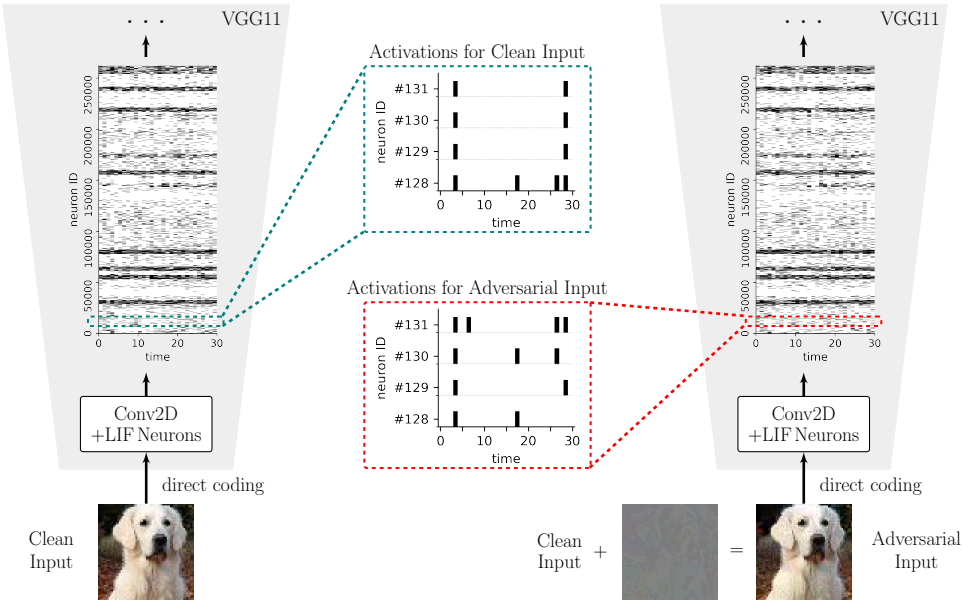
Security Analysis of Spiking Neural Networks



Spiking neural networks (SNNs) provide an energy-efficient alternative to a variety of artificial neural network (ANN) based AI applications. As the progress in neuromorphic computing with SNNs expands their use in applications, the problem of adversarial robustness of SNNs becomes more pronounced. Recent work [1] proposed a rigorous adversarial evaluation strategy to empirically emphasize the difficulty in achieving robust SNNs via spike-based backpropagation using surrogate gradients to approximate the discontinuous derivative of the spike function. In this project, we will build on this framework to construct a thorough benchmark for assessing SNN robustness.

[1] O. Özdenizci, R. Legenstein, "Adversarially robust spiking neural networks through conversion", Transactions on Machine Learning Research (TMLR), 2024.

Goals & Tasks

- Review of the state-of-the-art on adversarially robust SNNs.
- Simulating and benchmarking with an existing ensemble SNN attack framework [1].
- Extending this benchmark with stronger adversarial attack evaluations for SNNs.

Qualifications

- Interest in machine learning security.
- Experience with the Python based deep learning framework PyTorch.
- Registered to one of the following:
  - ✓ **Bachelor Thesis**
  - ✓ **Seminar Project**
  - ☐ **Master Thesis**

Contact

Ozan Özdenizci: oezdenizci@tugraz.at