Blur2Blur: Blur Conversion for Unsupervised Image Deblurring on Unknown Domains

Bang-Dang Pham[1] Phong Tran[2] Anh Tran[1] Cuong Pham[1,3] Rang Nguyen[1] Minh Hoai[1,4]

[1]VinAI Research, Vietnam [2]MBZUAI, UAE [3]Posts & Telecommunications Inst. of Tech., Vietnam [4]University of Adelaide, Australia

{v.dangpb1, v.anhtt152, v.rangnhm, v.hoainm}@vinai.io cuongpv@ptit.edu.vn the.tran@mbzuai.ac.ae

Abstract

In this supplementary PDF, we first provide the qualitative results obtained by methods with the Restormer backbone [11] and some additional qualitative results of each dataset to show our effectiveness in deblurring unknown-blur images compared to other baselines. Next, we illustrate the performance with different backbones for the Blur2Blur translator model. Then, we show the impact of the dataset size and the validation of our Blur Converter through classification model. Finally, we provide details of our collected PhoneCraft dataset and validate the video deblurring performance of Blur2Blur, demonstrating significant enhancements in hand movement visualization and thus leaving room for practical application. We also include our code and a video of sample deblurring results in the supplementary package.

1. Additional Qualitative Results

1.1. Restormer model

In Fig. 3 in the main paper, we omit the results with the Restormer backbone due to the space limit. We provide these results in this supplementary in Fig. 1. As can be seen, Restormer shows behavior similar to NAFNet. The original network produces blurry images that are close to the input images. However, when combined with Blur2Blur, it can successfully deblur the images and produce sharper outputs. From quantitative numbers, Restormer-based models perform slightly worse than the NAFNet-based counterparts.

1.2. Additional Deblurring Results

In this section, we provide additional qualitative figures comparing the image deblurring results of our Blur2Blur and other baselines. Figures 2, 3, and 4 show samples where $\mathcal{K}$ is built upon the GoPro dataset [4], with the Unknown set derived respectively from the REDS dataset [5], RB2V_Street [6], and RSBlur [7].

2. Blur2Blur Analysis

2.1. Backbone Experiments

We explore the integration of multi-scale architectures into the Blur2Blur mechanism by experimenting with different backbones. The UNet architecture [8] has been adapted to handle inputs at various scales, allowing for a more nuanced understanding of blur at multiple scales. Concurrently, we employed the NAFNet backbone in its original form, taking advantage of its robust feature extraction capabilities without modifications. The result on Tab. 1 shows that MIMO-UNet clearly surpasses the standard UNet, even in its modified form. Moreover, the results also reveal that the NAFNet does not perform as well as the multi-scale variants, highlighting the importance of multi-scale level optimization in the Blur2Blur framework for deblurring tasks.

Backbone	PSNR↑	SSIM↑
UNet [8]	22.54	0.732
MIMO-UNet [2]	26.98	0.812
NAFNet [1]	20.54	0.686

Table 1. Ablation studies with the Blur2Blur backbone.

2.2. Impact of Dataset Size

In Tab. 2, we validate the deblurring performance using GoPro-RB2V datasets, maintaining a fixed $\mathcal{B}$:$\mathcal{S}$ ratio of 6:4 while varying the dataset size across four different scales of the target dataset (α).

α	0.25	0.5	0.75	1.0
PSNR	25.45	25.93	26.32	26.98

Table 2. Affect of data size

Figure 1. Qualitative results of Restormer [11] on three datasets.

2.3. Validation on the Blur Converter

To evaluate the effectiveness of our blur converter, we do two classification experiments to determine the alignment of converted images with the KnownBlur domain ($\mathcal{K}$), using GoPro-RB2V settings as detailed in the main paper. The first (**Acc1**) used the pretrained Discriminator from our Blur2Blur framework, assessing if converted images belongs to ($\mathcal{K}$). For the second (**Acc2**), we synthesized a new dataset via the Blur Kernel Extractor F [9] using sharp GoPro images combined with blur kernels from the target datasets. We then trained a ResNet18 [3] to discern if the blur in converted images corresponded to ($\mathcal{K}$) or not. In both experiments, our method converts the input image to have the target known blur with near 100% accuracy (Tab. 3).

Model	Input	Converted
Acc1/Acc2(%)	11.24 / 6.85	86.67 / 96.53

Table 3. Blur converter validation

3. Real-world Application

3.1. Details of PhoneCraft collection

The data collection process for training Blur2Blur is actually inexpensive. Although the number of images required looks high (several thousand for each subset), they are mostly video frames and thus can be collected effectively. For example, in the PhoneCraft experiment above, we only need to collect 11 sharp videos and 12 blurry ones, with a total collection time of less than 2 hours. More specifically, the dataset contains more than 12500 diverse blurry images and 11000 sharp images.

3.2. Video Deblurring Performance

As mentioned in the main paper, to enrich our practical evaluation with more tangible visual examples and to demonstrate one real-world application of our Blur2Blur mode, we incorporated a video from the collected dataset. This video simulates scenarios with significant motion blur, which is common in dynamic environments. The clarity of visual details in such situations is crucial for various applications, including rehabilitation therapy. Accurate hand movement visualization is vital for tasks like hand pose detection and gesture-based interactive rehabilitation systems.

To evaluate our Blur2Blur model, we used a video with pronounced hand movements, pre-training the deblurring model on the RSBlur dataset. The results, demonstrated in video1.mp4, clearly show that our Blur2Blur framework significantly enhances visual clarity compared to using the pre-trained deblurring model alone. Moreover, to further assess the enhancement in hand movement recognition, we validated the deblurred videos using the Hand Pose Estimation model from MediaPipe[10]. The results, shown in the video, highlight a notable improvement in hand pose estimation when using our method. The enhanced sharpness and detail achieved by Blur2Blur enable more accurate and reliable recognition of hand poses. This demonstrates the potential of our Blur2Blur model in applications demanding high-fidelity visualization of hand movements, especially in advanced rehabilitation therapy tools that rely on precise hand movement tracking for effective patient care and recovery.

Besides that, we also provide the additional qualitative video deblurring result in PhoneCraft dataset is illustrated in video2.mp4.

References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Proceedings of the European Conference on Computer Vision*, 2022.

[2] 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 International Conference on Computer Vision*, 2021.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[4] 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.

[5] Seungjun Nah, Radu Timofte, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[6] Bang-Dang Pham, Phong Tran, Anh Tran, Cuong Pham, Rang Nguyen, and Minh Hoai. Hypercut: Video sequence from a single blurry image using unsupervised ordering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023.

[7] Jaesung Rim, Geonung Kim, Jungeon Kim, Junyong Lee, Seungyong Lee, and Sunghyun Cho. Realistic blur synthesis for learning image deblurring. In *Proceedings of the European Conference on Computer Vision*, 2022.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, 2015.

[9] Phong Tran, Anh Tuan Tran, Quynh Phung, and Minh Hoai. Explore image deblurring via encoded blur kernel space. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[10] Andrey Vakunov, Chuo-Ling Chang, Fan Zhang, George Sung, Matthias Grundmann, and Valentin Bazarevsky. Mediapipe hands: On-device real-time hand tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020.

[11] 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 Conference on Computer Vision and Pattern Recognition*, 2022.

Figure 2. Extra qualitative results on the REDS dataset.

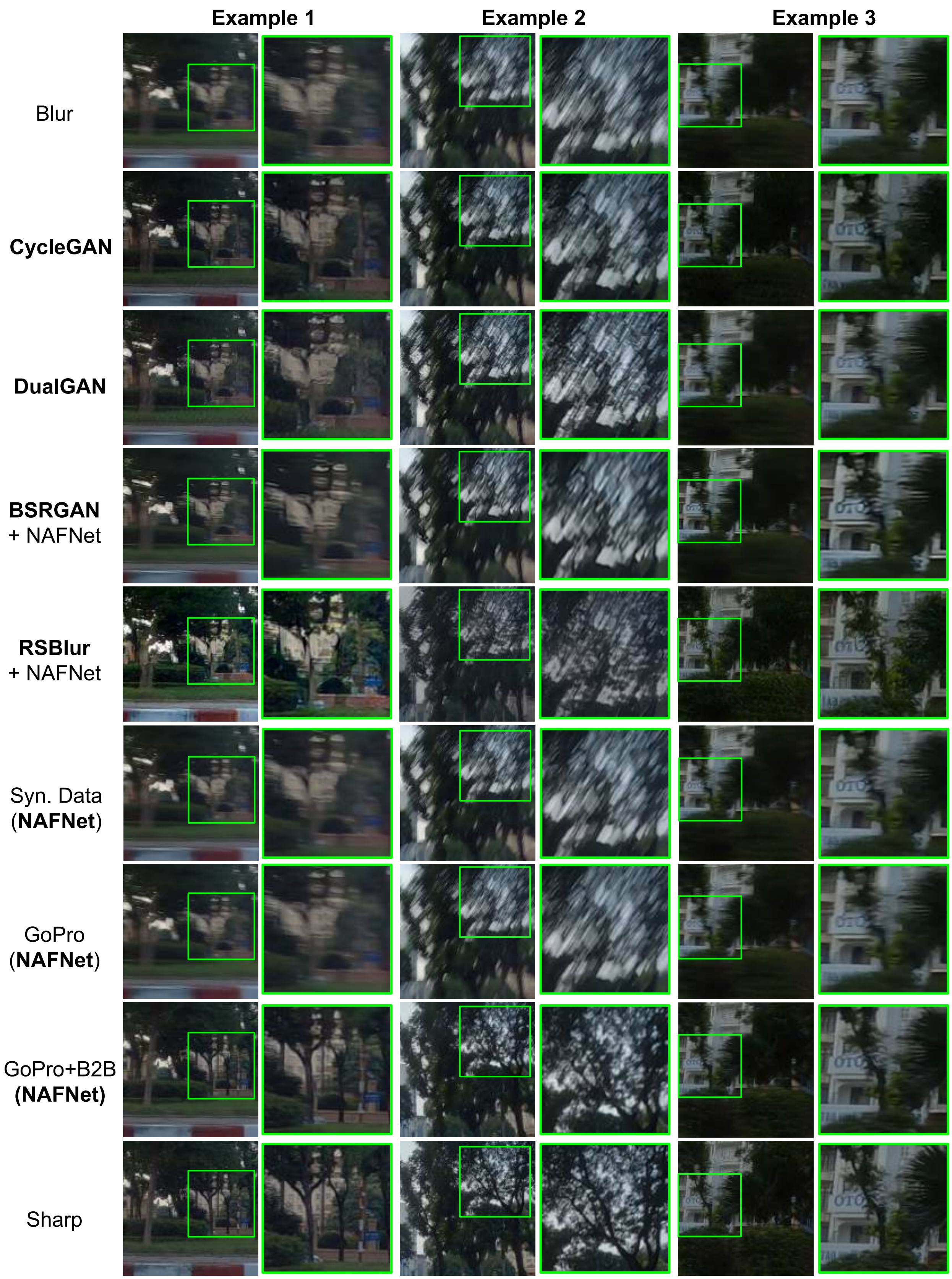

Figure 3. Extra qualitative results on the RB2V_Street dataset.

Figure 4. Extra qualitative results on the RSBlur dataset.