TinyDEVO: Deep Event-based Visual Odometry on Ultra-low-power Multi-core Microcontrollers

Alessandro Marchei[‡] ***Lorenzo Lamberti***[‡§] **Daniele Palossi**[‡§] **Luca Benini**[‡*]

ETH Zürich[‡] IDSIA, USI-SUPSI[§] University of Bologna[*]

Swiss National Science Foundation

RoboMix2 project (#10004854)

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

PULP Platform

Open Source Hardware, the way it should be!

@pulp_platform

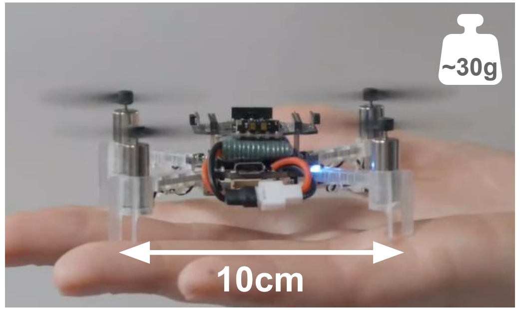
pulp-platform.org

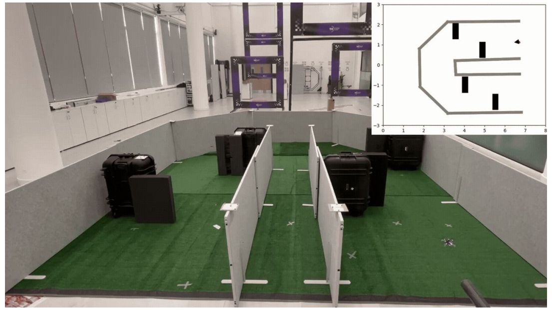
youtube.com/pulp_platform

Visual Odometry and Event-based cameras

Autonomous tiny robots and wearable devices rely on **state estimation** (e.g., pose, position, velocity, etc.).

Pocket-sized drones





Planning, localization, mapping

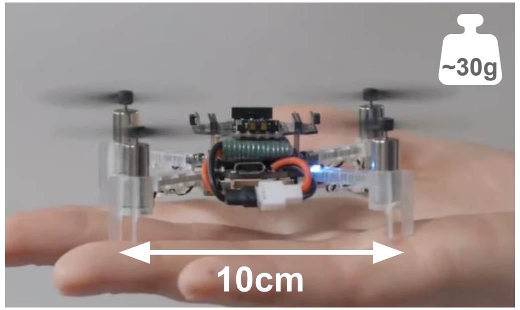
AR/VR devices, smart glasses

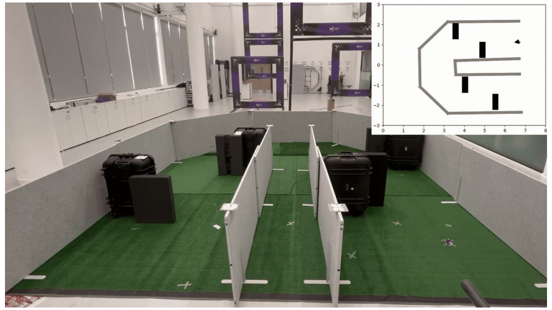
Head motion tracking & image rendering

Visual Odometry and Event-based cameras

Autonomous tiny robots and wearable devices rely on **state estimation** (e.g., pose, position, velocity, etc.).

Pocket-sized drones



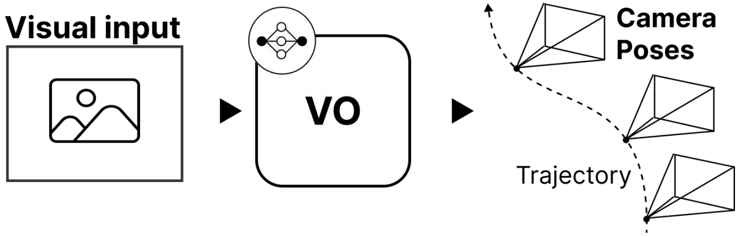


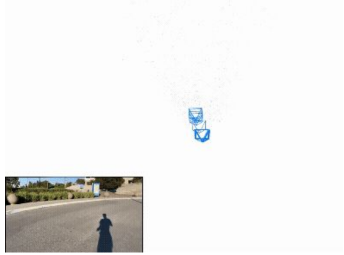
Planning, localization, mapping

AR/VR devices, smart glasses

Head motion tracking & image rendering

An effective solution is **visual odometry (VO)**

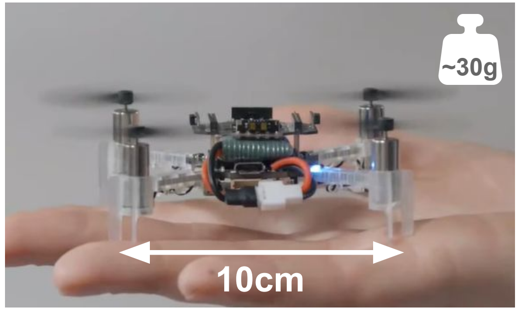


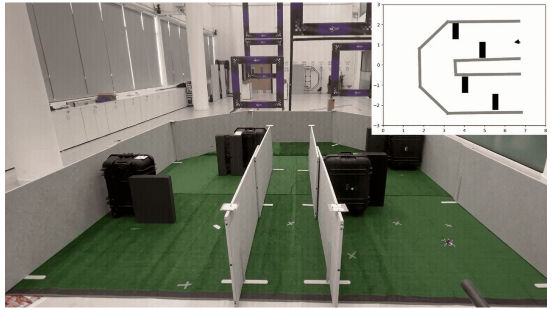


Visual Odometry and Event-based cameras

Autonomous tiny robots and wearable devices rely on **state estimation** (e.g., pose, position, velocity, etc.).

Pocket-sized drones



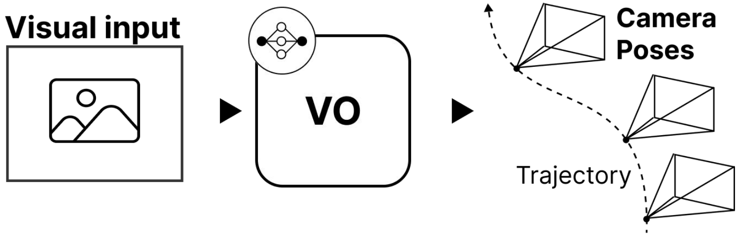


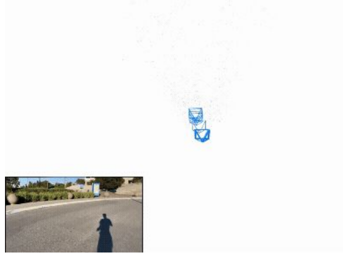
Planning, localization, mapping

AR/VR devices, smart glasses


Head motion tracking & image rendering

An effective solution is **visual odometry (VO)**





State-of-the-Art (SoA) VO use event-cameras

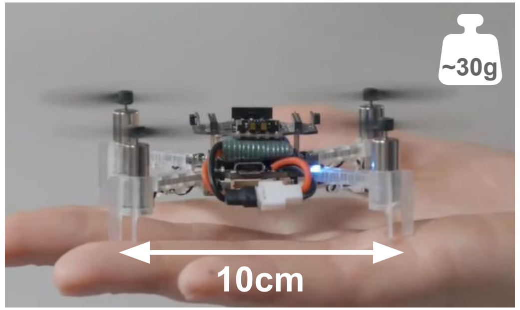

ADVANTAGES

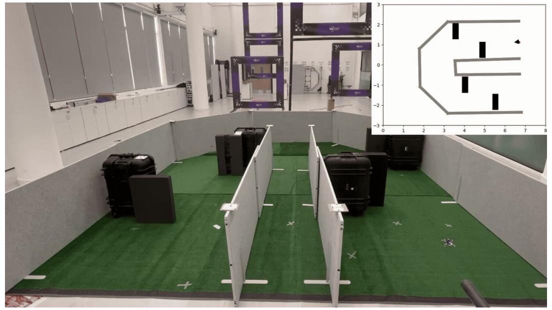
- **High dynamic range**
- **Motion-blur robustness**
- **Low power (<10mW)**

Visual Odometry and Event-based cameras

Autonomous tiny robots and wearable devices rely on **state estimation** (e.g., pose, position, velocity, etc.).

Pocket-sized drones



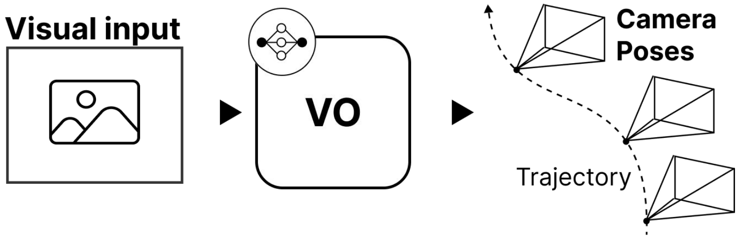


Planning, localization, mapping

AR/VR devices, smart glasses


Head motion tracking & image rendering

An effective solution is **visual odometry (VO)**



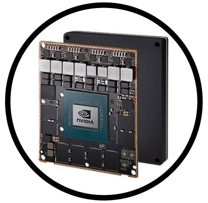
State-of-the-Art (SoA) VO use event-cameras


ADVANTAGES
- **High dynamic range**
- **Motion-blur robustness**
- **Low power (<10mW)**

But battery-powered and energy-constrained platforms **cannot afford State-of-the-Art event-based VO onboard.**

Challenges and contribution

Requirements VO algorithms



GPU

~250W

High power budget

733MB

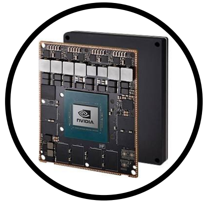
Large memory

vs.

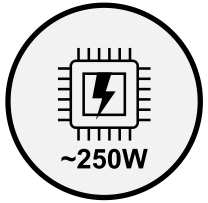
Constraints extreme edge devices

Challenges and contribution

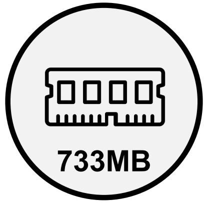
Requirements VO algorithms



GPU



High power budget

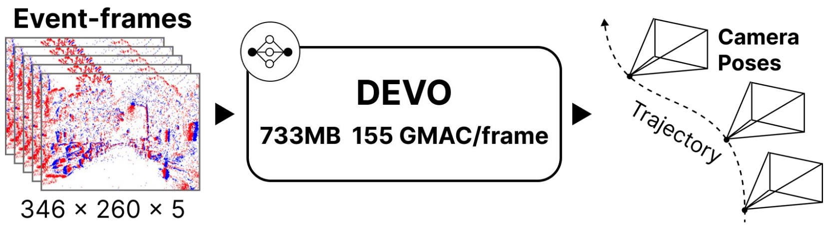


Large memory

vs.

Constraints extreme edge devices

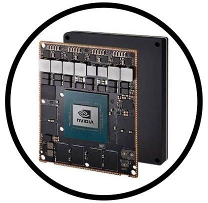
The SoA event-based VO algorithm, **DEVO** [1], runs on RTX4070 GPU (>250W) and requires 733MB of memory.



[1] S. Klenk et al., "Deep Event Visual Odometry," 3DV, 2024

Challenges and contribution

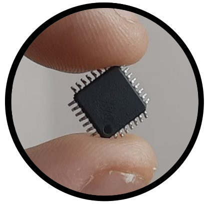
Requirements VO algorithms



- **GPU**
- **~250W** — **High power budget**
- **733MB** — **Large memory**

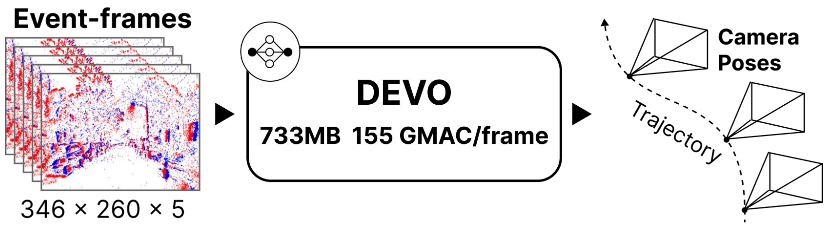
vs.

Constraints extreme edge devices



- **MCU**
- **<100mW** — **Limited computing power budget**
- **few MB** — **Limited on-chip memory**

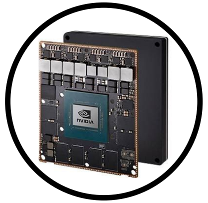
The SoA event-based VO algorithm, **DEVO** [1], runs on RTX4070 GPU (>250W) and requires 733MB of memory.



[1] S. Klenk et al., "Deep Event Visual Odometry," 3DV, 2024

Challenges and contribution

Requirements VO algorithms



GPU

~250W

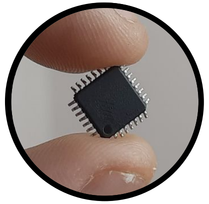
High power budget

733MB

Large memory

vs.

Constraints extreme edge devices



MCU

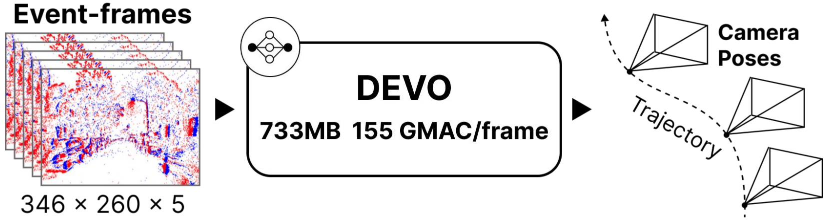
<100mW

Limited computing power budget

few MB

Limited on-chip memory

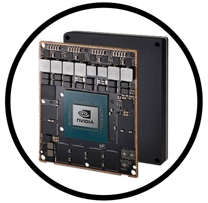
The SoA event-based VO algorithm, **DEVO** [1], runs on RTX4070 GPU (>250W) and requires 733MB of memory.



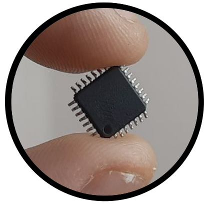
SoA event-based VO algorithm can not be deployed on ultra-low-power MCUs.

[1] S. Klenk et al., "Deep Event Visual Odometry," 3DV, 2024

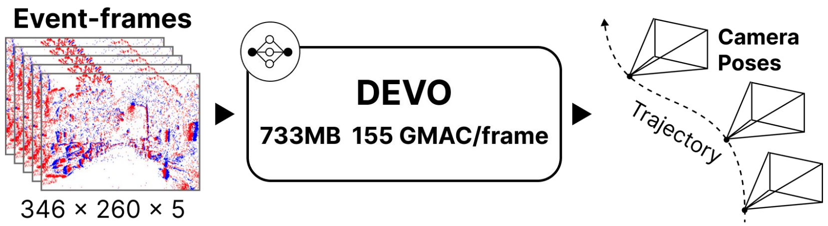
Challenges and contribution

Requirements VO algorithms		
 GPU	~250W High power budget	733MB Large memory

vs.

Constraints extreme edge devices		
 MCU	<100mW Limited computing power budget	few MB Limited on-chip memory

The SoA event-based VO algorithm, **DEVO** [1], runs on RTX4070 GPU (>250W) and requires 733MB of memory.



SoA event-based VO algorithm can not be deployed on ultra-low-power MCUs.

Until today!

Our contributions

1. **TinyDEVO:** event-based VO algorithm tailored to MCUs

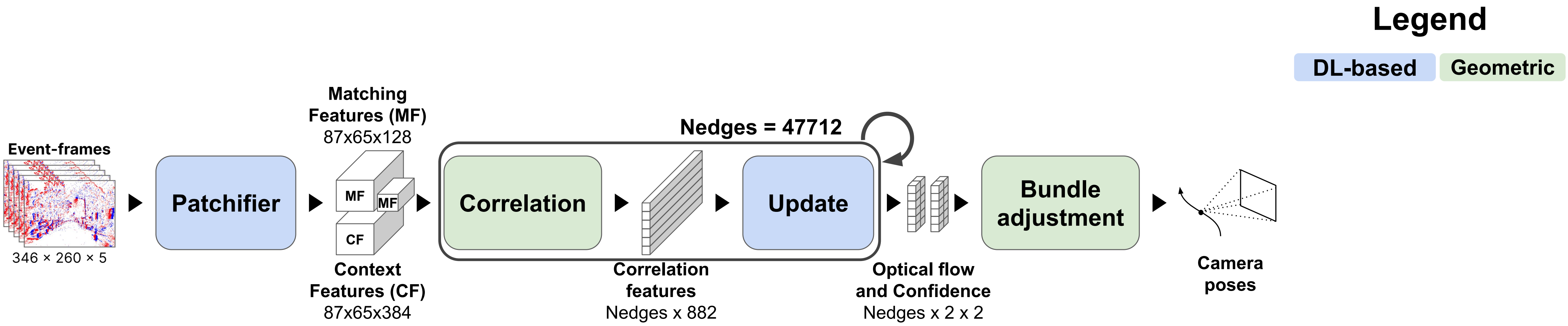
2. Energy-efficient implementation on an ULP multi-core RISC-V MCU.

3. **Trade-off analysis** between execution performance and VO's accuracy

[1] S. Klenk et al., "Deep Event Visual Odometry," 3DV, 2024

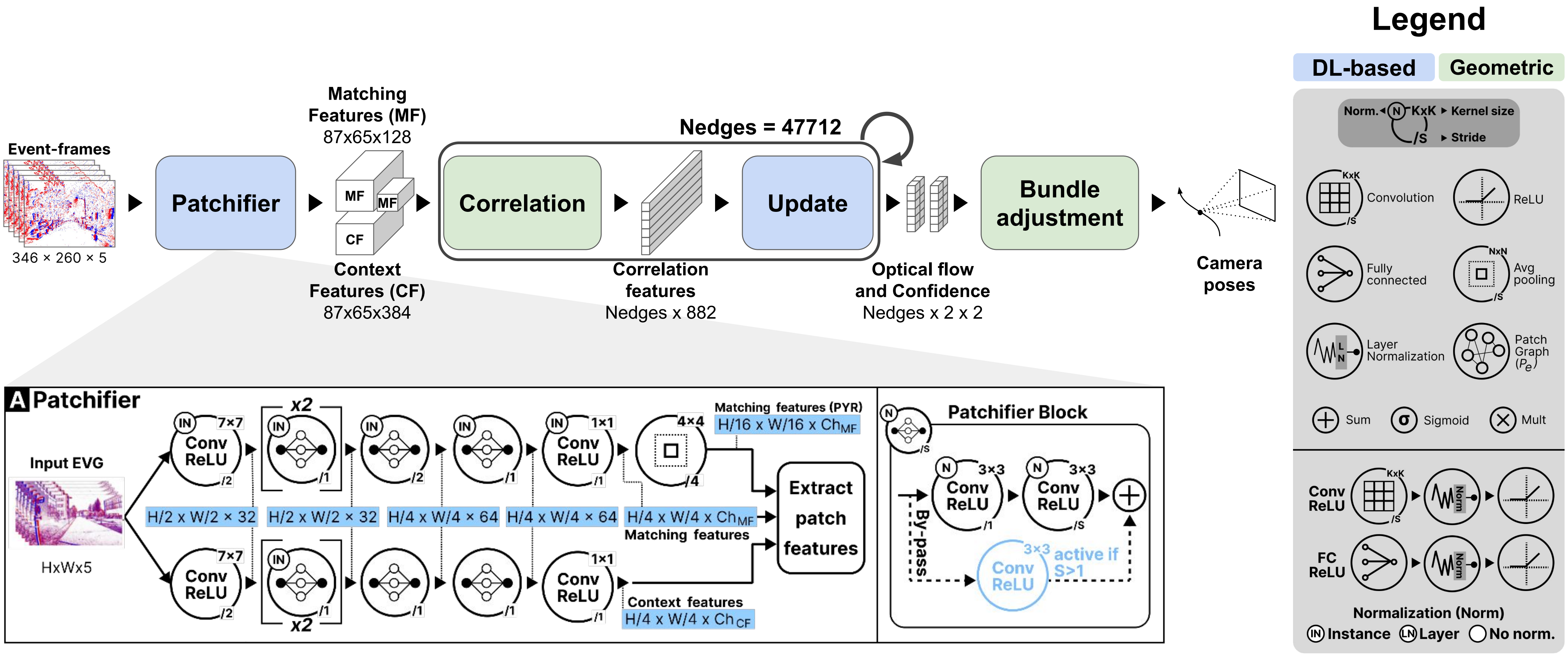
TinyDEVO: architecture optimization

Goal: reduce operations and memory requirements of SoA DEVO algorithm



TinyDEVO: architecture optimization

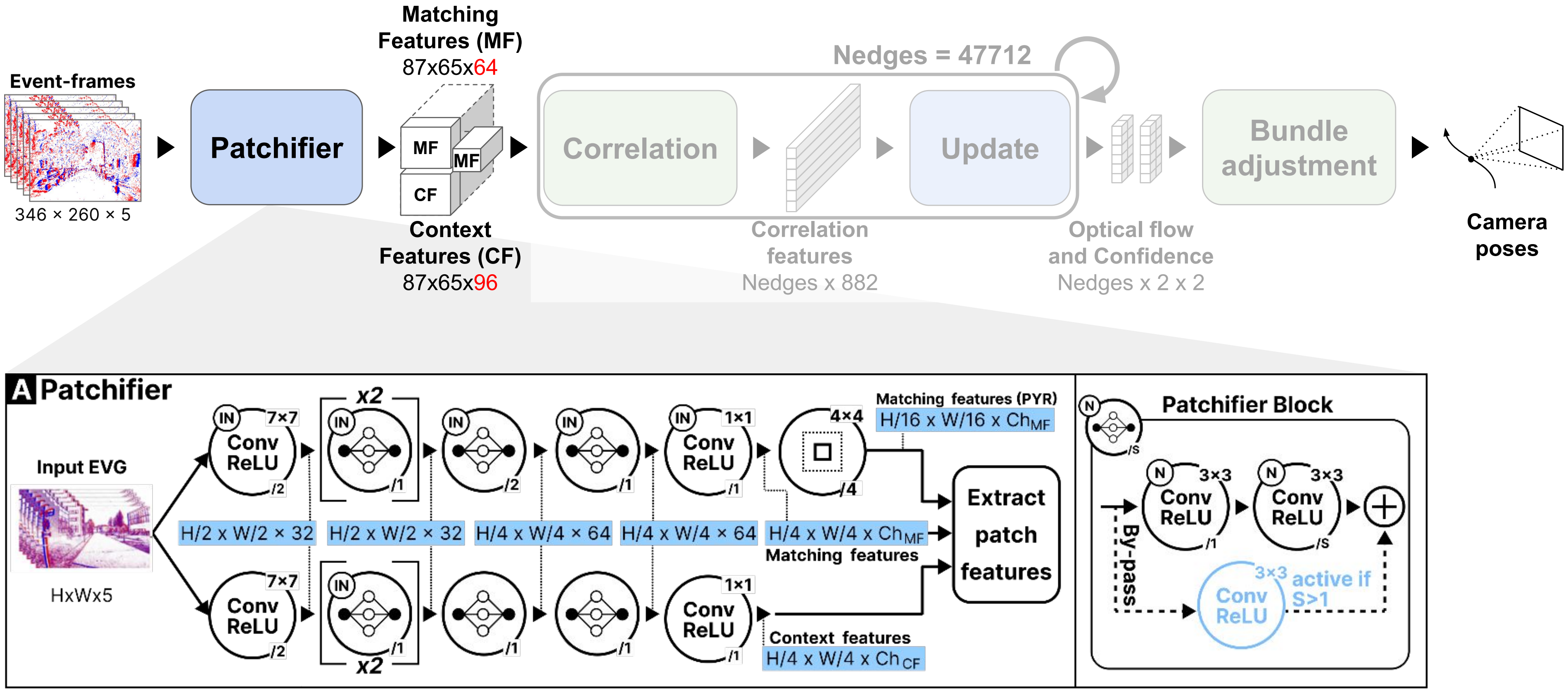
Goal: reduce operations and memory requirements of SoA DEVO algorithm



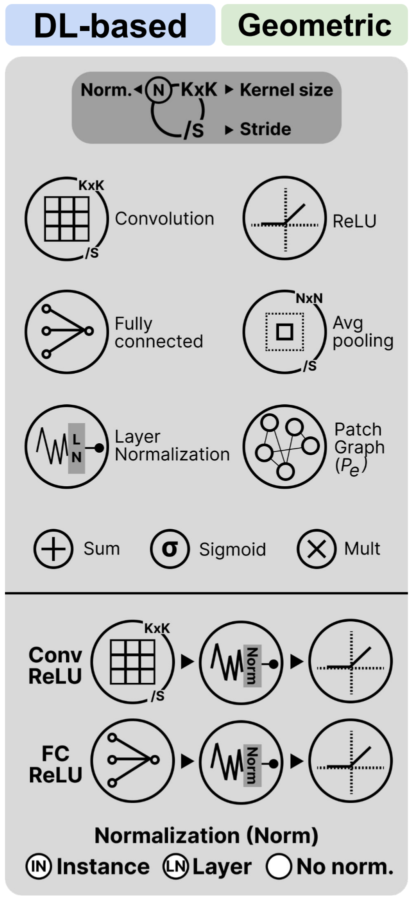
TinyDEVO: architecture optimization

Goal: reduce operations and memory requirements of SoA DEVO algorithm

1. Patchifier: output size reduction



Legend



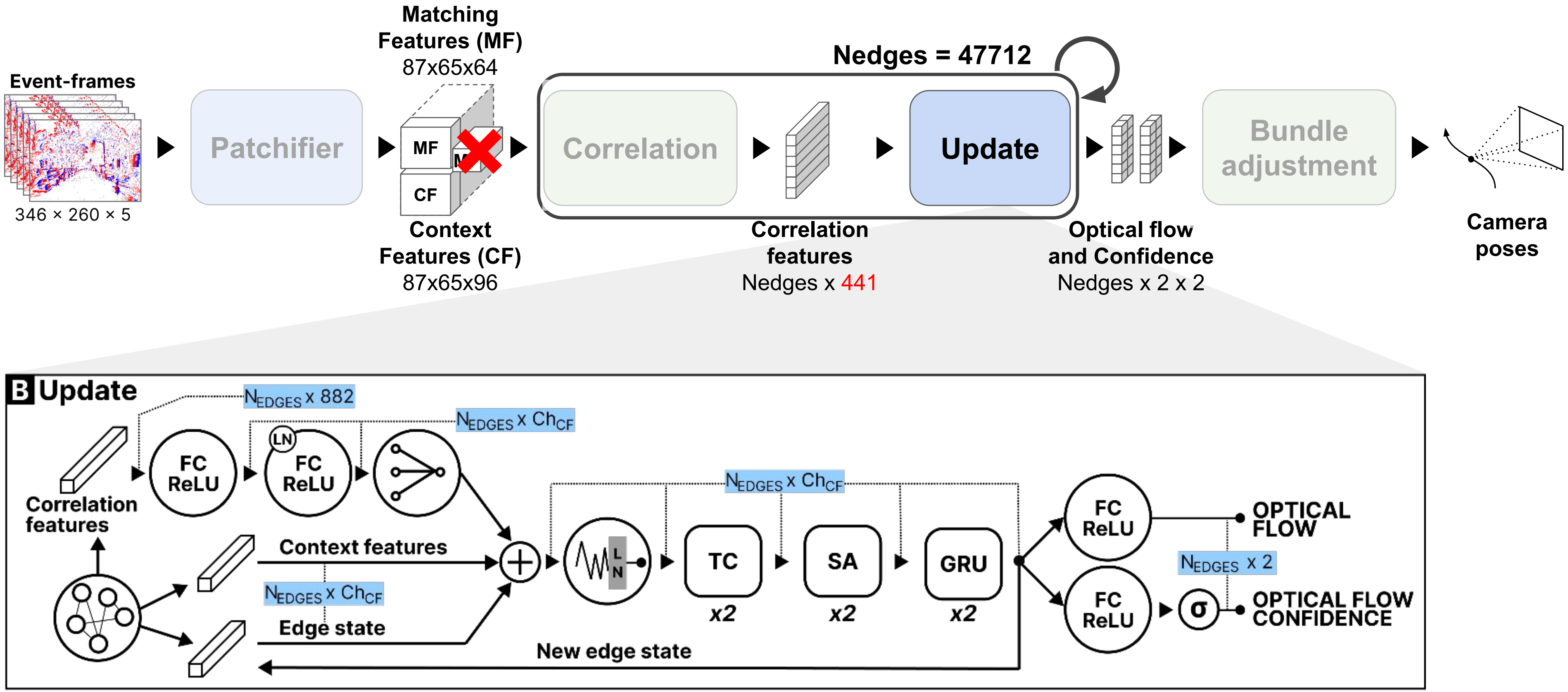
ETH zürich — Alma Mater Studiorum Università di Bologna — idsia Dalle Molle Institute for Artificial Intelligence USI-SUPSI

TinyDEVO: architecture optimization

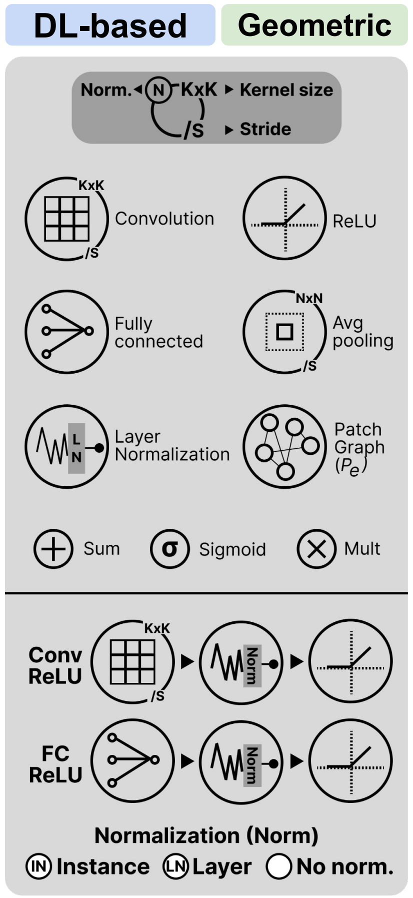
Goal: reduce operations and memory requirements of SoA DEVO algorithm

1. Patchifier: output size reduction

2. Update: architectural modifications



Legend

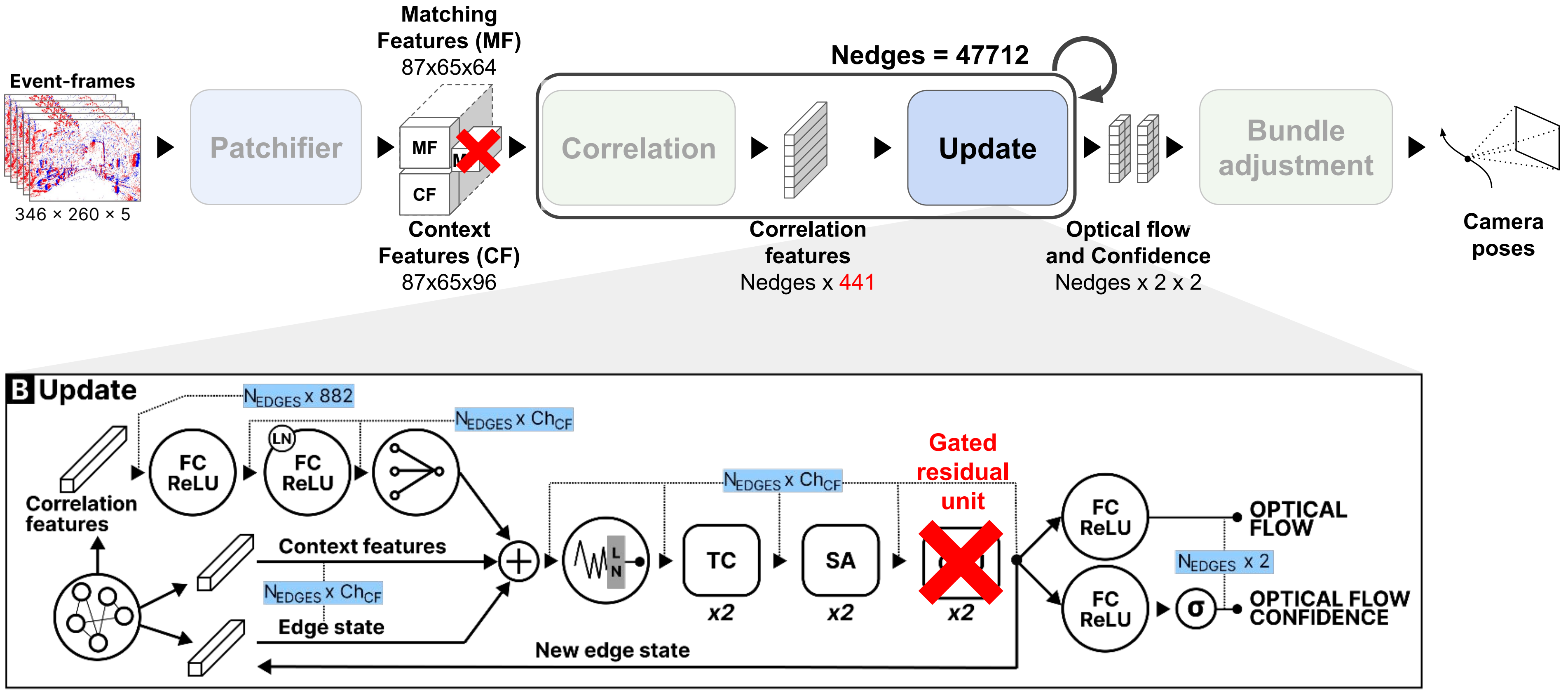


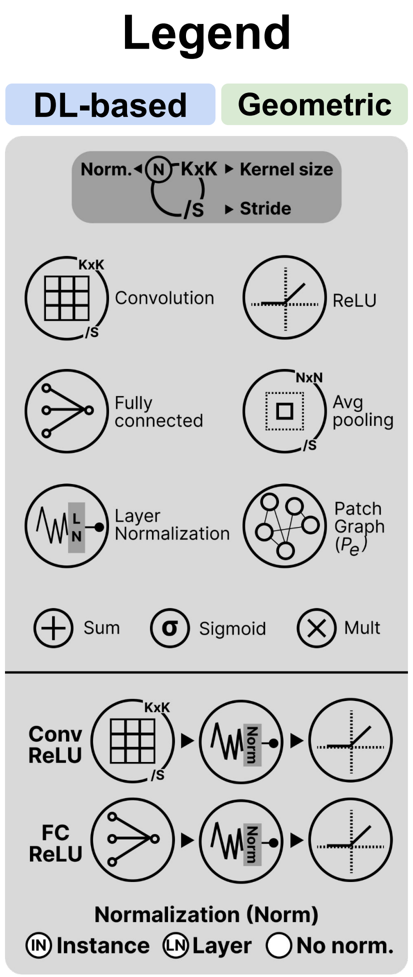
TinyDEVO: architecture optimization

Goal: reduce operations and memory requirements of SoA DEVO algorithm

1. Patchifier: output size reduction

2. Update: architectural modifications





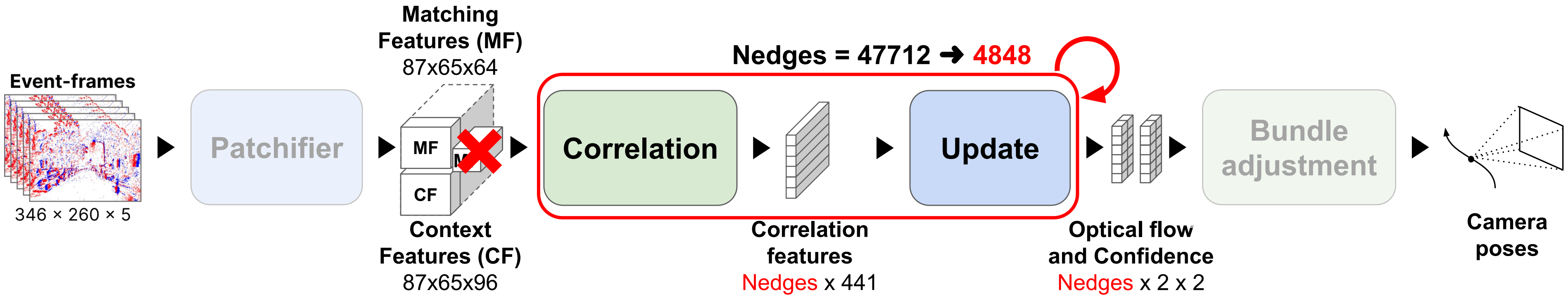
TinyDEVO: architecture optimization

Goal: reduce operations and memory requirements of SoA DEVO algorithm

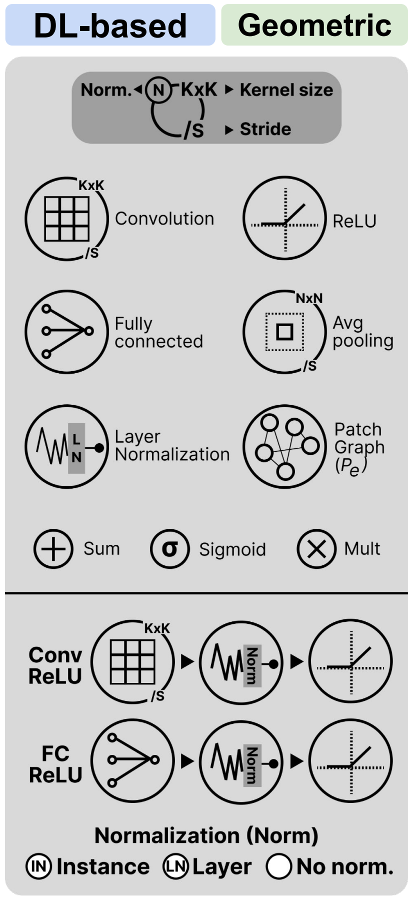
1. Patchifier: output size reduction

2. Update: architectural modifications

3. Reduce Correlation/Update iterations via parameter tuning (Npatches, Nedges, #frames in memory)



Legend



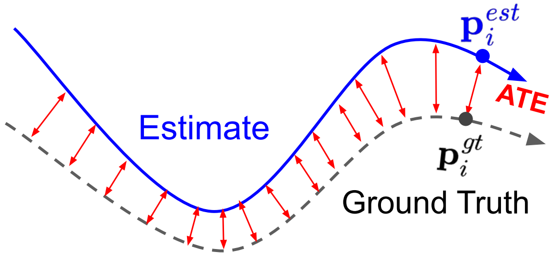
ETH zürich · Alma Mater Studiorum Università di Bologna · idsia Dalle Molle Institute for Artificial Intelligence USI-SUPSI

Results

Evaluated on three datasets: MVSEC, HKU, RPG

Model	ATE (cm) ↓			Peak Memory (MB)	Compute cost (Ops/frame)	Device	Power
	MVSEC	HKU	RPG				
DEVO (SoA)							
TinyDEVO (ours)							

Evaluation Metric: Absolute trajectory error (ATE)



$$\mathrm{ATE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{p}_i^{gt} - \mathbf{p}_i^{est}\right\|^2}$$

Results

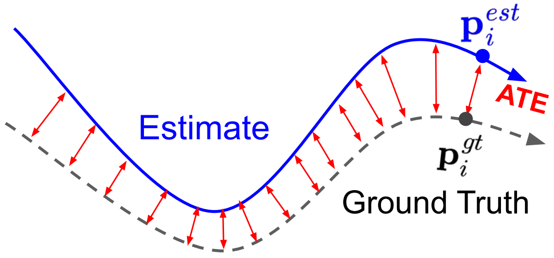
Evaluated on three datasets: MVSEC, HKU, RPG

Model	ATE (cm) ↓			Peak Memory (MB)	Compute cost (Ops/frame)	Device	Power
	MVSEC	HKU	RPG				
DEVO (SoA)	8.25	25.9	0.9				
TinyDEVO (ours)	27.0	45.3	4.9				

1.8x - 5.3x more

Average trajectory length		
MVSEC	HKU	RPG
31m	68m	10m

Evaluation Metric: Absolute trajectory error (ATE)



$$\text{ATE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{p}_i^{gt} - \mathbf{p}_i^{est}\right\|^2}$$

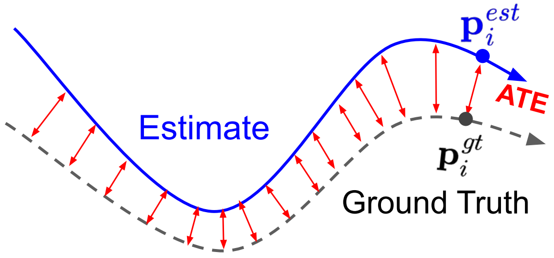
Results

Evaluated on three datasets: MVSEC, HKU, RPG

Model	ATE (cm) ↓			Peak Memory (MB)	Compute cost (Ops/frame)	Device	Power
	MVSEC	HKU	RPG				
DEVO (SoA)	8.25	25.9	0.9	733	155G	GPU	~250W
TinyDEVO (ours)	27.0	45.3	4.9	64	5.2G	MCU	<100mW
	1.8x - 5.3x more			11.5x less	29x less		

Average trajectory length		
MVSEC	HKU	RPG
31m	68m	10m

Evaluation Metric: Absolute trajectory error (ATE)



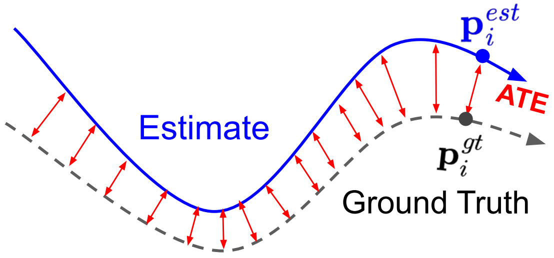
$$\text{ATE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{p}_i^{gt} - \mathbf{p}_i^{est}\right\|^2}$$

Results

Evaluated on three datasets: MVSEC, HKU, RPG

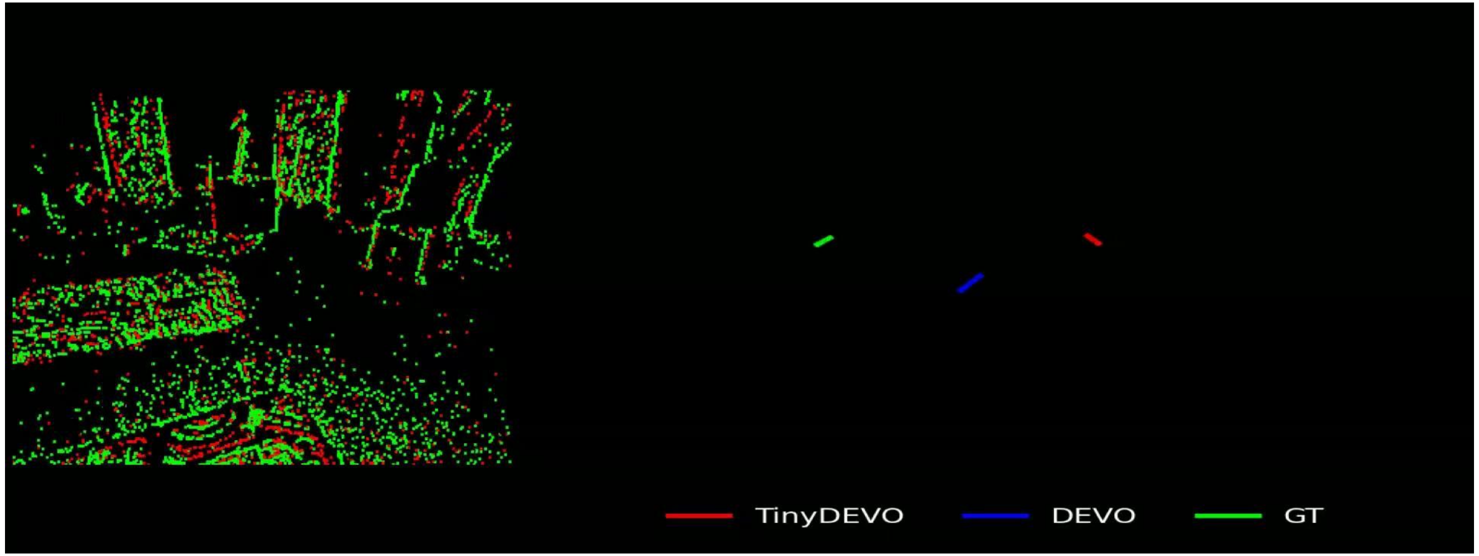
Model	ATE (cm) ↓			Peak Memory (MB)	Compute cost (Ops/frame)	Device	Power
	MVSEC	HKU	RPG				
DEVO (SoA)	8.25	25.9	0.9	733	155G	GPU	~250W
TinyDEVO (ours)	27.0	45.3	4.9	64	5.2G	MCU	<100mW
	1.8x - 5.3x more			11.5x less	29x less		

Evaluation Metric: Absolute trajectory error (ATE)



$$\mathrm{ATE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{p}_i^{gt} - \mathbf{p}_i^{est}\right\|^2}$$

MVSEC indoor scene #3

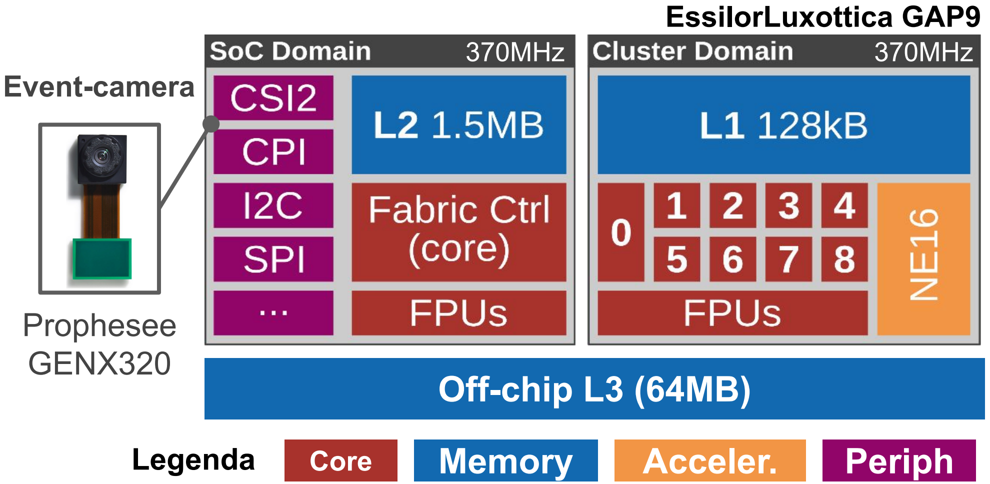


TinyDEVO vs DEVO:

- **<19cm ATE increase**
- **11.5x** less memory
- **29x** less operations

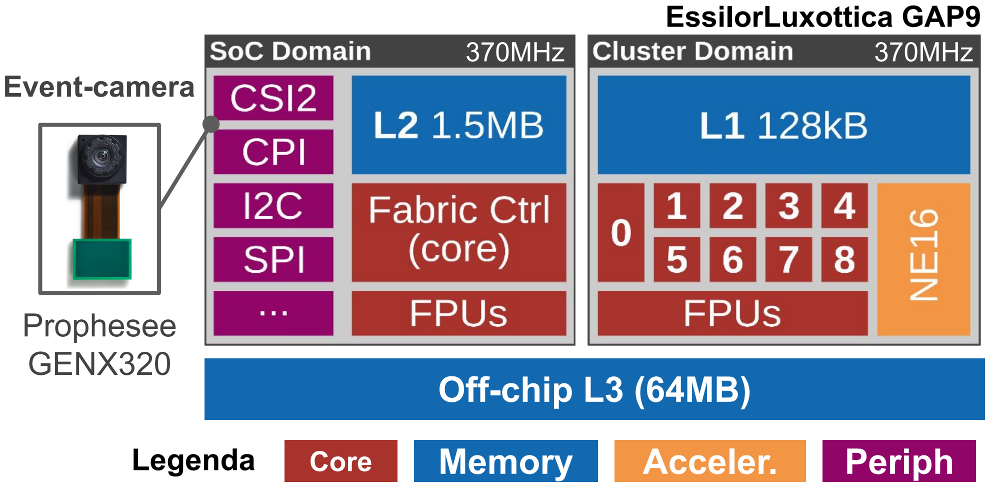
Deployment on a MCU

Target MCU: **GAP9** [2], a commercial embodiment of a **Parallel-ultra-low power (PULP) SoC**

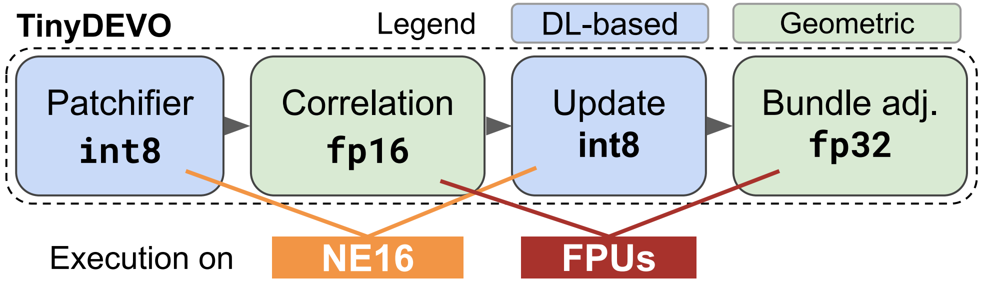


Deployment on a MCU

Target MCU: **GAP9** [2], a commercial embodiment of a **Parallel-ultra-low power (PULP) SoC**

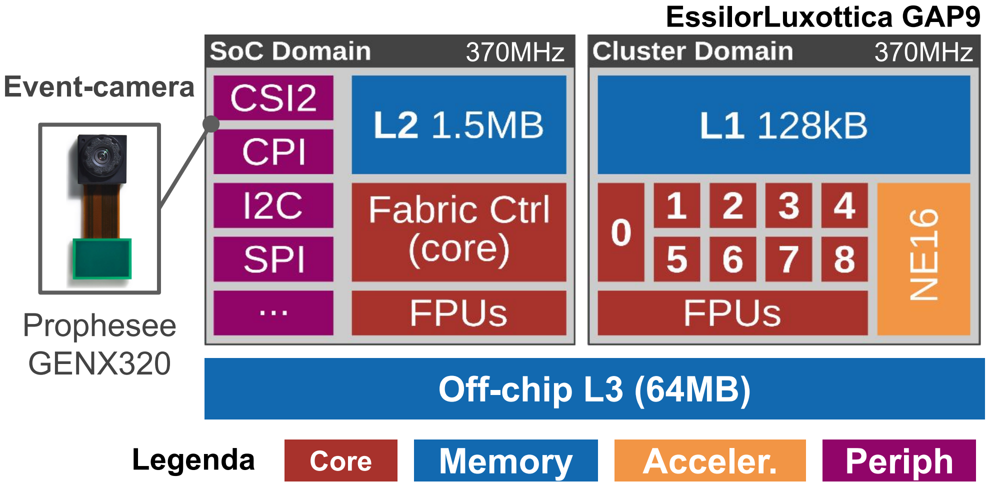


Mixed precision quantization & deployment

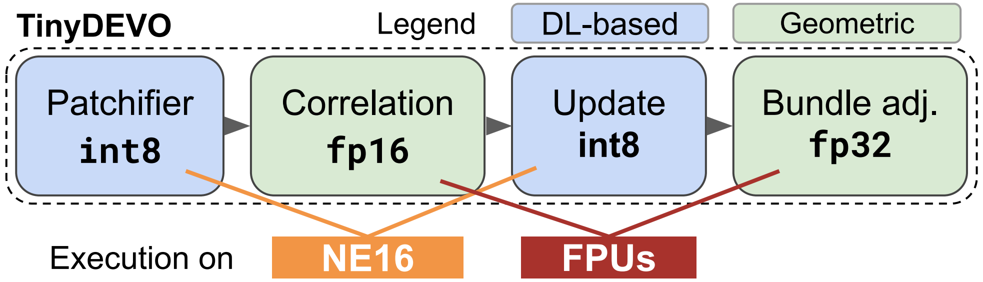


Deployment on a MCU

Target MCU: **GAP9** [2], a commercial embodiment of a **Parallel-ultra-low power (PULP) SoC**



Mixed precision quantization & deployment



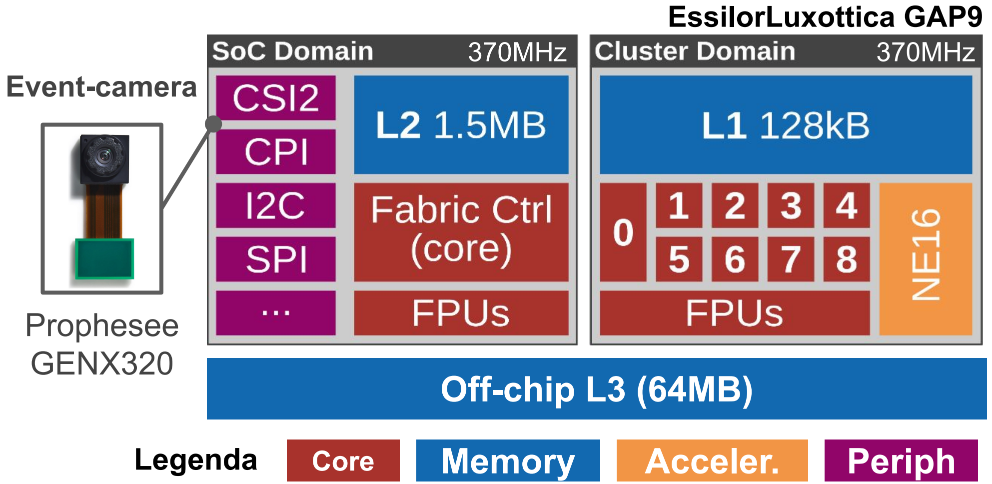
Model	Latency [s]	FPS	Power [mW]
DEVO [1]	45.00*	0.02*	Not deployable
TinyDEVO	0.85	**1.2**	**86**

*estimated

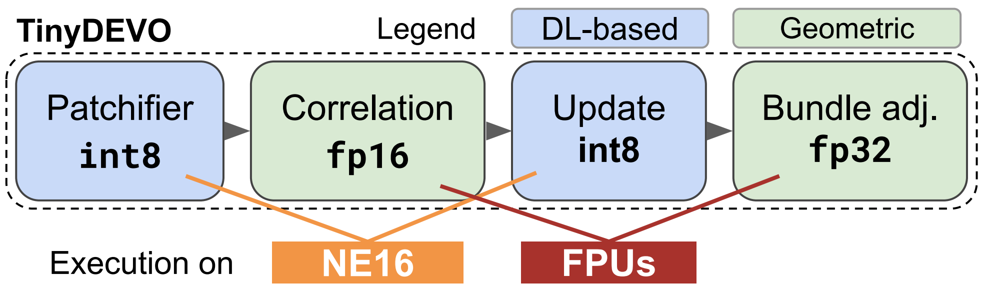
[1] S. Klenk et al., "Deep Event Visual Odometry," 3DV, 2024

Deployment on a MCU

Target MCU: **GAP9** [2], a commercial embodiment of a **Parallel-ultra-low power (PULP) SoC**

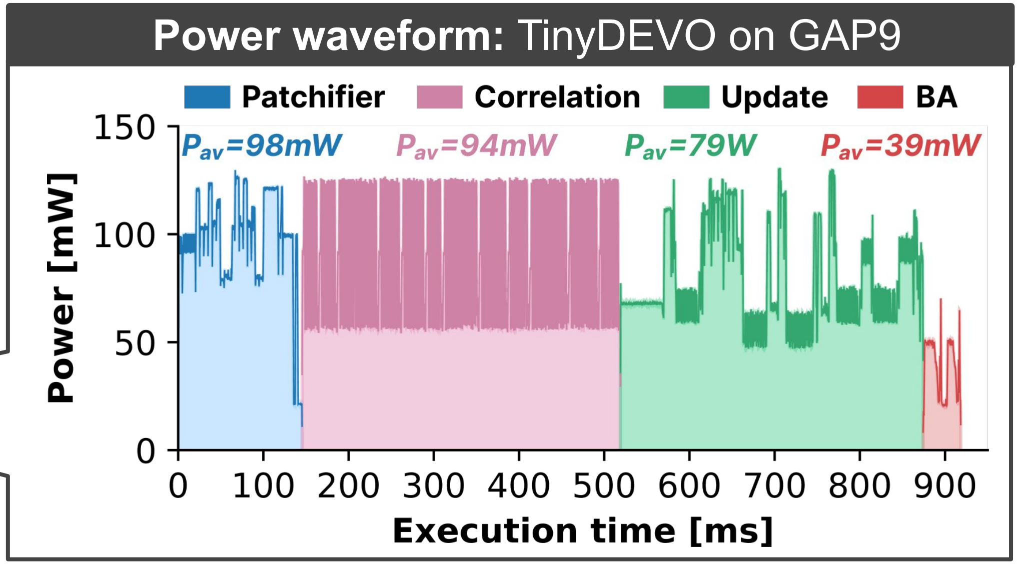


Mixed precision quantization & deployment



Model	Latency [s]	FPS	Power [mW]
DEVO [1]	45.00*	0.02*	Not deployable
TinyDEVO	0.85	1.2	86

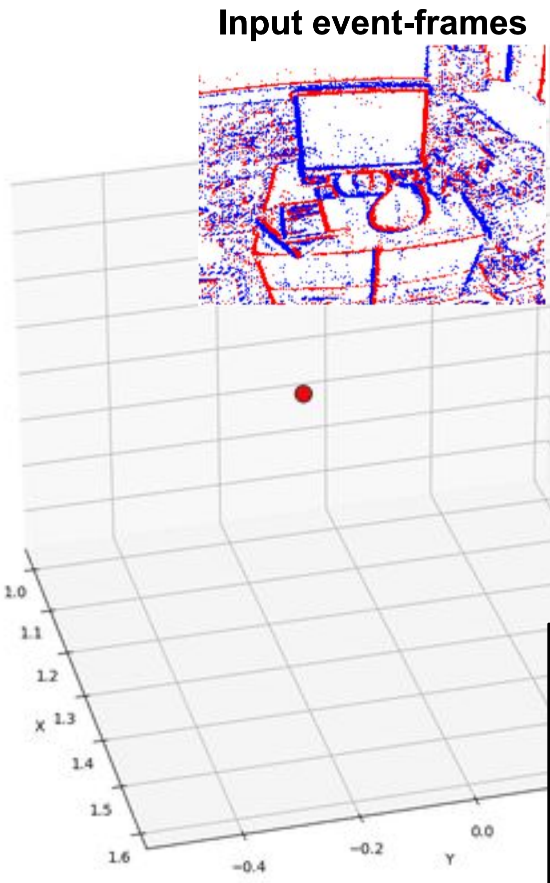
*estimated



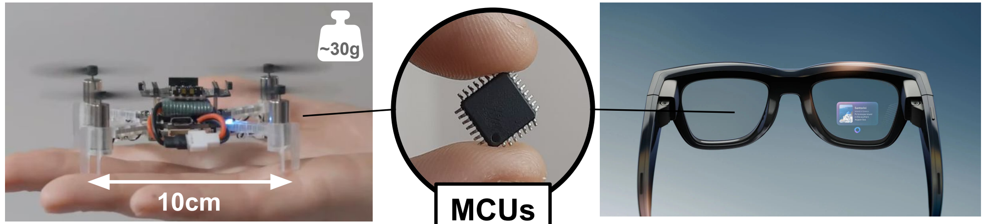
[1] S. Klenk et al., "Deep Event Visual Odometry," 3DV, 2024

Conclusion

We presented **TinyDEVO**, a event-based VO algorithm tailored to MCUs:



Custom PCB (GAP9 + GENX320 event-camera) on a Crazyflie nano-drone.



Paper

Video

Project

ETH zürich

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

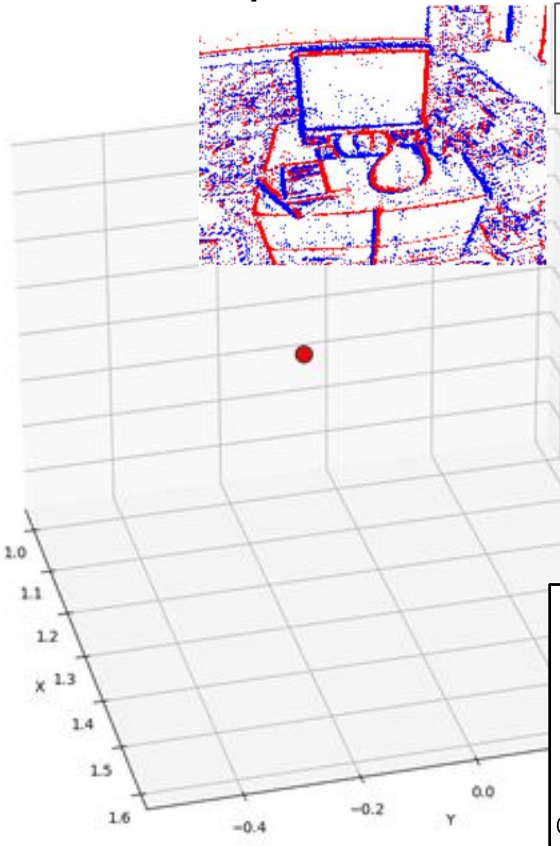
idsia Dalle Molle Institute for Artificial Intelligence USI-SUPSI

Swiss National Science Foundation
RoboMix2 project (#10004854)

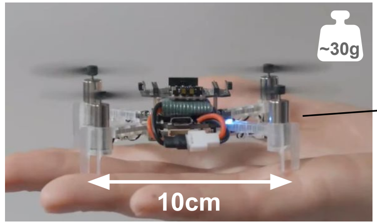
Conclusion

We presented **TinyDEVO**, a event-based VO algorithm tailored to MCUs:

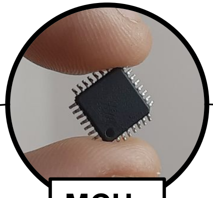
- Memory footprint: 63.8 MB **11.5x** less than DEVO
- Number of operations: 5.2 GMAC/frame **29x** less than DEVO
- Absolute trajectory error**:** 4.5-45cm **<19cm** higher than DE



Custom PCB (GAP9 + GENX320 event-camera) on a Crazyflie nano-drone.



Nano-sized drones


MCUs


Smart glasses

Paper

Video

Project

ETH zürich
ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA
idsia Dalle Molle Institute for Artificial Intelligence USI-SUPSI
Swiss National Science Foundation
RoboMix2 project (#10004854)

Conclusion

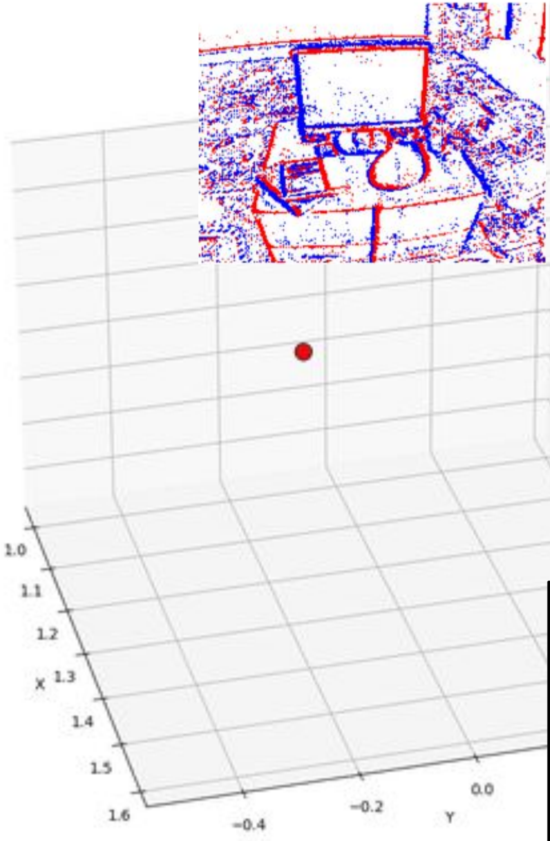
We presented **TinyDEVO**, a event-based VO algorithm tailored to MCUs:

- Memory footprint: 63.8 MB **11.5x** less than DEVO
- Number of operations: 5.2 GMAC/frame **29x** less than DEVO
- Absolute trajectory error**:** 4.5-45cm **<19cm** higher than DEV

When deployed on the GAP9 MCU:

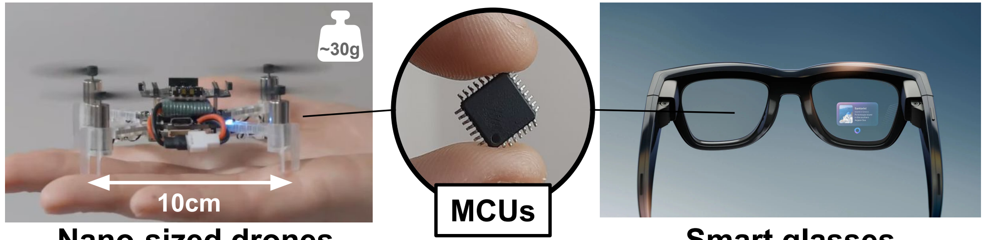
- Inference throughput: **1.2 FPS**
- Average power consumption: **86mW**

Input event-frames



Designed at ETH Zürich

Custom PCB (GAP9 + GENX320 event-camera) on a Crazyflie nano-drone.



Nano-sized drones

Smart glasses

Paper **Video** **Project**

ETH zürich · ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA · idsia Dalle Molle Institute for Artificial Intelligence USI-SUPSI · Swiss National Science Foundation

RoboMix2 project (#10004854)

Conclusion

We presented **TinyDEVO**, a event-based VO algorithm tailored to MCUs:

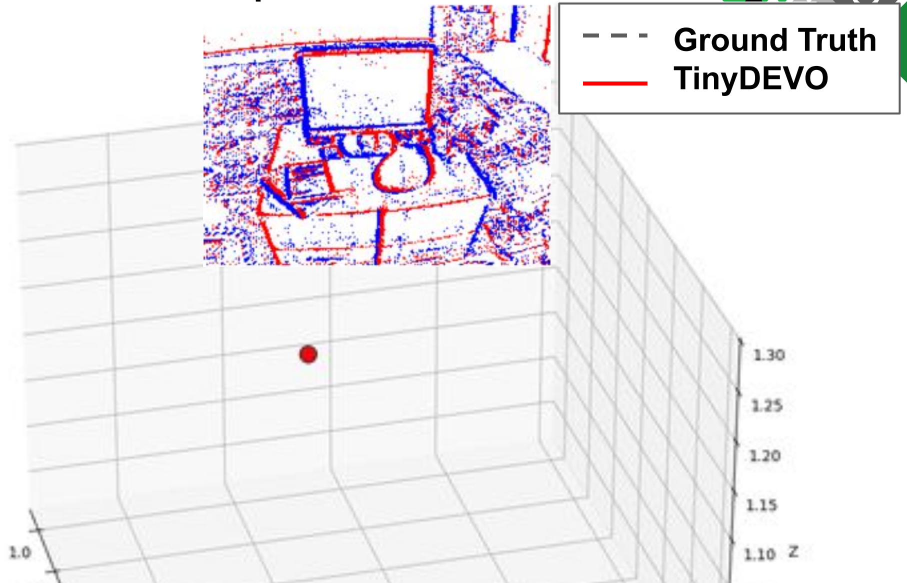
- Memory footprint: 63.8 MB **11.5x** less than DEVO
- Number of operations: 5.2 GMAC/frame **29x** less than DEVO
- Absolute trajectory error**:** 4.5-45cm **<19cm** higher than DE

When deployed on the GAP9 MCU:

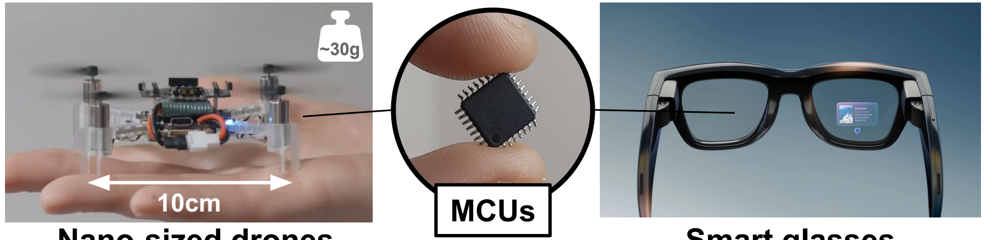
- Inference throughput: **1.2 FPS**
- Average power consumption: **86mW**

We demonstrated the feasibility of a SoA event-based VO pipeline on ULP devices.

Input event-frames



Custom PCB (GAP9 + GENX320 event-camera) on a Crazyflie nano-drone.



Nano-sized drones

Smart glasses

Paper

Video

Project

ETH zürich ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA idsia Dalle Molle Institute for Artificial Intelligence USI-SUPSI

Swiss National Science Foundation
RoboMix2 project (#10004854)

PULP

Parallel Ultra Low Power

Q&A

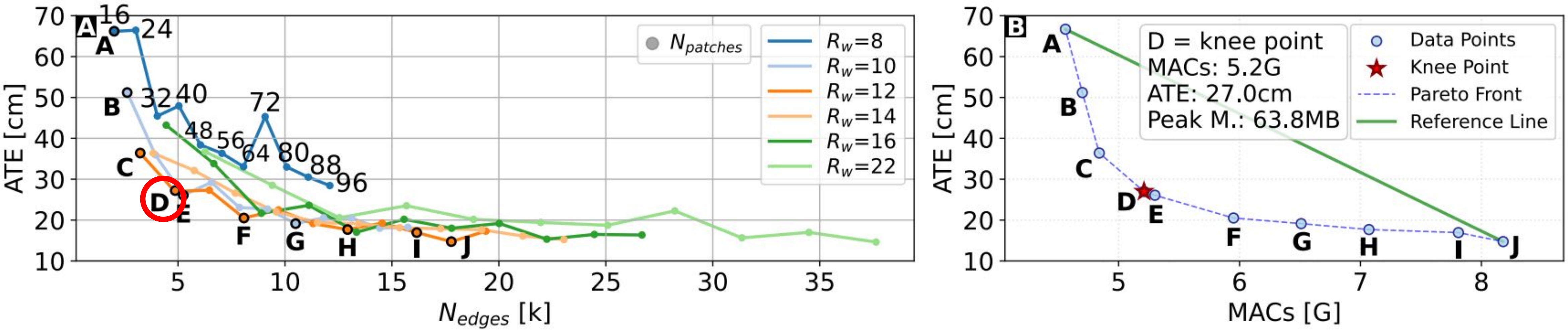
Thank you!

github.com/pulp-platform/pulp-dronet ☆ 600

https://github.com/pulp-platform http://pulp-platform.org @pulp_platform

https://www.youtube.com/c/PULPPlatform

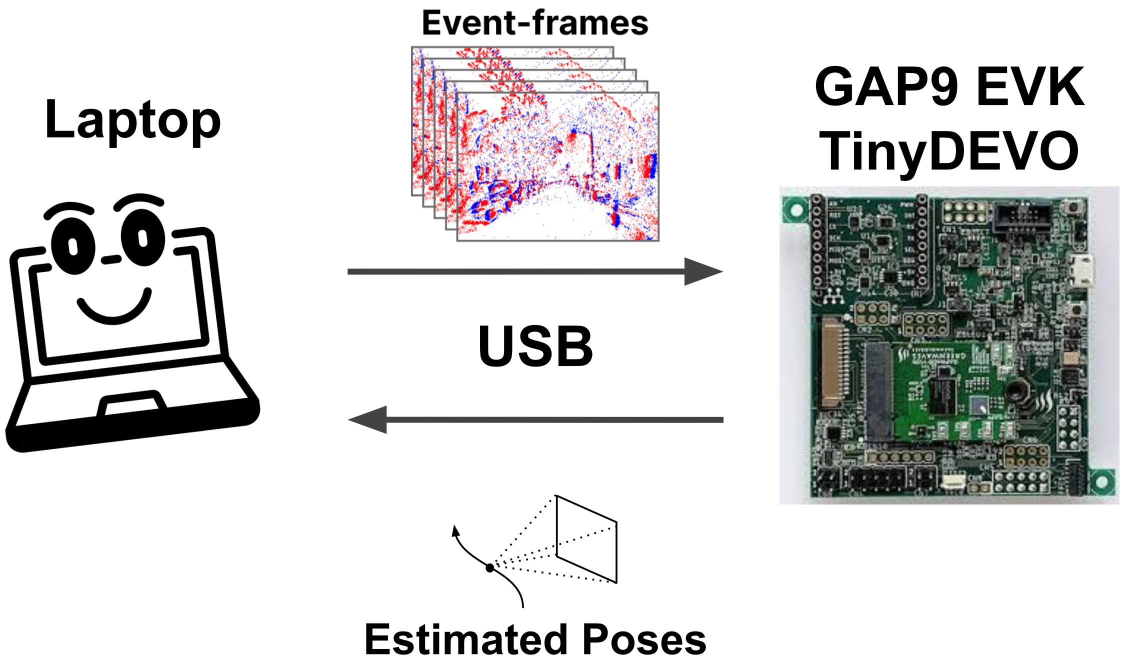
Trade-off: accuracy vs. latency

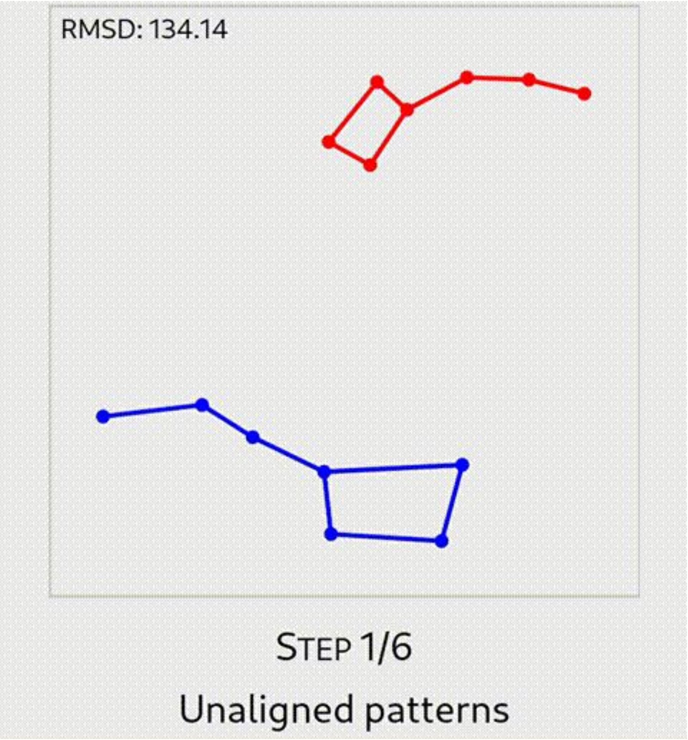


TinyDEVO: evaluation setup

1. Laptop stream input to GAP9
2. GAP9 returns the estimated poses

Umeyama alignment for evaluation:
1. Roto-translation
2. Rescaling





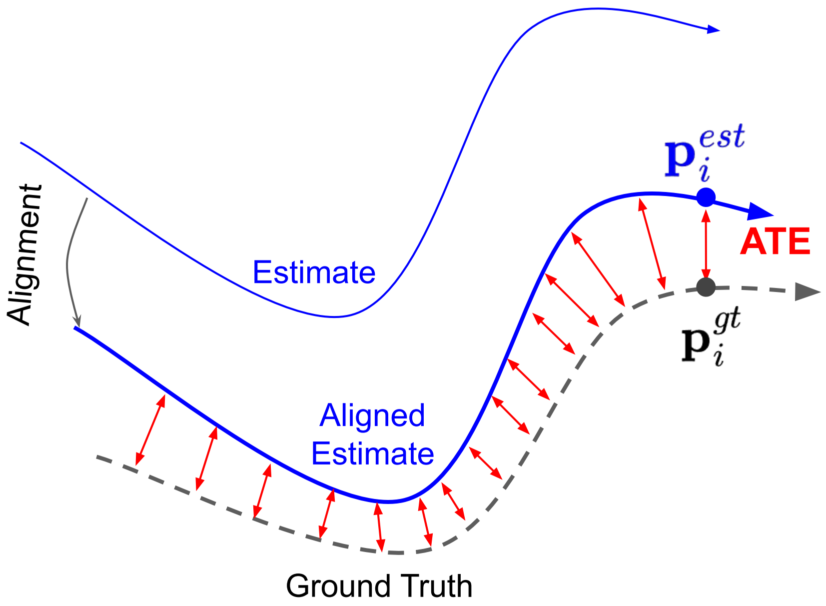
zpl.fi/aligning-point-patterns-with-kabsch-umeyama-algorithm/

Evaluation Setup

Evaluation metric

Absolute trajectory error (ATE)

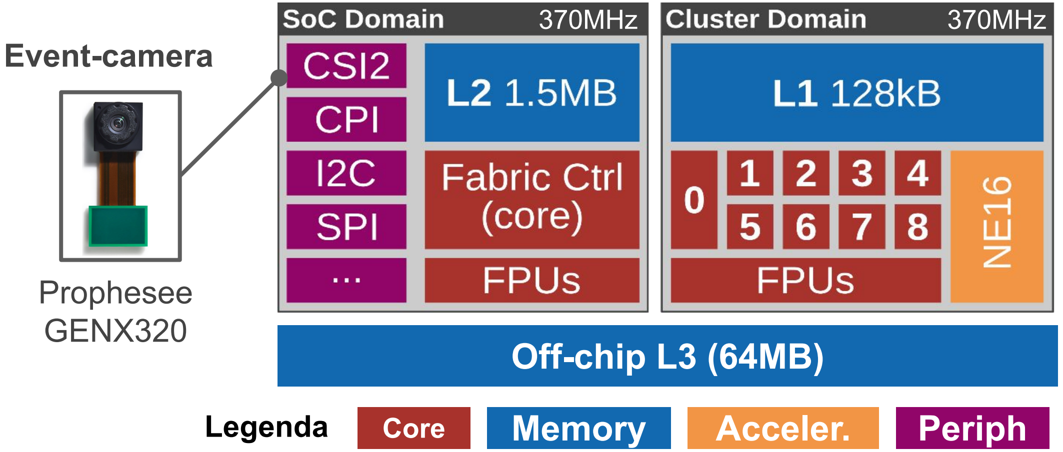
$$\mathrm{ATE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{p}_i^{gt} - \mathbf{p}_i^{est}\right\|^2}$$



Target MCU

GAP9 [2], a commercial embodiment of a **Parallel-ultra-low power (PULP**) SoC:

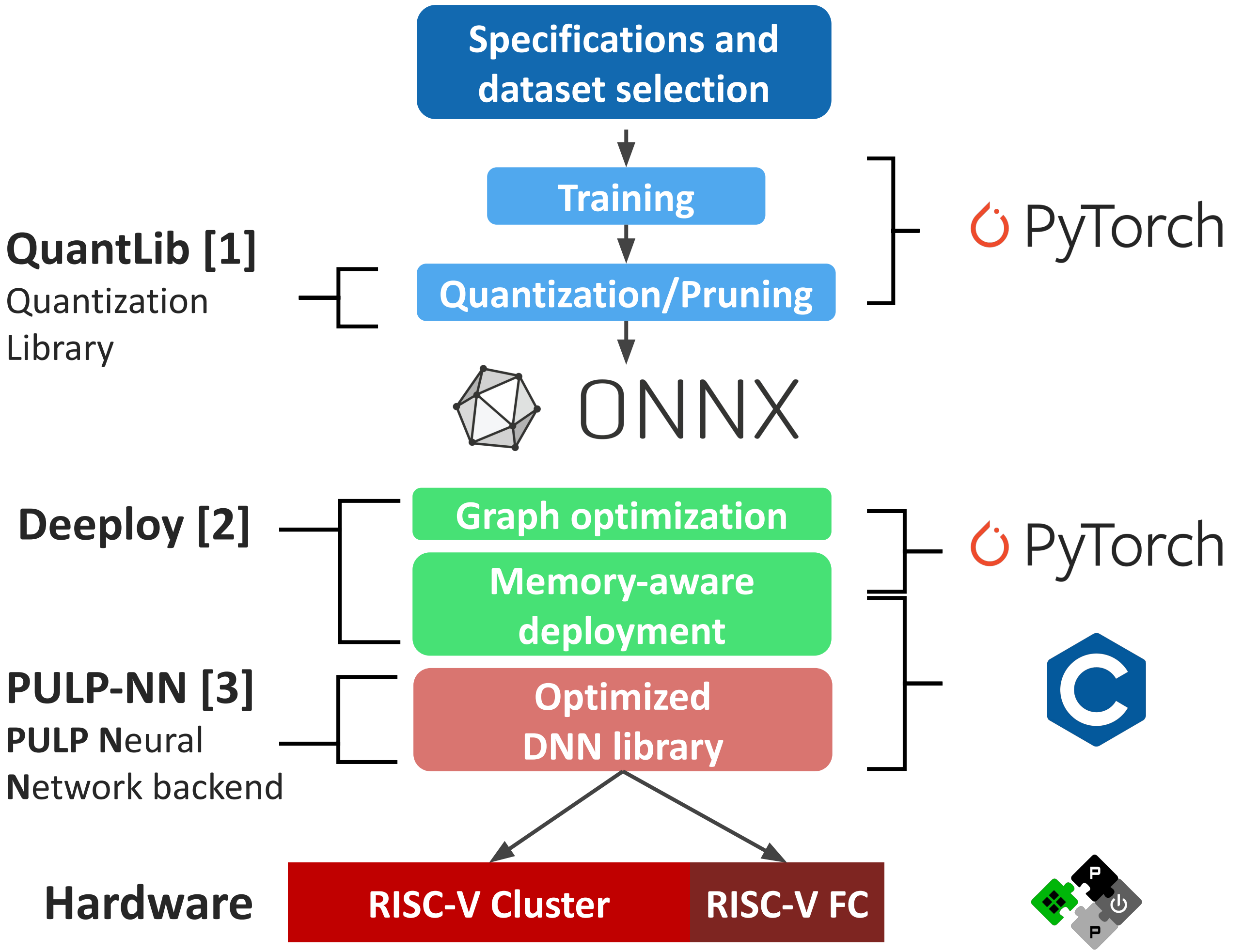
- 1+9 RISC-V cores
- Accelerator: NE16
- VDD 0.6 - 0.8V
- Max freq: 370MHZ



[2] D. Rossi, et al. "Vega: A ten-core SoC for IoT endnodes with DNN acceleration and cognitive wake-up from MRAM-based state-retentive sleep mode." JSSC, 2021

Deploying applications to PULP

A complete and automated vertical software stack



Quantization & model validation

github.com/pulp-platform/quantlib

Deployable C code

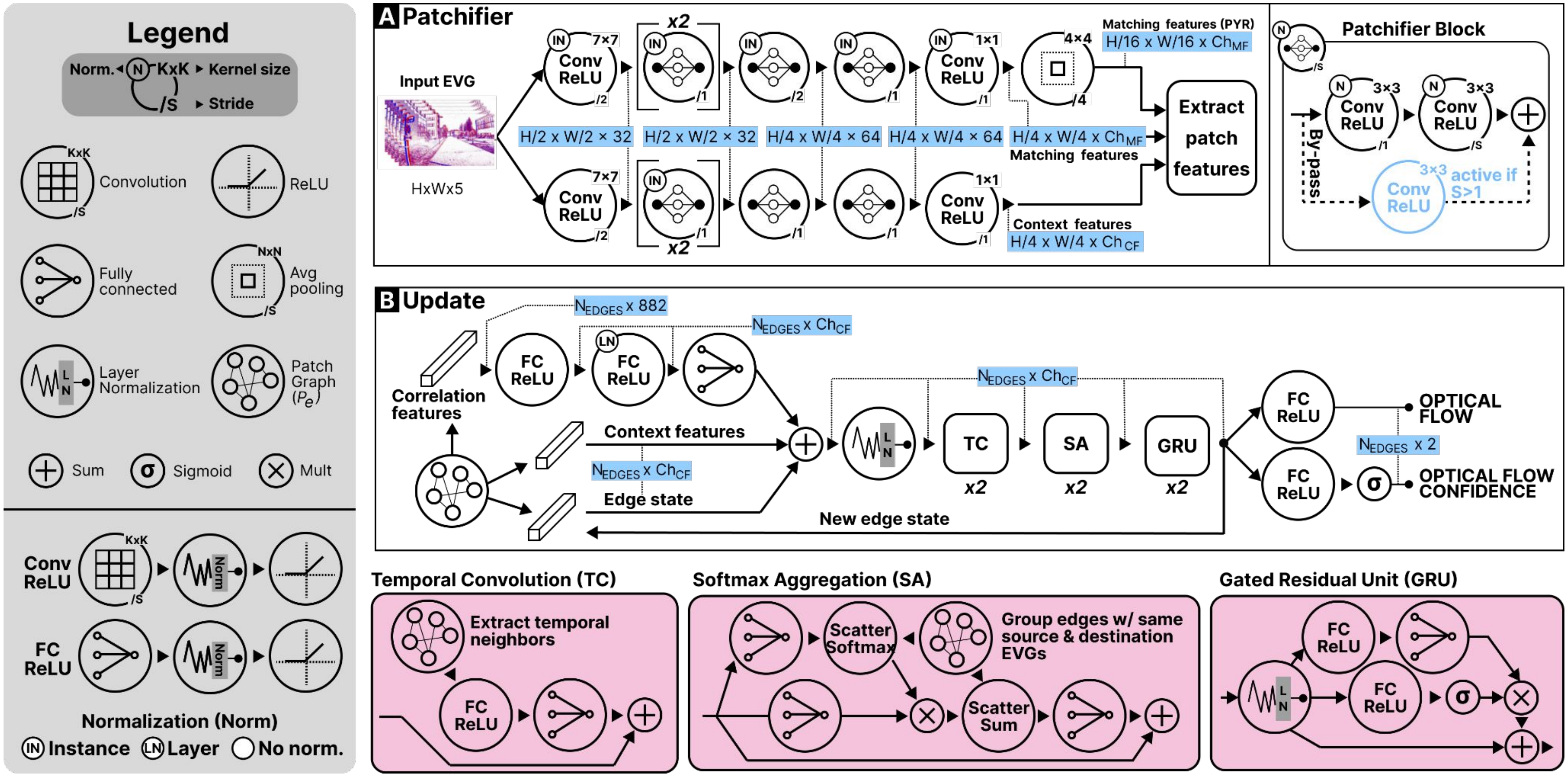
github.com/pulp-platform/deeploy

github.com/pulp-platform/pulp-nn

Optimized low-level CNN libraries

[1] SPALLANZANI, Matteo; LEONARDI, Gian Paolo; BENINI, Luca. Training quantised neural networks with ste variants: the additive noise annealing algorithm. CVPR. 2022. p. 470-479.
[2] M. Scherer *et al.*, “Deeploy: Enabling Energy-Efficient Deployment of Small Language Models On Heterogeneous Microcontrollers”, ArXiv, 2024.
[3] Garofalo, Angelo, et al. "PULP-NN: Accelerating quantized neural networks on parallel ultra-low-power RISC-V processors." Philosophical Transactions of the Royal Society A 378.2164 (2020): 20190155.

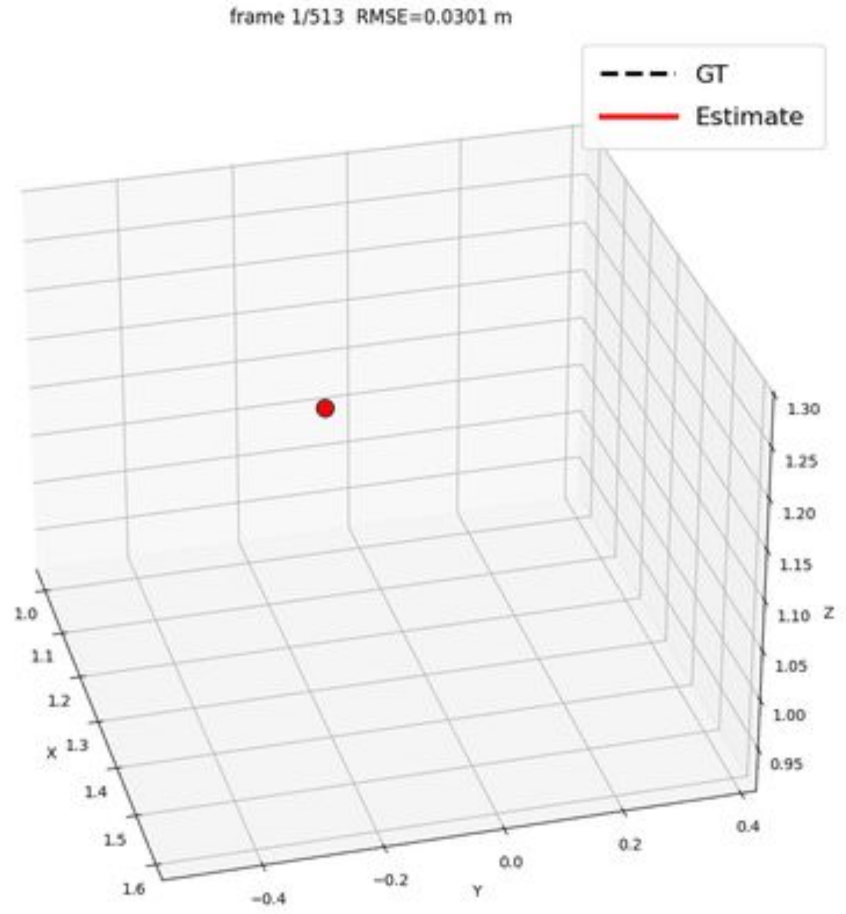
TinyDEVO



Legend
Norm.
KxK
Kernel size
/S
Stride
Convolution
ReLU
Fully connected
NxN
Avg pooling
Layer Normalization
Patch Graph (P_e)
Sum
Sigmoid
Mult
Conv ReLU
Norm
FC ReLU
Normalization (Norm)
IN Instance
LN Layer
No norm.
A Patchifier
Input EVG
HxWx5
7×7
Conv ReLU
/2
x2
/1
1×1
4×4
/4
H/2 x W/2 × 32
H/2 x W/2 × 32
H/4 x W/4 × 64
H/4 x W/4 × 64
H/4 x W/4 x Ch_{MF}
Matching features (PYR)
H/16 x W/16 x Ch_{MF}
Matching features
Extract patch features
Context features
H/4 x W/4 x Ch_{CF}
Patchifier Block
3×3
By-pass
active if S>1
B Update
N_{EDGES} x 882
N_{EDGES} x Ch_{CF}
Correlation features
FC ReLU
Context features
Edge state
N_{EDGES} x Ch_{CF}
N_{EDGES} x Ch_{CF}
TC
SA
GRU
x2
New edge state
OPTICAL FLOW
N_{EDGES} x 2
OPTICAL FLOW CONFIDENCE
Temporal Convolution (TC)
Extract temporal neighbors
Softmax Aggregation (SA)
Scatter Softmax
Group edges w/ same source & destination EVGs
Scatter Sum
Gated Residual Unit (GRU)

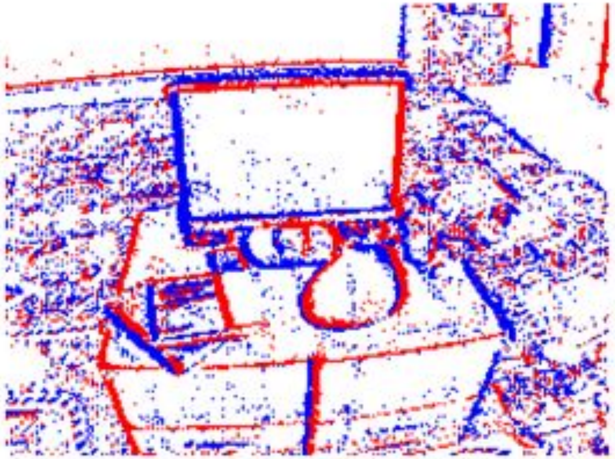
TinyDEVO: end-to-end prediction on GAP9

Evaluation on RPG dataset, monitor2 scene



RGB frames

Not processed by TinyDEVO



Event-frames

TinyDEVO's input

Vision-Shield

Bitcraze's extension PCB integrating the GENX320 event-camera

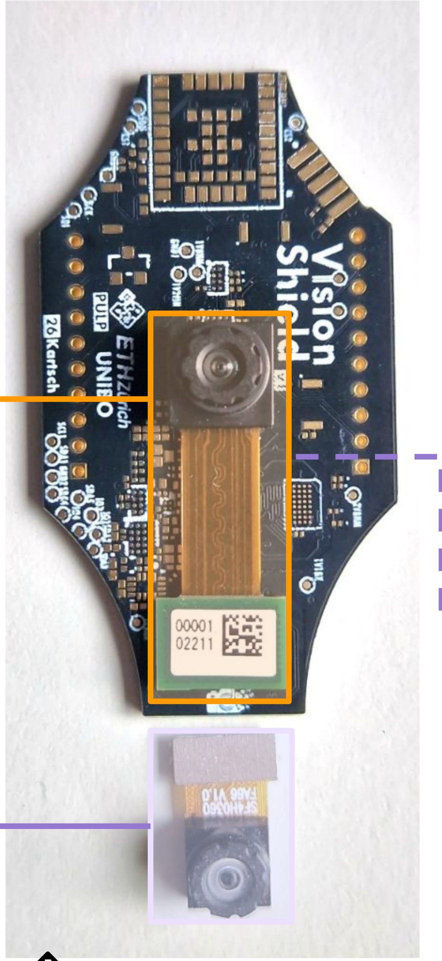
Designed at IIS - ETH Zürich

Vision-Shield

Bitcraze's extension PCB integrating the GENX320 event-camera

Prophesee Genx320 event-camera
- Resolution: 320x320
- MIPI-CSI2, CPI

Himax HM0360 camera
- Resolution: VGA
- RGB / Grayscale
- FPS: up to 60Hz @ VGA.



Designed at IIS - ETH Zürich

GAP9 SoC

Fabric controller (FC): 370 MHz

Cluster (CL): 370 MHz

Hardware acceleration: NE16

On-chip memories:
- 1536 kB L2
- 128 kB L1

Power:
- FC up to 14.8 mW
- CL up to 48.1 mW
- NE16 up to 48.1 mW

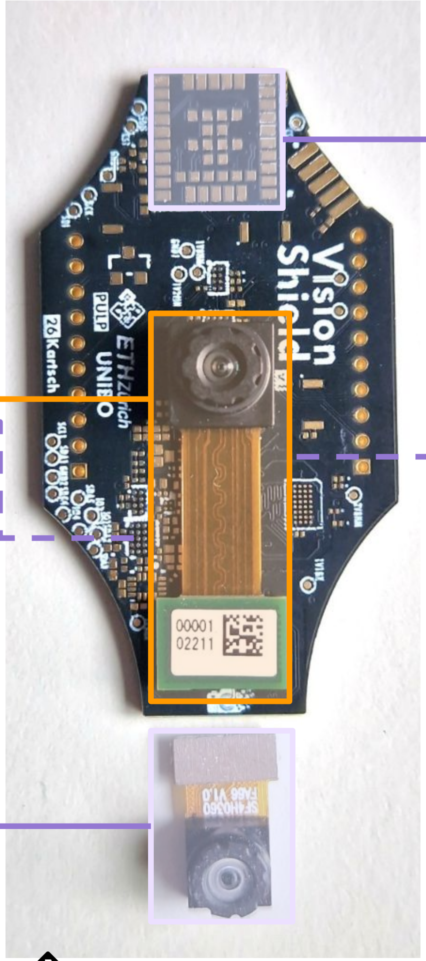
Vision-Shield

Bitcraze's extension PCB integrating the GENX320 event-camera

Prophesee Genx320 event-camera
- Resolution: 320x320
- MIPI-CSI2, CPI

Off-chip memories
- 32 MB PSRAM
- 64 MB Flash

Himax HM0360 camera
- Resolution: VGA
- RGB / Grayscale
- FPS: up to 60Hz @ VGA.



Designed at IIS - ETH Zürich

NINA-W102 Wi-Fi module
- Bandwidth up to 10 Mbit/s
- Power ~300mW

GAP9 SoC

Fabric controller (FC): 370 MHz

Cluster (CL): 370 MHz

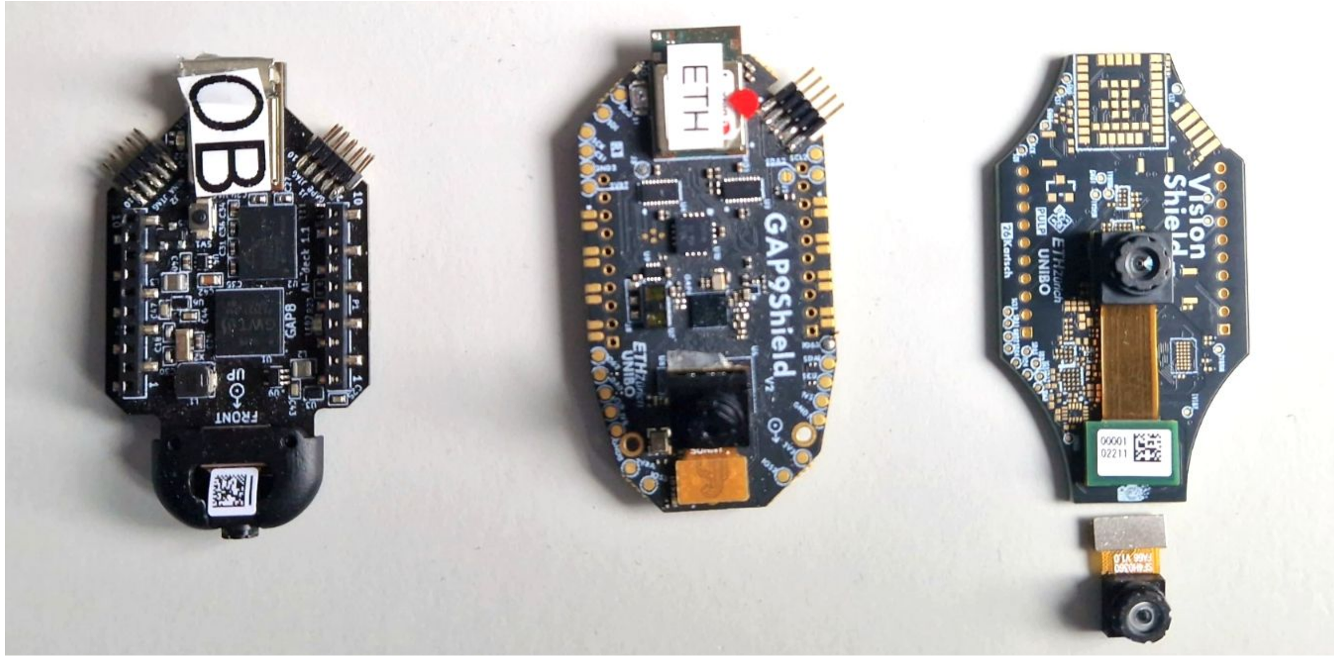
Hardware acceleration: NE16

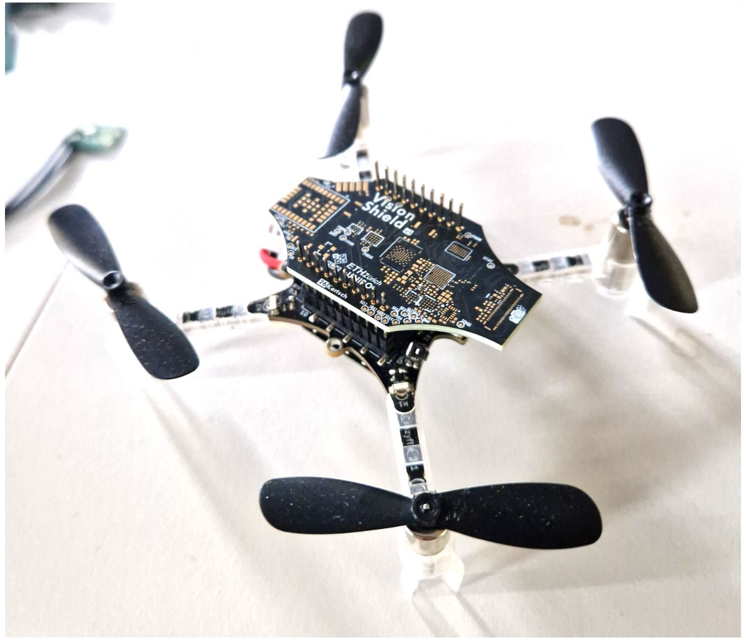
On-chip memories:
- 1536 kB L2
- 128 kB L1

Power:
- FC up to 14.8 mW
- CL up to 48.1 mW
- NE16 up to 48.1 mW

Vision-Shield vs. past generations

AI-Deck | **GAP9-Shield** | **Vision-Shield** | **Crazyflie 2.1 w/ Vision-shield**





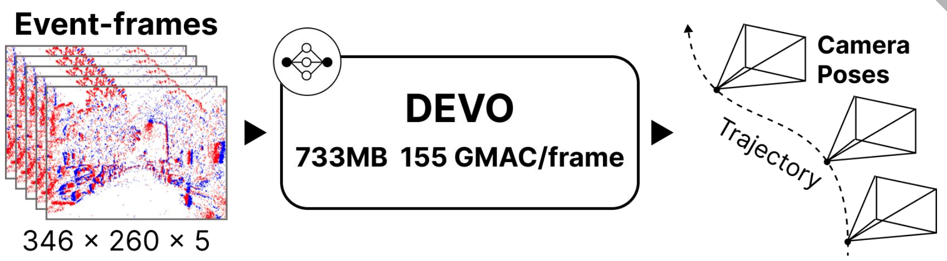
AI-Deck: *GAP8*, *Himax HM01B0*

GAP9-Shield: *GAP9*, *OV5647*

Vision-Shield: *GAP9*, *Himax HM0360*, *GENx320*

Challenge and contribution

The State-of-the-Art (SoA) event-based VO algorithm, DEVO [1] requires 733MB of memory, runs on RTX4070 (>250w).



Ultra-low-power (ULP) MCUs constraints		
<10cm Small form factor	<100mW Limited computing power budget	few MB Limited on-chip memory

Requirements	
Accurate perception	Real-time operation

SoA event-based VO algorithms are incompatible with MCUs.

Our contributions

1. Event-based VO algorithm tailored to MCUs: TinyDEVO
2. Energy-efficient implementation on an ULP multi-core RISC-V MCU.
3. Trade-off analysis between execution performance and VO's accuracy

ETH zürich — Alma Mater Studiorum Università di Bologna — idsia Dalle Molle Institute for Artificial Intelligence USI-SUPSI

[1] S. Klenk et al., "Deep Event Visual Odometry," 3DV, 2024