1. Dataset

To effectively train and evaluate a video deblurring model capable of generalizing to real-world conditions, we constructed a set of datasets that reflect diverse and challenging blur scenarios. Our dataset design includes both synthetic and real-world data, and distinguishes between two key sources of blur: object-induced motion and camera-induced motion. In doing so, we aim to capture the complexity and variability of real-world video blur far beyond what existing datasets offer.

We collected two categories of paired datasets. The first category—GoPro*, Sony*, and Nikon*—was created to capture real-world motion blur caused by dynamic objects and handheld camera shake. These videos were recorded using dual-camera setups, where one camera captures blurred frames at a slow shutter speed while the other captures sharp frames at a high shutter speed. This setup introduces several technical challenges. Synchronization between the two cameras is crucial, as temporal misalignment directly affects the quality of the paired data. We employed a standard cinematographic clapping technique to produce a temporal anchor point, and used professional editing tools (e.g., Adobe Premiere 2024 and DaVinci Resolve) to automatically align and trim frames with high precision.

Another major difficulty was exposure control. Because blurred frames are captured with longer exposures, they naturally admit more light, often resulting in overexposure. In contrast, the clean frames captured with short exposures can be severely underexposed, especially in low-light settings. This tradeoff is a fundamental challenge in capturing real-world blur: it often occurs under difficult lighting conditions, which cannot be easily reproduced in synthetic datasets. To mitigate this, we used shutter-priority mode to allow the camera to automatically adjust ISO and aperture, and in extreme lighting conditions, we applied manual ISO compensation or avoided capturing entirely. These practical limitations highlight the gap between synthetic blur generation methods and the challenges of real-world data collection.

A third challenge was spatial alignment. Because the dual cameras are separated by a small physical baseline, they inherently capture slightly different perspectives. We corrected for this parallax through rigid camera mounting and geometric post-processing. Specifically, we used ORB feature matching [19] to compute transformations for spatial alignment, ensuring high-quality frame pairs suitable for supervised learning.

The second category of datasets—GoProGim*, SonyGim*, and NikonGim*—was designed to capture realistic camera motion blur through intentional panning and motion trajectories. These datasets were collected using a DJI RS 4 gimbal, a popular stabilization device commonly used in social media and mobile video production. The gimbal allows us to simulate smooth but non-trivial camera movement. For the synthetic version (GoProGim*), we used high-frame-rate GoPro recordings and applied the recorded gimbal motion as a spatio-temporal impulse response to generate realistic blur. For SonyGim* and NikonGim*, we directly recorded real-world gimbal motion, capturing complex blur from both camera movement and foreground object motion. This approach allows us to build datasets that go beyond simplistic synthetic blur models and better reflect the challenges encountered in real-world video capture.

Together, these datasets form a comprehensive foundation for training and evaluating generalizable video deblurring models. They span a wide range of motion types, lighting conditions, and blur characteristics, enabling robust assessment of a model's performance in both controlled and uncontrolled environments.

2. Additional Qualitative Evaluation

Fig. 3 shows additional examples of blur images from REDs, RealBlur-R, and RSBlur datasets. The comparison results are shown in Fig. 4. These figures accompany Fig. **??**, **??** and **??** providing results for all the datasets in Table **??** and **??**. It can be observed, for example, that only the proposed method successfully deblurs the numeric characters without producing double images.

3. Additional Quantitative Evaluation

Figures 5 and 6 present frame-wise PSNR and SSIM scores across several test sequences, including both standard and gimbal-based real-world datasets. These plots offer a detailed temporal perspective of how each method performs over time, and reveal a clear advantage of our approach in terms of consistency and robustness.

Across all datasets—GoPro, BSD, Sony*, and Nikon*—the proposed method consistently achieves higher PSNR and SSIM values for nearly every frame when compared to existing baselines. Competing methods frequently exhibit performance fluctuations, with PSNR and SSIM values dropping significantly in challenging frames, such as those with fast motion or low visibility. In contrast, our model maintains stable deblurring quality from frame to frame, indicating a strong capacity to handle both subtle and complex blur patterns without degradation over time.

This consistency is even more pronounced in the gimbal-based datasets shown in Figure 6 (GoProGim*, SonyGim*, NikonGim*). These sequences include intricate motion trajectories induced by realistic camera panning and stabilization dynamics. While most baseline methods suffer severe drops in performance—especially in frames affected by overlapping camera and object motion—the proposed method demonstrates remarkable resilience, showing

(a) GoPro (b) BSD (c) Sony* (d) Nikon*

Figure 3. Examples of blur images in test dataset. Red squared regions are shown in details in Fig. 4.

steady frame-wise PSNR and SSIM levels without needing test-time adaptation or fine-tuning. This further supports the claim that our method generalizes effectively across both seen and unseen domains. We also include the results of Shift-Net, DADeblur, and Blur2Blur when they are fine-tuned with the GoProGim* dataset. The use of camera motion prior for the training of these method increased the performance, indicating the need of real-world motion prior of the GoProGim* dataset.

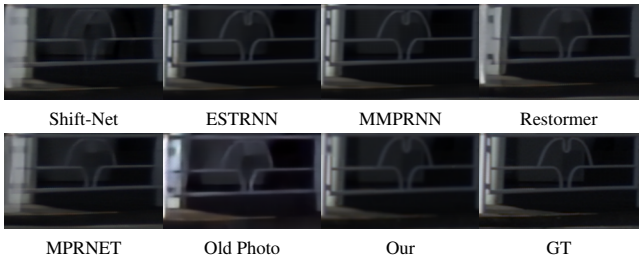
4. Ablation Study

To verify the contribution of each training component to generalization, we performed an ablation study and report the results in Tables 7 and 8. We compare three training configurations of our method: (1) using only the paired GoPro dataset, (2) using GoPro with unpaired real-world data (GoPro*, BSD), and (3) the full model trained with both paired GoPro and GoProGim* along with unpaired real-world data. The results show a clear and progressive improvement in both PSNR and SSIM as each data source is incorporated. When trained solely on paired synthetic data, performance on real-world test sets such as Sony*, Nikon*, and the gimbal-based datasets is significantly lower (e.g., 22.23 dB on Nikon* and 0.65 SSIM). Adding unpaired real blur data notably improves results across all domains, demonstrating that even without paired supervision, real-world blur plays a crucial role in bridging the domain gap. Finally, the full configuration yields the best results across all datasets—achieving 30.65 dB PSNR and 0.72 SSIM on NikonGim*, outperforming all previous variants. This confirms the effectiveness of our domain adaptation strategy and validates that incorporating unpaired real blur and motion-aware synthetic blur is essential for achieving robust, cross-domain generalization in real-world video deblurring.

5. Regarding Smoothness

The proposed method effectively restores blur caused by moving objects and camera motion. However, we observed that the results sometimes appear overly smooth. A plausible explanation is that the method predominantly restores the dominant motion blur, while less pronounced blur, such as that caused by out-of-focus regions, remains unresolved. Despite this, the structural details of objects, such as the letters on license plates, are accurately recovered. This opens the possibility of using simple post-processing techniques to enhance sharpness. In Fig. 7, we illustrate the impact of post-processing: (a) shows the results before post-processing (identical to Fig. 4 (d)), and (b) displays the results after applying unsharp masking and contrast enhancement. These post-processing techniques significantly improve sharpness, leveraging the intact structural elements in our results. In contrast, other SOTA models struggle due to broken structures. We are currently exploring ways to integrate this finding into our network architecture to enhance the sharpness

(a) GoPro dataset Video 6, frame 4

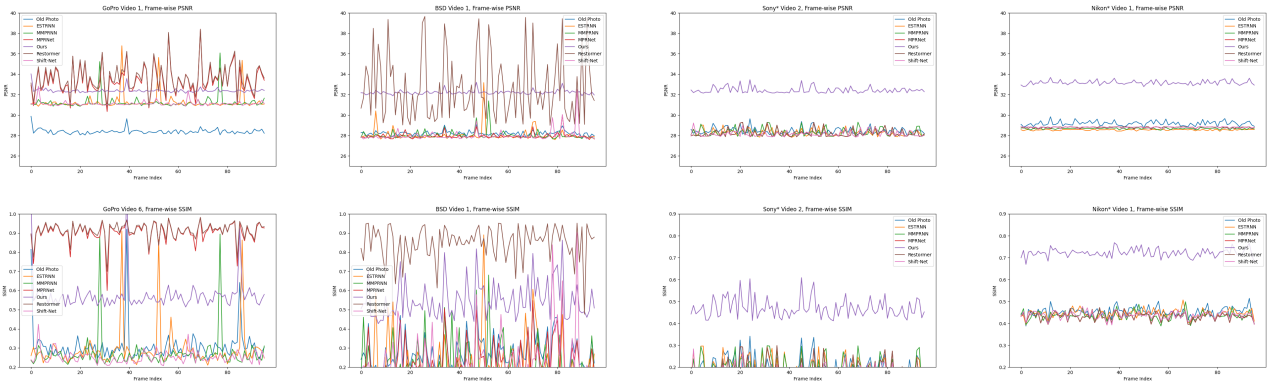


(b) BSD dataset Video 6, frame 4

(c) Sony* dataset Video 2, frame 320

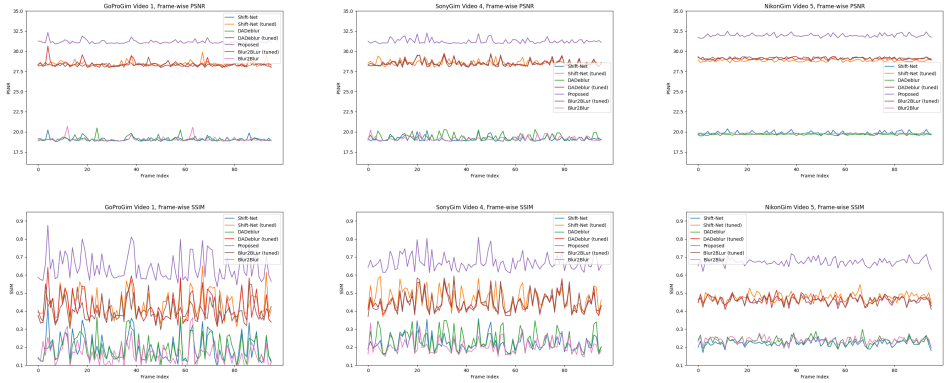
(d) Nikon* dataset Video 1, frame 18

Figure 4. Comparison of deblurring performance.



(a) GoPro Video 6 (b) BSD Video 6 (c) Sony* Video 2 (d) Nikon* Video 1

Figure 5. Frame-wise PSNR & SSIM of test set, top: PSNR, bottom: SSIM



(a) GoProGim* Video 1 (b) SonyGim* Video 4 (c) NikonGim* Video 5

Figure 6. Frame-wise PSNR & SSIM of test set with camera movement, top: PSNR, bottom: SSIM

Table 7. Ablation study: PSNR results when trained with different combinations of datasets.

method	training set		test set						
	paired	unpaired	GoPro	BSD	Sony*	Nikon*	GoProGim*	SonyGim*	NikonGim*
Proposed	GoPro		33.12	27.23	22.53	22.23			
Proposed	GoPro	BSD, GoPro*	33.15	32.76	33.07	33.14	25.25	24.56	23.97
Proposed	GoPro, GoProGim*	BSD, GoPro*	34.01	33.22	33.26	33.38	31.46	30.57	30.65

Table 8. Ablation study: SSIM results when trained with different combinations of datasets.

method	training set		test set						
	paired	unpaired	GoPro	BSD	Sony*	Nikon*	GoProGim*	SonyGim*	NikonGim*
Proposed	GoPro		0.98	0.82	0.62	0.65			
Proposed	GoPro	BSD, GoPro*	0.98	0.88	0.82	0.84	0.64	0.62	0.64
Proposed	GoPro, GoProGim*	BSD, GoPro*	0.98	0.91	0.84	0.85	0.86	0.71	0.72

None of the algorithms had previously encountered these blurred images, including our method which did not use images from the Nikon* and Sony* datasets as an unpaired dataset. Our proposed method was able to effectively restore these previously unseen blurred images, resulting in cleaner license plate numbers, lab signs, and guardrails in the restored images. While the restoration wasn't as pristine as in the synthetic GoPro case, which results indicates

(a) Before post-processing

(b) After post-processing

Figure 7. Effect of post-processing on over-smoothing, Nikon* dataset Video 1, frame 18.

that restoring real world blur is more challenging than synthetic blur. Despite the challenges of deblurring the real world blurs, our method outperformed other methods that produced blurry or broken objects.

6. Training Details

In all training phases, the Adam optimizer [4] was adapted with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ with a learning rate of 2×10^{-3} for the first 100 epochs, which has learning decay until zero. The patch size is 256×256, and the batch size is 80.