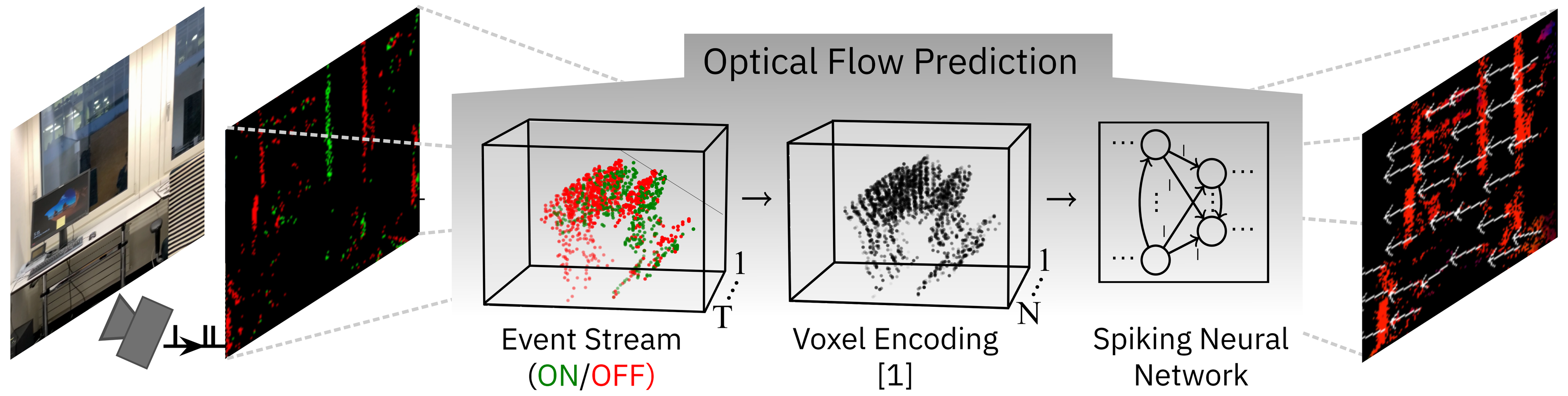
Neuromorphic Optical Flow and Real-time Implementation with Event Cameras

Yannick Schnider[1,2], Stanisław Woźniak[1], Mathias Gehrig[3], Jules Lecomte[4], Axel von Arnim[4], Luca Benini[2,5], Davide Scaramuzza[3], Angeliki Pantazi[1]
[1]IBM Research – Zurich, [2]ETH Zurich, [3]University of Zurich, [4]fortiss GmbH, [5]Università di Bologna

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Neuromorphic Optical Flow based on Event Data

Optical flow provides information on relative motion that is important for many computer vision applications. Neural networks yield **high accuracy** optical flow, yet their complexity is often prohibitive for application at the edge or in robots.



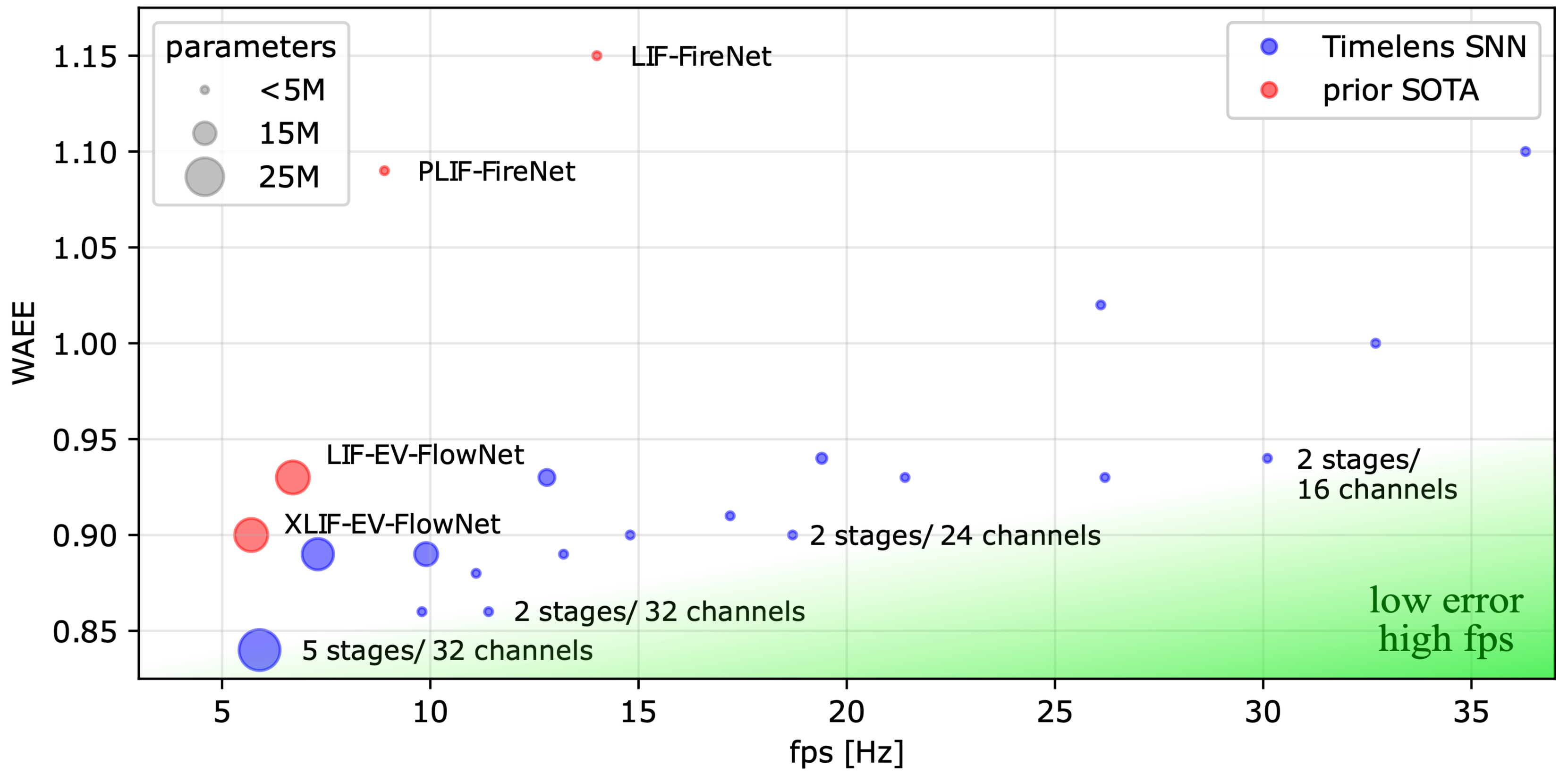
We propose an architecture that can operate both in a spiking and a non-spiking mode. **Model simplification** based on activity and latency analysis is performed to demonstrate high speed optical flow prediction for **real-time** deployments.

High-quality optical flow

	dt = 1	outdoor_day1		indoor_flying1		indoor_flying2		indoor_flying3		overall	
		AEE	$\%_{Out.}$	AEE	$\%_{Out.}$	AEE	$\%_{Out.}$	AEE	$\%_{Out.}$	WAEE	$\overline{\%}_{Out.}$
spiking	LIF-EV-FlowNet [4]	0.53	0.33	0.71	1.41	1.44	12.75	1.16	9.11	0.93	5.90
	XLIF-EV-FlowNet [4]	0.45	**0.16**	0.73	<u>0.92</u>	1.45	12.18	1.17	8.35	0.90	5.40
	LIF-FireNet [4]	0.57	0.40	0.98	2.48	1.77	16.40	1.50	12.81	1.15	8.02
	PLIF-FireNet [4]	0.56	0.38	0.90	1.93	1.67	14.47	1.41	11.17	1.10	7.00
	our SNN-Timelens	<u>0.44</u>	0.18	<u>0.70</u>	**0.79**	<u>1.30</u>	<u>9.41</u>	<u>1.05</u>	<u>6.00</u>	<u>0.84</u>	<u>4.10</u>
	our SNUo-Timelens	**0.39**	<u>0.17</u>	**0.64**	0.96	**1.17**	**7.71**	**0.96**	**4.92**	**0.76**	**3.44**
non-spiking	EV-FlowNet [4]	<u>0.47</u>	<u>0.25</u>	<u>0.60</u>	**0.51**	**1.17**	**8.06**	**0.93**	<u>5.64</u>	<u>0.78</u>	**3.61**
	RNN-EV-FlowNet [4]	0.56	1.09	0.62	0.97	1.20	8.82	**0.93**	**5.51**	0.83	4.10
	our sSNU-Timelens	**0.36**	**0.10**	**0.58**	<u>0.56</u>	<u>1.19</u>	<u>8.78</u>	<u>0.96</u>	6.11	**0.73**	<u>3.89</u>

Evaluation on the **MVSEC** dataset for comparable models trained on **UZH-FPV Drone Racing** dataset: AEE (the lower, the better), the percentage of outliers per sequence, and the overall weighted AEE (WAEE) as well as the average percentage of outliers. Best scores are in bold, while runner-ups are underlined. Horizontal line delimits **spiking** and **non-spiking** models.

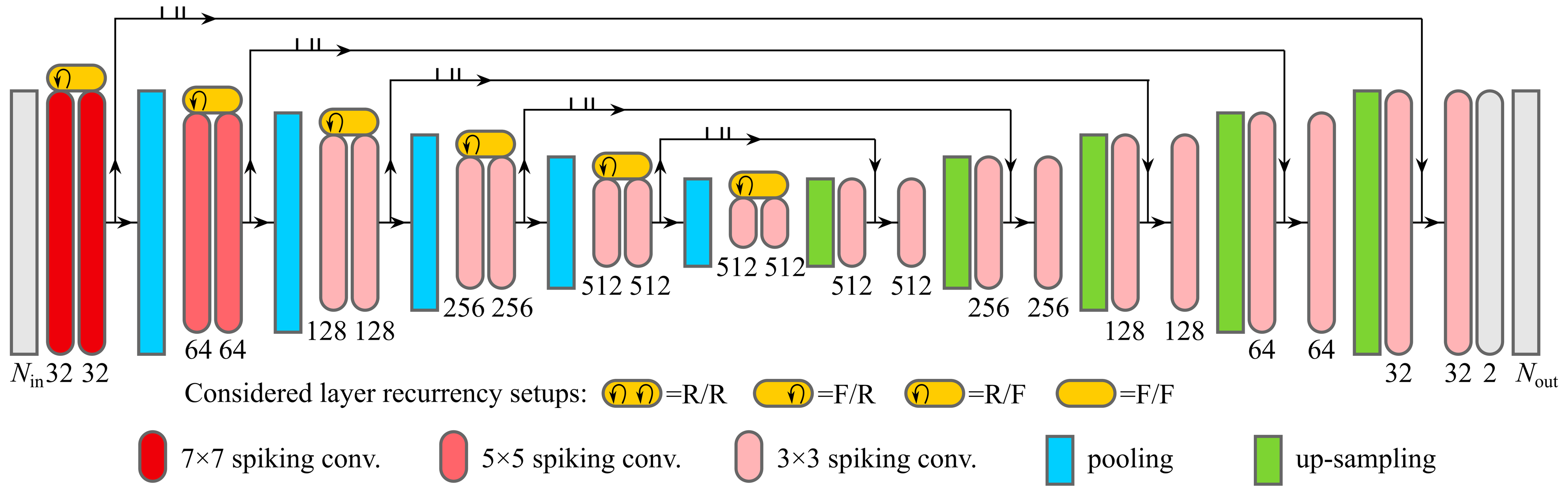
High-speed optical flow after model reduction



Model reduction results: Our model (blue) compared with prior state-of-the-art (SOTA) [4] (red) in our CPU setup. **Optical Flow quality** (WAEE) is plotted vs. **frames per second** (fps) while circle size indicates model size.

Network Architecture

Our network is based on the Timelens network [2] and reformulated as an **SNN** by incorporating spiking spatial convolutions featuring stateful neural cells and layer recurrency. Training is performed **self-supervised** via contrast maximization.

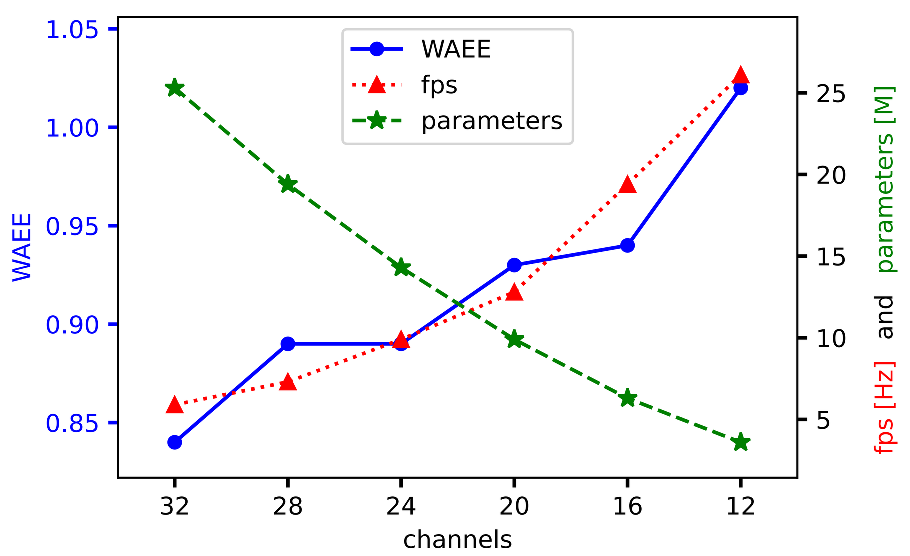


The biologically inspired neuronal cells [3] may operate in the following modes: **spiking** (SNN), **analog-valued spiking** (SNUo), **non-spiking** (sSNU).

Impact of connectivity

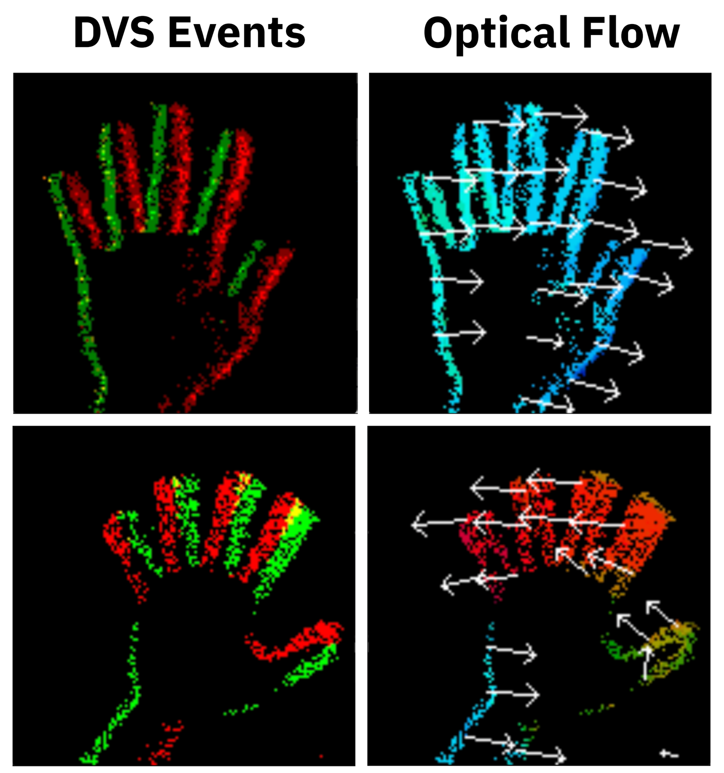
dt = 1	SNN-Timelens		sSNU-Timelens	
	WAEE	$\overline{\%}_{Outlier}$	WAEE	$\overline{\%}_{Outlier}$
R/F	**0.84**	**4.10**	<u>0.77</u>	<u>4.23</u>
F/R	<u>0.85</u>	4.36	1.11	8.44
R/R	0.89	<u>4.26</u>	**0.73**	**3.89**
F/F	1.12	5.89	1.18	9.26

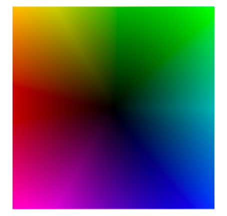
Impact of convolutional channels



Qualitative Results

Real-time predictions: aggregated DVS events and optical flow predicted by a **reduced SNN**-Timelens (0.32M) applied for different movements of a hand.





Color encoding used for Optical Flow representation

Conclusion and Contributions

- We design an optical flow architecture, enriched with spiking neurons operating with DVS-based inputs.
- We surpass the SOTA [4] for SNNs and ANNs on the MVSEC dataset by **6.1%** in spiking, **15.6%** in analog-valued spiking, and **5.5%** in non-spiking mode.
- We demonstrate **model reduction** from **20.4M to 0.32M** parameters with **0% penalty** in error with regard to the prior art [4], enabling real-time operation.

Demo video

https://youtu.be/jDGDxKabj0o

References

[1] "Unsupervised event-based learning of optical flow, depth and egomotion", A. Z. Zhu et al., CVPRW, Jun 2019
[2] "Time Lens: Event-based video frame interpolation", S. Tulyakov et al., CVPR, Jun 2021
[3] "Deep learning incorporating biologically inspired neural dynamics and in-memory computing", S. Woźniak et al., Nature Machine Intelligence, Jun 2020
[4] "Self-supervised learning of event- based optical flow with spiking neural networks", J. Hagenaars et al., NeurIPS, 2021