Master 2 internship proposal

ANITI CertifAI - Towards the certification of ML-based systems

J. Guérin, K. Delmas & J. Guiochet

Internship data

Title	Runtime Monitoring of Deep Neural Networks
Keywords	Machine Learning Safety, Fault Tolerance, Computer Vision
Technologies	Programming, Python, Pytorch, Scikit-learn, Git
Location	LAAS-CNRS, Toulouse, France
Level	Master 2 (MSc)
Duration/Dates	4-6 months
Student profile	Computer Science, Artificial Intelligence, Mathematics/Statistics
Salary	~600 euros/month
Supervisors	Joris Guérin (Espace-DEV), Kévin Delmas (ONERA), Jérémie Guiochet (LAAS-CNRS)

Internship description

With the recent progress in machine learning (ML) research, deep neural network (DNN) architectures are now used to address safety-critical tasks, such as self-driving cars, surgical robots, and drones landing. Online fault tolerance approaches, or runtime monitors, are a promising research direction to improve the safety of such systems. A runtime monitor is a component aiming to identify and reject unsafe data encountered at inference time.

In the past two years, our team has conducted many research about runtime monitoring of DNNs. Our recent efforts include the introduction of new evaluation criteria, the release of a new benchmarking library, and the development of new statistical models for efficient monitoring of DNN internal layers activations. In particular, we showed that different layers of the DNN represent data differently and require specific statistical modeling for efficient monitoring.

In this context, the intern is expected to participate to this ongoing effort by contributing to (at least) one of the following research directions:

- Propose and test new safety monitoring approaches for DNNs.
- Propose and test new ways to optimize the monitoring approach for a specific DNN layer.
- Propose and test new ideas for combining results of an ensemble of monitoring approaches, applied to different layers of the monitored DNN.

The intern will also contribute to maintain and extend the benchmarking library. Possible extensions can consist in adding new ML tasks, new datasets, new DNN architectures, and new recent monitoring approaches from the literature.

Application process

To apply, please send your **CV**, **cover letter**} and **transcripts** (marks) of the last 2 years to joris.guerin@ird.fr, kevin.delmas@onera.fr and jeremie.guiochet@laas.fr