TIA: Test, Improve, Assure Deep Neural Networks for Space

Fabrizio Pastore, Mohammed Attaoui
University of Luxembourg
fabrizio.pastore@uni.lu

ADCSS

October 23rd, 2024

Deep Neural Networks are becoming key components

Solar ejection classification

Space landscape segmentation

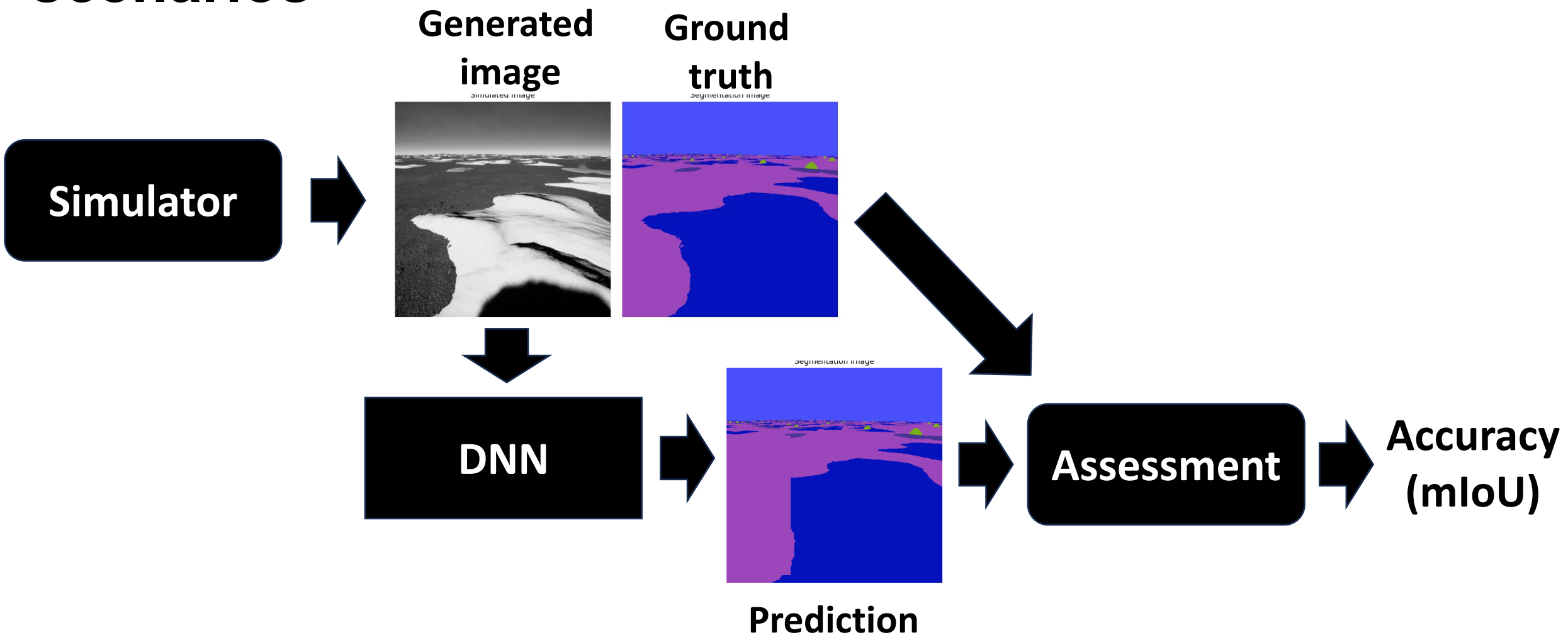
But we still have lot of doubts after traditional DNN training and testing process

- Did we miss testing any scenario where the DNN underperforms?
- How to improve the DNN at limited cost?
- Why does the DNN fail?

- **Did we miss testing any scenario where the DNN underperforms?**

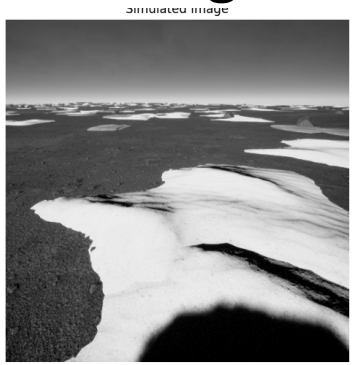
Solution: Test the DNN with additional (underperforming) scenarios

Test the DNN with additional (underperforming) scenarios



Test the DNN with additional (underperforming) scenarios

Generated image



VS

Real images

Unrealistic

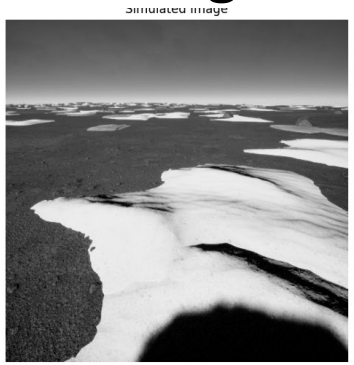
Out of distribution

DNN assessment is unreliable

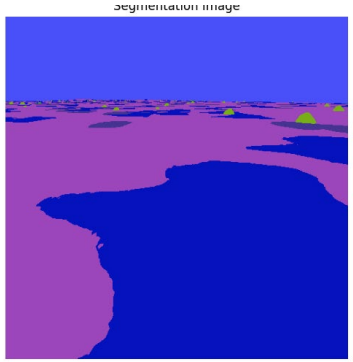
Test the DNN with additional (underperforming) scenarios

Simulator

Generated image



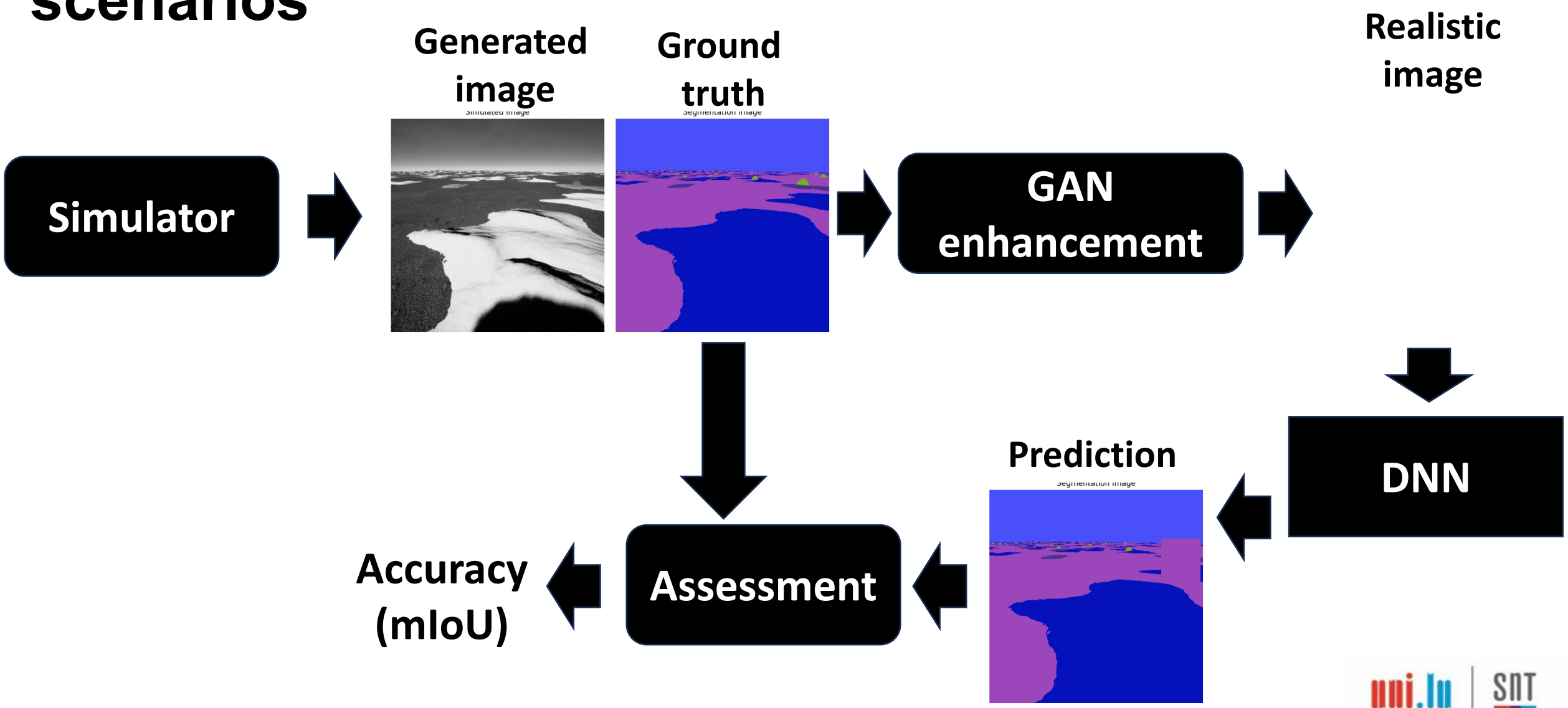
Ground truth



GAN enhancement

Trained with segmentations for real images

Test the DNN with additional (underperforming) scenarios

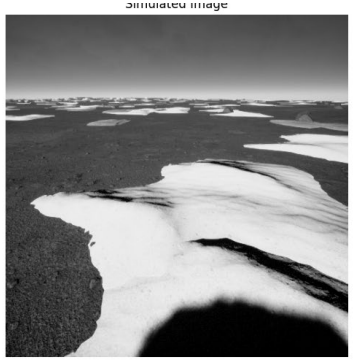


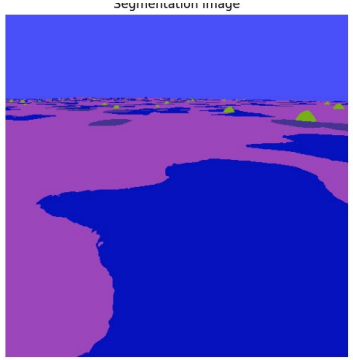
Test the DNN with additional (underperforming) scenarios

Generated image

Ground truth

Realistic image





Simulator

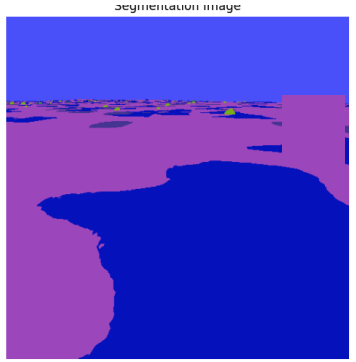
GAN enhancement

X=?
Y=?
cameraX=?

Parameters values?

DNN

Prediction



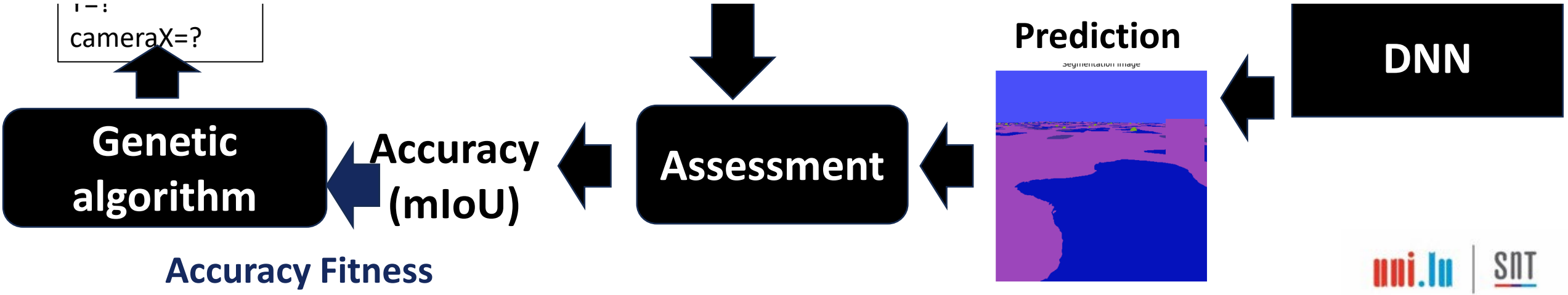
Assessment

Accuracy (mIoU)

Test the DNN with additional (underperforming) scenarios

Realistic

- Genetic algorithms evolve a population of individuals (here, simulator parameters) that minimise a given fitness
- In this case, they drive the generation of images that lead to worse prediction



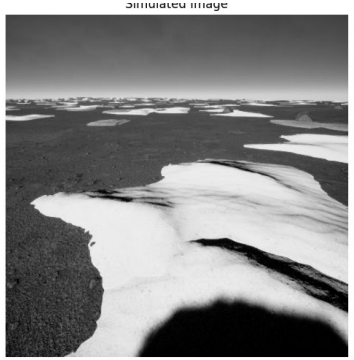
Test the DNN with additional (underperforming) scenarios

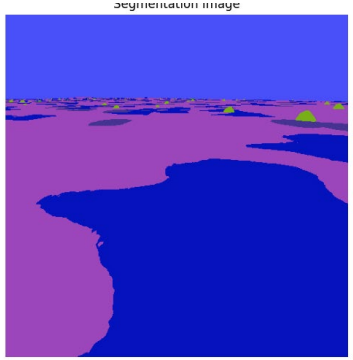
Generated image

Ground truth

Realistic image

Simulator

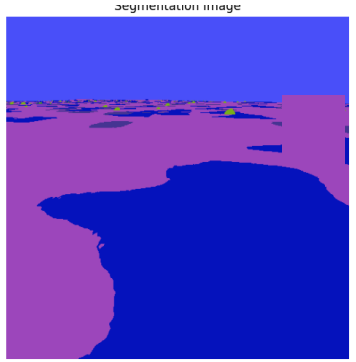




GAN enhancement

DNN

Prediction



Assessment

Accuracy (mIoU)

Genetic algorithm

X=?
Y=?
cameraX=?

Accuracy Fitness

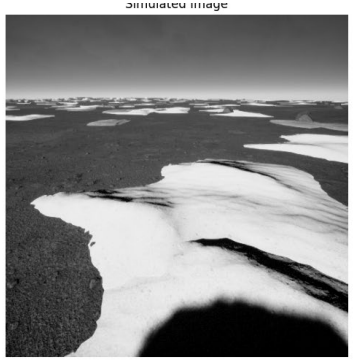
Test the DNN with additional (underperforming) scenarios

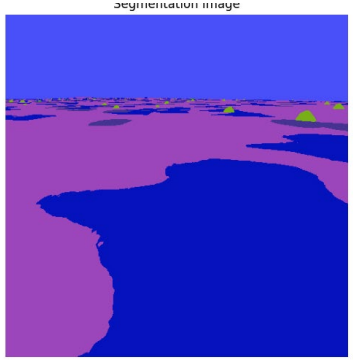
Generated image

Ground truth

Realistic image

Simulator





GAN enhancement

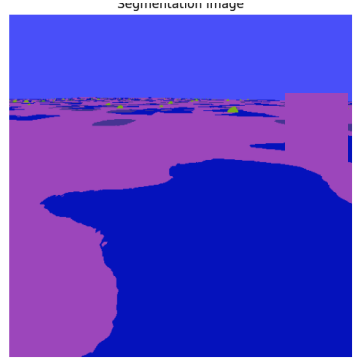
X=?
Y=?
cameraX=?

Generated images

Similarity Fitness

DNN

Prediction



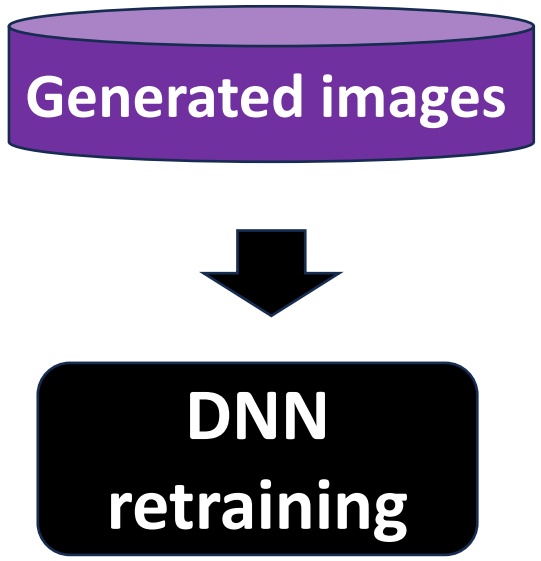
Assessment

Accuracy (mIoU)

Genetic algorithm

Accuracy Fitness

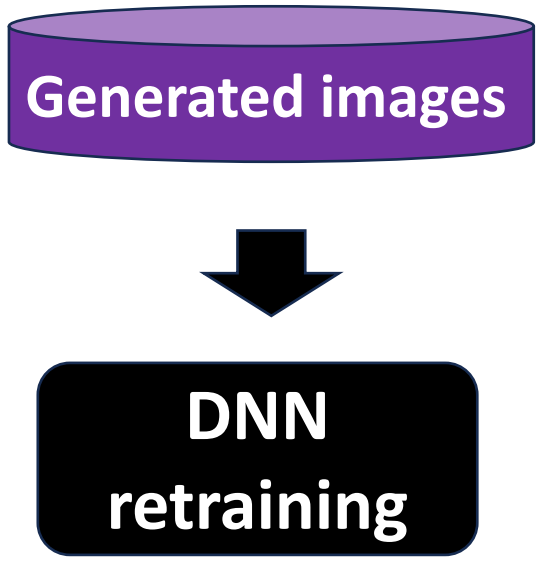
Test the DNN with additional (underperforming) scenarios



Diverse

Worst results

Improve the DNN at limited cost



Diverse

Worst results

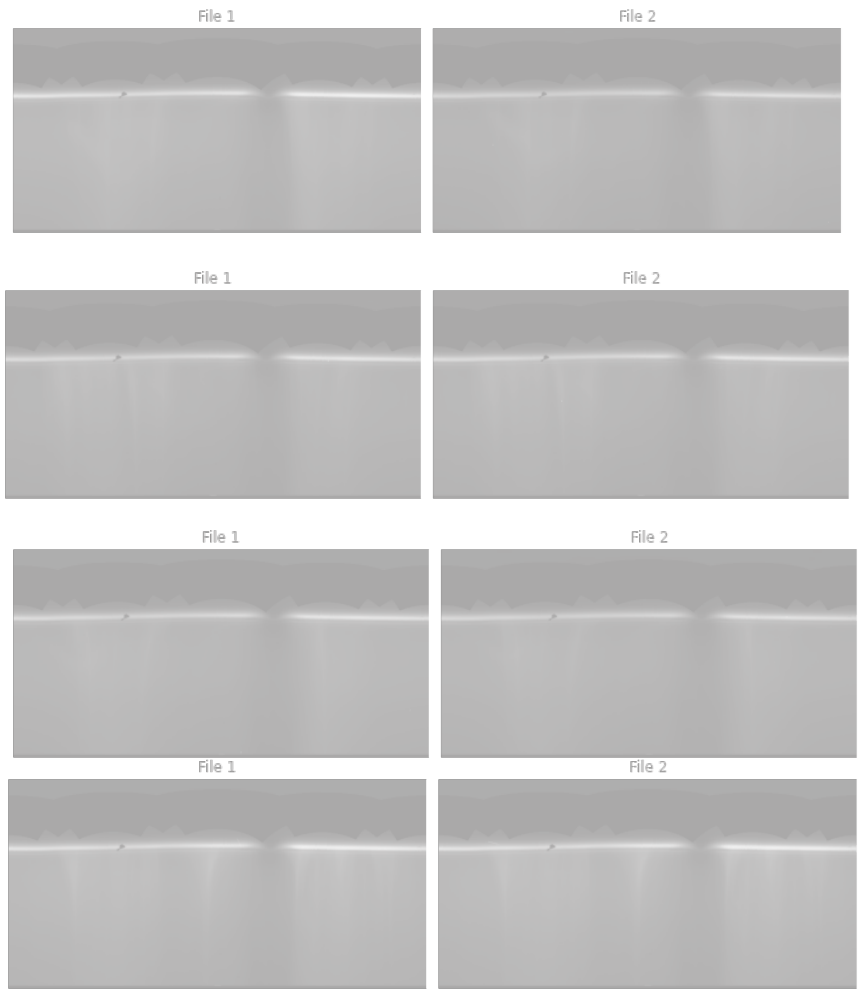
Retraining of DeepLabV3 for AI4Mars

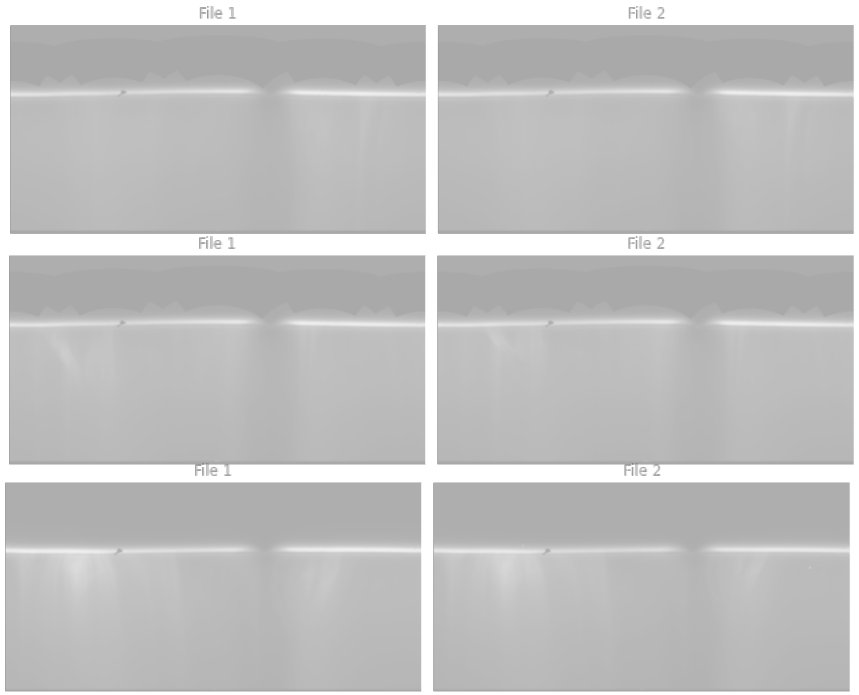
Retraining Set	Median	5th Percentile	1st Quartile	3rd Quartile	Average
Original DeeplabV3	0.47	0.17	0.34	0.59	0.48

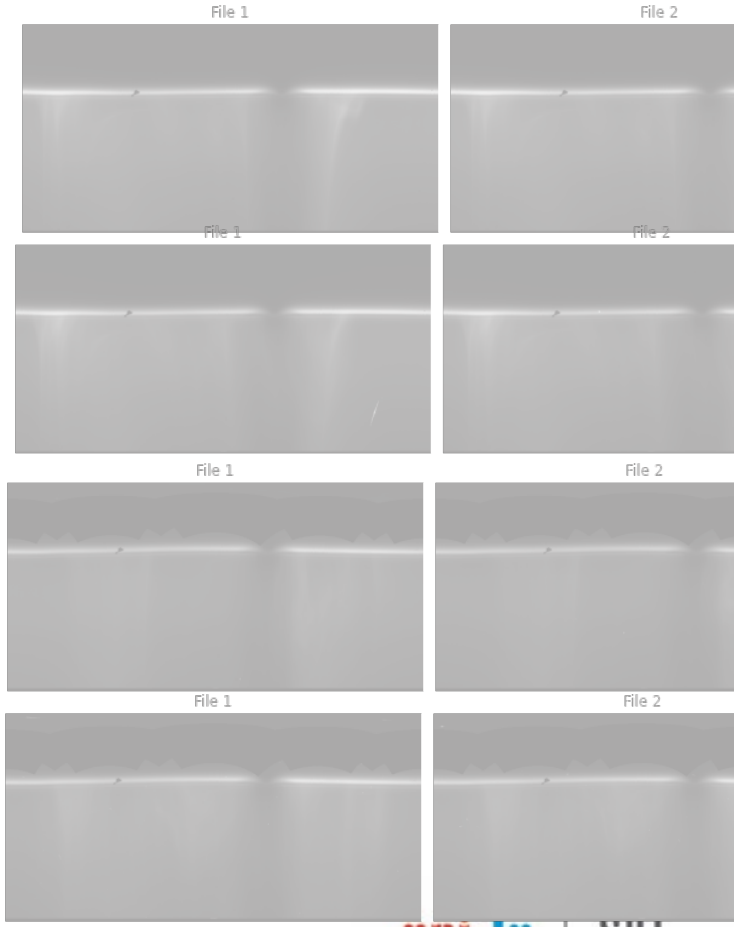
- ~~Did we miss testing any scenario where the DNN underperforms?~~
- ~~How to improve the DNN at limited cost?~~
- Why does the DNN fail?

Why does the DNN fail?

Hard to determine similarities manually





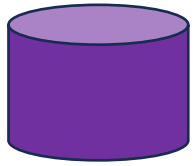


Solar Ejection Detection DNN

To group like a human use a DNN

Failure analysis pipeline

Failure-inducing images



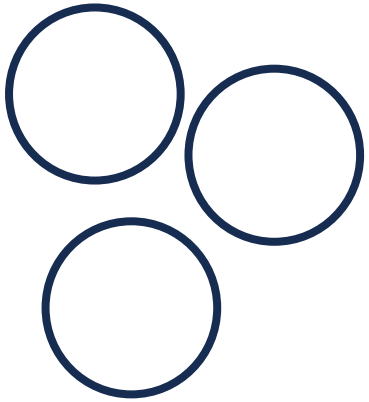
Prepro-cessing

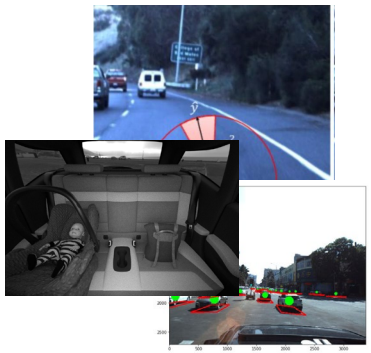
Feature Extraction (VGG 16)

Dimension Reduction (UMAP)

Clustering (DBSCAN)

Clusters with similar images





VGG 16 VS
ResNet VS
InceptionV3 VS
Xception VS
HUDD VS
LRP VS
AE VS

PCA
VS
UMAP
VS
None

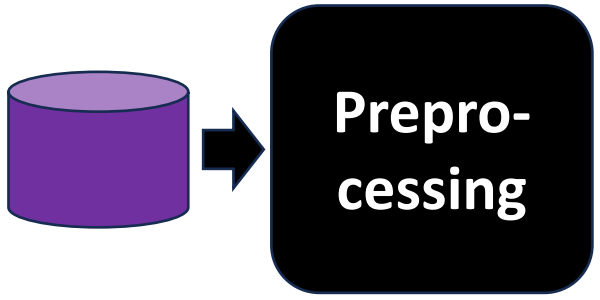
DBSCAN
VS
HDBSCAN
VS
KMeans

Visually inspect to determine failure scenarios

M. Attaoui, H. Fahmy, F. Pastore, and L. Briand. 2023. Black-box Safety Analysis and Retraining of DNNs based on Feature Extraction and Clustering. ACM Trans. Softw. Eng. Methodol.

Solar ejection preprocessing

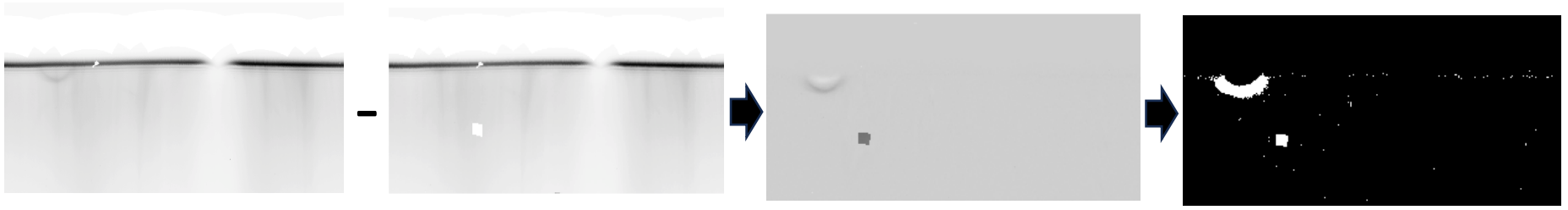
Failure-inducing images



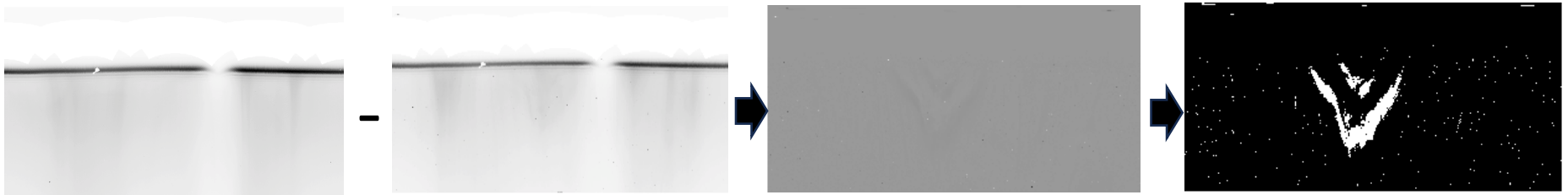
Difference

Normalize: divide by the most common pixel

Input 1

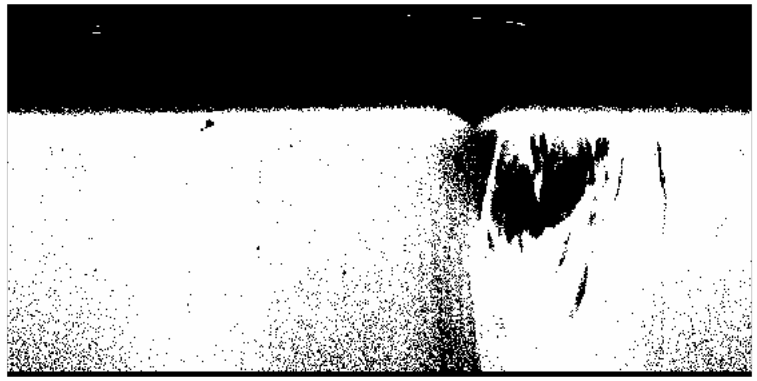


Input 2



Results with solar ejection

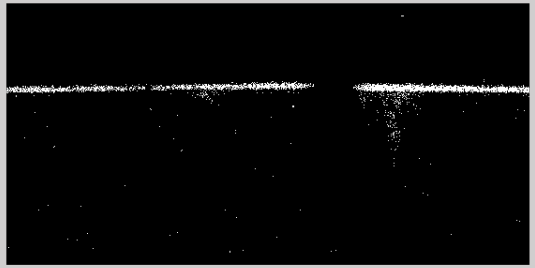
Images leading to wrong classifications



Interpretation: too much noise

Interpretation: no ejection

Images leading to correct classifications



Class: no ejection

Class: no ejection

Class: ejection

TIA: Test, Improve, Assure Deep Neural Networks for Space

Papers

fabrizio.pastore@uni.lu

SnT

UNIVERSITY OF LUXEMBOURG

TIA: Test, Improve, Assure Deep Neural Networks for Space

Fabrizio Pastore
University of Luxembourg
fabrizio.pastore@uni.lu