Back to Event Basics: Self-Supervised Learning of Image Reconstruction for Event Cameras via Photometric Constancy

Federico Paredes-Valles and Guido C. H. E. de Croon
(Poster Session Three, ID: 8305)

Code and models: mavlab.tudelft.nl/ssl_e2v/
Problem formulation

Event cameras and image reconstruction:

![Diagram showing the process of event camera and image reconstruction]

- **Brightness** → **Event camera** → **Event stream**
- **Event camera** → **NN** → **Brightness**
Problem formulation

Event cameras and image reconstruction:

Minimal training pipeline:
- Rebecq et al., TPAMI’19
- Stoffregen et al., ECCV’20
Problem formulation

Event cameras and image reconstruction:

Our goal:
To leverage our knowledge of the inner workings of event cameras to learn, in a self-supervised fashion, to perform image reconstruction without the need for any ground-truth or synthetic data.
Related work

**Flow - Brightness**
- Kim *et al.* (JSSC’08, ECCV’16)
  - EKF + Poisson int.
- Cook *et al.* (IJCNN’11)
  - Variational opt.
- Bardow *et al.* (CVPR’16)
  - Variational opt.

  - Joint estimation of flow and brightness
  - Computationally expensive
  - Hand-crafted regularizers

**Event integration**
- Reinbacher *et al.* (IJCV’18)
  - Manifold regularization
- Scheerlinck *et al.* (ACCV’18)
  - High-pass filter

**Machine learning**
- Rebecq *et al.* (TPAMI’19)
  - Synthetic labeled data
  - E2VID: Recurrent CNN
- Stoffregen *et al.* (ECCV’20)
  - Refined data augmentation
  - E2VID
  - State-of-the-art
  - Among many others...

TU Delft
Related work

Flow - Brightness
- Kim et al. (JSSC’08, ECCV’16)
  - EKF + Poisson int.
- Cook et al. (IJCNN’11)
  - Variational opt.
- Bardow et al. (CVPR’16)
  - Variational opt.

+ Joint estimation of flow and brightness
- Computationally expensive
- Hand-crafted regularizers

Event integration
- Reinbacher et al. (IJCV’18)
  - Manifold regularization
- Scheerlinck et al. (ACCV’18)
  - High-pass filter

+ Computationally efficient
- Artifacts due to unknown sensor parameters

Machine learning
- Rebecq et al. (TPAMI’19)
  - Synthetic labeled data
  - E2VID: Recurrent CNN
- Stoffregen et al. (ECCV’20)
  - Refined data augmentation
  - E2VID
  - State-of-the-art
- Among many others...
Related work

**Flow - Brightness**
- Kim et al. (JSSC’08, ECCV’16)
  - EKF + Poisson int.
- Cook et al. (IJCNN’11)
  - Variational opt.
- Bardow et al. (CVPR’16)
  - Variational opt.
  - Joint estimation of flow and brightness
  - Computationally expensive
  - Hand-crafted regularizers

**Event integration**
- Reinbacher et al. (IJCV’18)
  - Manifold regularization
- Scheerlinck et al. (ACCV’18)
  - High-pass filter
  + Computationally efficient
  - Artifacts due to unknown sensor parameters

**Machine learning**
- Rebecq et al. (TPAMI’19)
  - Synthetic labeled data
  - E2VID: Recurrent CNN
- Stoffregen et al. (ECCV’20)
  - Refined data augmentation
  - E2VID
  - State-of-the-art
  + Computationally efficient
  + High reconstruction accuracy
  - Sim-to-real gap

Among many others...
Related work

Flow - Brightness
- Joint estimation of flow and brightness
- Computationally expensive
- Hand-crafted regularizers

Cook et al. (IJCNN’11)
- Variational opt.

Bardow et al. (CVPR’16)
- Variational opt.

Event integration
- Computationally efficient
- Artifacts due to unknown sensor parameters

Reinbacher et al. (IJCV’18)
- Manifold regularization

Scheerlinck et al. (ACCV’18)
- High-pass filter

Machine learning
- Computationally efficient
- High reconstruction accuracy
- Sim-to-real gap

Rebecq et al. (TPAMI’19)
- Synthetic labeled data
- E2VID: Recurrent CNN

Stoffregen et al. (ECCV’20)
- Refined data augmentation
- E2VID
- State-of-the-art

Among many others...

We propose to come back to the theoretical basics of event cameras with a machine learning approach that leverages the optical flow - image brightness relation to learn to perform image reconstruction from real unlabeled event data while remaining computationally efficient.
Proposed framework
Self-supervised image reconstruction

Proposed training pipeline:
- FlowNet learns to estimate event-based optical flow by compensating for the motion blur in the input events (Zhu et al., CVPR’19).
- ReconNet learns to perform image reconstruction by predicting the brightness frames that best satisfy the input events and the estimated optical flow.

Self-supervised learning of optical flow via contrast maximization:
Proposed framework
Self-supervised image reconstruction

**Measured** brightness increment

\[ \Delta L = \sum_{e_i \in \Delta t} p_i C \]

The brightness change encoded in the events...

...is caused by the displacement of the spatial gradients of the brightness signal.

**Predicted** brightness increment

\[ \Delta L \approx \frac{\partial L}{\partial t} \Delta t \]

\[ \frac{\partial L}{\partial t} + \nabla L \cdot u = 0 \]

\[ \Delta L \approx -\nabla L \cdot u \Delta t \]

Generative model

\( L \doteq \log(I) \)

\( I \): image intensity
The brightness change encoded in the events...

\[ \sum_{e_i \in \Delta t} p_i C \approx -\nabla L \cdot u \Delta t \]

...is caused by the displacement of the spatial gradients of the brightness signal.

Proposed framework
Self-supervised image reconstruction
Training details

**Loss function:**

\[ \mathcal{L}_{\text{ReconNet}} = \sum_{k=0}^{S} \mathcal{L}_{\text{model}} + \lambda_2 \sum_{k=S_0}^{S} \mathcal{L}_{\text{TC}} + \lambda_3 \sum_{k=0}^{S} \mathcal{L}_{\text{TV}} \]

**Architectures:**

**Input representation:** voxel grid (Zhu, CVPR’19)

**FlowNet:**
- EV-FlowNet (Zhu et al., RSS’18)
- FireFlowNet (Ours)

**ReconNet:**
- E2VID (Rebecq et al., TPAMI’19)
- FireNet (Scheerlinck et al., WACV’20)

**Dataset:** UZH-FPV Drone Racing Dataset (Delmerico, ICRA’19).
Results

Event-Camera Dataset (Mueggler, IJRR’17)

| Method                  | MSE  | SSIM | LPIPS |
|-------------------------|------|------|-------|
| E2VID (Rebecq, TPAMI’19)| 0.08 | 0.54 | 0.37  |
| FireNet (Scheerlinck, WACV’20) | 0.06 | 0.57 | 0.29  |
| E2VID+ (Stoffregen, ECCV’20) | **0.04** | **0.60** | **0.27** |
| FireNet+ (Stoffregen, ECCV’20) | 0.06 | 0.51 | 0.32  |
| E2VID_E (Ours)           | 0.07 | 0.52 | 0.38  |
| E2VID_F (Ours)           | 0.06 | 0.55 | 0.37  |
| FireNet+ (Ours)          | 0.06 | 0.52 | 0.38  |
| FireNet_E (Ours)         | 0.06 | 0.51 | 0.41  |

High Quality Frames (Stoffregen, ECCV’20)

| Method                  | MSE  | SSIM | LPIPS |
|-------------------------|------|------|-------|
| E2VID (Rebecq, TPAMI’19)| 0.14 | 0.46 | 0.45  |
| FireNet (Scheerlinck, WACV’20) | 0.07 | 0.48 | 0.42  |
| E2VID+ (Stoffregen, ECCV’20) | **0.03** | **0.57** | **0.26** |
| FireNet+ (Stoffregen, ECCV’20) | 0.05 | 0.47 | 0.36  |
| E2VID_E (Ours)           | 0.07 | 0.44 | 0.47  |
| E2VID_F (Ours)           | 0.06 | 0.48 | 0.47  |
| FireNet+ (Ours)          | 0.06 | 0.46 | 0.47  |
| FireNet_E (Ours)         | 0.06 | 0.46 | 0.51  |

Subscripts “F” and “E” indicate whether our networks were trained together with FireFlowNet or EV-Flownet.

Close to SOTA performance!
Results

Event-Camera Dataset (Mueggler, IJRR’17)

| Model                        | MSE | SSIM | LPIPS |
|------------------------------|-----|------|-------|
| E2VID (Rebecq, TPAMI’19)     | 0.08| 0.54 | 0.37  |
| FireNet (Scheerlinck, WACV’20)| 0.06| 0.57 | 0.29  |
| E2VID+ (Stoffregen, ECCV’20) | 0.04| 0.60 | 0.27  |
| FireNet+ (Stoffregen, ECCV’20)| 0.06| 0.51 | 0.32  |
| E2VID_F (Ours)               | 0.07| 0.52 | 0.38  |
| E2VID_E (Ours)               | 0.06| 0.55 | 0.37  |
| FireNet_F (Ours)             | 0.06| 0.52 | 0.38  |
| FireNet_E (Ours)             | 0.06| 0.51 | 0.41  |

High Quality Frames (Stoffregen, ECCV’20)

| Model                        | MSE | SSIM | LPIPS |
|------------------------------|-----|------|-------|
| E2VID (Rebecq, TPAMI’19)     | 0.14| 0.46 | 0.45  |
| FireNet (Scheerlinck, WACV’20)| 0.07| 0.48 | 0.42  |
| E2VID+ (Stoffregen, ECCV’20) | 0.03| 0.57 | 0.26  |
| FireNet+ (Stoffregen, ECCV’20)| 0.05| 0.47 | 0.36  |
| E2VID_F (Ours)               | 0.07| 0.44 | 0.47  |
| E2VID_E (Ours)               | 0.06| 0.48 | 0.47  |
| FireNet_F (Ours)             | 0.06| 0.46 | 0.47  |
| FireNet_E (Ours)             | 0.06| 0.46 | 0.51  |

Subscripts “F” and “E” indicate whether our networks were trained together with FireFlowNet or EV-Flownet.
**Results**

Event-Camera Dataset (Mueggler, IJRR’17)

| Method                        | MSE | SSIM | LPIPS |
|-------------------------------|-----|------|-------|
| E2VID (Rebecq, TPAMI’19)      | 0.08| 0.54 | 0.37  |
| FireNet (Scheerlinck, WACV’20)| 0.06| 0.57 | 0.29  |
| E2VID+ (Stoffregen, ECCV’20)  | 0.04| 0.60 | 0.27  |
| FireNet+ (Stoffregen, ECCV’20)| 0.06| 0.51 | 0.32  |
| E2VID_F (Ours)                | 0.07| 0.52 | 0.38  |
| E2VID_E (Ours)                | 0.06| 0.55 | 0.37  |
| FireNet_F (Ours)              | 0.06| 0.52 | 0.38  |
| FireNet_E (Ours)              | 0.06| 0.51 | 0.41  |

High Quality Frames (Stoffregen, ECCV’20)

| Method                        | MSE | SSIM | LPIPS |
|-------------------------------|-----|------|-------|
| E2VID (Rebecq, TPAMI’19)      | 0.14| 0.46 | 0.45  |
| FireNet (Scheerlinck, WACV’20)| 0.07| 0.48 | 0.42  |
| E2VID+ (Stoffregen, ECCV’20)  | 0.03| 0.57 | 0.26  |
| FireNet+ (Stoffregen, ECCV’20)| 0.05| 0.47 | 0.36  |
| E2VID_F (Ours)                | 0.07| 0.44 | 0.47  |
| E2VID_E (Ours)                | 0.06| 0.48 | 0.47  |
| FireNet_F (Ours)              | 0.06| 0.46 | 0.47  |
| FireNet_E (Ours)              | 0.06| 0.46 | 0.51  |

Subscripts “F” and “E” indicate whether our networks were trained together with FireFlowNet or EV-Flownet.

Close to SOTA performance!
Results

**Event-Camera Dataset (Mueggler, IJRR’17)**

| Model                  | MSE  | SSIM | LPIPS |
|------------------------|------|------|-------|
| E2VID (Rebecq, TPAMI’19) | 0.08 | 0.54 | 0.37  |
| FireNet (Scheerlinck, WACV’20) | 0.06 | 0.57 | 0.29  |
| E2VID+ (Stoffregen, ECCV’20) | **0.04** | **0.60** | **0.27** |
| FireNet+ (Stoffregen, ECCV’20) | 0.06 | 0.51 | 0.32  |
| E2VID_F (Ours)          | 0.07 | 0.52 | 0.38  |
| E2VID_E (Ours)          | 0.06 | 0.55 | 0.37  |
| FireNet_F (Ours)        | 0.06 | 0.52 | 0.38  |
| FireNet_E (Ours)        | 0.06 | 0.51 | 0.41  |

**High Quality Frames (Stoffregen, ECCV’20)**

| Model                  | MSE  | SSIM | LPIPS |
|------------------------|------|------|-------|
| E2VID (Rebecq, TPAMI’19) | 0.14 | 0.46 | 0.45  |
| FireNet (Scheerlinck, WACV’20) | 0.07 | 0.48 | 0.42  |
| E2VID+ (Stoffregen, ECCV’20) | **0.03** | **0.57** | **0.26** |
| FireNet+ (Stoffregen, ECCV’20) | 0.05 | 0.47 | 0.36  |
| E2VID_F (Ours)          | 0.07 | 0.44 | 0.47  |
| E2VID_E (Ours)          | 0.06 | 0.48 | 0.47  |
| FireNet_F (Ours)        | 0.06 | 0.46 | 0.47  |
| FireNet_E (Ours)        | 0.06 | 0.46 | 0.51  |

Subscripts “F” and “E” indicate whether our networks were trained together with FireFlowNet or EV-Flownet.

Close to SOTA performance!
Conclusion

- We presented the first self-supervised learning-based approach to event-based image reconstruction.

- The framework can be extended in multiple ways (architectures, losses, optical flow algorithms, etc.).
  - Architectures
  - Optical flow algorithms
  - Other regularizers