DEEL
DEpendable & Explainable Learning

DEEL MAG #04

TOULOUSE & QUEBEC JOIN FORCES TO DEVELOP AI FOR CRITICAL SYSTEMS

IRT SAINT EXUPERY CANADA

In March, **IRT Saint Exupery Canada welcomed its first two data scientists**: Damien GRASSET (IRT employee) and Tapopriya MAJUMDAR ("Tapo", màd Scalian). Since their arrival, they have been investing in questions around **reinforcement learning**, with the aim of **creating links with Montreal's exceptional academic community** on these topics.

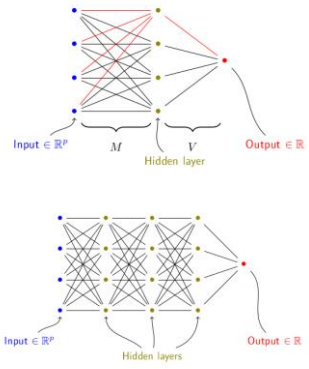
Two UQAM researchers have also been part-time at our office since January: Arthur CHARPENTIER and Marc-Olivier KILLIJIAN ("Marco"). Marco is one of the DEEL researchers in Quebec, and is interested in the **privacy chapter**. Arthur explores the **contribution of game theory and econometrics in reinforcement learning**. IRT Saint Exupery Canada is also building a collaboration with the team of Andréa LODI, MILA researcher involved in DEEL, on the use of **causal approaches in reinforcement learning**; Didier CHETELAT and Maxime GASSE, research officers at GERAD are engaged.

All these researchers are involved in the co-supervision and co-funding of two end-of-study trainees who will join IRT Saint Exupery Canada this summer: from ISAE Supaero and from ENSAE Paris, all engaged in a parallel Master's research (Operational Research in Toulouse, MVA in Saclay).

IRT Saint Exupery Canada is also working to create links with the four industries involved in the Quebec part of DEEL: Thales Canada, Bombardier.

 # Guillaume GAUDRON

FOCUS ON A PhD STUDENT, Joachim BONA-PELISSIER



We study the conditions that must be satisfied by a network and a learning sample to guarantee that all the parameters at which the objective function optimized during the learning phase is small are all contained in a set of small diameter. To do so, we establish **identifiability and stable recovery theorems**. Typically, when the architecture of the network is fixed, the theorem tells us how "diverse" the learning sample should be in order to get the stability property. The practical interest of such theorems is twofold: 1/ **Stability of the output and features in the latent space:** When the condition is satisfied and the optimization manages to find a good solution. The parameters and therefore the latent spaces and the prediction of the neural network are stably defined. In particular, they do not depend on the optimization algorithm, its initialization, the order of the sequence in the stochastic algorithm, the numerical tricks etc. This is a strong guaranty that provides confidence in the latent spaces (and the explainability using the latent spaces) and the output. 2/ **Stable recovery:** Assume that we know, through extra knowledge on underlying function learned by the neural network and expressivity theorems, that the prediction of the neural network can accurately approximate the underlying function. Then the stability property permits to control the error between the prediction of the neural network and the underlying function.

 # François MALGOUYRES, Joachim BONA-PELLISSIER, François BACHOC

ELECTRONIC COMPONENTS DATASET

Airbus has provided a dataset that comes from the production of electronic boards.

The dataset contains high **quality images** (1600x1200 pixels) of subset of electronic boards that are used to detect defects (missing components, solder short-cuts, mis-aligned components).

Thanks to the knowledge of boards design, it is possible to crop from these images all electronic components and create a components database with the type of component and its orientation.

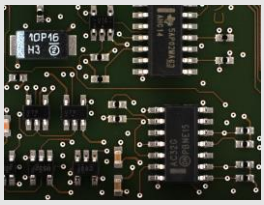


Figure 1

Figure 1 is an example of images that are inside the dataset.

Just like it has been done with the other datasets delivered to the DEEL project, it has been integrated to the deel datasets API for easy access by everyone in the core team.

Furthermore, this dataset represents an interesting study case for the numerous research projects that are currently being undertaken at DEEL: **explainability, fairness or even novelty detection.** In particular, in the context of **explainability of decision-taking algorithms**, the Gems-AI library been employed to globally explain the inner workings of a classifier trained to detect the type or the orientation of a given component.

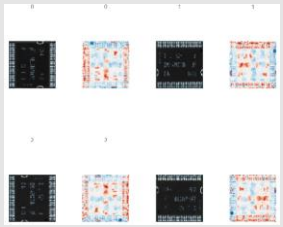


Figure 2

The F*igure 2* shows the global explanation of a rotation. It is interesting to notice that the classifier attached importance to pins and marks (like a human).

We also have ideas on the utilization of the dataset to see if a classifier or the detection of components is biased by the color or the number of pins.

 # Adrien GAUFFRIAU, Agustin Martin PICARD, Jean-Michel LOUBES, Mikaël CAPELLE

KEY DATES

28 May, Canceled	DEEL International Committee	➢ another meeting will be proposed
04 & 05 June	Certification mission Workshop	
29 & 30 Sept	MobiliT.AI	➢ postponed to spring 2021

PROJECT MANAGEMENT

Project Contract	R2V1 has been signed by 10/11 partners
Certification Mission	12-month prolongation new team member : Ludovic PONSOLLE (APSYS)
DEEL Team	MàD : Adrien ELFASSI (AIRBUS DS) 2 internships for 6 months : Eduardo DADALTO CAMARA GOME & Pedro COELHO