Supplementary Material For "WaveE2VID: FREQUENCY-AWARE EVENT-BASED VIDEO RECONSTRUCTION"

This supplementary document complements the main paper by providing additional results to support and visualize the proposed methods. First, we present visual examples of WaveE2VID to illustrate the temporal consistency of the reconstructed videos. Next, we offer a detailed qualitative and quantitative comparison of the proposed wavelet-guided (WG) models against their counterparts across various event-camera datasets.

1. EVALUATION ON WAVEE2VID

In Figure 1and Figure 2, we present two sequences from ECD. The visual results demonstrate that the proposed WaveE2VID can effectively reconstruct videos with notable temporal consistency, preserving smooth transitions and coherence across consecutive frames.

Figure 3, presents visual results from a high-speed and HDR dataset [1], comparing our model with three state-of-the-art methods: FireNet [2], E2VID [3], and HyperE2VID [4]. The results demonstrate that our model effectively reconstructs scene details at higher sensor resolutions, outperforming larger models

2. TESTING DATASETS

Here, for the evaluation of our proposed SPADE-WG and E2VID-WG, we used three event camera datasets. Details of each dataset are provided below.

2.1 HQF Dataset:

The high-quality frames (HQF) dataset [5] is a comprehensive event camera dataset designed to facilitate the evaluation of event-based video reconstruction methods and related computer vision tasks. It includes 14 diverse image sequences, encompassing both indoor and outdoor scenarios, with a wide range of motions and environmental conditions. The data was captured using two DAVIS240C cameras, each characterized by distinct noise and contrast threshold settings, ensuring robust variability across the dataset. Both cameras generate events and intensity frames from a shared 280×180 pixel sensor array, ensuring consistency in spatial resolution.

2.2 MVSEC Dataset:

The multi vehicle stereo event camera (MVSEC) [6] dataset is captured using stereo DAVIS346 cameras with a resolution of 346×260 pixels. It provides synchronized event streams, intensity frames, and high-precision ground truth from LiDAR and IMU sensors. The dataset includes a variety of scenes, such as urban streets, natural environments, and indoor spaces, recorded under diverse conditions, including daylight and nighttime scenarios with significant motion and lighting variations. In our experiments, we used both night and daylight scenes of this dataset.

2.3 ECD_FAST:

This dataset is part of ECD dataset [7] used in our paper. Ercan et al [4], introduced the fast subset of this dataset to access reconstruction quality under fast camera motion. In our work, we access the reconstruction

3. TRAINING CONFIGURATIONS

To ensure a consistent comparison, we trained E2VID-WG and SPADE-WG using the same training setup as WaveE2VID. The event streams were preprocessed into event tensors using a fixed event count strategy, with each tensor accumulating $N = 30{,}000$ events. Training was conducted over 300 epochs with a batch size of 8 for all methods. For SPADE-WG, we adopted the many-to-one training loss strategy from the original work [8], to accelerate the training process. E2VID-WG, however, employed the same training loss function as WaveE2VID.

4. RESULTS

Quantitative Results: To evaluate the performance of the proposed methods, we used three standard metrics: MSE (lower is better), SSIM (higher is better), and LPIPS (lower is better). As shown in TABLE 1 our methods achieve significant improvements across nearly all datasets.

Qualitative Results: Figure 4 provides a visual comparison of SPADE-WG and the original SPADE-E2VID [4] model. Rows 1–2 show reconstructions from the ECD dataset, rows 3-4 from the HQF, 5-6 from MVSEC, and 7-8 from ECD_Fast datasets. On the ECD dataset, SPADE-WG produces reconstructions free of artifacts with sharper edges. For HQF, the original SPADE-E2VID struggles to reconstruct images, while SPADE-WG performs effectively. On MVSEC and ECD_FAST, SPADE-WG outperforms the baseline, particularly under challenging scenarios (fast camera motion), showcasing its robustness.

In Figure 5 the reconstruction results of E2VID-WG vs original E2VID [5] are shown for all datasets. Overall, our E2VID-WG has better reconstruction quality with fewer artifacts around edges.

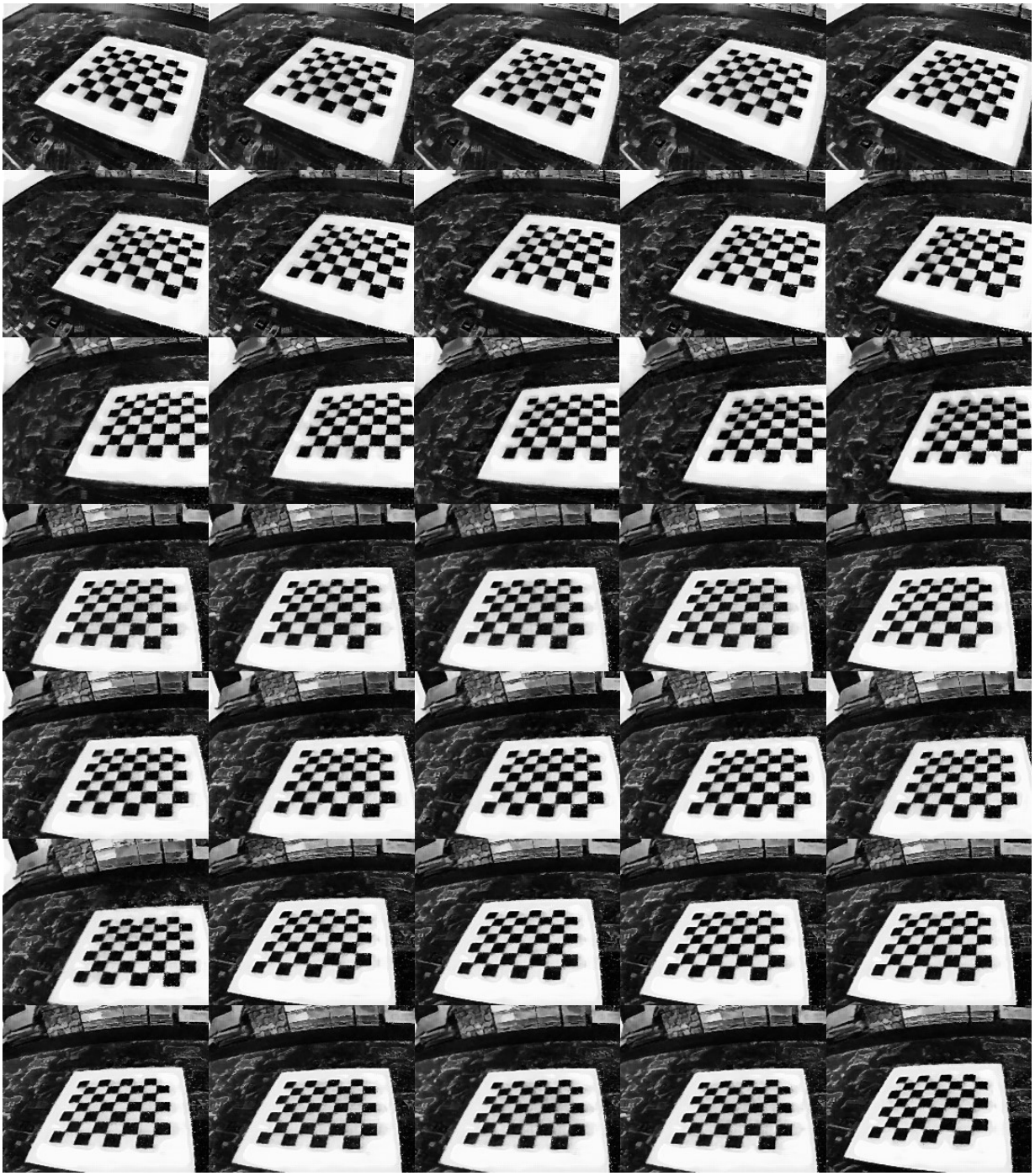

Figure 1: Qualitative analysis of reconstructed intensity image sequences from the calibration class of the ECD, highlighting temporal continuity.

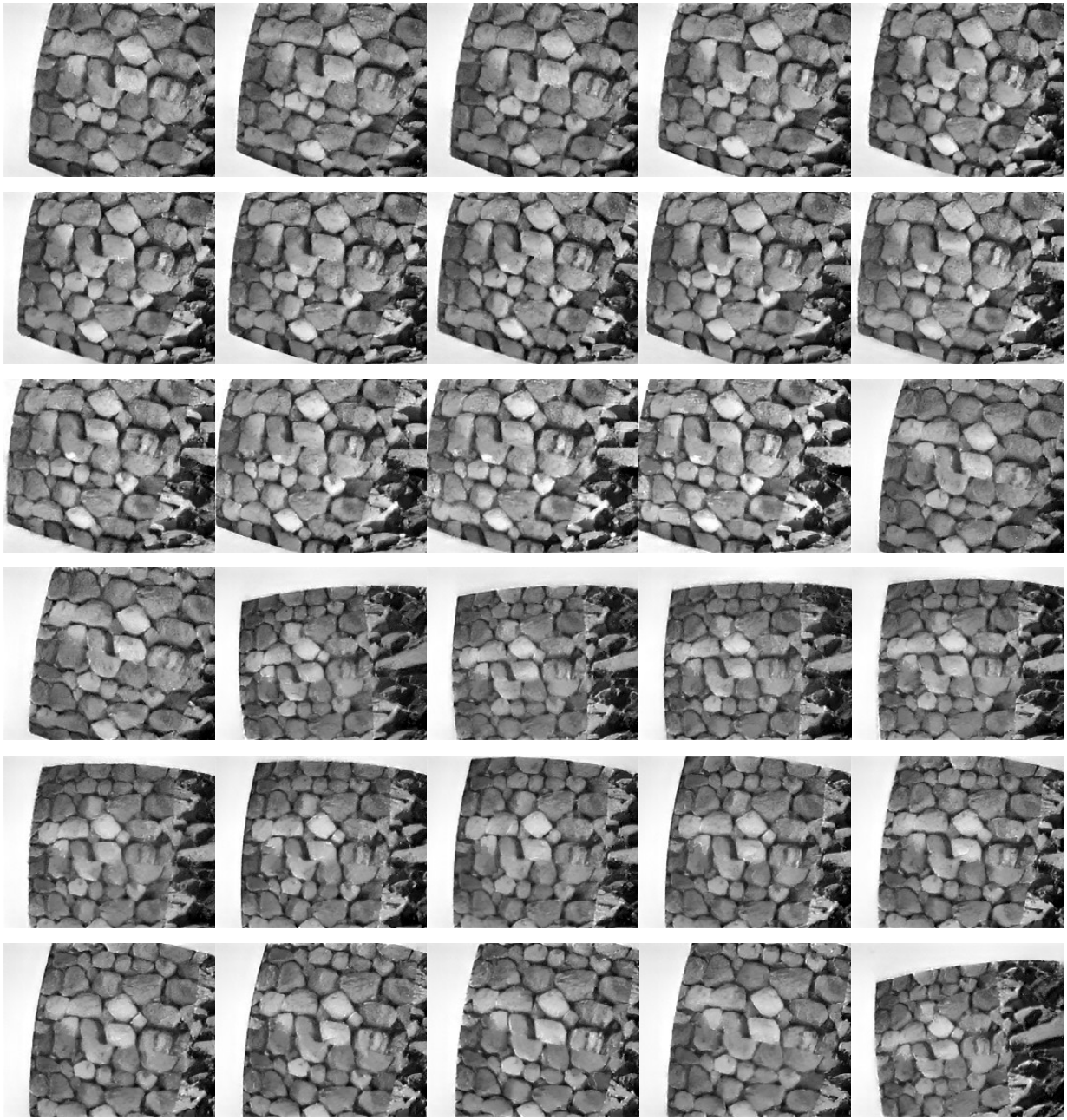

Figure 2: Visual results of a reconstructed images sequence from poster_6dof class of ECD.

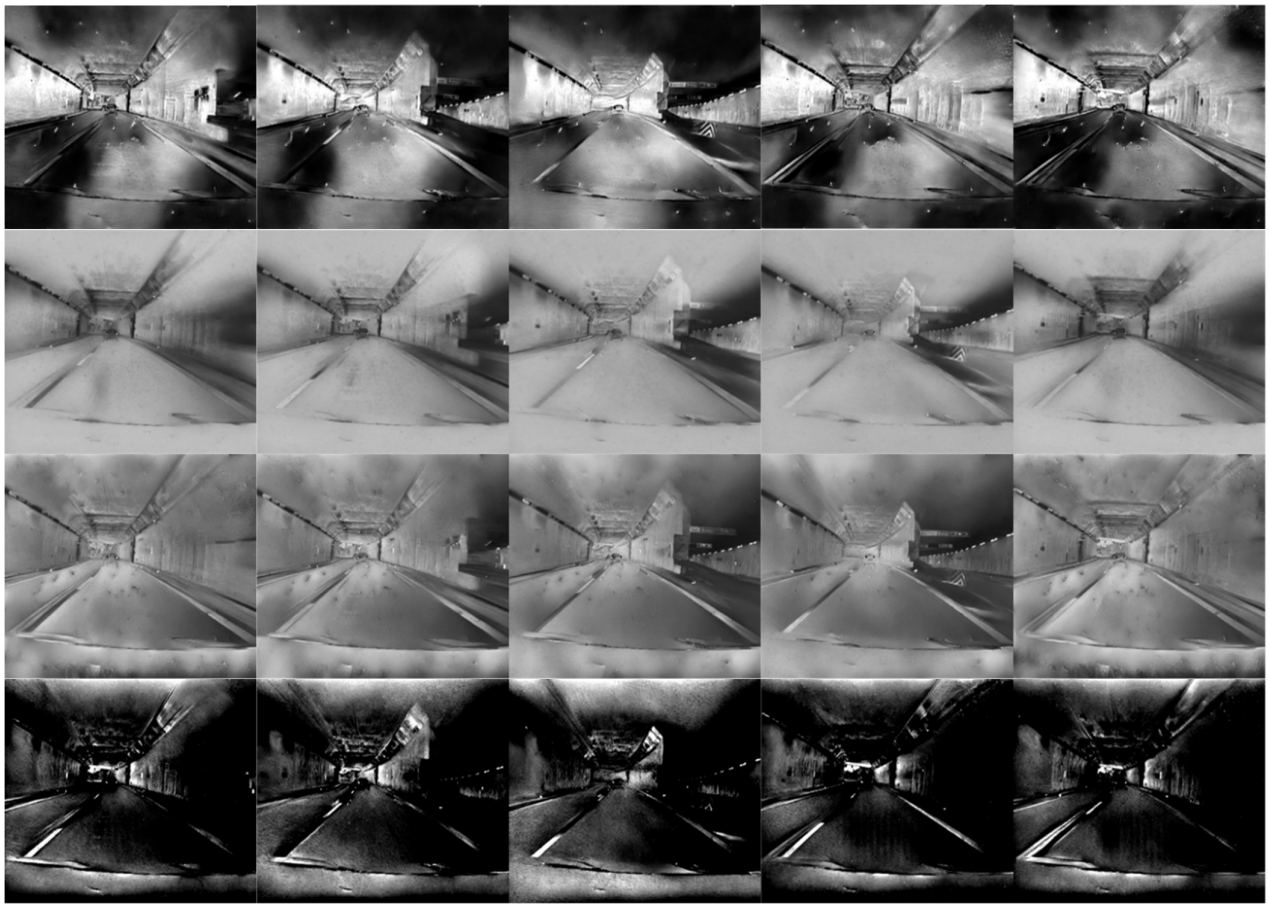

Figure 3: We show visual results on a sequence of high speed and HDR dataset. From top to bottom, we show results of WaveE2VID, FireNet, E2VID and HyperE2VID respectively.

TABLE 1: QUANTATIVE RESULTS OF PROPOSED METHODS AGAINST ORIGINAL METHODS. HERE BOLD VALUES INDICATE BEST-PERFORMED MODEL.

Methods	**MSE ↓**				**SSIM ↑**				**LPIPS ↓**			
	ECD	ECD_FAST	HQF	MVSEC	ECD	ECD_FAST	HQF	MVSEC	ECD	ECD_FAST	HQF	MVSEC
SPADE-E2VID	0.091	0.069	**0.074**	**0.128**	0.462	0.449	0.418	**0.266**	0.387	0.329	0.512	0.589
SPADE-WG	**0.084**	**0.058**	0.084	0.147	**0.475**	**0.466**	**0.443**	0.255	**0.254**	**0.310**	**0.329**	**0.581**
E2VID	0.064	**0.143**	0.098	0.205	0.502	0.374	0.468	0.241	0.426	0.413	**0.371**	0.644
E2VID-WG	**0.041**	0.166	**0.036**	**0.182**	**0.542**	**0.390**	**0.533**	**0.252**	**0.361**	**0.385**	0.382	0.644

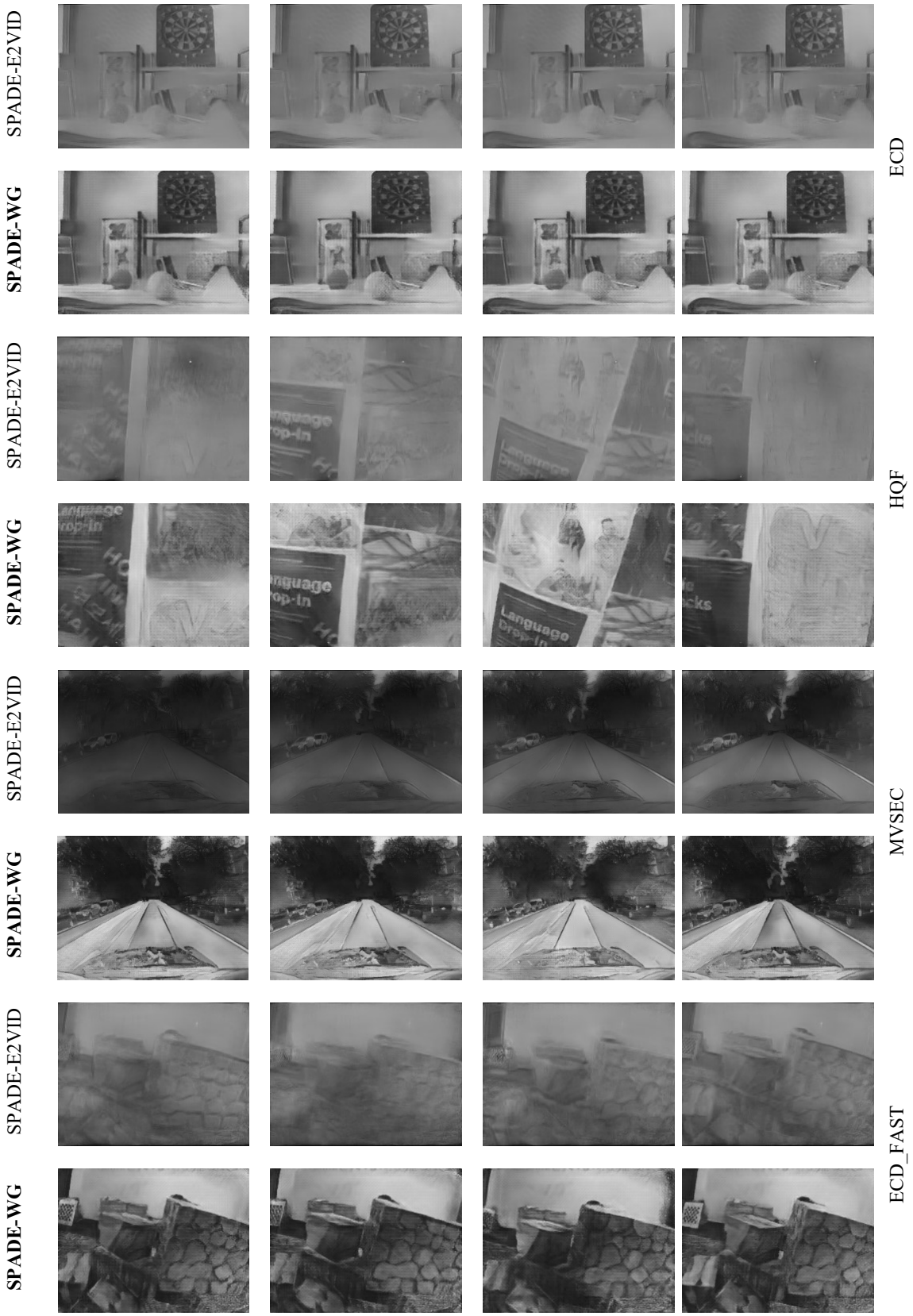

Figure 4: Reconstruction results of proposed SPADE-WG model on different event camera datasets. Overall, our model has better reconstruction quality with fewer artifacts

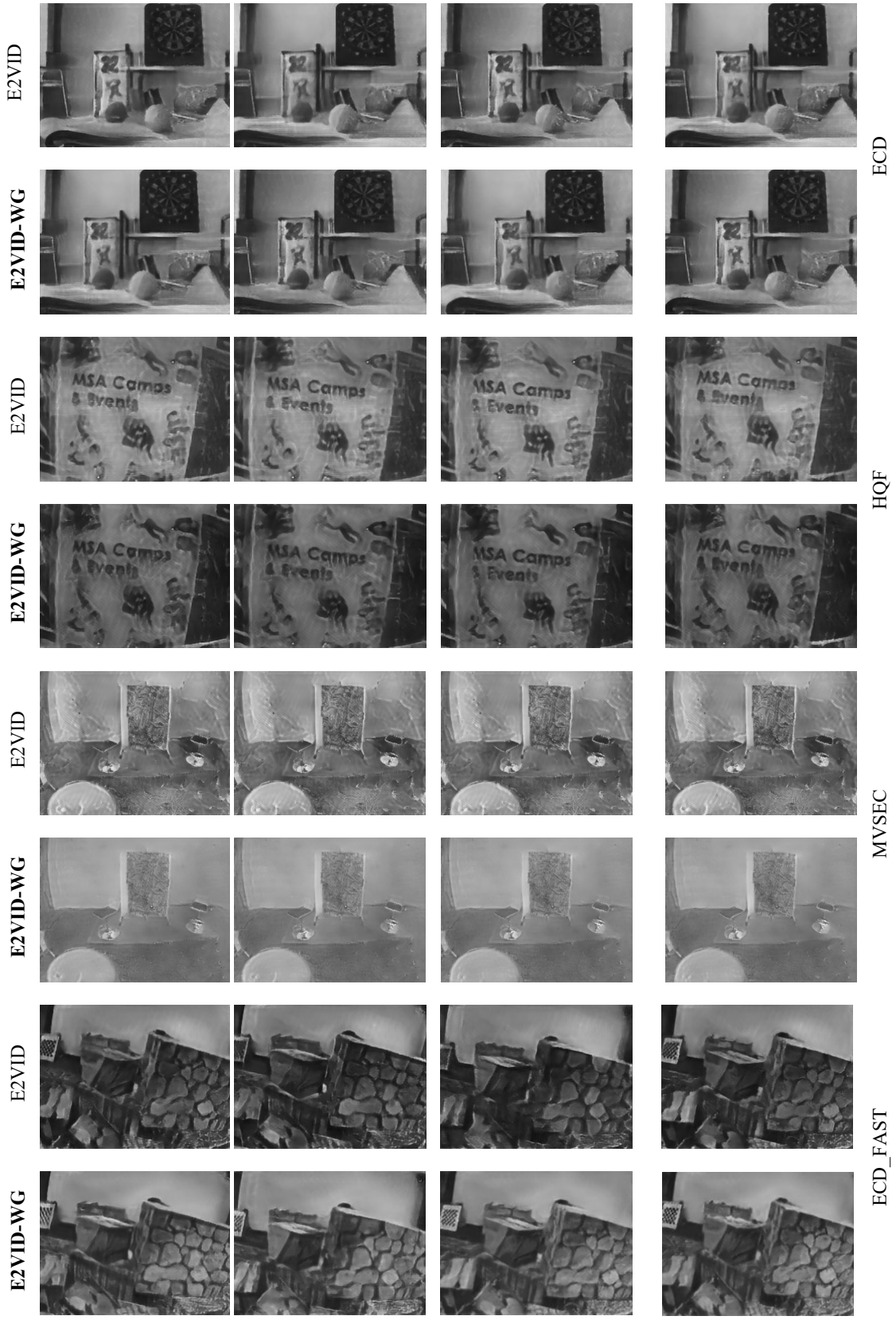

Figure 5: Reconstruction results of proposed E2VID -WG.

REFERENCES

[1] H. Rebecq, R. Ranftl, V. Koltun and D. Scaramuzza, "High Speed and High Dynamic Range Video with an Event Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 43, no. 9, pp. 1964-1980, 2019.

[2] C. Scheerlinck, H. Rebecq, D. Gehrig, N. Barnes, R. E. Mahony and D. Scaramuzza, "Fast Image Reconstruction with an Event Camera," in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Snowmass, CO, USA, 2020, pp. 156-163.

[3] H. Rebecq, R. Ranftl, V. Koltun and D. Scaramuzza, "High Speed and High Dynamic Range Video with an Event Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 43, no. 6, pp. 1964-1980, 2021.

[4] B. Ercan, O. Eker, C. Saglam, A. Erdem and E. Erdem, "HyperE2VID: Improving Event-Based Video Reconstruction via Hypernetworks," *IEEE Transactions on Image Processing,* vol. 33, pp. 1826-1837, 2024.

[5] T. Stoffregen, C. Scheerlinck, D. Scaramuzza, Drummond, B. Tom, K. Nick, M. Lindsay and R. Mahony, "Reducing the Sim-to-real Gap for Event Cameras," in *Computer Vision--ECCV*, 2020, pp. 534-549.

[6] A. Z. Zhu, D. Thakur, T. Ozaslan, B. Pfrommer, V. Kumar and K. Daniilidis, "The Multivehicle Stereo Event Camera Dataset: An Event," *IEEE Robotics and Automation Letters,* vol. 3, no. 3, pp. 2032-2039, 2018.

[7] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck and D. Scaramuzza, "The Event-Camera Dataset and Simulator: Event-based data for Pose Estimation, Visual Odometry, and SLAM," *The International Journal of Robotics Research,* vol. 36, no. 2, pp. 142-149, 2017.

[8] P. R. G. Cadena, Y. Qian, C. Wang and M. Yang, "SPADE-E2VID: Spatially-Adaptive Denormalization for Event-Based Video Reconstruction," *IEEE Transactions on Image Processing,* vol. 33, pp. 2488-2500, 2021.