DEEL

Dependable, Explainable & Embeddable Learning

IVADO

IRT SAINT EXUPÉRY

CRIAQ

ANITI Université Fédérale Toulouse Midi-Pyrénées

Institut intelligence et données

ONERA

AIRBUS

DEEL

Dependable, Explainable & Embeddable Learning

Applying the **DEEL Toolbox** on **LARD**

Mélanie Ducoffe, Maxime Carrere, Léo Féliers, Adrien Gauffriau, Christophe Gabreau, Jean-Brice Ginestet, Franck Mamalet, Claire Pagetti, Thierry Sammour

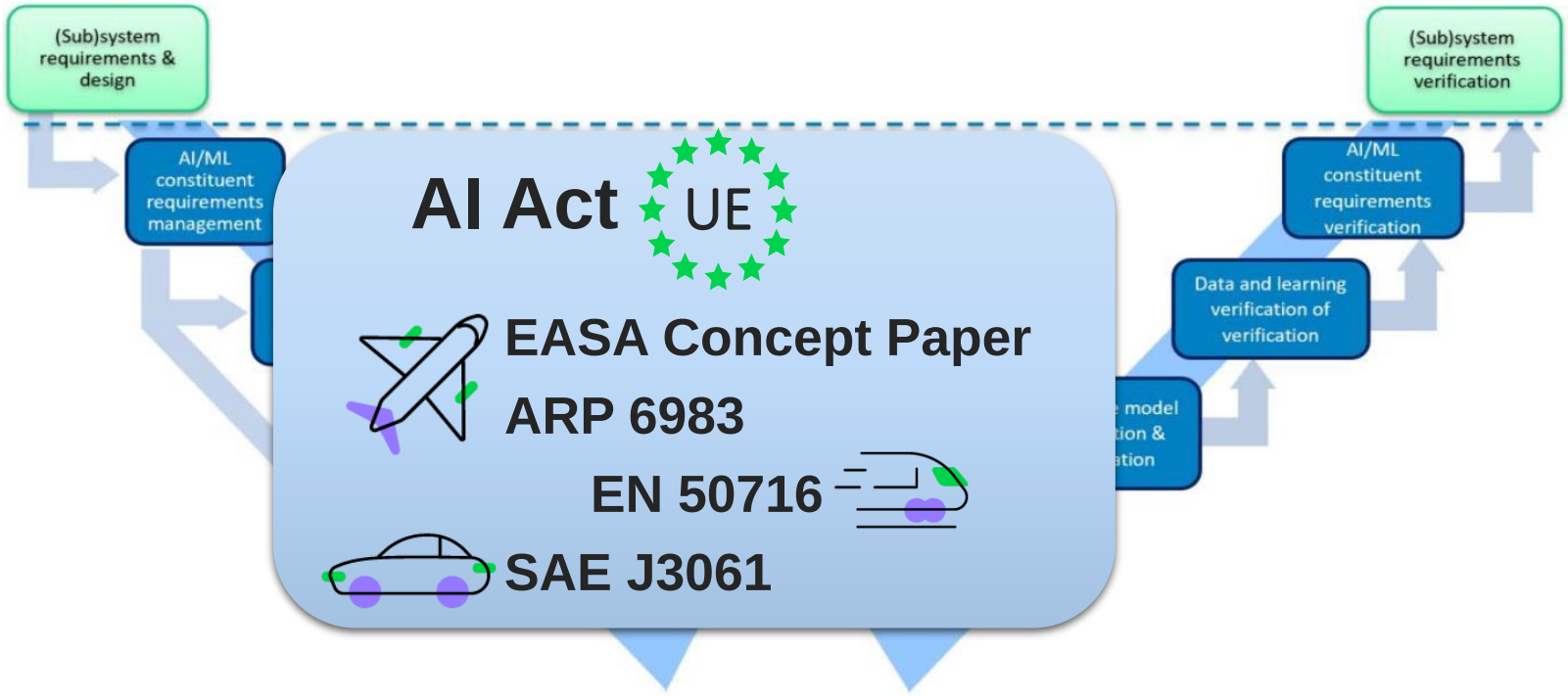
Thierry Daubos, Corentin Friedrich, Agustin Martin Picard, Vincent Mussot, Paul Novello, Yannick Prudent, David Vigouroux

DEEL Certification Mission

DEEL Core Team

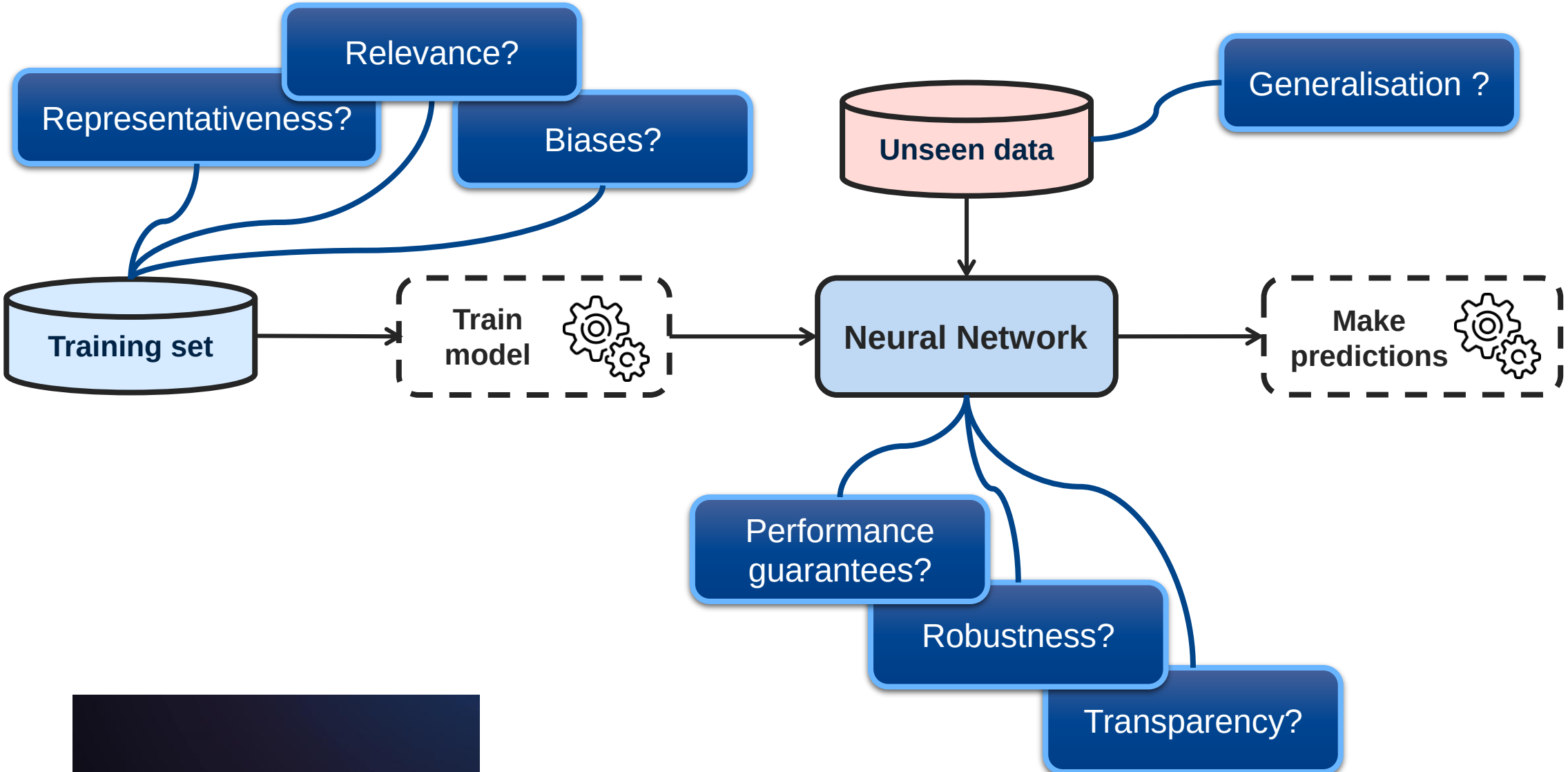
Norms & regulation

- *How to integrate IA in critical systems?*



Learning assurance from EASA

A typical ML process



DEEL

Dependable, Explainable & Embeddable Learning

LARD
Landing Approach Runway Detection

IVADO

IRT SAINT EXUPÉRY

CRIAQ

ANITI Université Fédérale Toulouse Midi-Pyrénées

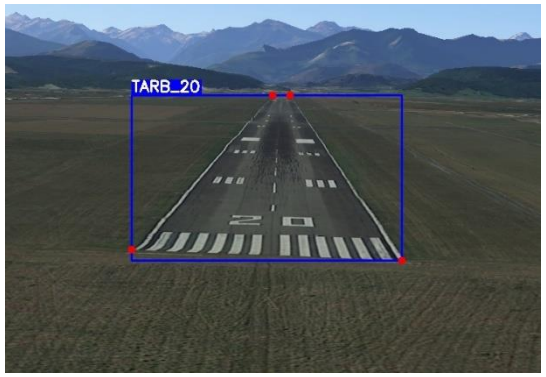
Institut intelligence et données

ONERA
THE FRENCH AEROSPACE LAB

AIRBUS

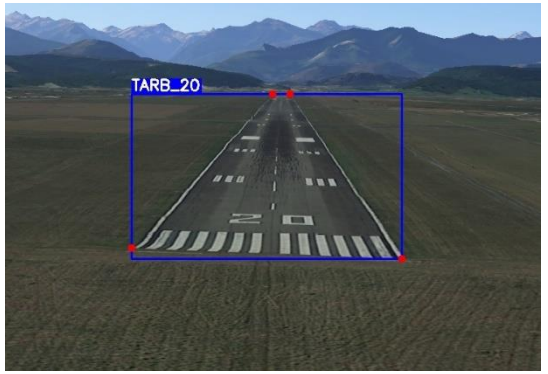
LARD

TASK: Autonomous vision-based landing

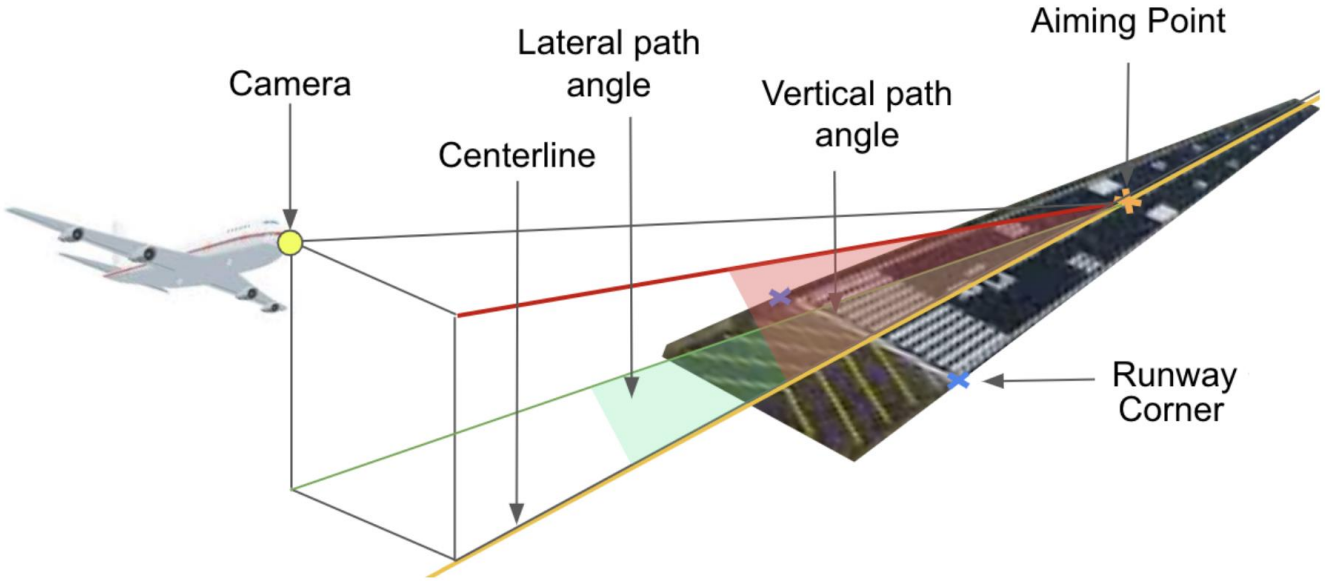


LARD: System-level ODD

TASK: Autonomous vision-based landing



Defining a Generic landing approach cone:



Parameter	range
Along track distance	[0.08, 3] NM
Vertical path angle	[-2.2, -3.8]°
Lateral path angle	[- 4, 4] °
Yaw	[-10,10] °
Pitch	[-8,0] °
Roll	[-10,10] °

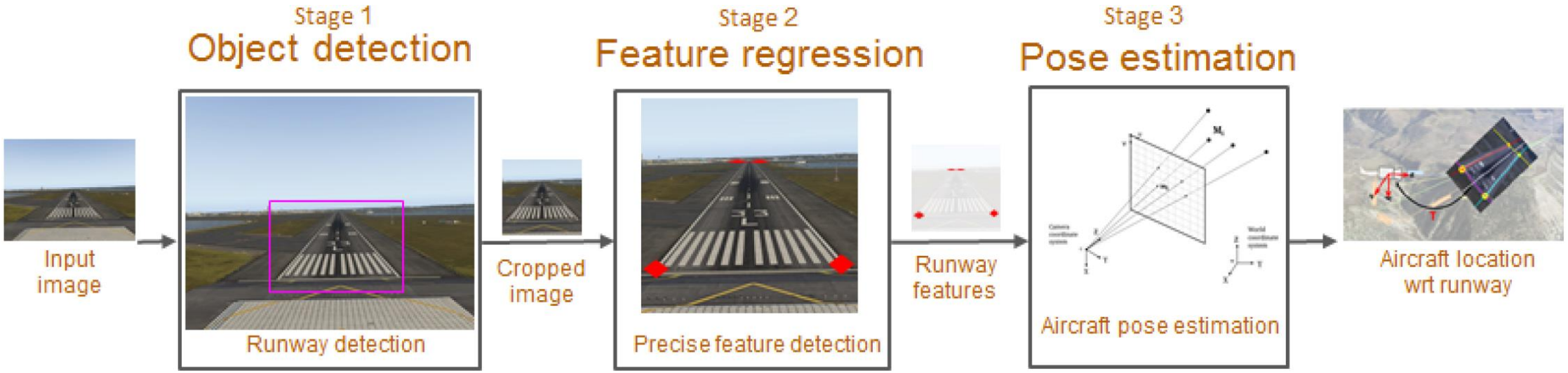
Operational Design Domain

LARD: ML Constituent

Intended function 1 (VBL intended function):

- Pose estimation of the aircraft with respect to the runway when the aircraft flies within the generic landing approach cone.

3- stage architecture:



LARD: ML Constituent

Refinement of System-level ODD

- Span a wide range of positions inside the cone
- Cover a variety of airports all around the world

ODD 2 (of VBL): *ODD 1 restriction*

1) Runway with a piano

*2) **Single runway** for which the current position is considered within the approach cone*

3) Runway fully visible on the image (no occlusion)

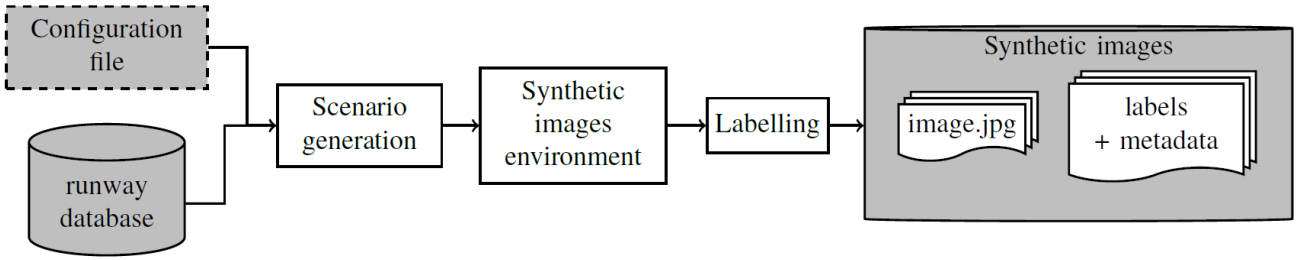
4) Optimal weather (clear daylight, no clouds)

Ensures **precision** of the pose estimation

Constraint from the generator

LARD: Dataset for Runway Detection

- **17 000 images** (2048x2448)
  From ~ **200+** different runways
- Videos of **real landing footage**
  1800 frames from **200** videos manually annotated
- **A scenario generator** for trajectories in Google Earth with **Automatic labeling** of images

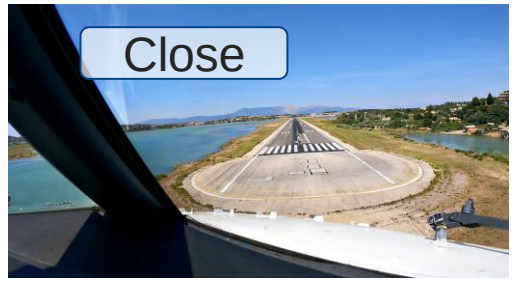


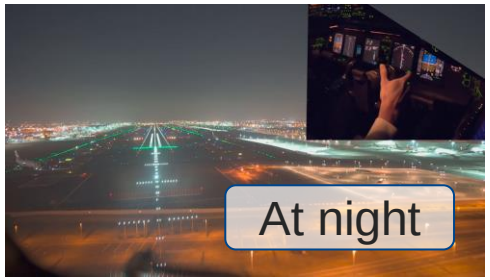
LARD: Dataset for Runway Detection

DEEL
Dependable, Explainable & Embeddable Learning

15K Synthetic images (train & test)

1 800 Real images (test only)





LARD in action

DEEL

Dependable, Explainable & Embeddable Learning

The Big Picture

IVADO

IRT SAINT EXUPÉRY

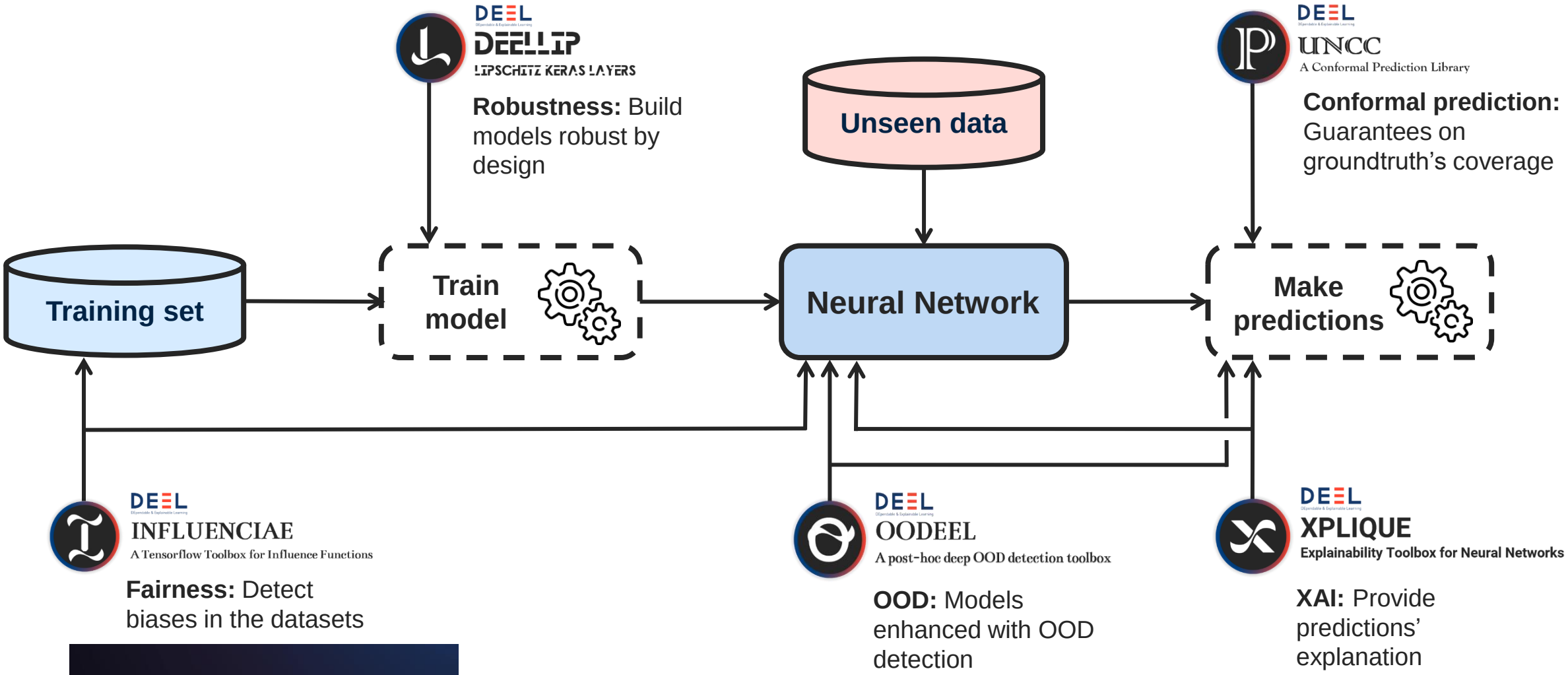
CRIAQ

ANITI Université Fédérale Toulouse Midi-Pyrénées

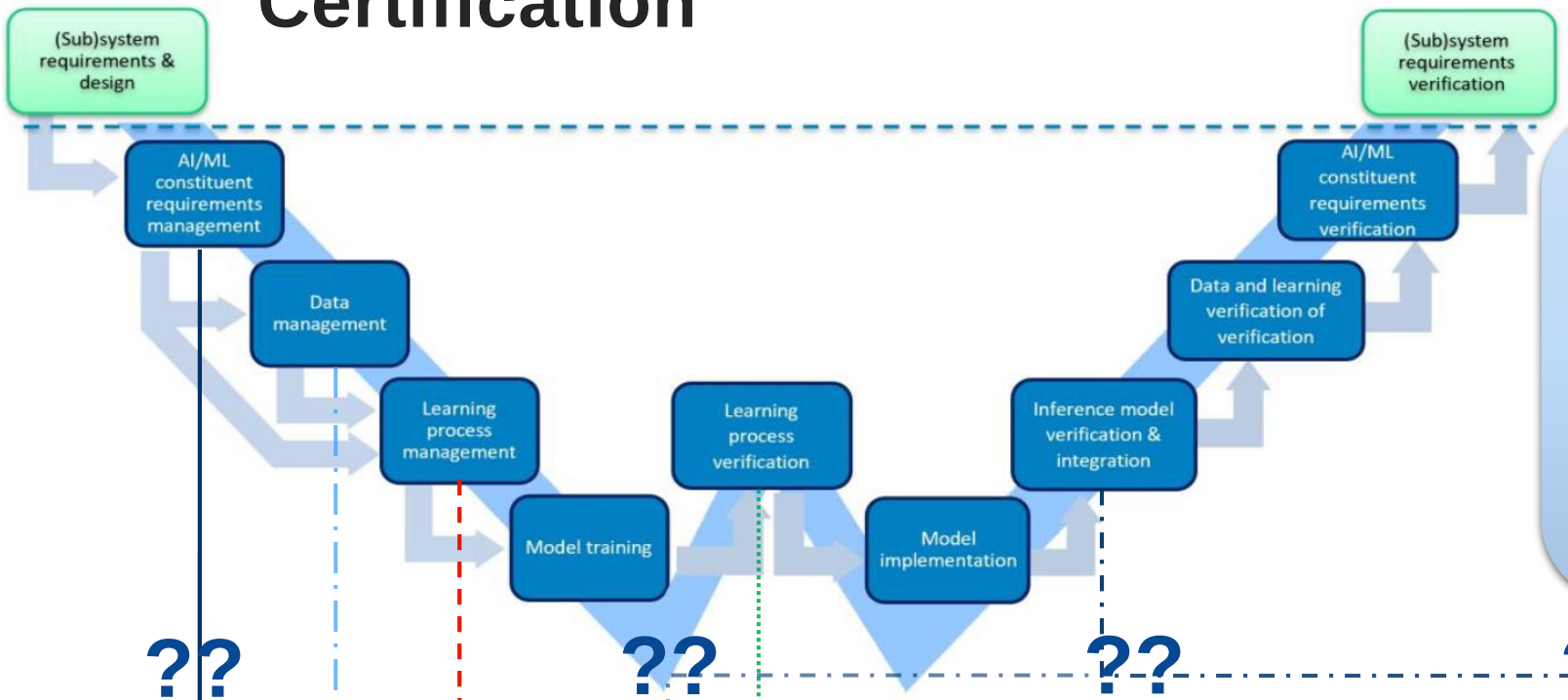
Institut intelligence et données

The Big Picture

DEEL
Dependable, Explainable & Embeddable Learning



Certification



AI Act UE

EASA Concept Paper

ARP 6983

EN 50716

SAE J3061

?? ?? ?? ?? ??

DEEL

Explainability | *Robustness* | *Fairness* | *Uncertainty Quantification* | *OOD Detection*

DEEL

How to choose the right method?

DEEL

Dependable, Explainable & Embeddable Learning

Starting with: Robustness

IVADO

IRT SAINT EXUPÉRY

CRIAQ

ANITI Université Fédérale Toulouse Midi-Pyrénées

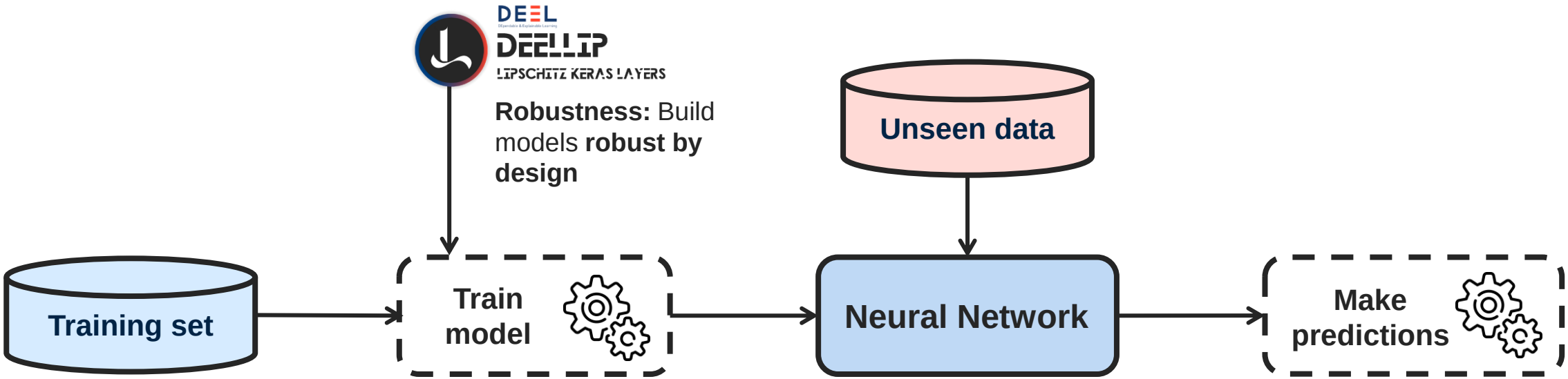
Institut intelligence et données

ONERA

THE FRENCH AEROSPACE LAB

AIRBUS

DEEL Libraries: Robustness



Attacks on LARD

LARD contains only 1 class (runway), we focus exclusively on **fabrication** and **vanishing** attacks

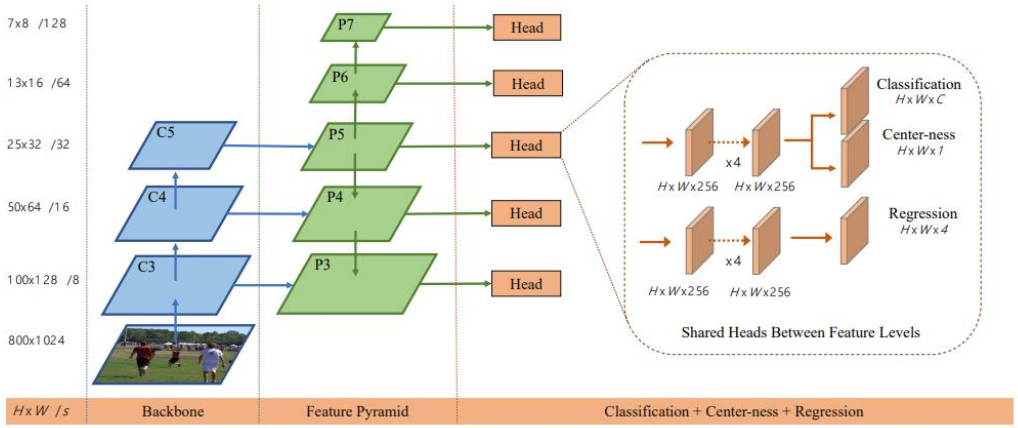
Original

Robust *by-design* model

DEEL
DEpendable & Explainable Learning
DEELLIP
LIPSCHITZ KERAS LAYERS

Layer	Original FCOS (using `torch`)	1- Lipschitz FCOS (using `deel-torchlip`)
Convolution	`Conv2d`	`SpectralConv2d`
Activation	`ReLU`	`GroupSort2`
Pooling	`MaxPool2d`	`ScaledL2NormPool2d`
Classification Loss	`sigmoid_focal_loss`	`sigmoid_focal_loss` + temperature factor

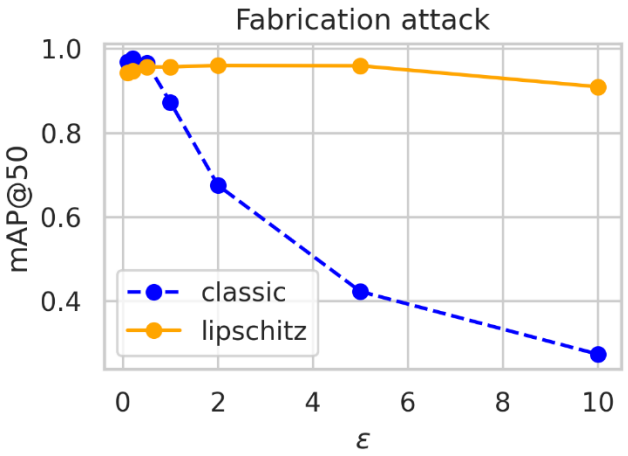
Only a few layers / loss need to be replaced to get a 1-Lipschitz object detector!



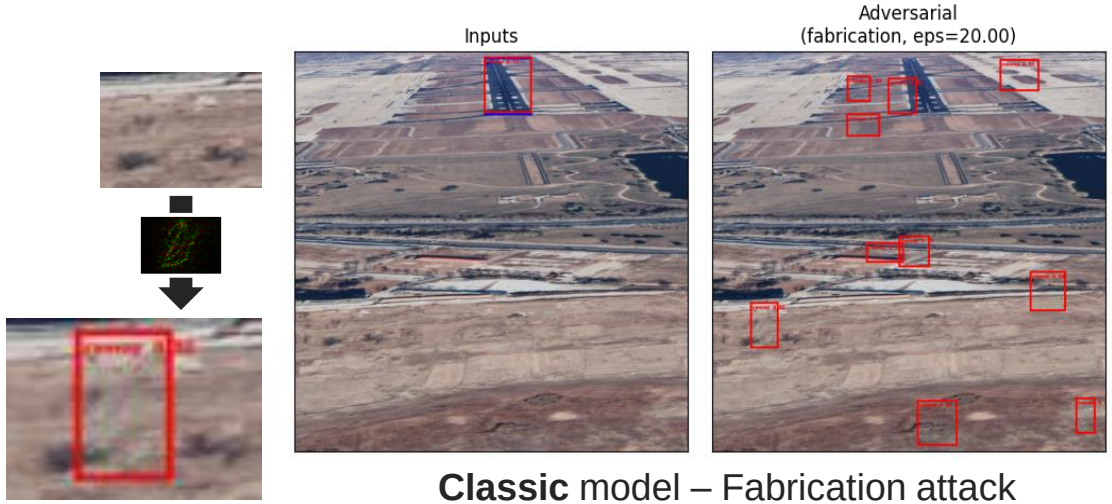
FCOS architecture. (Tian et al, 2019) "*FCOS: Fully Convolutional One-Stage Object Detection*".

Robustness to fabrication attack

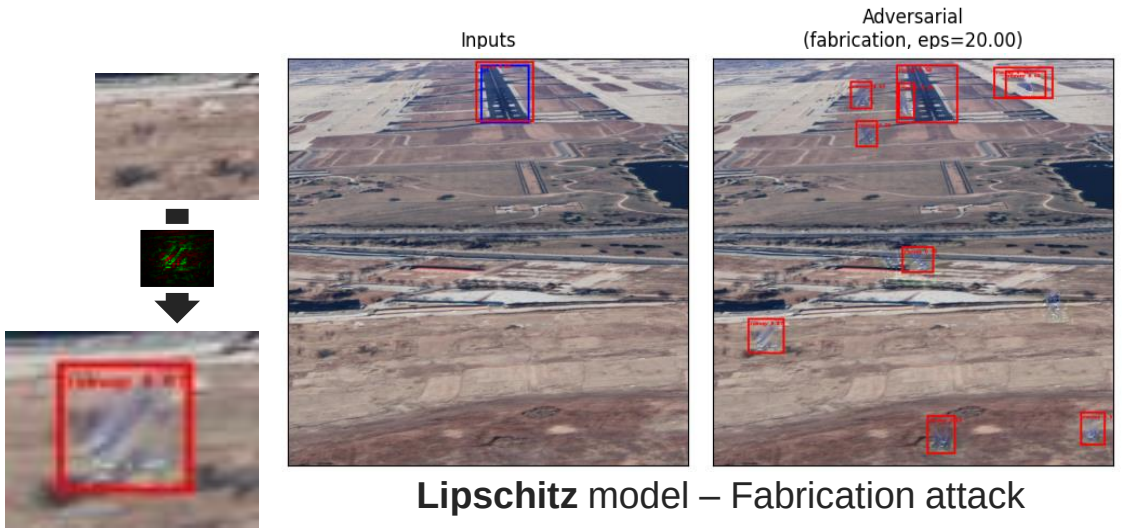
- **Objective:** Find minimal perturbation (of L2 norm ϵ) to trick the model into *detecting false targets* (randomly defined).



Robustness of classic vs Lipschitz equivalent models wrt L2 norm of *fabrication adversarial attacks*, evaluated using ***mAP@50*** *metric* (the higher the better).



Classic model – Fabrication attack

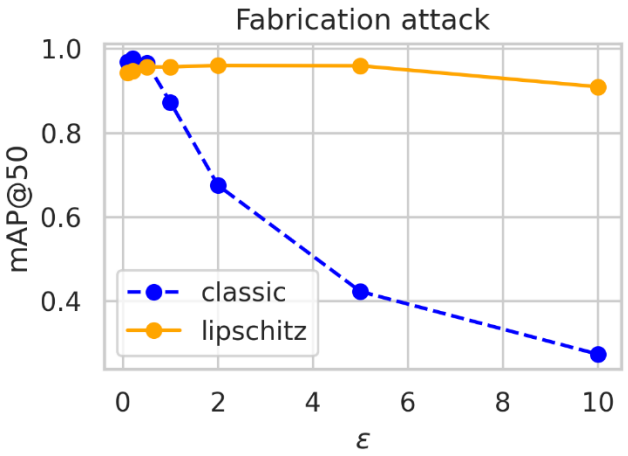


Lipschitz model – Fabrication attack

For very large perturbations, attacks are able to trick both models but the **modifications are only visible on Lipschitz model.**

Robustness to fabrication attack

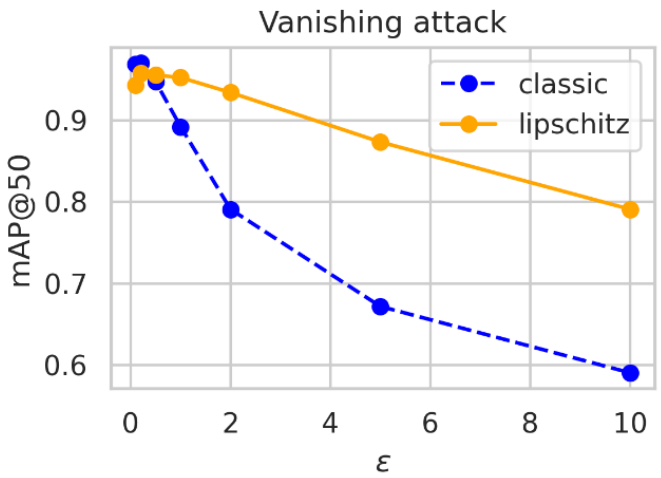
- **Objective:** Find minimal perturbation (of L2 norm ϵ) to trick the model into *detecting false targets* (randomly defined).



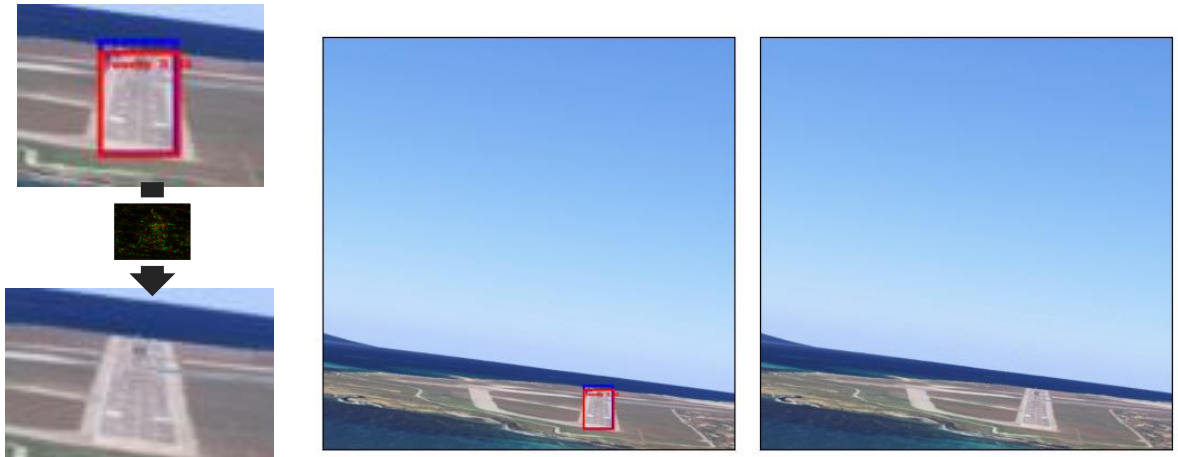
Robustness of classic vs Lipschitz equivalent models wrt L2 norm of *fabrication adversarial attacks*, evaluated using ***mAP@50*** *metric* (the higher the better).

Robustness to vanishing attack

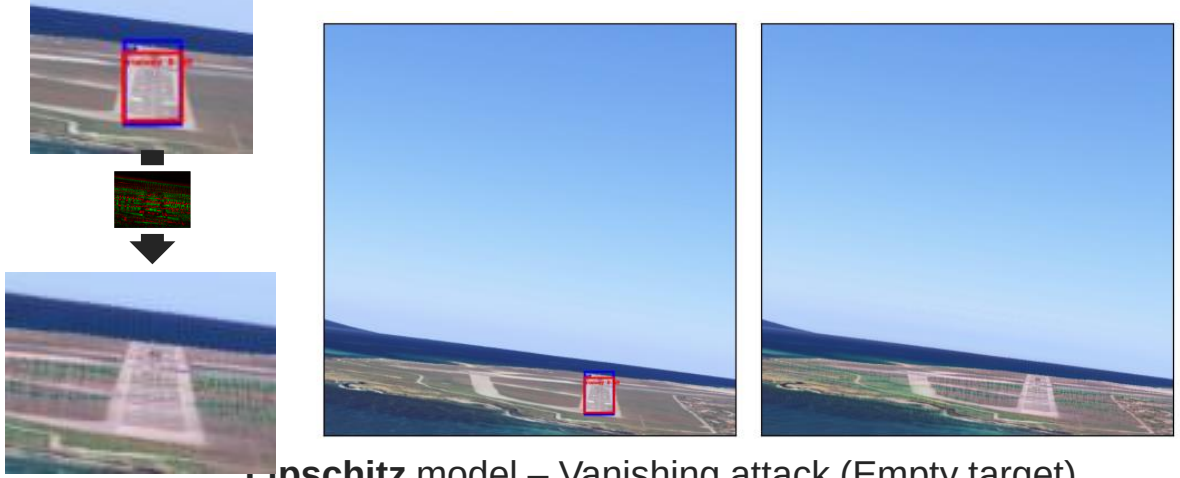
- **Objective:** Find minimal perturbation (of L2 norm ϵ) to trick the model into detecting no more targets.



Robustness of lipschitz vs classic equivalent models wrt L2 norm of *vanishing adversarial attacks*, evaluated using ***mAP@50*** *metric* (the higher the better).



Classic model – Vanishing attack (Empty target)

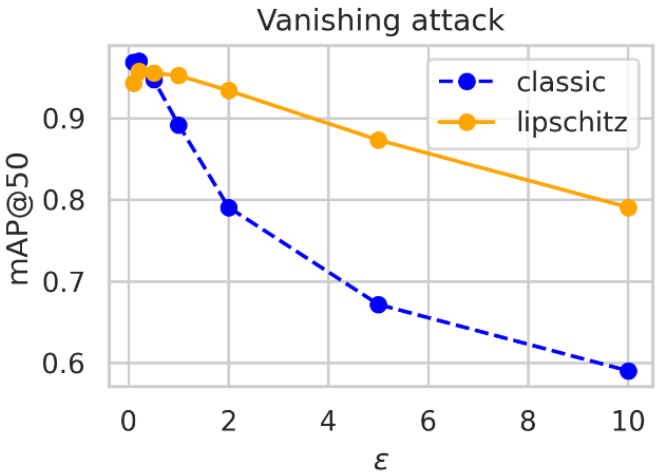


Lipschitz model – Vanishing attack (Empty target)

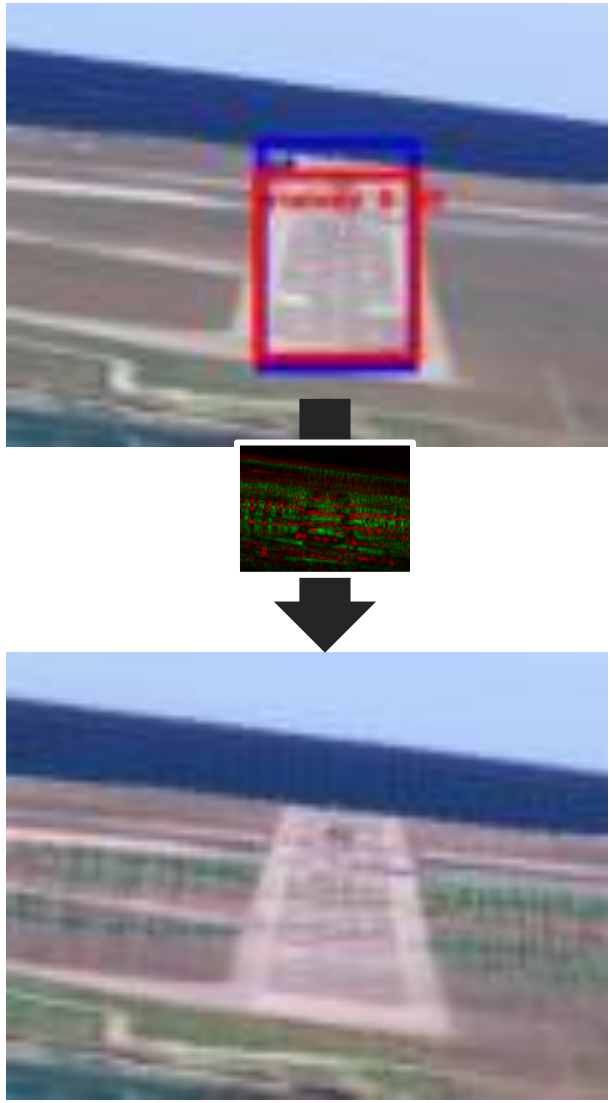
For very large values of ϵ (here 20), attacks are able to trick both models but the **edits are only visible on Lipschitz model.**

Robustness to vanishing attack

- **Objective:** Find minimal perturbation (of L2 norm ϵ) to trick the model into detecting no more targets.



Robustness of lipschitz vs classic equivalent models wrt L2 norm of *vanishing adversarial attacks*, evaluated using ***mAP@50*** *metric* (the higher the better).

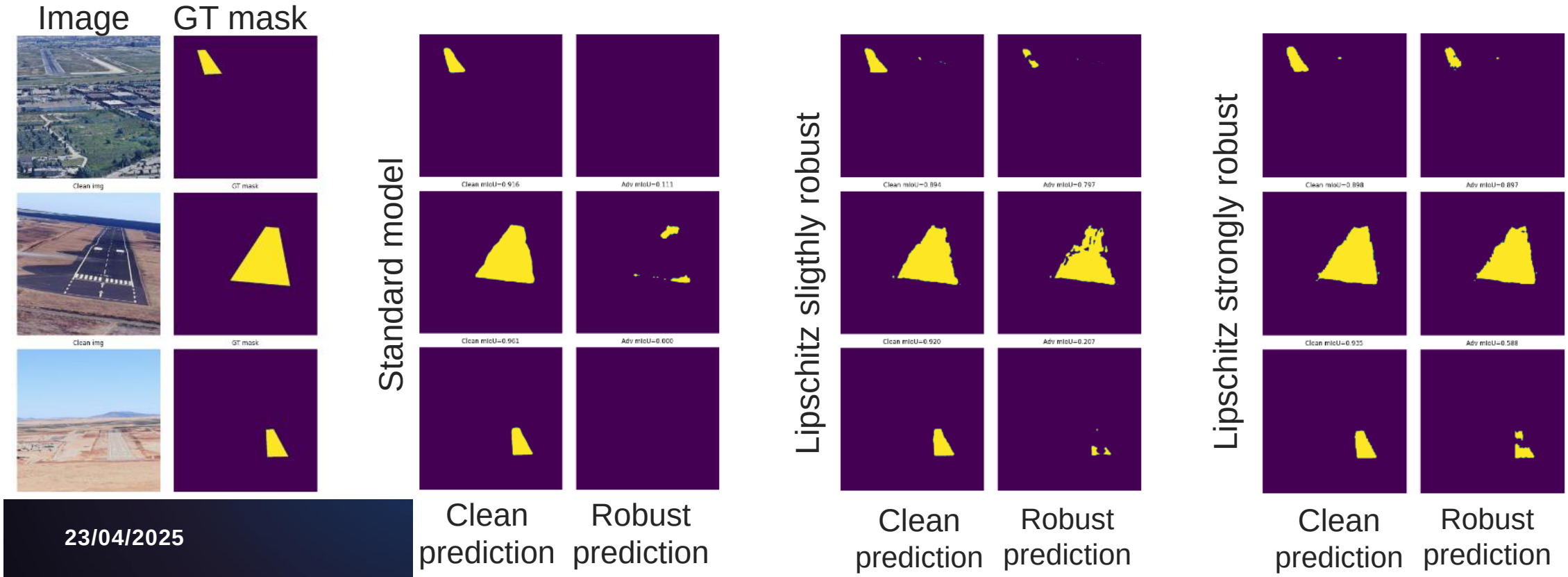


Robustness to vanishing attack

Segmentation

Model	Clean IoU	Robust IoU

Small *Fully Convolutional Network (FCN-8)*
Adversarial attack: vanishing objective
($\epsilon = 1.0$)



DEEL

Dependable, Explainable & Embeddable Learning

Looking for biases

IVADO

IRT SAINT EXUPÉRY

CRIAQ

ANITI Université Fédérale Toulouse Midi-Pyrénées

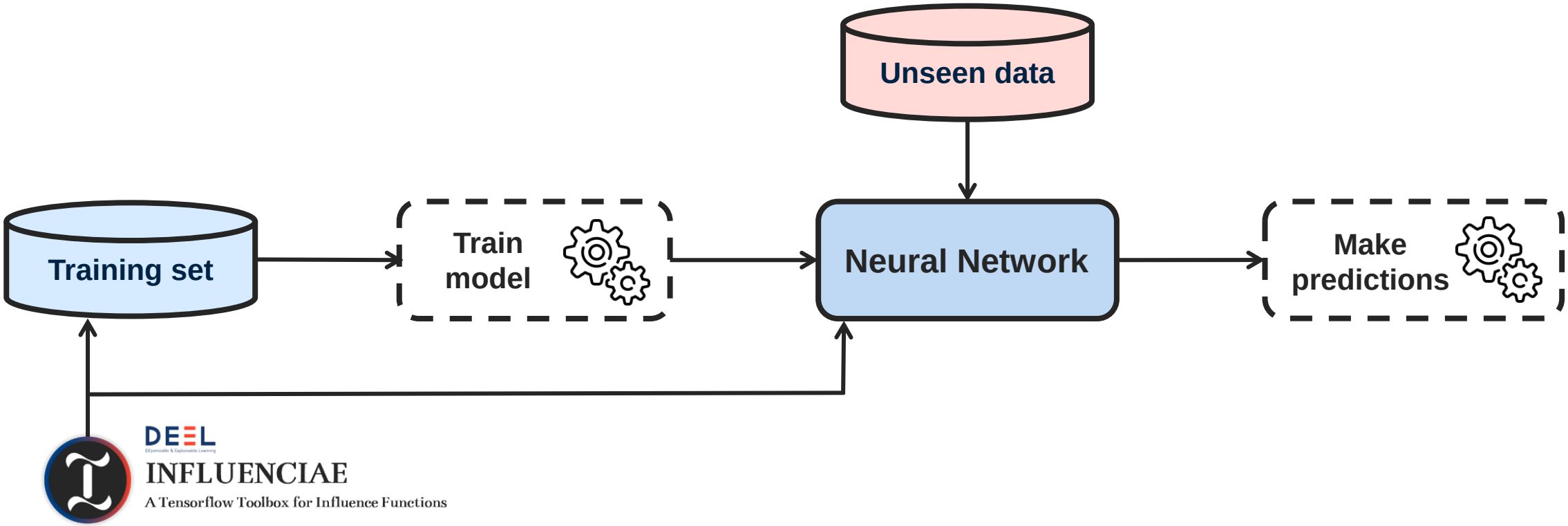
Institut intelligence et données

ONERA

THE FRENCH AEROSPACE LAB

AIRBUS

DEEL Libraries: Looking for biases



Fairness: Detect biases in the dataset

Influenciae

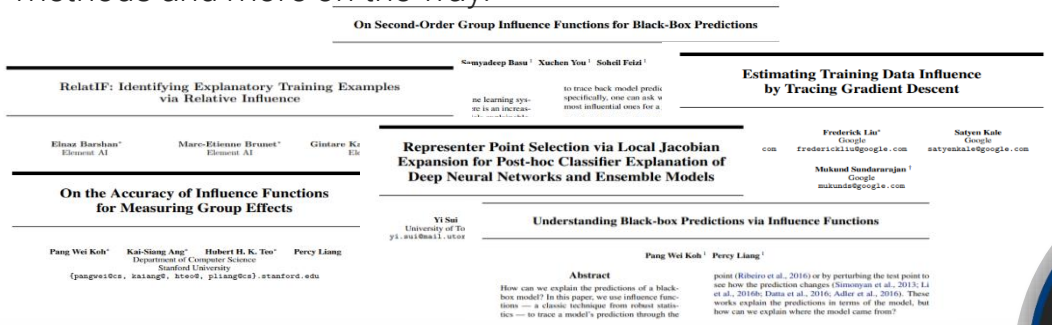
A Tensorflow Toolbox for Influence Functions

Agustin PICARD*, Lucas HERVIER, David VIGOUROUX, Thomas FEL

Optimized for Tensorflow / Keras ecosystem.

DEEL
Dependable, Explainable & Embeddable Learning

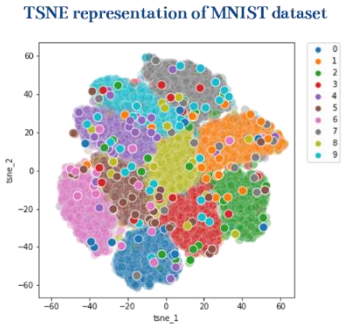
(1) State-of-the-art Influence Function implementations: 6 methods and more on the way!



(2) Tutorials 7 Colab notebooks to get your hands on the tool's possibilities !

```python
from deel.influenciae.common import InfluenceModel, ExactIHVP
from deel.influenciae.influence import FirstOrderInfluenceCalculator

influence_model = InfluenceModel(mnist_cnn, start_layer=-1, loss_function=unreduced_loss)

# Start by creating an object capable of efficiently computing inverse-hessian-vector products
ihvp_calculator = ExactIHVP(influence_model, train_dataset=ds_train)

# Create the influence calculator object
influence_calculator = FirstOrderInfluenceCalculator(influence_model, ds_train, ihvp_calculator)

influence = influence_calculator.compute_influence_values(ds_train.batch(32))
```

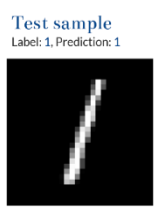
Learn to gain insights into your dataset!

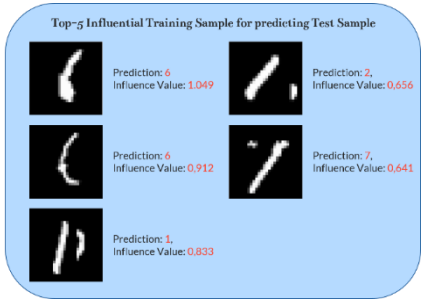


(3) Example-based XAI to gain knowledge on your model's prediction!

But also:

- ❖ **Understanding model-behavior**
- ❖ **Debugging models**
- ❖ **Mislabelling detection**





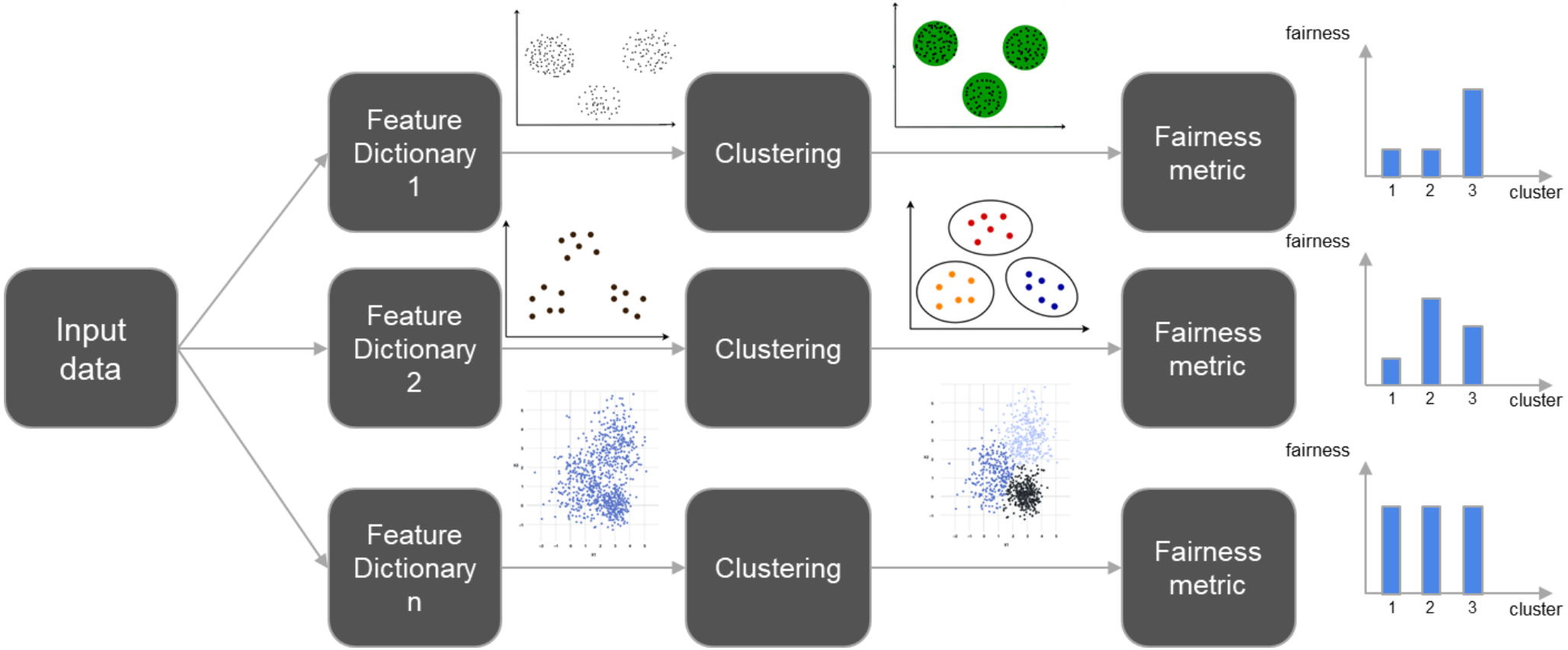
(4) On-going development

- ❖ **Benchmark tools**
- ❖ **Evaluation of the different methods**
- ❖ **Uncertainty estimation**
- ❖ **Implementation optimization**
- ❖ **…**

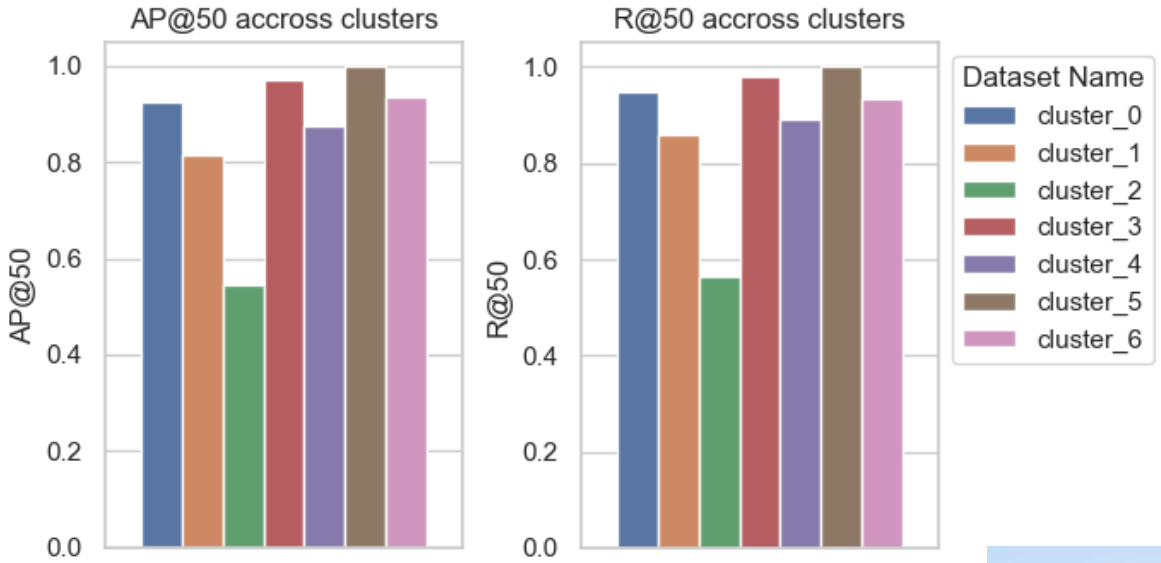
github.com/deel-ai/influenciae

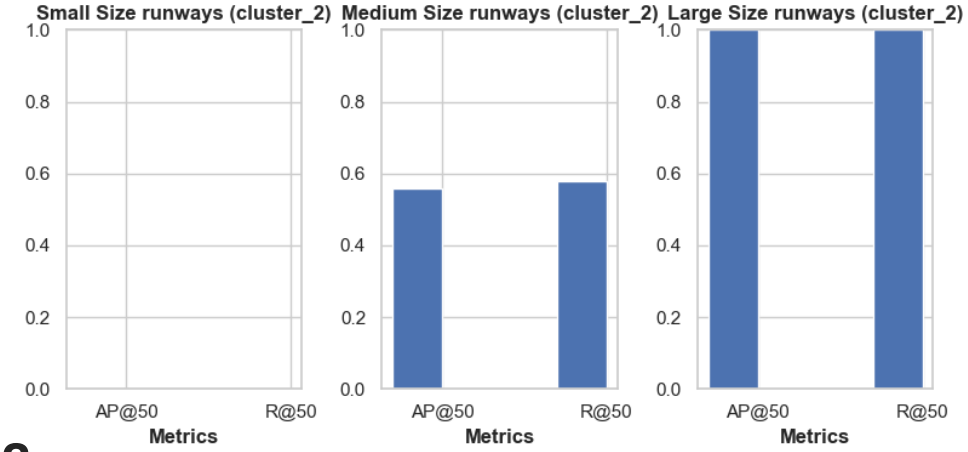
See also: github.com/deel-ai/xplique

Finding biases through *unsupervised clustering*



Insights from Color Clustering





Cluster_2

Cluster_5

DEEL

Dependable, Explainable & Embeddable Learning

Improving detection with OOD

IVADO

IRT SAINT EXUPÉRY

CRIAQ

ANITI Université Fédérale Toulouse Midi-Pyrénées

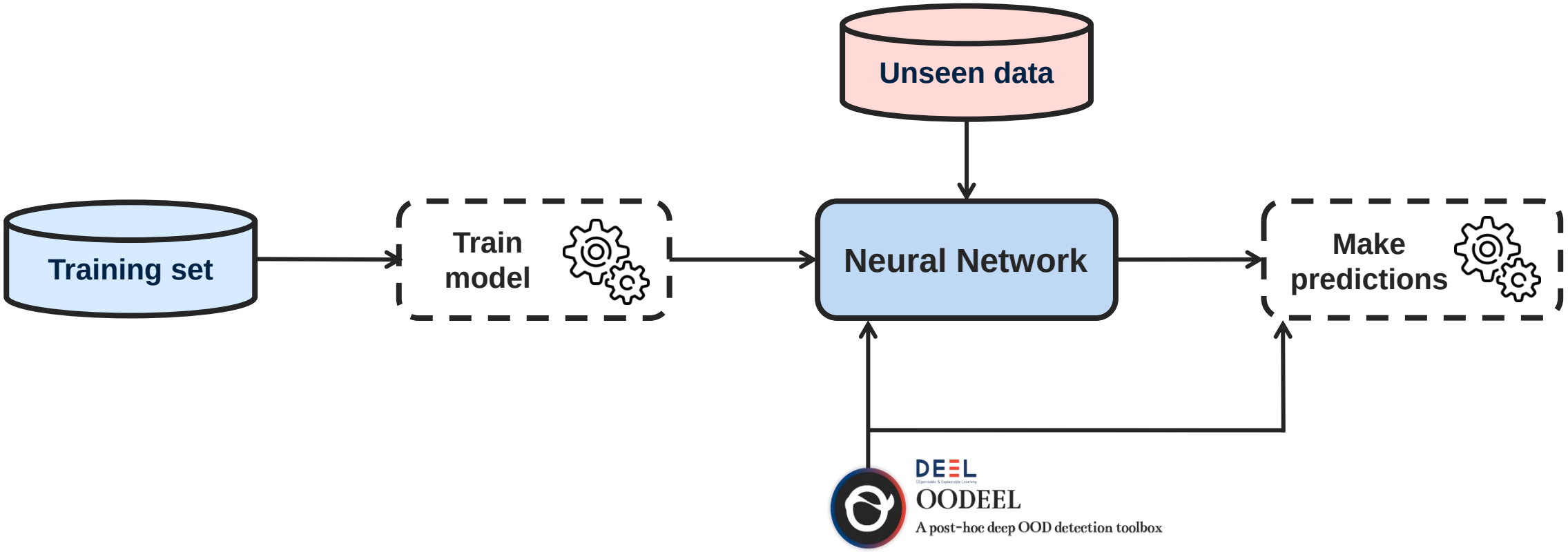
Institut intelligence et données

ONERA

THE FRENCH AEROSPACE LAB

AIRBUS

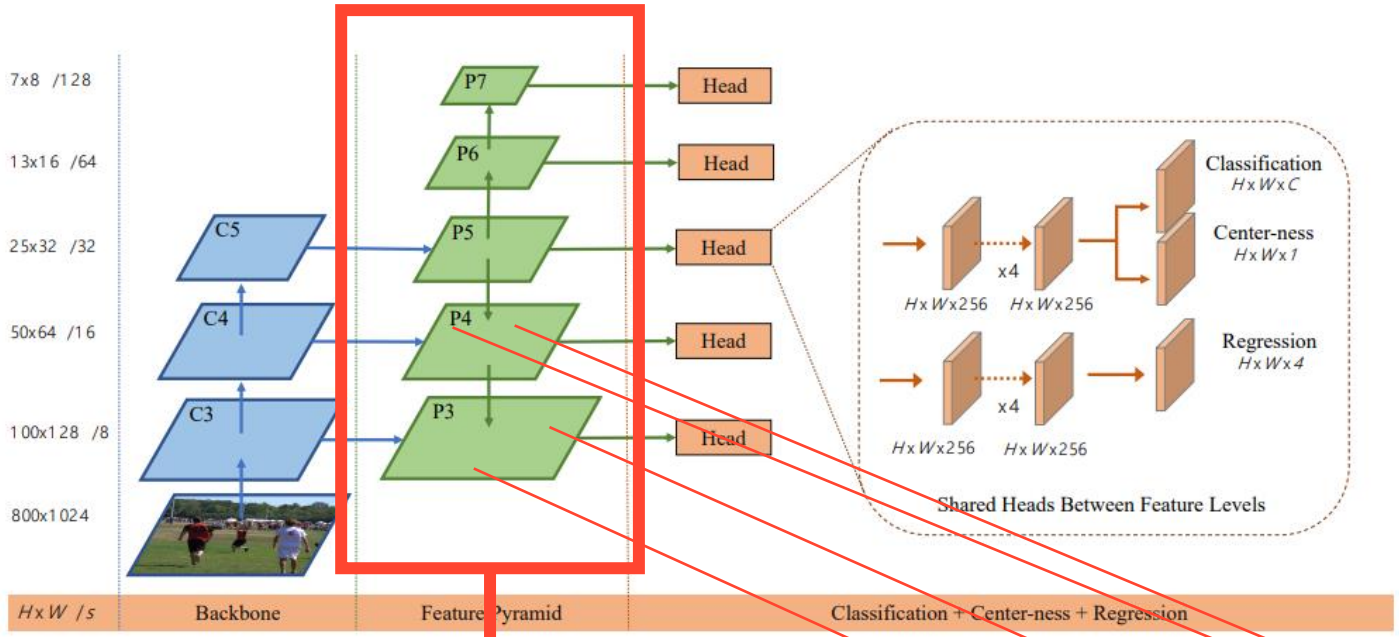
DEEL Libraries: OOD Detection



OOD: Models enhanced with OOD detection

Defining OOD in object detection

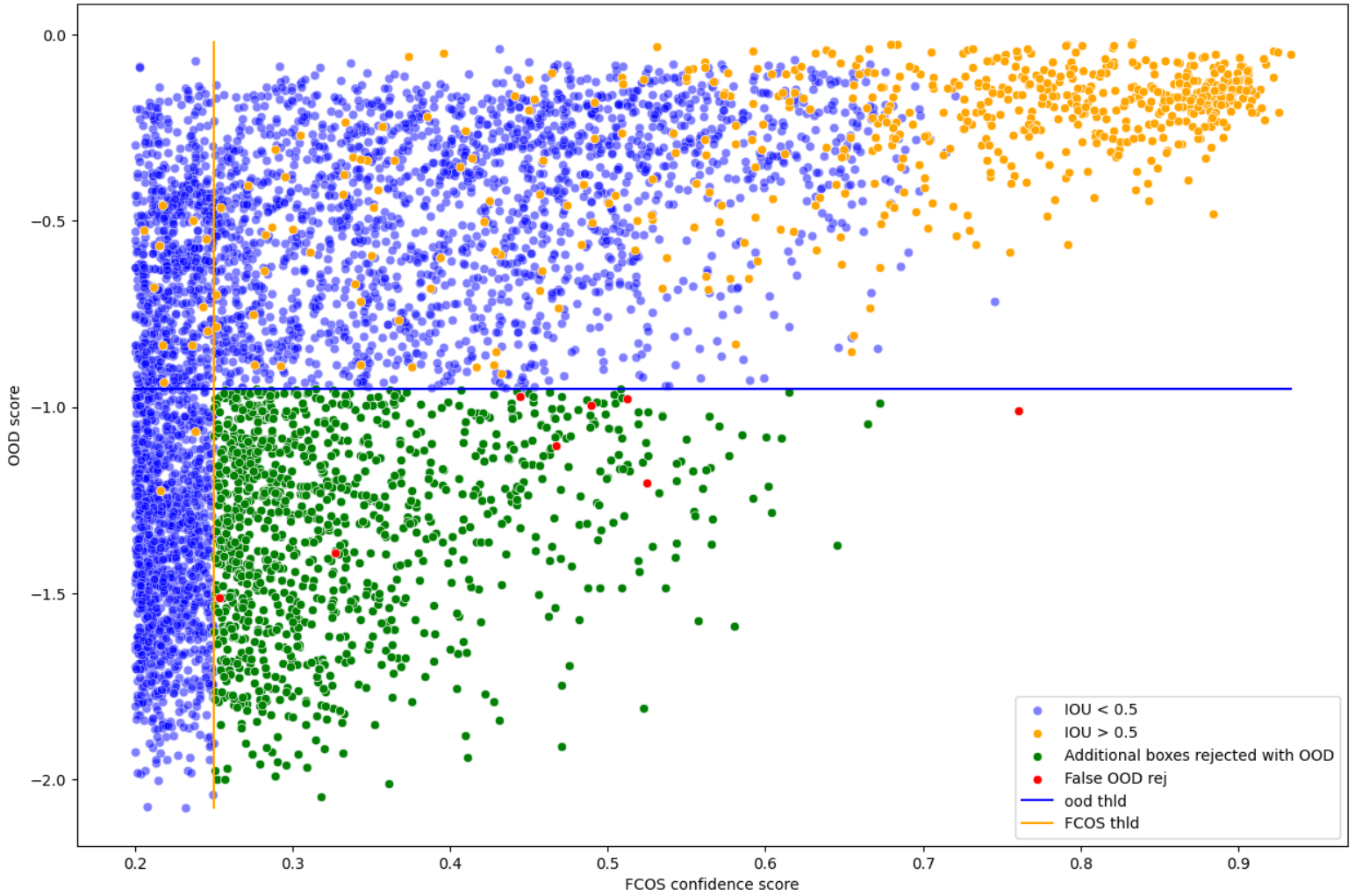
Using OODEEL for OOD in object detection



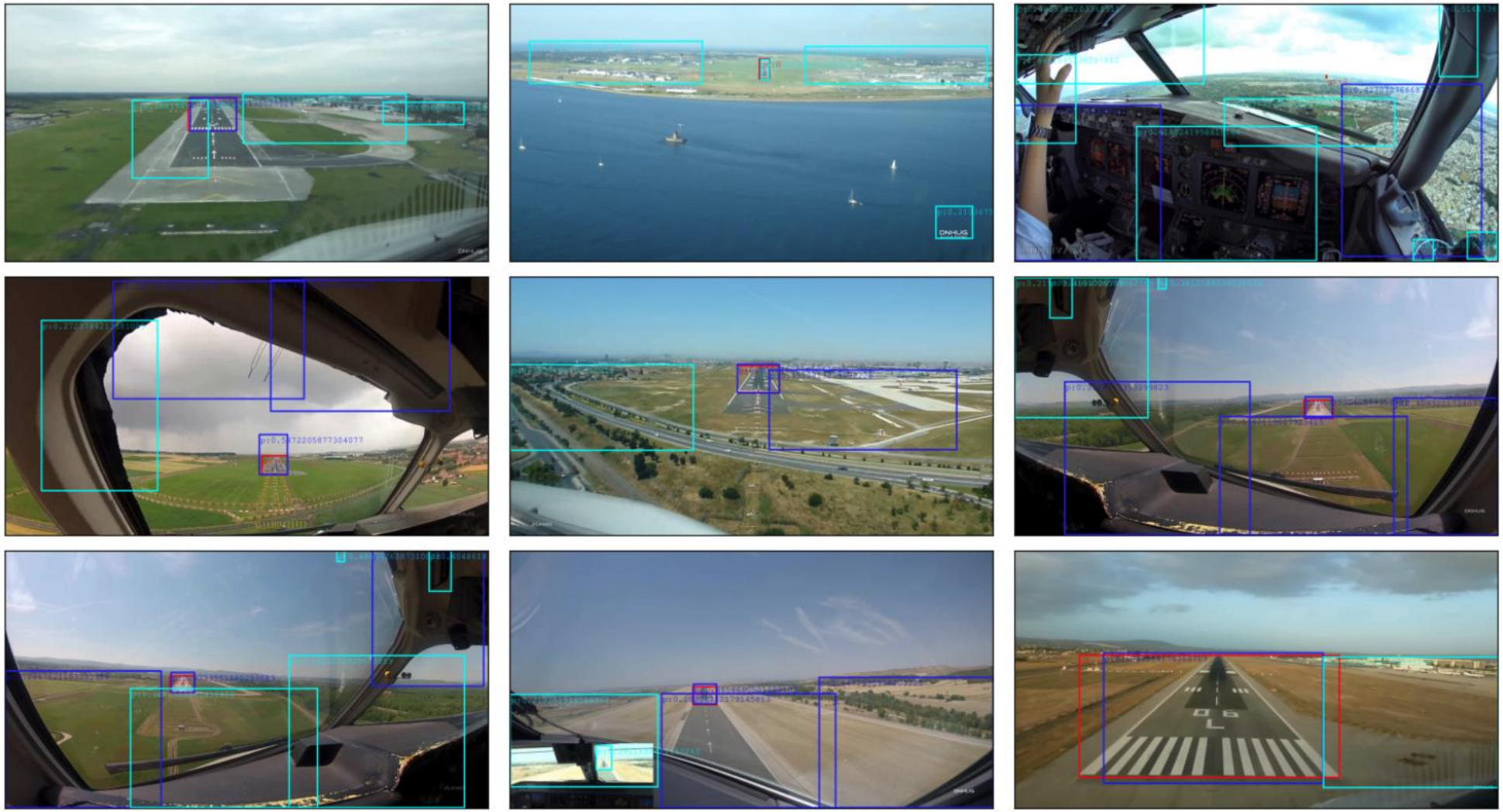
DEEL OODEEL

A post-hoc deep OOD detection toolbox

Using OODEEL for OOD in object detection



OOD in object detection: OOD bbox filtering



DEEL

Dependable, Explainable & Embeddable Learning

Obtaining guarantees on bounding boxes

IVADO

IRT SAINT EXUPÉRY

CRIAQ

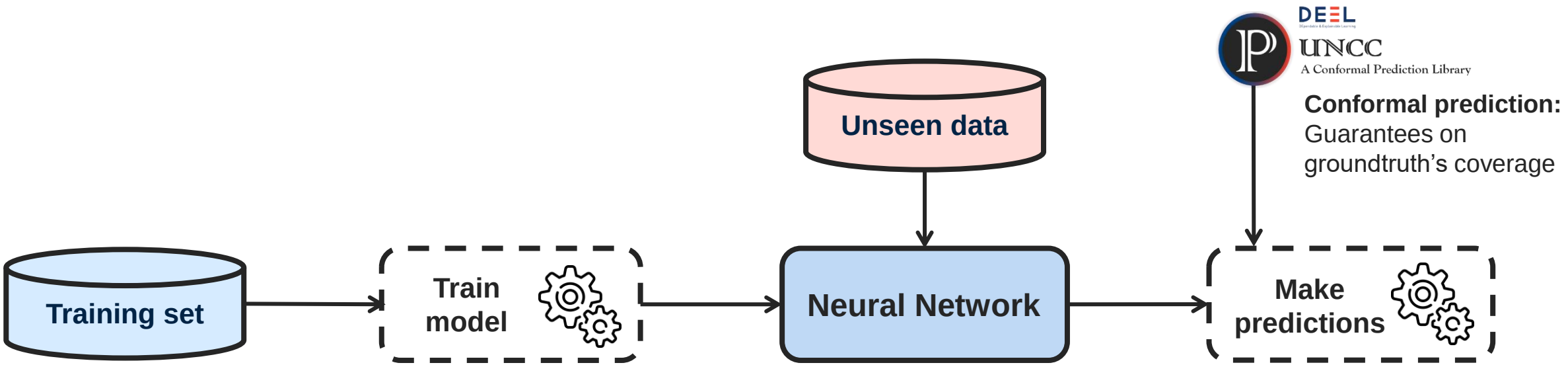
ANITI Université Fédérale Toulouse Midi-Pyrénées

Institut intelligence et données

ONERA THE FRENCH AEROSPACE LAB

AIRBUS

DEEL Libraries: Uncertainty quantification



DEEL
UNCC
A Conformal Prediction Library

Conformal prediction:
Guarantees on groundtruth's coverage

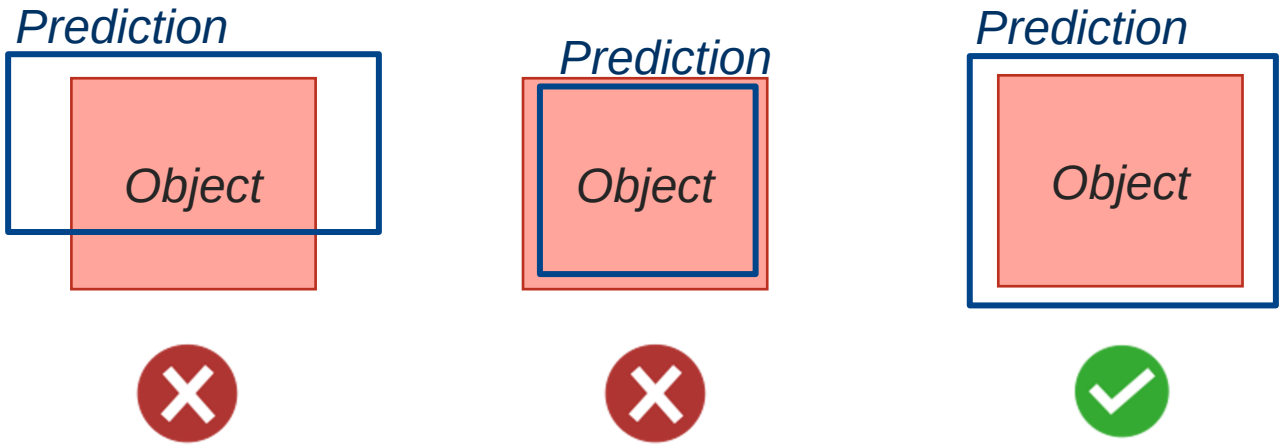
Conformal prediction

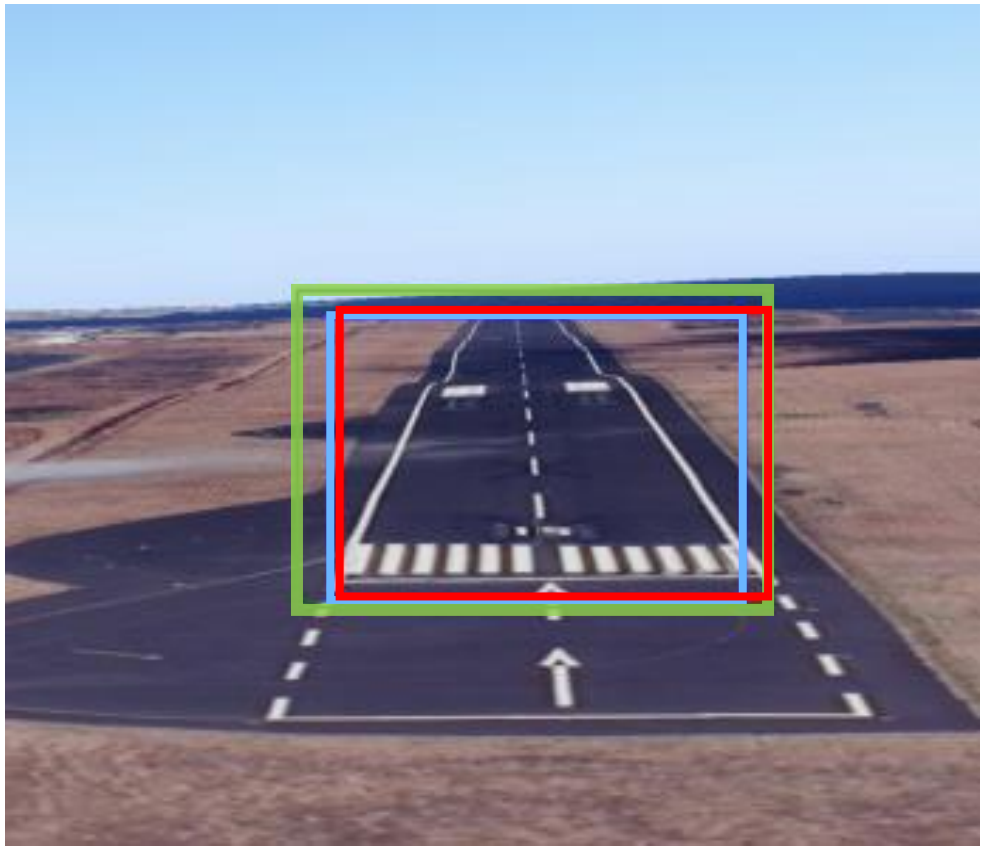
DEEL

PUNCC

A Conformal Prediction Library

- Objective { Provide a probabilistic guarantee that the **ground truth is fully enclosed** in the detection box





90% guarantee

Conformal prediction: α=0.1

DAAS_27_35_17, red: GT, blue: pred, green: conformal

EDDV_27R_35_17, red: GT, blue: pred, green: conformal

Lipschitz FCOS - Synthetic test set

DEEL

Dependable, Explainable & Embeddable Learning

Explaining decisions

IVADO

IRT SAINT EXUPÉRY

CRIAQ

ANITI Université Fédérale Toulouse Midi-Pyrénées

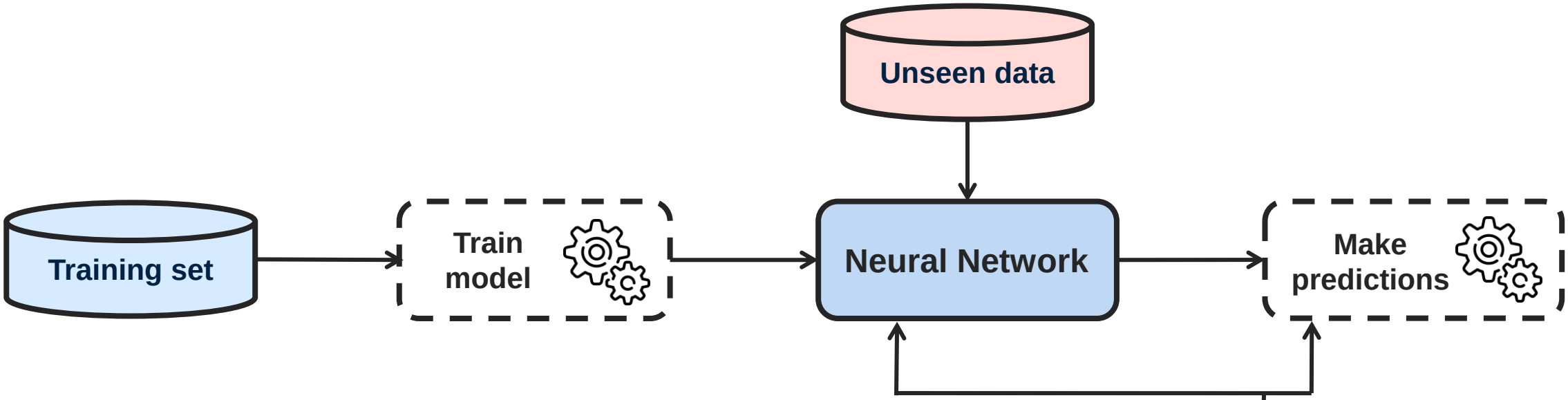
Institut intelligence et données

ONERA

THE FRENCH AEROSPACE LAB

AIRBUS

DEEL Libraries: The Big Picture

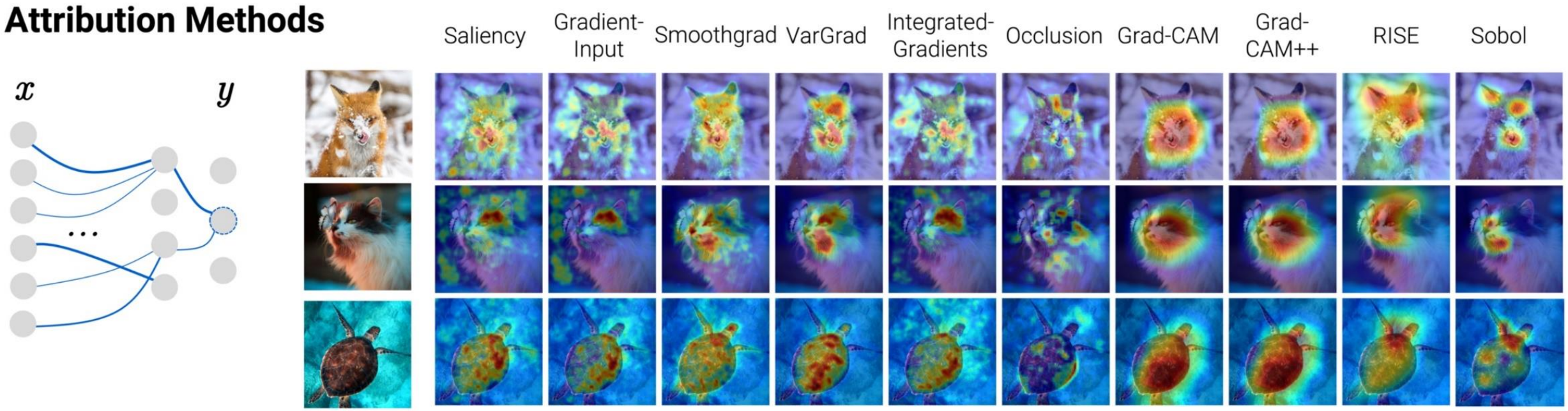


DEEL
XPLIQUE
Explainability Toolbox for Neural Networks

XAI: Provide explanations on predictions

XPlique: Attribution

– *'**Where** is the model looking?'*



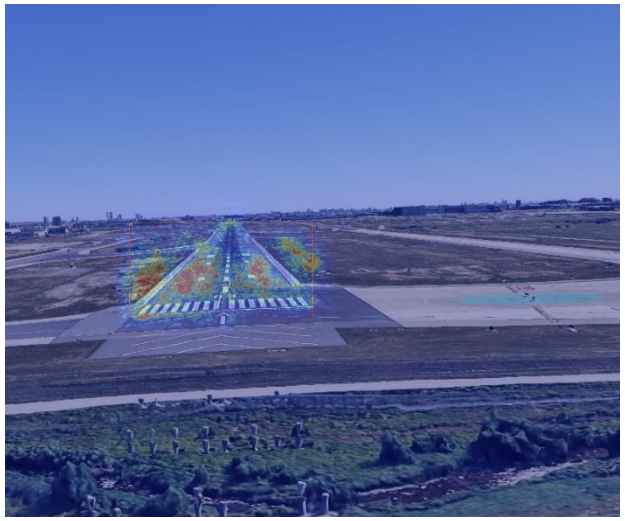
DEEL
XPLIQUE
Explainability Toolbox for Neural Networks

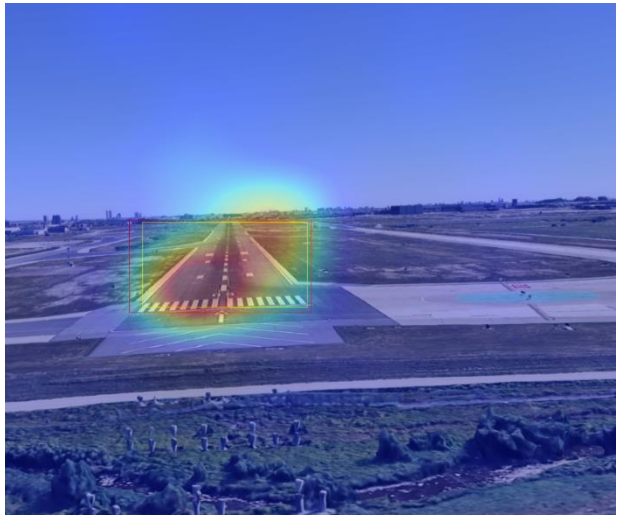
XPlique: Attributions

DEEL XPLIQUE Explainability Toolbox

DEEL Dependable, Explainable & Embeddable Learning

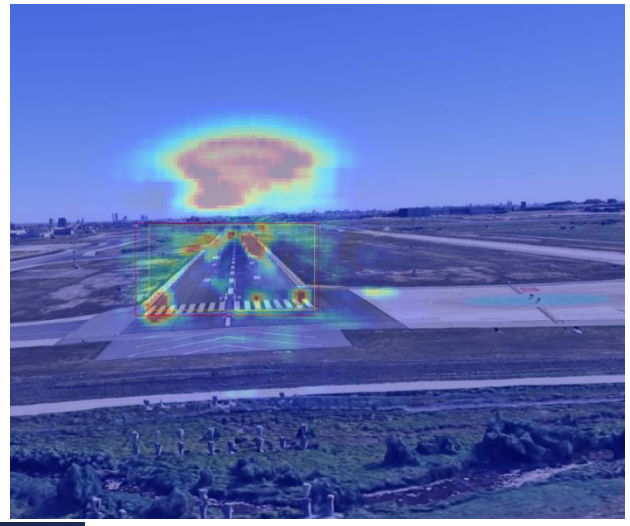
Saliency map

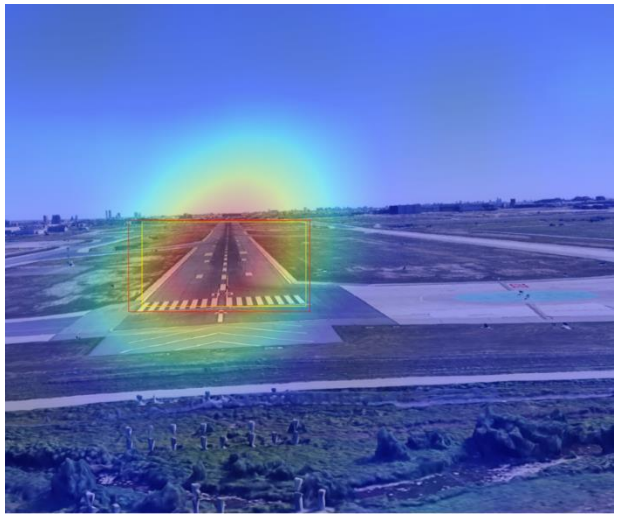




Rise

Occlusion

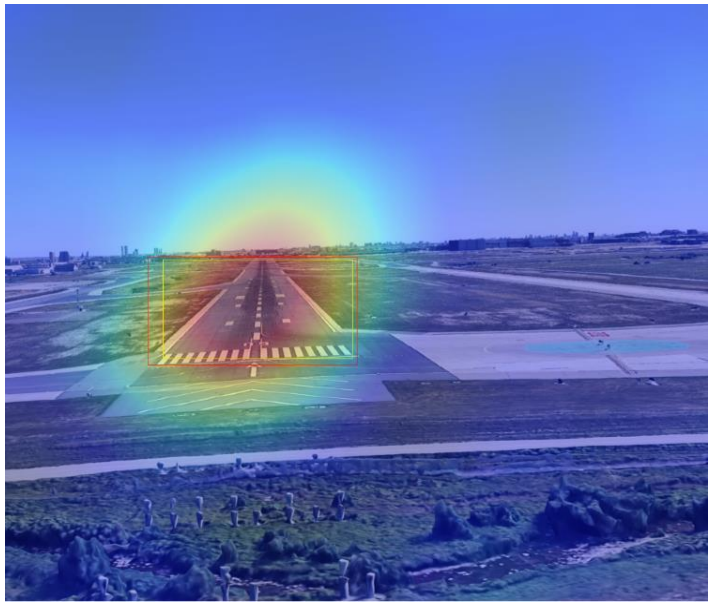




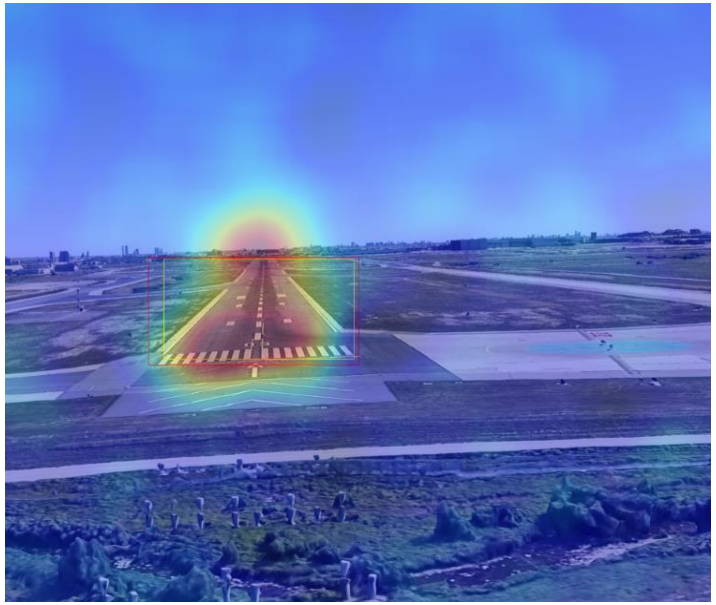
Sobol

Xplique: Attributions

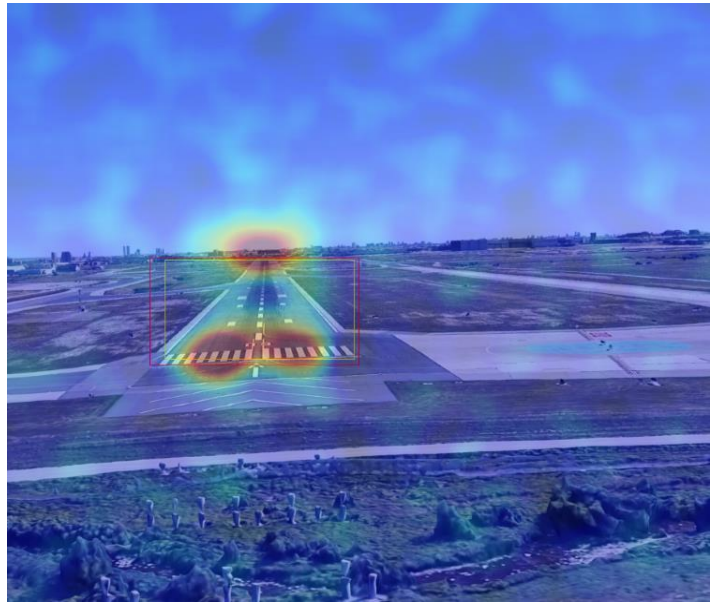
- Influence of method parameters
  - Grid size on Rise:



Rise : grid size 7 – 1K samples

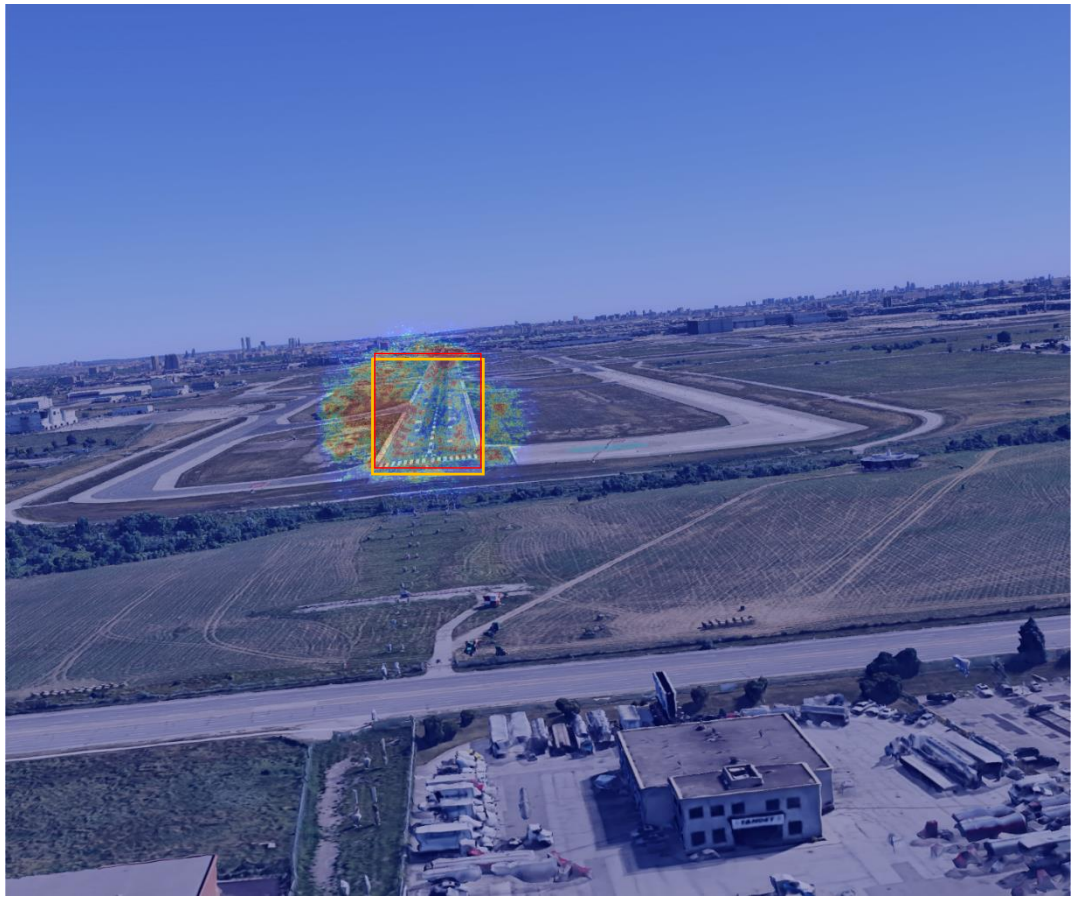


Rise : grid size 13 – 2K samples

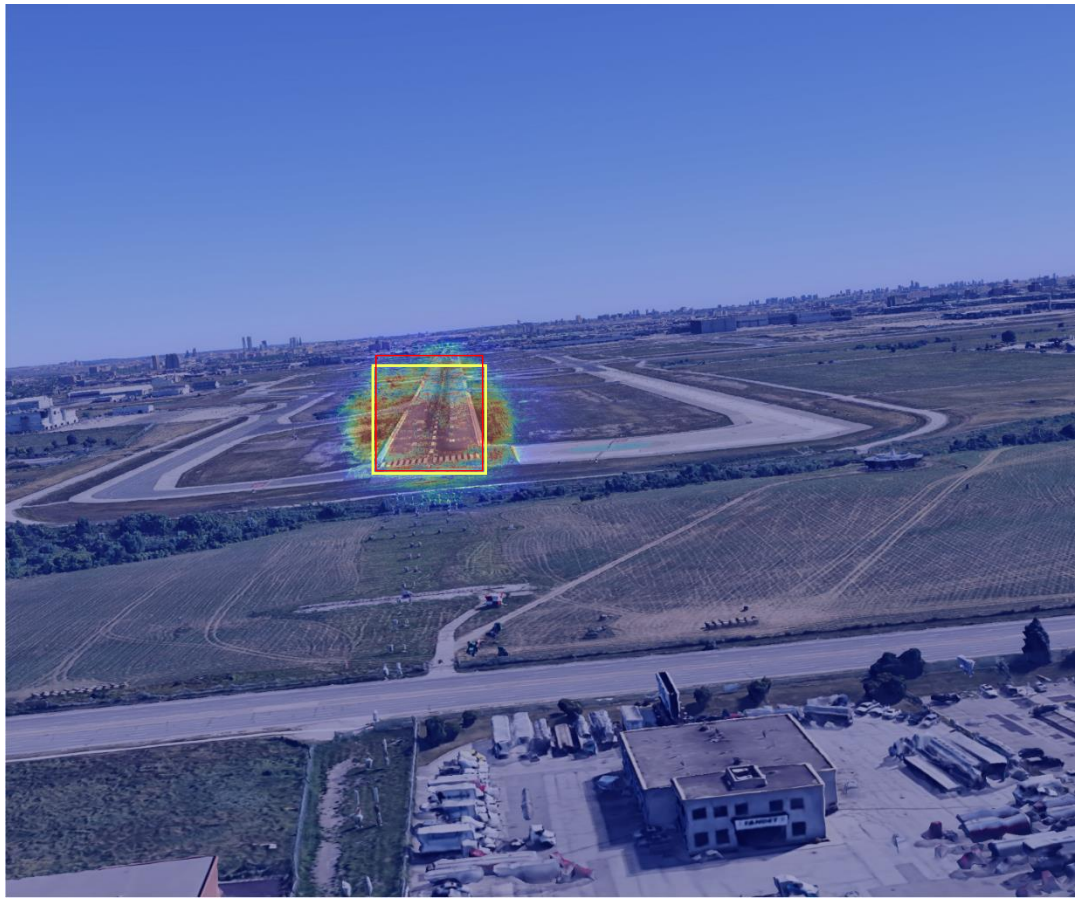


Rise : grid size 23 – 5K samples

Comparing explanations (TP)

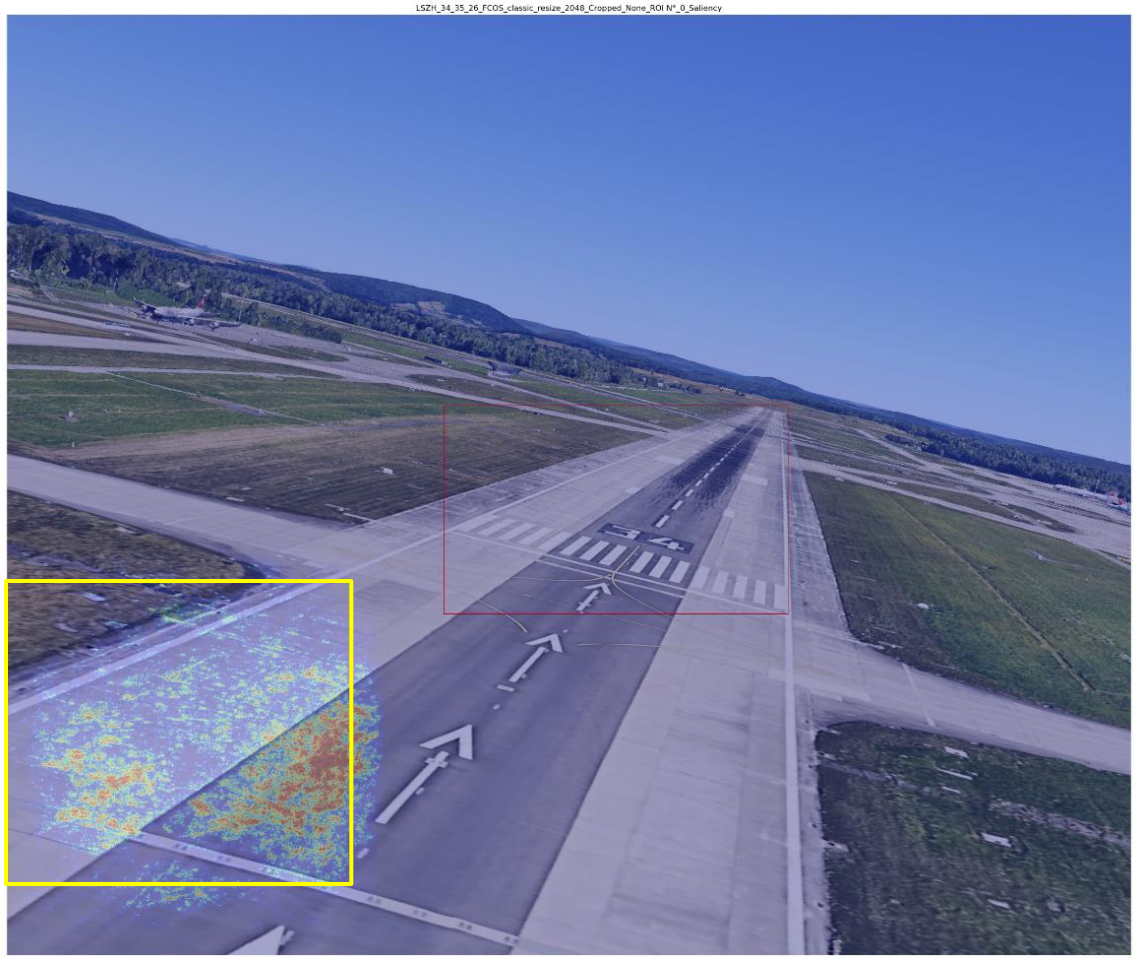


Saliency map for **Classic FCOS**

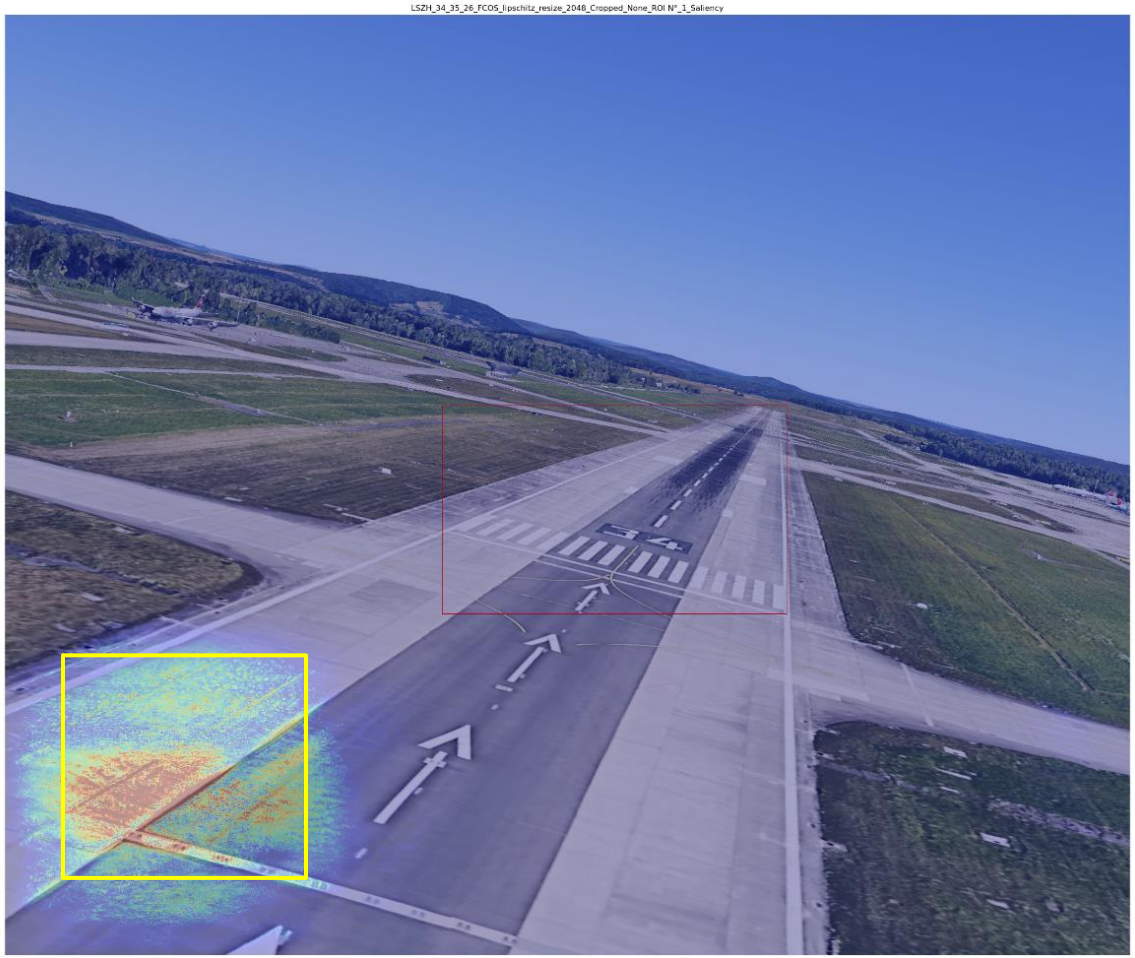


Saliency map for **Lipschitz FCOS**

Comparing explanations (FP)



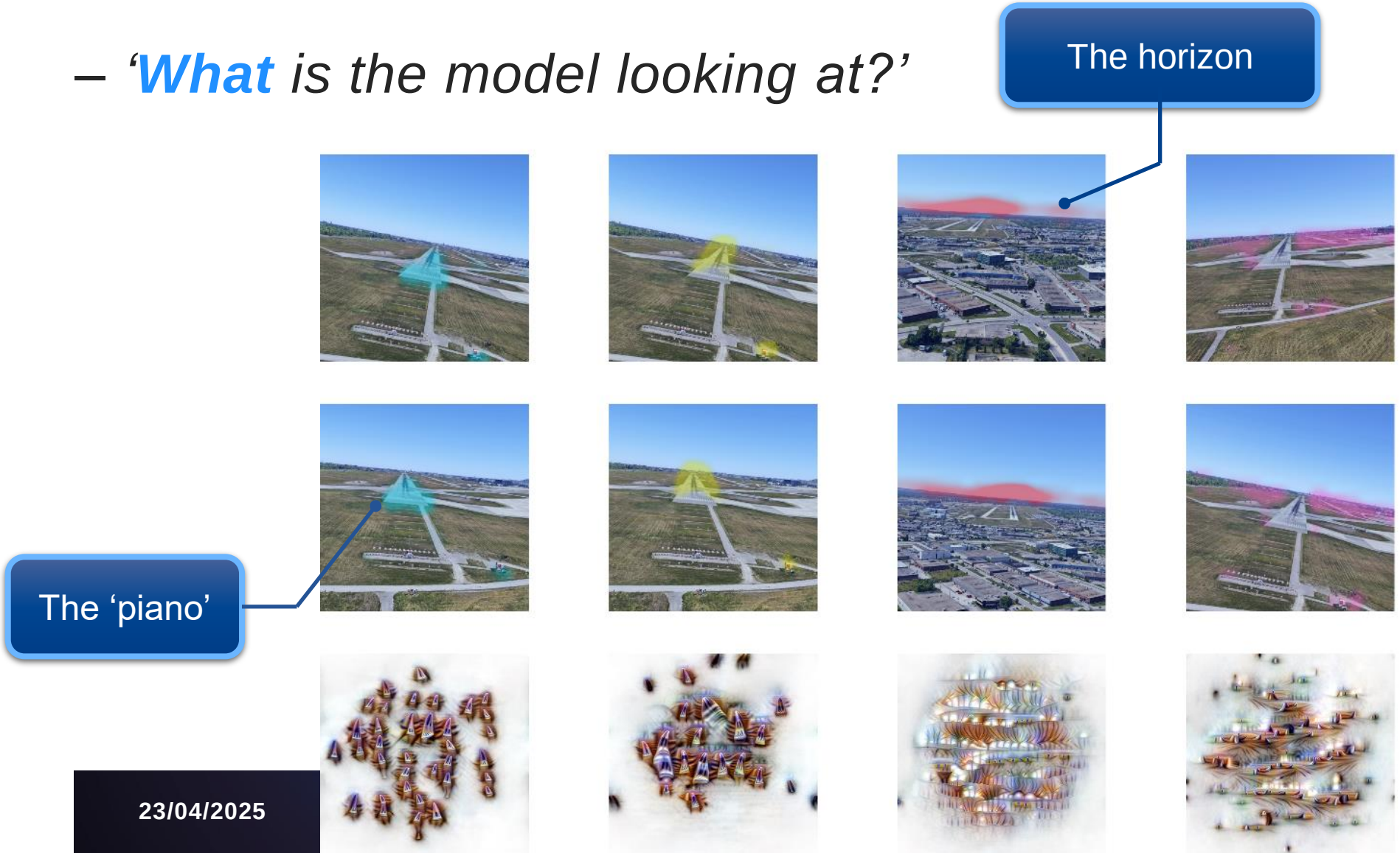
Saliency map for **Classic FCOS**



Saliency map for **Lipschitz FCOS**

LARD: Concepts

– *'**What** is the model looking at?'*

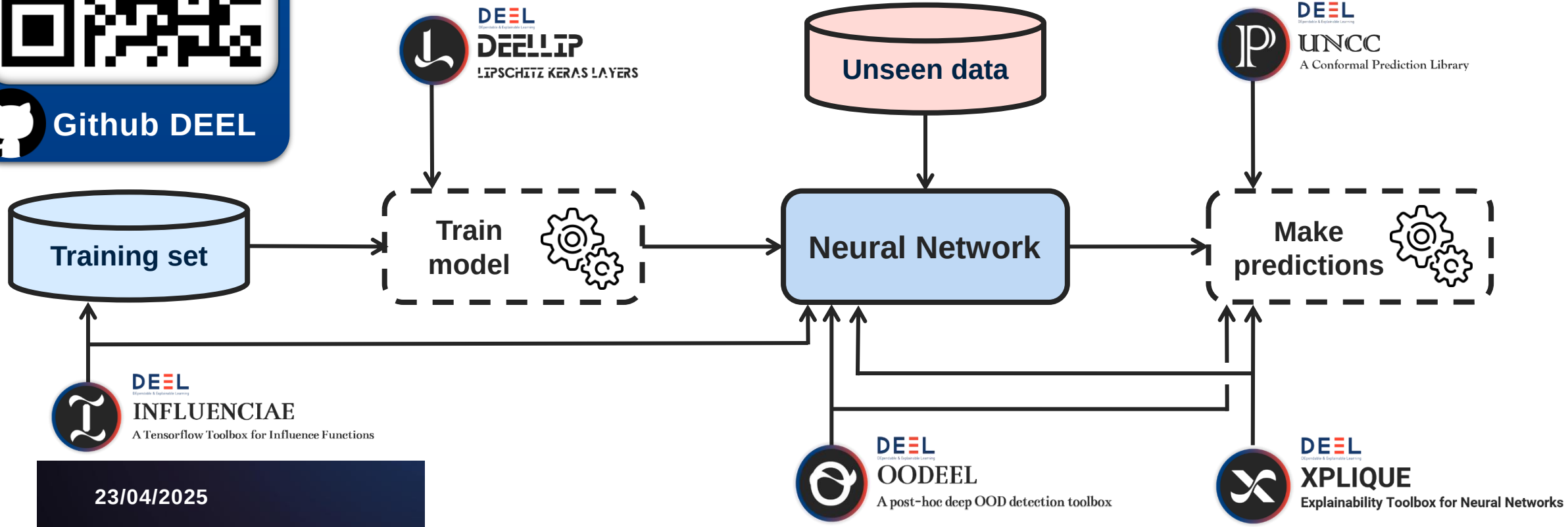


Thank you for your attention!

– Any questions?

DEEL
Dependable, Explainable & Embeddable Learning

Github DEEL



23/04/2025