EDeNN: Event Decay Neural Networks for low latency vision

CVSSP Centre for Vision, Speech and Signal Processing

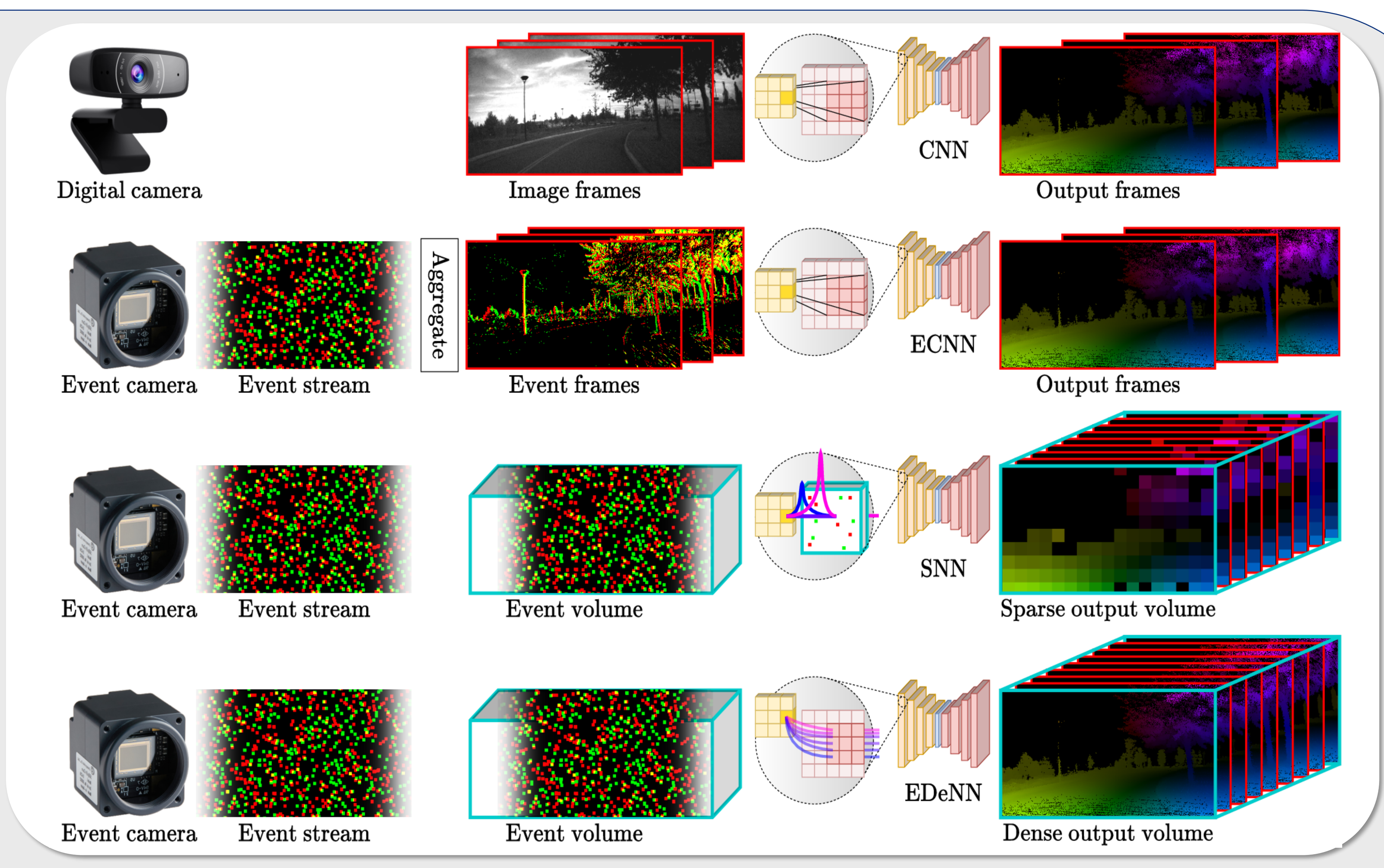
Celyn Walters (celynwalters@gmail.com)
Simon Hadfield (S.Hadfield@surrey.ac.uk)

University of Surrey

1 - Abstract

- A new approach for **deep learning** with **event cameras.**
- Operates **directly** on the event stream
  - No intermediate aggregation
  - **Densifies** sparse events via learned decay functions.
- Combines **responsiveness** of an SNN with the **efficiency** and spatial reasoning of a CNN.

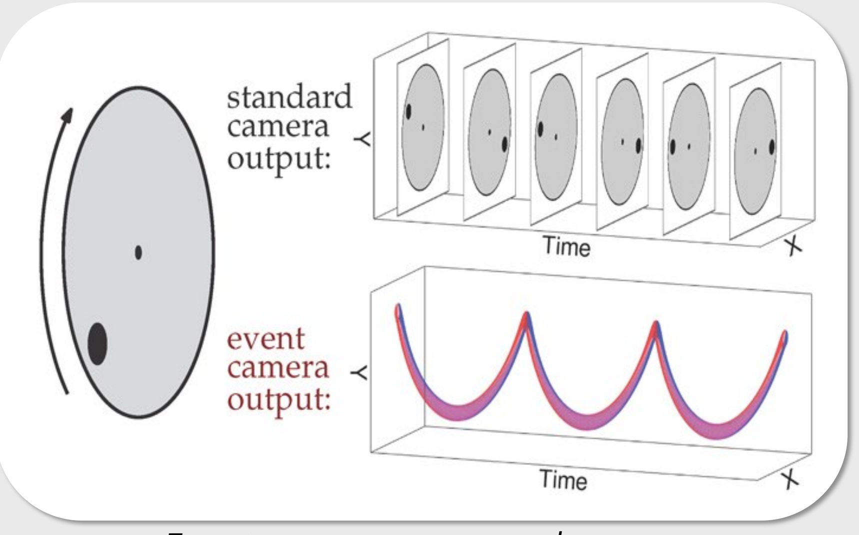
Code available



EDeNN vs other learning styles

2 - Background - Event Cameras

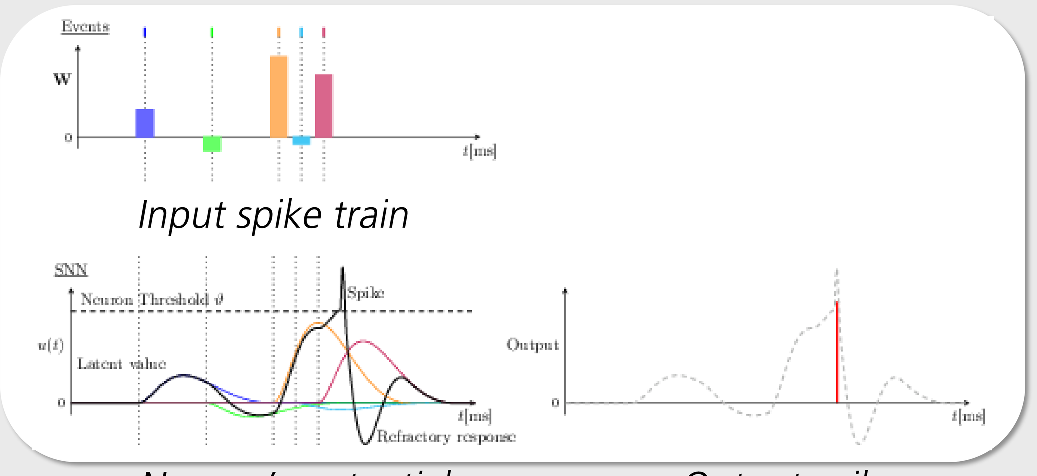
- Event cameras are **asynchronous** visual sensors
- Brightness changes cause **immediate** signals from the sensor, with no shutter based poling
- Numerous advantages: low-power, low bandwidth, high dynamic range, and low **world-to-sensor** latency
- Disadvantage: **No images**, unclear how to apply traditional computer-vision tools



Event camera vs normal camera

3 - Background - Spiking Neural Networks

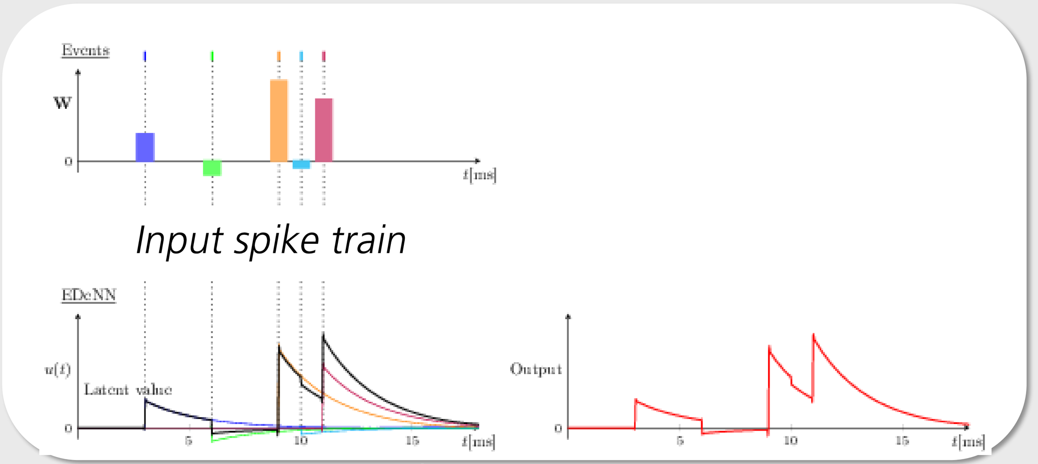
- SNN **input**: **spikes** of varying strength at arbitrary times
  - Strong parallels to event camera data
- Neuron aggregates spikes with weightings as potential function
- Thresholded potential gives **output spikes** for the next layer
- Many input spikes create 1 output spike
- "**Vanishing spike**" problem for deep networks



Input spike train

Neuron's potential *Output spikes*

4 - EDeNN overview



Input spike train

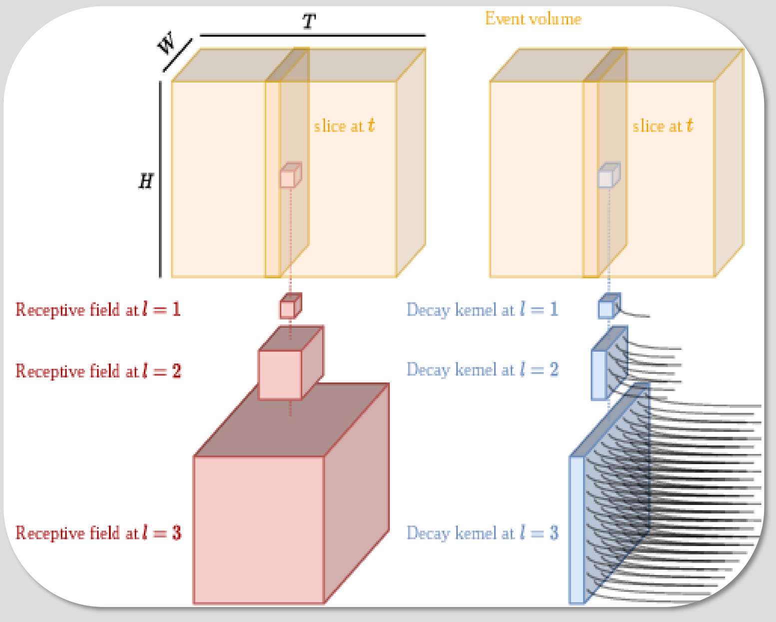
EDeNN potential *EDeNN output*

- **Output** is **continuous** potential function (unthresholded)
  - No vanishing spikes!
- Potential function aggregates exponential **decay of spikes**
  - No ramp up time or refractory period, learned decay rate
- Can apply to either spikes or continuous input

5 - EDeNN Details

5.1 - EDeC layer

- Event Decay Convolutions
  - CNN style **spatial convolution** kernel
  - SNN style **temporal decay** (learned per neuron)
- Spatiotemporal convolution - K^2+1 params (not K^3)



Spatiotemporal convolution (left) vs EDeC (right)

5.2 - Separability & streaming inference

- Specialist filter design has valuable properties
  - **Separable**: decompose spatial & temporal convs.
  - **Markov**: Output depends on input and prev. output
- Output at time t and layer l comprises **2 terms**
  - **Spatially convolved** input from layer l-1 at time t
  - **Temporally decayed** input from layer l at time t-1
  - See paper for full derivation

$$E^l(:,\bar{c},t) = \sum_{c\in C}\left[\boxed{K_c^{\bar{c}} * E^{l-1}(:,c,t)} + \boxed{\gamma E^l(:,c,t-1)}\right].$$

- **Training** can be performed in **parallel** over time
- **Inference** can be performed efficiently **online**
  - No waiting for future information as with 3D CNN

5.3 - Weighted partial convolution

- Partial convolutions used to **ignore empty** regions
  - Avoids wasted computation
  - Aids training stability
- **Weighted** to counter effect of **missing inputs**

$$\alpha^l(\boldsymbol{x},t) = \frac{2|\Omega|}{\sum_{\delta\in\Omega}\left[\mathbf{M}^{l-1}(\boldsymbol{x}+\delta\boldsymbol{x},t) + \mathbf{M}^l(\boldsymbol{x}+\delta\boldsymbol{x},t-1)\right]},$$

- Applied to spatial and temporal EDeC components
- Novel **reweighting** scheme based on **masked kernel weights** (not masked input values)

$$\hat{\alpha}^l(\boldsymbol{x},\bar{c},t) = \frac{\gamma|\Omega| + \sum_{c\in C}\sum_{\delta\in\Omega} K_c^{\bar{c}}(\delta\boldsymbol{x})}{[a+\gamma b]},$$

6 - Results - Angular velocity estimation

- Event camera dataset
- Estimate rate-of-change for roll, pitch and yaw
- Compared vs SOTA on **V**oxel, **A**ccumulated and **E**vent inputs

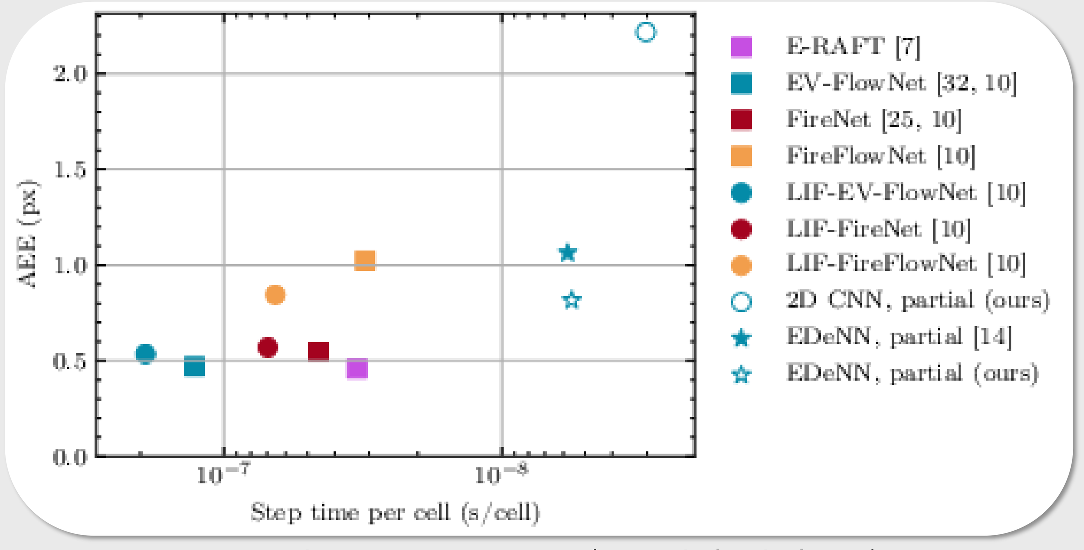
Approach	Data	Relative error	RMSE	Step time
ANN-6	V	0.22	59.00	–
ResNet-50	A	0.22	66.80	–
ResNet-50	V	0.15	36.80	–
SNN-6	E	0.26	66.32	0.15
EDeNN	E	**0.12**	**27.99**	**0.08**

State-of-the-art comparison for visual angular velocity estimation

- EDeNN **more accurate** than best **CNN**
- EDeNN **faster** than best **SNN**

7 - Results - Optical flow estimation

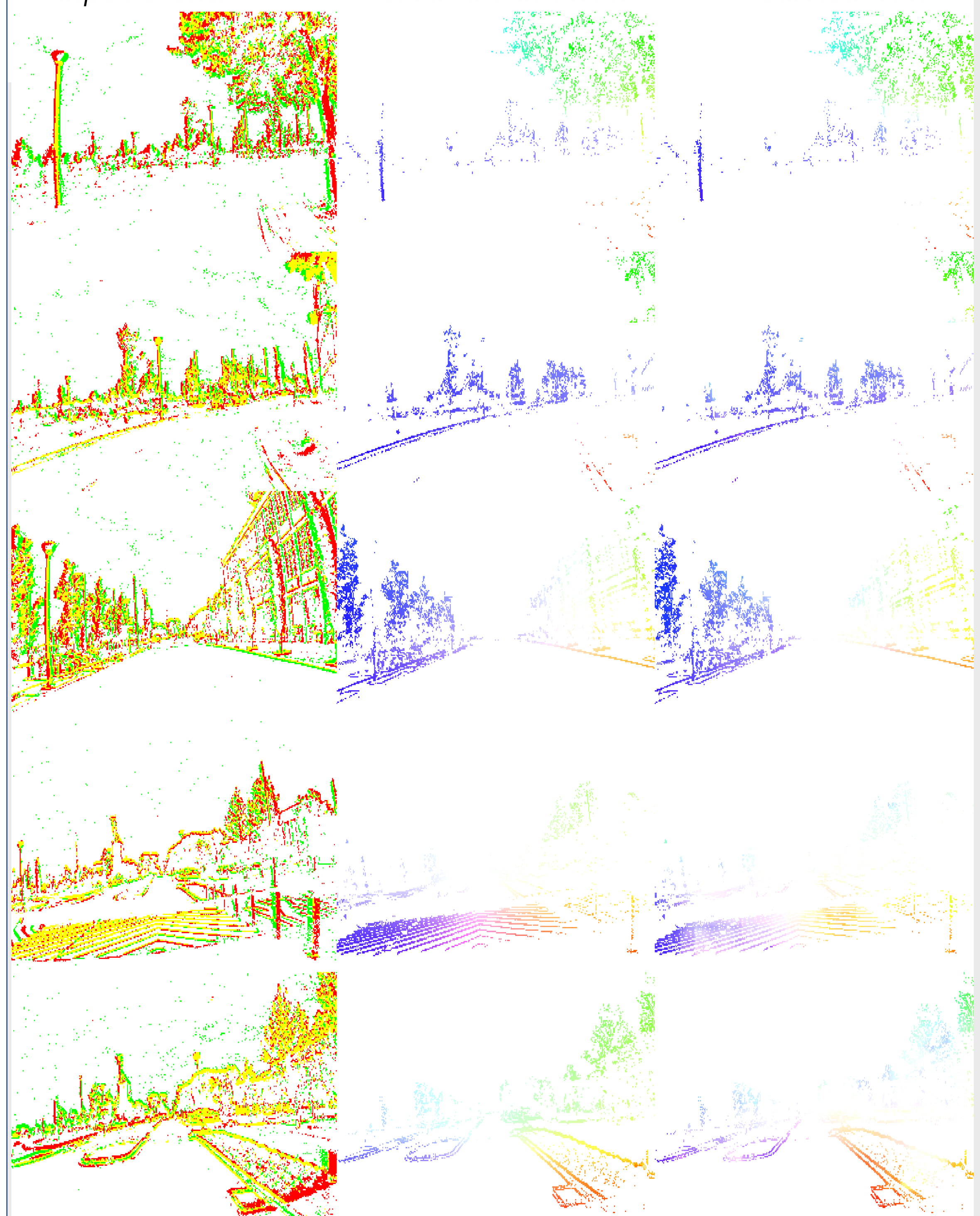
- MVSEC dataset
- Semi-dense optical flow estimation
- Extremely challenging for SNNs (vanishing spikes)



Run time vs accuracy (top right is best)

- EDeNNs achieve **comparable accuracy** to SOTA **CNNs**
- **Order of magnitude faster** runtime than second fastest

Input events *Ground truth flow* *Predicted flow*



Example optical flow results

- Flow direction (hue) matches well
- **Near stationary** scenes **challenging** (final row)
  - Few events with high overlap
  - Larger flow orientation errors
- EDeNN is **effective at complex geometry** like foliage
- Potential for future approaches to fuse events and RGB in EDeNNs

Funded by EPSRC grant "Reflexive robotics" (EP/S035761/1)

EPSRC Engineering and Physical Sciences Research Council