

HFR and HDR Video from Multi-Attenuated Spikes Using a Rapidly Rotating SpokeND Filter

Supplementary Material

Yakun Chang^{3,4#} Zhaojun Huang^{1,2#} Siqi Yang^{1,2,5} Yeliduosi Xiaokaiti^{1,2}
Shikui Wei^{3,4} Yao Zhao^{3,4} Tiejun Huang^{1,2} Boxin Shi^{1,2,6*}

¹ State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University

² National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

³ Institute of Information Science, Beijing Jiaotong University

⁴ Visual Intelligence +X International Cooperation Joint Laboratory of the MoE

⁵ Institute for Artificial Intelligence, Peking University, Peking University

⁶ PKU-AI² Robotics Joint Lab of Embodied AI

{ykchang, shkwei, yzhao}@bjtu.edu.cn, {huangzhaojun, yongqiye}@stu.pku.edu.cn,
{yousiki, tjhuang, shiboxin}@pku.edu.cn

In the supplementary material, we provide details of our hardware platform in Sec. 5.1, details of our simulator in Sec. 5.2, additional implementation details in Sec. 5.3, and additional qualitative results in Sec. 5.4.

5.1. Details of hardware platform

In this approach, we utilize the “Spike M1K40-H2-Gen3” spike camera, which supports a sampling rate of 20,000 Hz. The high-speed motor, “SOYG 3420 24V”, drives the filter rotation at 1800 RPM. The SpokeND filter is custom-fabricated with a 10 mm diameter, constructed by attaching multiple optical films with varying attenuation levels onto a transparent optical resin.

5.2. Details of simulator

Existing spike simulators [1, 2, 12, 13] are not designed to support multi-attenuated spike generation. To accommodate our experimental setup, we develop the multi-attenuated spike simulator tailored to our spoke-pattern attenuation model. This simulator takes video datasets as input and applies tone mapping to simulate realistic illumination in HDR scenes. To introduce multiple attenuation levels, we firstly generate the spoke-pattern mask according to our custom-designed SpokeND filter. The mask is then rotated by an angular step consistent with the actual rotational speed of the SpokeND filter to reflect temporal variation. For the incident light intensity, *i.e.*, the pixel value of the ground truth frame, is modulated by the mask at each pixel. For the spike generation, we set the sampling interval parameter to determine how many spike frames each video frame represents.

To bridge the domain gap between synthetic and real-world data, we augment the synthetic data by applying random Gaussian blur to each spoke-pattern mask, simulating

the lens defocus discussed in Section 3.2. Additionally, similar to the prior work [13], we incorporate both temporal and spatial noise simulations to enhance the realism of the synthetic spikes.

5.3. Additional implementation details

The ReGain module employs a modified U-Net [35] backbone enhanced with the self-attention mechanism. The encoder comprises a series of convolutional blocks with progressively increasing channel depth ($32 \rightarrow 64 \rightarrow 128 \rightarrow 256$), where downsampling is performed via strided convolutions. Each encoder layer consists of two convolutional blocks with Leaky ReLU activation. The bottleneck integrates a multi-head self-attention [38] mechanism to expand the receptive field, followed by two additional convolutional layers for deep feature refinement. In the decoder, skip connections are employed to facilitate high-fidelity spatial reconstruction. The ReFine module shares similar structures with ReGain module. The encoder consists of four convolutional stages with increasing channel depth ($4 \rightarrow 8 \rightarrow 16 \rightarrow 32$). Each encoder block stacks two convolutional layers with leaky ReLU activation and batch normalization [15]. The decoder mirrors the encoder in structure and includes bilinear upsampling followed by convolution, with skip connections from the encoder. The final output is produced via a 1×1 convolution followed by an upsampling layer to match the target resolution. Regarding the parameter settings in this approach, Δ_g is set to 0.5 ms, and K is empirically chosen as 4. Δ_f can be configured to any value greater than 0.5 ms, allowing the frame rate to reach up to 2000 FPS.

ReSt-Net is implemented using PyTorch [34] and trained on a single NVIDIA RTX 4090 GPU. We first train the ReGain module for 50 epochs, followed by training the ReFine module for 30 epochs. During both training stages, we use a batch size of 8 and adopt the Adam optimizer with

Equal contribution. * Corresponding author.

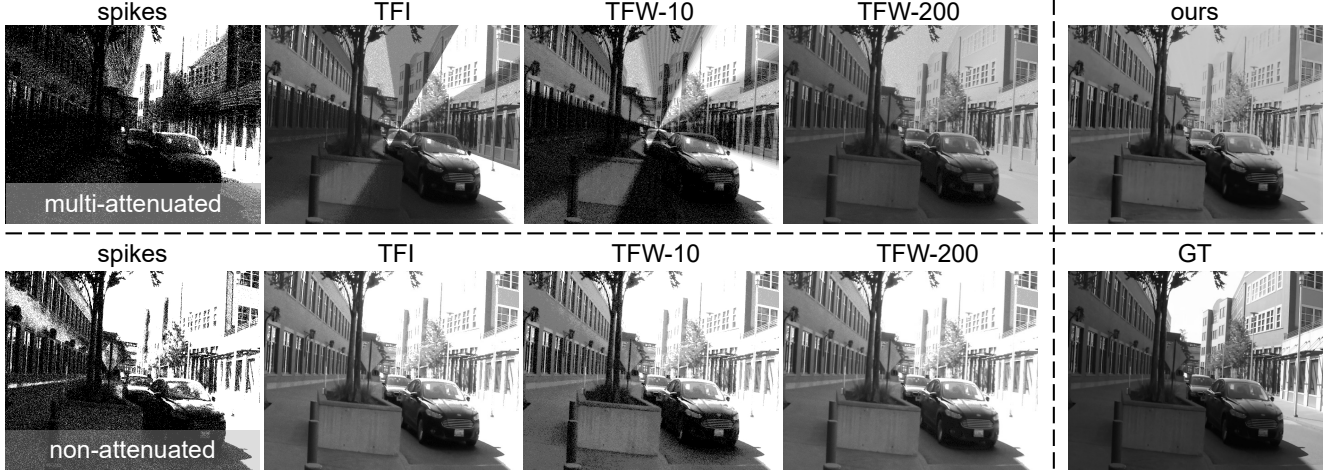


Figure 8. For Fig. 5 (a) in the main paper, we additionally simulate multi-attenuated and non-attenuated spikes to show the effectiveness of the SpokeND filter. The images reconstructed from non-attenuated spikes suffer from over-saturation.

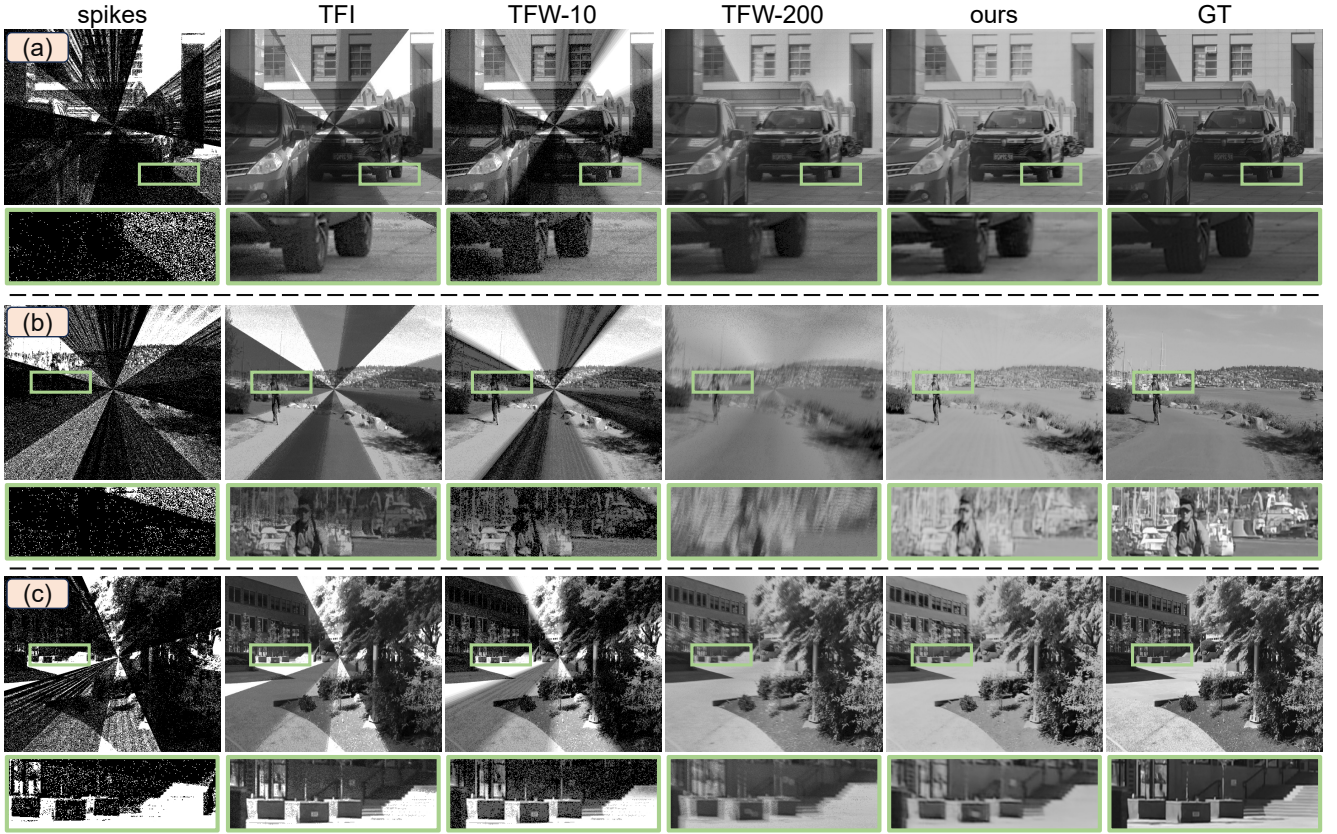


Figure 9. Additional visual equality comparison of synthetic data between the proposed method and compared methods, *i.e.*, TFI and TFW [14]. The ground truth image is tone-mapped for better visualization.

an initial learning rate of 0.0001. α_1 and α_2 are set to be 0.4 and 1.0. β_1 , β_2 , and β_3 are set to 0.4, 1.0 and 1.0, respectively.

The average inference time of ReST-Net is 21.53 ms per

frame, corresponding to an overall output rate of 46.44 FPS. In terms of model size, the ReGain module contains 19.3 million parameters (73.74 MB), while the ReFine module comprises 449K parameters (1.72 MB), demonstrating the



Figure 10. Additional visual equality comparison of synthetic data between the proposed method and compared methods, *i.e.*, TFI, TFW [14], and Spk2ImgNet [43].

efficiency of our two-stage architecture.

5.4. Additional qualitative results

Ablation study. To further validate the necessity of the SpokeND filter, we simulate HDR scenes as shown in Fig. 8. Without the SpokeND filter, the non-attenuated spikes exhibit saturated triggering in high-intensity regions. Under such conditions, the loss of information makes it impossible for all the methods to reconstruct textures in the saturated areas. More detailed ablation studies in Table 2 demonstrate effectiveness of attention module and the select of time window. Removing the attention module leads to noticeable performance degradation. Moreover, the input layer is designed to process a time window of $2K + 1$ spike frames, spanning 10 ms to cover a full attenuation cycle.

Table 2. More ablation studies.

Table A. Detailed ablation study						
Metrics	Ours($K = 9$)	w/o ReGain	w/o ReFine	w/o Attention	$K = 1$	$K = 3$
PSNR \uparrow	34.27	21.89	31.48	32.34	28.04	30.81
SSIM \uparrow	0.916	0.789	0.900	0.914	0.872	0.873

Compare with existing methods. We show more comparison results with existing methods on synthetic data in Fig. 9. As shown, our method reconstructs sharp video frames. Although TFW-200 [14] can also mitigate the spatial variation introduced by the rotating filter through temporal averaging, it tends to produce noticeably blurred results. We further demonstrate the utility of the SpokeND filter by employing standardized spike data for fair comparison with one state-of-the-art reconstruction method (*e.g.*, Spk2ImgNet [43]). As shown in Fig. 10, Spk2ImgNet [43] introduces motion-blur artifacts and noticeable noise, whereas our method produces cleaner reconstructions.

Motion limits. Aliasing and harmonic synchronization pose challenges for mechanical modulation. As illustrated in Fig. 11, we conduct experiments with a fan at 3 rotation speeds (240, 600, and 1800 RPM) to further investigate the motion limits: Our method is capable of handling a fan rotation speed of 240 RPM, where the fan rotates 1° in just 0.7 ms.

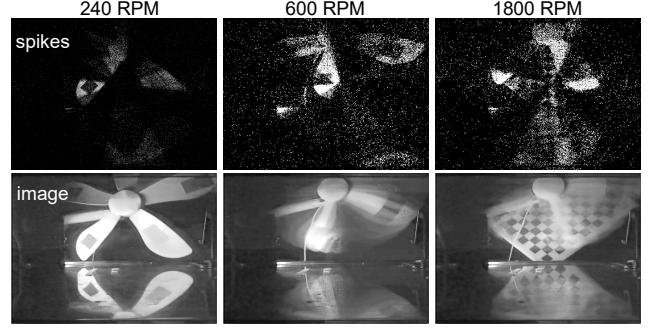


Figure 11. Test on a high-speed fan: Reconstructed results show motion blur at ultra-high rotational speeds.

References

- [1] Yakun Chang, Chu Zhou, Yuchen Hong, Liwen Hu, Chao Xu, Tiejun Huang, and Boxin Shi. 1000 FPS HDR video with a spike-rgb hybrid camera. In *Proc. of Computer Vision and Pattern Recognition*, pages 22180–22190, 2023. 3, 5, 1
- [2] Yakun Chang, Yeliduosu Xiaokaiti, Yujia Liu, Bin Fan, Zhaojun Huang, Tiejun Huang, and Boxin Shi. Towards HDR and HFR video from rolling-mixed-bit spikings. In *Proc. of Computer Vision and Pattern Recognition*, pages 25117–25127, 2024. 1, 3
- [3] Guanying Chen, Chaofeng Chen, Shi Guo, Zhetong Liang, Kwan-Yee K Wong, and Lei Zhang. HDR video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In *Proc. of International Conference on Computer Vision*, pages 2502–2511, 2021. 1, 3
- [4] Shiyan Chen, Chaoteng Duan, Zhaofei Yu, Ruiqin Xiong, and Tiejun Huang. Self-supervised mutual learning for dynamic scene reconstruction of spiking camera. page 2859–2866, 2022. 3
- [5] Shiyan Chen, Zhaofei Yu, and Tiejun Huang. Self-supervised joint dynamic scene reconstruction and optical flow estimation for spiking camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 350–358, 2023. 3
- [6] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. HDRUNet: Single image HDR reconstruction with denoising and dequantization. In *Proc. of Computer Vision and Pattern Recognition*, pages 354–363, 2021. 3
- [7] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. of ACM SIGGRAPH*, pages 1–10, 2008. 3
- [8] Sebastian Dille, Chris Careaga, and Yağız Aksoy. Intrinsic single-image HDR reconstruction. In *Proc. of European Conference on Computer Vision*, pages 161–177. Springer, 2024. 3
- [9] Yanchen Dong, Ruiqin Xiong, Jing Zhao, Jian Zhang, Xiaopeng Fan, Shuyuan Zhu, and Tiejun Huang. Joint demosaicing and denoising for spike camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 1582–1590, 2024. 3
- [10] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K

- Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics*, 36(6):1–15, 2017. 3
- [11] Yulia Gryaditskaya, Tania Pouli, Erik Reinhard, Karol Myszkowski, and Hans-Peter Seidel. Motion aware exposure bracketing for HDR video. In *Proc. of Computer Graphics Forum*, pages 119–130, 2015. 3
- [12] Liwen Hu, Rui Zhao, Ziluo Ding, Lei Ma, Boxin Shi, Ruiqin Xiong, and Tiejun Huang. Optical flow estimation for spiking camera. In *Proc. of Computer Vision and Pattern Recognition*, 2022. 1
- [13] Liwen Hu, Lei Ma, Yijia Guo, and Tiejun Huang. SCSim: A realistic spike cameras simulator. In *Proc. of International Conference on Multimedia and Expo*, 2024. 1
- [14] Tiejun Huang, Yajing Zheng, Zhaofei Yu, Rui Chen, Yuan Li, Ruiqin Xiong, Lei Ma, Junwei Zhao, Siwei Dong, Lin Zhu, et al. 1000× faster camera and machine vision with ordinary devices. *Engineering*, 2022. 1, 3, 4, 5, 7, 2
- [15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the International Conference on Machine Learning*, 2015. 1
- [16] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):144–1, 2017. 1, 3, 4
- [17] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep HDR video from sequences with alternating exposures. In *Proc. of Computer graphics forum*, pages 193–205, 2019. 3
- [18] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B Goldman, and Pradeep Sen. Patch-based high dynamic range video. *ACM Transactions on Graphics*, 32(6):202–1, 2013. 1, 4
- [19] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM Transactions on Graphics*, 22(3):319–325, 2003. 3
- [20] Joonsoo Kim, Zhe Zhu, Tien Bau, and Chenguang Liu. DCDR-UNet: Deformable convolution based detail restoration via u-shape network for single image HDR reconstruction. In *Proc. of Computer Vision and Pattern Recognition*, pages 5909–5918, 2024. 3
- [21] Phuoc-Hieu Le, Quynh Le, Rang Nguyen, and Binh-Son Hua. Single-image HDR reconstruction by multi-exposure generation. In *Proc. of Winter Conference on Applications of Computer Vision*, pages 4063–4072, 2023. 3
- [22] Byungju Lee and Byung Cheol Song. Multi-image high dynamic range algorithm using a hybrid camera. *Signal Processing: Image Communication*, 30:37–56, 2015. 3
- [23] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *Proc. of Computer Vision and Pattern Recognition*, pages 1651–1660, 2020. 3
- [24] Kede Ma, Hui Li, Hongwei Yong, Zhou Wang, Deyu Meng, and Lei Zhang. Robust multi-exposure image fusion: A structural patch decomposition approach. *IEEE Transactions on Image Processing*, 26(5):2519–2532, 2017. 3
- [25] Stephen Mangiat and Jerry Gibson. High dynamic range video with ghost removal. In *Proc. of Applications of Digital Image Processing*, pages 307–314. SPIE, 2010. 3
- [26] Stephen Mangiat and Jerry Gibson. Spatially adaptive filtering for registration artifact removal in HDR video. In *Proc. of International Conference on Image Processing*, pages 1317–1320, 2011. 3
- [27] Rafal K Mantiuk, Dounia Hammou, and Param Hanji. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625*, 2023. 8
- [28] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Proc. of Pacific Conference on Computer Graphics and Applications*, pages 382–390, 2007. 3
- [29] Srinivasa G Narasimhan and Shree K Nayar. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530, 2005. 1, 3
- [30] Manish Narwaria, Matthieu Perreira Da Silva, and Patrick Le Callet. HDR-VQM: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication*, 35:46–60, 2015. 8
- [31] Nayar and Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1168–1175. IEEE, 2003. 2, 3
- [32] Shree K Nayar and Tomoo Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *Proc. of Computer Vision and Pattern Recognition*, pages 472–479, 2000. 1, 3
- [33] Yutaro Okamoto, Masayuki Tanaka, Yusuke Monno, and Masatoshi Okutomi. Deep snapshot HDR imaging using multi-exposure color filter array. *The Visual Computer*, 40(5):3285–3301, 2024. 1, 3
- [34] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Proc. of Advances in Neural Information Processing Systems*. 2019. 1
- [35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 6, 1
- [36] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions on Graphics*, 31(6):203–1, 2012. 3
- [37] Shuo Chen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proc. of Computer Vision and Pattern Recognition*, 2017. 5
- [38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia

- Polosukhin. Attention is all you need. *Proc. of Advances in Neural Information Processing Systems*, 30, 2017. [6](#), [1](#)
- [39] Hongcheng Wang, Ramesh Raskar, and Narendra Ahuja. High dynamic range video using split aperture camera. In *IEEE 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Washington, DC, USA*. Citeseer, 2005. [2](#), [3](#)
- [40] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. [8](#)
- [41] Qingsen Yan, Lei Zhang, Yu Liu, Yu Zhu, Jinqiu Sun, Qinfeng Shi, and Yanning Zhang. Deep HDR imaging via a non-local network. *IEEE Transactions on Image Processing*, 29: 4308–4322, 2020. [3](#)
- [42] Siqi Yang, Zhaojun Huang, Yakun Chang, Bin Fan, Zhaofei Yu, and Boxin Shi. Real-data-driven 2000 FPS color video from mosaicked chromatic spikes. In *Proc. of European Conference on Computer Vision*, pages 1111–2222, 2024. [3](#)
- [43] Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream. In *Proc. of Computer Vision and Pattern Recognition*, pages 11996–12005, 2021. [3](#)
- [44] Jing Zhao, Ruiqin Xiong, Jiyu Xie, Boxin Shi, Zhaofei Yu, Wen Gao, and Tiejun Huang. Reconstructing clear image for high-speed motion scene with a retina-inspired spike camera. *IEEE Transactions on Computational Imaging*, 8:12–27, 2021. [1](#)
- [45] Jing Zhao, Ruiqin Xiong, Jian Zhang, Rui Zhao, Hangfan Liu, and Tiejun Huang. Learning to super-resolve dynamic scenes for neuromorphic spike camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 3579–3587, 2023. [3](#)
- [46] Junwei Zhao, Jianming Ye, Shiliang Shiliang, Zhaofei Yu, and Tiejun Huang. Recognizing high-speed moving objects with spike camera. In *Proc. of ACM MM*, pages 7657–7665, 2023. [3](#)
- [47] Rui Zhao, Ruiqin Xiong, Jian Zhang, Zhaofei Yu, Shuyuan Zhu, Lei Ma, and Tiejun Huang. Spike camera image reconstruction using deep spiking neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(6): 5207–5212, 2023. [1](#)
- [48] Yajing Zheng, Lingxiao Zheng, Zhaofei Yu, Boxin Shi, Yonghong Tian, and Tiejun Huang. High-speed image reconstruction through short-term plasticity for spiking cameras. In *Proc. of Computer Vision and Pattern Recognition*, pages 6358–6367, 2021. [1](#)
- [49] Lin Zhu, Siwei Dong, Tiejun Huang, and Yonghong Tian. A retina-inspired sampling method for visual texture reconstruction. In *Proc. of International Conference on Multimedia and Expo*, pages 1432–1437, 2019. [1](#), [3](#)
- [50] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. Retina-like visual image reconstruction via spiking neural model. In *Proc. of Computer Vision and Pattern Recognition*, pages 1438–1446, 2020. [1](#), [3](#)
- [51] Zhenkun Zhu, Ruiqin Xiong, Jing Zhao, Rui Zhao, Xiaopeng Fan, Shuyuan Zhu, and Tiejun Huang. High dynamic range imaging for dynamic scenes based on multi-level spike camera. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. [1](#)