tinyML® EMEA

Enabling Ultra-low Power Machine Learning at the Edge

June 26 - 28, 2023

TINY ML

www.tinyML.org

ETH zürich

CENTER FOR PROJECT-BASED LEARNING

Ultra-Fast, Energy-Efficient Neuromorphic Edge Processing For Event-Based and Frame-Based Cameras: ColibriUAV and Eye-tracking

Sizhen Bian, Lukas Schulthess, Georg Rutishauser, Alfio Di Mauro, Tommaso Polonelli, Sizhen Bian, Pietro Bonazzi, Luca Benini, **Michele Magno**

28th. Juni 2023, TinyML EMEA

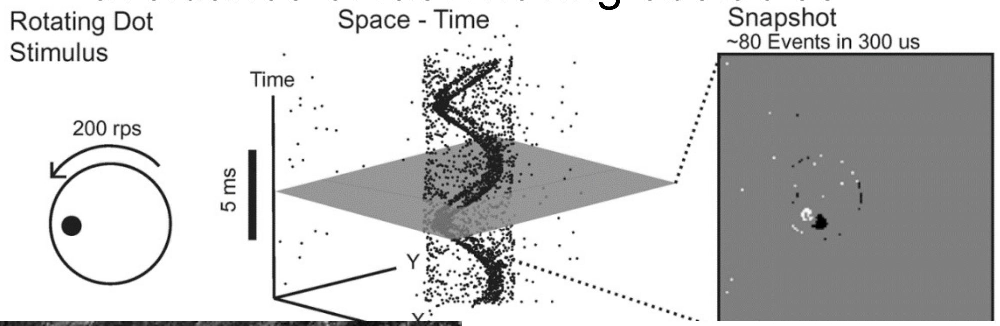
What have in common eye tracking and small-scale drone?

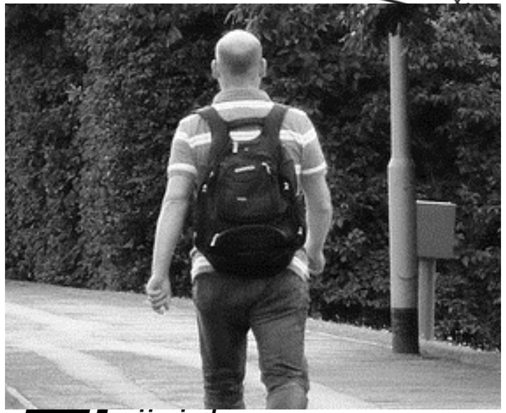
Both need to be extreme efficient and have low latency!

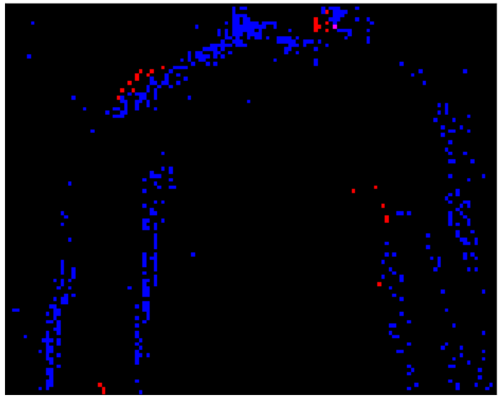
Exploiting 2 emerging technologies for efficiency and low latency for TinyML world.

Dynamic Vision Sensor

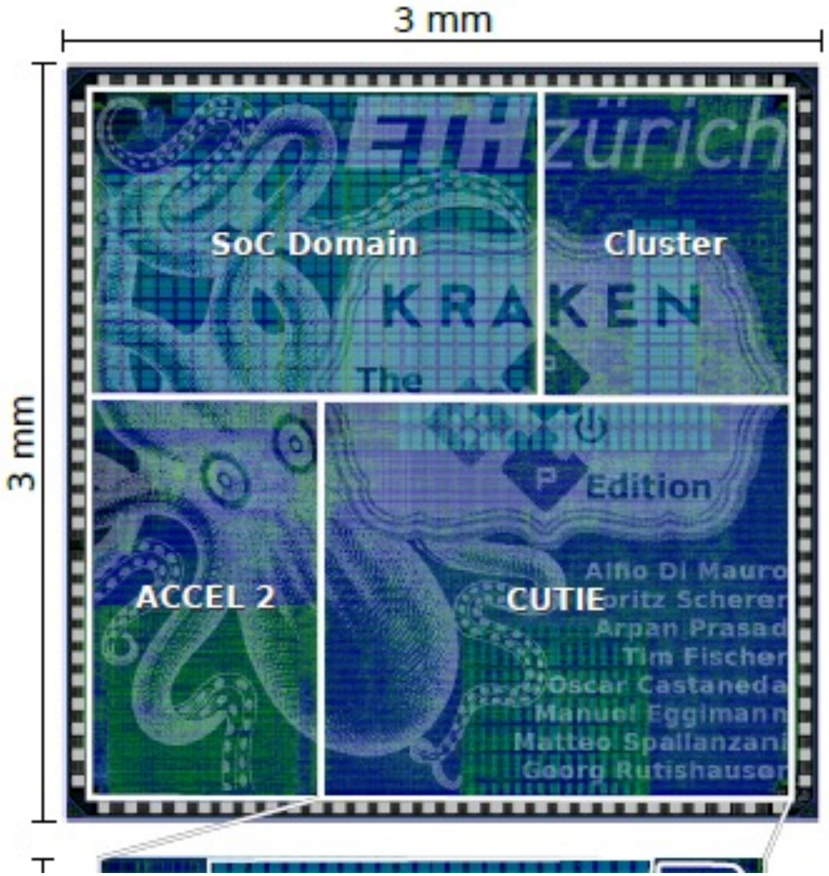
- Sparse images → Light weight, Fast
- Very Low latency of 3-5 milliseconds*
  - sufficient for reliable detection a
  - avoidance of fast-moving obstacles







Neuromorphic Hybrid Open Processors RISC-V



“*Extreme edge*” use case for TinyML: pico-size form factor UAVs

Advanced autonomous drone and vehicles

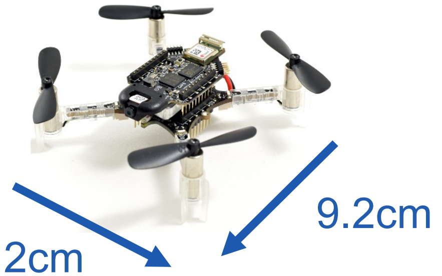
[1] A. Bachrach, “Skydio autonomy engine: Enabling the next generation of autonomous flight,” IEEE Hot Chips 33 Symposium (HCS), 2021

https://www.skydio.com/skydio-2-plus

Deployment in tight spaces?

Nano-drone

https://www.bitcraze.io/products/crazyflie-2-1



- 3D Mapping & Motion Planning
- Object recognition & Avoidance
- 0.06m2 & **800g of weight**
- Energy Capacity (Battery) **5410mAh**

KG

- Smaller form factor of 0.008m2
- Weight of **27g** (30X lighter)
- Battery capacity of **250mAh** (20X smaller)

KG

Can we fit sufficient “intelligence” in a **30X** smaller payload and **20X** lower energy budget?

Credits: Luca Benini

Relatd Work: UAVs with DVS Camera

Limitations:

- General processing platform for Event-data processing (Jetson TX2, Up-board), result in high energy consumption.
- Oldish interface to fetch events (USB).
- Neuromorphic processor is not for edge scenario (Loihi in best case or not low power and interconnected with Jetson Nvidia or Other high power platform).
- End-to-end evaluation limited available

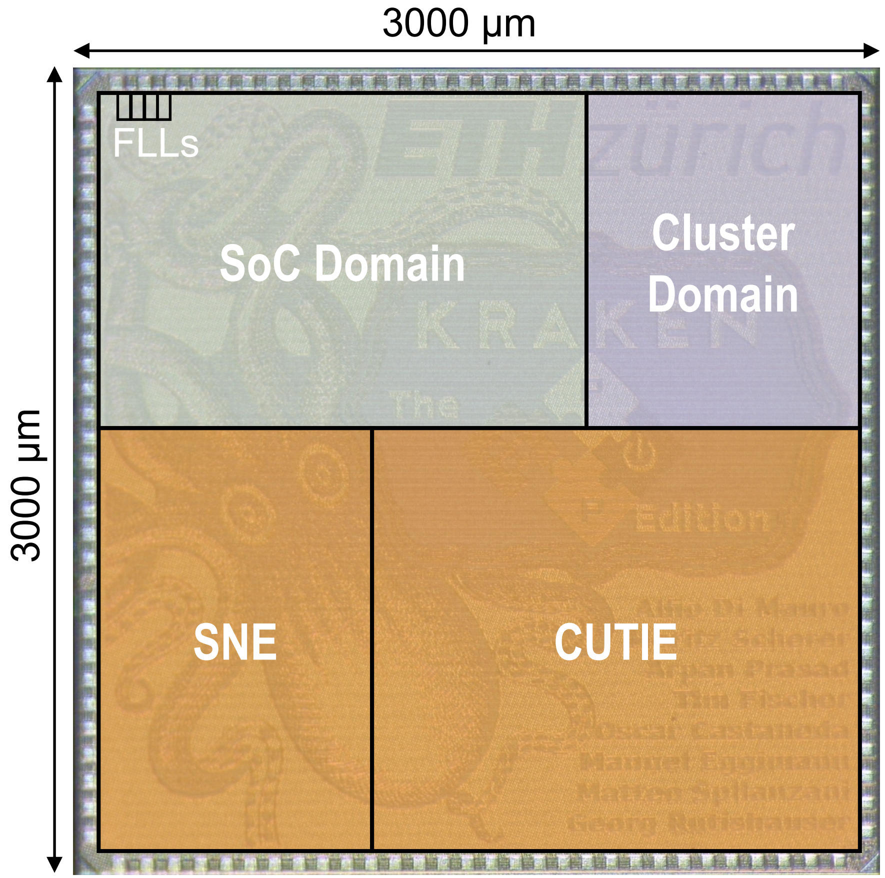
Contribution of this work:

- Construct a neuromorphic closed-loop UAV platform from event perception to decision-making, and benchmark the closed-loop latency and power consumption of various components in the event signal processing pipeline.

Bian, S., Schulthess, L., Rutishauser, G., Di Mauro, A., Benini, L., & Magno, M. (2023). ColibriUAV: An Ultra-Fast, Energy-Efficient Neuromorphic Edge Processing UAV-Platform with Event-Based and Frame-Based Cameras. *arXiv preprint arXiv:2305.18371*.

PULP-Kraken: Event-based Vision On The Extreme Edge

- RISC-V Cluster (8 Cores + 1)
- DVS Interface for hardware support of event-based vision
- SNE – energy-proportional spiking neural network accelerator
- CUTIE – dense ternary convolutional neural network accelerator
- PULPO – Floating point linear algebra accelerator



Technology	22 nm FDSOI
Chip Area	9 mm^2
SRAM SoC	1 MB
SRAM Cluster	128 KB
VDD range	0.55 V - **0.8 V**
Cluster Freq	**~370MHz**
SNE Freq	**~250MHz**
CUTIE Freq	**~140MHz**

Di Mauro, A., Scherer, M., Rossi, D., & Benini, L. (2022). Kraken: A direct event/frame-based multi-sensor fusion soc for ultra-efficient visual processing in nano-uavs. *arXiv preprint arXiv:2209.01065*.

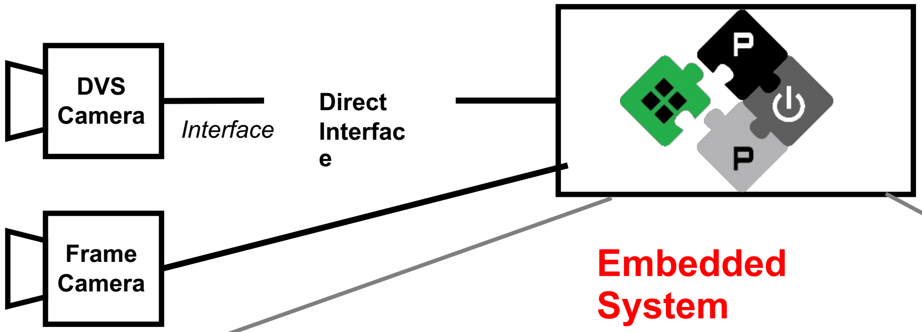
System Overview with the project

Demonstrator and in-field test

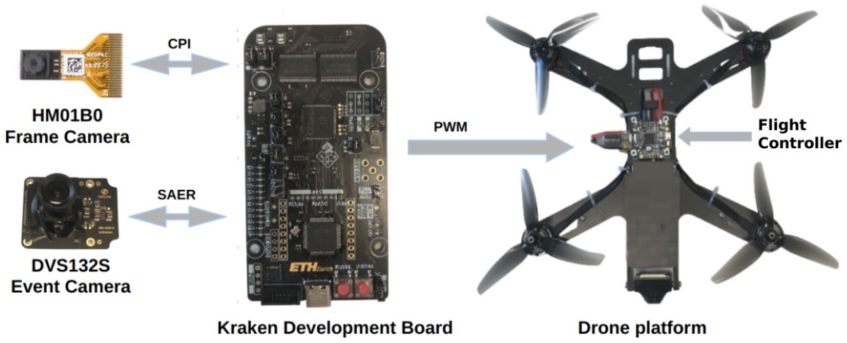
DVS Acquisition system

Kraken

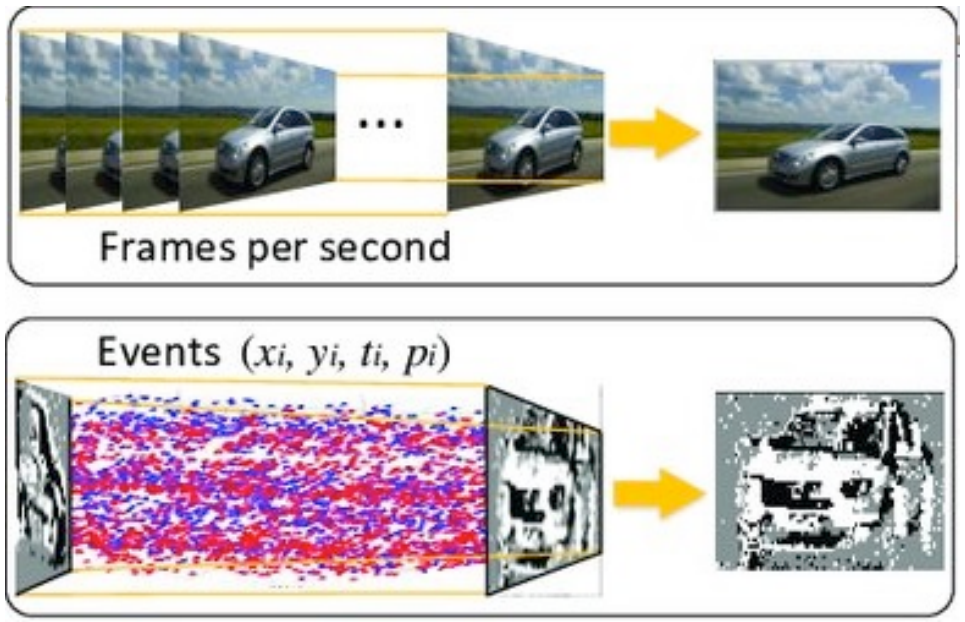
Sensor Interface



Embedded System



Algorithms



- Reduce the computational with respect to high-latency neural networks.*
- DVS interface integrated in next generation of embedded systems

Gao, S., Guo, G., Huang, H., Cheng, X., & Chen, C. P. (2020). An end-to-end broad learning system for event-based object classification. *IEEE Access*, *8*, 45974-45984.

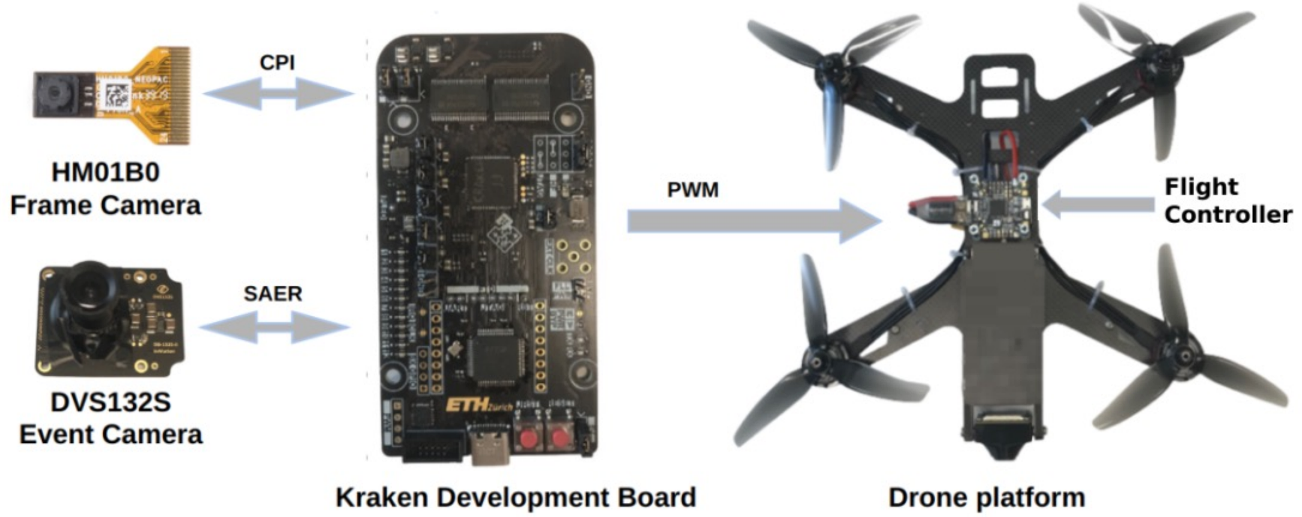
* Messikommer, N., Gehrig, D., Loquercio, A. and Scaramuzza, D., 2020. Event-based Asynchronous Sparse Convolutional Networks. *arXiv preprint arXiv:2003.09148*.

Implementation: ColibriUAV

- DVS132S
  -- 132x104, 1.2V Power Supply, 250 µW at 1K event frames per second.

- HM01B0
  -- 320x320, 1.5V/2.8V Power Supply, 4 mW at 60 frames per second.

- Drone
  -- Outlaw 270 racer (commercially available), 1720 g.
  -- F722-SE flight controller from Matek,



System architecture: ColibriUAV with Dual Camera and Kraken Development Board

Bian, S., Schulthess, L., Rutishauser, G., Di Mauro, A., Benini, L., & Magno, M. (2023). ColibriUAV: An Ultra-Fast, Energy-Efficient Neuromorphic Edge Processing UAV-Platform with Event-Based and Frame-Based Cameras. *arXiv preprint arXiv:2305.18371*.

Implementation: ColibriUAV

- Kraken development board

  -- Based on **Kraken**

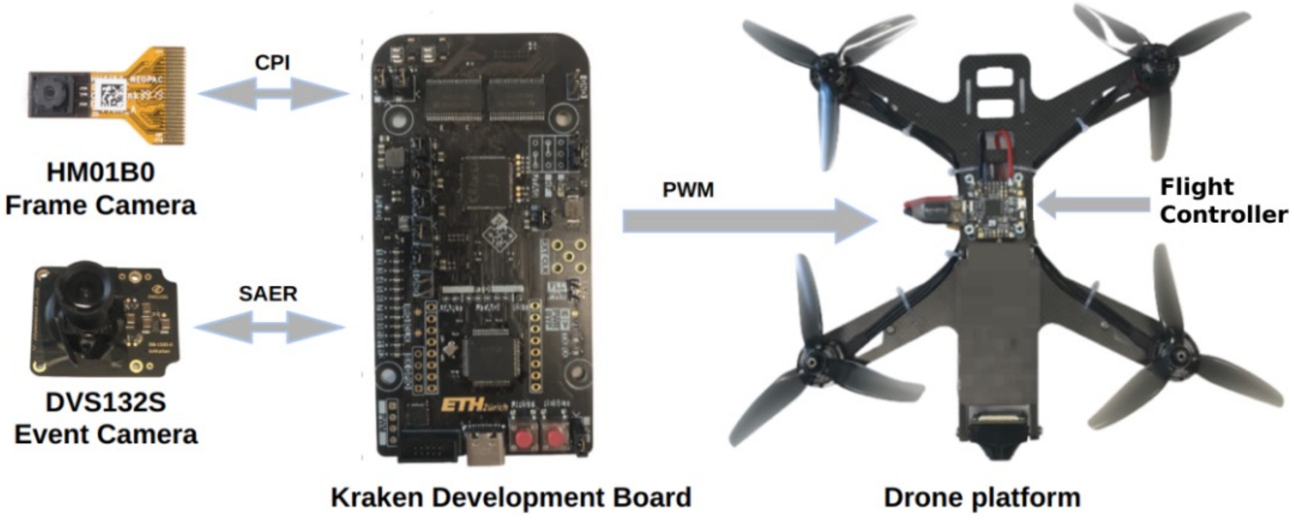
  -- Interfaces: CPI, SAER, PWM

  -- USB-C for JTAG and UART data transfer and supplies a 5 V power supply.

  -- Runtime-configurable power domain.

  -- Off-chip memory: a combined HyperFlash/HyperRAM chip and a quad-SPI flash memory chip.

  -- Arduino compatible headers for additional connectivity.



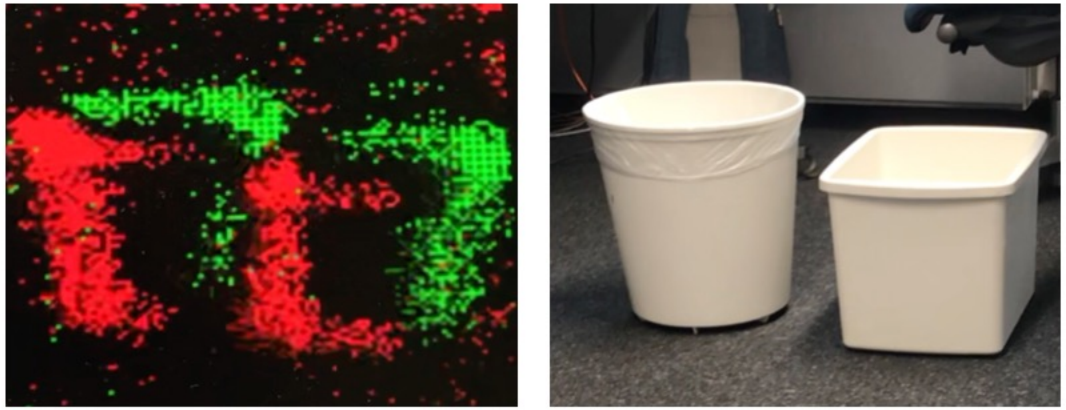
System architecture: ColibriUAV with Dual Camera and Kraken Development Board

Evaluation

- DVS Camera:

  -- Configurable threshold based ON/OFF event of each pixel

  -- Two simultaneous streams of output data:
  1, the X/Y address with one byte
  2, the ON/OFF events with one byte recording the events of four pixels (2x2). (thus, if **13728 events in one frame**, only 3432 (66x52) system clocks for the frame is needed)



Recordings at one particular event frame of the DVS132S (left) and an RGB camera (right)

-- Power consumption of the camera: **0.42 mW (0.36 mW for analog components and 0.06 mW for digital components.**

-- Event-frame sampling rate is configured on the Kraken SoC.

Evaluation:

- DVS Interface:

-- USB:
5V power suply yealding W-level data transmission. 1087 event-frame per second.

-- SAER on FPGA solution:
17.6 mW with 874 event-frame per second.

-- SAER on ColibriUAV:
10.656 mW with 7200 event-frame per second.

DVS INTERFACE

Modules[a]	Throughput[a] (efps[b])	Power[c] (mW)
USB [12], [15], [16], [18]	1087	over 1000 [d]
SAER on FPGA [10]	874	17.6
SAER on ColibriUAV	**7200**	**10.656[e]**

*Summarized in our paper.

The throughput and power of different interface solutions

Relatd Work: UAVs with DVS Camera

DRONES EQUIPPED WITH DYNAMIC VISION SENSORS (DVS)

	Computing Unit	Event Camera	interface	Algorithm	P_{inf} (mW) [a]	T_{inf} (ms) [a]	T_{samp} (ms) [b]	T_{close_loop} (ms) [c]	P_{close_loop} (mW) [c]
[11]-2020	NVIDIA Jetson TX2	two SEES1	USB2.0	Ego-motion compensation	1000+ (Watt level)	3.56	10	N/A [d]	1000+ (Watt level)
[15]-2022	NVIDIA Jetson TX2	DAVIS2404	USB2.0	Sparse Gated Recurrent Network	1000+ (Watt level)	500	≈ 1000-2000	N/A	1000+ (Watt level)
[12]-2015	Odroid U3 quad-core computer	two DVS128	USB2.0	event-based circle tracker, EKF	1000+ (Watt level)	4.5	N/A	N/A	1000+ (Watt level)
[16]-2020	UP Board	DAVIS 240C	USB2.0	hough transform, kalman filter	1000+ (Watt level)	12	3	N/A	1000+ (Watt level)
[17]-2021	Loihi, Kapohobay	DAVIS 240C	SAER [e]	Hough transform	1000+ (Watt level)	0.25	0.15	N/A	1000+ (Watt level)
[18]-2022	Intel NUC computer	two DAVIS346	USB3.0	graph-based optimization	1000+ (Watt level)	N/A	N/A	N/A	1000+ (Watt level)
Ours	**Kraken**	**DAV132S**	**SAER**	**SNN**	**35.6**[f]	**163 (131+32)**[f]	**300**	**163 (131+32)**[h]	**46.98 (35.6 + 11.076 + 0.3)**[f]

*Summarized in our paper.

Processor

Interface

Close-loop latency and power

Evaluation Result:

	Type	Size	Feature Size	Features	Stride
0	Input	128x128x2	-	-	-
1	Pool	32x32x2	4x4x1	2	4
2	Conv	32x32x16	3x3x2	16	1
3	Pool	16x16x16	2x2x1	16	2
4	Conv	16x16x32	3x3x1	32	1
5	Pool	8x8x32	2x2x1	32	2
6	Full	512	2048	512	-
7	Full	11	512	11	-

PERFORMANCE OF COLIBRIUAV ON REFERENCE NEURAL NETWORK

Modules[a]	Latency (ms)	Power (mW)[b]	Energy (mJ)[b]
DVS and SAER (single event-frame)	0.069	11.076[c]	0.0008
DVS and SAER (one window / 4350 event frames)	300	11.076	3.323
Preprocessing (Cluster)	131	34	4.5
Inference (SNE)	32	44	1.4
PWM(50% duty)	<0.001	0.3	<0.001
Total	163 [d]	46.98[e]	9.224

End-to-End latency: 163 ms.

-- 131ms on spike preprocessing in the cluster(WIP! Should be improved).

-- 32 ms on SNE for a complete SNN inference.

End-to-End energy consumption: 9.224 mJ

-- 3.323 mJ on DVS and interface when supplying one window of event frames for classification.

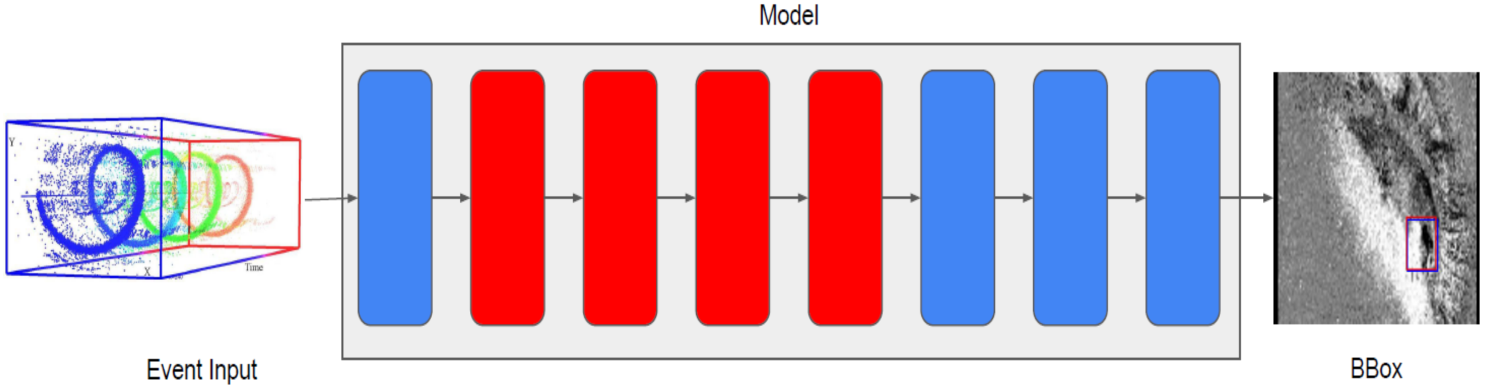
-- 5.9 mJ on Kraken (4.5 mJ on preprocessing and 1.4 mJ on Inference at SNE).

Eye-tracking with a truly low power platform

09-06-2023

Working in progress on eye tracking

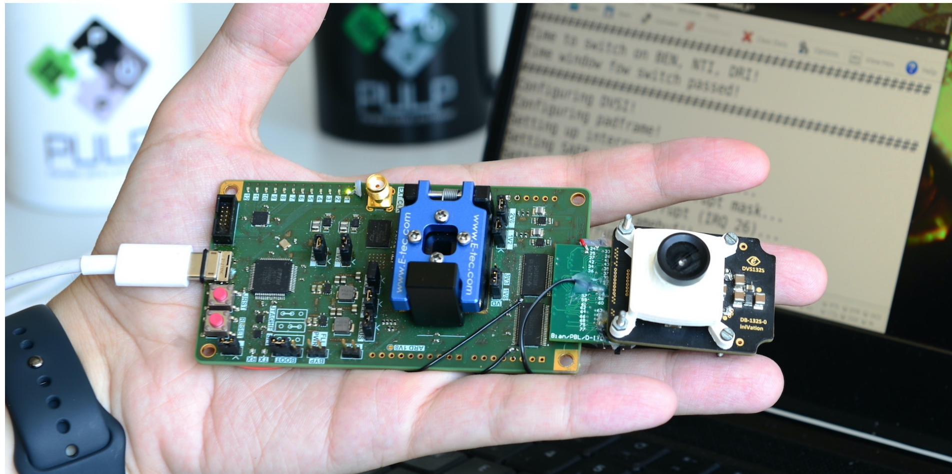
- Cooperation with Inivation.
  - Designing a yolov8 quantized CNN and yolo spiking (WiP)
  - Below 500KB of memory
  - Preliminary results of about 1.5ms (estimated time).



"*Neuromorphic Yolo*", Bonazzi P. , Bian, S. , Li Y. and Magno, M. , CVPR 2024

Innosuisse

inivation



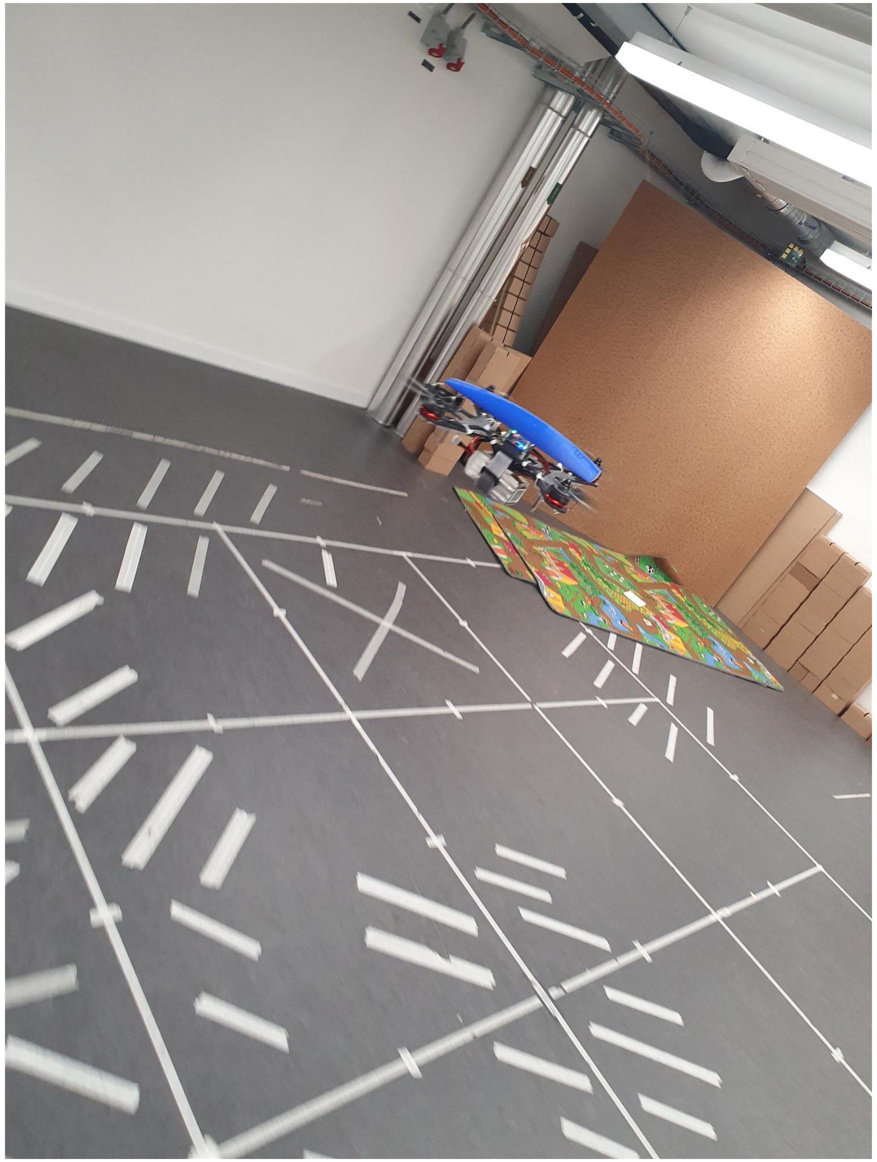
Conclusion

- **The first neuromorphic deep learning edge platform.**

- **Dedicated event interface (SAER) suppying state-of-the-art event transmission:**

- **Close-loop neuromorphic performance:**

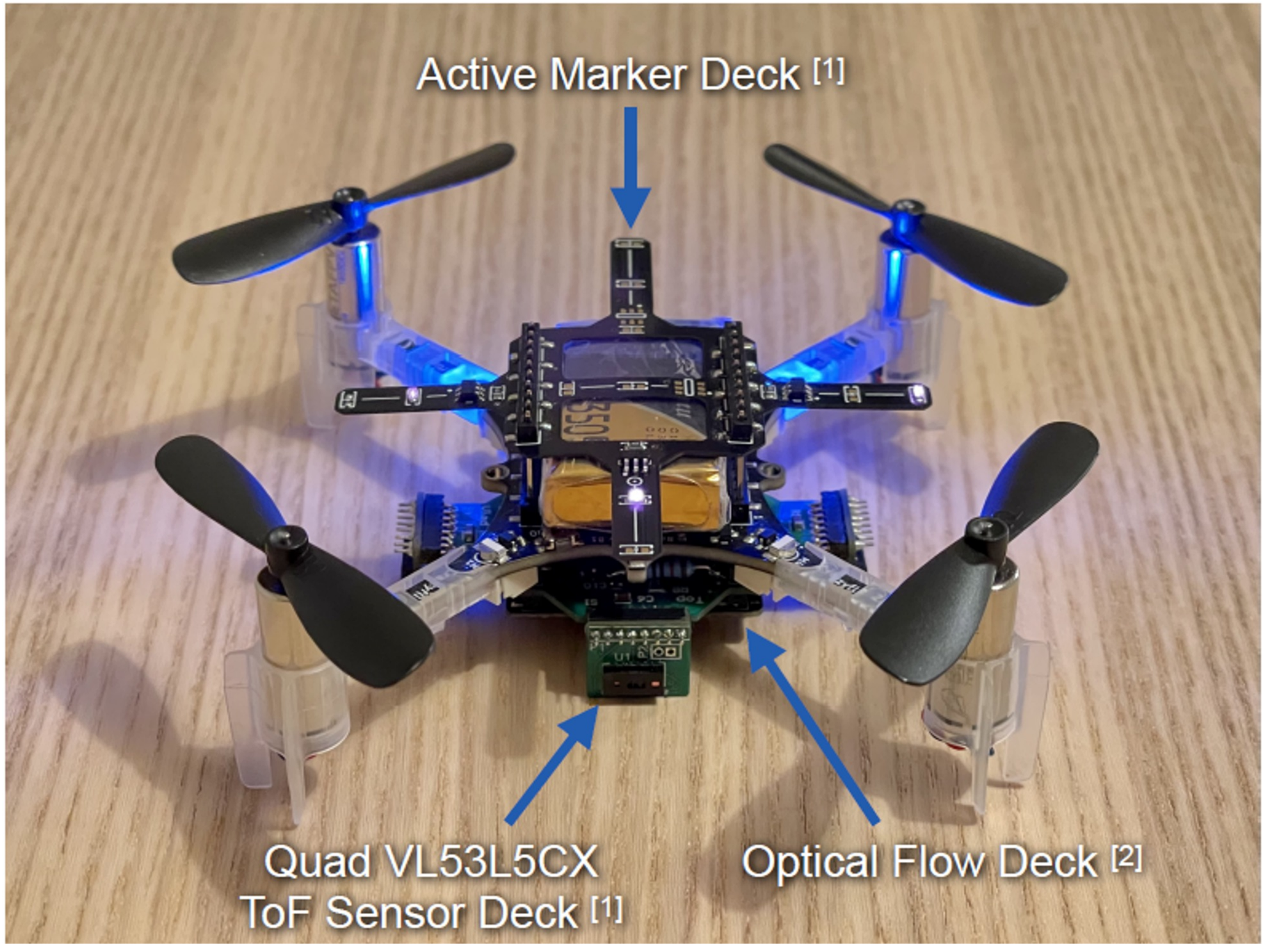
power of 46.98 mW, and energy consumption of 9.224 mJ for the end-to-end task from events to motor control



Nano-drone can be a good use case for TinyML and it is not (only) a toy

- Suitable for indoors and constricted environments
- Safe around humans
- Many applications including research, search and rescue missions, warehouse monitoring and others

Fully Onboard Autonomous navigation, with mapping, planning , object avoidance and landing (LIFVE DEMO!)



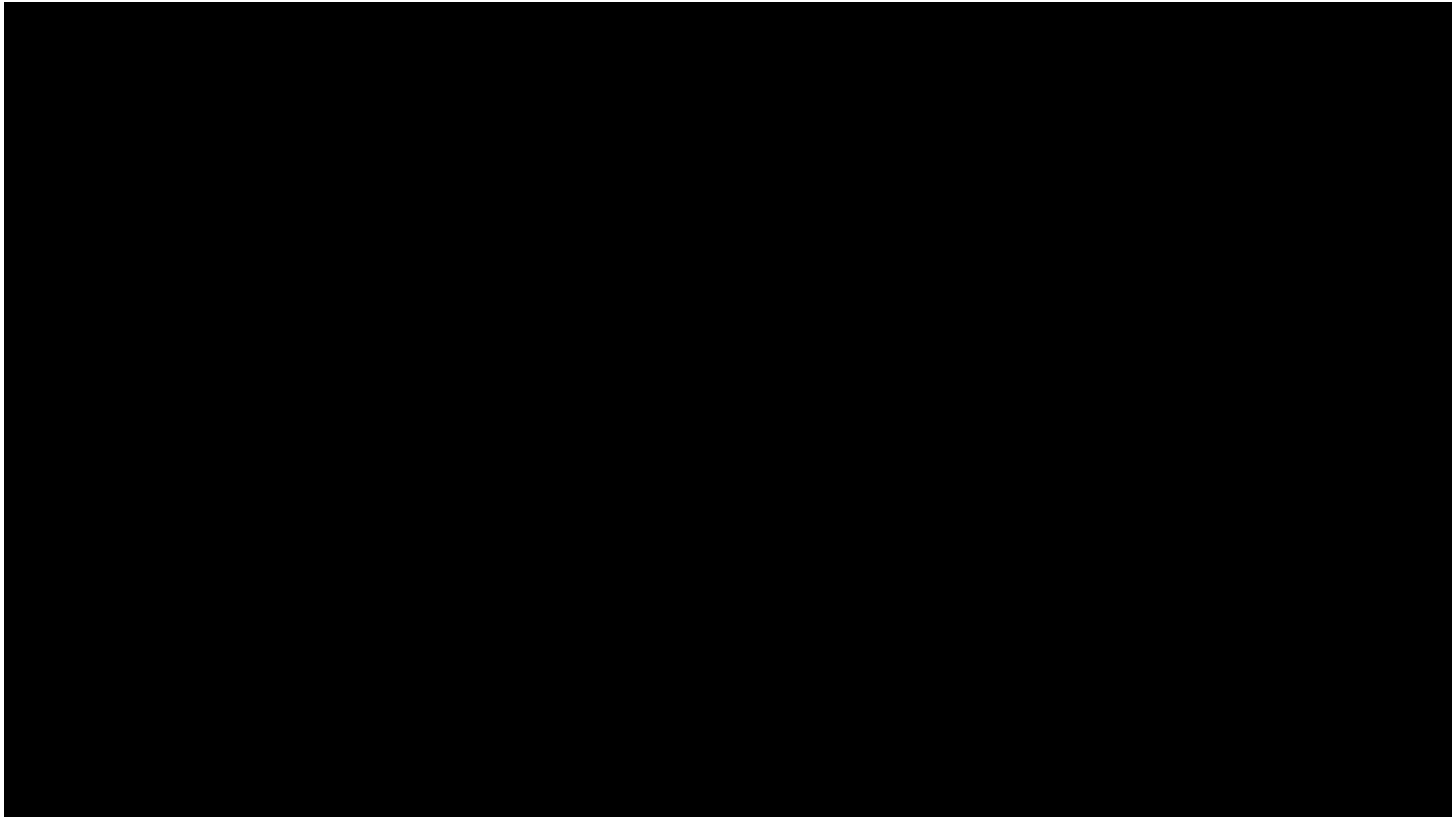
Each nano-UAV si equipped with:

1. **4x 64-pixels depth sensors**
2. **STM32 - controller**
3. **GAP9 - co-processor**
4. **NRF51 transceiver - intra-swarm communication**
5. **Active marker - ground truth**

[1] Custom

[2] Commercially available

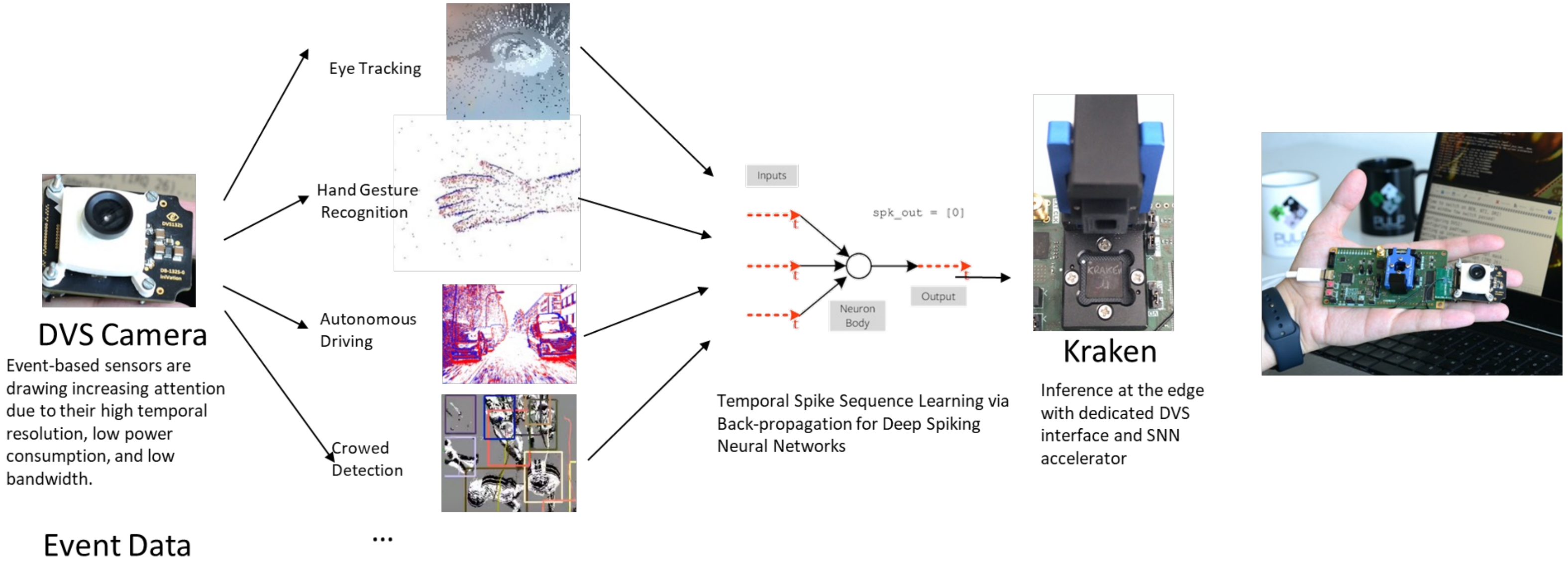
If the demo did not work…. Or for our video audience



Collaboration and Future Work

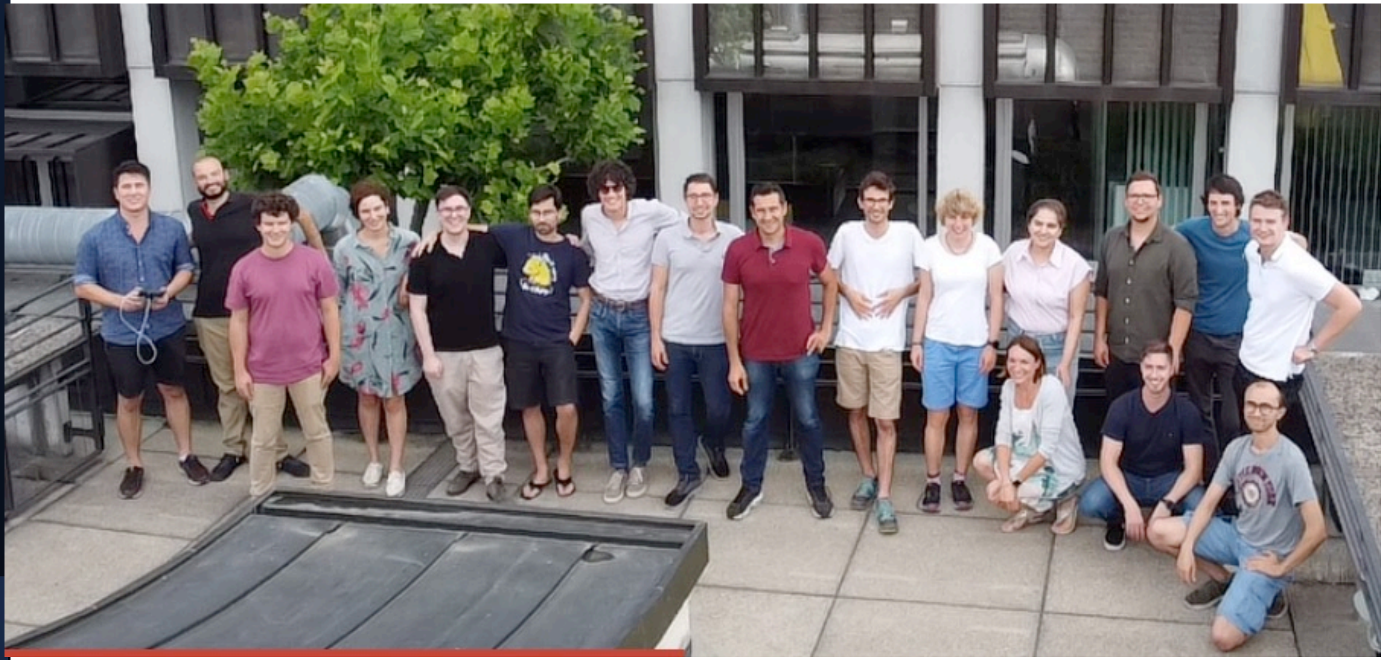
- Considering the latency and energy consumed during the preprocessing of event data caused by the limited neuromorphic resources on Kraken, an alternative edge neuromorphic platform will be explored.
- Collect an event dataset for obstacle avoidance / obejct detection with our ColibriUAV platform.
- Explore tinySNN to fit the edge neuromorphic hardware resources while preserving the accuracy.

Basic Pipeline:



Thank you
pbl.ee.ehtz.ch

Michele.magno@pbl.ee.ethz.ch



Copyright Notice

TINY ML EMEA
AMSTERDAM, NL
June 26-28,2023

This presentation in this publication was presented as a tinyML® EMEA Innovation Forum. The content reflects the opinion of the author(s) and their respective companies. The inclusion of presentations in this publication does not constitute an endorsement by tinyML Foundation or the sponsors.

There is no copyright protection claimed by this publication. However, each presentation is the work of the authors and their respective companies and may contain copyrighted material. As such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

tinyML is a registered trademark of the tinyML Foundation.

www.tinyml.org