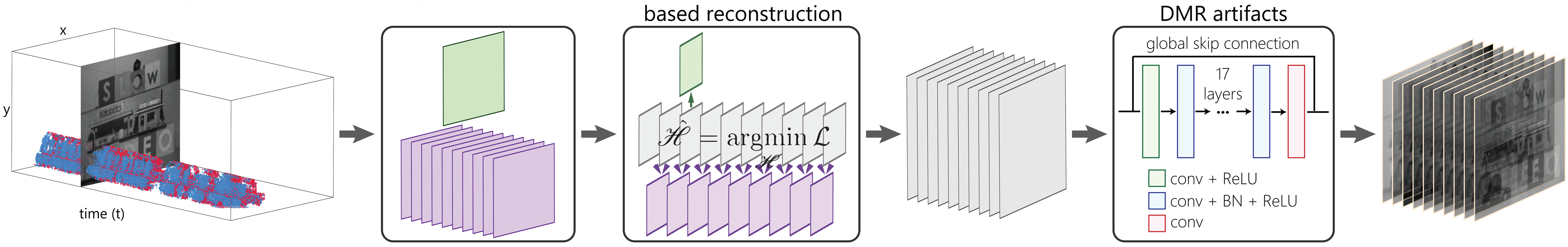
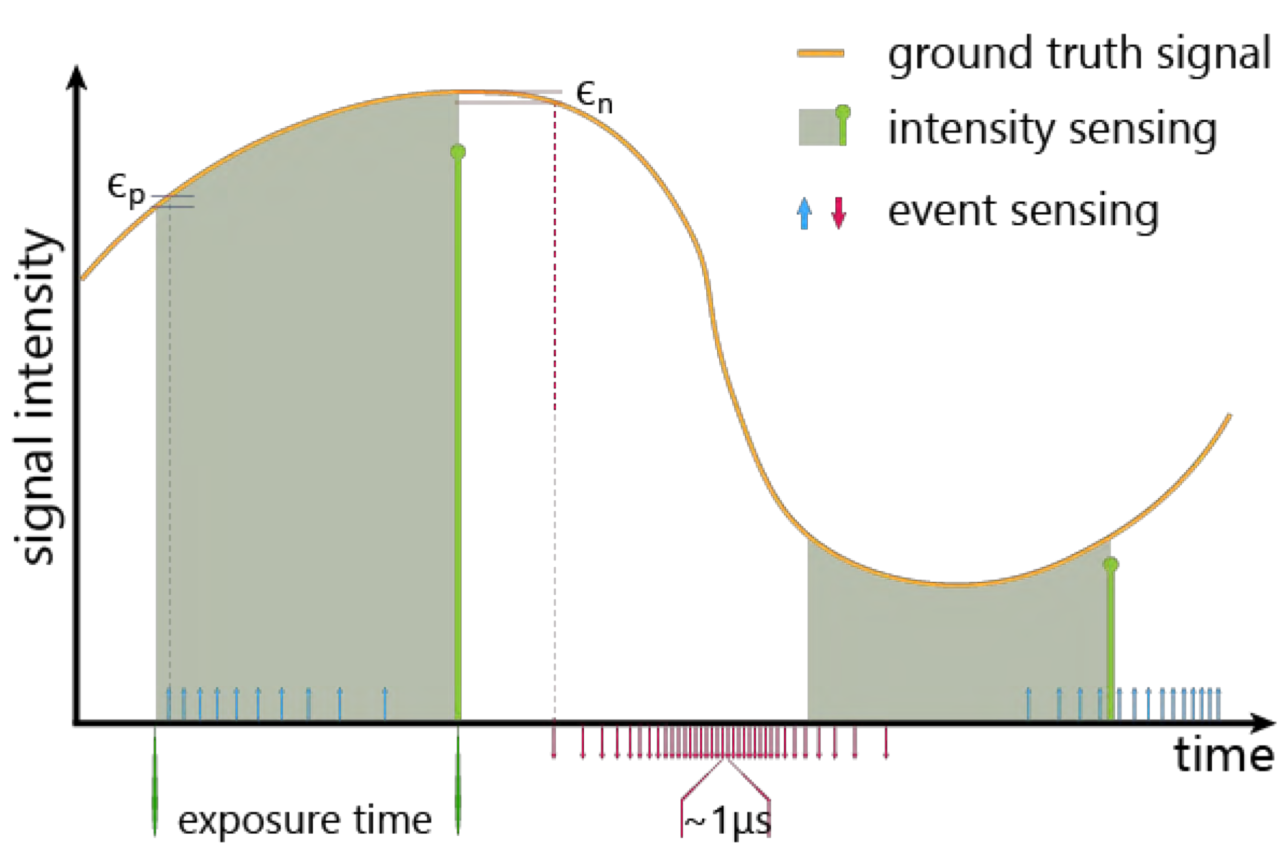
Event-driven Video Frame Synthesis

ZW. Wang[1], W. Jiang[1], K. He[1], B. Shi[2], A. Katsaggelos[1], O. Cossairt[1]
[1]Northwestern University [2]Peking University

captured data intensity image + events | **pre-processing** binning events to frames | **DMR** differentiable model-based reconstruction | **Residual "denoising"** learning to remove DMR artifacts | **final output** synthesized video



What's event camera?



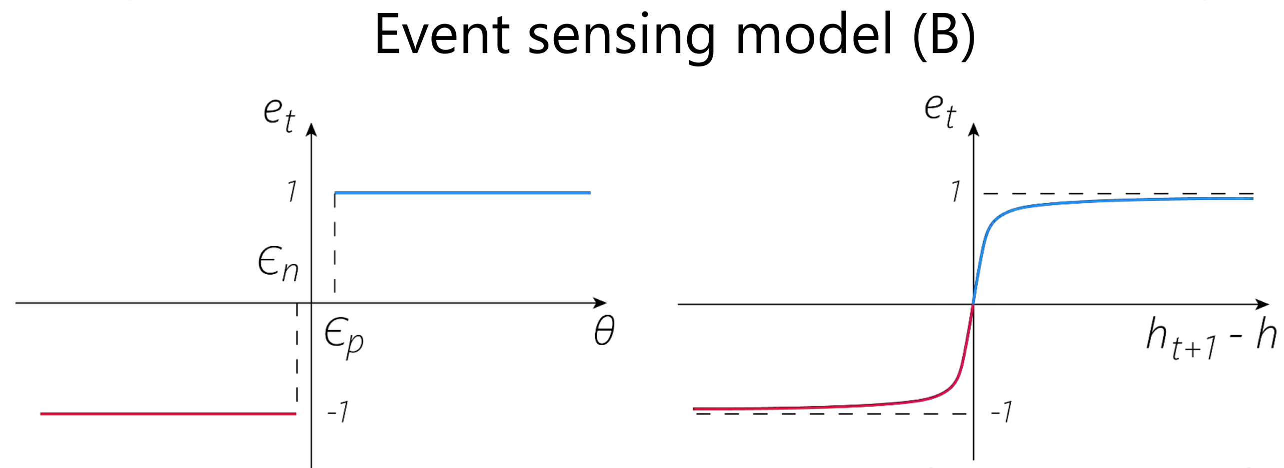
Conventional camera	Event camera
Synchronized frames I(x, y)	Independent pixels (x, y, p, t)
Intensity values	Brightness changes (binary)
Low speed: 30, 60 fps	High speed: <1us latency
Low dynamic range (60-90dB)	High dynamic range (120dB)
High power consumption	Low power consumption
Low noise	Severe noise

Intensity + events: differentiable modeling

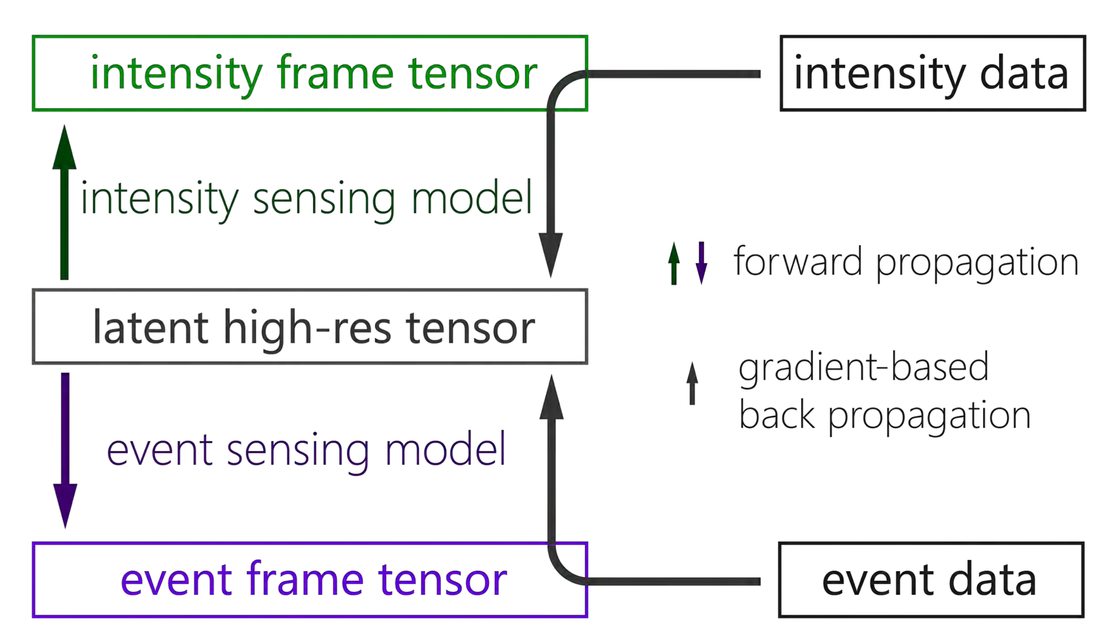
Intensity sensing model (A)

Case (intensity tensor notation)	Model
Interpolation ($\mathscr{F}^i$)	$\mathscr{F}_1^i = \mathscr{H}_1, \mathscr{F}_2^i = \mathscr{H}_d$
Prediction ($\mathscr{F}^p$)	$\mathscr{F}^p = \mathscr{H}_1$
Motion deblur ($\mathscr{F}^m$)	$\mathscr{F}^m = \frac{1}{d}\sum_{t=1}^{d} \mathscr{H}_t$

Event sensing model (B)



$$\mathscr{E}_t = \tanh\left\{\alpha[\mathscr{H}_{t+1} - \mathscr{H}_t]\right\}$$

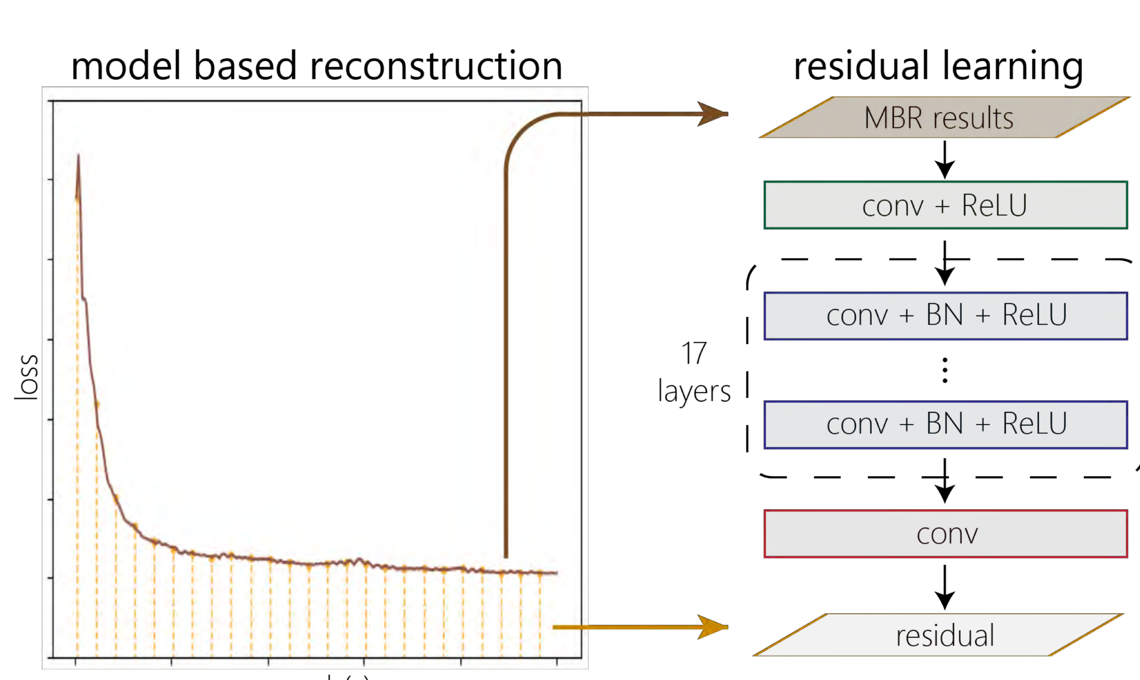


Objective and loss functions

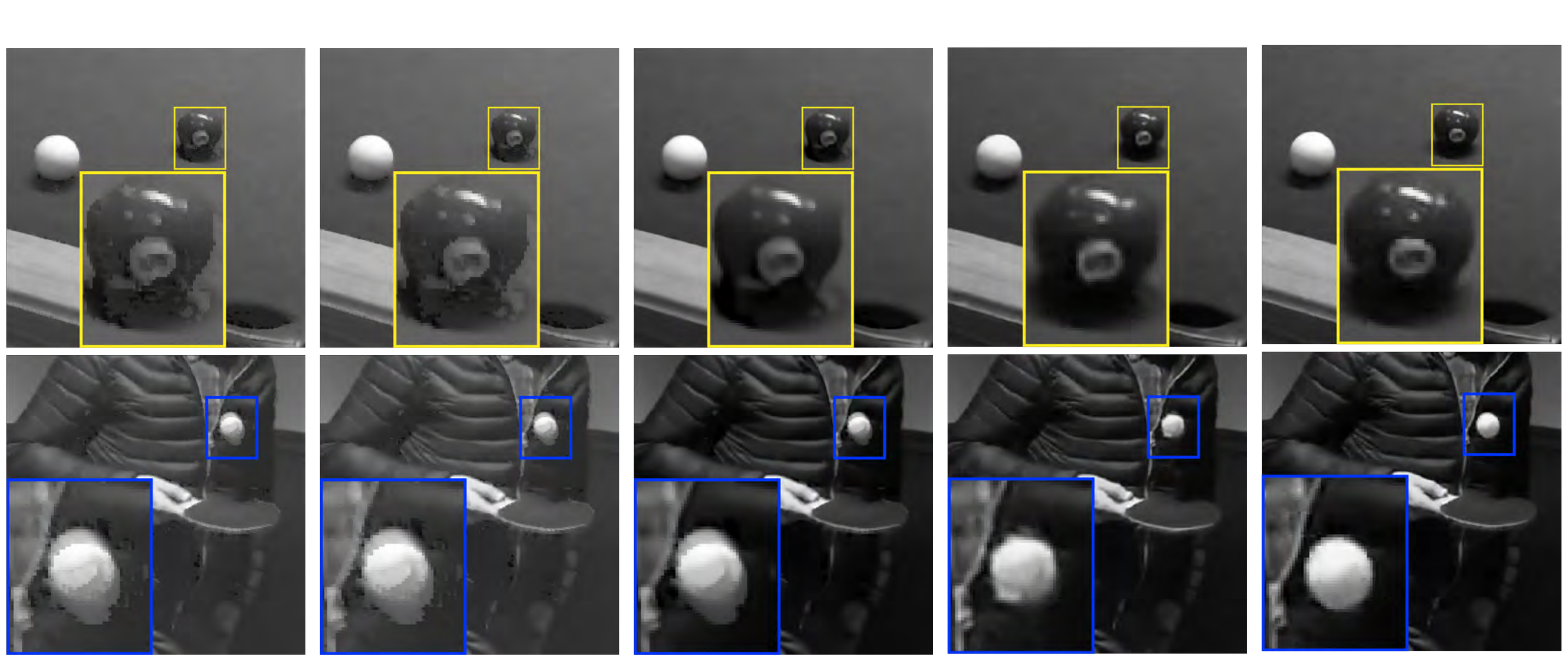
$$\hat{\mathscr{H}} = \operatorname*{argmin}_{\mathscr{H}} \mathcal{L}_{pix}(\mathscr{H}, \mathcal{F}, \mathcal{E}) + \mathcal{L}_{TV}(\mathscr{H}) \quad \mathcal{L}_{pix}(\mathscr{H}, \mathcal{F}, \mathcal{E}) = \mathbb{E}_{fpix}[\|\mathcal{F} - \mathcal{A}(\mathscr{H})\|_1] + \lambda_e \mathbb{E}_{epix}[\|\mathcal{E} - \mathcal{B}(\mathscr{H})\|_1]$$

$$\mathcal{L}_{TV}(\mathscr{H}) = \lambda_{xy}\mathbb{E}_{hpix}\left[\left\|\dot{\mathscr{H}}_{xy}\right\|_1\right] + \lambda_t \mathbb{E}_{hpix}\left[\left\|\dot{\mathscr{H}}_t\right\|_1\right]$$

Enhancement via residual learning



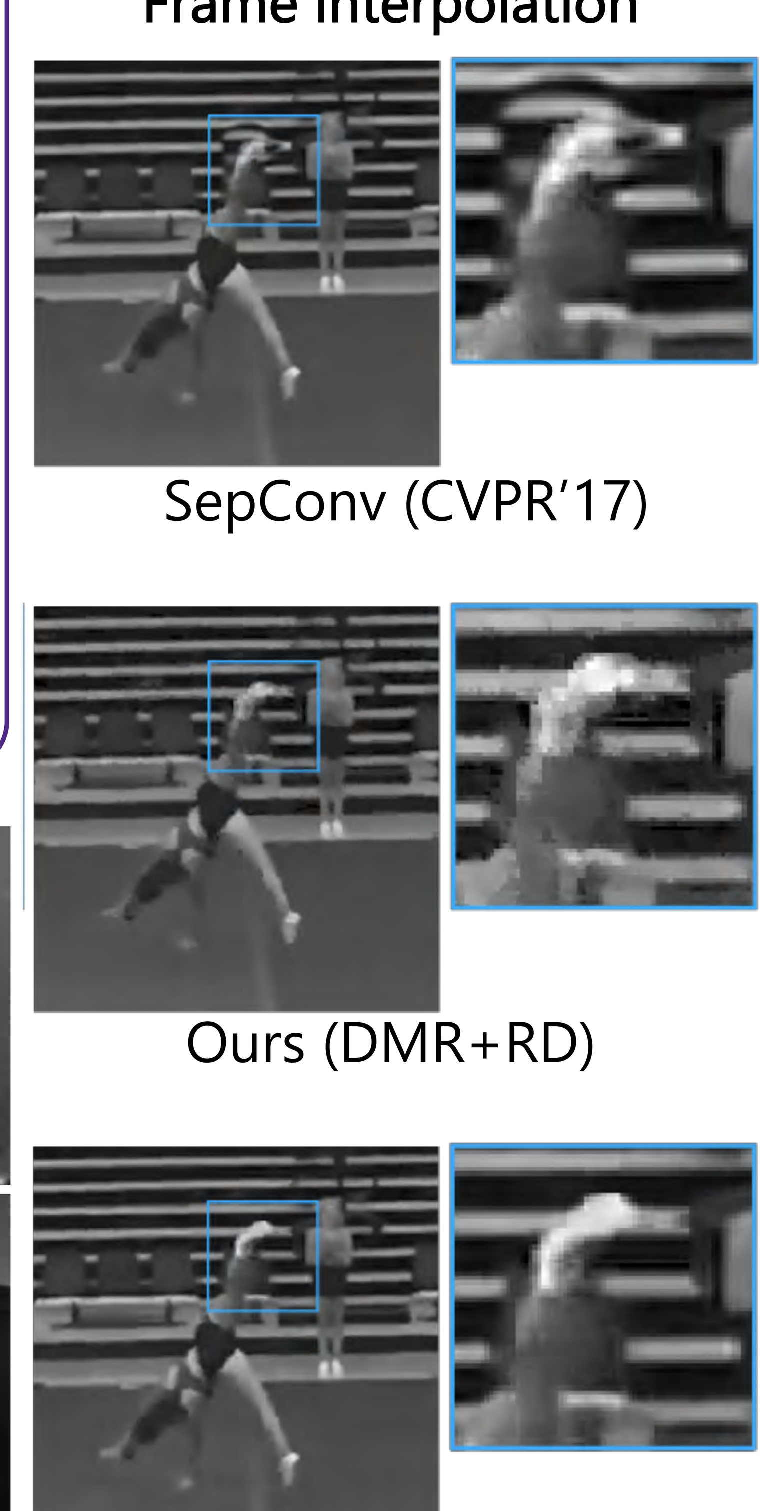
- Use CNN to learn the residual of DMR output w.r.t. ground truth
  - Designed to enhance DMR results
- Easy to train
  - Model DMR artifacts as residual "noise"
  - Actually beyond Gaussian denoising
- Single frame based
  - Interface well with DMR



Ours (DMR) | DnCNN | FFTNet | Ours (RD) | Ground truth

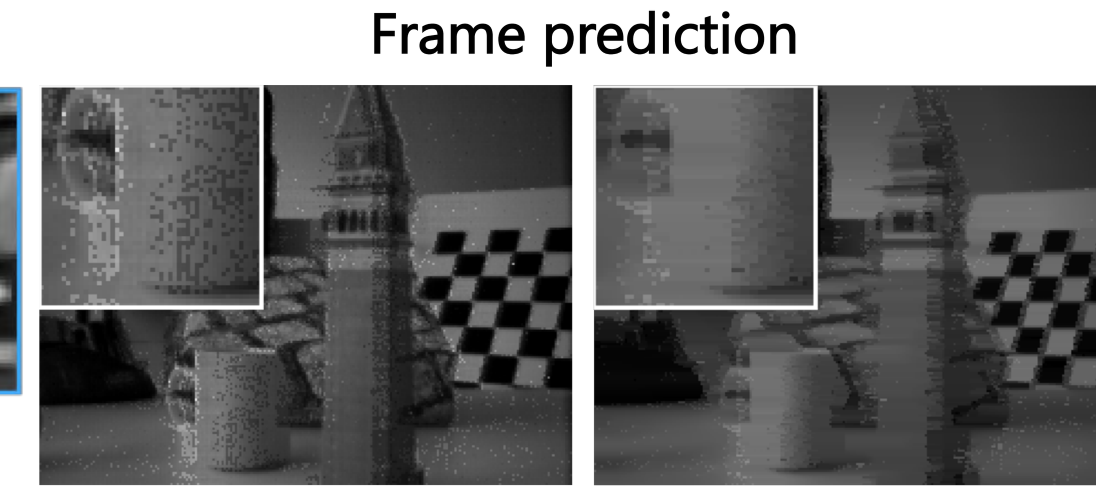
Results

Frame interpolation



SepConv (CVPR'17) | Ours (DMR+RD) | Ground truth

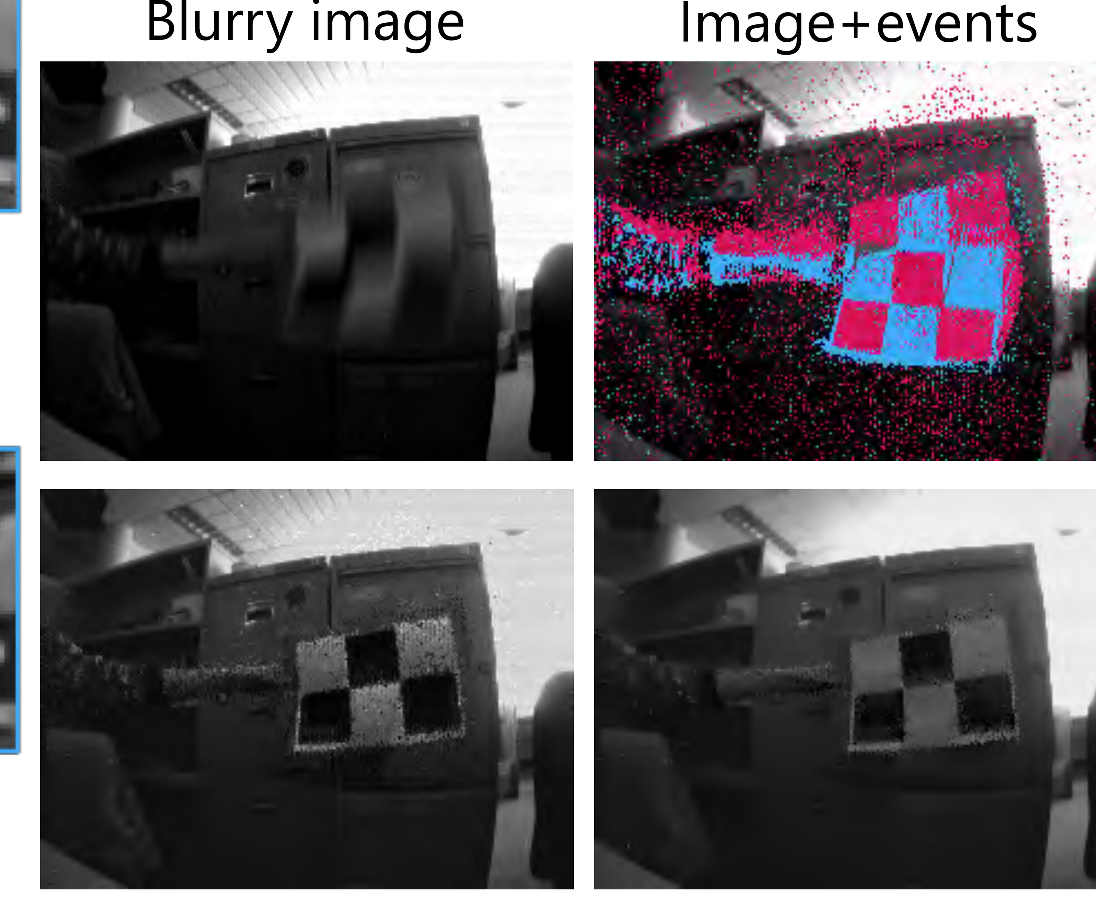
Frame prediction



PSNR: 23.33 SSIM: .771 — CF (ACCV'18) | PSNR: 25.12 SSIM: .831 — Ours (RD)

Motion deblur

Blurry image | Image+events



EDI (CVPR'19) | Ours (DMR)