# ETAP: Event-based Tracking of Any Point

Friedhelm Hamann, Daniel Gehrig, Filbert Febryanto, Kostas Daniilidis, Guillermo Gallego

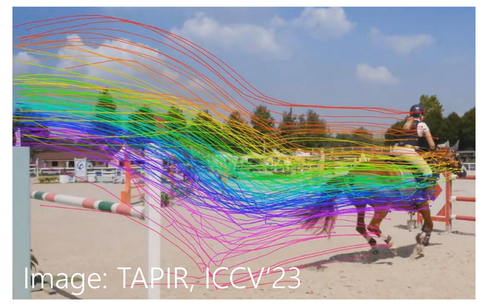
Technische Universität Berlin · Einstein Center Digital Future · Robotics Institute Germany · Science of Intelligence · Penn Engineering GRASP Laboratory (General Robotics, Automation, Sensing & Perception Lab)

CVPR Nashville June 11-15, 2025

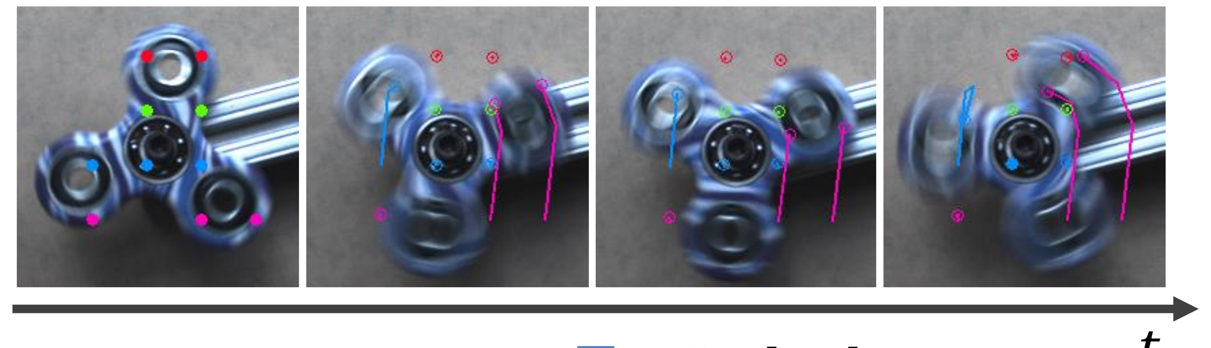
Highlight

## Introduction

We introduce the first method for **event-only point tracking**, overcoming limitations on RGB-based tracking.

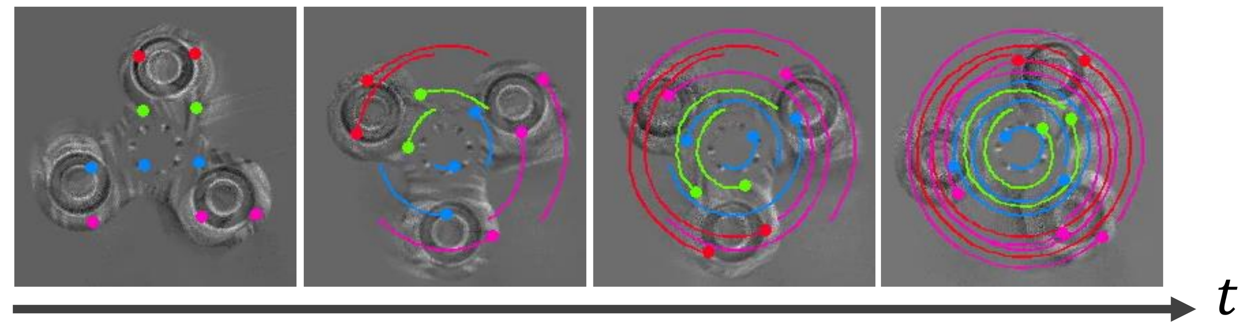



RGB-based point tracking works well in good conditions (well lit, strong colors, slow motion).




However, **it fails for challenging conditions**, like fast movements & low light.

**Solution**

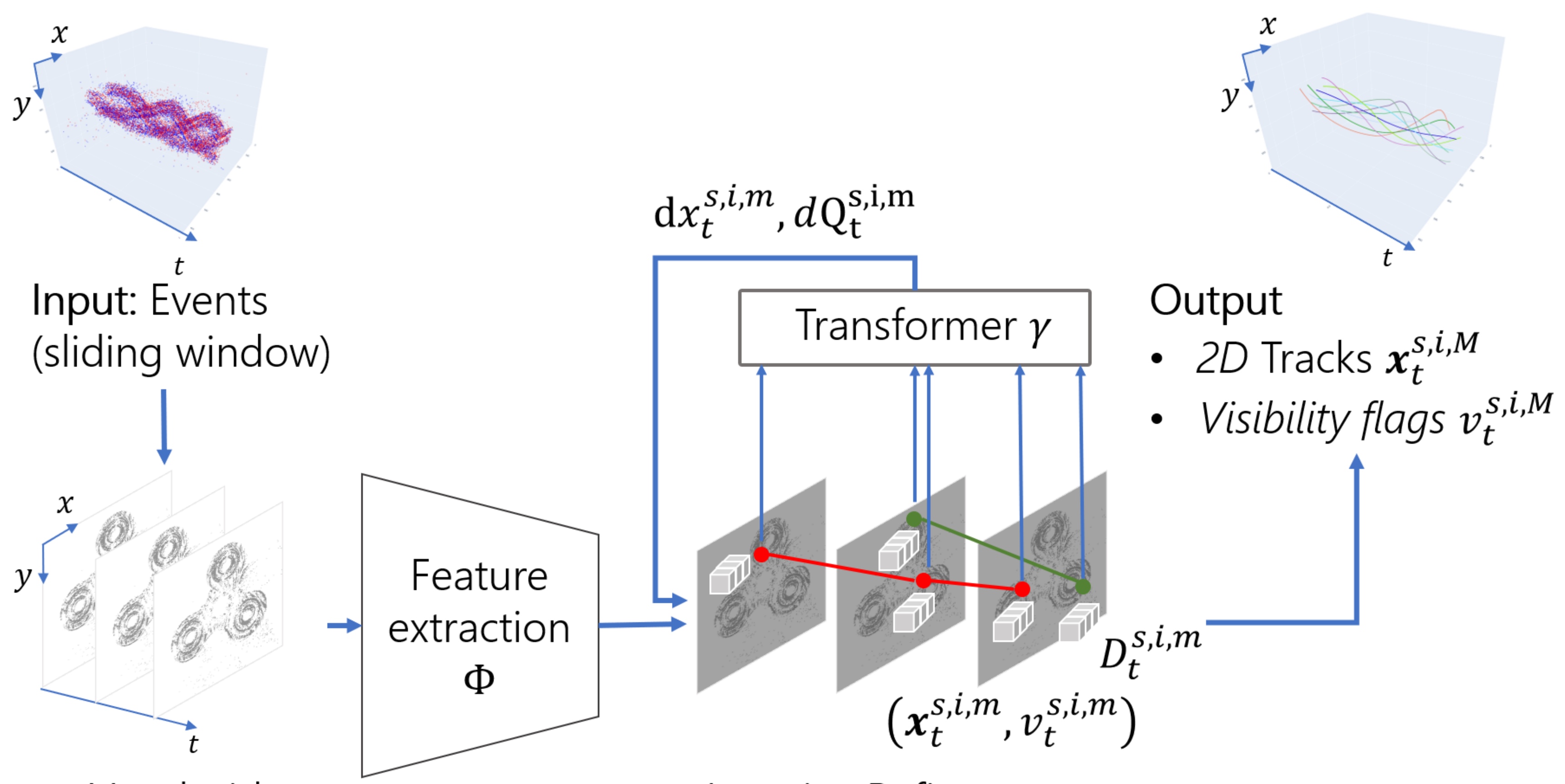



Event cameras handle these scenarios with their **high temporal resolution, minimal motion-blur and high dynamic range**.

## Method

The method tracks multiple points in parallel in a sliding window approach.

The input are raw **events** and **query points**, the output are **the 2D point tracks** and **visibility flags**.




## Summary

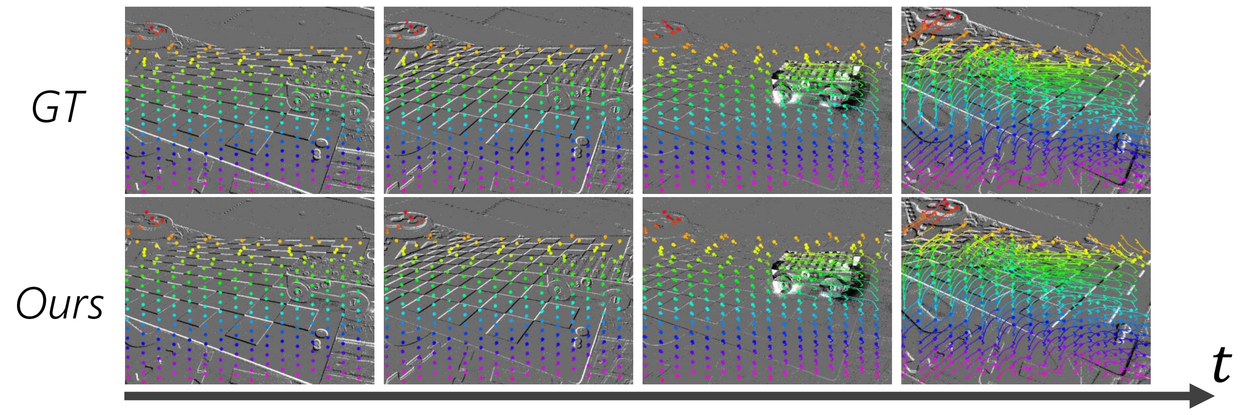
- Scaling synthetic event-generation combined with event specific losses leads **to strong event-only point trackers**.
- Results on new scenes proof **advantages over RGB** in challenging scenarios.
- We provide new ground truth fostering further development of event-based TAP.

Code
Dataset
Video

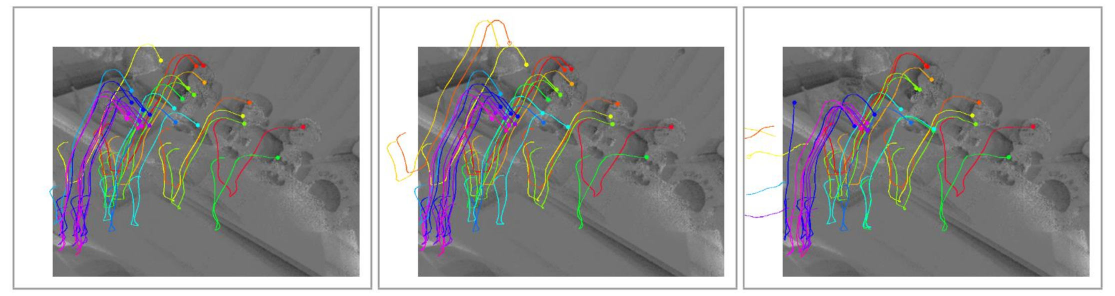
## Evaluation

**Strong cross-dataset generalization**, tested on six datasets.

1. We enable quantitative evaluation of the new event-based point tracking task, with **new ground truth** data for the datasets EVIMO2 and E2D2.



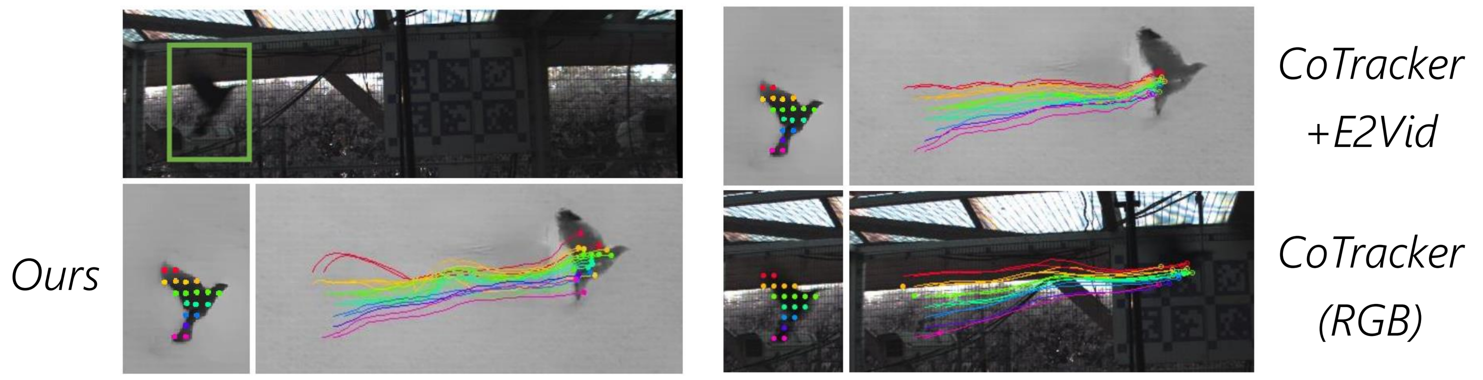

2. Evaluation on an established feature tracking benchmark (EDS/EC) shows **20% improvement** over the previously best method.



*EDS Example*

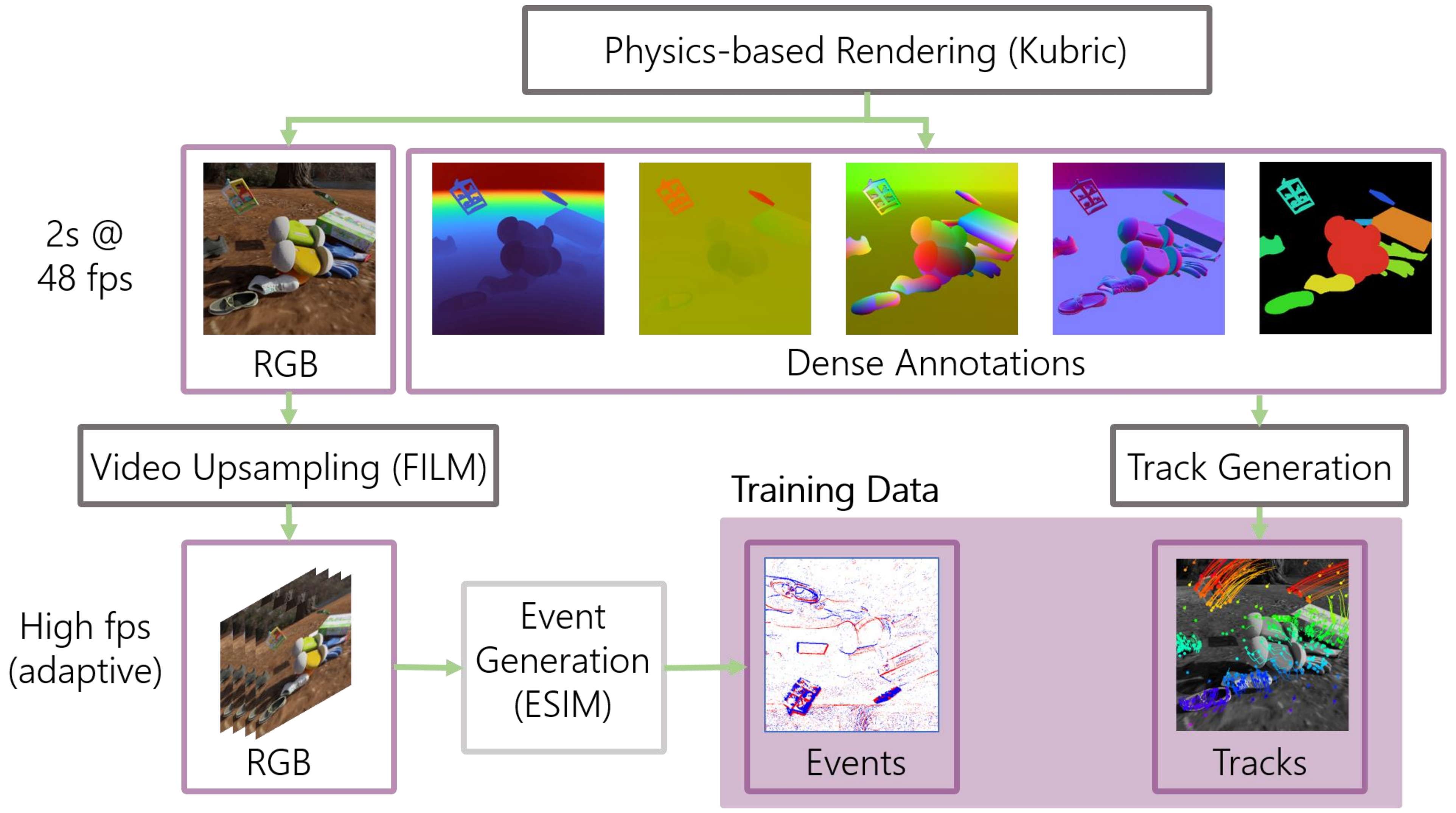
| Method | Input | EDS Feature Age ↑ | EDS Expected FA ↑ | EC Feature Age ↑ | EC Expected FA ↑ |
|---|---|---|---|---|---|
| ICP [32] | E | 0.060 | 0.040 | 0.256 | 0.245 |
| EKLT [21] | E+F | 0.325 | 0.205 | 0.811 | 0.775 |
| DDFT [44] | E+F | 0.576 | 0.472 | 0.825 | 0.818 |
| FE-TAP [38] | E+F | 0.676 | 0.589 | 0.844 | 0.838 |
| EM-ICP [63] | E | 0.161 | 0.120 | 0.337 | 0.334 |
| HASTE [3] | E | 0.096 | 0.063 | 0.442 | 0.427 |
| DDFT E2VID [44] | E | 0.589 | 0.495 | 0.794 | 0.786 |
| **ETAP w\o FA-loss (Ours)** | E | 0.698 | **0.599** | 0.885 | 0.879 |
| **ETAP (Ours)** | E | **0.704** | 0.598 | **0.888** | **0.883** |

3. Qualitative tests on a demanding scenario shows advantages over RGB-based tracking methods (**small, low-textured, fast, high deformation, HDR**).



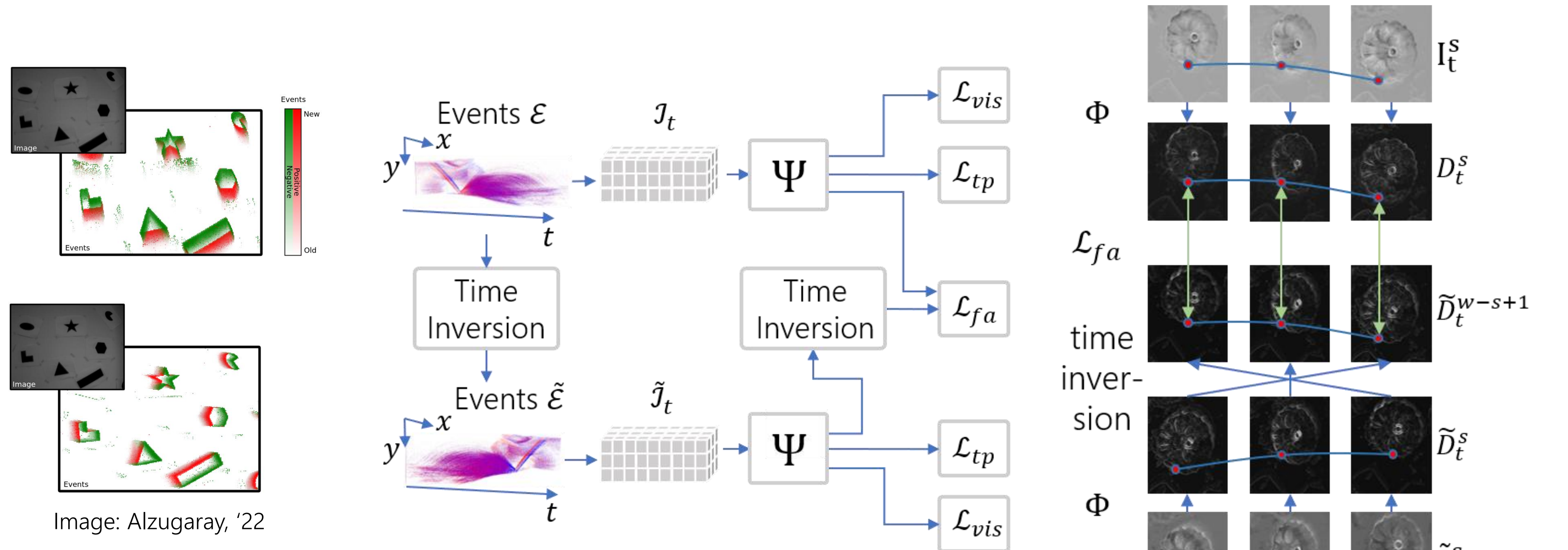

## Synthetic Training Data Generation

The model is trained solely on **synthetic event data** using a combination of the rendering engine **Kubric** and **Vid2e**.




## Training Pipeline and Loss Function

The model is trained with a combined loss of the common track prediction error (absolute distance between predicted and GT tracks), a cross-entropy loss on the visibility flags, **and a novel feature alignment loss**.




In event cameras the data **is a function of the scene motion**

At training time, we track the same scene forward and backward, **thereby inverting the flow, and maintaining appearance.**

We **enforce similarity** between the track features of forward and backward samples.