DEEL France-Quebec

COOPÉRATION FRANCE-QUÉBEC

NEWSLETTER #10

April 2021

TOULOUSE & QUEBEC JOIN FORCES TO DEVELOP AI FOR CRITICAL SYSTEMS

A NEW DYNAMISM IN IRT SAINT EXUPERY CANADA

The IRT Saint Exupéry Canada is based in Montreal within the MILA, a district with a high concentration of companies dedicated to artificial intelligence.

Since last February, **Igor CALDERAN** has taken the lead of the IRT Saint-Exupéry Canada by assuming the position of General Manager of the branch. With an engineering background and an MBA, Igor has spent the last 12 years at Bombardier Aerospace managing innovation, strategic projects and the technical, commercial and legal aspects of aerospace technology development.

Strengthened with a new dynamism, the team is now composed of 4 other people dedicated to support you in your artificial intelligence projects. Patrick Saint-Louis, Ph.D., for operations research, Damien Grasset, M.Sc., for reinforcement learning, Julien Caudroit, for communication, and Pierre Dupuy, for administration, complete Igor's team.

This team aims to build links with the Montreal AI ecosystem, the DEEL project partners, and will develop new initiatives in the coming months in different sectors.

Don't hesitate to contact the IRT Saint Exupéry Canada team, to explore new opportunities or if you have AI related projects.

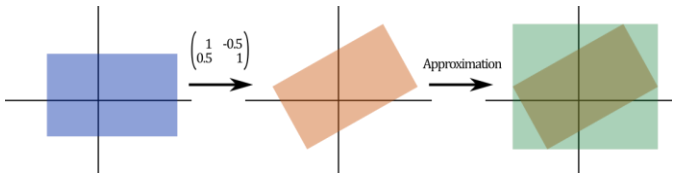
#Igor CALDERAN, Julien CAUDROIT, Pierre DUPUIS, Damien GRASSET, Patrick SAINT-LOUIS,

FINDING THE PLACE OF FORMAL METHODS IN AI

The effectiveness of formal methods, especially in critical systems, no longer needs to be proven, but how well do they apply to AI-based systems?

A possibility is to prove properties about the neural networks, such as their robustness. Indeed, being able to demonstrate the capacity of a network to defend against adversarial attacks could be valuable, as it shows that the model is able to perform well even with small perturbations on its input data.

Some of our work in that domain focuses on methods that build estimations of the transformations performed in a neural network. These methods, leveraging mature techniques like abstract interpretation, give mathematical proofs that these estimations are sound, which in turn provides guarantees about specific properties of the model.



Over-approximation using the Box abstract domain when applying a linear

However, this soundness comes at the cost of a loss of precision in the analysis due to over-approximations, as illustrated in the figure. The characterization of this precision loss constitutes an interesting and promising set of metrics to compare the effectiveness of each method in the domain.

This observation motivated one of our areas of work, where we are implementing various techniques, typically from the reachability analysis domain, but we are also designing a benchmark for the evaluation of the precision of current and hopefully future techniques.

 # (Mikaël CAPELLE), Vincent MUSSOT

FOCUS ON A PhD STUDENT, Thomas FEL

In recent years, Artificial Intelligence (AI) has experienced a notable growth. Properly used, this technology may offer great opportunities in many sectors. However, the origin of the results and the reasoning behind them are often opaque. In order to give confidence in such systems, the field faces the barrier of explicability. This barrier is particularly blocking, especially for functions requiring a high level of operational security such as autonomous trains.

Recently, many strategies have been proposed to help users visualize which features in the data allow models to reach their decisions. Some works [1] have highlighted that the results of some methods are problematic: the explanation is conditioned only by the input data and not by the model used.

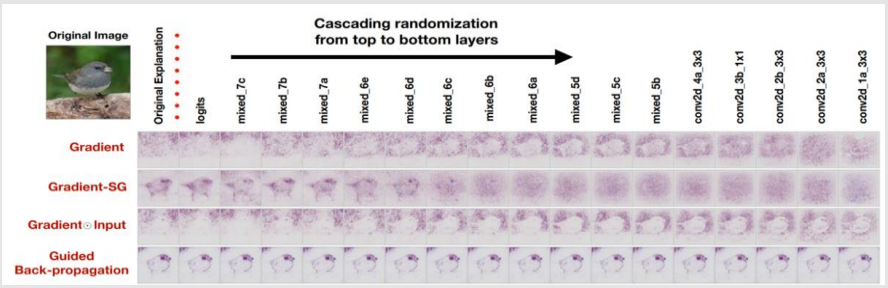


Figure 1. Explicability method results showing the decisive areas to model decision [1]. Random noise is progressively added to each layer and the explanation is computed. We note that even with a completely random model (last column "conv2d_1a_3x3"), the Guided Back-propagation explanation remains unchanged.

This reveals an important bias: although **some methods offer explanations that are satisfactory for users, they do not reflect the real functioning of the model.** The explanation that should give confidence becomes itself questionable. In parallel, several works have recently demonstrated **the interest of using human reasoning to guide the logic of machine learning models** [2].

These observations have led to a thesis jointly supervised by DEEL and the SNCF that aims to establish human-centered metrics for evaluating the explanations of a neural network model.

[1] Sanity check for Saliency map, NeurIPS 2018, [2] Learning what and where to attend, ICLR 2019

 # Thomas FEL, Laurent GARDES, Claire NICODEMME, Thomas SERRE, David VIGOUROUX

THE DEEL PROJET ON THE QUEBEC SIDE

Officially launched in September 2020, t**he DEEL Quebec project** is led by the CRIAQ and the IID of ULaval, with the collaboration of IVADO. This project involves **4 industrial partners** (Thales, CAE, Bombardier and Bell Helicopter) **and 5 universities**. The initiative addresses one of the key issues related to AI for aeronautics: certification, and is organized in 4 research axes:

Robustness: The ability of a system to operate outside of its usual conditions while maintaining a pre-determined level of performance.
Interpretability: The acceptance and certification of the learned model by the different actors involved depends on our understanding and confidence in these models.
Privacy by design: Ensuring data confidentiality is crucial when learning is to be done at a third party and to allow collaborative learning.
Certifiability: Certification ensures that the program will operate in the intended environment.

 # Giulio ANTONIOL, Fanny EYBOULET, François LAVIOLETTE, Yann PEQUINGOT, Lynda ROBITAILLE, Marion STOFFEL

KEY DATES & INFORMATIONS

Certification Mission	Next workshops : May, 5th & 6th – May, 26th & 27th
« Les Carrefours DEEL »	Next Carrefours DEEL: June, 3rd
MobiliT.AI 2021	Interactive and dynamic format : May, 10th to 12th → Registration
DEEL Discussion Group	Every two months, the next one → "Gabriel Laberge -Uncertainty in post-hoc explanations, and how it can be used to train diverse ensembles"
Future Intelligence	3-day event, June 1st to 3rd → More information