

Early detection modeling in Spiking Neural Networks for embedded stream-based image processing

Benoît Miramond, benoit.miramond@univ-cotedazur.fr¹

¹Université Côte d'Azur / LEAT, Sophia-Antipolis, France

The laboratory:

Benoît Miramond is Full Professor in Electrical Engineering at **LEAT** laboratory from Université Côte d'Azur (UCA). He leads the eBRAIN research group and develops an interdisciplinary research activity on embedded bio-inspired artificial intelligence and neuromorphic architectures, with a strong focus on spiking neural networks (SNNs).

1 Abstract

The increasing availability of **spatio-temporal data** has enabled significant advances in perception and pattern recognition systems. However, in many real-world and latency-critical applications, performance cannot be assessed solely through final recognition accuracy. Instead, the value of a prediction strongly depends on its **timeliness**, as reliable decisions must often be produced from **partially observed data**.

Event-based sensing and neuromorphic computing offer promising paradigms to address this challenge by providing asynchronous, high-temporal-resolution signals that naturally support **early detection** Gallego et al. (2020); Abderrahmane et al. (2020). At the same time, attention-based models have demonstrated strong capabilities for capturing temporal dependencies, but their dense and synchronous processing limits their suitability for embedded and low-latency scenarios Vaswani et al. (2017); Tay et al. (2020).

This work investigates **bio-inspired** approaches for early decision-making from spatio-temporal data, with a particular focus on **spiking neural networks (SNNs)** Gerstner et al. (2014). The objective is to analyze how event-driven sensing, spike-based computation, and attention-inspired mechanisms can be combined to achieve accurate and timely predictions while explicitly considering **hardware constraints**, including latency, memory usage, and energy consumption.

2 Context

Spatio-temporal perception plays a central role in many intelligent systems, including action recognition, visual understanding, and autonomous decision-making. In a wide range of real-world applications, decisions must be taken under strict temporal constraints and often from **incomplete or evolving observations**. In such settings, the relevance of a prediction depends not only on its correctness, but also on **when** it is produced, making **early detection** a key requirement rather than a secondary performance criterion.

Recent advances in deep learning have considerably improved the modeling of temporal dependencies. In particular, Transformer-based architectures relying on **attention mechanisms** have demonstrated strong performance in spatio-temporal perception tasks by integrating information over long temporal horizons Vaswani et al. (2017). However, the dense and synchronous nature of these models, combined with the quadratic complexity of self-attention, limits their applicability in real-time and embedded scenarios subject to tight **latency** and **energy** constraints Tay et al. (2020).

Event-based vision sensors offer an alternative sensing paradigm that naturally addresses some of these limitations. By encoding only changes in the visual scene, they produce sparse and asynchronous **spatio-temporal data** streams with very low latency Gallego et al. (2020). **Spiking neural networks (SNNs)**, which process information through spike-based communication, naturally align with this sensing modality and provide intrinsic temporal dynamics suitable for stream-based processing. These characteristics make SNNs promising candidates for early decision-making under constrained **hardware resources**.

Within this context, the e-BRAIN research group at the LEAT laboratory has developed a strong expertise in neuromorphic vision and algorithm–hardware co-design. Prior work includes the design of dedicated neuromorphic architectures for efficient SNN execution Abderrahmane et al. (2022), as well as analytical tools to characterize the effects of temporal spike-based encoding and quantization for static data Castagnetti et al. (2023b). Together, these contributions provide a solid methodological and technological foundation to investigate early decision-making mechanisms on **dynamic spatio-temporal data**, bridging perception performance, temporal behavior, and hardware efficiency.

3 Goals of the doctoral project

The main objective of this PhD project is to explore and analyze **bio-inspired** neural architectures for **early detection** from **spatio-temporal data** under realistic sensing and computational constraints. Rather than focusing exclusively on maximizing final recognition accuracy, the project aims to understand how reliable decisions can be produced as early as possible from **partially observed data**, and how prediction accuracy, latency, and confidence evolve over time as additional sensory information becomes available.

A first research direction focuses on the study of **spiking neural networks (SNNs)** for continuous spatio-temporal processing. Different neuron models, temporal coding strategies, and learning mechanisms will be explored, with particular attention to their impact on early decision-making and temporal dynamics Bellec et al. (2018); Fang et al. (2021). The objective is to characterize how spike-based representations support early inference under strict latency constraints.

A second research direction investigates the role of **attention mechanisms** in event-driven and spike-based models. Transformer architectures will serve as a reference framework to analyze the benefits and limitations of attention for early perception Vaswani et al. (2017); Tay et al. (2020). In this context, the project will explore attention-inspired mechanisms compatible with causal, spike-based computation, and assess their suitability for early decision-making.

A third research direction examines different strategies for building spiking models, with a particular focus on the trade-offs between **direct training of SNNs** using surrogate gradients and quantization-aware **ANN-to-SNN conversion** approaches. The project aims to compare these paradigms in terms of accuracy, temporal resolution, scalability, and suitability for early detection, especially when applied to spatio-temporal and event-based data.

Finally, the project explicitly addresses the interplay between **accuracy**, **decision latency**, and **hardware resources**. By building upon analytical energy models and neuromorphic execution platforms developed within the e-BRAIN group, the work will investigate how temporal quantization, neuron dynamics, and architectural constraints influence early decision-making under realistic energy and memory budgets.

The proposed approaches will be evaluated on a range of spatio-temporal datasets acquired from both conventional RGB cameras and **event-based vision sensors**, potentially involving multi-camera configurations. This evaluation will enable a systematic comparison of modeling choices for **early detection** in embedded and latency-critical perception systems.

4 State of the Art

Perception systems are traditionally designed and evaluated under the assumption that recognition accuracy is the primary performance criterion, independently of the time required to reach a

decision. In this paradigm, a prediction is considered valuable as long as it is correct, regardless of when it is produced. However, in many real-world perception and decision-making scenarios, the usefulness of a prediction is inherently time-dependent: a correct decision made too late may be functionally irrelevant or unsafe. This temporal dimension becomes particularly critical in settings involving continuously evolving observations and strict latency constraints, such as real-time perception or control systems [Gehrig and Scaramuzza \(2024\)](#). Beyond safety-critical scenarios, the importance of **early detection** has also been demonstrated in tasks such as **human behavior and action recognition**, where systems must operate on **partially observed spatio-temporal data** [Shao et al. \(2024\)](#). Despite this, most existing approaches remain primarily optimized and benchmarked with respect to their **final accuracy**, with limited analysis of how predictions evolve over time and how temporal performance is evaluated in practice [Yik et al. \(2025\)](#).

Event-based vision sensors provide access to visual information at very fine temporal resolutions through their asynchronous operation, enabling perception systems to react as soon as relevant events occur [Gallego et al. \(2020\)](#). This property makes them particularly well suited for **early perception** and has enabled early recognition in highly dynamic scenarios [Deniz et al. \(2023\)](#); [Rasamuel et al. \(2019\)](#). However, as emphasized in neuromorphic computing studies, the potential benefits in terms of latency and energy efficiency depend critically on how temporal information is exploited throughout the processing pipeline [Abderrahmane et al. \(2020\)](#). In practice, event streams are often processed through temporally aggregated representations prior to learning, which limits the exploitation of fine-grained temporal cues and shifts decision-making toward later stages of the observation.

To capture temporal context and long-range dependencies, **attention mechanisms** and **Transformer-based architectures** have been increasingly adopted for spatio-temporal perception tasks [Vaswani et al. \(2017\)](#). While these models achieve strong representational performance, they rely on dense and synchronous processing and exhibit **quadratic complexity** with respect to sequence length, posing major challenges for low-latency and embedded deployment under strict **hardware constraints**, particularly in terms of memory bandwidth and energy consumption [Tay et al. \(2020\)](#). Recent efforts have sought to alleviate these limitations through more efficient attention formulations, including spiking Transformer architectures that exploit temporal sparsity, such as ESTSformer [Lu et al. \(2025\)](#).

Moreover, in [Lee et al. \(2025\)](#) frequency domain analysis reveals why certain transformers designs work well for event-based data, which contains valuable high-frequency information but is also sparse and noisy. SpikePool transformers architecture is proposed to preserve meaningful high-frequency content while capturing critical features and suppressing noise, achieving a better balance for event-based data processing.

However, the energy evaluation in such works is typically based on accumulation or synaptic operation counts. The cost of **memory accesses**, which is known to dominate energy consumption in practical hardware implementations, is generally not explicitly considered [Lemaire et al. \(2022\)](#).

In contrast, **spiking neural networks (SNNs)** offer a **bio-inspired** and event-driven computational paradigm that naturally aligns with asynchronous sensing. By exploiting sparse spike-based communication and intrinsic temporal dynamics, SNNs enable **low-latency** and potentially **energy-efficient** processing [Gerstner et al. \(2014\)](#); [Yin et al. \(2021\)](#). Recent work has shown that SNNs can support **early decision-making** from event streams, as illustrated by the EEvAct framework for early event-based action recognition [Neumeier et al. \(2025\)](#).

A major line of research seeks to bridge the performance gap between ANNs and SNNs through **ANN-to-SNN conversion**. While early methods achieved near-lossless conversion for convolutional networks [Diehl et al. \(2015\)](#), recent studies have extended this paradigm to **Transformer architectures**, addressing non-linear components such as LayerNorm, GELU, and Softmax [Huang et al. \(2024\)](#); [Wang et al. \(2025\)](#). These approaches demonstrate that **spiking Transformers** can achieve high accuracy with a reduced number of timesteps. However, they introduce an implicit form of **temporal quantization**, controlled by the number of timesteps, whose impact on accuracy, latency, and energy is rarely analyzed explicitly, particularly for **dynamic and event-based**

data. Alternatively, direct training of SNNs using surrogate gradients enables full exploitation of spike-based temporal dynamics but incurs training and inference costs that scale linearly with time, limiting scalability for large architectures such as Transformers [Shi et al. \(2024\)](#).

This work is conducted within the **e-BRAIN research group at the LEAT laboratory** (UMR 7248, Université Côte d'Azur-CNRS) and builds directly upon established research directions of the team. In particular, the **analytical energy estimation model** proposed by [Lemaire et al. \(2022\)](#) will be used as a reference framework and adapted to the specific scenarios considered in this thesis, enabling hardware-realistic energy evaluation beyond proxy metrics. Similarly, the neuromorphic architecture **SPLEAT** [Abderrahmane \(2022\)](#), together with the QUALIA programming framework [Novac et al. \(2021\)](#), will serve as a concrete execution substrate to assess the behavior of spiking models under realistic hardware constraints.

Regarding **temporal quantization**, this thesis builds upon prior work which analyzed and controlled quantization effects for static data [Castagnetti et al. \(2023b,a\)](#). The contribution of this work is to investigate how these quantization effects manifest in **dynamic and spatio-temporal settings**, such as video sequences or event-based streams, and how they interact with early decision mechanisms.

5 Organization of the PhD

The PhD project is planned over a three-year period and is structured as follows:

- **Year 1:** in-depth literature review on early detection, spatio-temporal perception, event-based vision, and spiking neural networks; definition of evaluation protocols; implementation of representative baseline models using both conventional and event-driven data; preliminary experiments to analyze temporal decision behavior.
- **Year 2:** investigation of modeling strategies for early perception, including spiking architectures and attention-inspired mechanisms; comparison of different learning paradigms and temporal coding strategies; systematic analysis of trade-offs between accuracy, decision latency, and computational cost.
- **Year 3:** consolidation and optimization of the most promising approaches; extensive experimental validation on representative spatio-temporal perception tasks; evaluation under realistic hardware constraints, with particular attention to latency and energy efficiency.
- Dissemination of results through publications and presentations throughout the PhD.

6 Skills

The candidate should hold a Master's degree in electronic engineering, embedded systems, artificial intelligence, signal or image processing, neuromorphic engineering, or a closely related field.

A solid background in machine learning is expected, with interest or experience in spiking neural networks, temporal modeling, or bio-inspired computation considered an asset. Familiarity with embedded or hardware-aware constraints (latency, energy, real-time processing) is appreciated.

The candidate should demonstrate strong motivation for interdisciplinary research at the interface between algorithms and hardware, as well as the ability to work both independently and within a research team.

Proficiency in Python programming (e.g., PyTorch or equivalent frameworks) is required. Experience with embedded platforms, FPGA, or neuromorphic hardware is a plus. Fluent scientific communication in English is expected.

References

- A. Abderrahmane. *Event-Based Neuromorphic Architecture for Low-Power Spiking Neural Networks*. PhD thesis, Université Côte d'Azur, 2022.
- A. Abderrahmane et al. Event-based neuromorphic computing: A survey. *Proceedings of the IEEE*, 108(8):1363–1386, 2020.
- N. Abderrahmane, B. Miramond, E. Kervennic, and A. Girard. SPLEAT: SPiking Low-power Event-based ArchiTecture for in-orbit processing of satellite imagery. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10, July 2022. doi: 10.1109/IJCNN55064.2022.9892277. URL <https://ieeexplore.ieee.org/document/9892277?arnumber=9892277>.
- G. Bellec, D. Salaj, A. Subramoney, R. Legenstein, and W. Maass. Long short-term memory and learning-to-learn in networks of spiking neurons. *Advances in Neural Information Processing Systems*, 31, 2018.
- A. Castagnetti, A. Pegatoquet, and B. Miramond. Neural information coding for efficient spike-based image denoising. Technical Report arXiv:2305.11898, arXiv, May 2023a. URL <http://arxiv.org/abs/2305.11898>. arXiv:2305.11898 [cs] type: article.
- A. Castagnetti, A. Pegatoquet, and B. Miramond. Trainable quantization for Speedy Spiking Neural Networks. *Frontiers in Neuroscience*, 17, 2023b. ISSN 1662-453X. URL <https://www.frontiersin.org/articles/10.3389/fnins.2023.1154241>.
- D. Deniz, C. Fermüller, et al. Event-based vision for early prediction of manipulation actions. *arXiv preprint arXiv:2307.14332*, 2023.
- P. U. Diehl et al. Fast-classifying, high-accuracy spiking deep networks. *IJCNN*, 2015.
- W. Fang, Z. Yu, Y. Chen, T. Masquelier, T. Huang, and Y. Tian. Deep spiking neural networks with spike-based backpropagation. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5): 2090–2103, 2021.
- G. Gallego, T. Delbrück, G. Orchard, et al. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2020.
- D. Gehrig and D. Scaramuzza. Low-latency automotive vision with event cameras. *Nature*, 2024.
- W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge University Press, 2014.
- Z. Huang et al. Towards high-performance spiking transformers from ann to snn conversion. *ACM Transactions on Neural Networks and Learning Systems*, 2024.
- D. Lee, A. Sima, Y. Li, P. Stinis, and P. Panda. SpikePool: Event-driven Spiking Transformer with Pooling Attention, Oct. 2025. URL <http://arxiv.org/abs/2510.12102>. arXiv:2510.12102 [cs].
- A. Lemaire, B. Miramond, et al. An analytical estimation of spiking neural network energy consumption. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- C. Lu, H. Du, W. Wei, et al. Estformer: Efficient spatio-temporal spiking transformer. *Neural Networks*, 191:107786, 2025.
- M. Neumeier et al. Eevact: Early event-based action recognition with high-rate two-stream spiking neural networks. *arXiv preprint arXiv:2507.07734*, 2025.
- P.-E. Novac, G. Boukli Hacene, A. Pegatoquet, B. Miramond, and V. Gripon. Quantization and deployment of deep neural networks on microcontrollers. *Sensors*, 21(9), 2021. ISSN 1424-8220. doi: 10.3390/s21092984. URL <https://www.mdpi.com/1424-8220/21/9/2984>.

- M. Rasamuel, L. Khacef, L. Rodriguez, and B. Miramond. Specialized visual sensor coupled to a dynamic neural field for embedded attentional process. In *2019 IEEE Sensors Applications Symposium (SAS)*, pages 1–6, 2019. doi: 10.1109/SAS.2019.8705979.
- Y. Shao et al. Spatio-temporal early prediction based on multi-objective reinforcement learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- X. Shi et al. Efficient training of spiking transformers via spike firing approximation. *arXiv preprint arXiv:2411.16061*, 2024.
- Y. Tay, M. Deghani, D. Bahri, and D. Metzler. Efficient transformers: A survey. *ACM Computing Surveys*, 55(6):1–28, 2020.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.
- J. Wang et al. Training-free ann-to-snn conversion for high-performance spiking transformers. *AAAI*, 2025.
- J. Yik et al. Neurobench: A framework for benchmarking neuromorphic computing algorithms and systems. *arXiv preprint arXiv:2304.04640*, 2025.
- B. Yin, F. Corradi, and S. M. Bohté. A survey of spiking neural network architectures and their applications. *Neurocomputing*, 426:189–203, 2021.