Parameterized Blur Kernel Prior Learning for Local Motion Deblurring

Supplementary Material

In this supplementary material, we provide more details about applying the selective scanning mechanism (S6) to vision data, then we analyze the impact of different length threshold selections, and present additional visual comparison results on test images.

1. Selective Scanning Mechanism

In the global branch of our dual-branch deblurring network, we utilize the State Space Model (SSM) to capture long-range dependencies of blurry image features. A crucial modules in SSM is the Selective Scanning Mechanism (S6). Here we provide a more detailed introduction to it.

The traditional Mamba [3] framework is designed for processing one-dimensional sequential data, which is particularly suitable for tasks involving long-sequence modeling, such as Natural Language Processing (NLP). However, when it comes to the non-sequential structure of 2D vision images, directly applying the selective scanning mechanism in Mamba is problematic due to the lack of the essential sequential arrangement in 2D data. To address this issue, VMamba [7] proposes the visual state space model backbone with linear time complexity, specifically tailored for vision tasks. Based on the approaches in [4, 7], we introduce the 2D selective scanning mechanism for processing the input image features. As illustrated in Figure 1, the 2D image feature is flattened into 1D sequences by scanning along four distinct directions, the spatial signal is converted into four sequences in parallel, then processed by the discrete state-space equation described in the main paper:

$$\begin{aligned} &h_t = \bar{\boldsymbol{C}} h_{t-1} + \bar{\boldsymbol{D}} x_t, \quad y_t = \boldsymbol{E} h_t + \boldsymbol{F} x_t, \\ &\bar{\boldsymbol{C}} = e^{\Delta \boldsymbol{C}}, \quad \bar{\boldsymbol{D}} = (\Delta \boldsymbol{C})^{-1} \left(e^{\Delta \boldsymbol{C}} - I\right) \cdot \Delta \boldsymbol{D}, \end{aligned} \tag{1}$$

By integrating information from pixels scanned in different directions, the S6 module effectively captures long-range dependencies with linear complexity. Finally, the output sequences are reshaped back into the same structure as the input features.

2. Selection of the Length Threshold

To obtain the fine-grained binary blur mask, we categorize each pixel into either the sharp or blurry region based on its kernel length parameter. Specifically, the kernel length reflects the strength of the blur kernel: a larger length indicates a severe degree of blur, while a smaller length suggests that the pixel belongs to a sharp region. Based on this observation, pixels with a kernel length below the threshold are categorized as sharp, whereas those exceeding the threshold are marked as blurry. Therefore, the length threshold plays a critical role in defining the boundary between sharp and blurry regions, directly influencing the characteristics of the resulting blur mask.

Threshold	PSNR	SSIM	PSNR$_w$	SSIM$_w$
3	36.49	0.9335	29.13	0.8852
5	**36.53**	**0.9340**	**29.15**	**0.8856**
7	36.42	0.9328	28.98	0.8837

Table 1. The results of different length thresholds for deblurring performance. The threshold of 5 demonstrated the best performance, indicating an optimal balance between capturing blurred regions and preserving the quality of sharp areas.

Generally, a **lower threshold** will generate a wider blur mask, capturing even mildly blurry areas. This ensures that all potential blurry regions are included and reduces the risk of missing subtle blurry areas, but a larger number of pixels are classified as blurry, resulting in a more inclusive blur mask. Furthermore, a lower threshold may introduce unnecessary interference in sharp regions, leading to over-processing or artifacts. Conversely, a **higher threshold** restricts the blur mask to focus on significantly blurry areas, only regions with a higher degree of blur are classified as blurry, resulting in a narrower blur mask. While this minimizes interference with sharp areas and improves processing efficiency, it risks overlooking mildly blurred regions, potentially leaving some degradations unaddressed.

Therefore, the selection of the length threshold requires careful consideration, as it directly affects the balance between coverage and precision. As illustrated in Figure 2, a lower threshold produces a mask that covers a wider area, which can lead to over-processing and artifacts in already sharp regions. On the other hand, a higher threshold restricts the mask to highly blurry areas, potentially neglecting subtle blur and resulting in incomplete restoration. To address this trade-off, we evaluated the impact of different length thresholds on deblurring performance, as shown in Table 1. Among the tested values, a threshold of 5 achieved the best performance, keeping an optimal balance between accurately capturing blurry regions and preserving the integrity of sharp areas. This choice ensures comprehensive restoration while maintaining the quality of the sharp background. **Based on these findings, we adopt a length threshold value of 5 in our experiments.** This carefully chosen threshold effectively guides the deblurring network, ensuring precise and high-quality restoration.

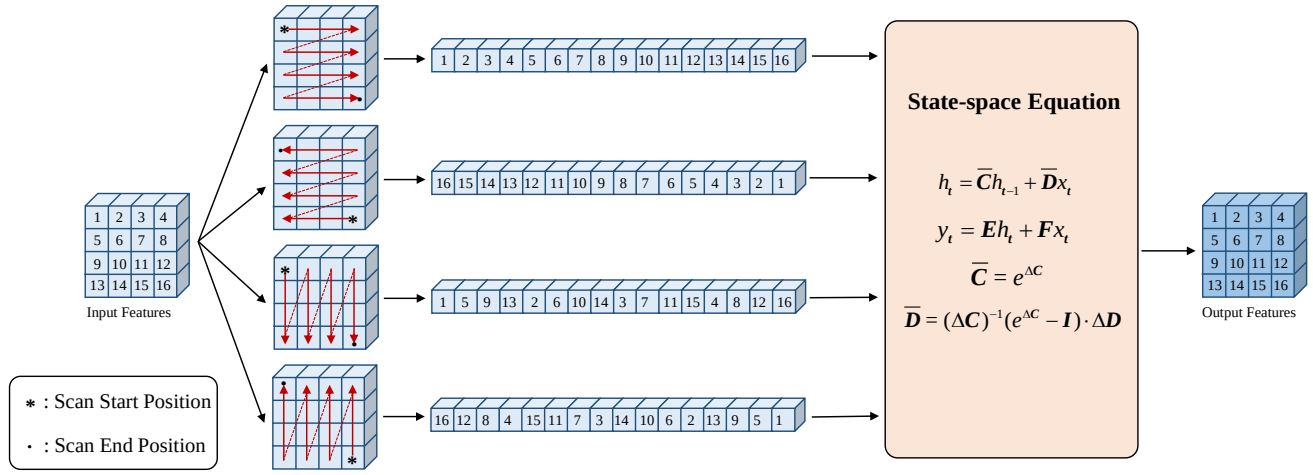
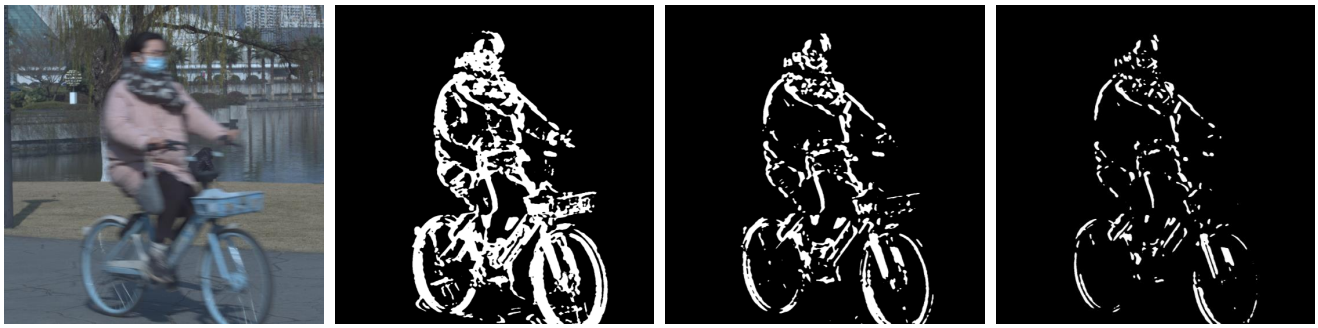


Figure 1. The illustration of the selective scanning mechanism. The input 2D image feature is flattened into 1D sequences by scanning along four distinct directions, then the converted sequences are processed by the discrete state-space equation. Finally, the output sequences are reshaped back into the same structure as the input features.



Local Blur Image | Threshold = 3 | Threshold = 5 | Threshold = 7

Figure 2. Blur masks generated by different length thresholds. The selection of the length threshold value directly affects the balance between coverage and precision: a lower threshold produces a mask that covers a wider area, while a higher threshold restricts the blur mask to focus on significantly blurry regions.

3. Additional Visual Comparison Results

We present additional visual results on the ReLoBlur testset. The comparison methods include MIMO-Unet [2], Restormer [8], NAFNet [1], LBAG+ [5] and LMD-ViT [6]. As shown in Figure 3 - 6, we can see that our method consistently delivers superior visual results, effectively restoring sharp edges and intricate details while maintaining the natural appearance of the image. Notably, our approach excels at recovering complex textures and fine structures, particularly in regions heavily affected by motion blur. In contrast, other methods often leave residual blur or introduce artifacts. Furthermore, our method preserves the sharpness of the unaffected background. These results demonstrate the effectiveness of our approach in addressing local motion blur with high visual fidelity.

The outstanding performance of our method is attributed to several key innovations. First, the proposed parameterized motion kernel modeling improves the accuracy of kernel estimation and effectively captures motion blur characteristics. The fine-grained blur mask derived from the estimated kernel length ensures targeted deblurring, focusing on blurry areas while preserving the sharpness of unblurred regions. The dual-branch network leverages Mamba for capturing global dependency and a mask-guided CNN for precise local restoration. Additionally, the integration of a shared memory bank enhances the learning of motion priors, ensuring consistency and robust representation across the network components. These advancements jointly enable sharper and artifact-free visual results, surpassing existing methods.

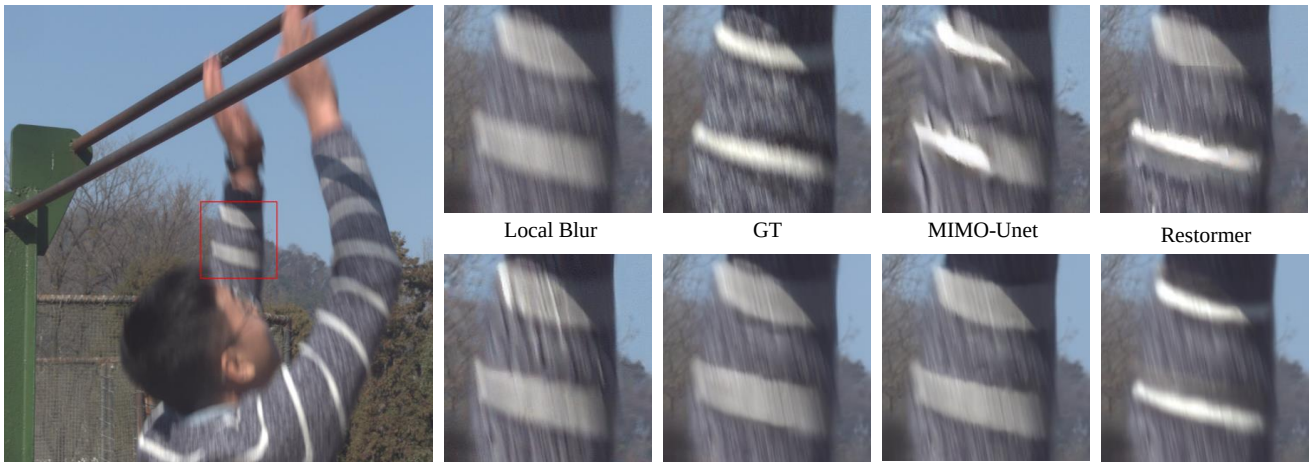


Figure 3. Visual comparisons on the ReLoBlur testset, the models are trained on the ReLoBlur dataset. From left to right: local blurry image, ground-truth, results by MIMO-Unet [2], Restormer [8], NAFNet [1], LBAG+ [5], LMD-ViT [6] and PGDN (ours).

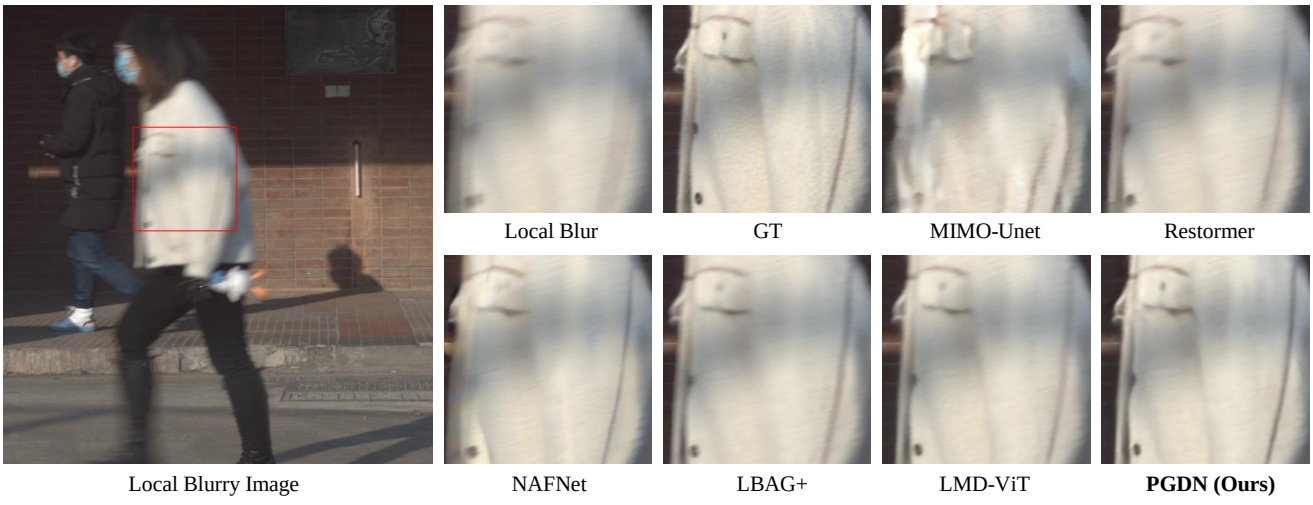


Figure 4. Visual comparisons on the ReLoBlur testset, the models are trained on the ReLoBlur dataset. From left to right: local blurry image, ground-truth, results by MIMO-Unet [2], Restormer [8], NAFNet [1], LBAG+ [5], LMD-ViT [6] and PGDN (ours).

Figure 5. Visual comparisons on the ReLoBlur testset, the models are trained on the ReLoBlur dataset. From left to right: local blurry image, ground-truth, results by MIMO-Unet [2], Restormer [8], NAFNet [1], LBAG+ [5], LMD-ViT [6] and PGDN (ours).

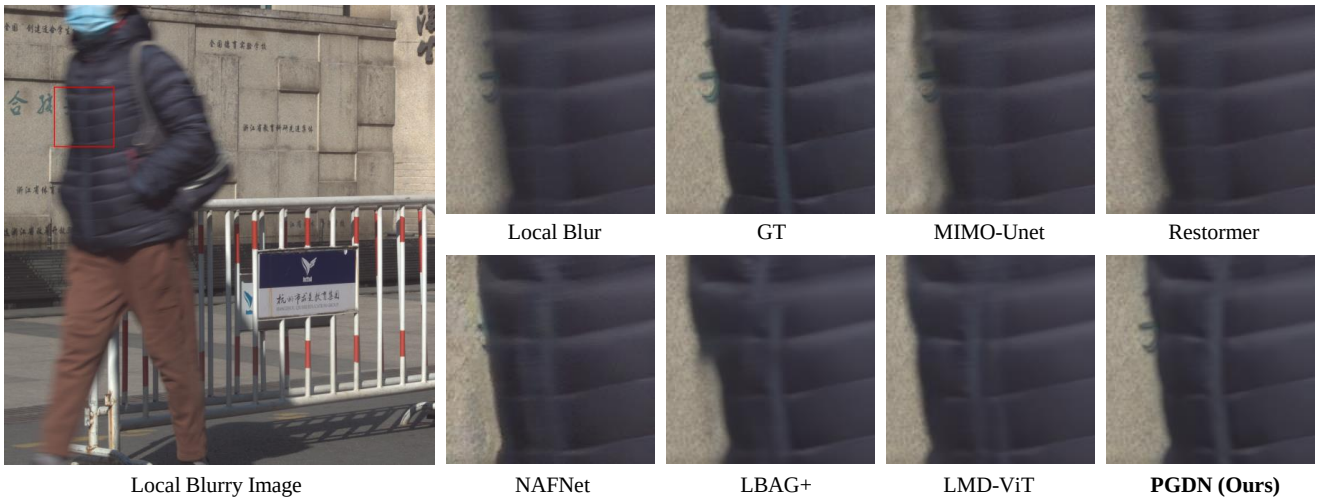


Figure 6. Visual comparisons on the ReLoBlur testset, the models are trained on the ReLoBlur dataset. From left to right: local blurry image, ground-truth, results by MIMO-Unet [2], Restormer [8], NAFNet [1], LBAG+ [5], LMD-ViT [6] and PGDN (ours).

References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 2, 3, 4

[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 IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 2, 3, 4

[3] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 1

[4] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. *arXiv preprint arXiv:2402.15648*, 2024. 1

[5] Haoying Li, Ziran Zhang, Tingting Jiang, Peng Luo, Huajun Feng, and Zhihai Xu. Real-world deep local motion deblurring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1314–1322, 2023. 2, 3, 4

[6] Haoying Li, Jixin Zhao, Shangchen Zhou, Huajun Feng, Chongyi Li, and Chen Change Loy. Adaptive window pruning for efficient local motion deblurring. In *The Twelfth International Conference on Learning Representations*, 2024. 2, 3, 4

[7] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. Vmamba: Visual state space model. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 1

[8] 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*, pages 5728–5739, 2022. 2, 3, 4