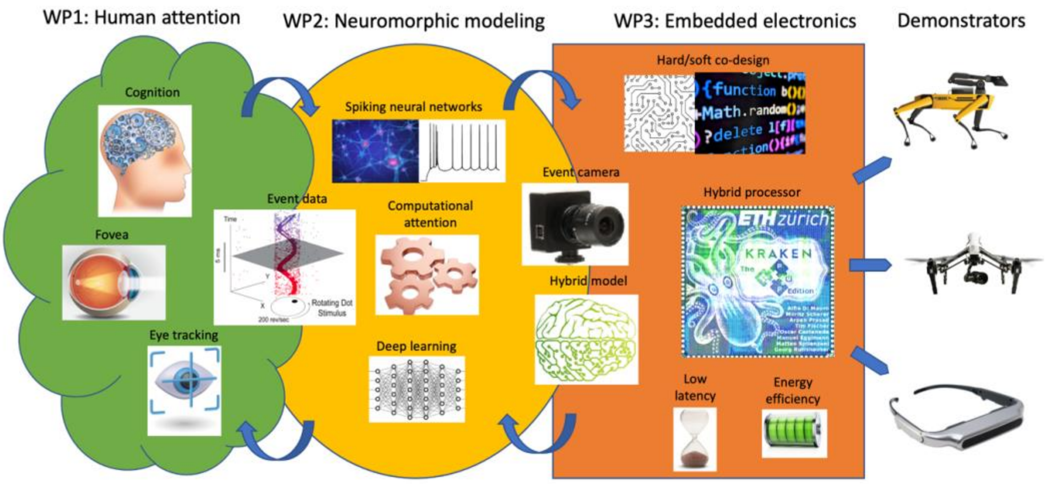
Attention models for event data

MSc research internship proposal – Mar-Aug 2024 (6 months)
+ PhD opportunity – Oct 2024-Sept 2027 (36 months)



Context

The field of embedded computer vision has become increasingly important in recent years as the demand for low-latency and energy-efficient vision systems has grown [Thiruvathukal et al. 2022]. One of the key challenges in this area is developing intelligent vision systems that can efficiently process large amounts of visual data while still maintaining high accuracy and reliability. One important characteristic of today's computer vision systems is that most of them are frame-based. Conventional video sensors record the entire image with a given rate and resolution [Hopkins et. al 2018]. The original rationale for sensing a scene this way is that the transmission or recording is intended to be viewed by a human observer who may be looking closely at any part of the moving image. Frame-based video contains a huge amount of redundant data and requires enormous computational power to process. As stated in [Hopkins et. al 2018], biological vision sensors, however, are quite different from frame-based cameras. They do not sample images at a uniform rate, nor at a uniform resolution. The human eye has a small high-resolution region (the fovea) in the center of the field of vision, and a much larger peripheral vision, which has much lower resolution, combined with an increased sensitivity to movement. Therefore, limited resources are deployed to extract the most salient information from the scene without wasting energy capturing the entire scene at the highest resolution.

This efficient process has inspired original deep learning models such as [Lukanov et al. 2021], as well as the recent development of a variable-resolution event sensor [Serrano-Gotarredona et al., 2022]. Event-based sensors mimic the biological retina: they asynchronously measure per-pixel brightness changes and output a stream of events that encode the time, location, and sign of the brightness changes (positive or negative). In addition to eliminating redundancy, they benefit from several advantages over conventional frame cameras, from which they fundamentally differ. Furthermore, the human eye is primarily sensitive to changes in the luminance falling on its individual sensors. These changes are processed by layers of neurons in the retina through to the retinal ganglion cells that generate action potentials, or *spikes*, whenever a significant change is detected. Then these spikes propagate through the optic nerve to the brain. This approach focuses the resources on specific areas of the image that convey the most useful information. Given the core objective of computer vision systems, it seems natural to sense the world with bio-inspired sensors.

This internship work will take place in the context of the ANR-funded project NAMED (Neuromorphic Attention Models for Event Data) that will start on February 1st, 2024.

Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis
I3S - UMR7271 - UNS CNRS
2000 route des Lucioles - Les Algorithmes - bât. Euclide B - 06900 Sophia Antipolis - France
Tél +33 (0)4 89 15 43 00 - Fax +33 (0)4 89 15 43 01 - dir@i3s.unice.fr - www.i3s.unice.fr

Adresse postale : Laboratoire I3S - CS 40121 - 06903 Sophia Antipolis CEDEX

Scientific objectives

The objective of this internship research project is to design and implement attention models adapted to event data. The first step will consist in studying state-of-the-art attentional mechanisms in deep networks and their link with cognitive attention as implemented in the brain. Cognitive attention refers to the selective processing of sensory information by the brain based on its relevance and importance to the current task or goal. It involves the ability to focus one's attention on specific aspects of the environment while filtering out irrelevant or distracting information. In particular, the study will distinguish between both top-down and bottom-up attention. The second step will be the design an attention architecture that will allow selectively focusing on relevant regions while ignoring irrelevant part, which will depend on the target task (e.g., segmentation, object tracking, obstacle avoidance, etc.). The model will be based either on standard deep networks, or on spiking neural networks, based on previous work developed in the team [GIT]. Spiking Neural Networks are a special class of artificial neural networks, where neurons communicate by sequences of asynchronous spikes. Therefore, they are a natural match for event-based cameras due to their asynchronous operation principle. This selection of regions will result in less data usage and smaller models (frugal system).

In the third step, we will evaluate the impact of the attention mechanism on the general performance of the system. The target metrics will obviously depend on the selected task, and will include accuracy, MIOU, complexity, training time, inference time, etc. of the network.

Job information

Location

Université Côté d'Azur, Sophia Antipolis (Nice area) France

Type of contract

Internship, duration 4-6 months

Job status

Full-time

Candidate profile

Master 2 in Computer Science (Machine Learning, Computer Vision, AI) or Mathematics or Computational Neuroscience. Programming skills in Python/C++, interest in research, machine learning, bio-inspiration and neurosciences are required.

Salary

Standard French internship allowance ("gratification") by CNRS (around 600 € per month)

PhD opportunity

Yes

Offer starting date

Between February 1 and April 15, 2024

Application deadline

The earlier the better, no later than Jan 31, 2024

How to apply

Send a resume (CV), transcript of grades, and motivation letter to the contact given below

Contact

Prof. Jean Martinet
Université Côte d'Azur / I3S / CNRS | Polytech Nice Sophia, Campus SophiaTech, Sophia-Antipolis
jean.martinet@univ-cotedazur.fr, +334-89-15-43-86

References

[GIT] https://github.com/ameliegruel/FoveationStakes_DVS

[Bulzomi 2023] Hugo Bulzomi, Amélie Gruel, Jean Martinet, Takeshi Fujita, Yuta Nakano, et al.. **Object Detection for Embedded Systems Using Tiny Spiking Neural Networks: Filtering Noise Through Visual Attention**. International Conference on Machine Vision and Applications (MVA), Jul 2023, Hamamatsu, Japan.

[Gruel et al., 2023] Amélie Gruel, Dalia Hareb, Antoine Grimaldi, Jean Martinet, Laurent Perrinet, Bernabé Linares-Barranco and Teresa Serrano-Gotarredona. **Stakes of Neuromorphic Foveation: a promising future for embedded event cameras**. Biological Cybernetics Special Issue: What can Computer Vision learn from Visual Neuroscience? 2023.

[Gruel et al., 2023] Amélie Gruel, Alfio di Mauro, Robin Hunziker, Luca Benini, Jean Martinet, Michele Magno. **Embedded neuromorphic attention model leveraging a novel low-power heterogeneous platform**. IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS 2023), Hangzhou, China. June 2023.

[Gruel et al., 2023] Amélie Gruel, Jean Martinet, Michele Magno. **Simultaneous neuromorphic selection of multiple salient objects for event vision**. IJCNN 2023, Queensland, Australia. June 2023.

[Gruel et al., 2022, 1] Amélie Gruel, Dalia Hareb, Jean Martinet, Bernabé Linares-Barranco, Teresa Serrano-Gotarredona. **Neuromorphic foveation applied to semantic segmentation**. CVPR 2022 workshop "NeuroVision: What can computer vision learn from visual neuroscience?", New Orleans, Louisiana, June 2022. Best poster award.

[Gruel et al., 2022, 2] Amélie Gruel, Antonio Vitale, Jean Martinet, Michele Magno: Neuromorphic Event-Based Spatio-temporal Attention using Adaptive Mechanisms. AICAS 2022: 379-382.

[Gruel and Martinet, 2021] Amélie Gruel, Jean Martinet: Bio-inspired visual attention for silicon retinas based on spiking neural networks applied to pattern classification. CBMI 2021: 1-6

[Hopkins et al. 2018] Hopkins M, Pineda-Garcıa G, Bogdan PA, Furber SB. 2018 Spiking neural networks for computer vision. Interface Focus8: 20180007.

[Iacono et al. 2019] M. Iacono, G. D'Angelo, A. Glover, V. Tikhanoff, E. Niebur and C. Bartolozzi. Proto-object based saliency for event-driven cameras, IEEE/RSJ IROS,doi: 10.1109/IROS40897.2019.8967943.

[Itti et al. 1998] Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on PAMI, 20(11), 1254-1259.

[Lukanov et al. 2021] Hristofor Lukanov, Peter König, and Gordon Pipa, Biologically Inspired Deep Learning Model for Efficient Foveal-Peripheral Vision. Front. Comput. Neurosci. (15) 2021.

[Serrano-Gotarredona et al. 2022] T. Serrano-Gotarredona, F. Faramarzi and B. Linares-Barranco: Electronically Foveated Dynamic Vision Sensor. IEEE International Conference on Omni-Layer Intelligent Systems COINS 2022.

[Thiruvathukal et al. 2022] Thiruvathukal, George K., and Yung-Hsiang Lu. "Efficient computer vision for embedded systems." Computer 55.4 (2022): 15-19.