Supplementary Materials of "Seeing Through Blur: Tackling Defocus in Spike-Based Imaging"

Xiantao Ma[2,*] Siwei Dong[3,*] Lin Zhu[1,2,†] Lizhi Wang[1] Hua Huang[1]
[1]Beijing Normal University [2]Beijing Institute of Technology [3]Peking University

A. Incorporating Sensor Noise into Neuron Membrane Dynamics

Spike-based representations constitute a central component of neuromorphic vision systems, as they encode scene dynamics through temporally precise, event-driven signals. A key challenge in constructing reliable spike representations lies in the hardware characteristics of spike cameras: pixel-wise response variations and fixed pattern noise (FPN) can accumulate during temporal integration, leading to inconsistencies across pixels and time.

To obtain a stable and physically meaningful spike representation, we incorporate a pixel-dependent FPN correction factor directly into the membrane update of each IF neuron. By calibrating the contribution of every incoming spike at the moment it is integrated, this neuron-level mechanism naturally compensates for sensor non-uniformities and temporal response fluctuations, yielding a more robust and noise-resilient representation without introducing additional processing stages. Furthermore, because the correction operates at the spike level, the resulting membrane dynamics remain consistent across different temporal integration lengths. This makes the formulation inherently compatible with multi–time-scale training, enabling the representation to adapt seamlessly to diverse defocus patterns and spike accumulation behaviors.

Compared with applying FPN correction after reconstruction, this spike-wise formulation acts directly on the raw event stream, ensuring that each spike is corrected prior to temporal aggregation. Such early-stage calibration preserves temporal fidelity, prevents reconstruction-induced artifacts, and ensures that the correction accumulates implicitly with spike events. Consequently, the proposed mechanism provides a physically consistent and computationally efficient way to produce stable and robust spike-based representations.

The derivation of the correction factor follows the calibration methodology introduced in [7]. For completeness, we summarize below the essential steps used to obtain the pixel-wise coefficient $\gamma(x, y)$ used in our implementation.

A.1. Uniform-Field Measurements

Following the calibration procedure of [7], uniform illumination scenes are captured at three intensity levels: dark (L_1), medium (L_2), and bright (L_3). For each illumination level, a deviation matrix is computed as:

$$D_k(x, y) = \frac{1}{I_k(x, y)} - 1, \tag{9}$$

where $I_k(x, y)$ is the normalized spike firing rate at pixel (x, y) under illumination L_k. These deviation matrices characterize the inverse-response distortion of individual pixels.

A.2. Offset and Gain Noise Components

As in [7], the pixel-wise offset noise parameter is estimated using the dark and medium fields:

$$f_{o1}(x, y) = \frac{L_2 \cdot \dfrac{D_{\text{medium}}(x, y)}{D_{\text{dark}}(x, y)} - L_1}{1 - \dfrac{D_{\text{medium}}(x, y)}{D_{\text{dark}}(x, y)}}. \tag{10}$$

Similarly, the gain noise parameter is estimated from the bright-field deviation:

$$f_{\text{test}}(x, y) = \frac{D_{\text{light}} \cdot (L_3 + f_{o1})}{D_{\text{light}}(x, y) \cdot (L_3 + f_{o1}(x, y))}. \tag{11}$$

These two terms describe the dominant components of fixed pattern noise in neuromorphic sensors, following the formulation presented in [7].

A.3. Final Correction Factor

The final correction factor used in membrane integration is therefore obtained as:

$$\gamma(x, y) = \frac{1}{f_{\text{test}}(x, y) + I \cdot \dfrac{f_{o1}(x, y)}{D_{\text{light}} \cdot (L_3 + f_{o1}(x, y))}}. \tag{12}$$

This expression serves as a pixel-wise normalization factor compensating for both offset and gain variations.

*Equal contribution.

†Corresponding author: Lin Zhu (linzhu@bnu.edu.cn).

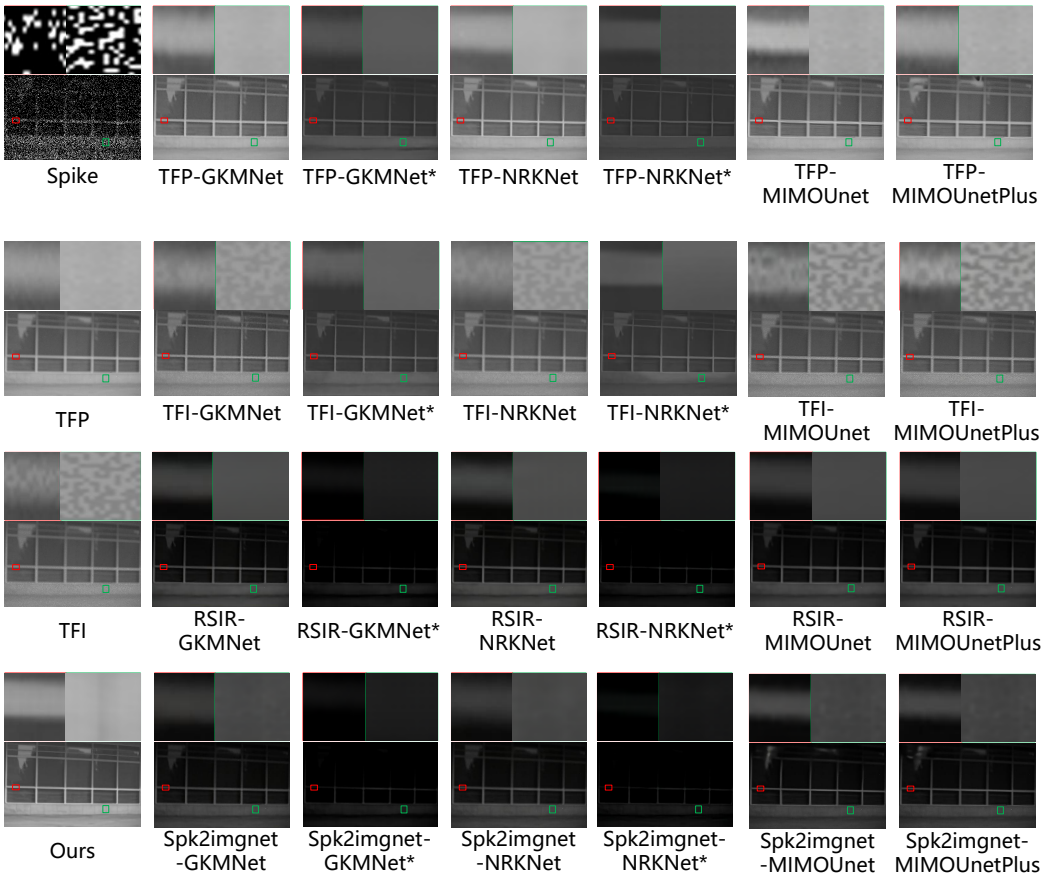

Figure 10. Visual comparison of TFI [6], TFP [6], RSIR [7], Spk2imgnet [5] followed by GKMNet [3], NRKNet [4], MIMOUnet [2], MIMOUnetPlus [2] and our method on window scene.

A.4. Simplified Form Used in the Main Paper

To keep the training pipeline lightweight and stable, the main paper adopts a simplified version of the above calibration procedure, following the ratio-based approximation in [7]. Given a response deviation matrix $R(x, y)$ measured under uniform illumination, we select a reference pixel (x_m, y_m) and compute

$$\gamma(x, y) = \frac{R(x_m, y_m)}{R(x, y)}. \tag{13}$$

This approximation retains the normalization behavior while avoiding multi-field calibration.

A.5. Practical Role in Our Model

In our framework, the correction factor $\gamma(x, y)$ is directly embedded into the membrane update of each IF neuron,

$$V(t) = V(t-1) + \gamma(x, y)\, S(t), \tag{14}$$

forming a unified integration scheme in which FPN compensation is naturally applied to every incoming spike. Because the correction operates at the level of individual spike contributions, the accumulated membrane potential inherently balances pixel-wise response variations over time without requiring additional processing steps or auxiliary calibration modules.

This spike-wise integration mechanism aligns seamlessly with the multi–time-scale training strategy adopted in our model: regardless of the temporal window length, the membrane potential aggregates an FPN-corrected stream of spikes, ensuring consistent feature formation across different integration horizons. As a result, the model benefits from a physically grounded and computationally efficient correction process that mitigates integration imbalance (Sec. 3 of the main paper) while remaining lightweight and training-compatible.

B. More Experimental Results

B.1. More Qualitative Results

We compare the deblurring and reconstruction results of our model with those of recent state-of-the-art deep reconstruction networks (Spk2ImgNet [5], RSIR [7]) that have under-

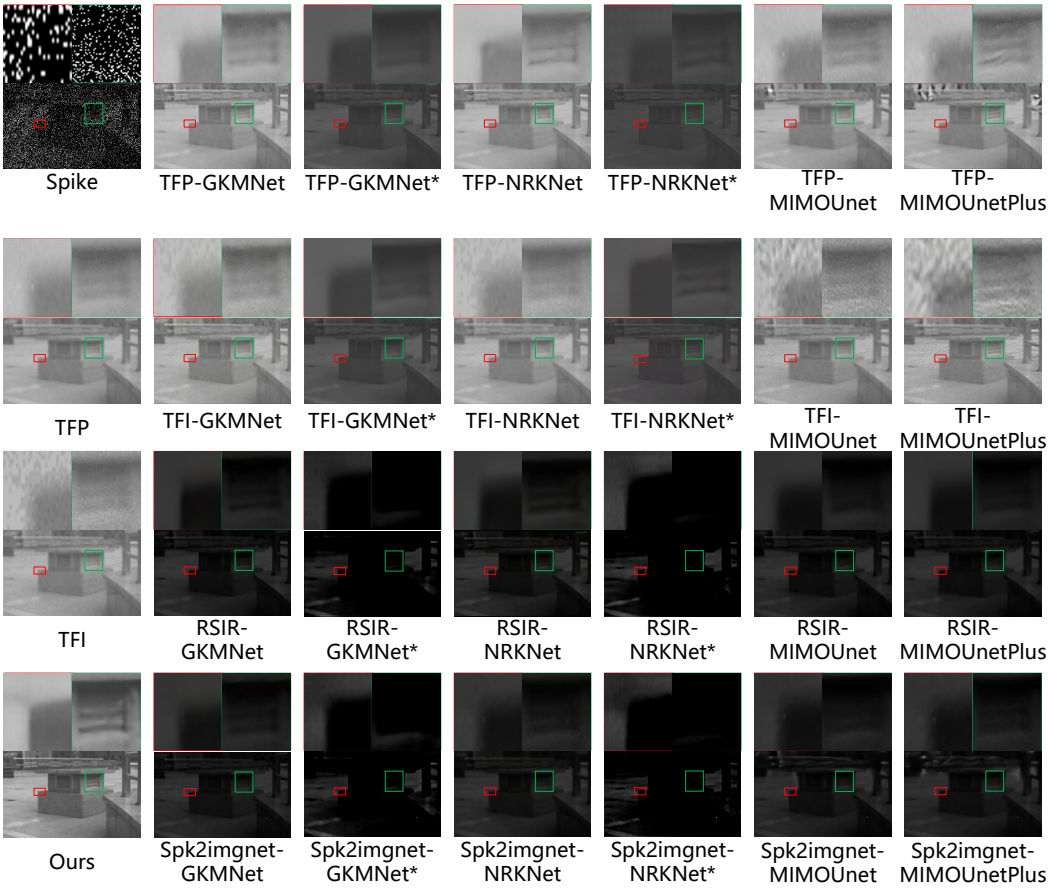

Figure 11. Visual comparison of TFI, TFP, RSIR, Spk2imgnet followed by GKMNet, NRKNet, MIMOUnet, MIMOUnetPlus and our method on the ventilation grille scene.

gone defocus deblurring processing in Fig. 10 and Fig. 11. In the Window Scene, only our method successfully reconstructs the mortar joint textures between bricks, while all other approaches fail due to defocus blur homogenization. Additionally, our method achieves the sharpest edge details. In the Ventilation Grille Scene, our model produces the clearest edge structures and richest texture details. Notably, it is the only method capable of recovering fine floor tile seams in distant regions. Our method achieves state-of-the-art defocus removal quality, conclusively demonstrating the infeasibility of the reconstruction-then-deblurring approach.

B.2. Complexity

Table 3 reports the parameter size of each method for an input spatial resolution of 400×256. Our model contains 9.498M parameters, achieves an average inference latency of 154 ms, with a peak memory consumption of 3.15 GB (an RTX 3090 GPU, 400×256), which is faster than NRKNet-representative work (0.3s, 60M parameters) in the defocus deblurring field. Overall, it remains reasonably sized for practical deployment.

Table 3. Parameters Comparison

Ours:9.50M	**Methods**			
	GKMNet	NRKNet	MIMOUnet	MIMOUnetPlus
RSIR	1.66M	6.34M	7.06M	16.35M
Spk2ImgNet	5.31M	9.99M	10.71M	20.00M

C. Dataset Details

In this section, we provide a detailed description of the datasets employed in our study. We generate our synthetic test data by performing spike simulation on the DPDD dataset, and additionally captured 75 real-world test sequences with varying scenes and depths using a spike camera.

Figure 12. Detailed description of our simulated test dataset. We construct our synthetic test dataset by generating simulated spike data with corresponding RGB labels through a pipeline incorporating cropping, symmetry, flipping, and spike simulation operations.

C.1. Simulated Data

Since there is currently no existing defocus spike dataset, we adopt a simulation approach using RGB datasets to generate the required spike data and corresponding labels. We utilize the DPDD dataset [1], which consists of aligned defocus and focus images captured by a DP camera. For the training data, we employ the entire DPDD training set as our training dataset. At each step, one defocus image and its corresponding label are randomly selected from the training set, followed by data augmentation techniques including random cropping, flipping, and symmetry operations. The cropped defocus image is then processed through a spike simulator [7] to generate simulated defocus spike data. For the simulated test data, we randomly cropped images from the DPDD test set to generate the test spike data. Fig. 12 illustrates examples of our simulated test dataset, where "Image" column displays the input blurry images, "Spike" column presents the visualized spike plane generated by the spike simulator, "Tag" column shows the corresponding ground truth labels.

C.2. Real Data

To evaluate the model's performance in real-world scenarios, we captured 75 sets of spike sequences with varying degrees of blur using a Spike Camera in diverse real-world environments. Each sequence has a resolution of 400×250 and contains objects at different depths.

Fig. 13 presents our real-world test dataset, where "Spike" shows the visualization of the spike planes, and "TFP" displays the reconstructed scene using TFP with a window length of 200 spikes, illustrating the original scenes. For each example scene, we demonstrate 3 to 9 different focal settings.

D. Limitations

Our method is designed primarily to model defocus-related behaviors in spike data, and the training setup is correspondingly tailored to typical indoor and moderately illuminated scenes. Under extremely bright or extremely dark conditions, such as the spotlight case illustrated in Fig. 14, the characteristics of the spike responses deviate substantially from the distribution observed in standard acquisition settings. Since these extreme situations fall outside the scope of the present training regime, the model may not fully capture their atypical spike patterns. Extending the framework to explicitly account for such lighting extremes, for example through dedicated data augmentation or illumination-aware modeling, represents a natural direction for future exploration.

E. Code and Dataset Availability

To support reproducibility and further research, we will release the full implementation of our method, including data preprocessing scripts, network configurations, training pipelines, and evaluation tools.

In addition, the dataset used in this work, together with the corresponding calibration files and metadata, will be made publicly available upon the publication of the paper. Links to the repository and detailed usage instructions will be provided in the camera-ready version.

References

[1] Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*, pages 111–126. Springer, 2020. 5

[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] Yuhui Quan, Zicong Wu, and Hui Ji. Gaussian kernel mixture network for single image defocus deblurring. *Advances in Neural Information Processing Systems*, 34:20812–20824, 2021. 2

[4] Yuhui Quan, Zicong Wu, and Hui Ji. Neumann network with recursive kernels for single image defocus deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5754–5763, 2023. 2

[5] Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. Spk2imgnet: Learning to reconstruct dynamic scene from continuous spike stream. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11996–12005, 2021. 2

[6] Lin Zhu, Siwei Dong, Tiejun Huang, and Yonghong Tian. A retina-inspired sampling method for visual texture reconstruction. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1432–1437. IEEE, 2019. 2

[7] Lin Zhu, Yunlong Zheng, Mengyue Geng, Lizhi Wang, and Hua Huang. Recurrent spike-based image restoration under general illumination. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 8251–8260, 2023. 1, 2, 5

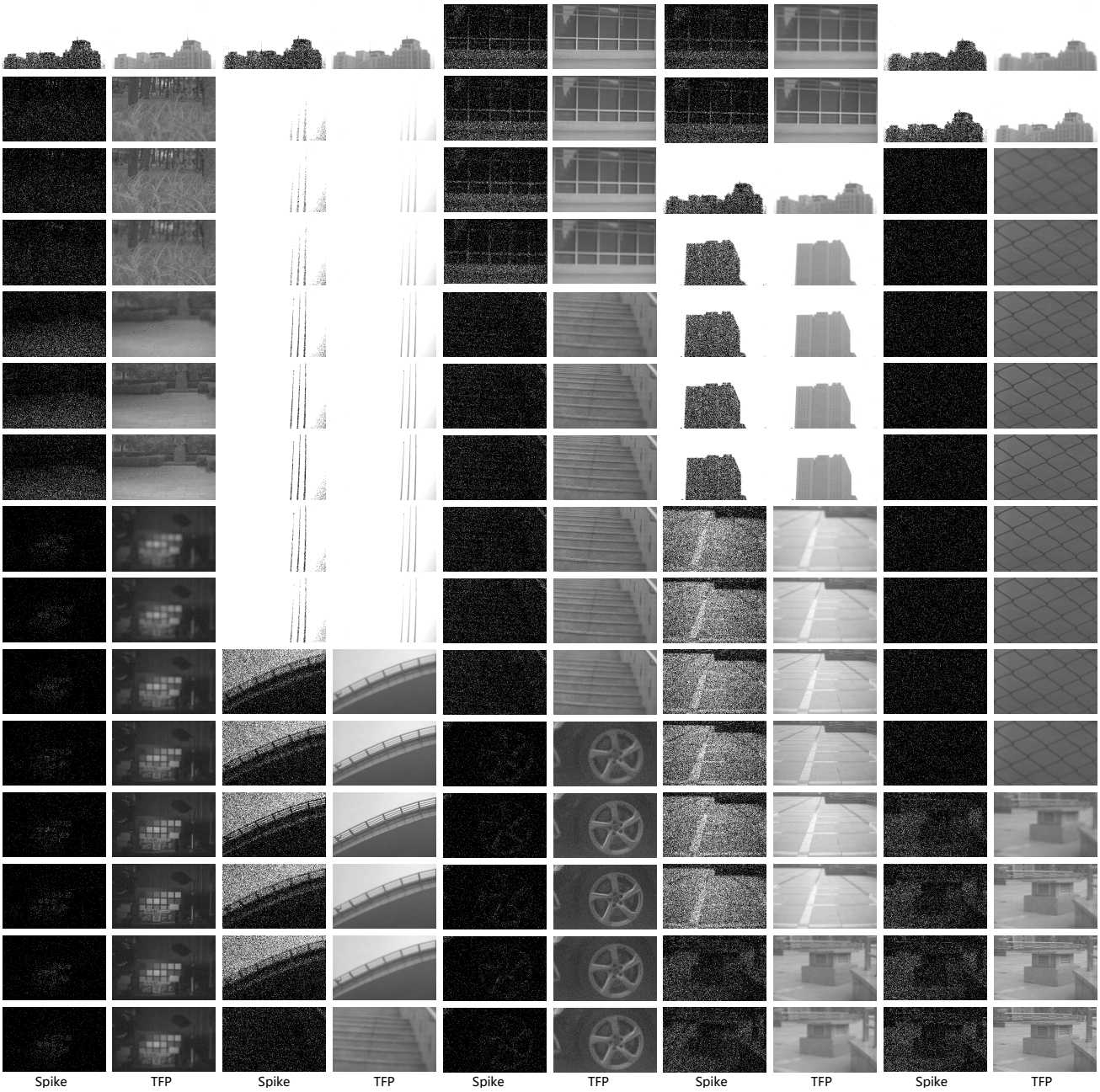

Figure 13. Detailed description of our real test dataset. For each scene, we employed 3 to 9 distinct focal lengths to capture variations in defocus characteristics.

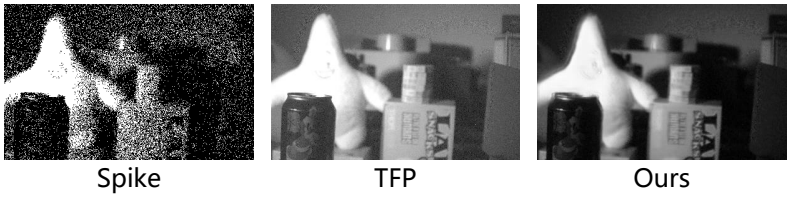

Figure 14. Failure case.