (BME), which comprises a five-stage encoder-decoder design with Multi-Scale Feature Fusion (MSFF). The right figure is the architecture of MSFF. E_k denotes the output of the encoder at stage k, and $\tilde{E}_k$ denotes its output after MSFF.

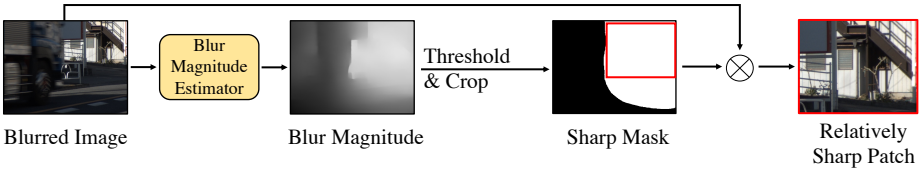


Fig. 4: Illustration of the Relative Sharpness Detection Module. We use the Blur Magnitude Estimator to obtain the blur magnitudes for a blurred image and crop a relatively sharp patch based on an adaptive sharpness threshold.

trajectories as $\mathcal{F} = [u; v] \in \mathbb{R}^{H\times W\times 2}$. Lastly, the blur magnitude map $G \in \mathbb{R}^{H\times W}$ corresponding to B can be obtained by

$$G = \frac{1}{\tau}\sqrt{u^2 + v^2}, \tag{4}$$

where τ serves as a normalization term, set to the maximum value of the blur magnitudes, and thus $G \in [0, 1]$. To this end, a dataset $\{B_k, G_k\}_{k=1}^{K}$ is constructed to train the BME, where $K = 2,103$ for the GoPro training set.

Next, we take the trained BME to obtain the blur magnitudes for every frame in the testing blurred video to sift through relatively sharp patches. As shown in Figure 4, given a blurred video $V^{(i)}$ in the test set, BME predicts a blur magnitude map $M_t^{(i)} \in \mathbb{R}^{H\times W}$ for each blurred frame $V_t^{(i)}$ as

$$M_t^{(i)} = BME(V_t^{(i)}), \tag{5}$$

where t denotes the frame index. To crop a relatively sharp patch from $V_t^{(i)}$, we use an adaptive sharpness threshold $\eta^{(i)}$ to binarize $M_t^{(i)}$ with magnitude larger than $\eta^{(i)}$ set to 1 and obtain the sharp mask. We then crop a sharp patch by the regionprop library [1] based on the mask to generate a relatively sharp patch $\tilde{S}_t^{(i)}$ of size 256×256. Here, $\eta^{(i)}$ is determined to ensure that the number of relatively sharp patches extracted from the video $V^{(i)}$ reaches $r\%$ of the number of total

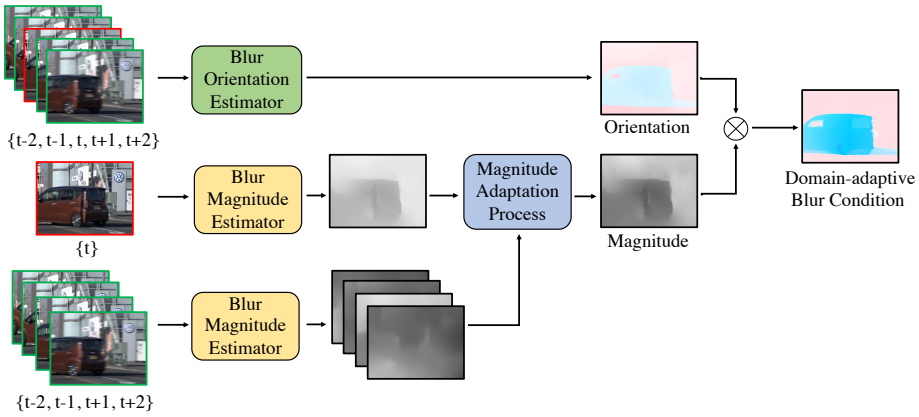


Fig. 5: Illustration of Domain-adaptive Blur Condition Generation Module. Given a pseudo-sharp patch and the collocated patches in its neighboring frames, we use the Blur Orientation Estimator to generate domain-specific blur orientations and the Blur Magnitude Estimator to generate domain-specific blur magnitudes. The blur magnitudes estimated from neighboring patches are used to modulate the pseudo-sharp patch by the Magnitude Adaptation Process. In the end, the domain-specific blur orientations and magnitudes are used for blurring.

frames in $V^{(i)}$, meaning the top $r\%$ relatively sharp patches in the video. In our work, r is set to 20.

3.2 Domain-adaptive Blur Condition Generation Module (DBCGM)

With the selected pseudo-sharp patches, we utilize a blurring model to generate their blurred versions for fine-tuning a deblurring model. In this work, we use ID-Blau [57], a conditional diffusion-based blurring model that takes a sharp image S and a controllable blur condition map $C = (x, y, z) \in \mathbb{R}^{H \times W \times 3}$ to synthesize a blurred image B as

$$B = \textbf{ID-Blau}(S, C), \tag{6}$$

where x, y, and $z \in \mathbb{R}^{H \times W}$ denote horizontal and vertical blur orientations, and blur magnitudes, respectively. Although randomly generated blur conditions can be used to produce blurred images by ID-Blau, they do not conform to the test data distribution in the target domain, which means the generated blurred images do not agree with the blur patterns that exist in test videos, leading to little performance gain for deblurring models. Therefore, as shown in Figure 5, we propose a Domain-adaptive Blur Condition Generation Module (DBCGM) that can generate domain-specific blur conditions by leveraging motion cues implicitly provided in blurred videos during inference.

In a blurred video, continuous motion across consecutive frames reveals motion blur trajectories during capturing, and the degree of blurriness implies the blur intensities during exposure. Therefore, the proposed DBCGM includes a Blur Orientation Estimator (BOE) and the previously mentioned BME to generate domain-specific blur conditions for blurring. First, given a pseudo-sharp patch $\tilde{S}_t^{(i)}$ in the t-th frame of the i-th video, we take it and its collocated patches from the two previous and two next frames $\{\tilde{S}_{t-2}^{(i)}, ..., \tilde{S}_{t+2}^{(i)}\}$ to calculate the motion trajectory map $\tilde{F}_t^{(i)} = [\tilde{u}; \tilde{v}] \in \mathbb{R}^{H \times W \times 2}$ as

$$\tilde{\mathcal{F}}_t^{(i)} = \sum_{n=-2}^{1} f(\tilde{S}_{t+n}^{(i)}, \tilde{S}_{t+n+1}^{(i)}), \tag{7}$$

where f denotes the pre-trained optical flow estimator [52]. Next, we obtain the domain-specific blur orientations $\tilde{O}_t^{(i)} \in \mathbb{R}^{H \times W \times 2}$ tailored for $\tilde{S}_t^{(i)}$ as

$$\tilde{O}_t^{(i)} = \frac{\tilde{\mathcal{F}}_t^{(i)}}{\sqrt{\tilde{u}^2 + \tilde{v}^2}}. \tag{8}$$

To generate domain-adaptive blur magnitudes for $\tilde{S}_t^{(i)}$, we first estimate blur magnitudes of $\tilde{S}_t^{(i)}$ by the BME as $M_t^{(i)} = BME(\tilde{S}_t^{(i)})$. Since $\tilde{S}_t^{(i)}$ is relatively sharp, we intend to modulate its blur magnitudes by considering blur patterns rendered in its neighboring collocated patches $\{\tilde{S}_{t-2}^{(i)}, \tilde{S}_{t-1}^{(i)}, \tilde{S}_{t+1}^{(i)}, \tilde{S}_{t+2}^{(i)}\}$. To achieve this, we use the average blur magnitudes of neighboring collocated patches to adjust $M_t^{(i)}$ by the Magnitude Adaptation Process as

$$\tilde{M}_t^{(i)} = \text{Norm}(M_t^{(i)}) \cdot \text{Avg}(M_{t-2}^{(i)}, M_{t-1}^{(i)}, M_{t+1}^{(i)}, M_{t+2}^{(i)}), \tag{9}$$

where $M_t^{(i)}$ is normalized in the range of $[0, 1]$ and multiplied by the blur magnitude average to generate domain-specific blur magnitudes $\tilde{M}_t^{(i)} \in \mathbb{R}^{H \times W}$. At last, a domain-specific blur condition $\tilde{C}_t^{(i)}$ that combines blur orientations $\tilde{O}_t^{(i)} \in \mathbb{R}^{H \times W \times 2}$ and magnitudes $\tilde{M}_t^{(i)} \in \mathbb{R}^{H \times W}$ is used to blur $\tilde{S}_t^{(i)}$ by

$$\tilde{B}_t^{(i)} = \textbf{ID-Blau}(\tilde{S}_t^{(i)}, \tilde{C}_t^{(i)}), \tag{10}$$

where $\tilde{B}_t^{(i)}$ is the generated blurred patch. The pseudo-training pair $\{\tilde{B}_t^{(i)}, \tilde{S}_t^{(i)}\}$ is then utilized to update a deblurring model for domain adaptation. Ultimately, we collect these pairs from the total N blurred videos in the target domain to fine-tune a deblurring model for domain adaptation.

Loss Function: In the proposed scheme, we only need to optimize the BME. We use the L1 loss,

$$\mathcal{L} = \mathcal{L}_1(M, G), \tag{11}$$

where M is the predicted blur magnitudes, and G is the calculated blur magnitudes using Equation 4, used as the ground truth.

The authors from the universities in Taiwan completed the experiments on the datasets.

Table 1: Evaluation results on the deblurring datasets, including BSD [63], RealBlur [44], and RVBD [3], where "Baseline" and "+Ours" denote the video deblurring performances w/o or w/ our domain adaptation method.

Model		BSD-1ms8ms		BSD-2ms16ms		BSD-3ms24ms		RealBlur		RBVD	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ESTRNN	Baseline	25.57	0.747	24.64	0.726	26.01	0.748	25.87	0.773	24.47	0.725
	+Ours	**29.44**	**0.843**	**28.36**	**0.820**	**28.32**	**0.810**	**27.64**	**0.816**	**26.83**	**0.764**
MMP-RNN	Baseline	21.63	0.620	21.26	0.605	22.74	0.597	24.65	0.639	22.81	0.780
	+Ours	**29.17**	**0.797**	**26.95**	**0.750**	**26.77**	**0.707**	**27.69**	**0.663**	**25.81**	**0.822**
DSTNet	Baseline	25.42	0.821	23.50	0.760	24.68	0.788	26.57	0.750	23.15	0.768
	+Ours	**28.69**	**0.868**	**27.11**	**0.830**	**26.69**	**0.838**	**27.74**	**0.798**	**25.66**	**0.808**
Shift-Net	Baseline	25.00	0.837	23.75	0.807	24.98	0.819	26.01	0.797	23.98	0.870
	+Ours	**28.75**	**0.888**	**26.31**	**0.854**	**26.92**	**0.852**	**27.71**	**0.854**	**25.35**	**0.891**

4 Experiments

4.1 Implementation Details

Blurring Model We choose a diffusion-based blurring ID-Blau [57] as the blurring network to generate blurred images. Note that we use ID-Blau with its original training settings. For optimizing the BME, we use the Adam optimizer [18] with the initial learning rate of $1e^{-3}$, gradually decayed to $1e^{-4}$ by the cosine annealing strategy. We resize images to 320×320 and adopt random flipping and rotation for data augmentation with a batch size of 16 for training the model 50 epochs. We use the GoPro training set [32], which contains 22 training videos with $2,103$ blurred and sharp image pairs for optimizing ID-Blau and BME.

Video Deblurring Models We adopt four state-of-the-art video deblurring models, including ESTRNN [63], MMP-RNN [56], DSTNet [37], and Shift-Net [21] to validate the effectiveness of the proposed domain adaptation method. We regard the GoPro training set as the source domain and five real-world deblurring datasets as target domains, including BSD-1ms8ms [63], BSD-2ms16ms [63], BSD-3ms24ms [63], RealBlur [44], and RBVD [3] test sets. The BSD test set has three subsets: 1ms8ms, 2ms16ms, and 3ms24ms, indicating two exposure times: the former of which is for sharp images and the latter of which is for blurred ones. Each subset contains 20 videos with $3,000$ blurred images. The RealBlur test set consists of 50 videos with 980 blurred images, taken in low-light environments. The RBVD test dataset consists of 7 videos with 246 blurred images, encompassing outdoor scenes, indoor scenes, and high-frequency charts. In the domain adaptation process, we fine-tune each deblurring model for 10 epochs on pseudo-training pairs generated by the proposed RSDM and DBCGM, as we use each model's original loss functions for training.

4.2 Experimental Results

Quantitative Analysis We compare the performance of four video deblurring models, including ESTRNN [63], MMP-RNN [56], DSTNet [37], and Shift-

Table 2: Ablations on the effectiveness of the RSDM and DBCGM.

	Pseudo-Sharp Patches		Blur Condition Generation			PSNR GAIN	
	Random	RSDM(Ours)	Random	Optical-Flow	DBCGM(Ours)		
(a)						25.57	+0.00
(b)	✓		✓			23.88	-1.69
(c)	✓			✓		25.51	-0.06
(d)	✓				✓	29.01	+3.44
(e)		✓	✓			24.32	-1.25
(f)		✓		✓		26.19	+0.62
(g)		✓			✓	**29.44**	**+3.87**

Net [21], on five real-world deblurring datasets with (+Ours) or without (Baseline) using the proposed domain adaptation scheme in Table 1. It shows that our method can consistently and significantly improve the performance for these video deblurring models by 4.61dB, 3.90dB, 2.57dB, 1.92dB, and 2.31dB in PSNR on average on the BSD-1ms8ms, BSD-2ms16ms, BSD-3ms24ms, RealBlur and RBVD test sets, respectively. Particularly, our method can obtain up to 3.87dB, 7.54dB, 3.61dB, and 3.75dB performance gains for deblurring models: ESTRNN, MMP-RNN, DSTNet, and Shift-Net, respectively. These experimental results show that our method can generate domain-adaptive training pairs to effectively fine-tune deblurring models originally trained with a synthetic dataset for those real-world blurred videos.

Qualitative Analysis In Figure 6, we compare deblurred results with or without using our method on several blurred video frames chosen from the BSD test set. Our domain-adaptive scheme can significantly improve the visual quality of the deblurred results, exemplified by the regions of the face, text, and pedestrians. Figure 6 shows comparisons of the deblurred results with or without using our method on blurred examples on the RBVD and RealBlur test sets. The visual quality of deblurred results using our method is significantly better, exemplified by the regions of the pillar, text, and building. These visualization results show that our method can truly help boost the deblurring performance compared to the baseline methods. It also attests to the effectiveness of the proposed domain adaptation scheme to produce domain-adaptive training pairs for model fine-tuning in the target domain.

4.3 Ablation studies

To analyze the impact of the proposed components in our domain adaptation method, we adopt ESTRNN [63] as the tested deblurring model and analyze its deblurring performances on the BSD-1ms8ms dataset with various ablative settings. First, we discuss the effectiveness of the RSDM and DBCGM. Next, we discuss the designs in RSDM and DBCGM. Lastly, we compare the proposed method with the existing domain adaptation method [29] for video deblurring.

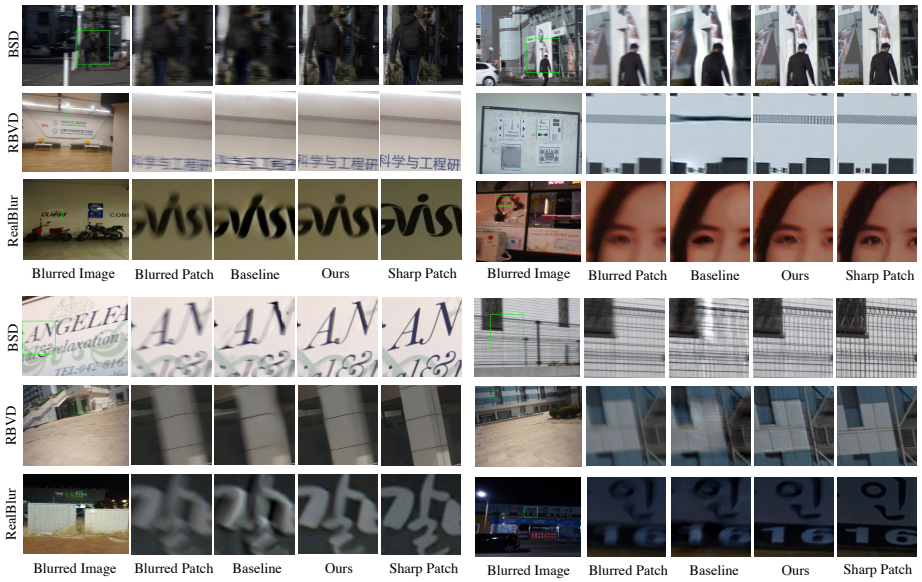


Fig. 6: Qualitative results on the BSD datasets, RBVD dataset and RealBlur dataset with DSTNet(top left), MMP-RNN(top right), Shift-Net(bottom left), and ESTRNN(bottom right).

Effects of RSDM and DBCGM Table 2 compares the effectiveness of the RSDM and DBCGM to the video deblurring performance. Table 2(a) shows the deblurred results on BSD-1m8ms using the GoPro pre-trained model, considered the baseline setting. First, we analyze the effectiveness of the proposed DBCGM with randomly selected patches in blurred videos to be pseudo-sharp patches (Table 2(b), (c), and (d)). To blur pseudo-sharp patches using ID-Blau, we can use randomly sampled blur conditions (Table 2(b)) or take optical flows as blur conditions (Table 2(c)). As seen, these two cases do not improve the performance at all. In contrast, using the proposed DBCGM (Table 2(d)) can largely improve the baseline by 3.44dB, which demonstrates that even with random patches cropped from blurred video frames, the deblurring model can still exploit domain-specific blur conditions generated by the DBCGM to achieve better performance. Next, we analyze the effectiveness of the proposed RSDM. In Table 2, we can compare the settings (b) vs. (e), (c) vs. (f), and (d) vs. (g), which shows that using the RSDM with different blur condition generation methods can all boost the deblurring model. Most of all, the combination of the RSDM and DBCGM achieves the best performance, achieving a 3.87dB gain compared to the baseline. In addition, we visualize the blurred results in Figure 7 using different blur condition generation methods. It shows that the proposed DBCGM can generate domain-specific blurred images more realistic and consistent with

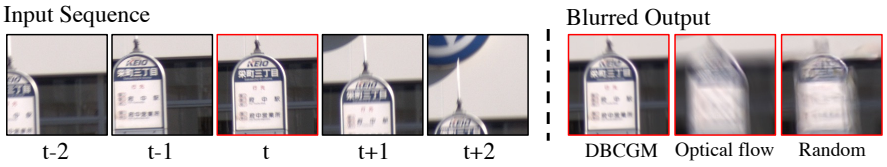


Fig. 7: Visualization of blurred images obtained using ID-Blau with different blur condition generation methods: DBCGM, optical flows, and random sampling.

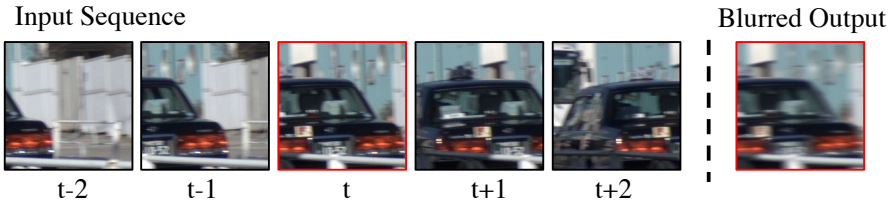
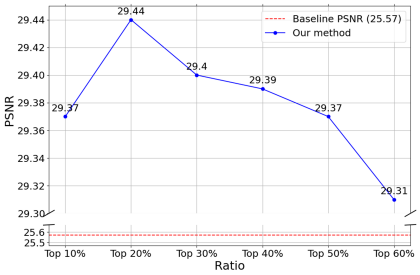


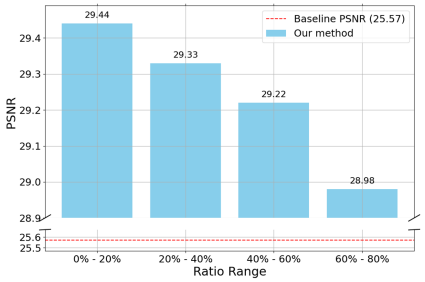
Fig. 8: Example of a blurred image generated from a blurry image using our method.

video frames of a moving scene than optical flows or randomly sampled blur conditions.

Effects of the adaptive sharpness threshold for the RSDM The adaptive sharpness threshold $\eta^{(i)}$ is determined for the video $V^{(i)}$ in the RSDM to pick the top $r\%$ relatively sharp patches. Choosing a smaller ratio can result in sharper patches to be pseudo-sharp patches. However, it may not provide sufficient training data for adaptation. In contrast, selecting a larger ratio would include more blurred patches as pseudo-sharp patches, leading to inferior pseudo-training pairs generated, where those blurred patches are too similar to their blurred versions. Therefore, we experiment to analyze the effect of using different ratios $r\%$ on the deblurring performance for adaptation. Figure 9a shows the deblurring performance of using different ratios from 10% ($3,000 \times 10\% = 300$ patches) to 60% ($3,000 \times 60\% = 1,800$ patches). As seen, setting $r = 20$ achieves the best performance. Choosing a larger ratio would include more inferior training pairs, providing less help for deblurring. In addition, we conduct another experiment by using patches in different ratio ranges for blurring, including $0\% - 20\%$, $20\% - 40\%$, $40\% - 60\%$, and $60\% - 80\%$. For example, $20\% - 40\%$ means we select the top 40% relatively sharp patches but exclude the top 20% ones. Based on Figure 9b, using patches with more blur to generate pseudo-training pairs is not as beneficial as those with less blur to obtain quality pairs. That said, compared to the baseline, intentionally selecting patches with more blur in the proposed domain adaptation scheme still improves the deblurring performance. Figure 8 shows a blurred example using a patch with much blur.



(a) Performance of using different ratios for the threshold in RSDM



(b) Performance of using different ratio ranges for picking pseudo-sharp patches

Fig. 9: Illustration of the PSNR results when we set different threshold for RSDM.

Table 3: Comparison between our method and Liu *et al.* [29].

Model	BSD-1ms8ms	BSD-2ms16ms	BSD-3ms24ms	RealBlur	RBVD
Baseline	25.57	24.64	26.01	25.87	24.47
Liu *et al.* [29]	25.58	24.53	25.15	26.12	24.83
Ours	**29.44**	**28.36**	**28.32**	**27.64**	**26.83**

4.4 Comparison with an existing domain adaptation method

Considering little work regarding domain adaptation for video deblurring, we only compare the proposed method with [29], which uses meta-learning to achieve domain adaptation. Table 3 shows that our method significantly outperforms [29] on all the benchmark datasets.

5 Conclusion

We proposed a domain adaptation scheme for video deblurring based on a conditional diffusion-based blurring model to achieve test-time fine-tuning in unseen domains. To generate domain-adaptive training pairs for updating deblurring models during inference, we proposed two modules: the Relative Sharpness Detection Module (RSDM) and the Domain-adaptive Blur Condition Generation Module (DBCGM). The RSDM extracts relatively sharp patches from blurred input frames as pseudo-sharp images. The DBCGM generates domain-specific blur conditions for blurring the pseudo-sharp images. Lastly, the generated pseudo-sharp and blurred pairs are used to calibrate a deblurring model for better performance. Extensive experimental results have demonstrated that our approach can significantly improve state-of-the-art video deblurring methods, offering performance gains of up to 7.54dB on various real-world video deblurring datasets.

Acknowledgments

This work was supported in part by the National Science and Technology Council (NSTC) under grants 112-2221-EA49-090-MY3, 111-2628-E-A49-025-MY3, 112-2634-F002-005, 112-2221-E-004-005, 113-2923-E-A49-003-MY2, 113-2221-E-004-001-MY3 and 113-2622-E-004-001 This work was funded in part by Qualcomm through a Taiwan University Research Collaboration Project.

References

1. https://scikit-image.org/docs/stable/auto_examples/segmentation/plot_regionprops.html
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: ECCV (2020)
3. Chao, Z., Hang, D., Jinshan, P., Boyang, L., Yuhao, H., Lean, F., Fei, W.: Deep recurrent neural network with multi-scale bi-directional propagation for video deblurring. In: AAAI (2022)
4. Chen, C.F., Panda, R., Fan, Q.: Regionvit: Regional-to-local attention for vision transformers. In: ICLR (2022)
5. Chen, L., Wei, Z., Jin, X., Chen, H., Zheng, M., Chen, K., Jin, Y.: Deliberated domain bridging for domain adaptive semantic segmentation. In: NeurIPS (2022)
6. Chi, Z., Wang, Y., Yu, Y., Tang, J.: Test-time fast adaptation for dynamic scene deblurring via meta-auxiliary learning. In: CVPR (2021)
7. 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: ICCV (2021)
8. Chu, X., Tian, Z., Wang, Y., Zhang, B., Ren, H., Wei, X., Xia, H., Shen, C.: Twins: Revisiting the design of spatial attention in vision transformers. In: NeurIPS (2021)
9. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv (2018)
10. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. In: NeurIPS (2021)
11. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2021)
12. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: ICML (2017)
13. Gao, H., Tao, X., Shen, X., Jia, J.: Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: CVPR (2019)
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020)
15. Hoyer, L., Dai, D., Van Gool, L.: HRDA: Context-aware high-resolution domain-adaptive semantic segmentation. In: ECCV (2022)
16. Ji, B., Yao, A.: Multi-scale memory-based video deblurring. In: CVPR (2022)
17. Jiang, B., Xie, Z., Xia, Z., Li, S., Liu, S.: Erdn: Equivalent receptive field deformable network for video deblurring. In: ECCV (2022)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2017)

19. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: ICCV (2019)
20. Lai, X., Tian, Z., Xu, X., Chen, Y., Liu, S., Zhao, H., Wang, L., Jia, J.: Decouplenet: Decoupled network for domain adaptive semantic segmentation. In: ECCV (2022)
21. Li, D., Shi, X., Zhang, Y., Cheung, K.C., See, S., Wang, X., Qin, H., Li, H.: A simple baseline for video restoration with grouped spatial-temporal shift. In: CVPR (2023)
22. Li, D., Xu, C., Zhang, K., Yu, X., Zhong, Y., Ren, W., Suominen, H., Li, H.: Arvo: Learning all-range volumetric correspondence for video deblurring. In: CVPR (2021)
23. Li, W., Liu, X., Yuan, Y.: Sigma: Semantic-complete graph matching for domain adaptive object detection. In: CVPR (2022)
24. Li, Y.J., Dai, X., Ma, C.Y., Liu, Y.C., Chen, K., Wu, B., He, Z., Kitani, K., Vajda, P.: Cross-domain adaptive teacher for object detection. In: CVPR (2022)
25. Liang, J., Cao, J., Fan, Y., Zhang, K., Ranjan, R., Li, Y., Timofte, R., Van Gool, L.: Vrt: A video restoration transformer. In: arXiv (2022)
26. Liang, J., Fan, Y., Xiang, X., Ranjan, R., Ilg, E., Green, S., Cao, J., Zhang, K., Timofte, R., Van Gool, L.: Recurrent video restoration transformer with guided deformable attention. In: NeurIPS (2022)
27. Liang, Z., Li, Z., Zhou, S., Li, C., Loy, C.C.: Control color: Multimodal diffusion-based interactive image colorization. arXiv (2024)
28. Lin, J., Cai, Y., Hu, X., Wang, H., Yan, Y., Zou, X., Ding, H., Zhang, Y., Timofte, R., Van Gool, L.: Flow-guided sparse transformer for video deblurring. In: ICML (2022)
29. Liu, P.S., Tsai, F.J., Peng, Y.T., Tsai, C.C., Lin, C.W., Lin, Y.Y.: Meta transferring for deblurring. In: BMVC (2022)
30. Ma, J., Liang, J., Chen, C., Lu, H.: Subject-diffusion: Open domain personalized text-to-image generation without test-time fine-tuning. arXiv (2023)
31. Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.Y., Ermon, S.: Sdedit: Guided image synthesis and editing with stochastic differential equations. arXiv (2021)
32. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: CVPR (2017)
33. Nah, S., Son, S., Lee, J., Lee, K.M.: Clean images are hard to reblur: A new clue for deblurring. In: ICLR (2022)
34. Nah, S., Son, S., Lee, K.M.: Recurrent neural networks with intra-frame iterations for video deblurring. In: CVPR (2019)
35. Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., Chen, M.: Glide: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv (2021)
36. Pan, J., Bai, H., Tang, J.: Cascaded deep video deblurring using temporal sharpness prior. In: CVPR (2020)
37. Pan, J., Xu, B., Dong, J., Ge, J., Tang, J.: Deep discriminative spatial and temporal network for efficient video deblurring. In: CVPR (2023)
38. 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: ECCV (2020)
39. Purohit, K., Rajagopalan, A.N.: Region-adaptive dense network for efficient motion deblurring. In: AAAI (2020)
40. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: ICML (2021)

41. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.: Exploring the limits of transfer learning with a unified text-to-text transformer. JMLR (2020)
42. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I.: Zero-shot text-to-image generation. In: ICML (2021)
43. Ranftl, R., Bochkovskiy, A., Koltun, V.: Vision transformers for dense prediction. In: ICCV (2021)
44. Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: ECCV (2020)
45. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: CVPR (2023)
46. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., Salimans, T., et al.: Photorealistic text-to-image diffusion models with deep language understanding. NeurIPS (2022)
47. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: ICLR (2021)
48. Su, S., Delbracio, M., Wang, J., Sapiro, G., Heidrich, W., Wang, O.: Deep video deblurring for hand-held cameras. In: CVPR (2017)
49. Suin, M., Purohit, K., Rajagopalan, A.N.: Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: CVPR (2020)
50. Suin, M., Rajagopalan, A.N.: Gated spatio-temporal attention-guided video deblurring. In: CVPR (2021)
51. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: CVPR (2018)
52. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: ECCV (2020)
53. Tsai, F.J., Peng, Y.T., Lin, Y.Y., Tsai, C.C., Lin, C.W.: Stripformer: Strip transformer for fast image deblurring. In: ECCV (2022)
54. Tsai, F.J., Peng, Y.T., Tsai, C.C., Lin, Y.Y., Lin, C.W.: Banet: A blur-aware attention network for dynamic scene deblurring. IEEE TIP (2022)
55. Voynov, A., Aberman, K., Cohen-Or, D.: Sketch-guided text-to-image diffusion models. In: SIGGRAPH (2023)
56. Wang, Y., Lu, Y., Gao, Y., Wang, L., Zhong, Z., Zheng, Y., Yamashita, A.: Efficient video deblurring guided by motion magnitude. In: ECCV (2022)
57. Wu, J.H., Tsai, F.J., Peng, Y.T., Tsai, C.C., Lin, C.W., Lin, Y.Y.: Id-blau: Image deblurring by implicit diffusion-based reblurring augmentation. In: CVPR (2024)
58. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. In: NeurIPS (2021)
59. Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: CVPR (2019)
60. Zhang, H., Xie, H., Yao, H.: Spatio-temporal deformable attention network for video deblurring. In: ECCV (2022)
61. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: ICCV (2023)
62. Zhang, Z., Hoai, M.: Object detection with self-supervised scene adaptation. In: CVPR (2023)
63. Zhong, Z., Gao, Y., Zheng, Y., Zheng, B., Sato, I.: Real-world video deblurring: A benchmark dataset and an efficient recurrent neural network. IJCV (2022)
64. Zhou, K., Li, W., Lu, L., Han, X., Lu, J.: Revisiting temporal alignment for video restoration. In: CVPR (2022)

65. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. In: ICLR (2021)