Unblur-SLAM: Dense Neural SLAM for Blurry Inputs

Supplementary Material

This supplementary material provides additional experimental results and further details on the experiments and methodology. The accompanying video further showcases experimental results.

6. Experiment Details

6.1. Ablation of the Fallback Mechanism

To demonstrate the effectiveness and necessity of our fallback mechanism under severe motion blur, we conduct an ablation study on the ReplicaBlurry dataset. This dataset simulates severe motion blur by averaging 36 Replica frames. We specifically select the sequence with the largest variation in motion blur degree for this evaluation.

As shown in Table 7, the inclusion of the fallback mechanism improves the reconstruction quality, raising the PSNR from 29.38 dB to 29.94 dB, demonstrating its critical role in robustly handling extreme blur cases.

Scenario	PSNR [dB]
With Fallback Mechanism	**29.94**
Without Fallback Mechanism	29.38

Table 7. **Ablation of the fallback mechanism.** ReplicaBlurry dataset simulates severe motion blur by averaging 36 Replica frames. We select the sequence with the largest variation in motion blur degree to demonstrate the effectiveness and necessity of our fallback mechanism.

6.2. Additional Information

In the qualitative experiments in Fig. 5, we run Deblur Gaussian SLAM [13] to compare the performance of its system with our system in TUM dataset [48]. It can be seen that, compared to methods based on multiple sub-frames, it is not precise for mild blurring and easily falls into local minima. Especially in the online setting, the rendered images inherently have a certain degree of blurriness due to insufficient observations.

The synthetic dataset ArchViz [52] in Table 1 was obtained through communication with the MBA-SLAM authors. For fair comparison and reasonable blur modeling, we use a deblurring neural network that was not trained on defocus datasets for preliminary blur removal. Meanwhile, to avoid introducing overfitting to synthetic data, we only used one sequence from the synthetic dataset. Specifically, this neural network was trained only on the Red [39], GoPro [38], and Office3 sequences (from the ReplicaBlurry [13] dataset) after manually filtering out frames containing moving objects.

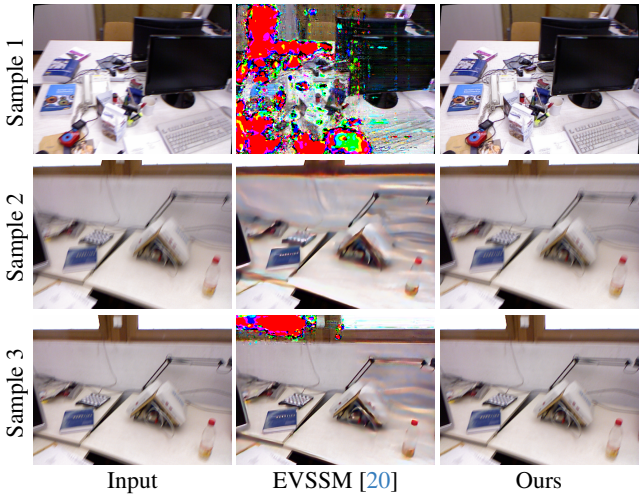

Figure 8. Comparison of input images and deblurred results

In Figure 4, we demonstrate the ATE performance on the MCD [45] s2r16 sequence to show that under our proposed training method, the single-image deblurring neural network can achieve good 3D consistency, thereby significantly improving multi-view pose estimation performance. We present the trajectory plots of Droid-SLAM [49] and our method combined with Droid-SLAM.

Table 3 shows that insufficient training data prevents the network from learning rigorous intermediate frame extraction, causing the deblurring neural network to acquire 3D-inconsistent sharpening abilities. In complex scenes, Our* augmented solely with one synthetic blur sequence cannot extract reasonable intermediate frames or less-blurred intermediate frames in more general scenarios, thus resulting in potential 3D inconsistencies.

6.3. Comparison to EVSSM

We show additional results in comparison to the recent offline method EVSSM [20]. We used the publicly available weights of EVSSM which was trained on the GoPro dataset [38]. The results are shown in Figure 8. EVSSM struggles with motion blur from moving objects along their edges, producing strange color patches and noise as camera motion and object motion occur jointly. In contrast, Unblur-SLAM decouples camera motion and object motion, and their separate processing makes our approach more robust.

7. Design Details

7.1. Thresholds:

Thresholds are an effective solution for handling various blur levels and are fixed for all experiments (Sec. 4), based

on the statistical distribution of values in the large-scale dataset (Sec. 7.3). During mapping, we employ a maximum iteration limit and robustly handle 7-12 consecutive blurry frames on the Deblur-NeRF dataset[29]. The deblurring network (Sec. 3.2) could identify the blurred frame implicitly with more computation time.

Method	PSNR [dB]	FPS
Ours_no_threshold	27.18	0.41
Ours	**29.46**	**0.74**

The table compares the average metrics to a variant without threshold on the fr1_desk, fr2_xyz and fr3_office sequences, which relies on the deblurring network for judgment.

7.2. Notion Table:

In brief, λ is the weight for the loss terms of a single frame, excluding λ_{reg} (Eq. (15)). Here is a notion overview:

Symbol	Definition	Context
x	Pixel coordinates x = (u, v) on image plane	All
f	Focal length used in defocus kernel calculation	Eq. (1)
d(t)	Camera depth at time t during exposure	Eq. (1)
σ_Δ	Standard Deviation of the Score Differences	Sec. 7.3
C, D_r	Rendered 2D RGB and depth image	Eq. (3)
d, d'	Original Gaussian depth(d) and updated tracker's depth(d')	Eq. (4)
R_k, t_k	Rotation and translation of the k-th virtual sub-frame	Sec. 3.5
θ_k, ρ_k	Lie algebra parameters for trajectory interpolation	Sec. 3.5
$w_{\text{sharp/deblur/fail/f}}$	Weights for sharp, deblurred, and failed frames	Eq. (14)-(15)
F, L_f	Union and Loss term of sharp, deblurred, and failed frames	Eq. (14)-(15)
a_{exposure}, b_{exposure}	Learnable affine Scale and Shift parameters for exposure	Eq. (8)

7.3. Blur Quantification and Recognition Details

This section provides further details on the dataset construction and metric evaluation introduced in Sec. 3.1 of the main paper.

Dataset construction. Our comprehensive blur detection benchmark incorporates real-world datasets alongside two custom semi-synthetic datasets. Following "Image as an IMU" [6], we first apply the Eden [61] video interpolation model to the ScanNet [9] dataset to generate intermediate frames. To physically model motion blur, we convert images from sRGB to linear space using inverse gamma correction, synthesize motion blur by averaging N interpolated frames, and then convert back to sRGB space:

$$I_{\text{blur}} = \gamma^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\gamma(I_{\text{linear}}^{(i)})\right) \tag{16}$$

where $\gamma(\cdot)$ and $\gamma^{-1}(\cdot)$ represent the gamma correction and its inverse. The middle interpolated frame serves as the ground truth sharp frame to ensure physical correctness.

The second semi-synthetic dataset combines sequences from the RED [39], GoPro [38], and ReplicaBlurry [13] datasets, with manual annotation to remove frames containing moving objects or inherent blur.

Blurry frames detection. To evaluate the 39 blur detection metrics, we designed a proxy measurement on a selected dataset with paired sharp and blurry images. The improvement score $\overline{\Delta}$ quantifies the magnitude of the quality difference between sharp and blurred images:

$$\overline{\Delta} = \frac{1}{M}\sum_{i=1}^{M}|s_{\text{sharp}}^{(i)} - q_{\text{blur}}^{(i)}| \tag{17}$$

where q denotes quality scores, and M is the total number of sharp/blurry pairs.

Effect size uses Cohen's d [8] to measure standardized separation: $d = \overline{\Delta}/\sigma_\Delta$. Additionally, accuracy a measures the percentage of correctly ranked blurry images. The final consistency score combines accuracy and effect size by measuring $(a \times |d|)/100$. We select the blur detection metric that achieves the highest consistency score as our blur detector, which is ARNIQA [2] (Table 8).

Metric	Acc. [%]	Effect size	Consistency score
ARNIQA-CSIQ [2]	92.93	**37.01**	**34.39**
MANIQA-KADID [56]	96.55	6.35	6.13
CLIPIQA+ [50]	96.58	2.26	2.18
MUSIQ [15]	95.13	2.14	2.03
BRISQUE [34]	79.78	1.33	1.06

Table 8. Blur detection metric. We compared the accuracy and consistency scores of 39 blur metrics and selected 5 representative ones for reporting. We chose the metric with the highest consistency score among them as our blur detector (Sec. 3.1).

8. Additional qualitative results

We conducted additional qualitative experiments on various datasets to demonstrate the robustness of our method. As shown in Figure 9, our approach exhibits improved detail recovery in various scenarios.

9. Limitations and Future Work

The physically constrained deblurring neural network training method proposed in Sec. 3.2 is unsuitable for images where the linear RGB color space cannot be accurately restored. These images typically undergo complex optimization algorithms by mobile phone pipelines, rendering them nearly impossible to reverse-engineer. Furthermore, while the proposed fallback mechanism for severe blur modeling renders multiple virtual sub-frames, the memory footprint grows explosively as the number of sub-frames increases. This memory constraint inherently limits its further expansion and scalability within online methodologies.

Nevertheless, this work proposes a reasonable engineering pipeline to implement a theoretical framework aimed at promoting progress within the community. Given the rapid

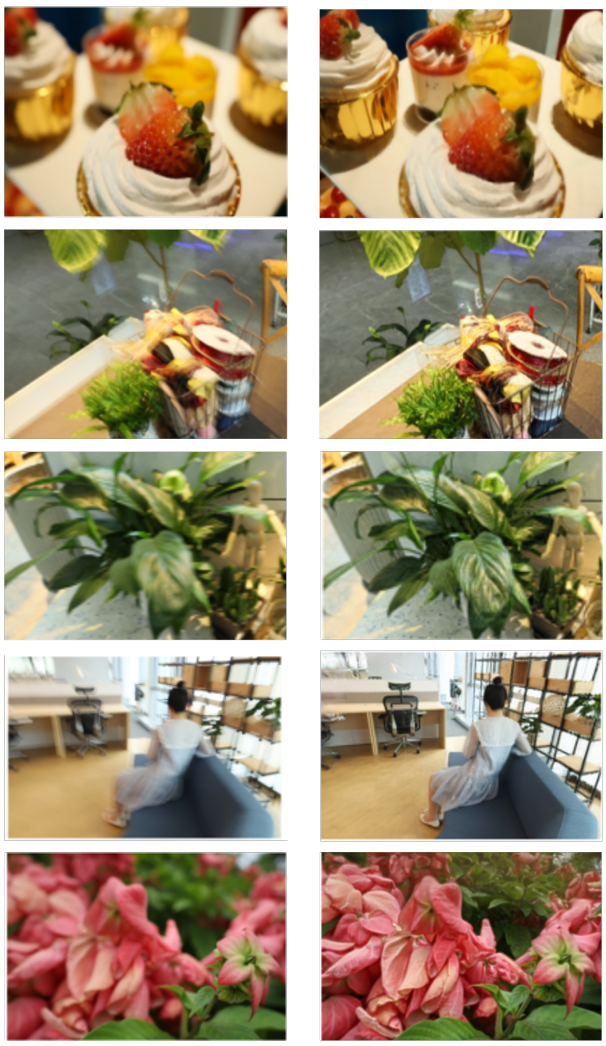

Figure 9. **Qualitative results on various datasets.** The left column shows the input images, while the right column displays the rendered images of our reconstructed 3D models.

development in the SLAM field, pure neural network-based approaches represent a highly promising direction for the further development of this theory.