
Asymmetric Dual-Lens Video Deblurring

—Supplementary Material—

Zeyu Xiao Xinchao Wang
National University of Singapore

Overview

This supplementary document is organized as follows:

Section 1 provides more detailed ablation studies.

Section 2 offers a more qualitative results.

Section 3 provides more discussions.

1 Ablation Study

In the main text, due to page limitations, we are unable to provide a comprehensive discussion of the core module designs. Therefore, we present these results again in Table 1. A detailed analysis is provided as follows.

Table 1: Ablation of ALM, DC, and RMC variants on RealMCVSR.

Method		RealMCVSR	
		PSNR↑	SSIM↑
ALM	(a) w/o ALM	25.89	0.9145
	(b) Concat+ChannelAtt.	26.08	0.9150
	(c) $k = 1$	26.29	0.9155
	(d) $k = 3$	26.34	0.9157
	(e) $k = 5$	26.32	0.9156
	(f) DAT [1]	26.28	0.9154
	(g) MASA-SR [2]	26.19	0.9152
DC	(h) w/o DC	25.84	0.9140
	(i) Concat+ChannelAtt.	26.13	0.9149
	(j) w/o F^{Ref}	25.99	0.9144
RMC	(k) w/o RMC	25.75	0.9139
	(l) w/o Refine	25.95	0.9143
	(m) w/o Warping	26.13	0.9148
	(n) Concat+Resblock	26.20	0.9151
	(o) EDVR [4]	26.31	0.9156

To further analyze the impact of each module, we visualize the qualitative results in Figure 1, where each image corresponds to the removal of a specific component from the proposed AsLeD-Net. When the ALM module is removed, the reconstructed image exhibits noticeable detail loss and blurriness, as highlighted by the red rectangular box. This indicates that ALM is essential for refining textures using relevant reference features. The absence of the DC module results in significant structural

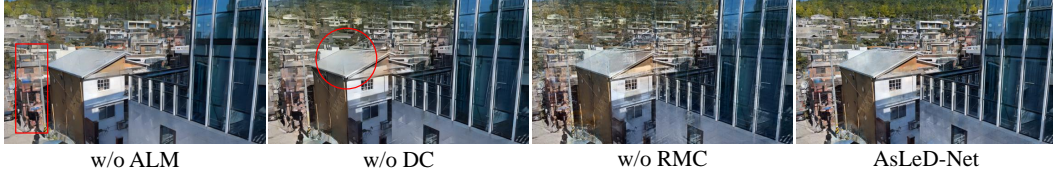


Figure 1: The results obtained by removing different core modules.

degradation, as marked by the red circle, demonstrating that DC is crucial for maintaining spatial consistency and reducing misalignment between adjacent frames. Without the RMC module, severe artifacts emerge, revealing the model’s inability to properly reconstruct temporal information, which suggests that RMC plays a pivotal role in ensuring frame-to-frame coherence and mitigating motion inconsistencies. The rightmost image, generated by the complete AsLeD-Net, exhibits the highest visual quality with sharp textures, well-preserved details, and minimal artifacts, confirming that each module contributes to the overall performance.

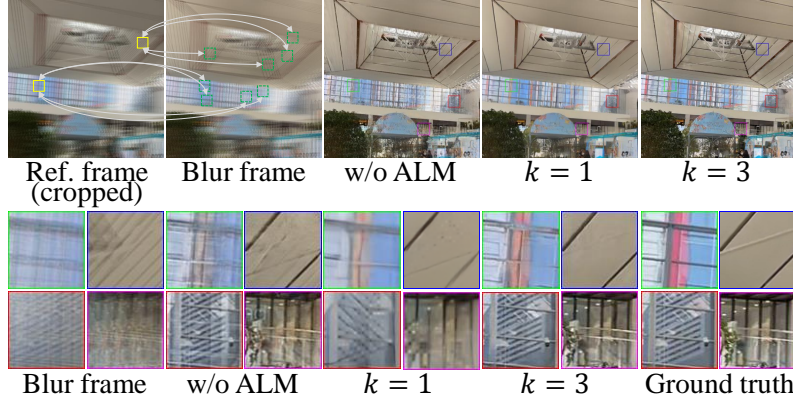


Figure 2: The results obtained by different variants of the ALM module.

Effectiveness of the ALM module. Table 1 presents an ablation study on the ALM module, which enhances structural detail transfer via K -nearest neighbor aggregation. Different design choices and variants are evaluated to assess their impact on deblurring performance. Method (a) represents the baseline without ALM, achieving a PSNR of 25.89 dB and an SSIM of 0.9145, which serves as the lower bound. Method (b) replaces ALM with a simple concatenation and channel attention mechanism, slightly improving performance to 26.08 dB PSNR and 0.9150 SSIM, demonstrating that while attention mechanisms contribute to feature refinement, they are less effective than explicit local matching. Methods (c)-(e) explore different values of K in ALM, where increasing K allows more reference features to contribute to the aggregation. Specifically, setting $K = 1$ (method (c)) achieves a PSNR of 26.29 dB and an SSIM of 0.9155. In comparison, $K = 3$ (method (d)) further improves results to 26.34 dB and 0.9157, indicating that selecting multiple nearest neighbors enhances structural detail transfer. However, increasing K to 5 (method (e)) leads to a slight drop to 26.32 dB PSNR and 0.9156 SSIM, suggesting excessive reference aggregation may introduce noise or redundant information. Methods (f) and (g) compare ALM with alternative detail aggregation strategies. Method (f) employs the DAT mechanism [1], achieving a competitive 26.28 dB PSNR and 0.9154 SSIM, slightly below the best-performing ALM configuration, suggesting that ALM provides a more effective structural matching approach. Method (g) utilizes the MASA-SR framework [2], yielding 26.19 dB PSNR and 0.9152 SSIM, which, while better than the baseline, underperforms compared to ALM-based designs, reinforcing the advantage of ALM’s targeted local matching.

Figure 2 presents the final results of several representative variants, where the visualized outcomes correspond to the quantitative results. The comparison highlights the impact of different ALM configurations on deblurring quality. The baseline without ALM exhibits noticeable structural distortions and insufficient detail restoration. Increasing K in ALM leads to progressively enhanced structural consistency and sharper details, with $K = 3$ achieving the best balance between effective detail aggregation and noise suppression. Overall, the results demonstrate that ALM significantly

enhances deblurring quality, with $K = 3$ providing the optimal trade-off between feature refinement and noise control.

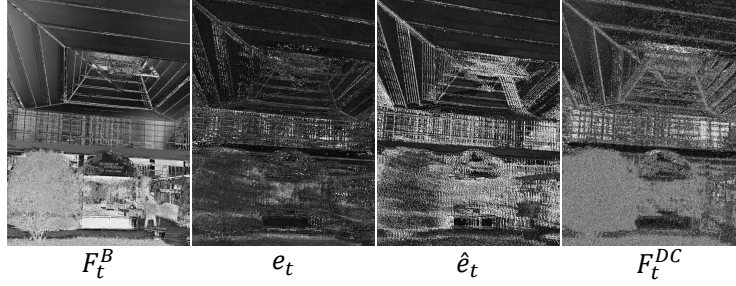


Figure 3: Feature maps in the DC module.

Effectiveness of the DC module. Table 1 presents an ablation study on the DC module, which ensures spatial consistency and reduces misalignment in the deblurring process. Different design choices and module variants are evaluated to analyze their impact on reconstruction quality. Method (h) removes the DC module entirely, resulting in a PSNR of 25.84 dB and an SSIM of 0.9140, which serves as the lower bound for this study. The absence of DC leads to inconsistencies in feature alignment, causing a degradation in structural fidelity. Method (i) replaces DC with a simple concatenation and channel attention mechanism, leading to an improved PSNR of 26.13 dB and an SSIM of 0.9149. This indicates that attention mechanisms contribute to feature enhancement but are less effective than explicit difference compensation in maintaining spatial consistency. Method (j) removes $F_t^{t,ref}$, the reference-guided compensation feature, leading to a PSNR of 25.99 dB and an SSIM of 0.9144. Compared to the entire DC module, the drop in performance confirms the crucial role of reference-guided compensation in reducing misalignment artifacts and enhancing fine details.

Figure 3 visualizes feature maps from different stages within the DC module, revealing its impact on the deblurring process. Notably, more error-related features are activated as the features are fed to the DC module, indicating enhanced feature representation. This suggests that the DC module effectively captures and refines crucial structural details regarding the error, leading to improved deblurring performance.

Effectiveness of the RMC module. Table 1 presents an ablation study on the RMC module, which enhances structural consistency through feature refinement. Removing RMC entirely (method k) results in the lowest performance, with a PSNR of 25.75 dB and an SSIM of 0.9139, establishing a lower bound. Omitting the Refine step (method l) slightly improves the results to 25.95 dB PSNR and 0.9143 SSIM, indicating that refinement contributes to better feature enhancement. Removing Warping (method m) achieves 26.13 dB PSNR and 0.9148 SSIM, suggesting that explicit warping aids in accurate feature alignment. Replacing RMC with a concatenation plus residual block structure (method n) results in 26.20 dB PSNR and 0.9151 SSIM, showing that while residual learning improves feature integration, it is slightly less effective than the complete RMC formulation. Finally, using EDVR (method o) achieves the best performance with 26.31 dB PSNR and 0.9156 SSIM, suggesting that leveraging EDVR’s temporal-spatial alignment further enhances restoration quality. Overall, the results validate the importance of RMC’s warping and refinement mechanisms, demonstrating that they play a crucial role in improving deblurring performance. While alternative strategies like concatenation and residual learning contribute to feature enhancement, they remain slightly inferior to the complete RMC module, confirming its effectiveness in structural consistency restoration.

Additionally, we visualize the optical flow and observe that after applying the refining step, the details in the optical flow become clearer, and object contours are more accurate, further confirming the effectiveness of refinement in improving structural consistency (as is shown in Figure 4). Overall, the results validate the importance of RMC’s warping and refinement mechanisms, demonstrating that they play a crucial role in improving deblurring performance.

2 Qualitative Results

In Figure 5 and Figure 6, we present additional visual results to provide a more comprehensive evaluation of our method. By comparing our approach with existing baselines across diverse test

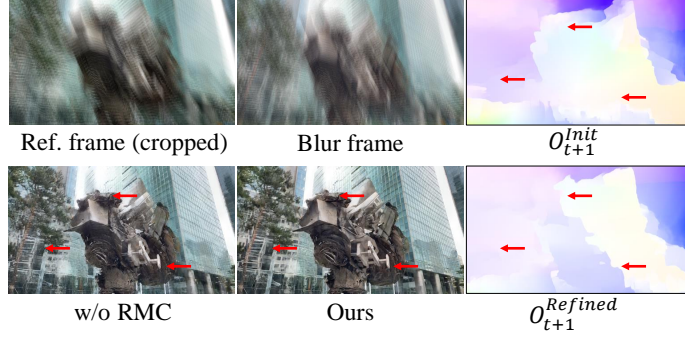


Figure 4: The results obtained by the RMC module.

cases, we observe that our method consistently produces sharper details, more accurate textures, and improved structural coherence. In particular, our model excels in handling challenging motion patterns, effectively mitigating common artifacts such as ghosting, blurring, and texture distortion. Compared to alternative methods, our approach better reconstructs fine details, especially in regions with high-frequency textures or intricate object boundaries, where competing models often produce over-smoothed or distorted results.

Figure 7 presents additional optical flow estimation results. Our method exhibits better temporal consistency, as indicated by the smoother and more coherent motion trajectories across consecutive frames.

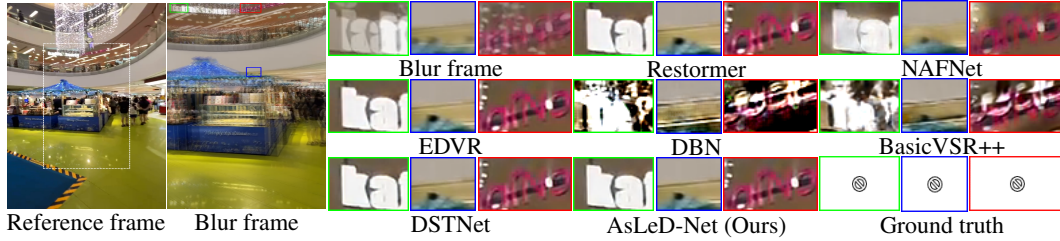


Figure 5: Qualitative comparison of AsLeD performance on real-world blurry scenes captured using an iPhone 14 Plus.

3 Limitations and Discussions

Although trained on simulated blur, our method performs strongly across a range of blur levels and real-world videos. Nevertheless, frame averaging is a coarse approximation: it cannot model exposure-dependent effects, non-uniform motion trajectories, rolling shutter, or lens/optical distortions, and may introduce ghosting under large inter-frame motion (Fig. 8). The approximation is frame-rate dependent; given the relatively low fps of RealMCVSR, averaging may under-represent realistic motion blur and reduce data diversity. In extreme conditions (e.g., fast motion, large occlusions, or low light), we observe incomplete texture recovery and occasional ghosting. Importantly, our method does not assume perfect geometric calibration or require spatial/color pre-alignment between dual-lens views; in real captures (iPhone 14 Plus) we preserve cross-view shifts and photometric differences and rely on content-aware, data-driven feature alignment, while ensuring frame-level temporal synchronization. Finally, the current model has a relatively high parameter count, increasing computational cost and complicating deployment on mobile or edge hardware; resource limits also prevented reproduction of a few memory-intensive baselines, which we view as a practical rather than methodological constraint.

We will pursue four directions. (1) Develop exposure-aware and motion-aware blur synthesis and collect paired real videos under controlled capture settings. (2) Profile and optimize efficiency via lightweight ALM/DC/RMC variants, pruning, distillation, and hardware-aware design. (3) Explore calibration-robust learning objectives and, where beneficial, optional geometric/color pre-alignment

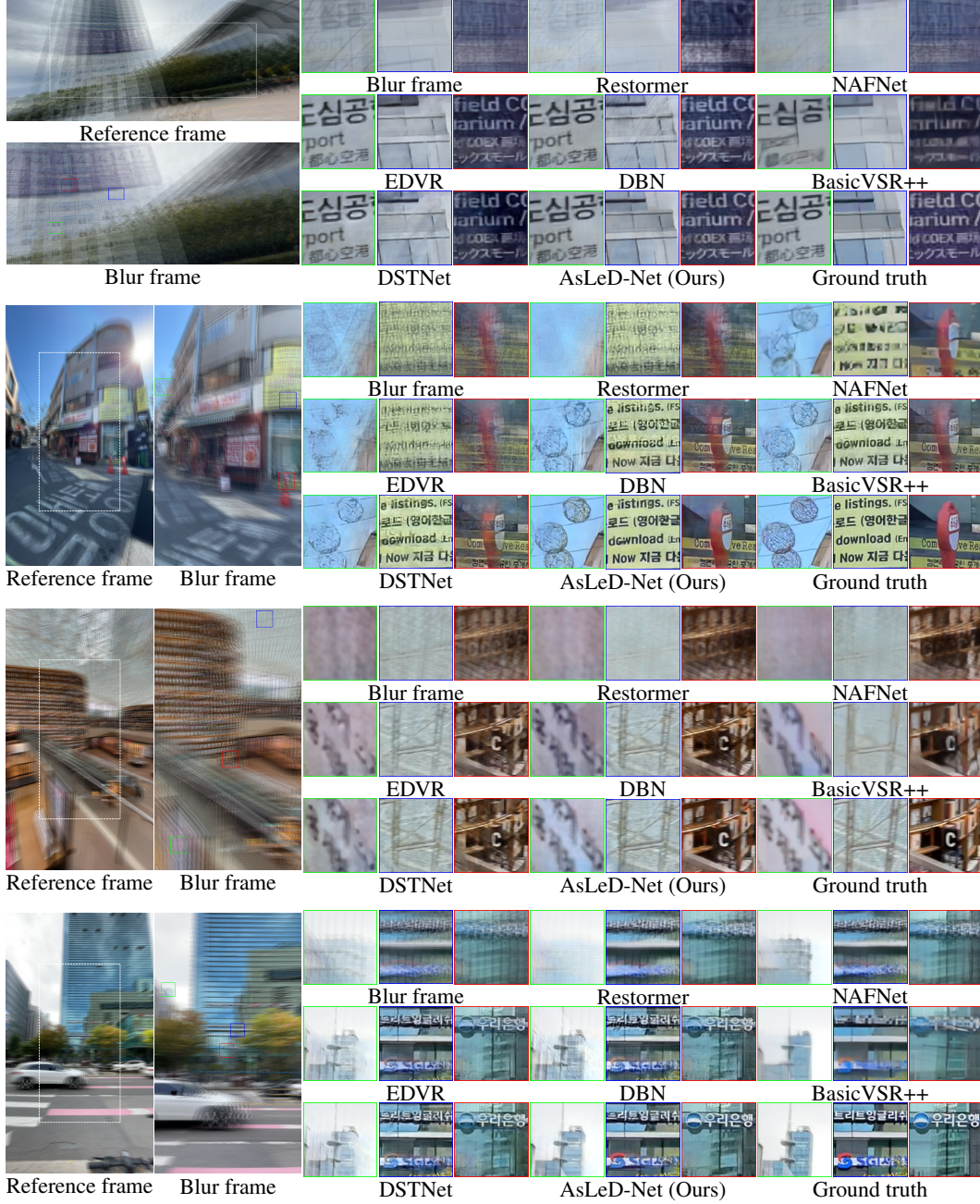


Figure 6: Qualitative comparison of AsLeD performance on the RealMCVSR dataset.

to enhance resilience. (4) Investigate adaptive inference that modulates computation by estimated blur severity, exposure, and scene dynamics, thereby narrowing the simulation-to-reality gap and enabling real-time or resource-constrained deployment. In addition, we will broaden cross-device and cross-dataset evaluation and include non-reference metrics on real captures to provide a fuller assessment.

Deblurring technologies offer broad benefits for assistive access, mobile imaging and video, telemedicine (e.g., endoscopy and handheld ultrasound), safer navigation for autonomous systems in low light or shake, and AR/VR capture and archival restoration. By improving temporal coherence and structural fidelity, systems like AsLeD-Net can reduce eye strain and provide more reliable inputs for detection, tracking, and recognition. Risks include expanded surveillance, erosion of

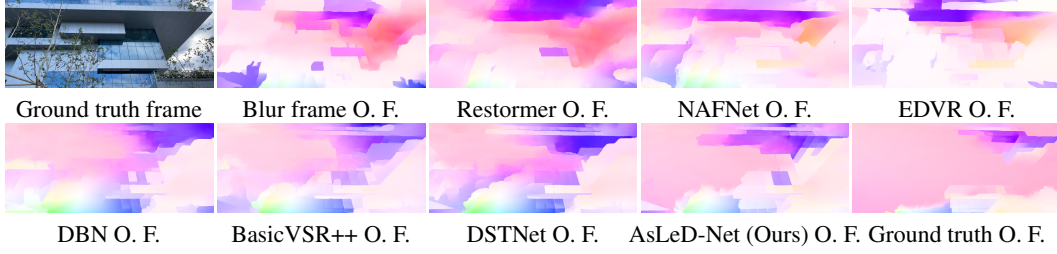


Figure 7: Temporal consistency comparison on the AsLeD task. The optical flow (O. F.) are estimated using the pre-trained RAFT [3].

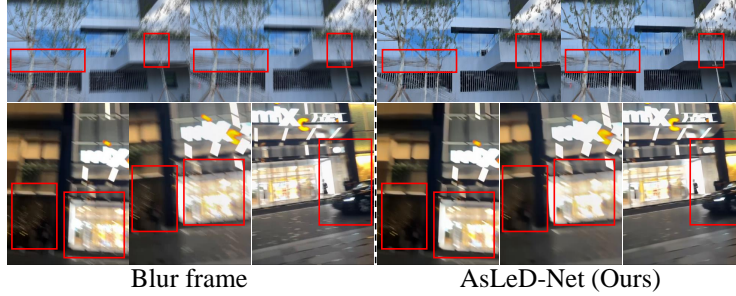


Figure 8: Failure case. From left to right in sequence are patches cropped from the ground truth image, EvTexture, and MamEVSr.

practical obscurity, unintended recovery of sensitive details, and misuse of hallucinated content; device and population shifts can yield uneven performance and bias, and in safety critical settings overreliance may create unwarranted confidence. Resource and environmental costs also matter, motivating energy tracking and efficiency through lightweight variants, pruning, distillation, and hardware aware optimization.

References

- [1] Jiezhong Cao, Jingyun Liang, Kai Zhang, Yawei Li, Yulun Zhang, Wenguan Wang, and Luc Van Gool. Reference-based image super-resolution with deformable attention transformer. In *ECCV*, 2022.
- [2] Liying Lu, Wenbo Li, Xin Tao, Jiangbo Lu, and Jiaya Jia. Masa-sr: Matching acceleration and spatial adaptation for reference-based image super-resolution. In *CVPR*, 2021.
- [3] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *ECCV*, pages 402–419, 2020.
- [4] Xintao Wang, Kelvin C. K. Chan, Ke Yu, Chao Dong, and Chen Change Loy. EDVR: video restoration with enhanced deformable convolutional networks. In *CVPRW*, 2019.