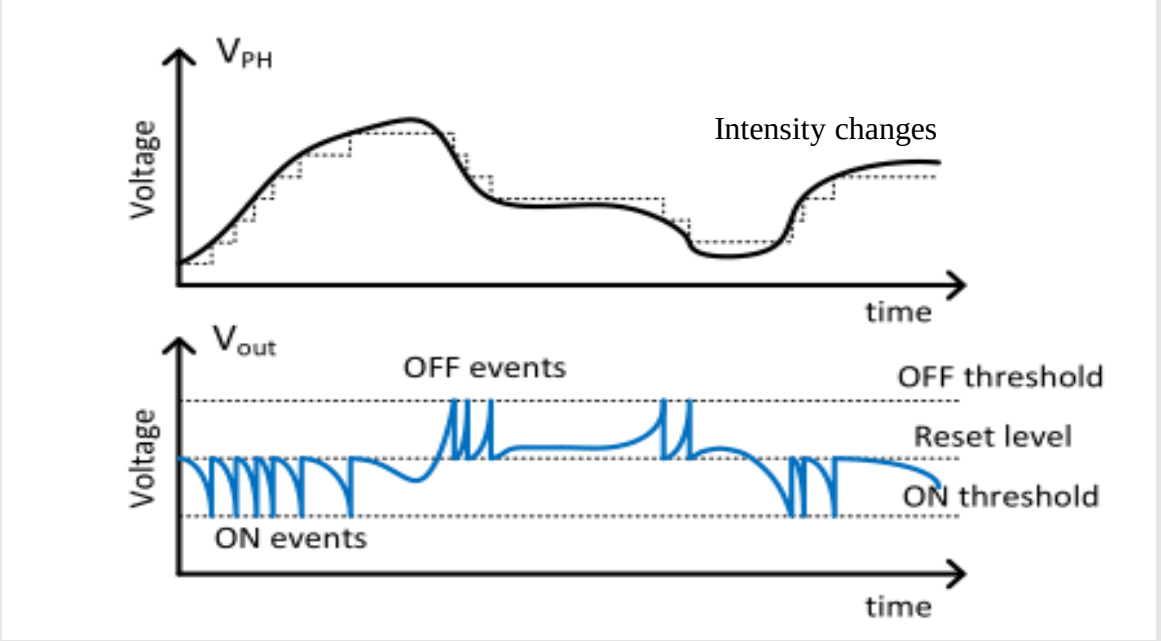
ICASSP 2020

Intensity-Image Reconstruction For Event Cameras Using Convolutional Neural Network

Yongwei Chen, Weitong Chen, Xixin Cao, Qianting Hua
Peking University

Introduction

- **What is an event camera?**
- Dynamic Vision Sensor (DVS)
- Pixels measure intensity changes asynchronously and independently
- Event （location X,Y， Polarity，Time）



Generate events at a pixel when intensity changes exceed the threshold

Working principle of event cameras

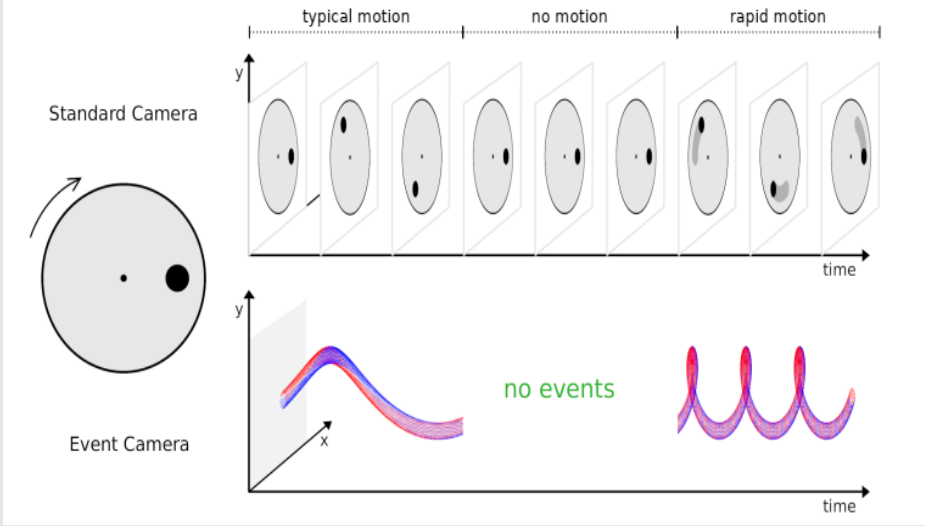
Introduction

Advantages of event cameras

- High dynamic range
- No motion blur
- low redundancy
- Low latency (μs)
- High temporal resolution (μs)

Purpose of research:

- Advantages of using event cameras.
- Compatible with existing frame-based algorithms
- Screen display and human observation



Event cameras VS standard cameras

HDR

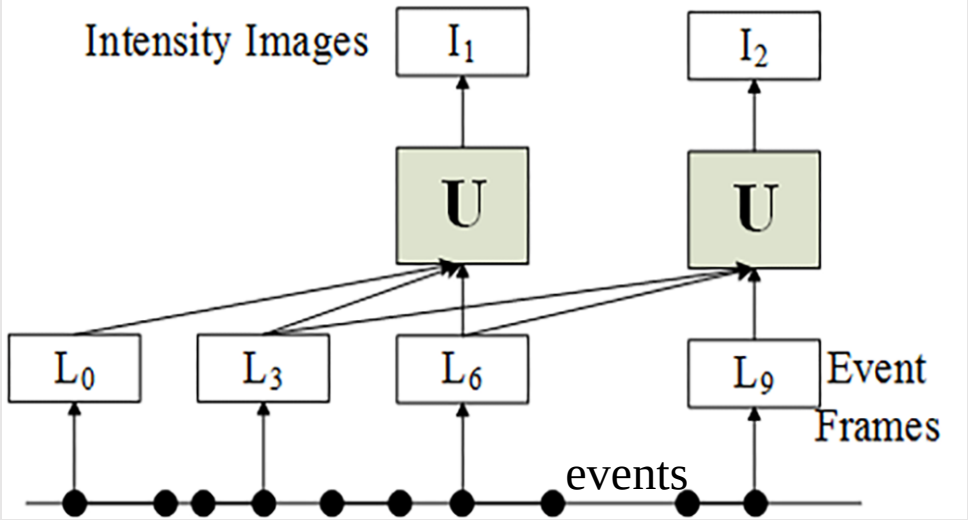
Method

Two steps

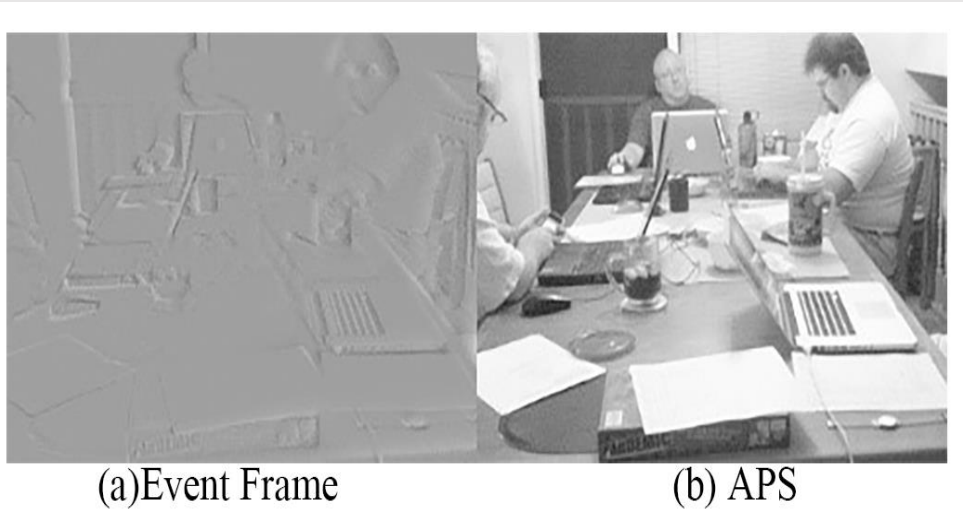
- Convert the events into event frames
- U-net (U) network transforms event frames into intensity images (I).

Some details

- The U-net is trained on simulated datasets.
- The simulator ESIM generate event streams and time-stamped frame images(APS).
- Perceptual loss



Overview of our approach



A pair of event frame (one channel of the input image) and APS image (the ground truth) in the training dataset.

Method

An attenuation method to generate event frames.

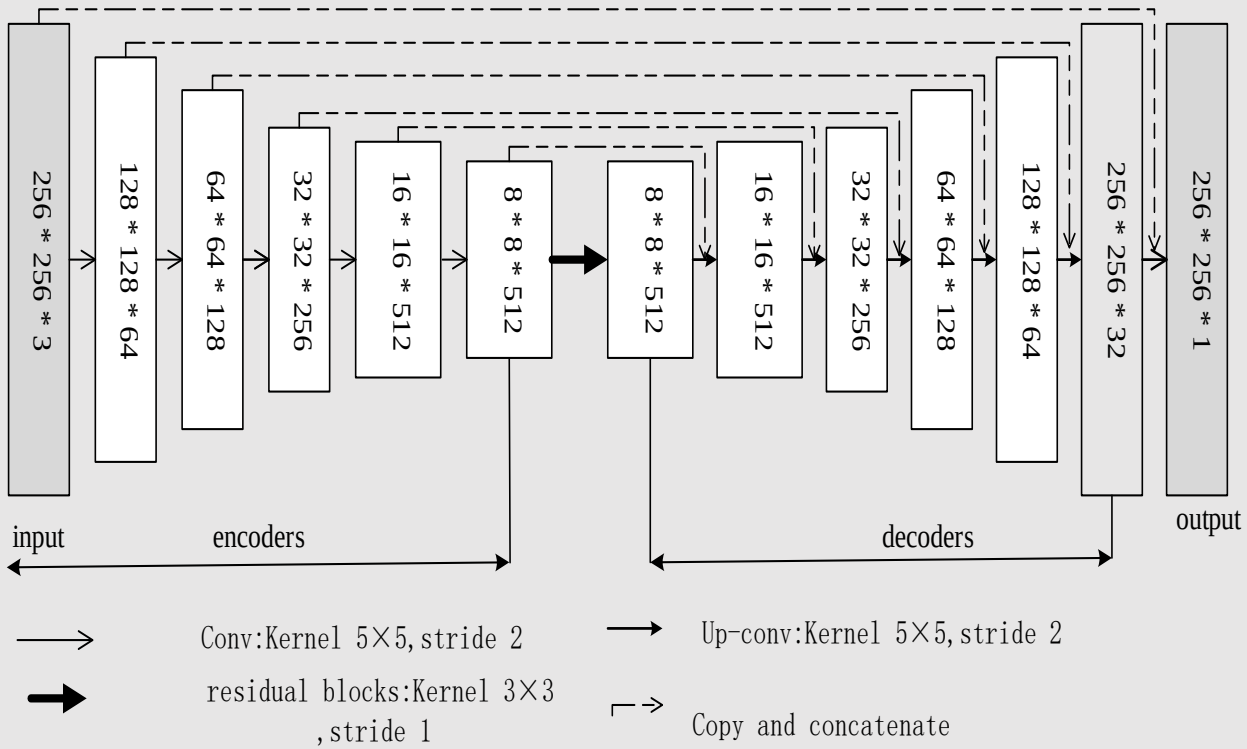
- If the same pixel (x, y) generates several events e_0, e_1, e_2 ... $e_k \ldots e_n$.
- The threshold of e_k is expressed as C_k.
- $L_k(x, y)$is the intensity value of e_k
- β is an attenuation coefficient less than 1.
- $L_k(x, y)$ is affected by all events that are generated by pixel (x, y)
- The magnitude of influence is reflected in β^n.
- Historical information is attenuated.

$$
\begin{cases}
L_0(x,y) = C_0 \\
L_1(x,y) = L_0(x,y) \times \beta + C_1 \\
L_k(x,y) = L_{k-1}(x,y) \times \beta + C_k \\
L_n(x,y) = L_{n-1}(x,y) \times \beta + C_n \\
= C_0 \times \beta^n + C_1 \times \beta^{n-1} + \ldots C_k \times \beta^{n-k} + \ldots + C_n \times \beta^0
\end{cases}
$$

Method

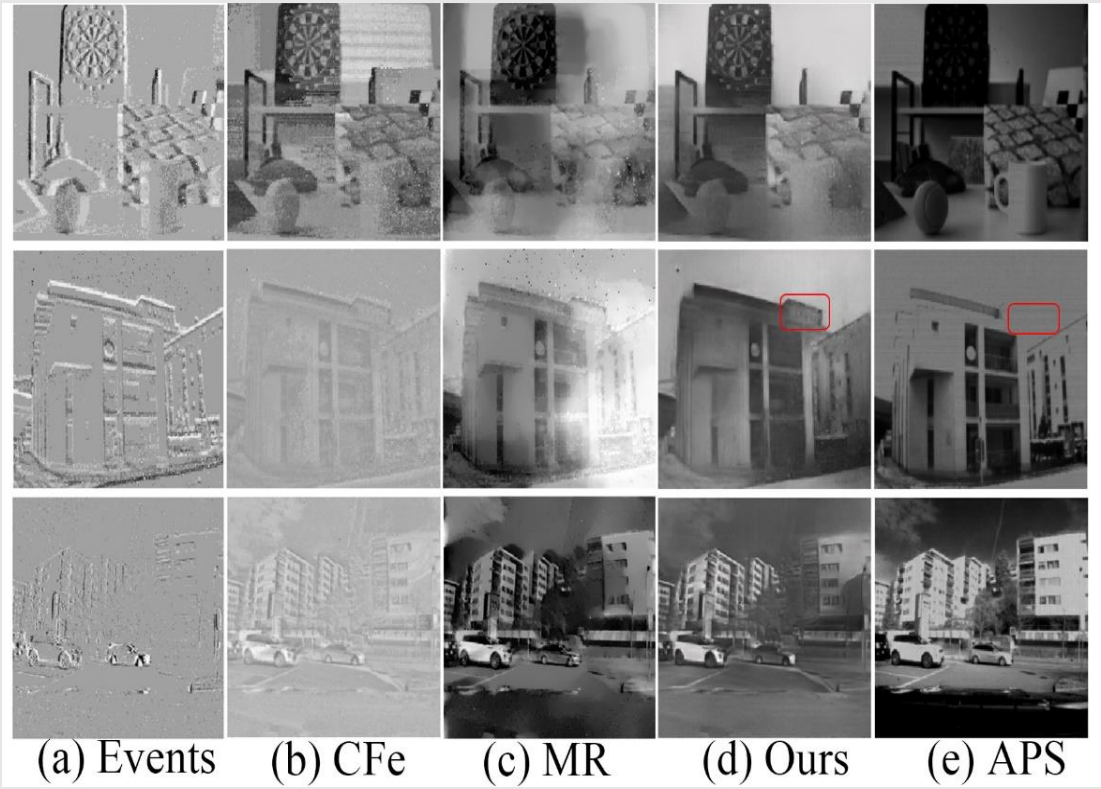
U-net network

- The network has 5 encoders, 1 residual block, 5 decoders and a final convolutional layer
- Skip connections



U-net

Experiments and Evaluation



Generated images of different model. The column (a) shows events, and the column (e) shows APS images of DAVIS for reference, and other columns show the generated images.

Experiments and Evaluation

Table 1 *quantitative comparison using SSIM and FSIM*

Model	SSIM	FSIM
CFe	0.55	0.68
MR	0.64	0.71
Ours	0.78	0.89

(a)Events (b)MR (c)CFe (d)Ours

Result in noise dataset.

Thanks for listening！

Intensity-Image Reconstruction For Event Cameras Using Convolutional Neural Network