Istituto Nazionale di Fisica Nucleare

MoEDAL NTD
Data Augmentation
& Automatic pit detection

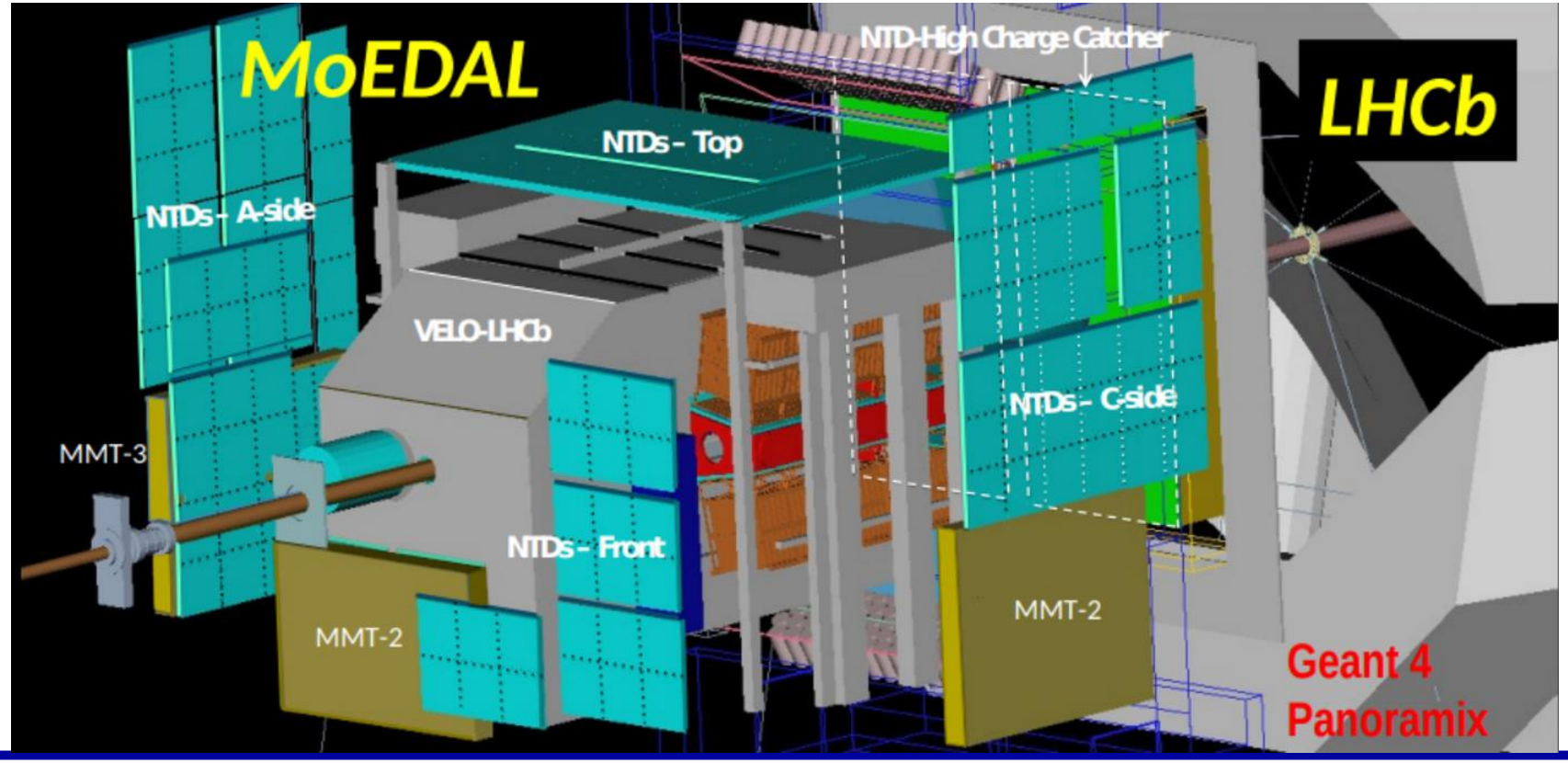
Giuseppe Levi
INFN BOLOGNA

MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

MoEDAL is a CERN experiment (LHC-IP8) dedicated to the search for highly ionizing exotic particles such as monopoles.

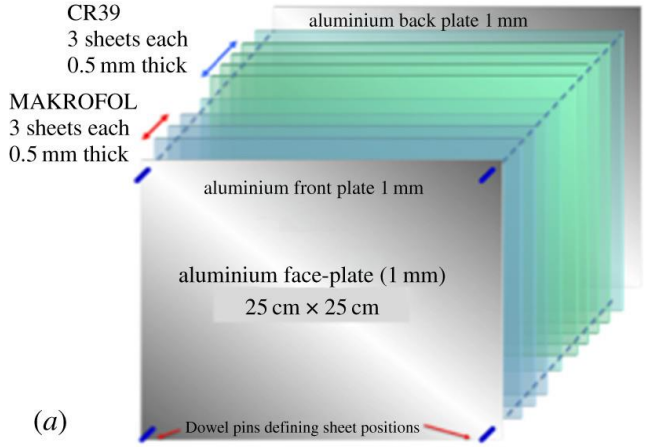
it consists mainly of passive detectors.

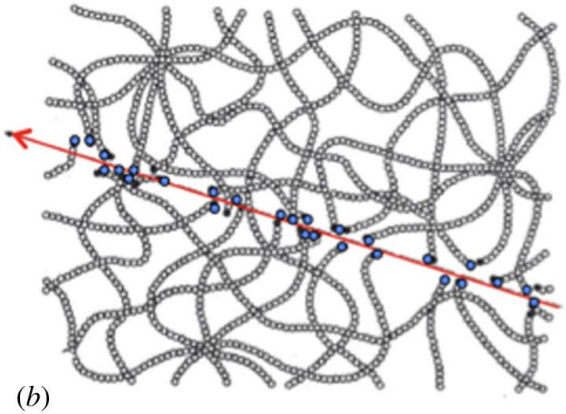


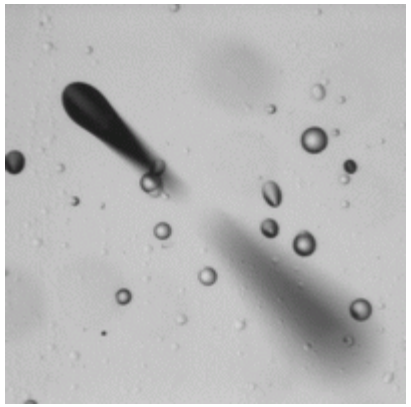
MoEDAL NTD Data Augmentation

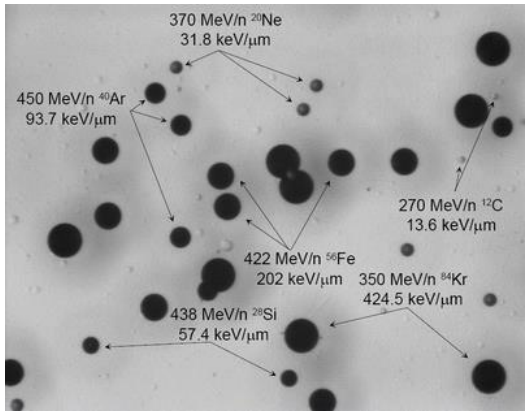
Istituto Nazionale di Fisica Nucleare

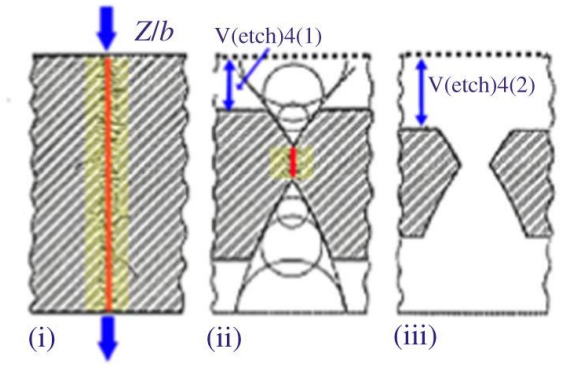
Nuclear Track Detectors are passive plastic sheets. The tracks are developed by a chemical attack and revealed through a microscope.

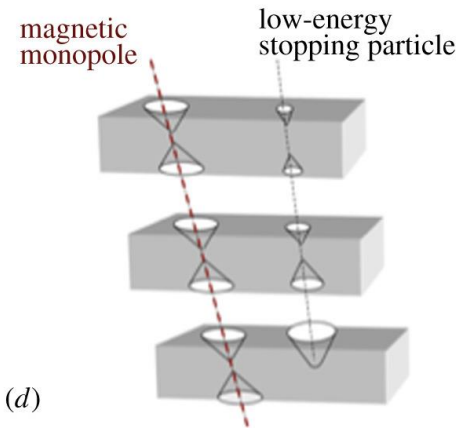


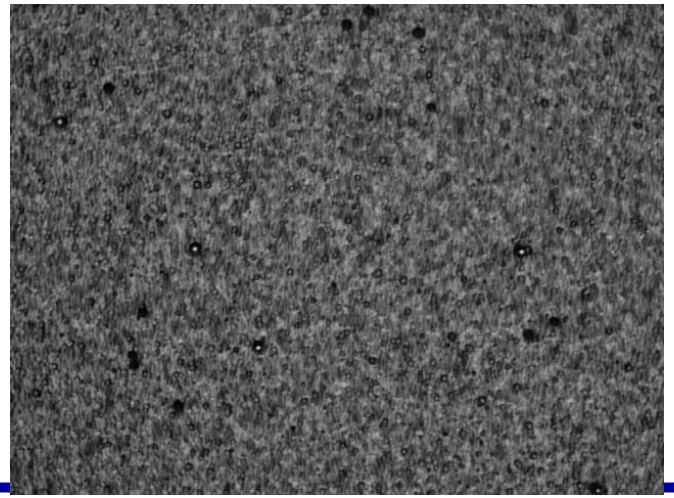












Istituto Nazionale di Fisica Nucleare

MoEDAL NTD Data Augmentation

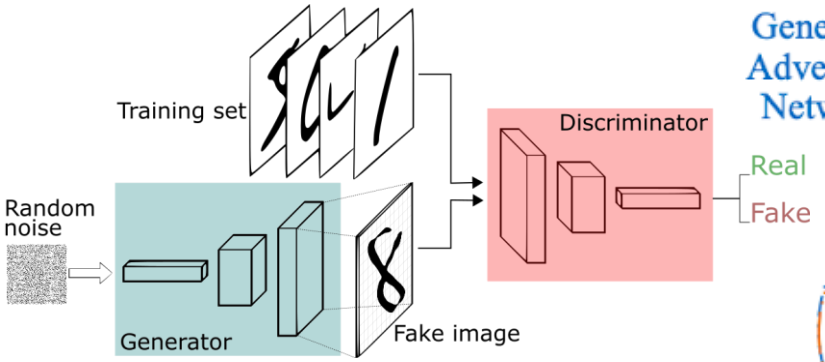
We have an enormous amount of images (frames) to analyze for NTD.
Scanning 20m^2 of detector with a 2x2 mm FOV means that we have about 5M frames for each side. (8.8 M with a 1.5mm step)
If we want to make this process automatic we should divide it into steps e.g.:

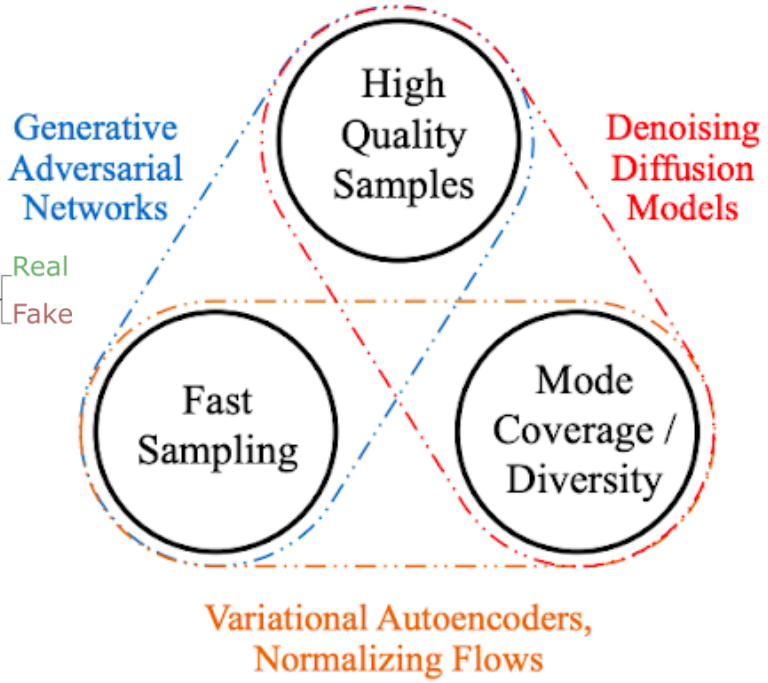
DAQ (exposure + etching) → Bologna Microscope → Data storage
Pit identification → to select frames with tracks
Measure → (Calibration + Data)

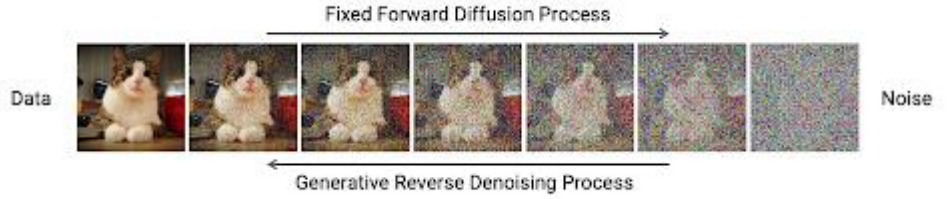
AI can be used along with CV for Pit Identification but a large number of labeled images tipically in the order of from 1000 to 10000 for each CLASS.

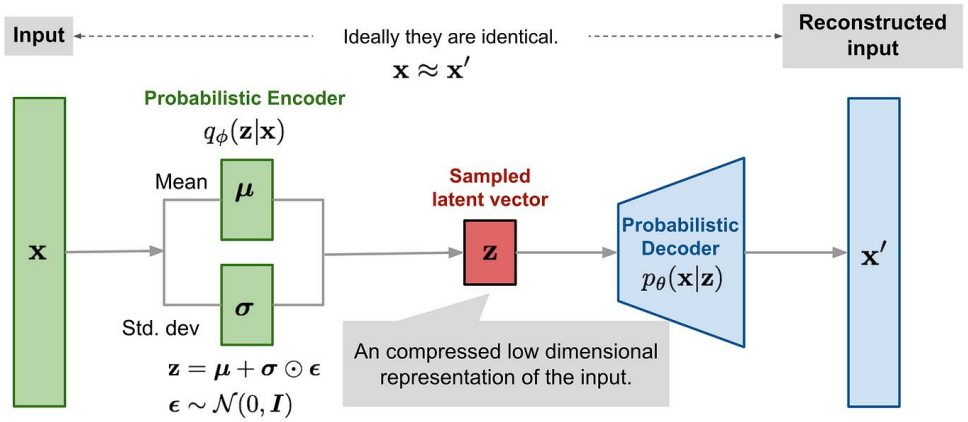
MoEDAL NTD Data Augmentation

Apart from the convenctional data augmentation techniques (rotation,mirroring) there is now the possibility to GENERATE «syntetic data» from a small sample.







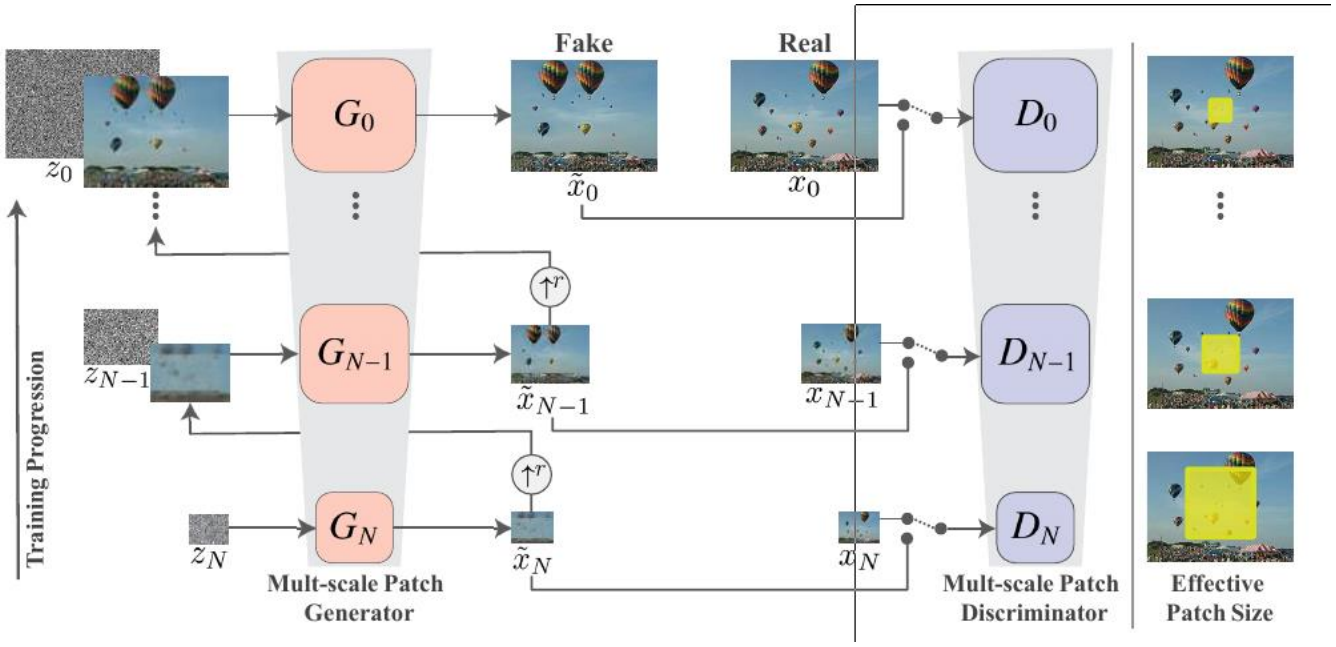


MoEDAL NTD Data Augmentation

GAN is well tested technology and are simpler to implement and train than diffusion models.

Normally a GAN would need a large number of images to be trained BUT in 2019 a new technique has been introduced to train a GAN starting from a single image. (Rott et al 2019)

The main point is to create a pyramid of GAN networks sampling the image at different scales.



MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

SinGAN: Learning a Generative Model from a Single Natural Image

Tamar Rott Shaham
Technion

Tali Dekel
Google Research

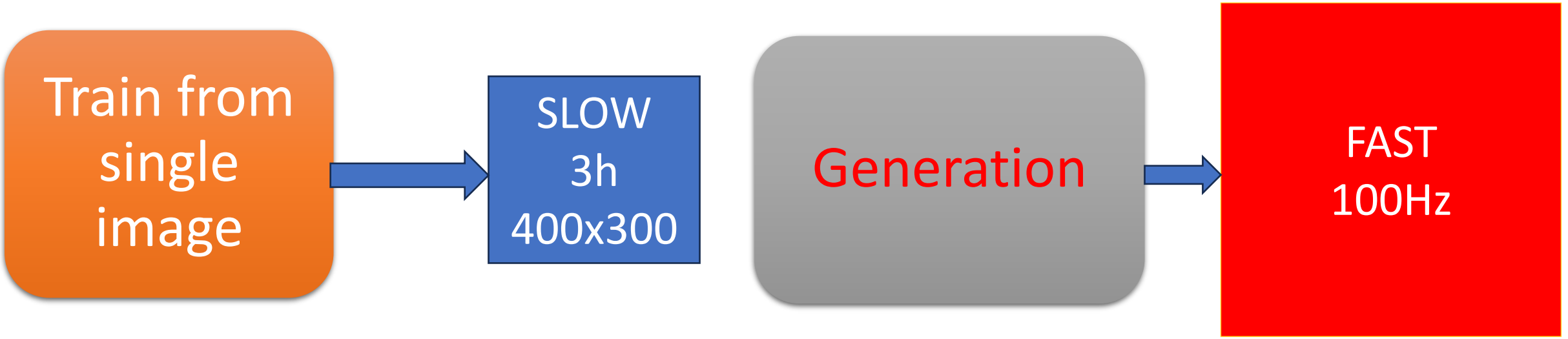
Tomer Michaeli
Technion

***Single* training image**

Random samples from a *single* image

MoEDAL NTD Data Augmentation

Limitation on image size is due to memory. Full use of GPU (up to 86/80 W)
(Now on CNAF HPC machine we can go up to 1000x1000 px.)

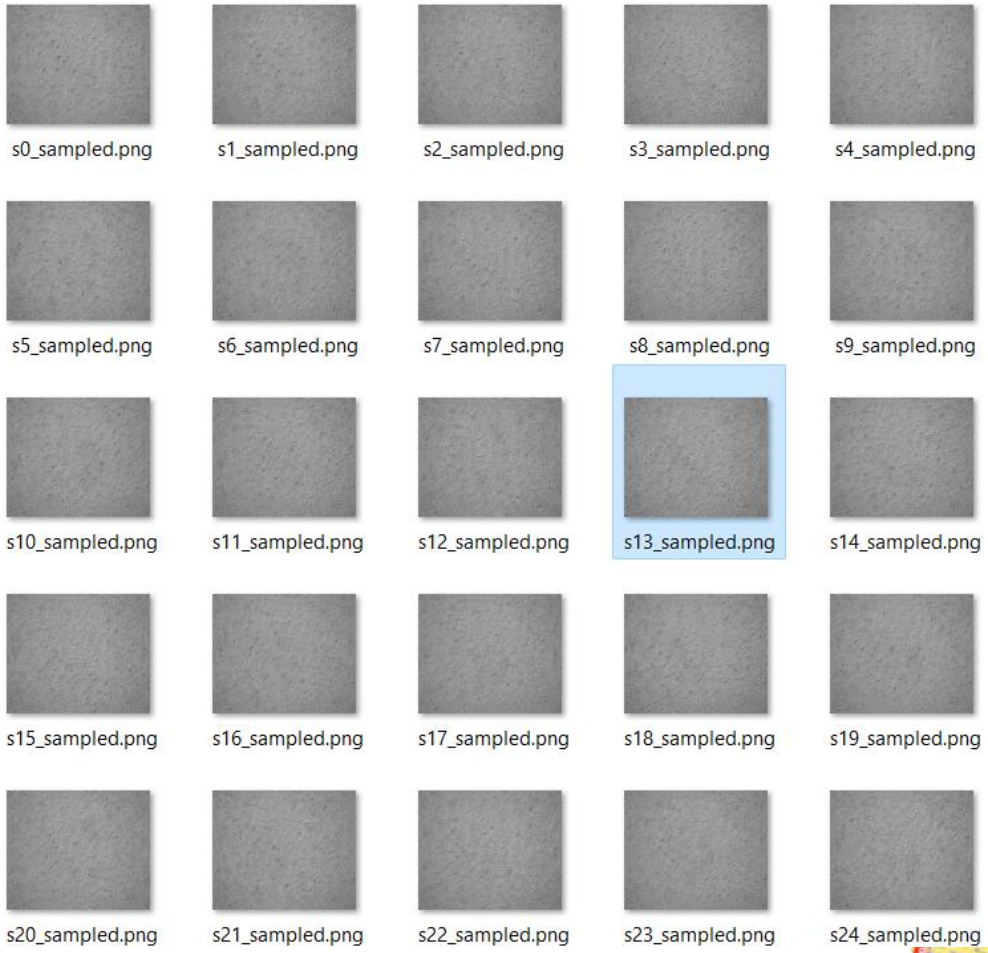


MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

From one backgroud:

MANY

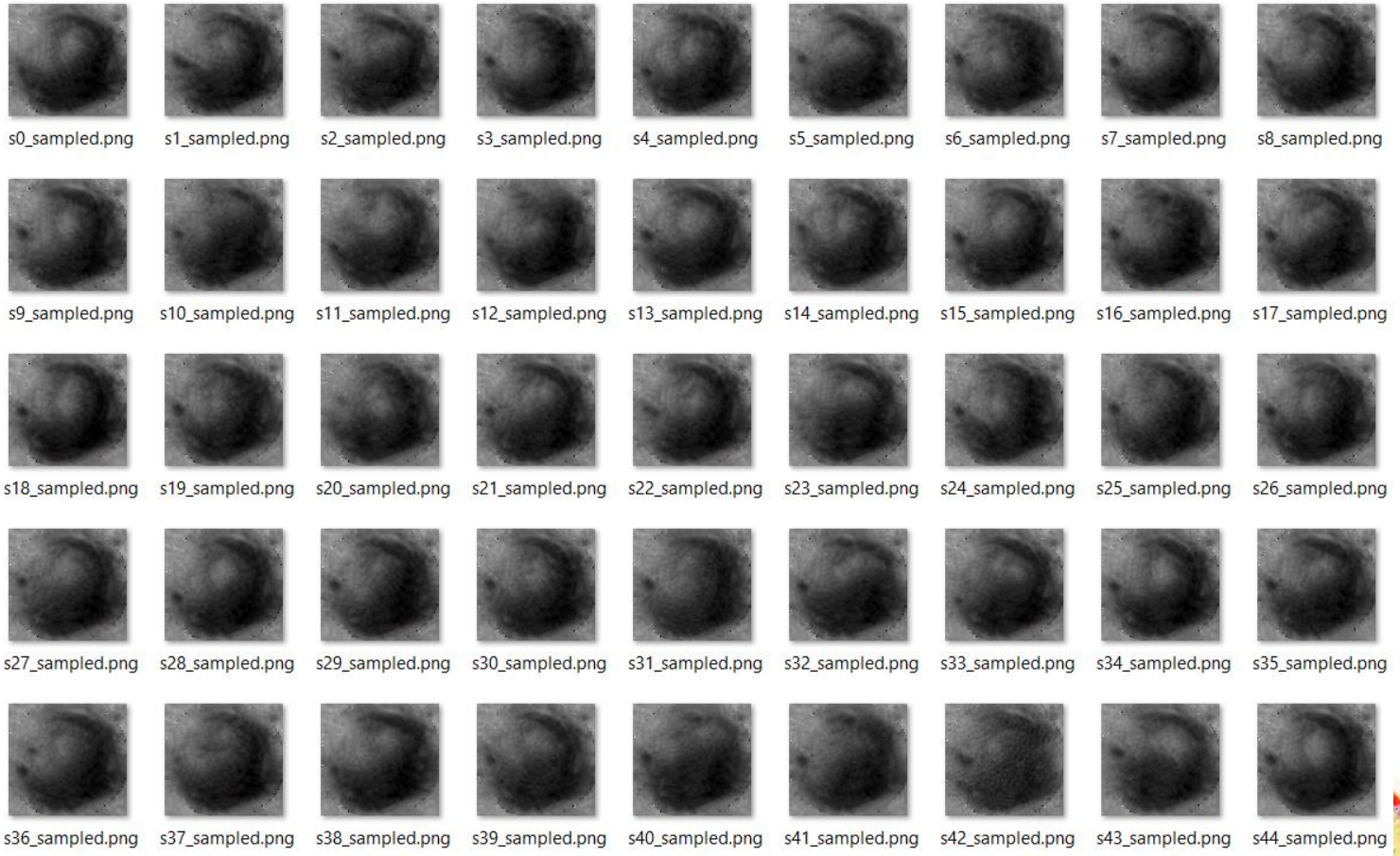


MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

From few pits:

MANY

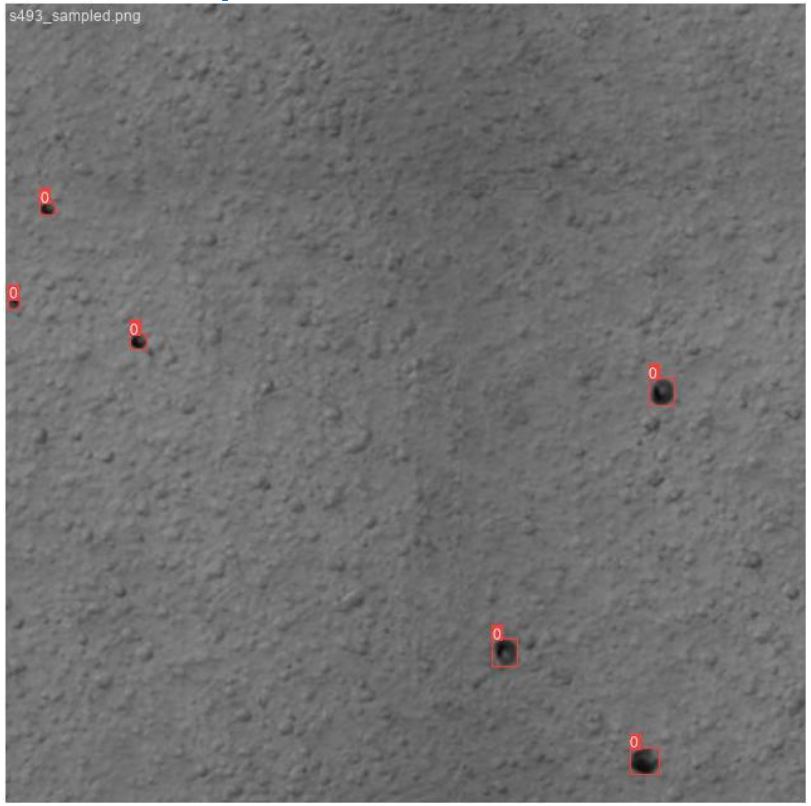


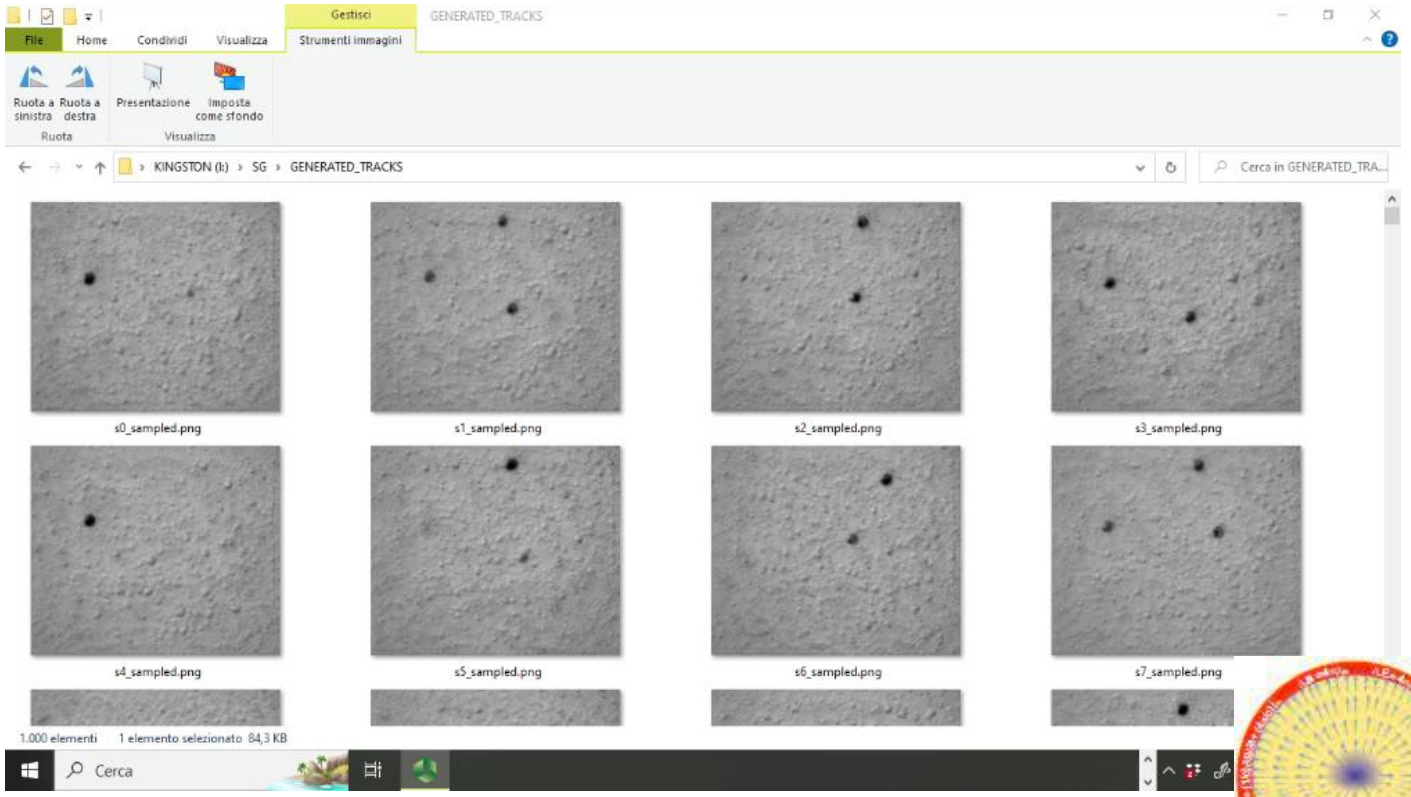
MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

We can then combine backgrounds and pits to obtain many images AND generate also LABELS.

In the process data are even more augmented with scaling and reflections.

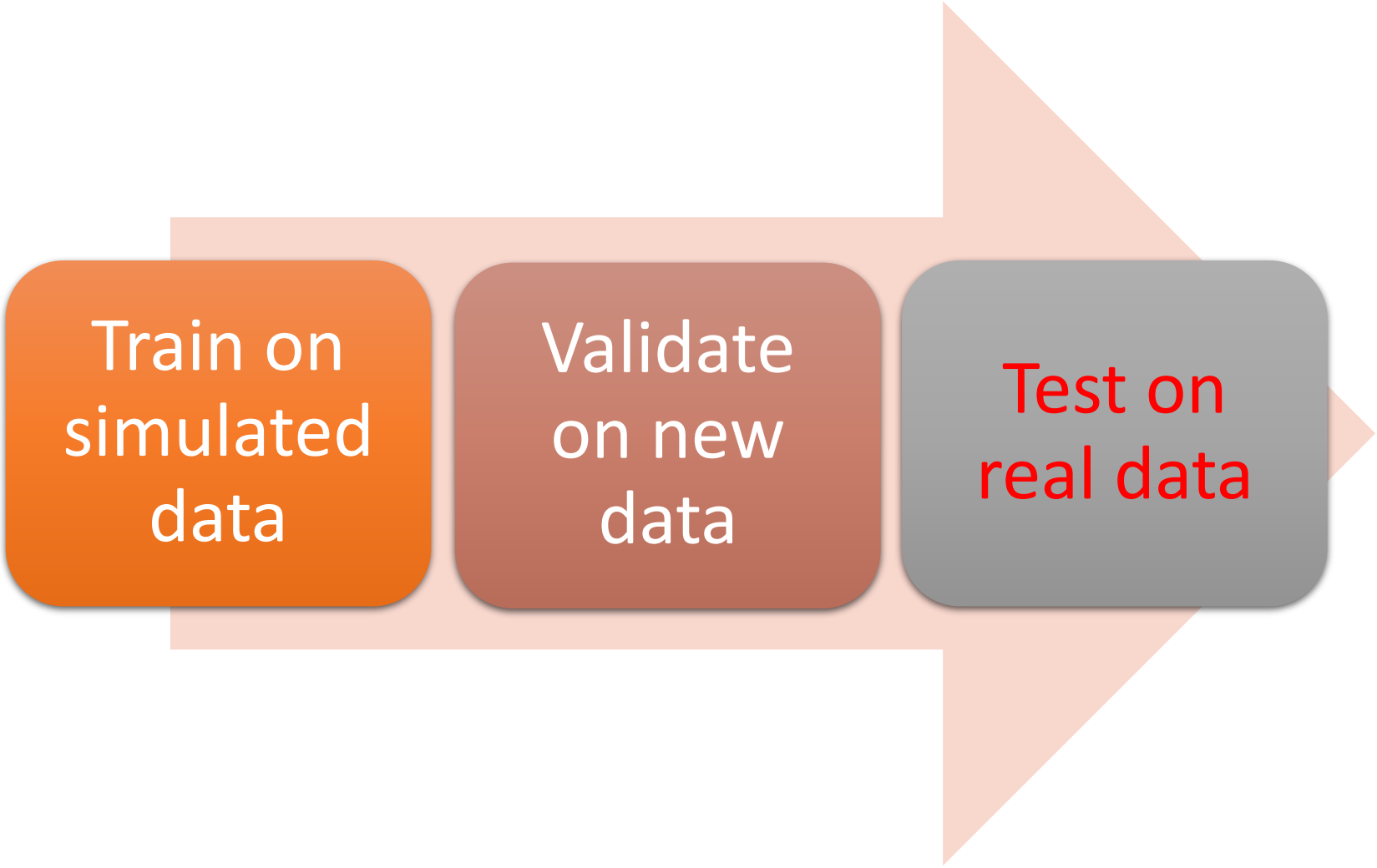




MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

We have to:

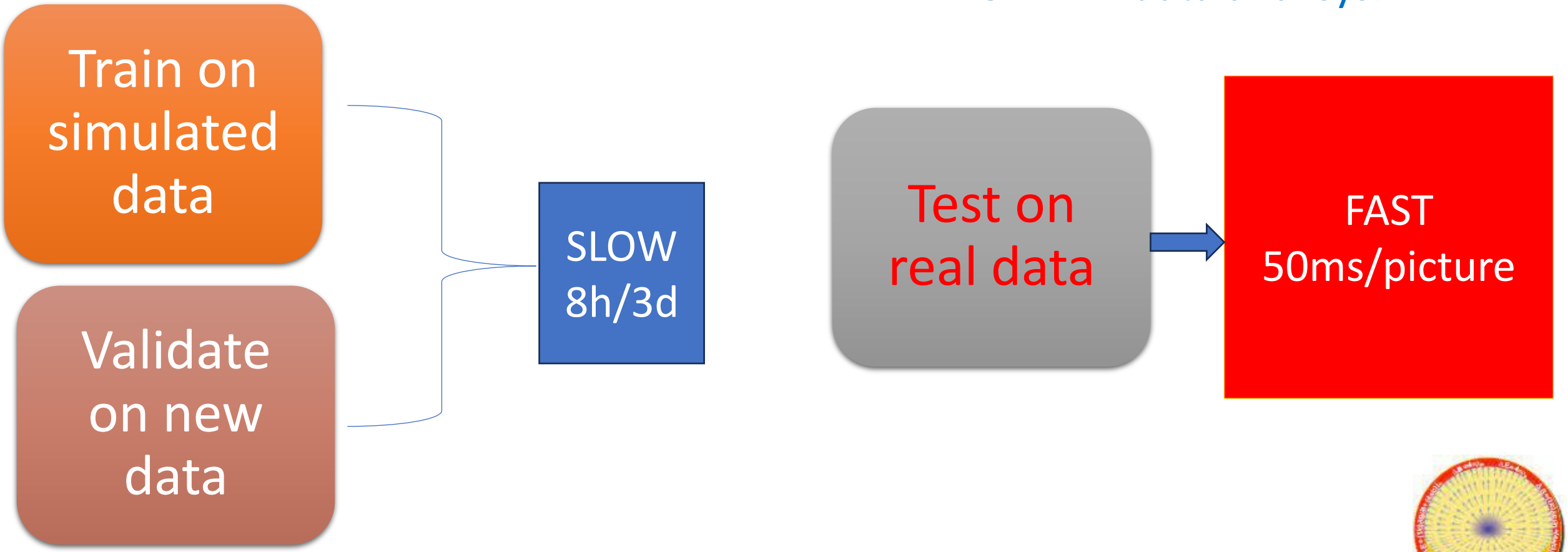


MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

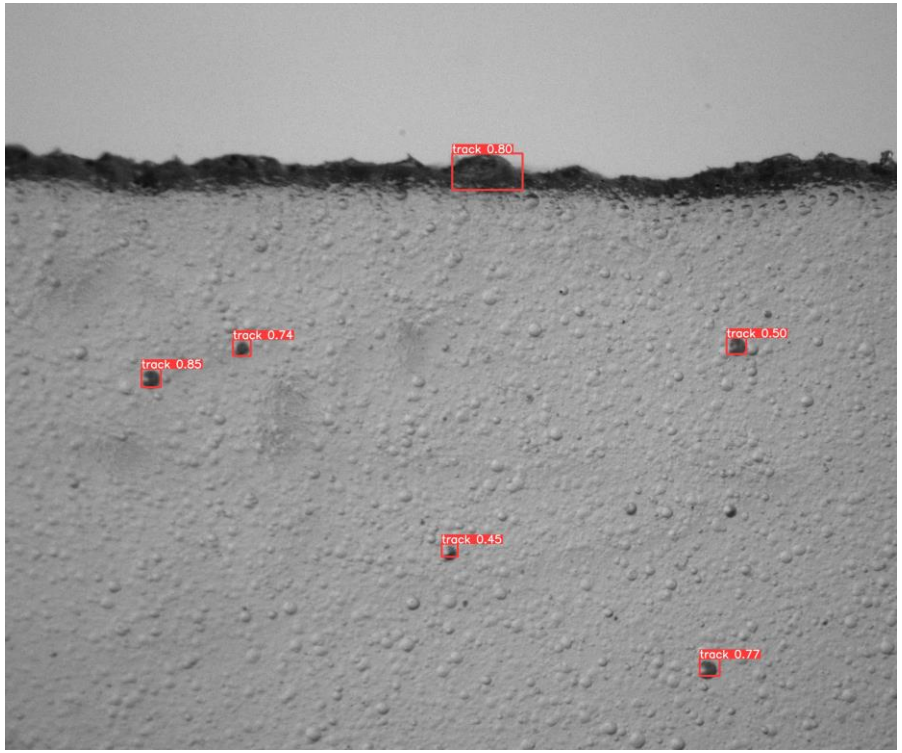
We used «yolov5» networks from small to xlarge with 300 epochs

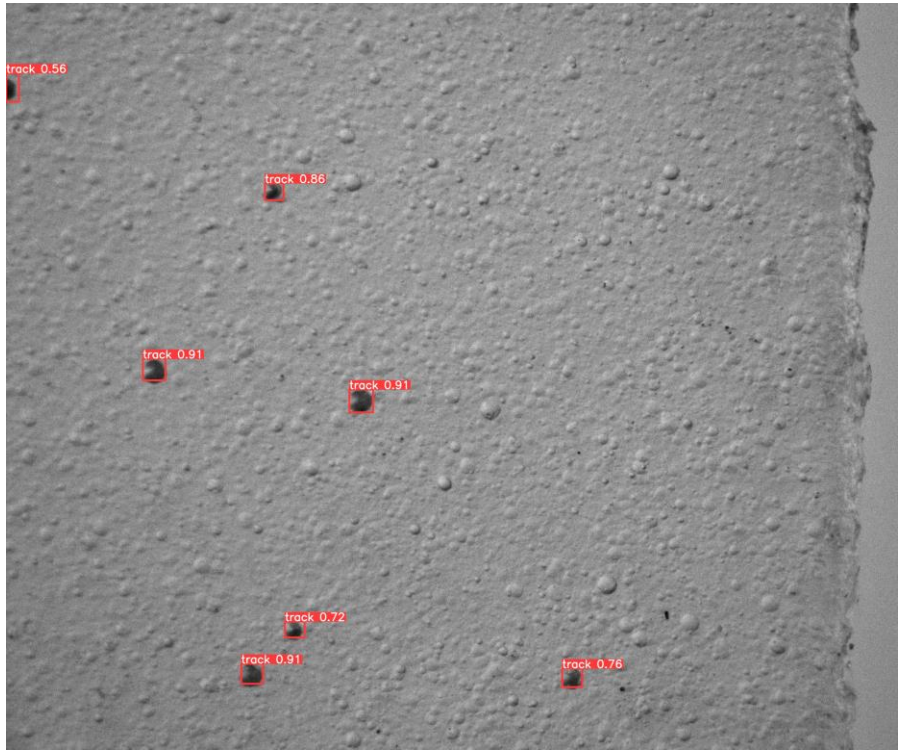
Fast detection permits ONLINE data analisys.



MoEDAL NTD Data Augmentation

Even if it was trained on small (400x300) images the network works nicely on REAL (2448x2048) images.

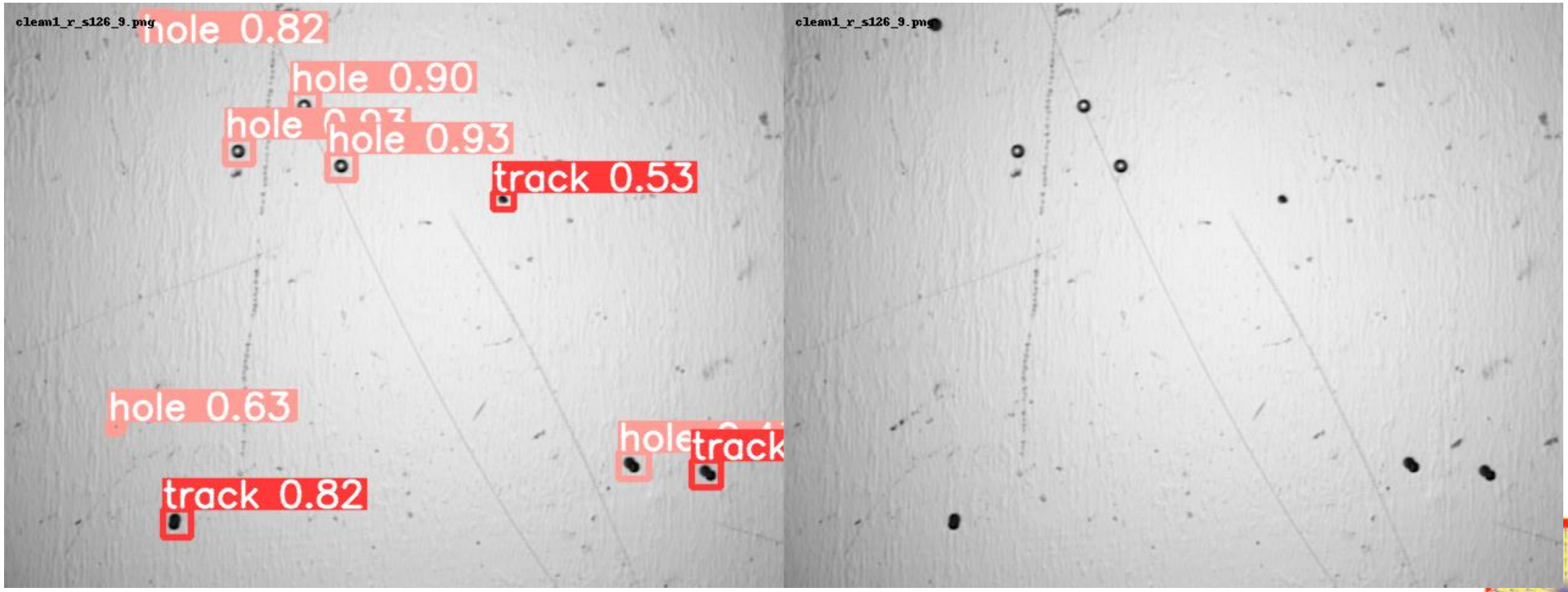




MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

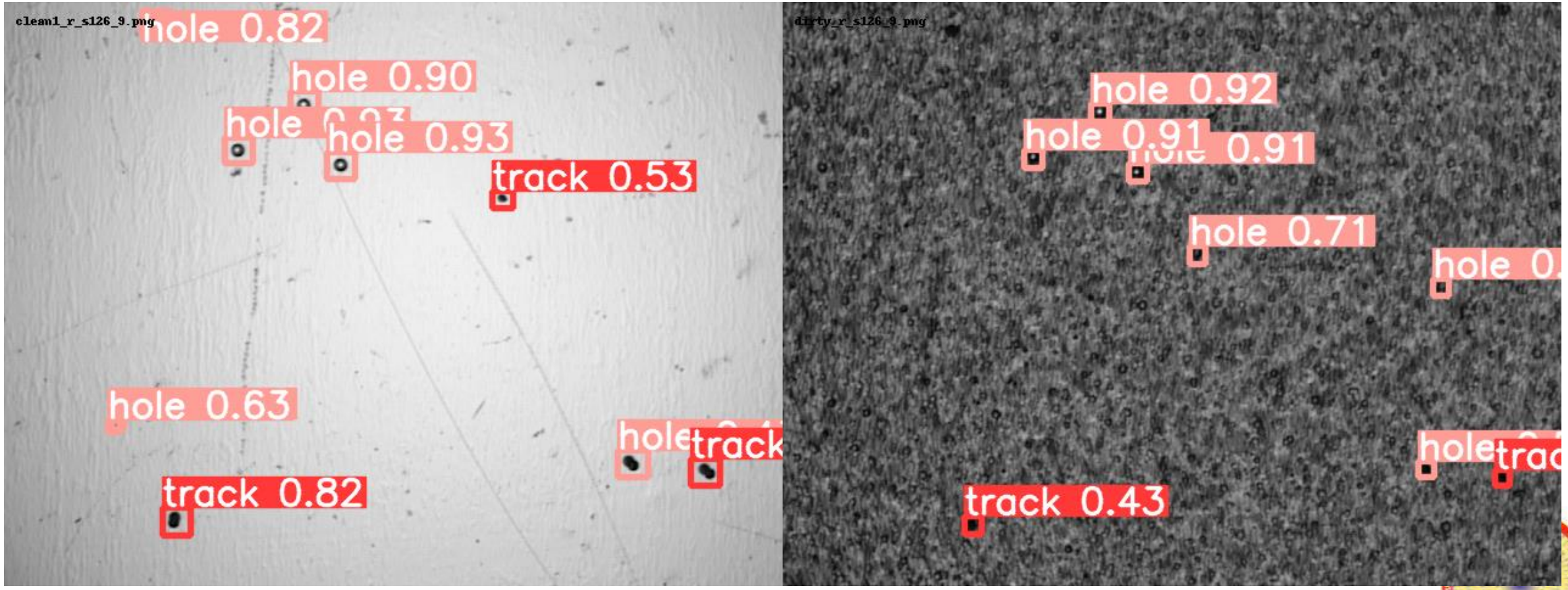
Test has been done also on data and images from the AI group:



MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

Test has been done also on data and images from the MoEDAL AI-group:



MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

Very Preliminary results (manually verified):

Folder	Model	No. of images	No. of Tracks	Correctly identified	False –ve (misses)	False +ve
Co_clear_exp	YoloV5_Small	91	509	449 (88 %)	61 (12 %)	252 (49 %)
YoloV5_Ex large	YoloV5_Large	50	274	231 (84 %)	43 (16 %)	15 (5 %)

MoEDAL NTD Data Augmentation

Istituto Nazionale di Fisica Nucleare

First test of :

- Generation of syntetic images (400x300 due to memory limitation)
- Object detection network traing on syntetic images
- Object detection on REAL data

Has been successful!

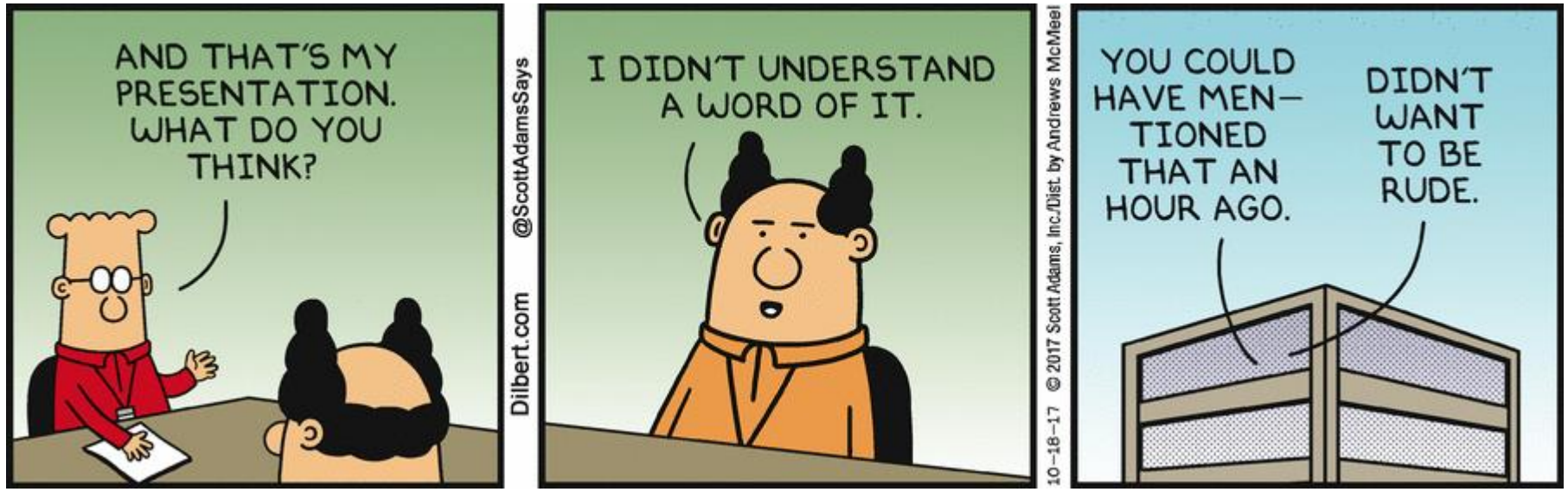
We will now fine tune the analisys pipeline in order to deploy a ‘real time’ image analysis system for the Bologna Microscope.

Referrences:

https://github.com/tamarott/SinGAN

https://github.com/ultralytics/yolov5/releases

MoEDAL NTD Data Augmentation



Giuseppe Levi,
University of Bologna, INFN
Levi_at_unibo.it