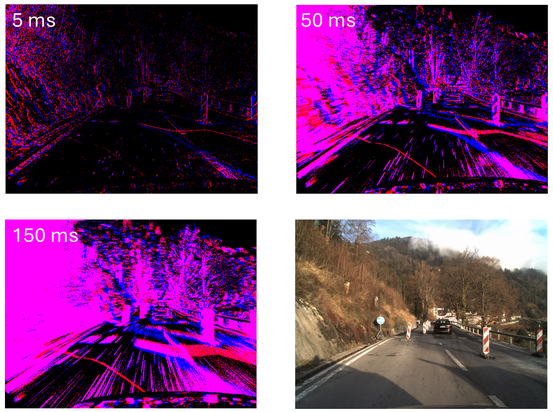
A Limited Number of Events Is Not Informative Enough for Prediction.

Event-Cameras record brightness changes, and many regions remain invisible over short periods.

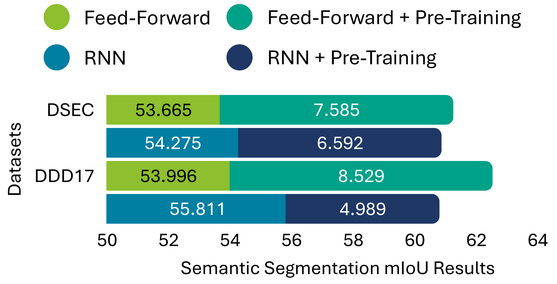


Labeled Event Sequences Are Limited For Training RNNs.

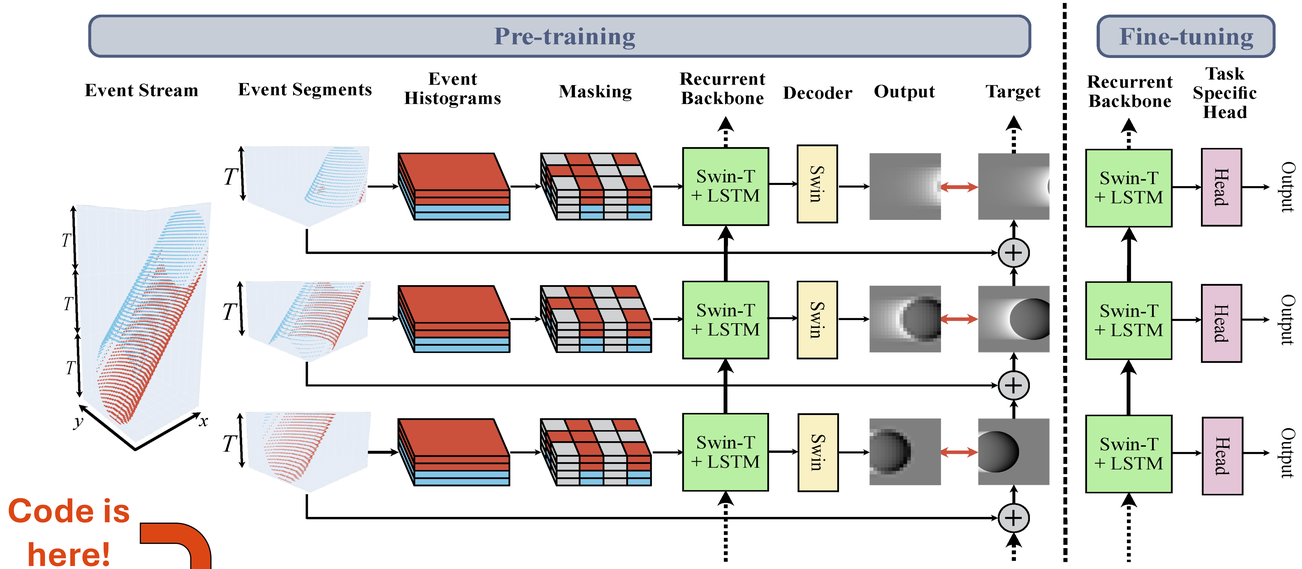
✅ Sequence models can process information beyond a limited window of events.

❌ Training requires several long streams of data.

Existing pre-training methods focus only on spatial information and serve better as initial weights for feed-forward models.



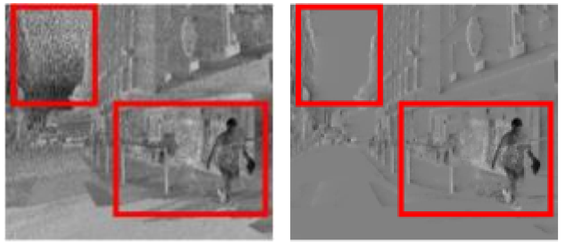
There is a lack of task-specific Event-Camera data to train sequence models. This can be mitigated by pre-training with a special target that contains spatiotemporal information.



Code is here!

Our Pre-Training Target Contains Spatiotemporal Information.

We propose a new event-to-image method that produces targets resembling grayscale images.



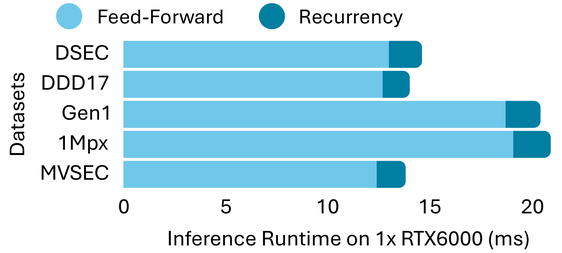
- Tube masking and reconstruction encourage the model to learn temporal information.
- Our target shows low motion blur and high robustness to noise; essential for pre-training.

Results; SOTA Everywhere!

RNN · RNN + Existing Pre-Training · Ours · Previous SOTA

DESC	mIoU	54.3	+5.5	+8.4	+7.0
DDD17		55.8	+3.9	+9.4	+6.8
Gen1	mAP	48.5	+1.1	+3.1	+1.9
1Mpx		45.4	+4.6	+5.2	+3.0
MVSEC	RMS	7.90	-0.61	-1.24	-1.03

Downstream Perception Tasks Results



Temporally-Enhanced Self-Supervised Pretraining for Event Camera

Mohammad Mohammadi, Ziyi Wu, Igor Gilitschenski

ICCV OCT 19-23, 2025 HONOLULU HAWAII

TORONTO INTELLIGENT SYSTEMS LAB · UNIVERSITY OF TORONTO · VECTOR INSTITUTE | INSTITUT VECTEUR