



Camera Intelligence

A Computational Photography Lab @ PKU



Spatio-Temporal Interactive Learning for Efficient Image Reconstruction of Spiking Cameras

Bin Fan¹ # Jiaoyang Yin^{2,3} # Yuchao Dai⁴ Chao Xu¹ Tiejun Huang^{2,3} Boxin Shi^{2,3}

¹ Nat'l Key Lab of General AI, School of Intelligence Science and Technology, Peking University

² State Key Lab of Multimedia Info. Processing, School of Computer Science, Peking University

³ Nat'l Eng. Research Ctr. of Visual Technology, School of Computer Science, Peking University

⁴ School of Electronics and Information, Northwestern Polytechnical University

Contributed equally to this work as first authors

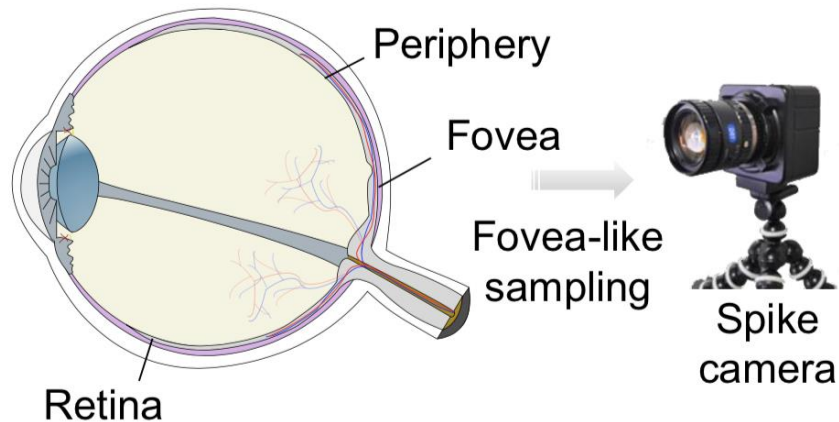


Spiking Camera Introduction

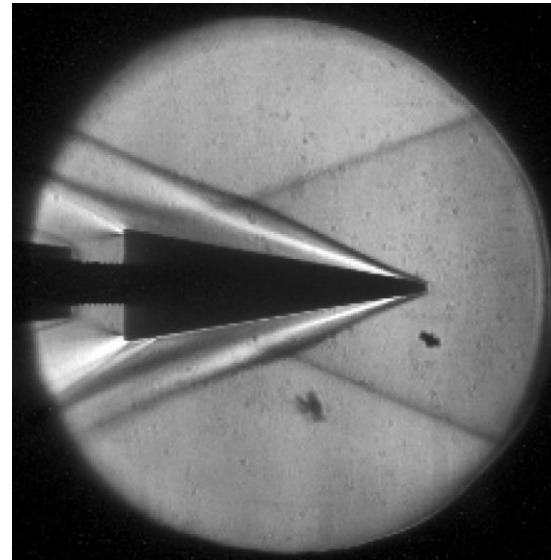


Designed by Peking university, the spiking camera achieves integral sampling with 40,000Hz by emulating the central fovea's sampling mechanism in the retina.

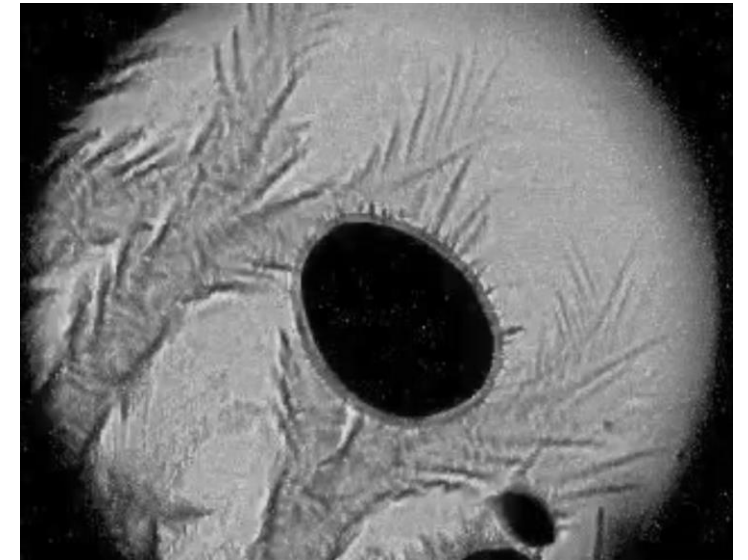
- High temporal resolution
- High dynamic range
- Low latency

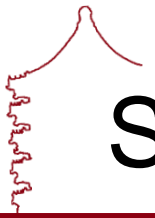


Hypersonic flow in aerodynamics

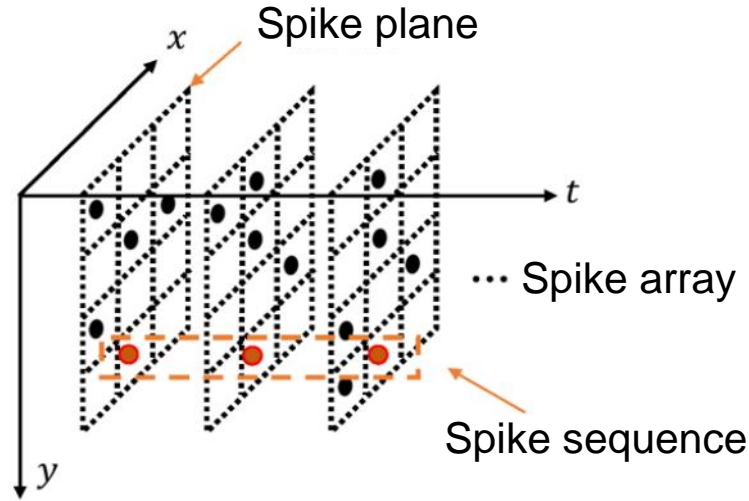


Phase change process in liquid metal

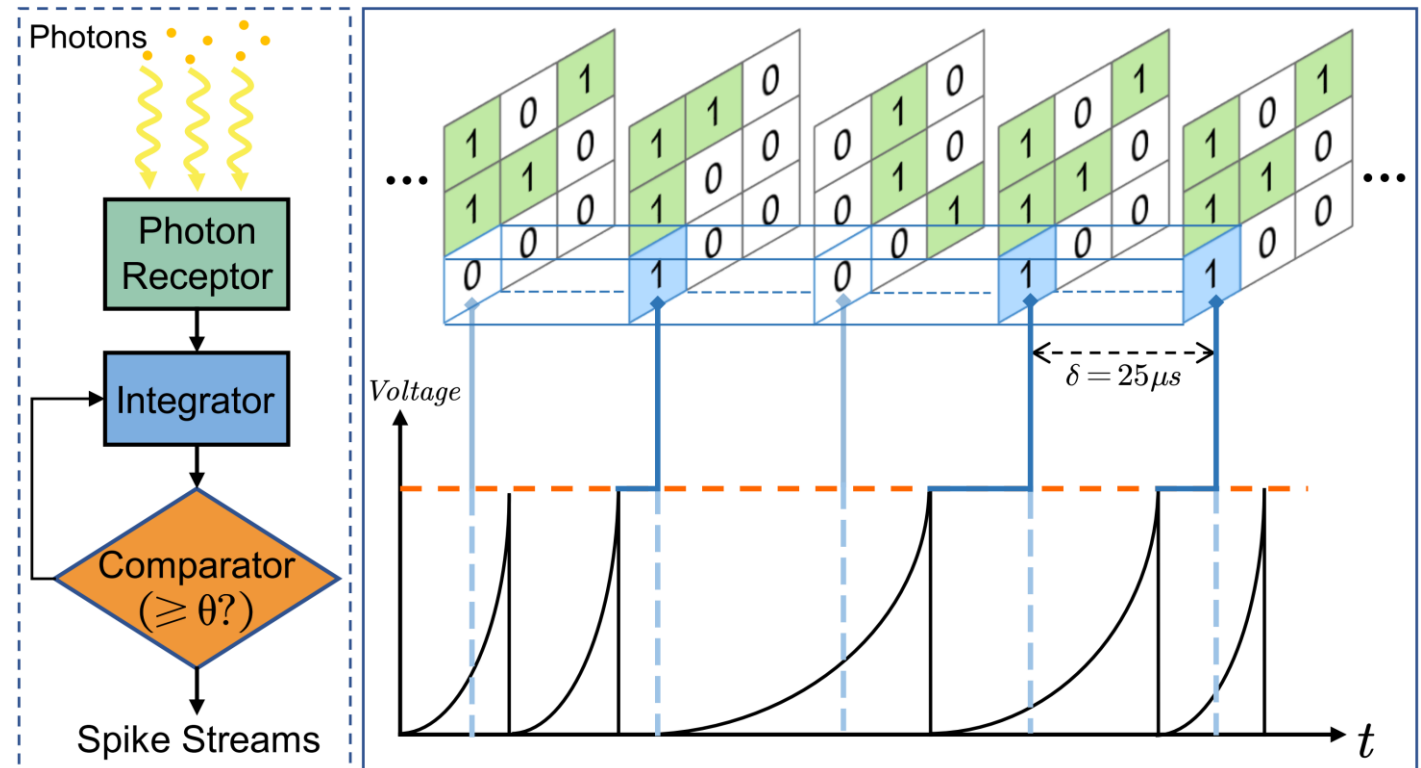




Spike Camera Working Mechanism



For a pixel, the light intensity is accumulated, if the accumulated intensity reaches the dispatch threshold θ , a **spike** is fired and the accumulator is reset.



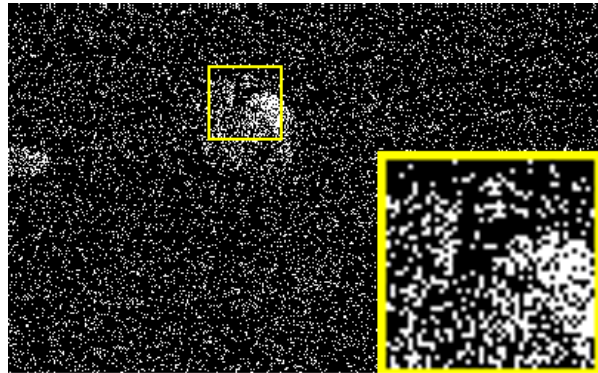


Prior Works



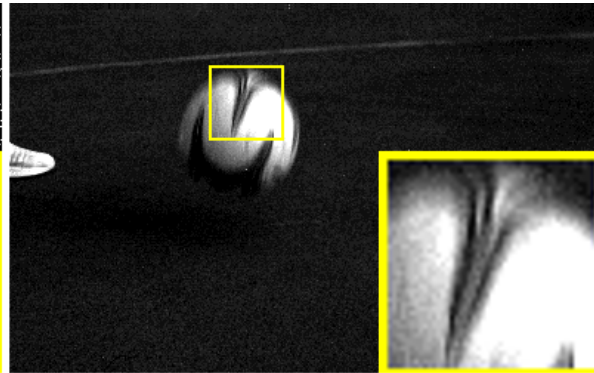
Conventional method

Spike Streams



TFP

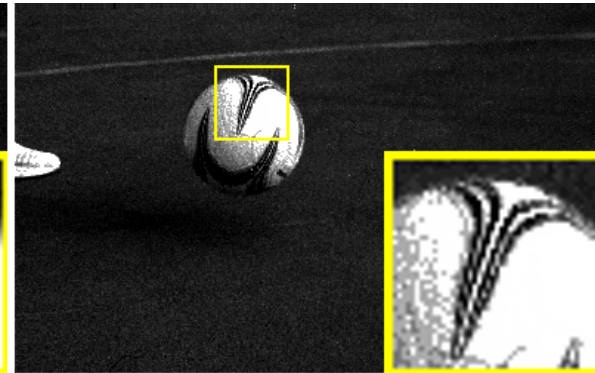
Texture from playback



- HDR imaging
- **Dependent on manual selection of window length**
- **Motion blur or noise**

TFI

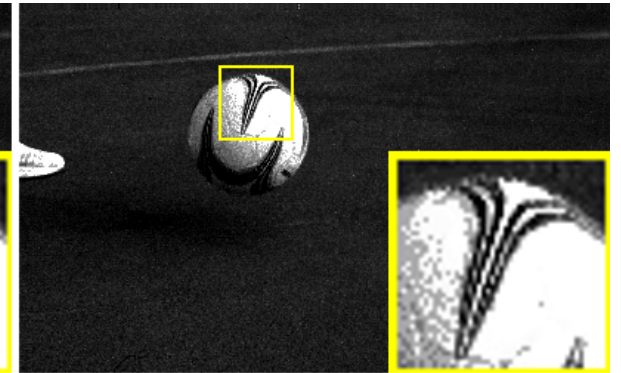
Texture from Inter-spike Interval



- Capable of handling high-speed scenes
- **Loss of texture detail**

TFSTP, etc

Other methods that mimic human visual mechanisms



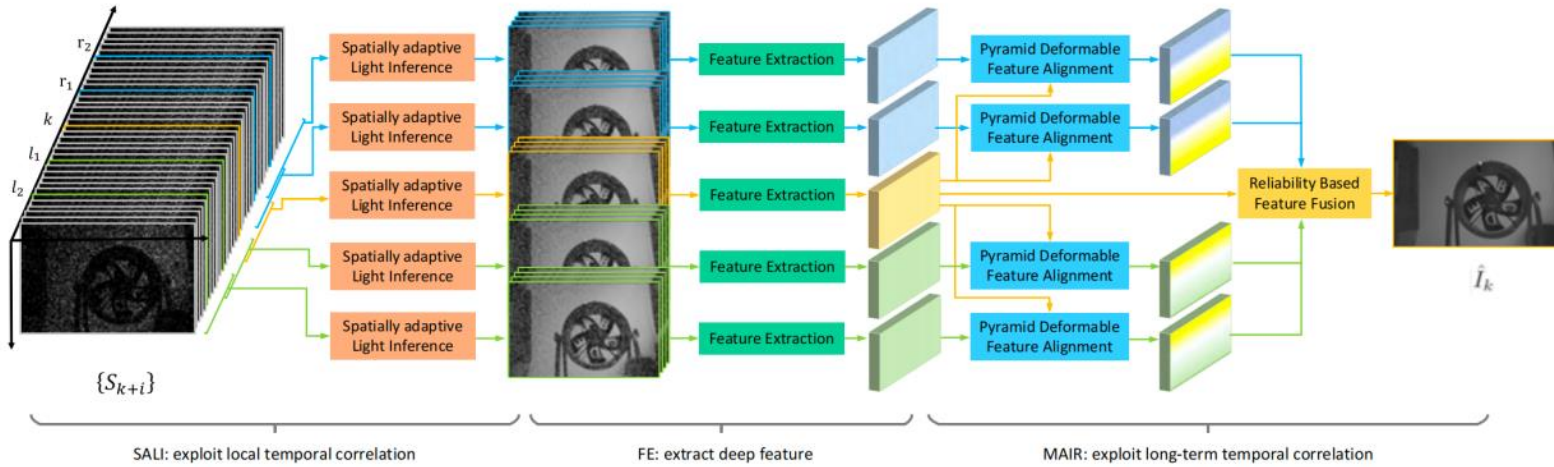
- **Complex algorithm and hard to replicate**
- **Unsatisfactory results with noise**



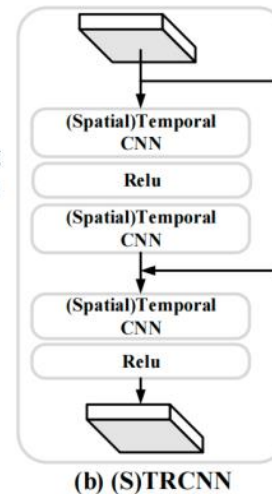
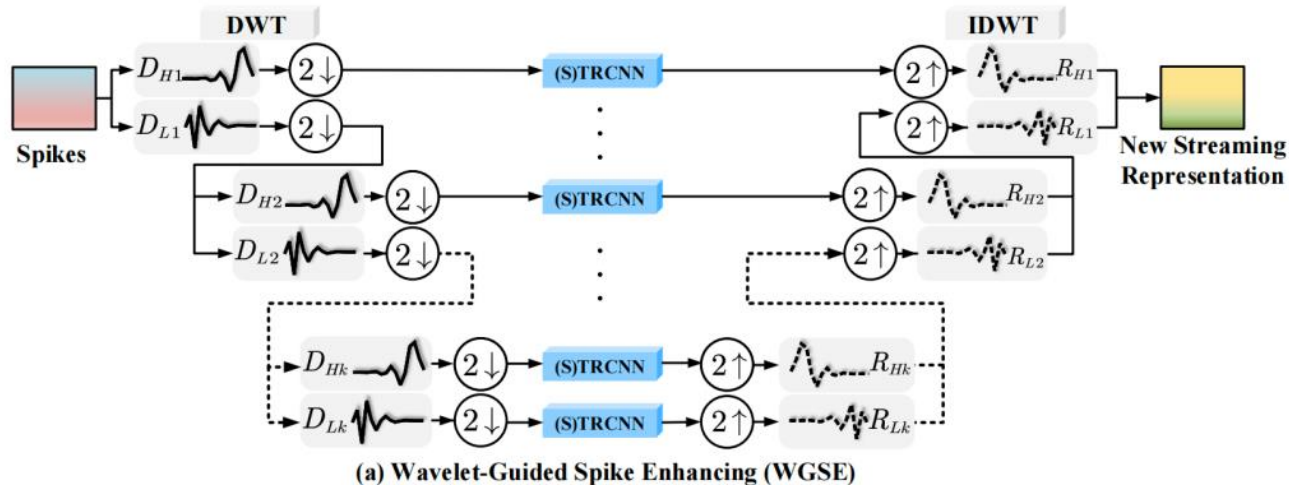
Related Works



Learning-based method



Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream [Zhu *et al.*, CVPR 21]



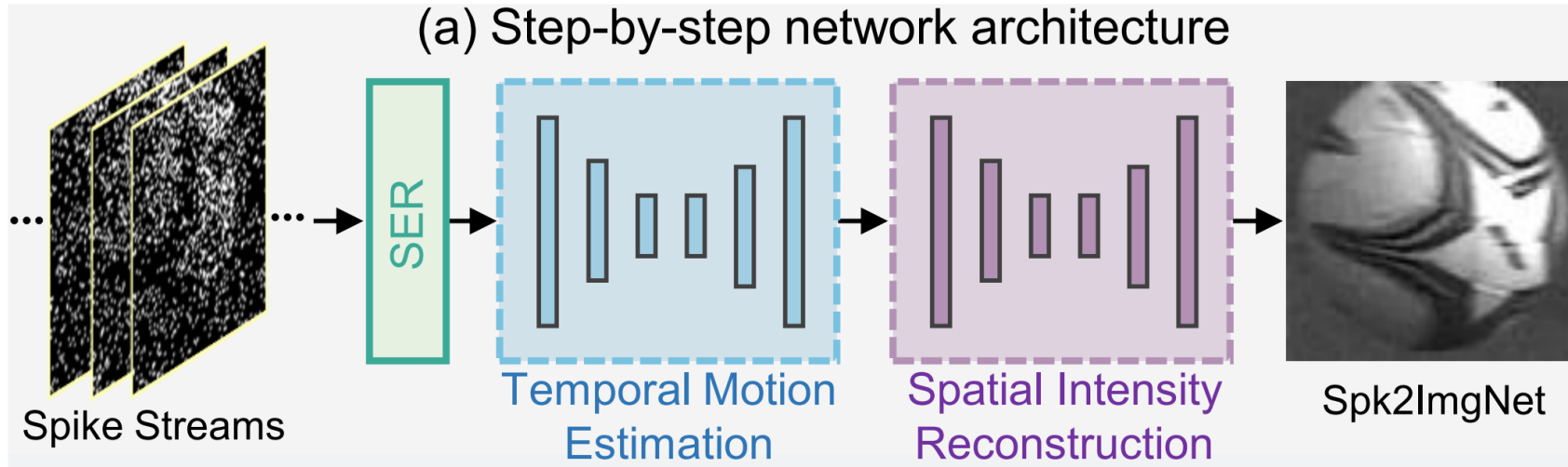
WGSE: Learning temporal-ordered representation for spike streams based on discrete wavelet transforms [Zhang *et al.*, AAAI 23]



Motivation

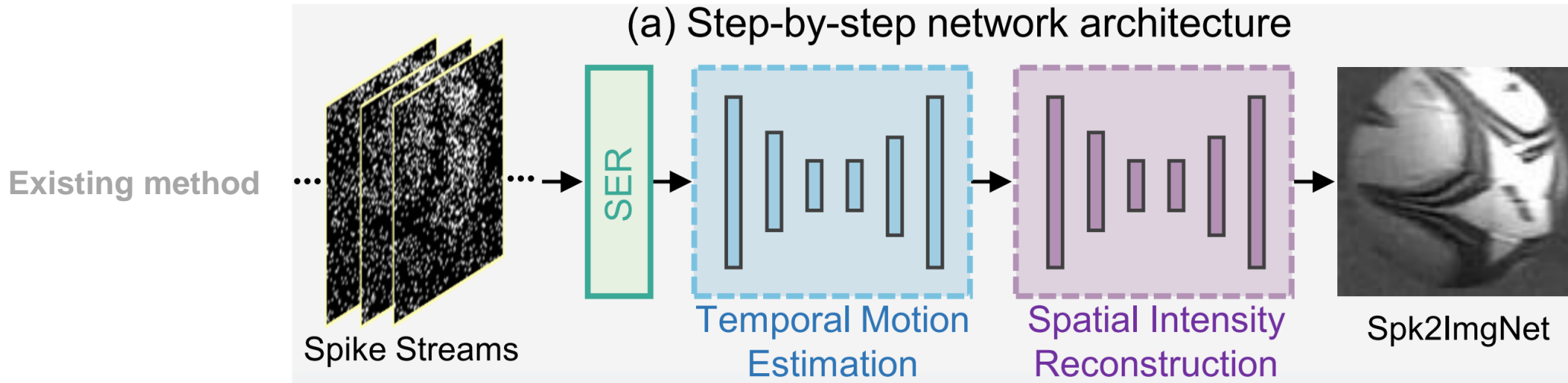


Existing method ...





Motivation

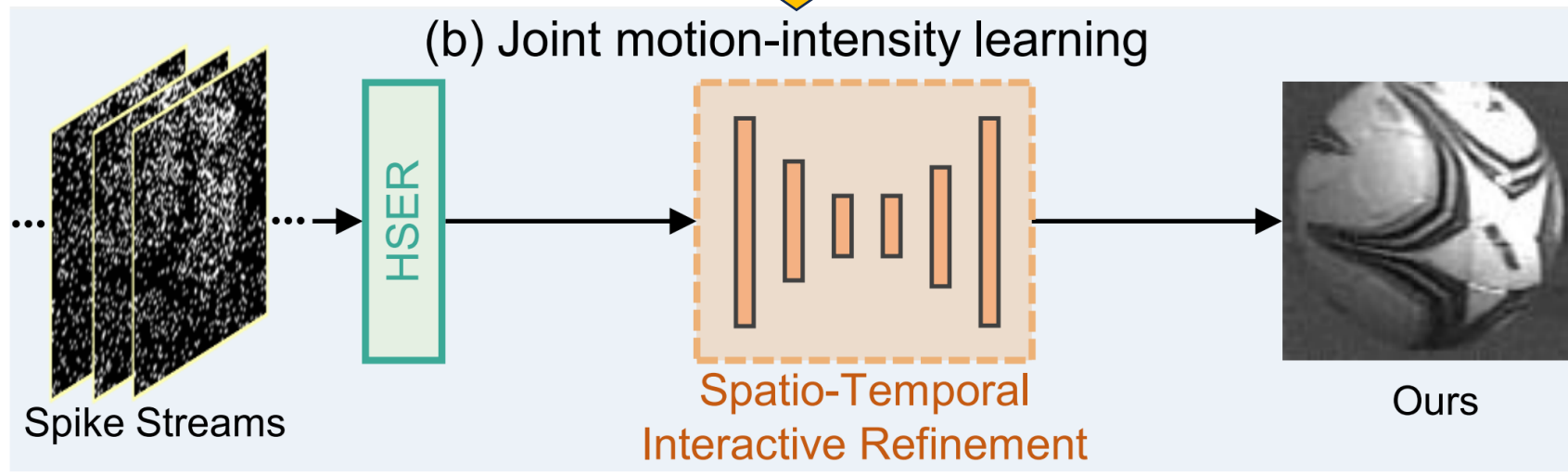
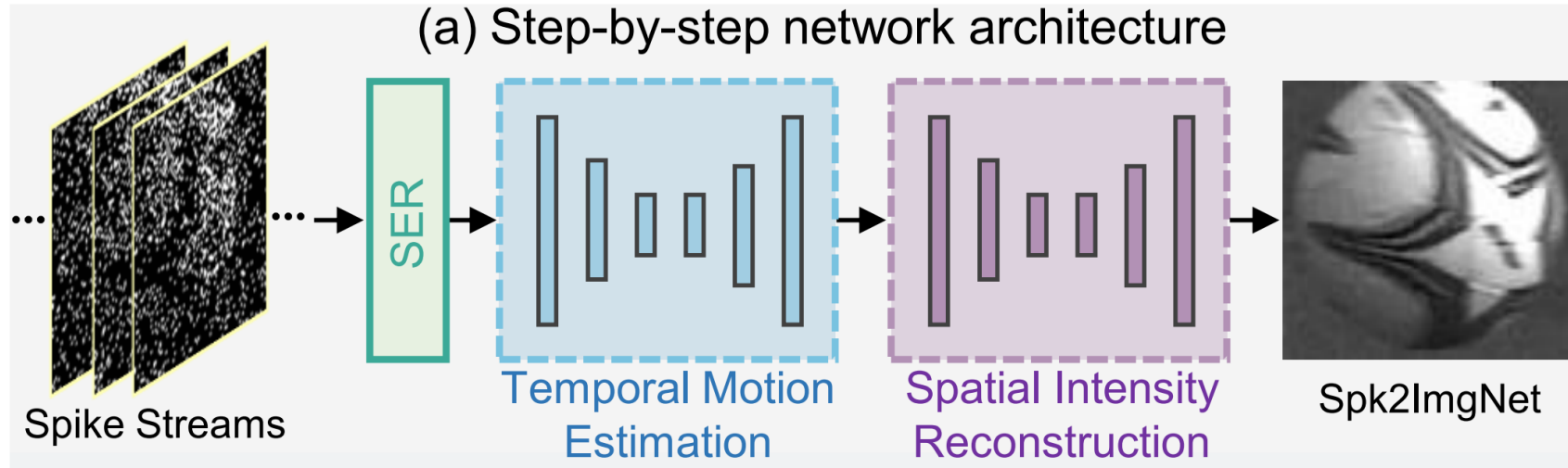


High model complexity and low inference efficiency

Fails to address the “Chicken-Egg Problem”



Motivation

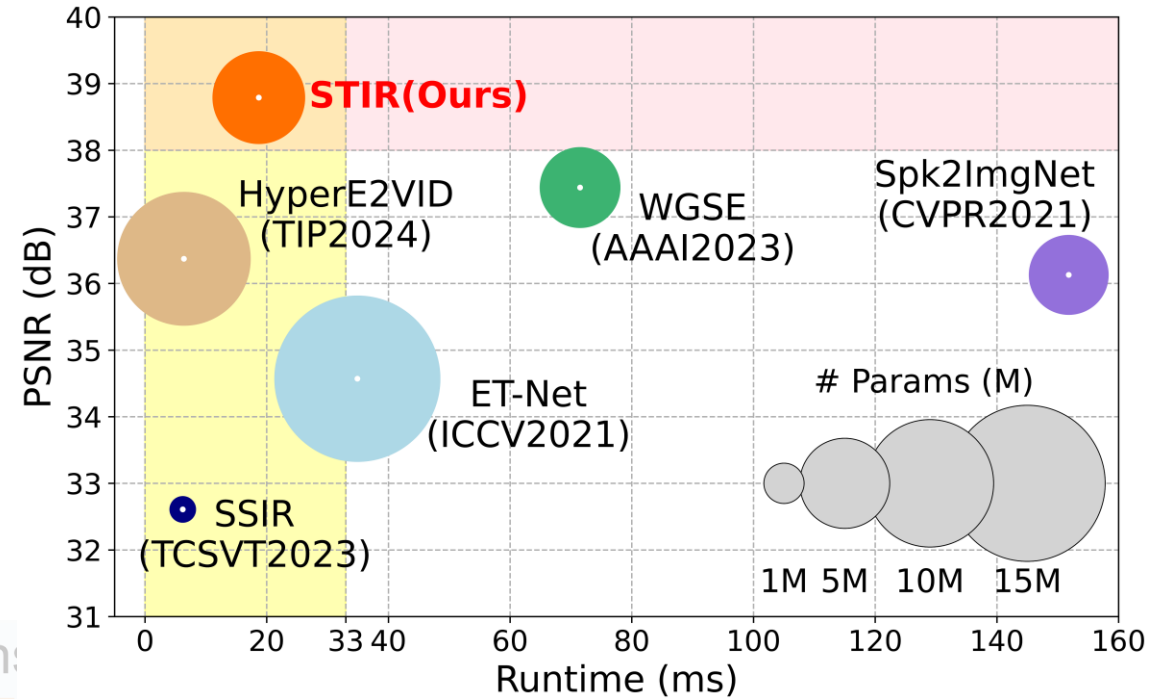




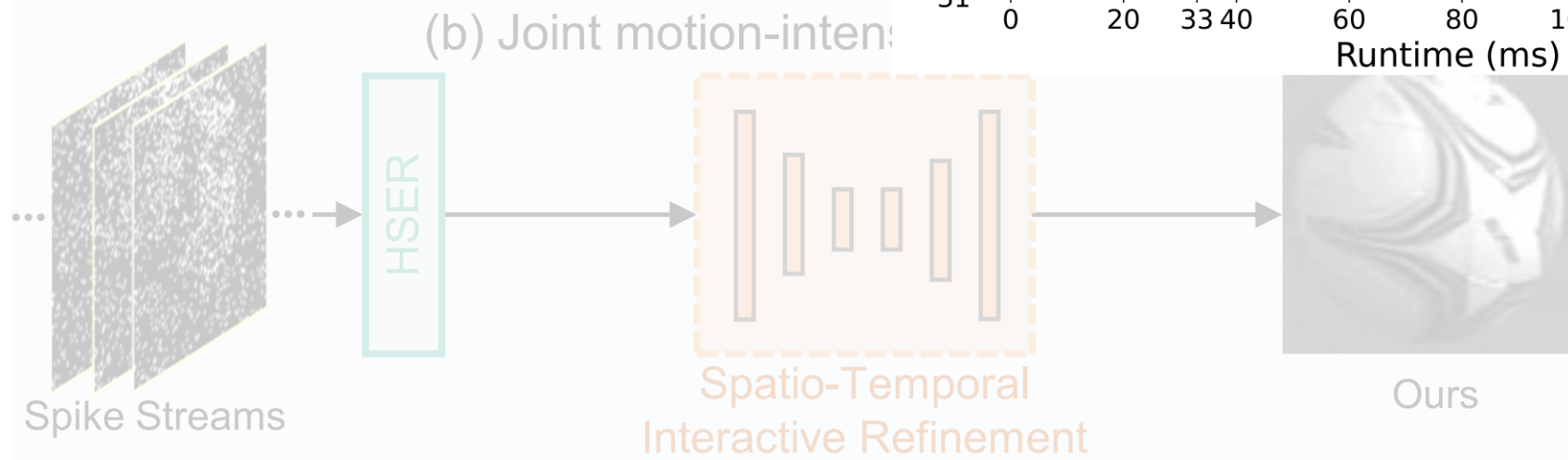
Motivation



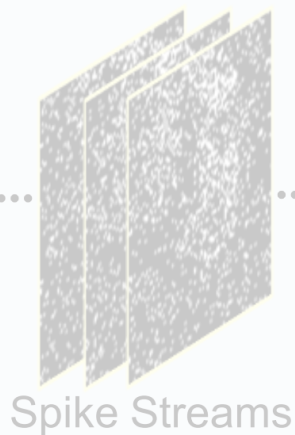
- Lower model complexity and faster inference
- Higher image reconstruction quality
- More conducive to real-time application deployment



(b) Joint motion-intens



Proposed method

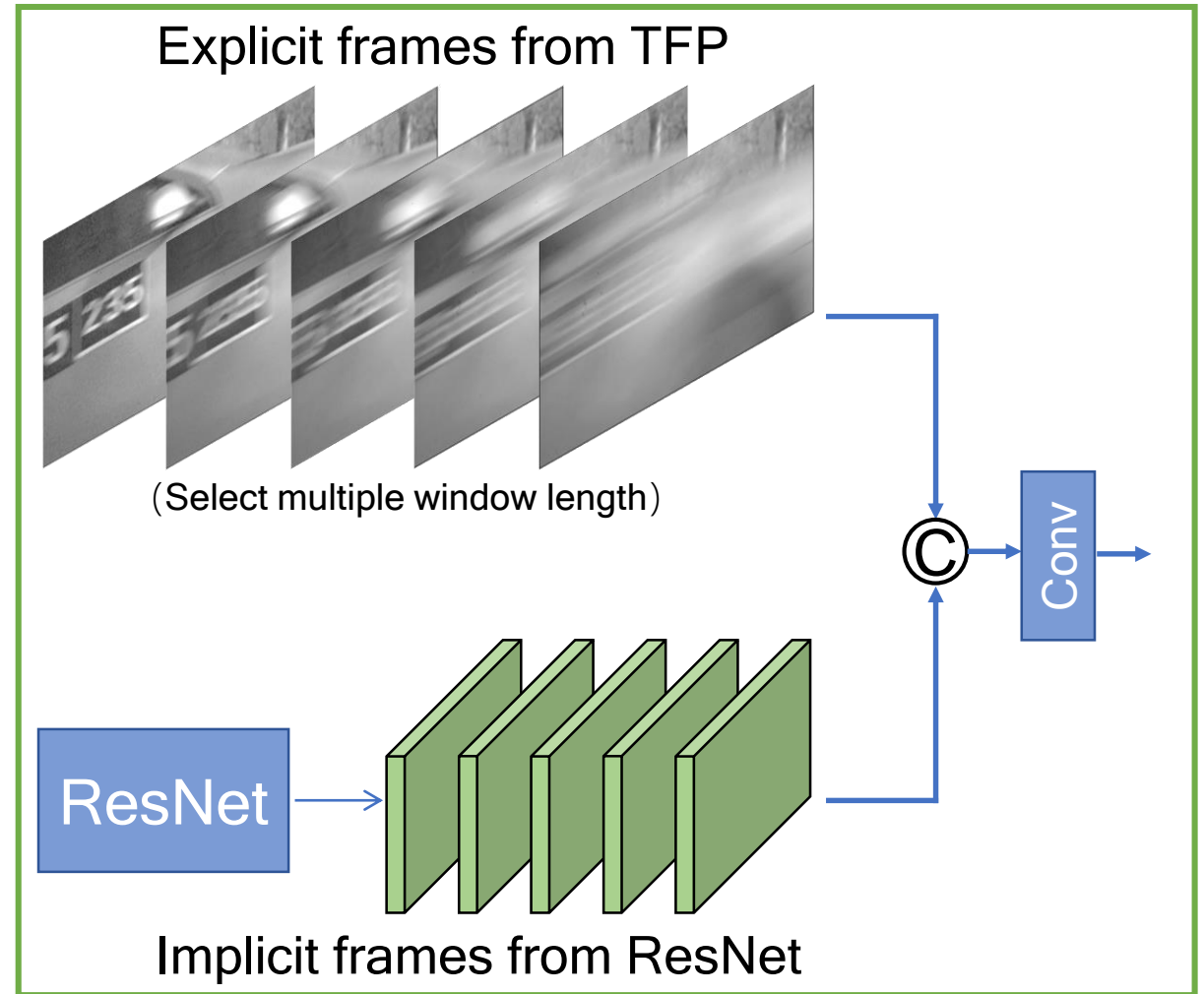
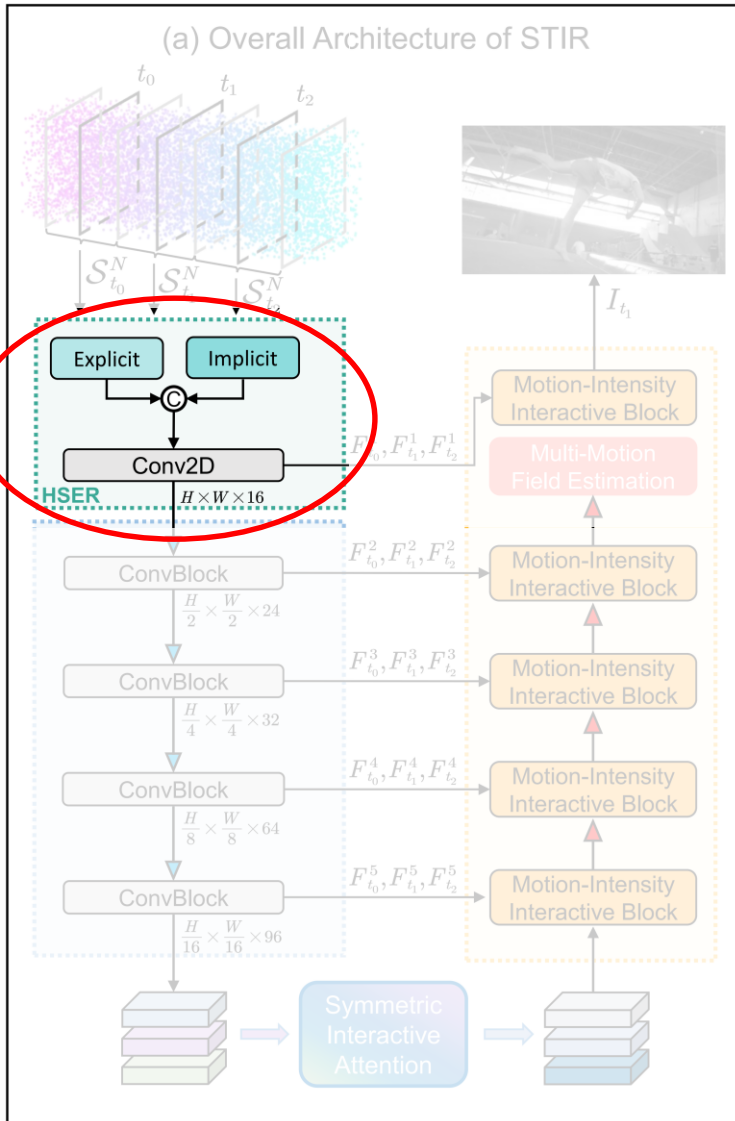


Spatio-Temporal
Interactive Refinement



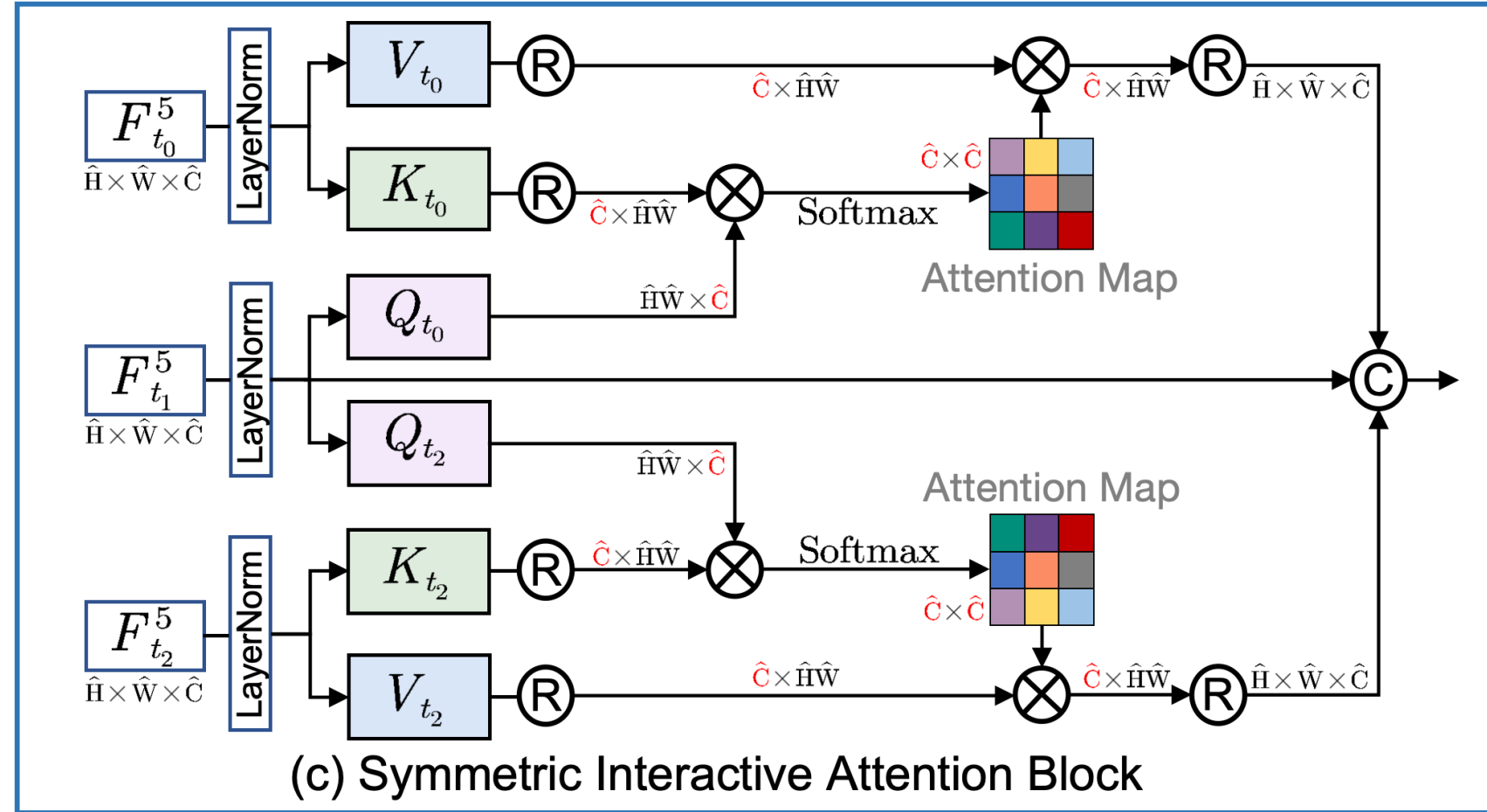
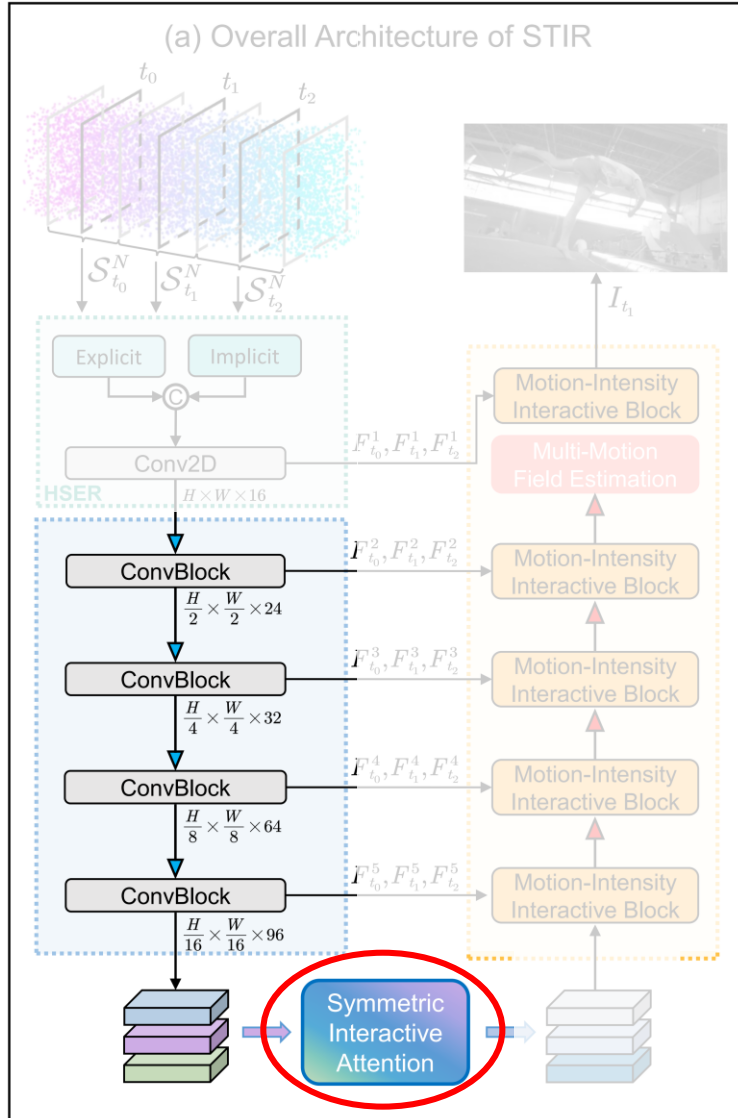
Ours

Our Approach

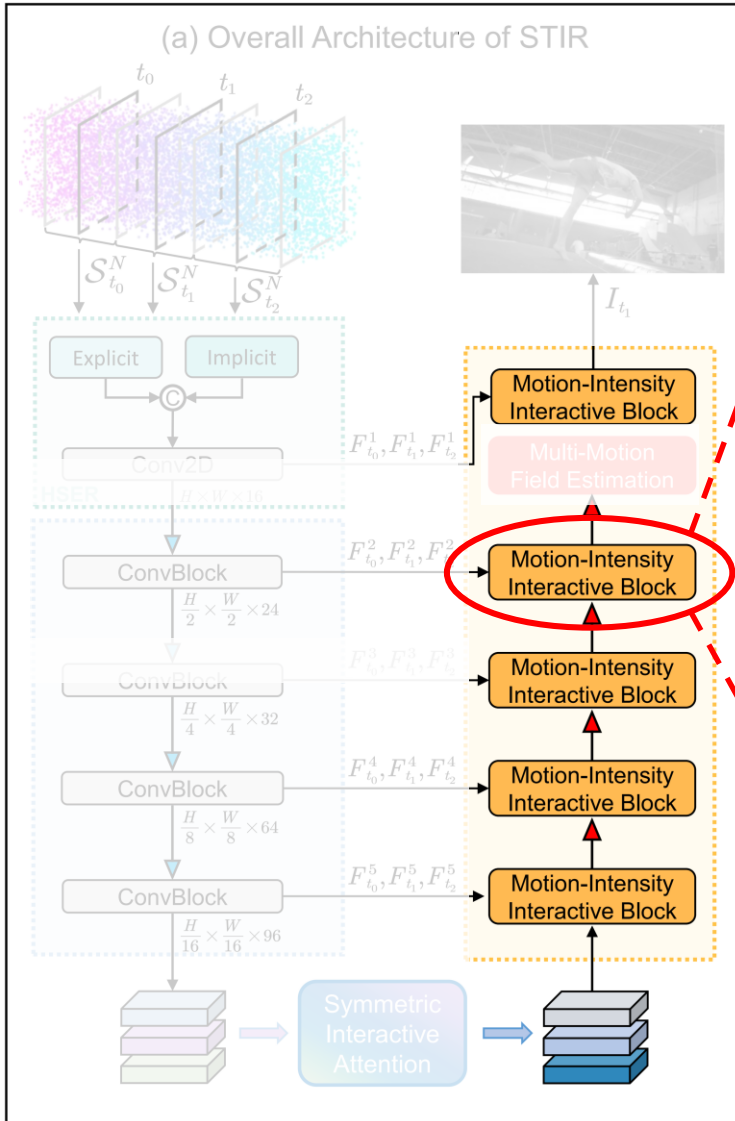


HSER: Hybrid Spike Embedding Representation

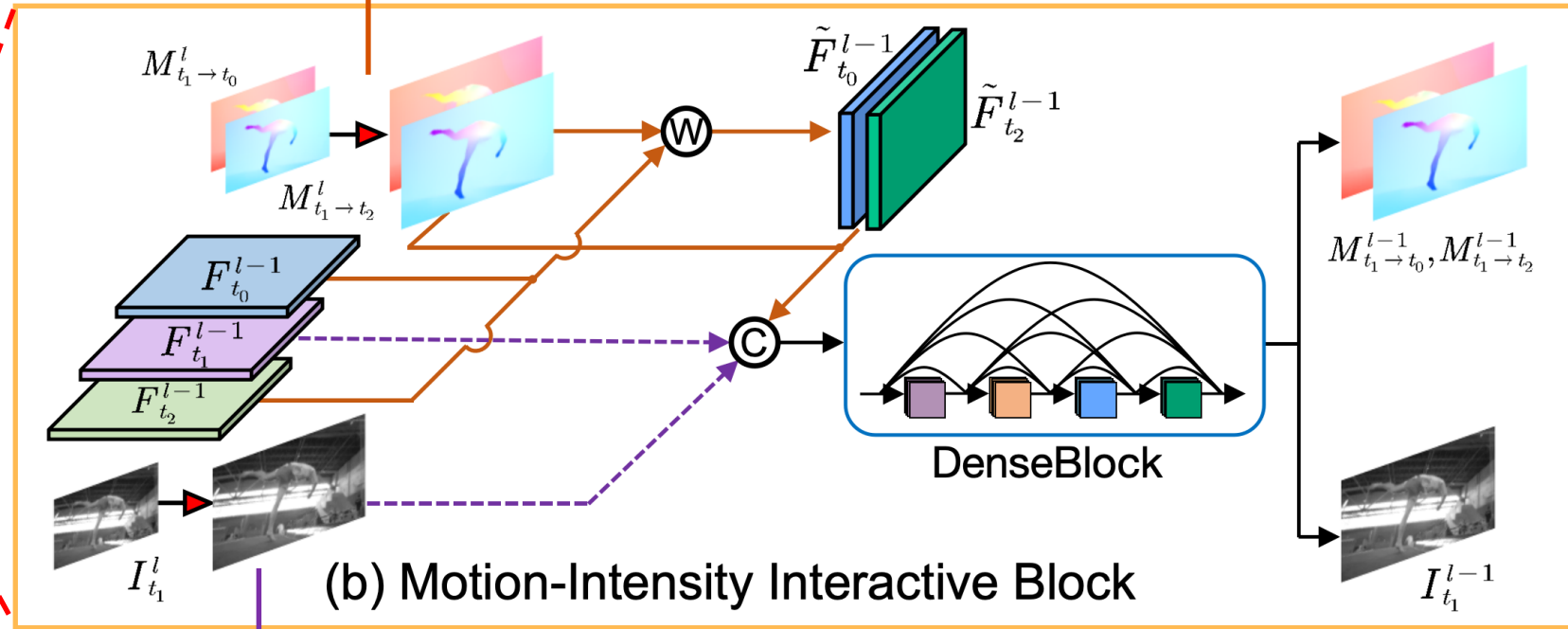
Our Approach



Our Approach

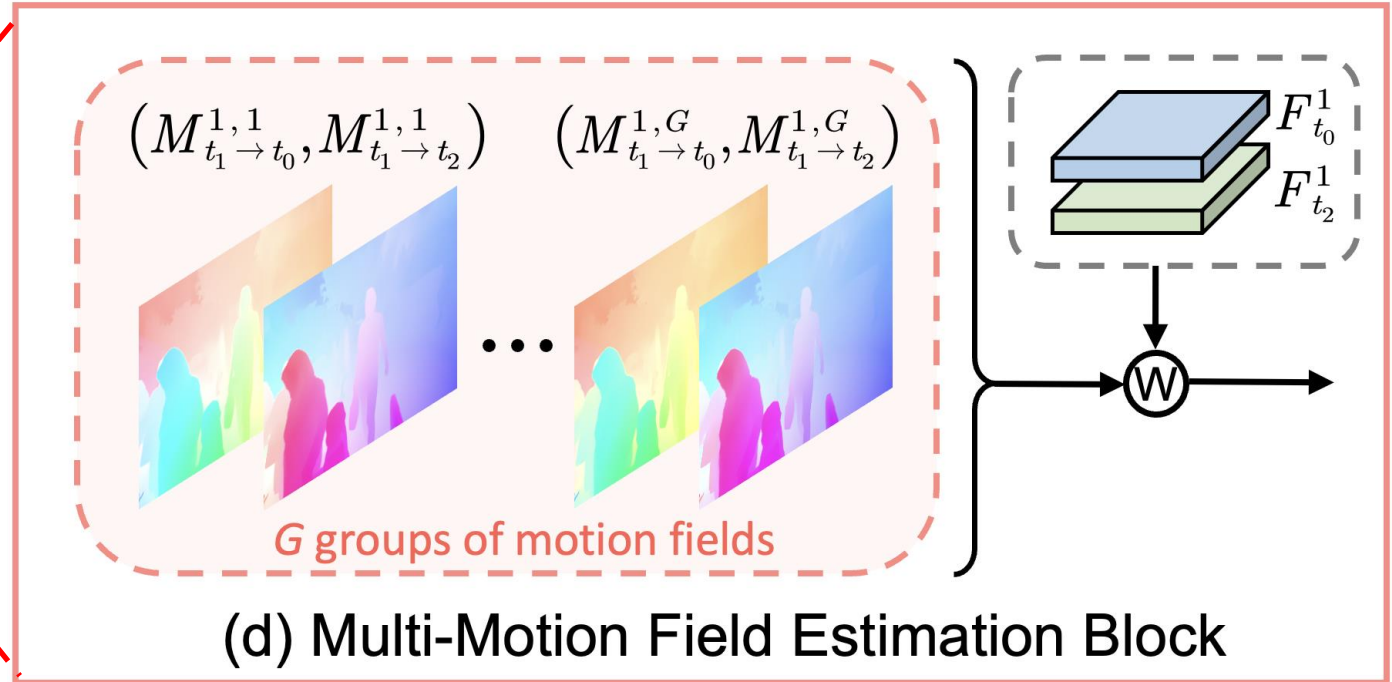
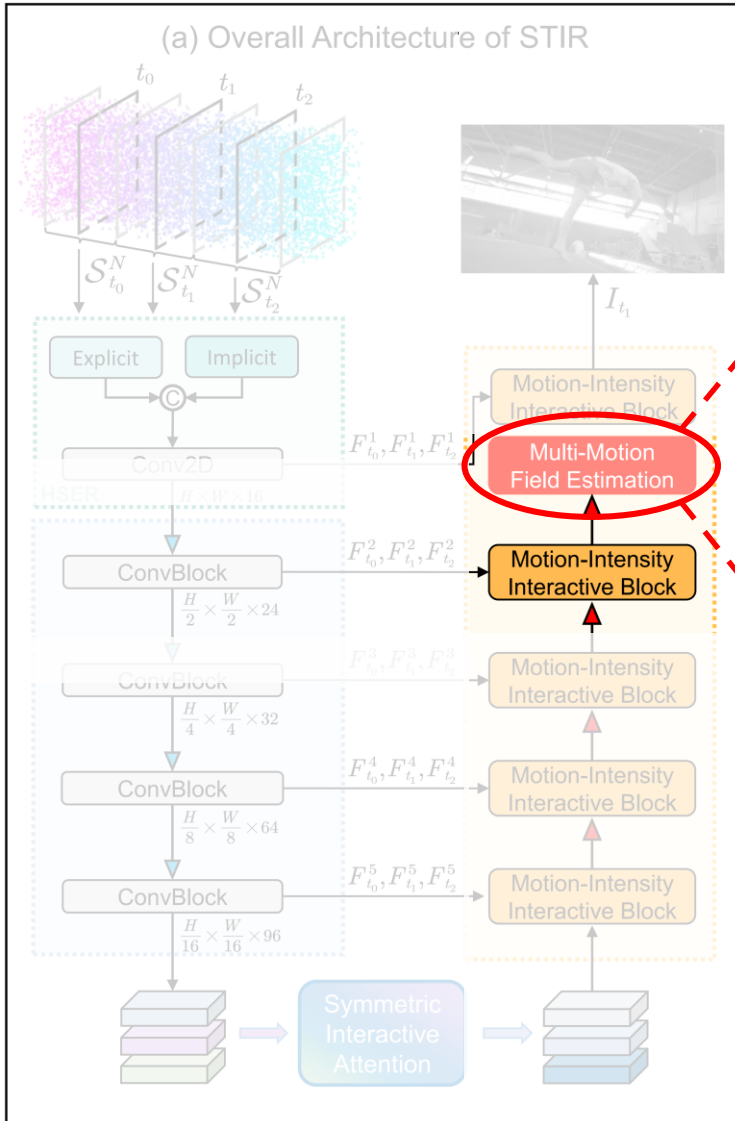


warping-based inter-frame feature alignment



synthesis based intra-frame feature filtering

Our Approach





Quantitative Results



- **SOTA** performance with **fast inference speed** and **low cost**

Method	Params (M)	Memory (G)	FLOPs (T)	Synthetic Dataset					Real Dataset	
				PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	NIQE \downarrow	BRISQUE \downarrow	NIQE \downarrow	BRISQUE \downarrow
TFP [65]	–	–	–	25.35	0.690	0.2547	5.970	43.074	9.342	45.202
TFI [65]	–	–	–	18.50	0.638	0.2590	4.518	44.933	10.09	58.309
TFSTP [63]	–	–	–	20.68	0.618	0.2761	5.348	51.697	10.92	64.566
SSIR [57]	0.38	10.4	0.24	32.61	0.919	0.0500	3.467	15.664	5.750	25.341
ET-Net [48]	16.7	17.7	0.52	34.57	0.938	0.0535	3.400	17.155	6.512	17.393
HyperE2VID [10]	10.7	6.87	0.43	36.37	0.947	0.0506	3.126	16.774	6.306	17.020
Spk2ImgNet [55]	3.76	14.6	9.17	36.13	0.950	0.0294	3.084	<u>15.348</u>	5.662	16.518
WGSE [52]	3.85	19.7	3.93	<u>37.44</u>	<u>0.958</u>	<u>0.0241</u>	<u>3.032</u>	15.555	<u>5.620</u>	<u>16.154</u>
STIR (Ours)	5.11	<u>9.20</u>	<u>0.42</u>	38.79	0.966	0.0183	2.915	14.835	5.394	15.854

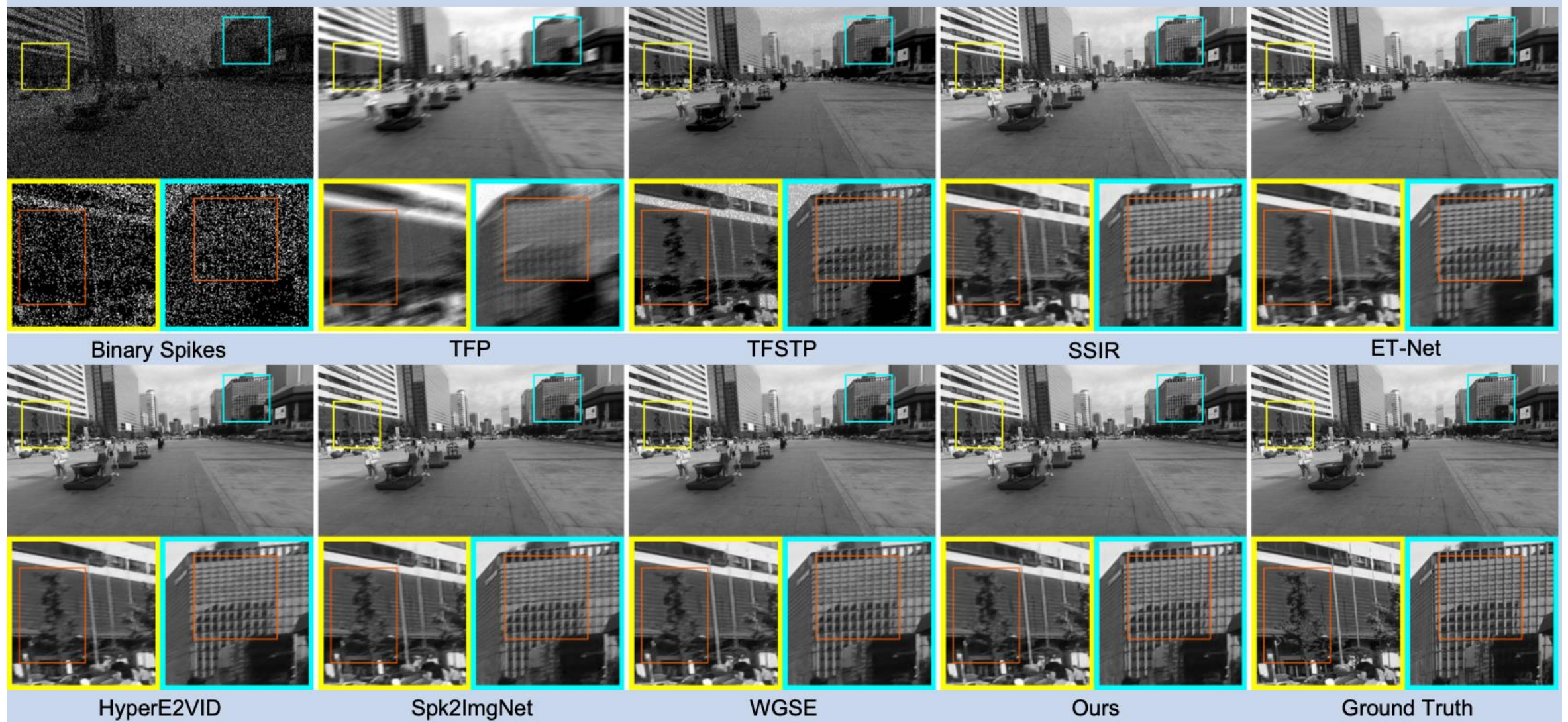


Evaluation on Synthetic Data



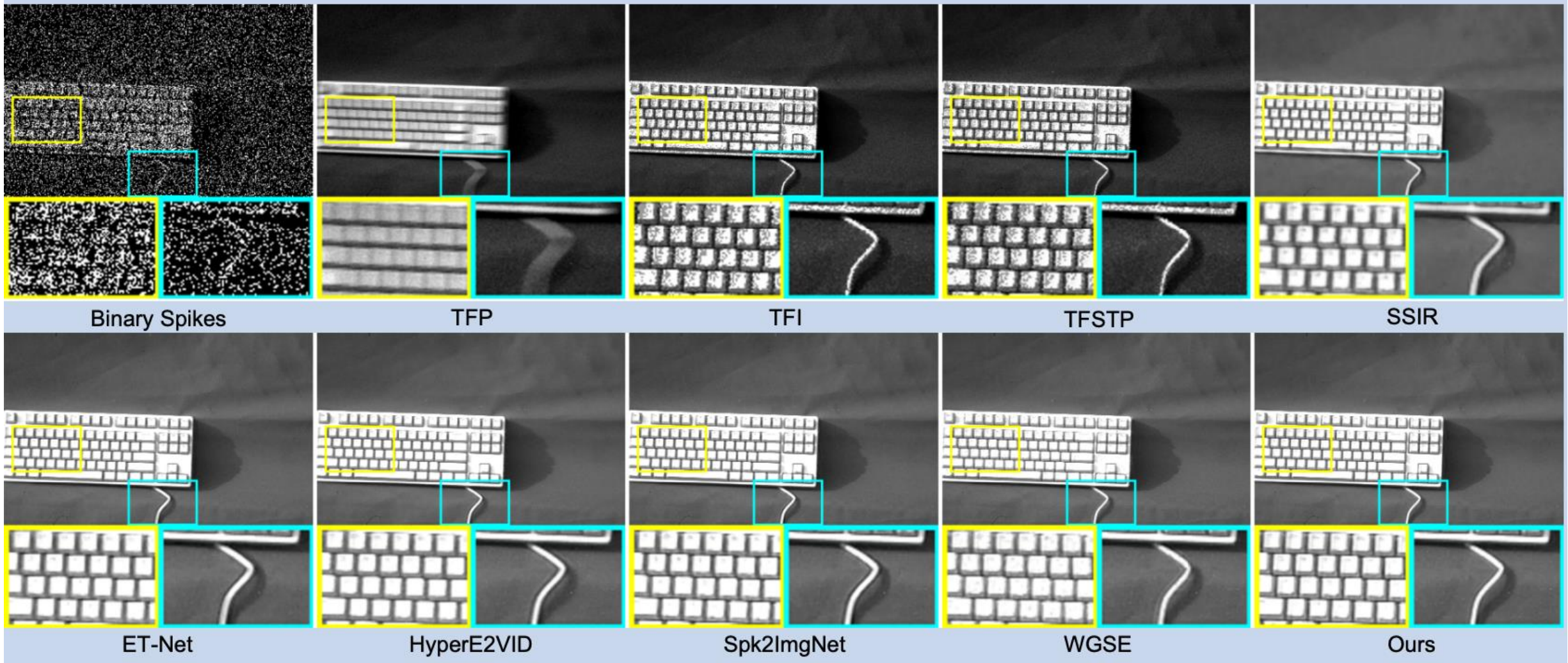


Evaluation on Synthetic Data



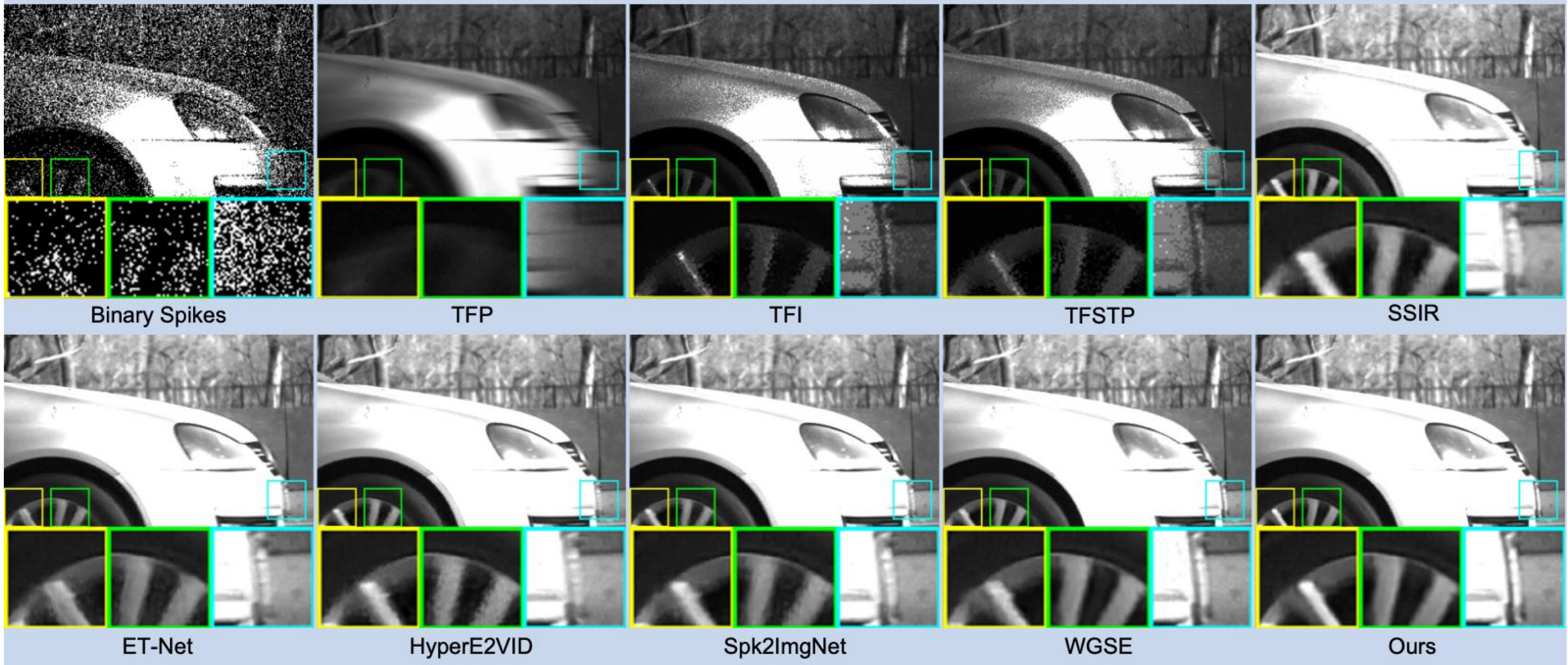


Evaluation on Real Data



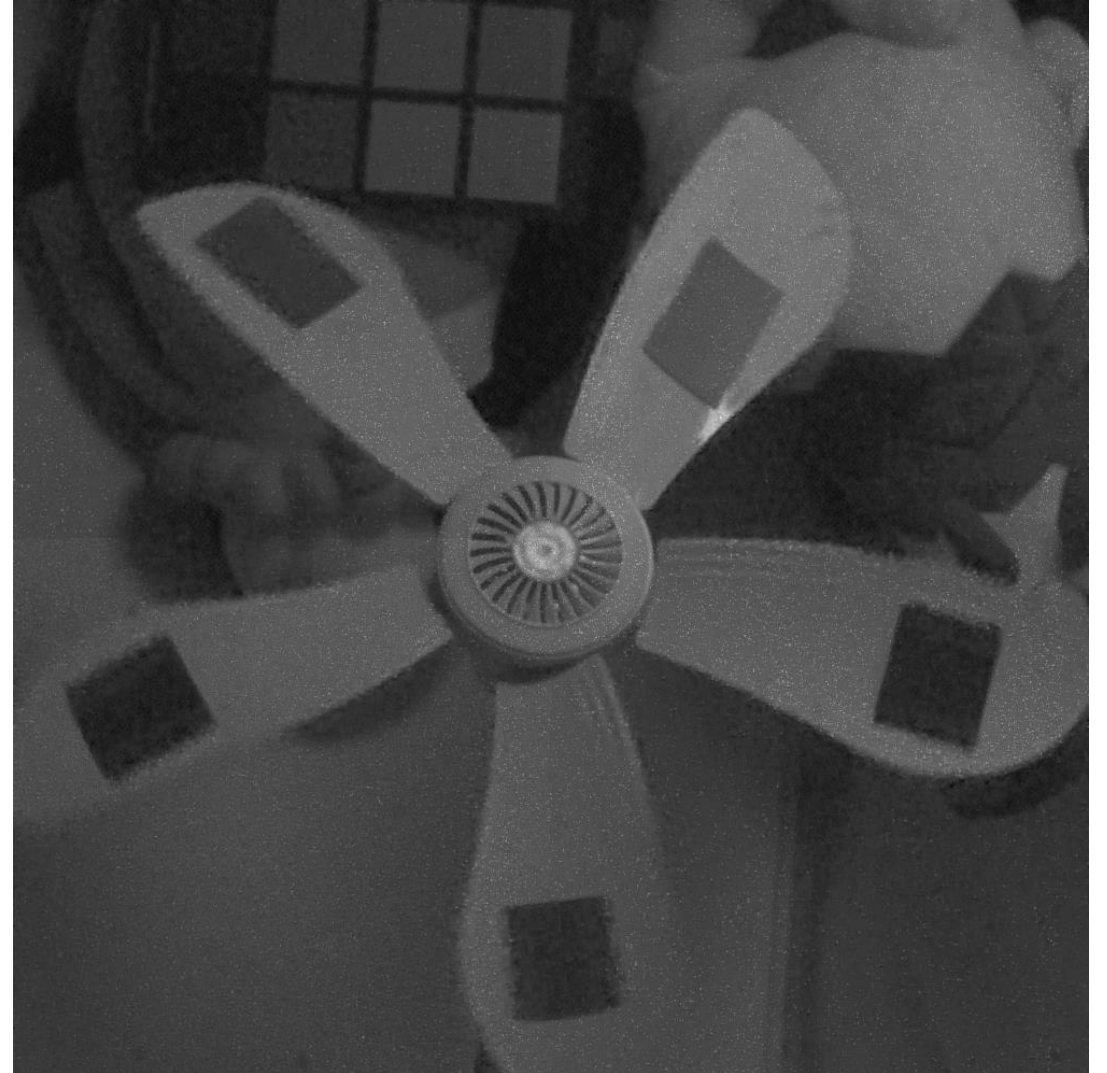
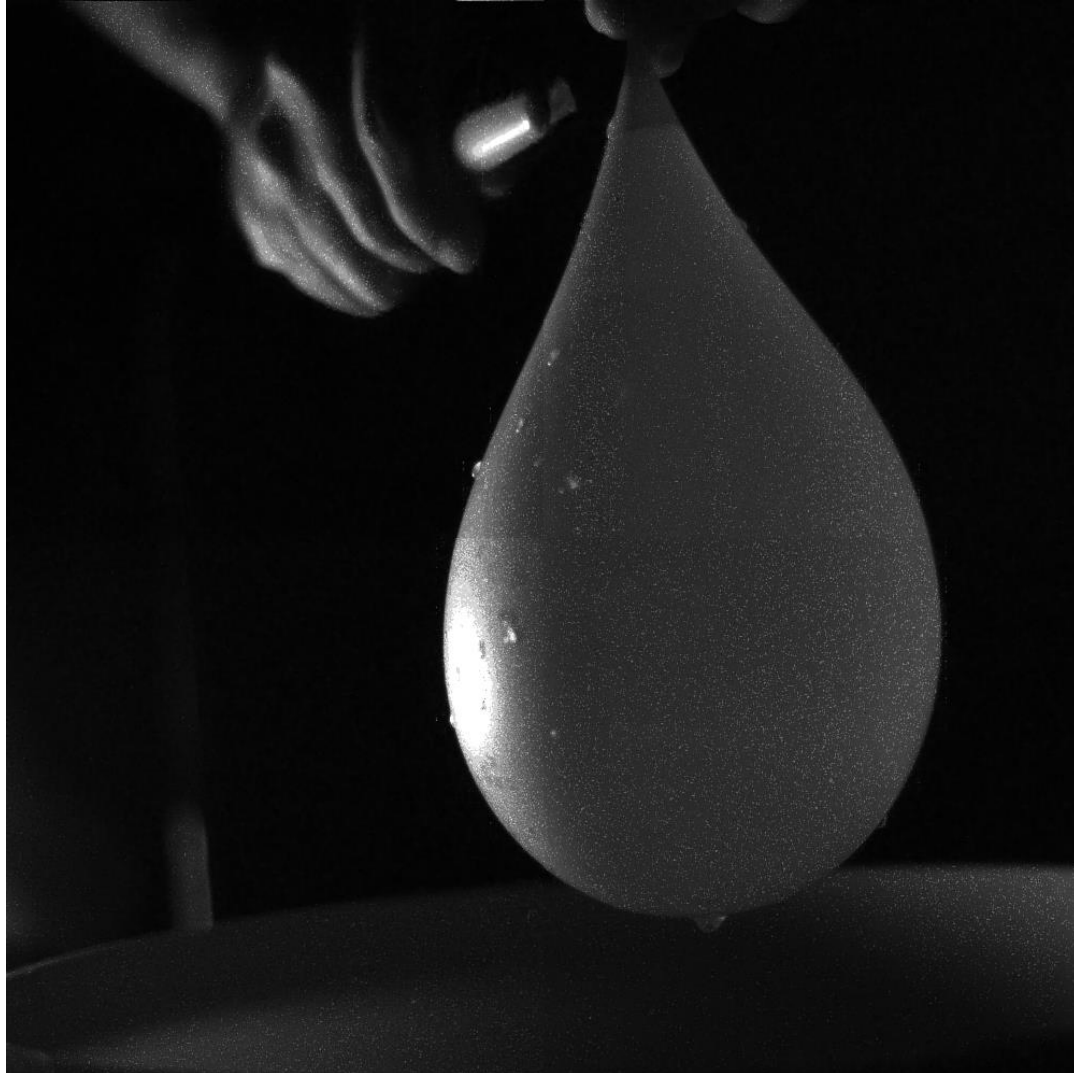


Evaluation on Real Data





High-frame-rate Video Reconstruction





Thank You for Listening!

Code link: <https://github.com/GitCVfb/STIR>