

# Meta Transferring for Deblurring

Po-Sheng Liu<sup>1</sup>

bensonliu.cs09@nycu.edu.tw

Fu-Jen Tsai<sup>2</sup>

fjtsai@gapp.nthu.edu.tw

Yan-Tsung Peng<sup>3</sup>

ytpeng@cs.nccu.edu.tw

Chung-Chi Tsai<sup>4</sup>

charles0184@gmail.com

Chia-Wen Lin<sup>2</sup>

cwlin@ee.nthu.edu.tw

Yen-Yu Lin<sup>1</sup>

lin@cs.nycu.edu.tw

<sup>1</sup> National Yang Ming Chiao Tung University, Taiwan

<sup>2</sup> National Tsing Hua University, Taiwan

<sup>3</sup> National Chengchi University, Taiwan

<sup>4</sup> Qualcomm Technologies, Inc., United States

---

## Abstract

Most previous deblurring methods were built with a generic model trained on blurred images and their sharp counterparts. However, these approaches might have sub-optimal deblurring results due to the domain gap between the training and test sets. This paper proposes a reblur-deblur meta-transferring scheme to realize test-time adaptation without using ground truth for dynamic scene deblurring. Since the ground truth is usually unavailable at inference time in a real-world scenario, we leverage the blurred input video to find and use relatively sharp patches as the pseudo ground truth. Furthermore, we propose a reblurring model to extract the homogenous blur from the blurred input and transfer it to the pseudo-sharps to obtain the corresponding pseudo-blurred patches for meta-learning and test-time adaptation with only a few gradient updates. Extensive experimental results show that our reblur-deblur meta-learning scheme can improve state-of-the-art deblurring models on the DVD, REDS, and RealBlur benchmark datasets. The source code is available at [https://github.com/po-sheng/Meta\\_Transferring\\_for\\_Deblurring](https://github.com/po-sheng/Meta_Transferring_for_Deblurring).

## 1 Introduction

Dynamic scene deblurring aims to recover sharp images from blurred ones caused by camera shakes or moving objects. It is challenging to recover images with such blur since dynamic scene blur is often non-uniform and directional. Using methods with some prior assumptions [1, 20, 53] or blur uniformity [0] can only achieve limited performance. Existing learning-based methods for single image deblurring [6, 12, 23, 32, 34, 35, 36] or video deblurring [11, 8, 14, 15, 18, 21, 26, 29, 38, 39] could obtain better results by training these models in a supervised manner, learning to deblur images from training with blurred and

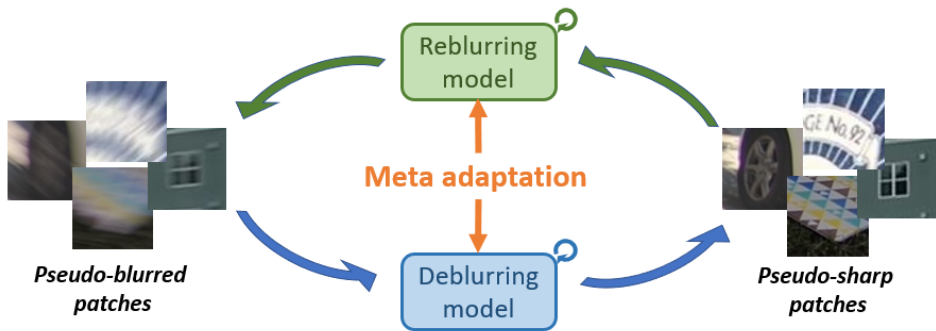


Figure 1: Illustration of the proposed reblur-deblur meta-transferring scheme. We utilize the reblurring model to generate pseudo-blurred and pseudo-sharp pairs to facilitate meta-learning for the test-time adaptation. Note that pseudo-sharp patches are selected from blurred video frames.

sharp image pairs. Although significant progress has been made using these supervised-learning methods, they did not consider exploiting rich internal information in test data and tend to be sub-optimal during testing. Therefore, we aim to build a domain adaptation strategy that allows test-time adaptation to improve existing deblurring methods. Few deblurring works have been done using meta-learning since ground truth is not accessible during testing. Chi *et al.* [5] proposed a self-supervised meta-auxiliary learning strategy that meta-trains and meta-tests the model by restoring the input blurred image. However, it could only achieve sub-optimal adaptation since blur self-restoration may not help to deblur. By contrast, we observe that a scene appearing in multiple frames in a video often has different blurring degrees. Some local patches are relatively sharp, whereas some are relatively blurred. Therefore, we propose to leverage these patches to synthesize sharp and blurred pairs to enable meta-learning for dynamic scene deblurring during testing without access to the ground truth like [5] but in a different approach. More specifically, this work proposes a reblur-deblur meta-transferring scheme to generate pseudo-blurred and pseudo-sharp pairs to achieve test-time adaptation in meta-learning, as shown in Fig. 1. Our reblurring model can transfer blurred patterns homogeneous to the task to pseudo-sharp patches selected to synthesize pseudo-blurred patches, which can serve as pseudo-sharp-and-blurred pairs as the support set in meta-learning. Our contributions are three-fold. First, we propose a novel reblur-deblur meta-transferring scheme that can generate pseudo-blurred and pseudo-sharp pairs for meta-learning. Second, the proposed scheme facilitates meta-learning for dynamic scene deblurring without extra training data needed. Third, extensive experimental results show that our method improves the performance of existing deblurring models on various datasets, including DVD [28], REDS [17], and RealBlur-J [25].

## 2 Related Works

### 2.1 Dynamic scene deblurring

Previous dynamic scene deblurring methods can be roughly divided into single image deblurring [6, 12, 23, 30, 31, 32, 34, 35, 36] and video deblurring [11, 8, 14, 15, 18, 21, 26,

[29, 38, 39]. The main difference is the number of input frames. Video deblurring utilizes consecutive frames to help deblurring, which can obtain more information than image deblurring. However, most existing methods fail to consider internal information from testing data and ignore the problem of domain gap between training and testing sets, which would be sub-optimal during testing.

## 2.2 Meta learning on low-level vision tasks

Previous low-level vision methods commonly train a generic model using training data without considering the rich internal information in test data. Recently, a few methods developed for low-level vision tasks have been proposed [6, 24, 27]. Super-resolution (SR) methods [24, 27] utilized MAML [9] to efficiently adapt to test images for better SR results, where they used the patch-recurrence property intrinsic to an image itself, downsampling images to generate low- and high-resolution training pairs during testing. Since finding such an intrinsic characteristic from a blurred image is challenging, Chi *et al.* [6] proposed a meta-auxiliary learning method, where they meta-trained the deblurring model to learn the input blurred image in a self-supervised manner to help deblur. Thus, it can enable the model to adapt to test blurred images without using their ground-truth counterparts. Nevertheless, restoring the input blurred image and deblurring it essentially works in a contradictory fashion, which may only produce sub-optimal results. Unlike meta-auxiliary learning for deblurring proposed in [6] requires attaching additional layers to train the model, our work builds a reblurring model that can transfer blur patterns from relative-blurred to relative-sharp patches. It can synthesize pseudo-blurred and pseudo-sharp pairs for deblurring models without modifying them to facilitate meta-learning.

## 2.3 Reblurring for deblurring.

Some recent deblurring works [6, 19, 22, 37] were proposed to generate additional blurred images to train the models. Zhang *et al.* [37] used an extra real blurred dataset to produce synthetic blurred images using generative adversarial networks. Park *et al.* [22] used a recurrent method to produce images with various blurring degrees. However, the generated data for training may not help upon testing. To enable test-time adaptation, Nah *et al.* [19] utilizes a reblurring loss to constrain deblurring results using a pre-trained reblurring model based on the observation that clean images are hard to reblur. Chen *et al.* [6] chose to self-supervise a model for deblurring, similar to [6]. In contrast, we leverage the internal information in testing videos to generate domain-aware pseudo-blurred and pseudo-sharp pairs, enabling meta-learning better than [6, 6].

# 3 Proposed Method

## 3.1 Overview

The paper proposes a reblur-deblur meta-transferring scheme for test-time adaptation, composed of meta-training and meta-testing phases.

In meta-training, we consider deblurring each video  $V_i$  in the training set as a task, where  $V_i$  contains  $N$  blurred frames  $B_n$  and their sharp counterparts  $S_n$  ( $n = \{1, 2, \dots, N\}$ ). Figure 2 depicts the proposed reblur-deblur meta-transferring training scheme, where the inner update trains a deblurring model on the support set generated by the reblurring model, and the

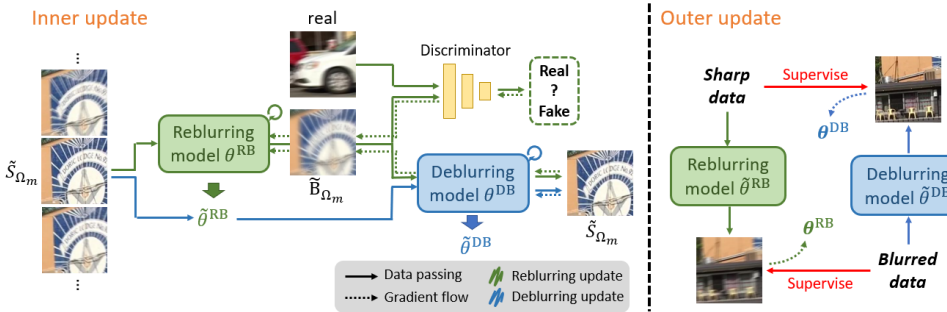


Figure 2: Illustration of our reblur-deblur meta-transferring training scheme. In the inner update, we use an adapted reblurring model to generate a domain-aware support set, enabling meta-learning to update the deblurring model. In the outer update, we evaluate the adapted models by query set to obtain meta-learned weights  $\theta^{RB}$  and  $\theta^{DB}$ .

outer update trains the deblurring and reblurring models adapted to the query set. Like [5], not needing the ground truth at inference as well, we instead use the proposed reblurring model to generate  $M$  pseudo-blurred patches  $\tilde{B}_{\Omega_m}$  from pseudo-sharp patches  $\tilde{S}_{\Omega_m}$  selected from  $\{B_n\}_{n=1}^N$ . Based on our observation that a scene often has various blurring degrees in multiple frames in a blurred video, we can pick relatively sharp patches in blurred frames to be pseudo-sharp patches. To choose sharp patches, we propose to measure the blurring degree by a self-shift method, and those with the least blur are selected. These  $M$  pseudo-blurred and pseudo-sharp pairs  $\{\tilde{B}_{\Omega_m}, \tilde{S}_{\Omega_m}\}_{m=1}^M$  are used as the support set. For the query set, we use the whole  $V_i = \{B_n, S_n\}_{n=1}^N$  to train the reblurring and deblurring models.

In meta-testing, given any testing video  $V_i^{test}$ , we generate the support set as in the meta-training phase, not needing the ground truth. Next, we explain how to create the support set for meta-learning in Sec. 3.2 and the architecture of the reblurring model in Sec. 3.3. At last, Sec. 3.4 details the proposed reblur-deblur meta-transferring scheme for test-time adaptation.

## 3.2 Generation of the Support Set

To enable meta-learning for the test-time adaptation, we generate pseudo-blurred and pseudo-sharp pairs  $\{\tilde{B}_{\Omega_m}, \tilde{S}_{\Omega_m}\}_{m=1}^M$  from each blurred video. These pseudo-sharp patches  $\tilde{S}_{\Omega_m}$  are fed into a reblurring model to generate pseudo-blurred patches  $\tilde{B}_{\Omega_m}$ . The support set contains  $\{\tilde{B}_{\Omega_m}, \tilde{S}_{\Omega_m}\}_{m=1}^M$ .

To search for relatively sharp patches in a blurred video, we propose to measure the blurring degree by a self-shift method. Based on our observation, sharp images usually have stronger edges with larger gradients than blurred images. Fig. 3 illustrates the self-shift method, where four images are generated by shifting the input image one pixel to the right, top, top-right, and top-bottom, in four directions. We obtain a self-shift score by averaging four PSNR values calculated by the four shifted images with the input image as the reference. A higher score indicates that the shifted images are similar to their original image (Fig. 3 (a)), implying it is blurred. In comparison, a lower score represents that the input image is sharp. We randomly select  $M$  locations to crop image patches, where the relatively sharp patches are chosen with the lowest shift scores among all the collocated patches in all the input video

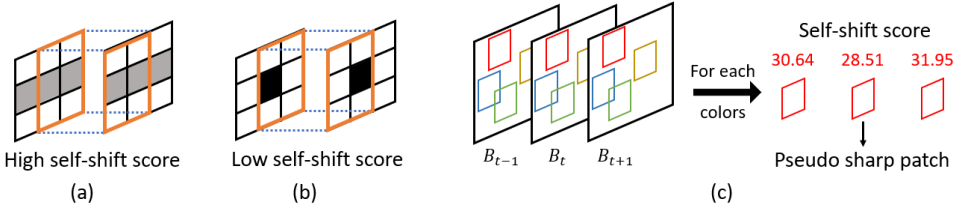


Figure 3: The proposed self-shift method. The self-shift compares the shifted image and its original input on PSNR. (a) A high self-shift score implies blur. (b) A low self-shift score implies sharpness. (c) We select pseudo-sharp patches from blurred video frames with low self-shift scores.

frames, shown in Fig. 3 (c). Finally, we feed  $\{\tilde{S}_{\Omega_m}\}_{m=1}^M$  into the reblurring model to generate pseudo-blurred patches  $\{\tilde{B}_{\Omega_m}\}_{m=1}^M$  to generate the support set.

### 3.3 Reblurring Model

Fig. 4 shows the architecture of the proposed video reblurring network. For each pseudo-sharp patch  $\tilde{S}_{\Omega_m}$ , we choose collocated patches in the five frames neighbored by the sharp patch  $\{\tilde{S}_{\Omega_{m,t-2}}, \dots, \tilde{S}_{\Omega_{m,t+2}}\}$ , where  $\tilde{S}_{\Omega_{m,t}} = \tilde{S}_{\Omega_m}$ , and  $t$  denotes the frame index, to generate the corresponding pseudo-blurred patch  $\tilde{B}_{\Omega_m}$ . In the reblurring model, the five collocated patches  $\{\tilde{S}_{\Omega_{m,t-2}}, \dots, \tilde{S}_{\Omega_{m,t+2}}\} \in \mathbb{R}^{T \times H \times W \times 3}$  are fed into an encoder to generate embedding features  $\mathbf{F} \in \mathbb{R}^{T \times H' \times W' \times C}$ , where  $T$ ,  $H$ ,  $H'$ ,  $W$ ,  $W'$ , and  $C$  denote the number of patches ( $T = 5$  here), height, embedded feature height, width, embedded feature width, and channel dimensions, respectively. Inspired by [20], we propose a cross-frame fusion block (CFFB), containing regional self-attention (RSA), cross self-attention (CSA), and local self-attention (LSA) [20].

The CFFB divides  $\mathbf{F}$  into non-overlapping regional tokens with a window size of  $r$  as  $\mathbf{F} \in \mathbb{R}^{T \times N \times D}$ , where  $N = \frac{H'W'}{r^2}$  and  $D = C \times r^2$  denote the number of tokens and dimensions. To reduce the channel dimensions, we use an embedding layer to embed  $\mathbf{F}$  into  $\mathbf{F} \in \mathbb{R}^{T \times N \times D'}$ , where  $D' = 2C$ . We then feed  $\mathbf{F}$  into RSA [20], which performs self-attention on regional tokens to generate  $\mathbf{F}^{RSA} \in \mathbb{R}^{T \times N \times D'}$ , where  $T$  is regarded as the number of batches. To utilize cross-frame information, we propose CSA, which considers the center frame patch tokens  $F_t^{RSA} \in \mathbb{R}^{N \times D'}$  from  $\mathbf{F}^{RSA} \in \{F_{t-2}^{RSA}, \dots, F_{t+2}^{RSA}\}$  as the query to perform self-attention with  $F_{t'}^{RSA} \in \mathbb{R}^{N \times D'}$ , where  $t' \in \{t-2, \dots, t+2\}$ , i.e.,

$$Q = F_t^{RSA} W^q; \quad K_{t'} = F_{t'}^{RSA} W^k; \quad V_{t'} = F_{t'}^{RSA} W^v; \\ F_{t'}^{CSA} = \mathbf{Softmax}\left(\frac{Q(K_{t'})^T}{\sqrt{D'}}\right)V_{t'}, \quad (1)$$

where  $W^q$ ,  $W^k$ , and  $W^v \in \mathbb{R}^{D' \times D'}$  denote the embedding weights of the query, key, and value. We concatenate the output features followed by a linear layer with dimensions  $\mathbb{R}^{5D' \times D'}$  to generate  $\tilde{F}_t^{CSA} \in \mathbb{R}^{N \times D'}$  as

$$\tilde{F}_t^{CSA} = \mathbf{Linear}(\mathbf{Concat}([F_{t-2}^{CSA}, F_{t-1}^{CSA}, F_t^{CSA}, F_{t+1}^{CSA}, F_{t+2}^{CSA}])). \quad (2)$$

After CSA, we replace the  $F_t^{RSA}$  in  $\mathbf{F}^{RSA}$  with  $\tilde{F}_t^{CSA}$  to obtain a new  $\mathbf{F}^{RSA} \in \mathbb{R}^{T \times N \times D'}$ , where  $\tilde{F}_t^{CSA}$  has fused the neighboring information. Lastly, we feed  $\mathbf{F}^{RSA}$  and initial input features

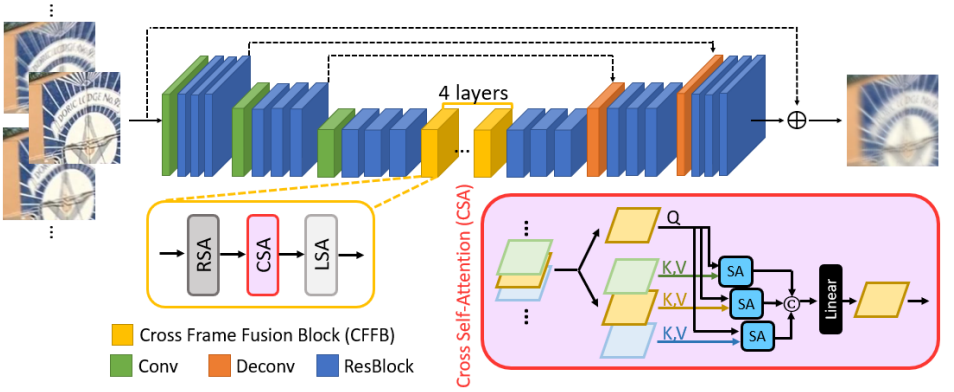


Figure 4: We use cross self-attentions to transfer the extracted blurred pattern from the collocated patches in the neighboring frames of the input pseudo-sharp patch to generate a domain-aware pseudo-blurred and sharp pair.

$\mathbf{F}$  into LSA to obtain  $\mathbf{F}^{LSA} \in \mathbb{R}^{T \times N \times D'}$  like [2]. After four CFFBs, the middle frame patch tokens  $F_t^{LSA} \in \mathbb{R}^{N \times D'}$  are fed into a linear layer to increase channel dimensions from  $D'$  to  $D$  and folded into  $F^O \in \mathbb{R}^{H' \times W' \times C}$  followed by a decoder to generate the pseudo-blurred patch  $\tilde{B}_{\Omega_m} \in \mathbb{R}^{H \times W \times 3}$ .

The reblurring model needs to be pre-trained, where we use blurred and sharp image pairs in the training set with the Charbonnier loss [13] and adversarial loss [14] derived from an additional discriminator, i.e., five consecutive sharp frames as input and the middle blurred counterpart as the ground truth. To address the domain gap of the pre-trained reblurring model, we also update the reblurring model and discriminator during testing. Therefore, we generate realistic pseudo pairs  $\{\tilde{B}_{\Omega_m}, \tilde{S}_{\Omega_m}\}_{m=1}^M$  as the support set by transferring blurred patterns at inference to enable meta-learning.

### 3.4 Reblur-Deblur Meta-Transferring

Here, we introduce the meta-learning process for the reblurring and deblurring models.

**Meta training:** We consider deblurring each video  $V_i$  in the training set as a task, where  $V_i = \{B_n, S_n\}_{n=1}^N$  are available during training. We initialize the reblurring model (parameterized by  $\theta^{RB}$ ), discriminator ( $\theta^D$ ), and deblurring model ( $\theta^{DB}$ ) with their pre-trained weights. For each video  $V_i$ , we first initialize the reblurring, discriminator, and deblurring weights  $\hat{\theta}_i^{RB}$ ,  $\hat{\theta}_i^D$ , and  $\hat{\theta}_i^{DB}$  by  $\theta^{RB}$ ,  $\theta^D$ , and  $\theta^{DB}$ , respectively. As shown in Fig. 2, the meta-process can be split into two stages: one is the inner update to obtain the reblurring and deblurring models using the support set, and the other is the outer update to evaluate the ability of adapted models by the query set.

In the inner update, we first update the reblurring model to generate the support set for the deblurring model. We choose  $M$  pseudo-sharp patches  $\{\tilde{S}_{\Omega_m}\}_{m=1}^M$  from  $\{B_n\}_{n=1}^N$  by the self-shift method to generate pseudo-blurred patches  $\{\tilde{B}_{\Omega_m}\}_{m=1}^M$  from the reblurring model  $f^{RB}(\cdot)$  parameterized by  $\hat{\theta}_i^{RB}$  as

$$\tilde{B}_{\Omega_m} = f^{RB}(\{\tilde{S}_{\Omega_{m,t-2}}, \dots, \tilde{S}_{\Omega_{m,t+2}}\}; \hat{\theta}_i^{RB}), \quad (3)$$

where the five consecutive frame patches from the pseudo-sharp patch are concatenated as

the input. To generate realistic and homogenous blurred patterns, we use the discriminator parameterized by  $\tilde{\theta}_i^D$  to calculate the adversarial loss  $\mathcal{L}_{adv}$ . Here, the pseudo-blurred patch  $\tilde{B}_{\Omega_m}$  is considered a faked blurred image, and we select a real blurred patch by the self-shift method with a high score. Additionally, we add a cycle loss based on cycle consistency from the deblurring model  $f^{DB}(\cdot)$  parameterized by  $\theta^{DB}$  as

$$\mathcal{L}_{cycle} = \mathcal{L}(f^{DB}(\tilde{B}_{\Omega_m}; \theta^{DB}), \tilde{S}_{\Omega_m}), \quad (4)$$

where  $\mathcal{L}$  is the loss function adopted in the deblurring model. It represents that deblurring the generated pseudo-blurred patch by the reblurring model should obtain its corresponding pseudo-sharp. For updating the reblurring model, we first update the discriminator  $\tilde{\theta}_i^D$  by  $\mathcal{L}_{adv}$  and then update the reblurring model by the total reblurring loss as  $\mathcal{L}_{RB} = \mathcal{L}_{adv} + \lambda \mathcal{L}_{cycle}$ , where we set  $\lambda = 0.01$  to focus more on the blur-transferring ability of the discriminator, i.e.,

$$\tilde{\theta}_i^{RB} = \tilde{\theta}_i^{RB} - \alpha \nabla_{\tilde{\theta}_i^{RB}}(\mathcal{L}_{RB}), \quad (5)$$

where we iteratively update the  $\tilde{\theta}_i^D$  and  $\tilde{\theta}_i^{RB}$  by the  $M$  pseudo pairs with the learning rate of  $\alpha$ . At last, the reblurring model ( $\tilde{\theta}_i^{RB}$ ) can effectively transfer blurred patterns extracted from the test domain to generate realistic pseudo-blurred patches.

After the adaptation of reblurring model  $\tilde{\theta}_i^{RB}$ , we generate the support set  $\{\tilde{B}_{\Omega_m}, \tilde{S}_{\Omega_m}\}_{m=1}^M$  by Eq. 3 and iteratively update the deblurring model using these  $M$  pairs as

$$\tilde{\theta}_i^{DB} = \tilde{\theta}_i^{DB} - \beta \nabla_{\tilde{\theta}_i^{DB}}(\mathcal{L}(f^{DB}(\tilde{B}_{\Omega_m}; \tilde{\theta}_i^{DB}), \tilde{S}_{\Omega_m})), \quad (6)$$

where  $\mathcal{L}$  is the loss function used by the deblurring method, and  $\beta$  is its learning rate.

In the outer update, we verify the adaptability of adapted models for each video  $V_i$ . We utilize all the blurred and sharp images  $\{B_n, S_n\}_{n=1}^N$  in  $V_i$  as the query set to perform meta-update based on the gradient calculated using the query set as

$$\begin{aligned} \theta^{RB} &= \theta^{RB} - \alpha \nabla_{\theta^{RB}} \sum_{n=1}^N \mathcal{L}_{char}(f^{RB}(S_n; \tilde{\theta}_i^{RB}), B_n), \\ \theta^{DB} &= \theta^{DB} - \beta \nabla_{\theta^{DB}} \sum_{n=1}^N \mathcal{L}(f^{DB}(B_n; \tilde{\theta}_i^{DB}), S_n), \end{aligned} \quad (7)$$

where  $L_{char}$  is the Charbonnier loss [13]. Since we regard deblurring on each video  $V_i$  as a task, the meta-training process is repeated on all the available training videos until the model converges. Through Eq. 7, we can obtain meta-learned weights  $\theta^{RB}$  and  $\theta^{DB}$ , more transferable to a new task (video) compared to the original pre-trained weights.

**Meta testing:** Given any testing video  $V_i^{test} = \{B_n\}_{n=1}^N$ , we take meta-learned reblurring ( $\theta^{RB}$ ), deblurring ( $\theta^{DB}$ ) weights and pre-trained discriminator ( $\theta^D$ ) weights as initialization. Similar to the meta-training phase, we choose  $M$  pseudo-sharp patches  $\{\tilde{S}_{\Omega_m}\}_{m=1}^M$  from  $\{B_n\}_{n=1}^N$  by the self-shift method and iteratively update the reblurring model using Eq. 5 to obtain the adapted reblurring model ( $\tilde{\theta}_i^{RB}$ ). Next, we use the reblurring model ( $\tilde{\theta}_i^{RB}$ ) to generate the support set to update the deblurring model by Eq. 6. Lastly, we deblur the whole testing video  $V_i^{test}$  by the adapted deblurring model ( $\tilde{\theta}_i^{DB}$ ). Because our support set is generated by pseudo-blurred and pseudo-sharp patches  $\{\tilde{B}_{\Omega_m}, \tilde{S}_{\Omega_m}\}_{m=1}^M$ , we can enable meta-learning on the deblurring task without the ground truth. The detailed training and testing processes are shown in the supplementary.

Table 1: Evaluation results on three datasets and four SOTA deblurring models. ‘‘Baseline’’ means the deblurring results obtained using the original models pre-trained on GoPro. ‘‘Meta’’ means the results using ‘‘Baseline’’ with our reblur-deblur meta-transferring scheme.

		DVD		REDS		RealBlur-J	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
<b>MIMOUNet+</b> [6]	Baseline	29.43	0.914	26.43	0.859	27.63	0.837
	Meta	<b>29.70</b>	<b>0.917</b>	<b>26.73</b>	0.859	<b>28.11</b>	<b>0.851</b>
<b>MPRNet</b> [32]	Baseline	29.68	0.918	26.85	0.864	28.70	0.873
	Meta	<b>30.04</b>	<b>0.921</b>	<b>27.05</b>	0.864	<b>28.75</b>	<b>0.876</b>
<b>Restormer</b> [35]	Baseline	29.67	0.916	26.93	0.867	28.96	0.879
	Meta	<b>30.01</b>	<b>0.921</b>	<b>27.21</b>	0.867	<b>29.07</b>	<b>0.885</b>
<b>CDVD-TSP</b> [21]	Baseline	30.86	0.938	27.17	0.891	28.69	0.873
	Meta	<b>30.97</b>	<b>0.939</b>	<b>27.52</b>	<b>0.892</b>	<b>29.15</b>	<b>0.887</b>

## 4 Experiments

In this section, we evaluate the effectiveness of the proposed method. We first explain datasets and implementation details in Sec. 4.1. We then demonstrate the quantitative and qualitative results in Sec. 4.2. In the end, we conduct ablation studies in Sec. 4.3.

### 4.1 Datasets and Implementation Details

We trained all the compared models on the GoPro [16] training set, containing 22 videos, totally having 2,103 training blurred and sharp image pairs in both the pre-training and meta-training stages. At the meta-testing time, we use three datasets to demonstrate the effectiveness of the proposed test-time adaptation scheme, including DVD [28] testing set (10 videos, having 1,000 images), REDS [17] validation set (30 videos, having 3,000 images), and RealBlur-J [25] testing set (50 videos, having 980 images). First, we pre-trained the reblurring model on the GoPro training set with a batch size of eight for 1,000 epochs by Adam optimizer [18], where the initial learning rate is set to  $10^{-4}$ , and the decay is  $10^{-8}$  with the cosine annealing strategy. We adopted random cropping, flipping, and rotating for data augmentation. Next, for meta-training, we randomly cropped  $256 \times 256$  patches for the support and query sets. Note that the support set contains pseudo-sharp patches selected from these cropped patches and then pseudo-blurred patches generated by the pseudo-sharps using the reblurring model. We set the learning rate  $\alpha$  to  $10^{-6}$  in the reblurring model and  $\beta$  to  $2.5 \times 10^{-6}$  in the deblurring model. All methods are tested on an Nvidia A5000 GPU.

### 4.2 Experimental Results

**Quantitative Analysis:** We employed our scheme on four different state-of-the-art deblurring models, including MIMOUNet+ [6], MPRNet [32], Restormer [35], and CDVD-TSP [21]. The first three models are for image deblurring, and the last is for video deblurring. In Table 1, we compare the results obtained with or without the proposed reblur-deblur meta-transferring scheme. As can be seen, our scheme can improve existing deblurring methods



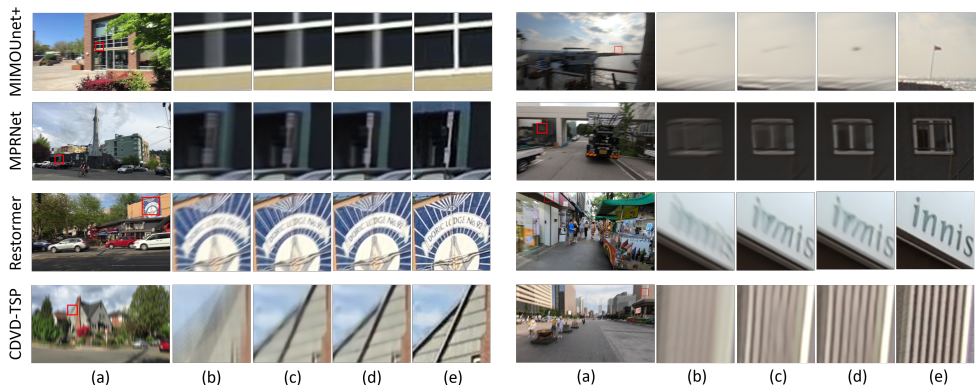


Figure 5: Qualitative comparisons of deblurring results on DVD (left) and REDS (right) datasets. (a) Blurred input. (b) A zoom-in blurred patch, and its deblurred results using (c) Baseline and (d) Baseline /w our meta-transferring scheme. (e) Ground truth.

Table 2: Comparisons among different losses used in the reblurring model on DVD dataset.

	Baseline	Cycle loss	GAN loss	GAN+Cycle loss
<b>MIMOUnet+</b>	29.43	29.58	29.62	29.70 (+0.27)
<b>MPRNet</b>	29.68	29.83	30.02	30.04 (+0.36)
<b>Restormer</b>	29.67	29.81	29.90	30.01 (+0.34)
<b>CDVD-TSP</b>	30.83	30.94	30.94	30.97 (+0.14)

by 0.27dB, 0.29dB, and 0.28dB in PSNR on average on the DVD, REDS, and RealBlur-J datasets, respectively, indicating that it achieves the test-time adaptation for better performance.

**Qualitative Analysis:** Fig. 5 shows visual comparisons of results obtained using the original state-of-the-art methods with or without using the proposed reblur-deblur meta-transferring scheme on public benchmark datasets. As shown, using our meta-transferring scheme restores the blurred content better, indicating that the proposed approach synthesizing images with homogenous blur to each task at inference time for test-time adaptation can effectively boost the deblurring performance.

### 4.3 Ablation Studies

This section analyzes the ablation studies we conducted for the proposed approach. Table 2 shows the effectiveness of different loss terms. Next, we compared meta-testing with fine-tuning and different numbers of pseudo pairs included in the support set, shown in Fig. 6.

**Analyses of the adopted losses in reblurring model:** Our approach adopts the adversarial and cycle losses to update the reblurring model, as in Eq. 5. In Table 2, we analyze the effects of using these losses for deblurring on the DVD test set. “Baseline” here denotes the performance using the original models. “Cycle loss” and “GAN loss” mean updating the reblurring model with either the cycle loss or adversarial loss used in the inner update. As observed, using either the cycle loss or GAN loss improves the performance, and adopting both losses leads to more performance gain.

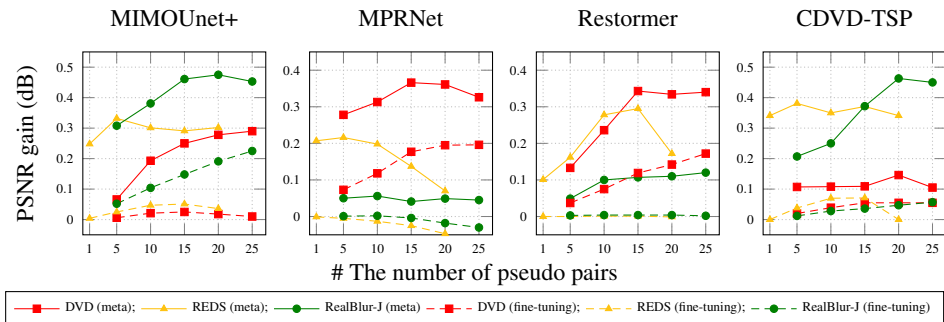


Figure 6: Ablations on meta-transferring vs. fine-tuning and analyses on different numbers of pseudo pairs included in the support set. The x-axis shows the number of pseudo pairs, and the y-axis shows the PSNR gain. Solid and dashed curves represent meta-testing and fine-tuning.

**Meta-testing and number of pseudo pairs used:** Fig. 6 compares our meta-transferring scheme with fine-tuning, which means directly fine-tuning the models with the pseudo pairs instead of meta-learning. It shows that adopting our scheme outperforms merely fine-tuning for all the state-of-the-art methods on the benchmark datasets. We can also observe that including more pseudo pairs in the support set can increase the performance gain but with diminishing marginal benefits since selecting more pseudo-sharp patches may include ones with much blur and harm the performance. In our experiment,  $M$  is set to 20 for DVD and RealBlur-J and 10 for REDS. We use more pairs in DVD [28] and RealBlur-J [25] since they have less blur for finding good pseudo-sharp patches to help test-time adaptation.

**Computation cost and limitation:** For meta-testing, each video needs to run the inner update for ten iterations on the REDS dataset and twenty on the DVD and RealBlur-J datasets. Take MIMO-UNet+ as an example. We need additional 0.1 seconds on average for inferring one image. A limitation of the proposed scheme is that when we deal with a strongly blurred video containing few relative-sharp patches, the reblur-deblur meta-transferring process would be less effective, possibly leading to poor performance.

## 5 Conclusion

This paper proposed a novel reblur-deblur meta-transferring scheme to facilitate meta-learning without ground truth at inference time. We combine a reblurring model to generate pseudo-blurred patches from selected pseudo-sharp patches as the support set, enabling meta-learning to update deblurring models during testing. Extensive experiments have shown that our method can improve existing deblurring models on benchmark datasets, including DVD, REDS, and RealBlur-J.

## Acknowledgements

This work was supported in part by National Science and Technology Council (NSTC) under grants 111-2628-E-A49-025-MY3, 109-2221-E-009-113-MY3, 110-2634-F-006-022, 110-2634-F-002-050, 111-2634-F-007-002, 111-2221-E-004-010, and 110-2622-E-004-001. This work was funded in part by Qualcomm Technologies, Inc., through a Taiwan University Research Collaboration Project, under Grant NAT-487844 and MediaTek. We thank to National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) in Taiwan for providing computational and storage resources.

## References

- [1] Dawit Mureja Argaw, Junsik Kim, Francois Rameau, and In So Kweon. Motion-blurred video interpolation and extrapolation. In *AAAI Conference on Artificial Intelligence*, 2021.
- [2] Chun-Fu Chen, Rameswar Panda, and Quanfu Fan. Regionvit: Regional-to-local attention for vision transformers. *Proc. Int'l Conf. Learning Representations*, 2022.
- [3] Huaijin Chen, Jinwei Gu, Orazio Gallo, Ming-Yu Liu, Ashok Veeraraghavan, and Jan Kautz. Reblur2deblur: Deblurring videos via self-supervised learning. In *IEEE International Conference on Computational Photography*, 2018.
- [4] Liang Chen, Faming Fang, Tingting Wang, and Guixu Zhang. Blind image deblurring with local maximum gradient prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.
- [5] Zhixiang Chi, Yang Wang, Yuanhao Yu, and Jin Tang. Test-time fast adaptation for dynamic scene deblurring via meta-auxiliary learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [6] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
- [7] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *ACM Trans. on Graphics*, 2009.
- [8] Senyou Deng, Wenqi Ren, Yanyang Yan, Tao Wang, Fenglong Song, and Xiaochun Cao. Multi-scale separable network for ultra-high-definition video deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
- [9] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proc. Int'l Conf. Machine Learning*, 2017.
- [10] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C. Courville. Improved training of wasserstein gans. In *Proc. Neural Information Processing Systems*, 2017.
- [11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *Proc. Int'l Conf. Learning Representations*, 2015.

- [12] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.
- [13] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017.
- [14] Dongxu Li, Chenchen Xu, Kaihao Zhang, Xin Yu, Yiran Zhong, Wenqi Ren, Hanna Suominen, and Hongdong Li. Arvo: Learning all-range volumetric correspondence for video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [15] Songnan Lin, Jiawei Zhang, Jinshan Pan, Zhe Jiang, Dongqing Zou, Yongtian Wang, Jing Chen, and Jimmy Ren. Learning event-driven video deblurring and interpolation. In *European Conference on Computer Vision*, 2020.
- [16] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [17] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [18] Seungjun Nah, Sanghyun Son, and Kyoung Mu Lee. Recurrent neural networks with intra-frame iterations for video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.
- [19] Seungjun Nah, Sanghyun Son, Jaerin Lee, and Kyoung Mu Lee. Clean images are hard to reblur: A new clue for deblurring. *Proc. Int'l Conf. Learning Representations*, 2022.
- [20] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Deblurring images via dark channel prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
- [21] Jinshan Pan, Haoran Bai, and Jinhui Tang. Cascaded deep video deblurring using temporal sharpness prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [22] Dongwon Park, Dong Un Kang, and Se Young Chun. Blur more to deblur better: Multi-blur2deblur for efficient video deblurring. *arXiv preprint arXiv:2012.12507*, 2020.
- [23] Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *European Conference on Computer Vision*, 2020.
- [24] Seobin Park, Jinsu Yoo, Donghyeon Cho, Jiwon Kim, and Tae Hyun Kim. Fast adaptation to super-resolution networks via meta-learning. In *European Conference on Computer Vision*, 2020.

- [25] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *European Conference on Computer Vision*, 2020.
- [26] Wang Shen, Wenbo Bao, Guangtao Zhai, Li Chen, Xiongkuo Min, and Zhiyong Gao. Blurry video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [27] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [28] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [29] Maitreya Suin and AN Rajagopalan. Gated spatio-temporal attention-guided video deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [30] Fu-Jen Tsai\*, Yan-Tsung Peng\*, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Banet: Blur-aware attention networks for dynamic scene deblurring. In *arXiv preprint arXiv:2101.07518*, 2021.
- [31] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *European Conference on Computer Vision*, 2022.
- [32] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
- [33] Yanyang Yan, Wenqi Ren, Yuanfang Guo, Rui Wang, and Xiaochun Cao. Image deblurring via extreme channels prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017.
- [34] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021.
- [35] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
- [36] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.
- [37] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

- [38] Zhihang Zhong, Ye Gao, Yinqiang Zheng, and Bo Zheng. Efficient spatio-temporal recurrent neural network for video deblurring. In *European Conference on Computer Vision*, 2020.
- [39] Shangchen Zhou, Jiawei Zhang, Jinshan Pan, Haozhe Xie, Wangmeng Zuo, and Jimmy Ren. Spatio-temporal filter adaptive network for video deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.