

THAVELSAN



EVREAL: Towards a Comprehensive Benchmark and Analysis Suite for Event-based Video Reconstruction



Burak Ercan, Onur Eker, Aykut Erdem, and Erkut Erdem



- A fundamental task in event-based vision:
 Reconstructing intensity images from events
- It enables us to:
 - Visualize and interpret the scene being captured
 - Use frame-based computer vision methods for downstream tasks (recognition, calibration, ...)
- Not completely solved yet
- Evaluation has challenges on its own ...

Issues with Evaluation

- Lack of a standard evaluation procedure
- Benchmark datasets limited in scale and scope
- Paradox of ground truth: We want to evaluate especially when reference frames are low quality:





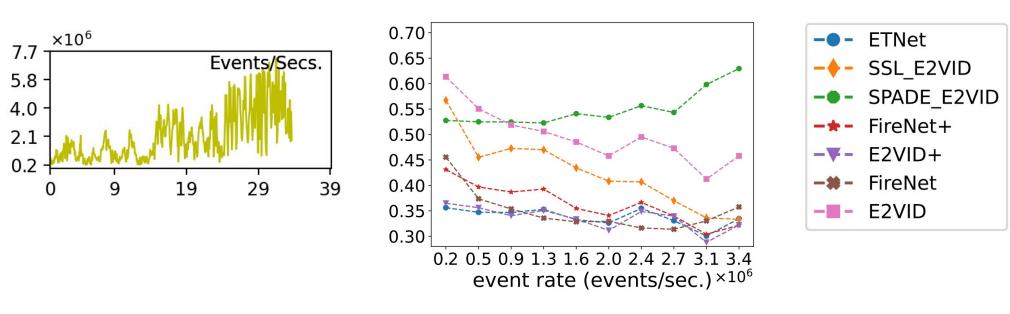


Motion blur

Low Dynamic Range

Low Light

Assessing **robustness** under changing conditions:



Subjective nature of image quality assessment

Evaluate on Downstream Intensity Image Downstream Tasks No-Reference Intensity Image 0 2 6 Result 5 Settings of Classification Metrics Post-processing Datasets 1. Select **Pre-trained** • Normalization Neural Network Networks Prediction Evaluation Object Detection **Neural Network Pre-processing** Pipeline → Data Flow Settings **Event** Representation · · · · > Settings Grouping Pre-processing Representation Events Fixed-number Camera Calibration Fixed-duration

What We Propose with EVREAL

- A unified evaluation pipeline
- A result analysis tool to visualize and compare reconstructions and scores
- Large set of real-world test sequences
- Full-reference and no-reference metrics
- Evaluation in challenging scenarios
- Robustness analysis under varying conditions
- Evaluation on downstream tasks
- Evaluation of **computational efficiency**

Evaluation Details

Models: E2VID, FireNet, E2VID+,
FireNet+, SPADE-E2VID,
SSL-E2VID, ET-Net, HyperE2VID*

Datasets: ECD, MVSEC, HQF, BS-ERGB, ECD-FAST, MVSEC-NIGHT, HDR

Image Quality Metrics: MSE, SSIM, LPIPS, BRISQUE, NIQE, MANIQA

Main Takeaway from Experimental Results

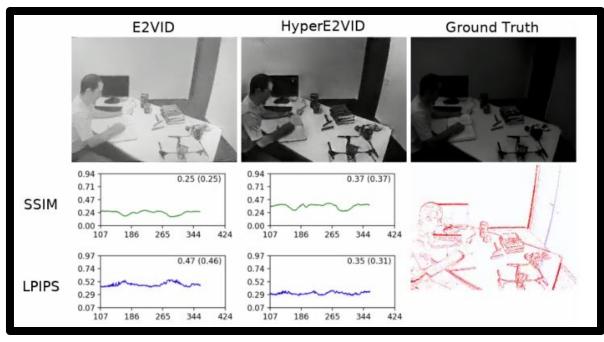
 Good scores on standard benchmark sequences do not always indicate good scores on challenging scenarios, varying conditions, and downstream tasks.

Summary & Conclusions

- EVREAL can serve as a **valuable resource** for researchers and practitioners.
- Using EVREAL yielded insightful observations under varying settings, challenging scenarios, and downstream tasks.
- Results imply that standard benchmark sequences are not enough and further analysis needed.
- We believe our work will contribute to the development of more effective and robust models.

Note: In our interactive webpage, we also present results of a **new state-of-the-art model**, **HyperE2VID!**







Acknowledgements: This work was supported in part by KUIS Al Center Research Award, TUBITAK-1001 Program Award No. 121E454, and BAGEP 2021 Award of the Science Academy to A. Erdem.