LSE-NeRF: Learning Sensor Modeling Errors for Deblured Neural Radiance Fields with RGB-Event Stereo

Supplementary Material

7. BADNeRF [36] details

BADNeRF [36] uses the same blur model as Deblur-NeRF [15] in Eq. (9). Since camera positions can be inaccurate, it parameterizes camera poses as exponential maps for optimization using gradients from the RGB loss. For training, it linearly interpolates consecutive camera poses to obtain camera poses within the exposure time for rendering synthetic blurry images using the blur model. These synthetic images are, then, used to compare them with the real blurry images to calculate the RGB loss.

8. Sensitivity analysis - embedding dimension

We study the impact of the embedding dimension size in Tab. 6. As shown, the choice of the embedding dimension shows minor differences. The differences are minor but our choice of D=32 provides the best overall results.

Embedding dimension size	SSIM↑	PSNR↑	LPIPS↓
$D = 8$	0.785	24.216	0.382
$D = 16$	*0.788*	*24.504*	0.383
$D = 32$	**0.790**	*24.504*	**0.374**
$D = 64$	0.787	**24.548**	*0.376*
$D = 128$	0.785	24.500	0.386

Table 6. **Sensitivity analysis - embedding dimension –** We report quantitative metrics for embedding dimensions. The differences are minor but our choice of D=32 provides the best overall results.

9. Detailed network architecture

We use the Instant Neural Graphics Primitives (Instant-NGP) [21] backbone with its default configuration. We use 16 hashgrid levels, starting with a minimum resolution of 16×16, and with a maximum resolution of 2046×2046. We use a hashmap size of 19, with each level having two feature dimensions. We then use a Multi-Layer Perceptron (MLP) with two layers each with 64 neurons to convert hash encodings to a 16-dimensional feature, where 1 dimension represents density, and the rest is used as input to the color MLP. For the color MLP head, we use three layers, again each with 64 neurons.

10. Additional detail on data collection

When collecting our data, we record our RGB data in a 12-bit High Dynamic Range (HDR) raw image to make the best use of our RGB sensor. However, for ease of utilization and compatibility with existing non-HDR pipelines, we convert them to conventional RGB images. Specifically, to convert an HDR image in $[0, 65535]$ to a conventional RGB image in $[0, 255]$, we apply again gamma mapping, after clipping the dynamic range of the HDR sensor to its 95-th percentile to avoid focusing too much on saturated pixels, except for the 'Engineer Building' and 'House' sequences where we set it to 50-th and 70-th percentile, respectively—we use a different percentile because of the wide dynamic range of these two scenes due to shadows and the sun. For the gamma mapping, we start from the standard value of 2.4, and lower or enhance its value until the scene looks natural. We thus write:

$$I_{RGB} = 255 \times \min\left(1, \max\left(0, \left(\frac{I_{HDR}}{b}\right)^{\frac{1}{k}}\right)\right). \quad (14)$$

We provide the b and k values used for each scene in Tab. 7. We further list the number of RGB images paired with event streams.

Scenes	k	b	Num Images
Bag	2.2	37937	378
Bicycle	2.8	65487	372
Courtyard	2.2	65458	442
Dragon Max	1.0	48809	569
Engineer Building	1.8	39550	547
House	2.4	65523	408
Lab	1.0	65529	569
Grad Lounge	1.8	21169	369
Presentation Room	1.7	12512	468
Teddy Grass	1.0	38158	569

Table 7. **Dataset statistics –** We report the k and b values used to convert HDR images into RGB. We also report the number of frames associated with event streams. We set the gamma mapping value, k, by either enhancing or reducing the value manually to look natural, starting from 2.4. For the clipping value for HDR images, b, we set it to the 95-th percentile for each scene, except for 'Engineer Building' and 'House', which we set to the 50-th and 70-th percentile because of the wide dynamic range of these scenes.

11. Qualitative examples for EVIMOv2 [2]

In addition to the qualitative examples that we provide in the main paper, we show qualitative examples in Fig. 6.

As shown, our results provide the sharpest reconstructions. It is interesting to note that, while in Tab. 2 both BAD-NeRF [36] with our embeddings and Our method without embeddings provide similar results according to PSNR, using events provide qualitatively sharper reconstructions. As most of the images are without detailed textures, this difference is not as pronounced in terms of quantitative metrics.

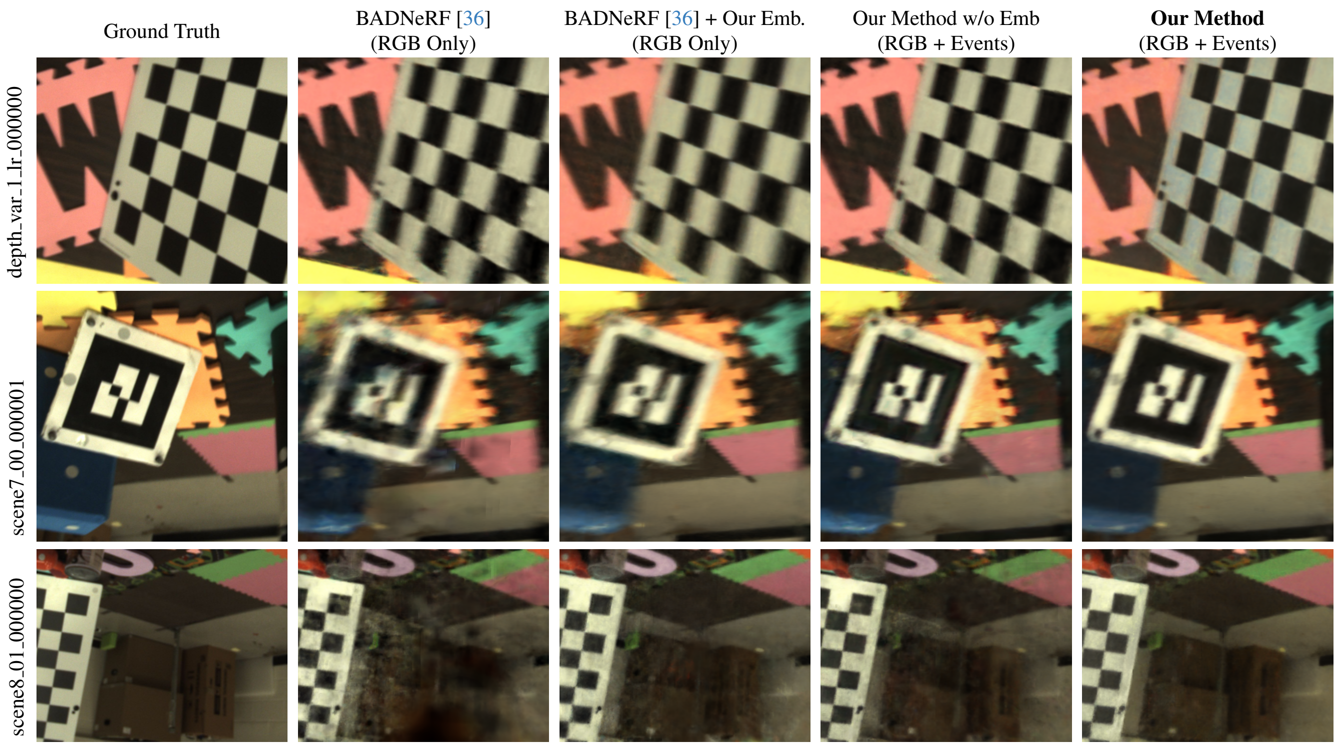

Figure 6. **EVIMOv2 [2] qualitative examples –** We show qualitative examples of zoomed-in reconstruction cutouts from EVIMOv2 [2]. As shown, our results provide the sharpest reconstructions.