DEEL
DEpendable & Explainable Learning

DEEL MAG #08

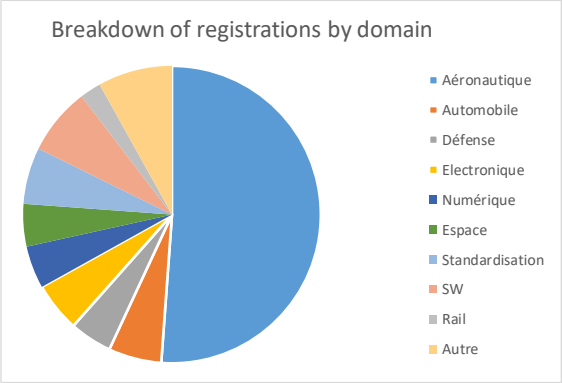
TOULOUSE & QUEBEC JOIN FORCES TO DEVELOP AI FOR CRITICAL SYSTEMS

WORKSHOP « MACHINE LEARNING IN CERTIFIED SYSTEMS »

On January 14th and 15th, **DEEL organized the Machine Learning in Certified Systems (MLCS)** workshop: https://mlcertifiedsystems.deel.ai/

MLCS addressed several scientific challenges and recent works related to certification of machine learning. Among the covered topics were generalization bounds, uncertainty quantification, formal methods for critical software verification, numerical aspects, and explainability.

The event was a great success with 500 registered participants (maximum allowed) from all over the world (32 countries represented). The audience was a perfect balance between industrials and academics; all the transportation domains were represented (with aeronautics accounting for 50% of them). Videos and slides are available on the website. We gratefully thank the speakers and all the people that contributed to this event.



 # Grégory FLANDIN, Adrien GAUFFRIAU, Sébastien GERCHINOVITZ

NDT DATASET FOR OUT OF DISTRIBUTION (OOD) DETECTION

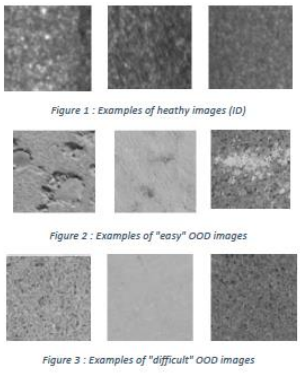


Figure 1 : Examples of heathy images (ID)

Figure 2 : Examples of "easy" OOD images

Figure 3 : Examples of "difficult" OOD images

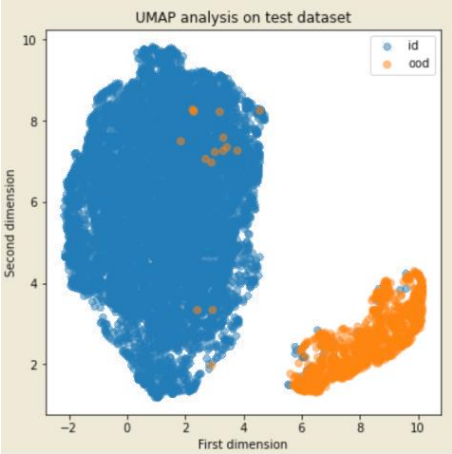
Safran has released a **dataset of images for the OOD detection challenge**. This dataset, called **NDT (Non Destructive Testing), is dedicated to the visual inspection of aeronautical parts**. All the images of the dataset are gray-level patches of size 64*64. The dataset is composed of nearly 20000 images of healthy parts, which are considered as the in-distribution (ID) images. The OOD images have been taken from a public database, called SDNET2018 (Dorafshan et al., Data in Brief, 2018), and show cracks in walls and pavements. Although they are different, the textures of the ID and the OOD images are very close. The NDT dataset contains nearly 1700 images of OOD, which are divided into two categories: one considered as « easy », the other considered as « difficult ». Assessing the performances on each OOD category will be very important to evaluate the robustness of the methods developed in the OOD detection challenge.

 # Camille CHAPDELAINE

KEY DATES & INFORMATIONS

Certification Mission	Next workshops : 3rd & 4th Match – 31th & 1st April
Reinforcement learning School	Free and entirely virtual, total of 6 days : March 25th-26th, April 1st- 2sd and April 8th-9th
« Les Carrefours DEEL »	5th edition → 4th March
MobiliT.AI 2021	Interactive and dynamic format : May 10th to 12th

DETECTING OUT-OF-DISTRIBUTION (OOD) SAMPLES VIA VARIATIONAL AUTO-ENCODER



Deep neural networks are often trained with strong assumptions. For instance, the test data distribution is assumed to be similar to the training data distribution. However, when employed in real-world tasks, the data distribution can be sometimes different and OOD detection is important to **prevent AI systems from making prediction errors**.

An OOD is an unknown observation and can be a novelty, anomaly or outlier sample.

In DEEL, we are currently studying different approaches to detect OOD instances. A promising approach is to use **Variational Auto-Encoder (VAE).**

Our research is focused on:

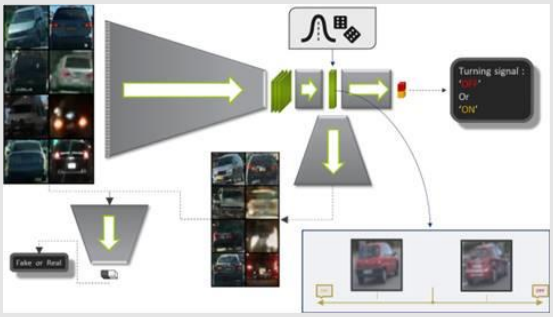
- Exploration of several VAE models (from Euclidian latent space to Riemannian latent space),

- Detecting unknown observations based on latent space exploration (visual exploration: PCA, T-SNE, UMAP and machine learning algorithms applied to latent space: One-class SVM, K-Means…),

- Detecting unknown observations based on loss estimation (reconstruction loss, negative log-likelihood, likelihood regret...).

 # Adrien EL FASSI

EXPLORING DECISION FUNCTION FRONTIER TO IMPROVE INTERPRETABILITY AND ROBUSTNESS OF DEEP NEURAL NETWORK USING GAN AND VAE FRAMEWORKS



The always increasing number of parameters of Deep neural network models is a double-edge sword: on one hand it can approximate any function easily, on the other hand it is also more prone to overfit training set, or to learn unwanted bias due to data collection protocol issues. Thus, **improving our capability to assess closely what is going on inside a DNN is utterly important**. **More specifically, we would like to have an in-depth understanding of how the DNN inner representations of inputs evolve around the learnt decision function.**

With this objective in mind, we have adapted previous works related to **GANs and VAEs in order to incorporate an ability for the model** to generate input-like objects from DNN inner representations, while preserving the decision capability of the original architecture. This is possible with the use of negative feedbacks able to stabilize the optimization of the various losses. At the end, it could be shaped as a plug-and-play module able to “open-up” any learnt representation to increase robustness of critical industrial ML systems.

 # Raphaël PUGET