AI for analysis exo-Jupiters climate ensemble simulations

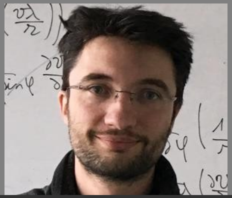
Who ?

Merieme Bourenane

Simplice Donfack

Martial Mancip

Felix Sainsbury-Martinez



Pascal Tremblin

Hiba Taher

How to categorized event in results of ensembles of HPC simulations with visualization plots and Deep Learning ?

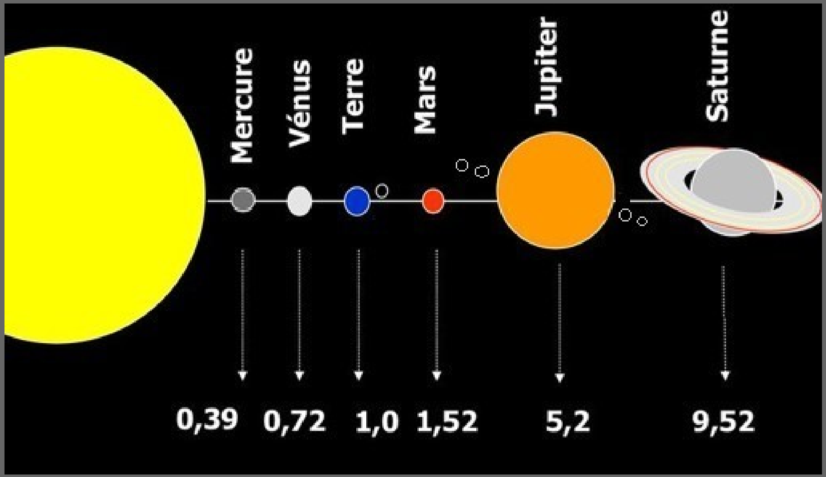
How to detect positions of physical events with computer vision technics ?

Simulations

Simulations of Jupiters-like planets near their sun : between 12 and 334 Astro Units.

Hot jupiters have synchronized orbits.

Sun	Hot Jupiters	Base	Cold Jupiters
	< 0.048 UA	0.048 UA	0.048 < < 0.334UA



8 simulations (Dynamico model) ; 3300 saved steps, resolution 1 degré (10000 cells).

(every 1.2 e6 secondes = 50 10^3 days) => ~600G of data

2018 publication :

F. Sainsbury-Martinez et Al. « Idealised simulations of the deep atmosphere of hot jupiters: deep, hot, adiabats as a robust solution to the radius inflation problem. »

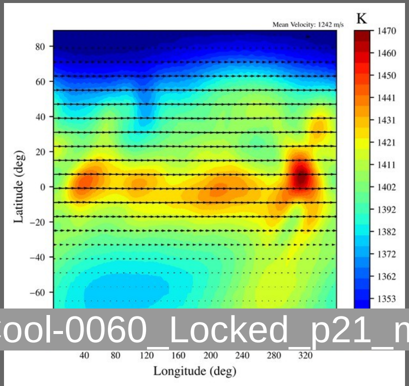
Simulations

We observe an inflation of hot exo-jupiters (like a rugby ball). Why ?

Matplotlib plots

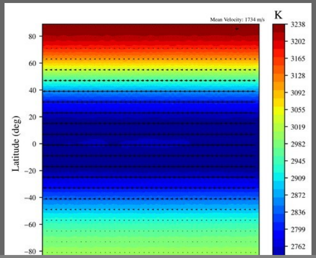
5 categories :

"asymetric"



0060-Cool-0060_Locked_p21_mean_4

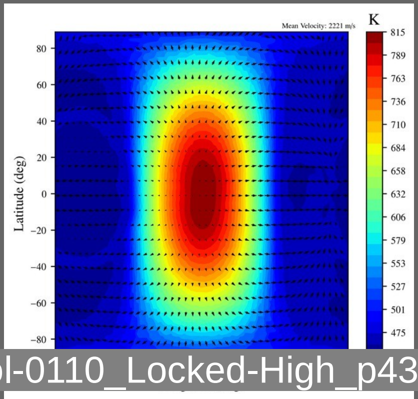
"banded"



0021-Hot-0021_Locked_p14_mean_4

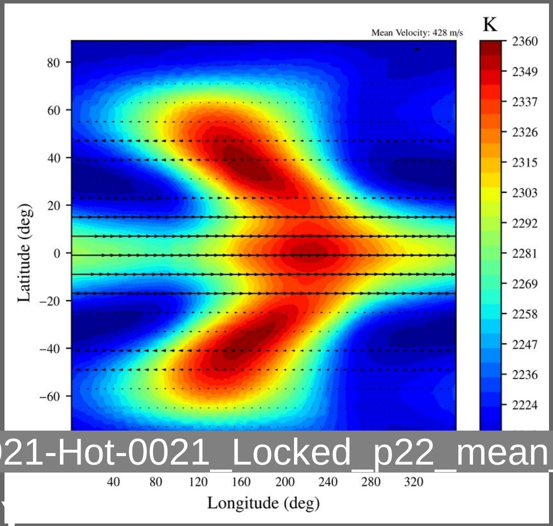
"locked"

Hot spot in the front face



0110-Cool-0110_Locked-High_p43_mean_4

"butterfly" and "jet"



0021-Hot-0021_Locked_p22_mean_1

Mean over 200 time steps (suppression of turbulence)

AI for understand those simulations ?

- With an ensemble of HPC simulations,
  - and visualizations that gives as a low precision model (RGB = 3 x [0-255]),
- Can we build an convolution neural network (Computer vision) to automatically get event map of simulations ?
- And better understand global physics of those simulations ?
- Second major property is the position of each event.
- We start from this post-hoc analysis, but plan for in-situ analysis in the future.

Data : The big work.

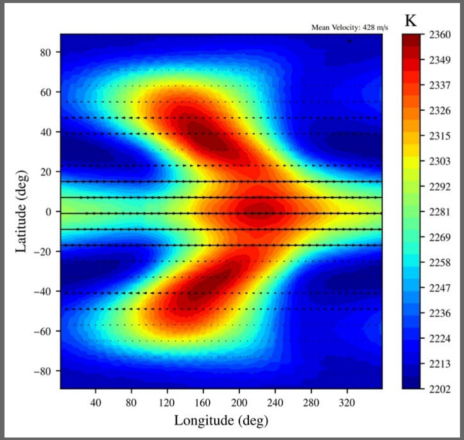
We can’t deal with scientific visualisation for AI directly. Here we have an overlap of Temp and Wind field (for the « jets »).

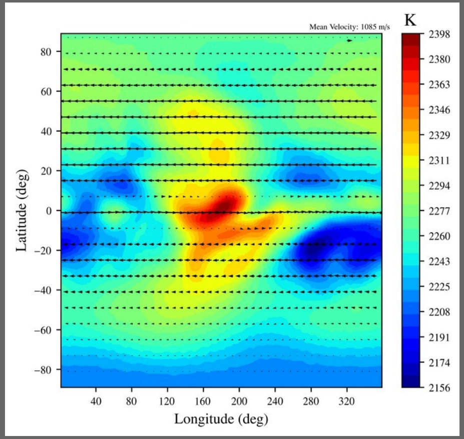
Results as accuracy of fits are low.

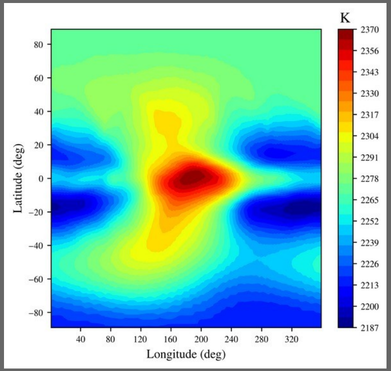
We had to split variables.

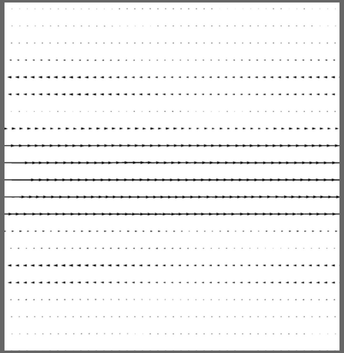
But Wind field are still not good.

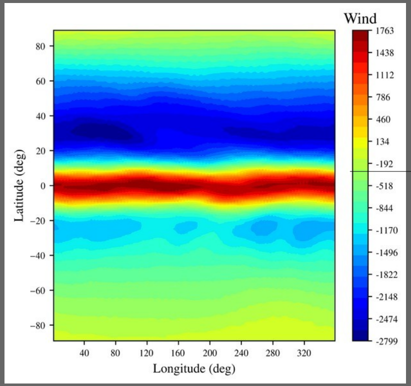
We have used longitudinal coordinates of wind for better results and extract pictures only with Pillow module in memory.

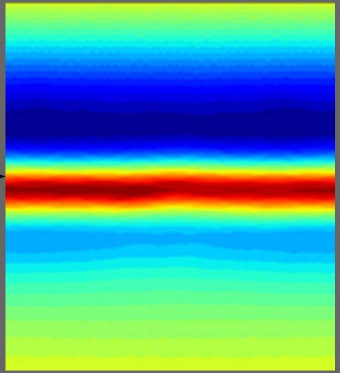












AI Tools for understand those simulations.

MAISON DE LA SIMULATION

CNN model for multi-catégorial classification

for earth atmospheric simulations.

Ryan Lagerquist, Amy McGovern and David John Gagne II (2019). *Deep Learning for Spatially Explicit Prediction of Synoptic-Scale Fronts.* American Meteorological Society.

We use TensorFlow inside Docker container to ease moving of AI core :

Docker Tensorflow on node with 5 K40 and another with 13 K80.

Fidle Course in french Grenoble – IDRIS :

https://gricad-gitlab.univ-grenoble-alpes.fr/talks/fidle

Projet Dockerfidle : https://gitlab.maisondelasimulation.fr/dataia/dockerfidle

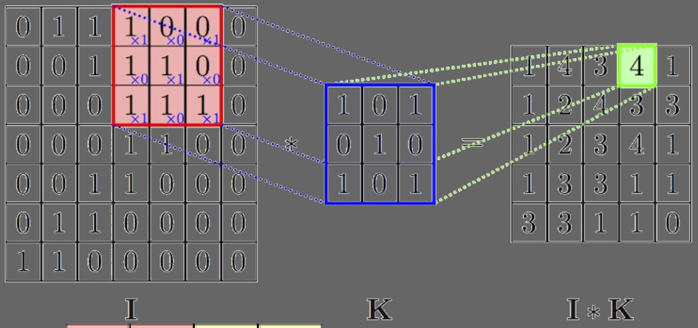
CNNs do stochastic optimization.

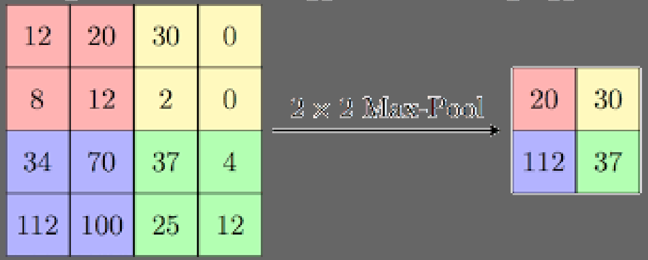
Data => reductions to use one GPU(s) and parallelisation with data or replication

Model Keras: => Fit with 80 % of dataset => Validation 20 % => save best parameters

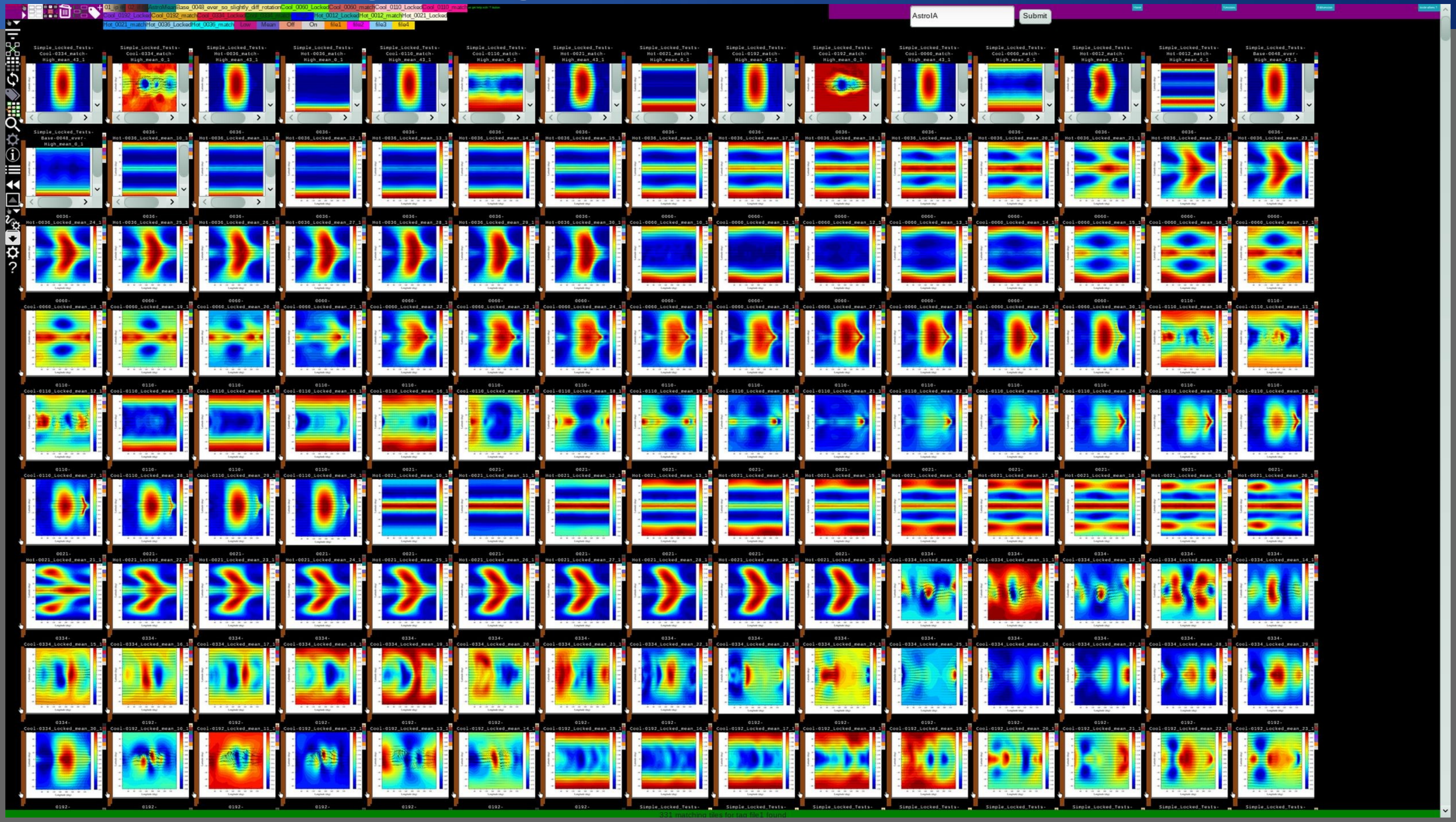
Layer (type)	Output Shape	Param #	
input_1 (InputLayer)	[(180, 360, 3)]	0	3 = « RGB »
conv2d (Conv2D)	(176, 356, 16)	1216	
max_pooling2d (MaxPooling2D)	(88, 178, 16)	0	
dropout (Dropout)	(88, 178, 16)	0	
conv2d_1 (Conv2D)	(84, 174, 32)	12832	…
flatten (Flatten)	(8064)	0	
dense (Dense)	(128)	1032320 …	
dense_2 (Dense)	(4)	260	4 = « asymetric, banded, locked, butterfly categories »

Total params: 1,208,612





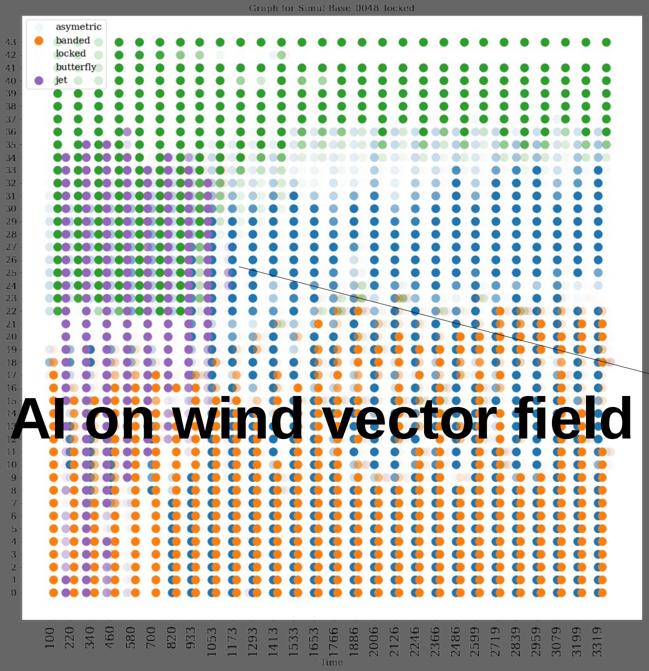
Labeling / Tags with TiledViz tool



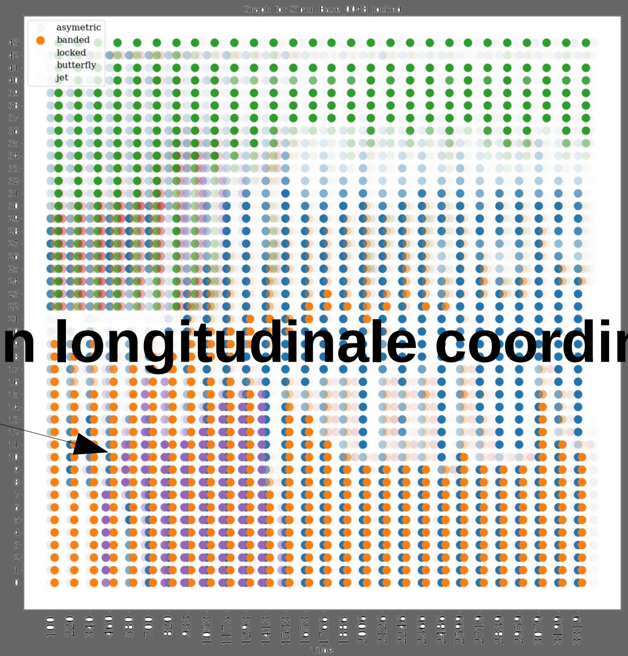
Tiled Visualizations for display walls by M. Mancip et Al.

Event maps

- We get multi-categorial maps for each simulation
- CNN AIs on vector fields have not worked so we have used 2D plots of longitudinal coordinate of the wind :



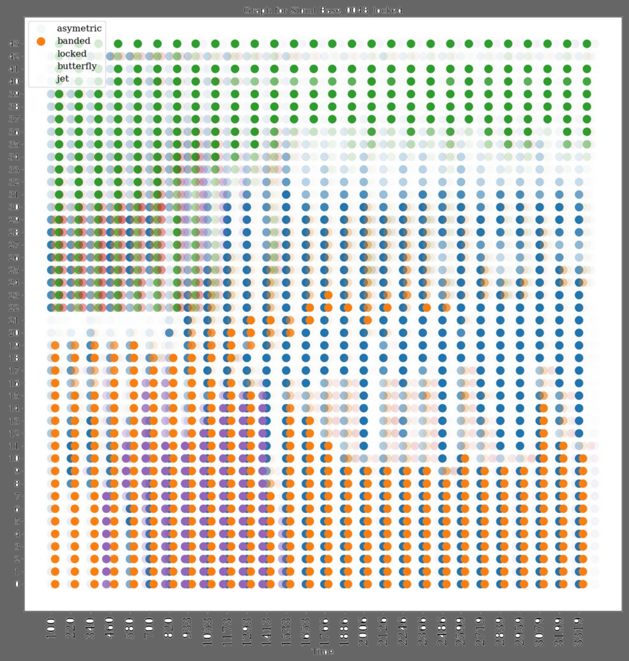
AI on wind vector field

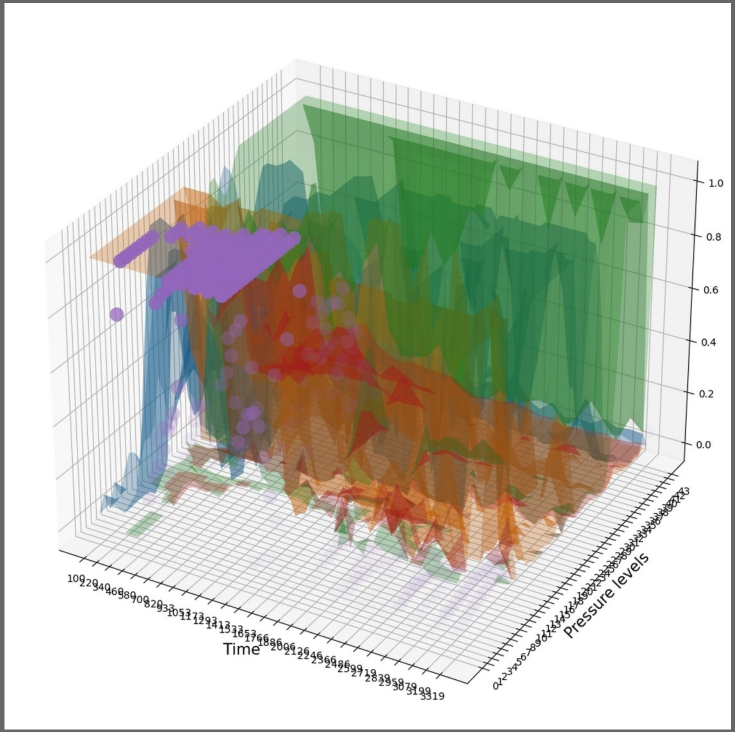


AI on longitudinale coordinate

Event maps

- 3D Vision with matplotlib widget in jupyter notebooks :

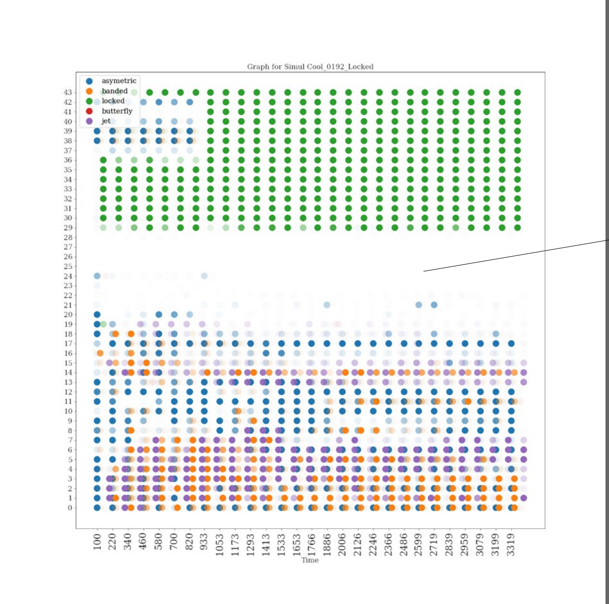


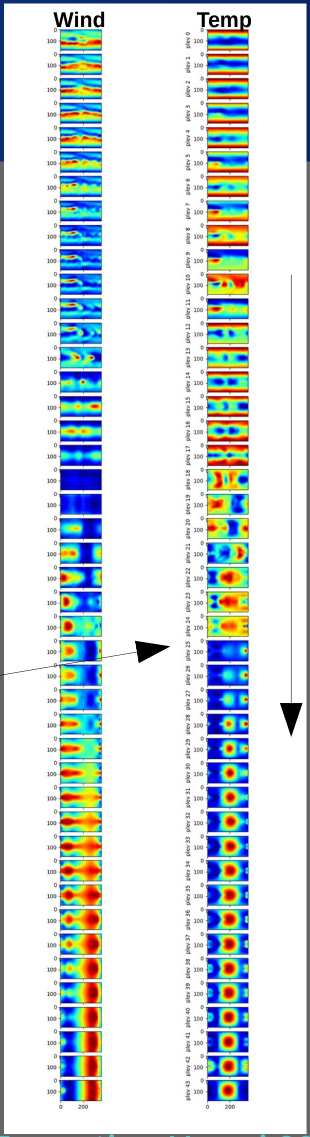


Unknown event areas

Science is in unknown areas.

- What events is in this blank areas for Cold_0192 simulations ?





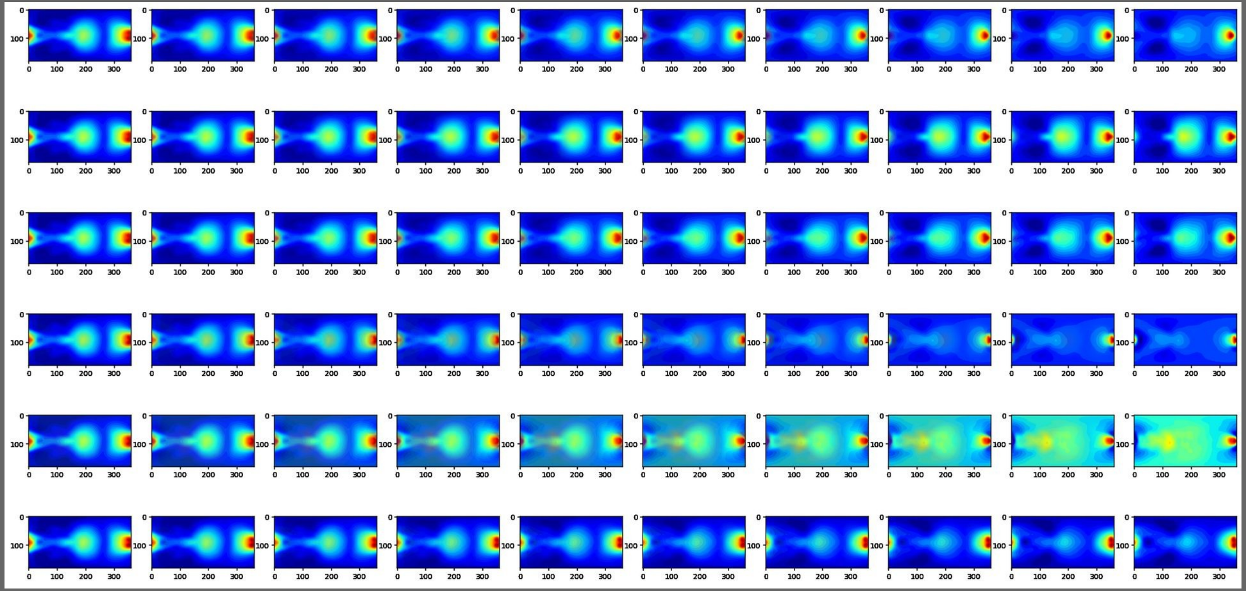
L
e
v
e
l

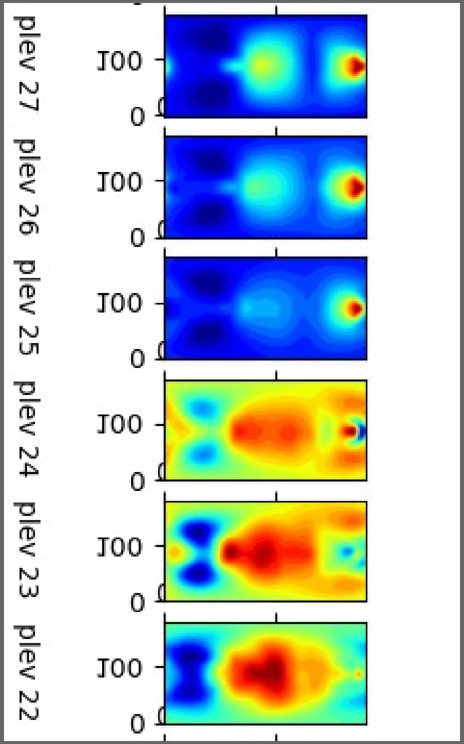
Wind Temp

New Tags for inversion

New events (hot spots in the back face) :
« inversion down » and « up »

Rare events. Need oversampling :
Here only picture interpolations (linear) gives good pseudo results.



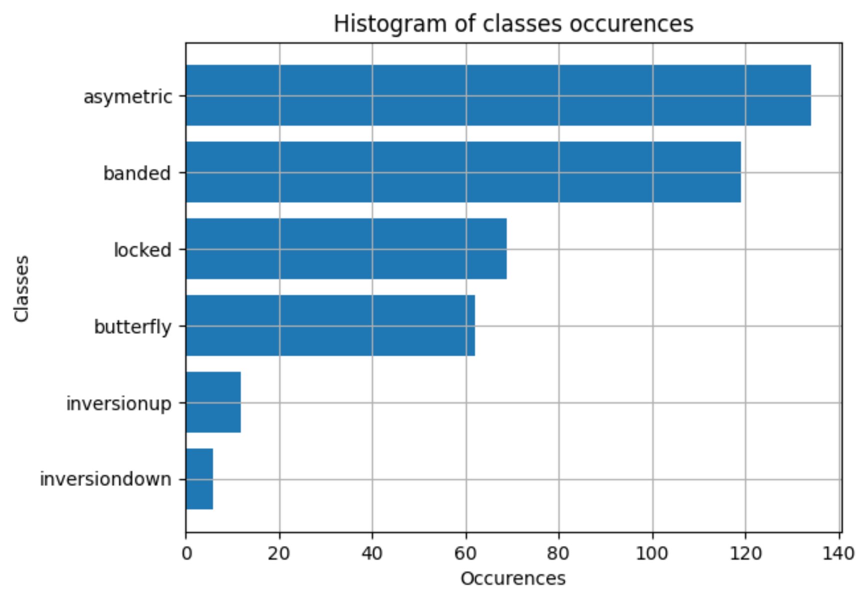


Levels

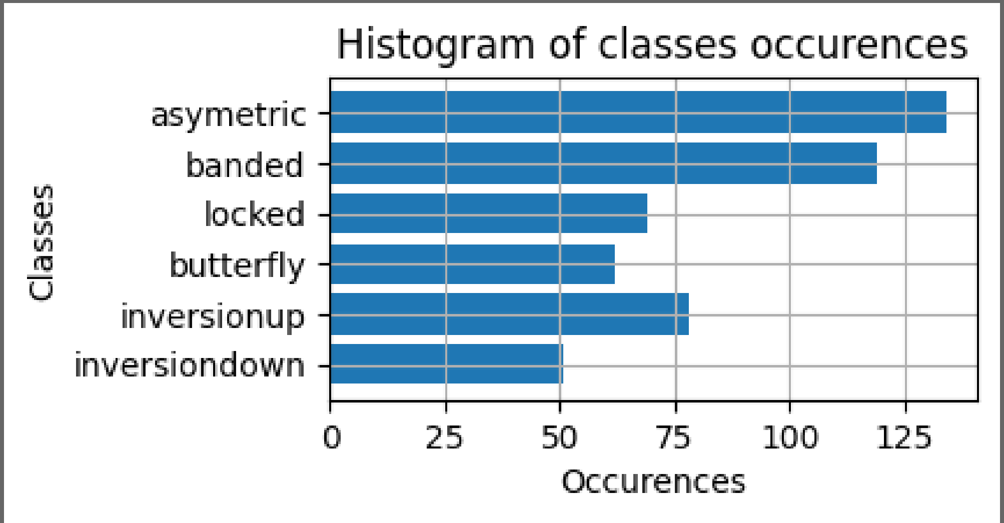
Oversampling

Tags inversion down and up frequency in training set must be balanced.

Before oversampling



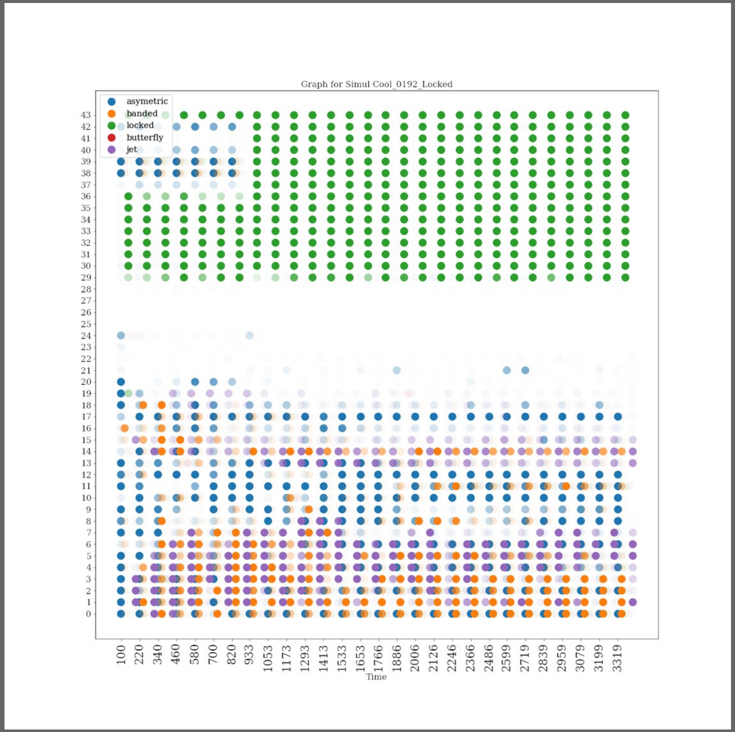
After oversampling



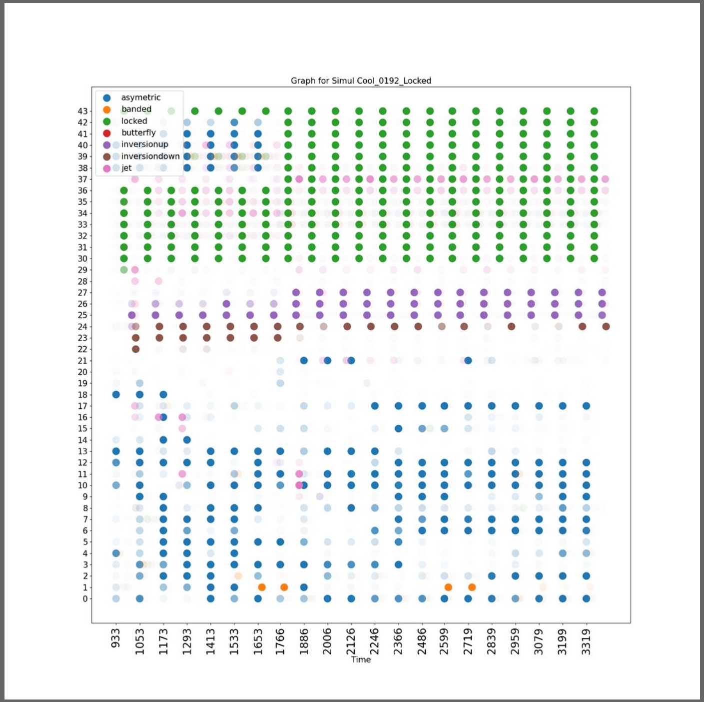
Maps with inversion down/up

Wind Temp

Without inversion down/up



With inversion down/up :
New hot spot in the back face.



3D visualization to confirm results

Wind Temp

Helmoltz decomposition for the wind

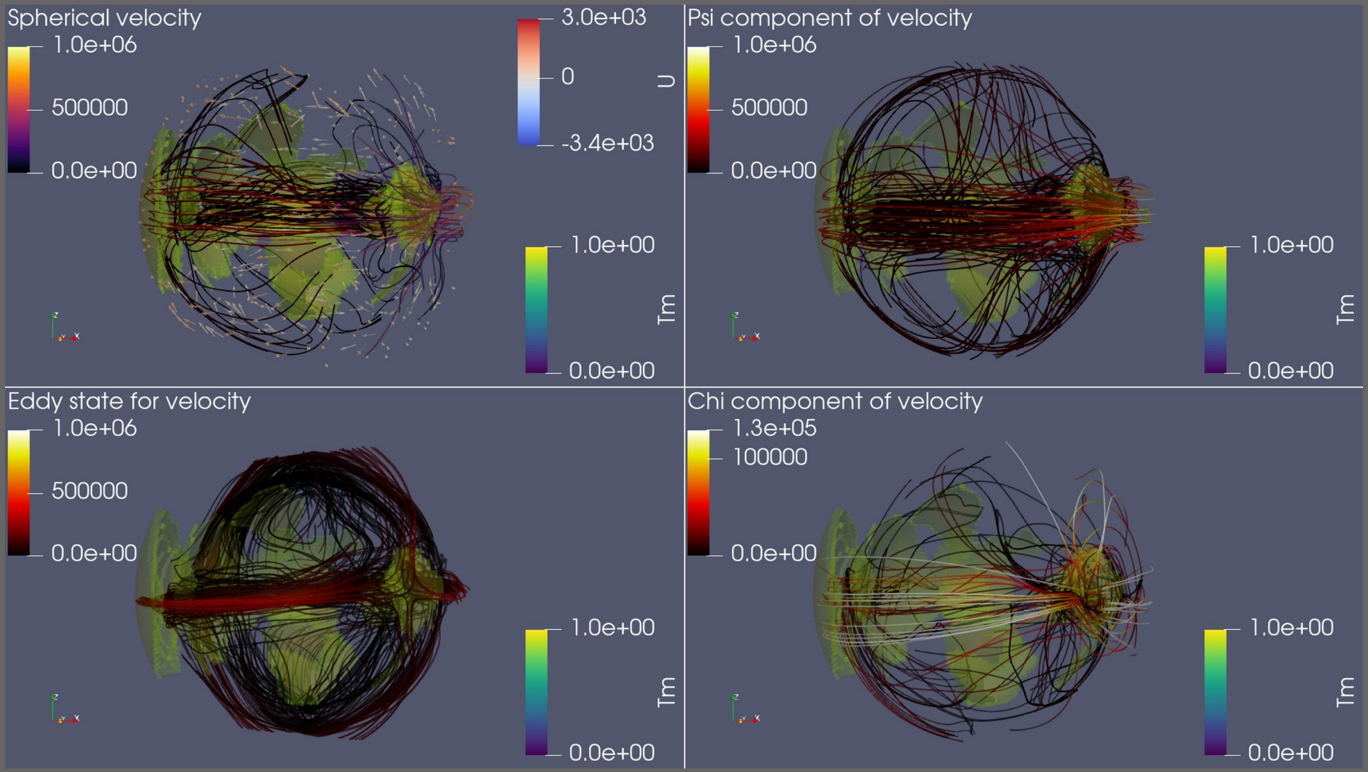
and plot of temperature

with isovolumes.

In the left : front hot spot

In the right : back one.

We see the heat transfer.



On our Youtube
https://www.youtube.com/watch?v=toDpbqer2e4

Colormaps

Colormaps : Perceptually uniform, sequential (linear) or rainbow palettes ?

Hue is the component that distinguishes “different colors” in a non-technical sense (RGB).

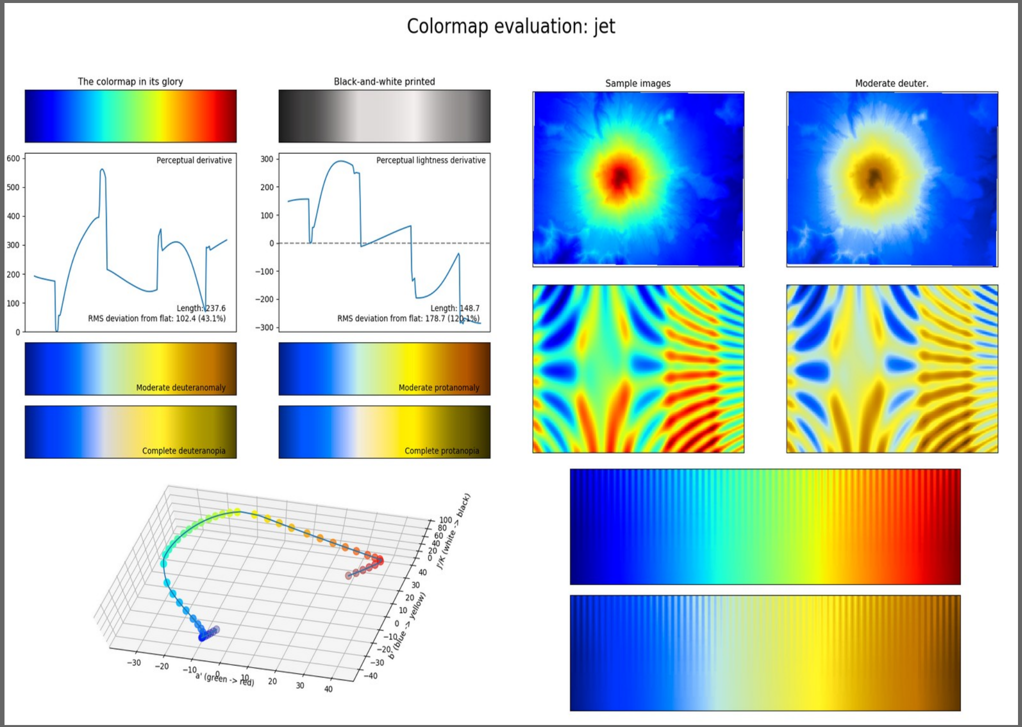
Saturation (or chroma) is the colorfulness. Two colors with different hues will look more distinct when they have more saturation.

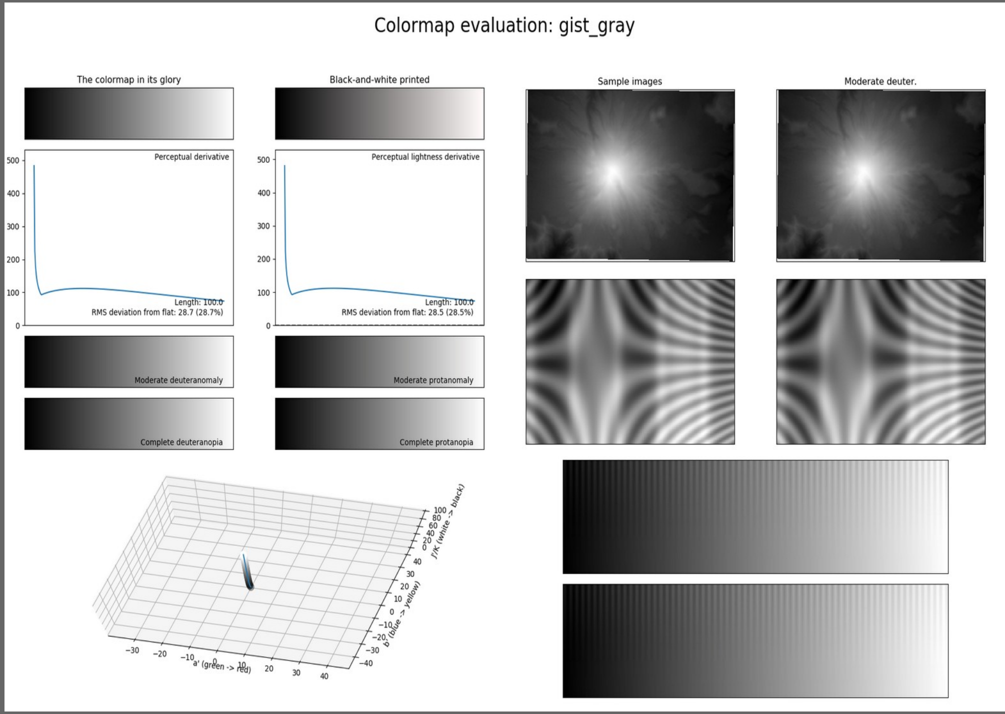
And lightness corresponds to how much light is emitted (or reflected, for printed colors), ranging from black to white.

What parameter is more relevant as zero layer with our CNN ?

Some colormaps

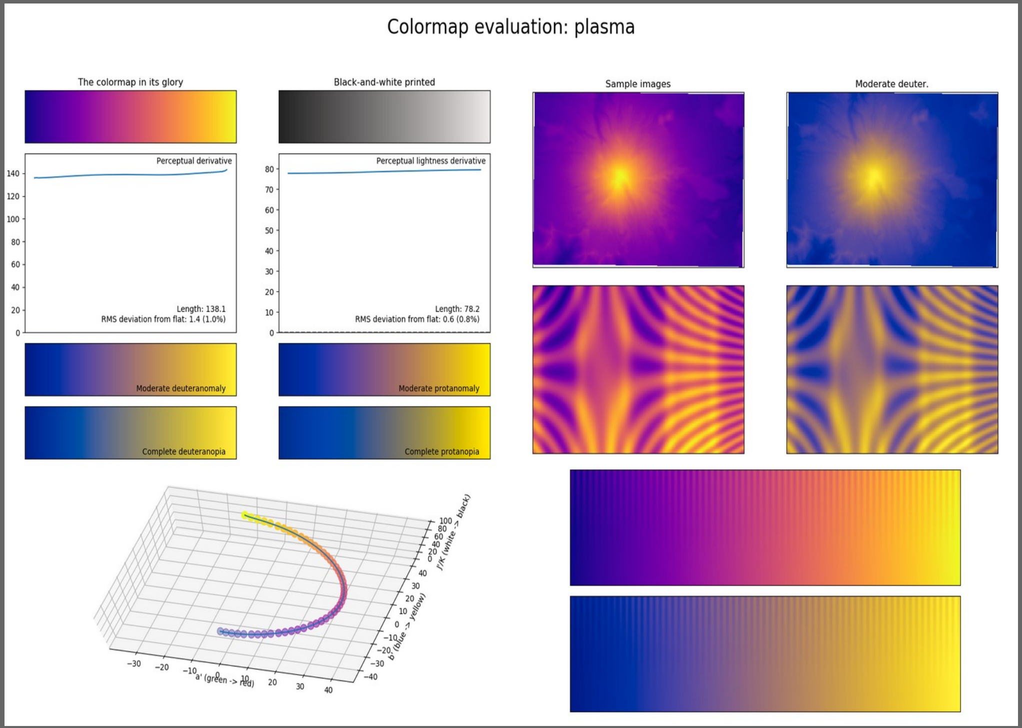
- Matplotlib colormaps.
- Miscellaneous and Sequential

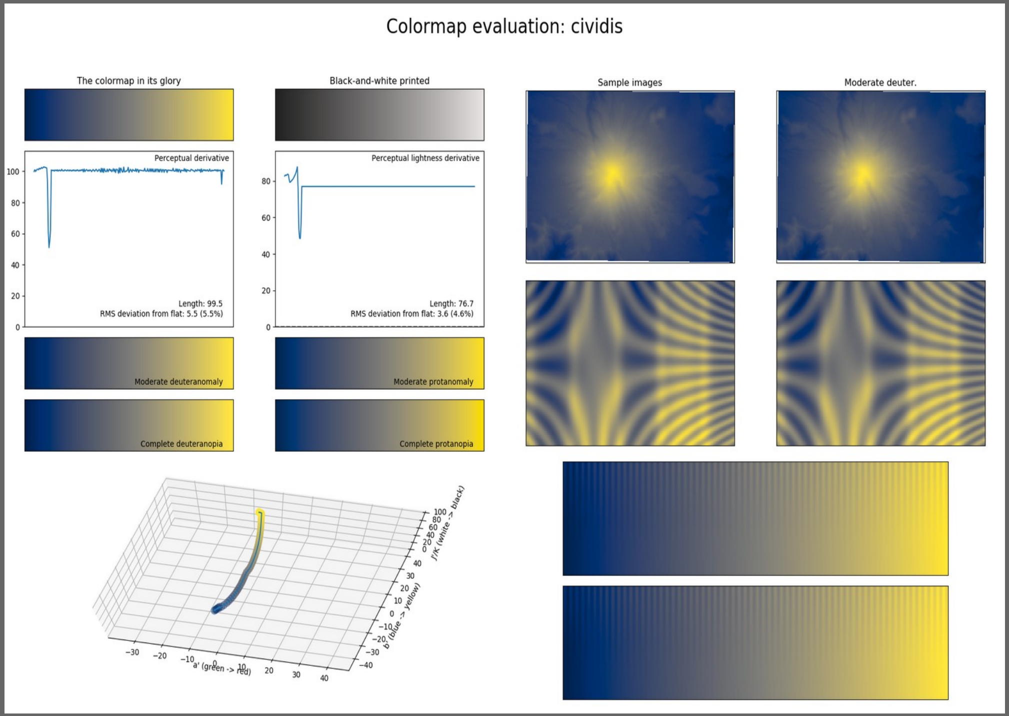




Some colormaps

- Perceptual colormaps





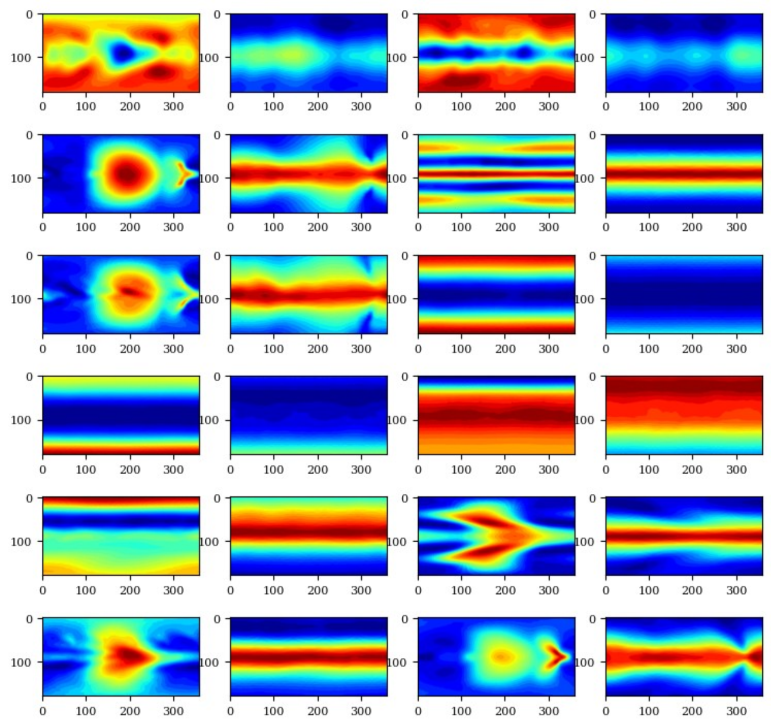
Colormaps

Colormaps (plot process) may be considered as the first fixed layer of the AI.

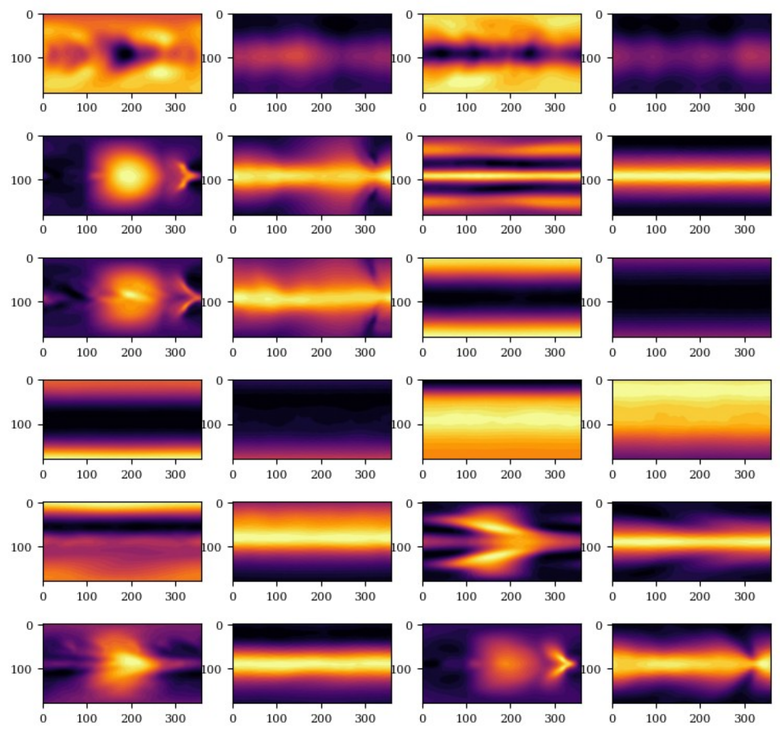
Then it is an hyper-parameter.

Do the perceptual colormaps give better inferences ?

« Jet » is always bad !

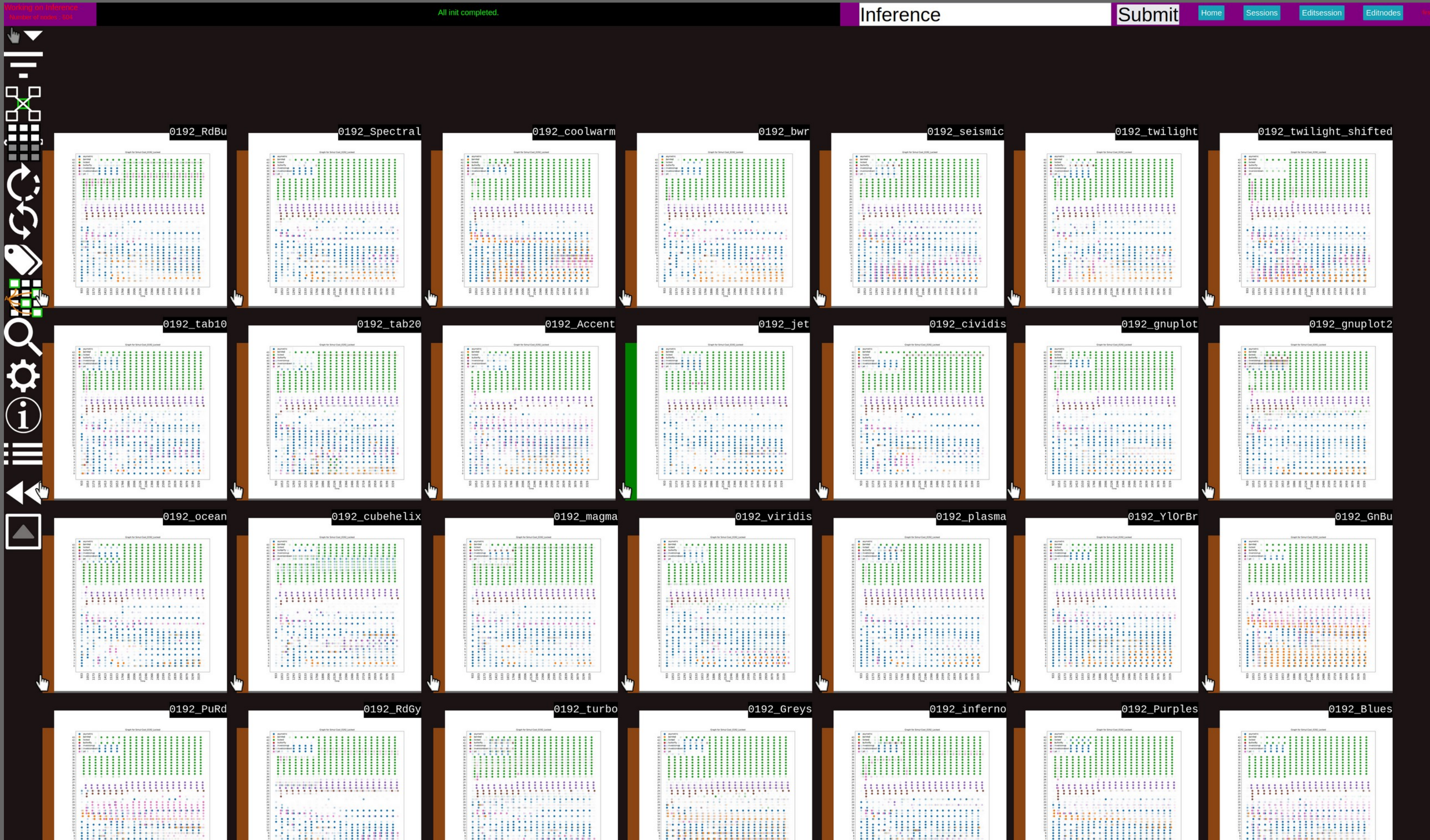


inferno



Inferances

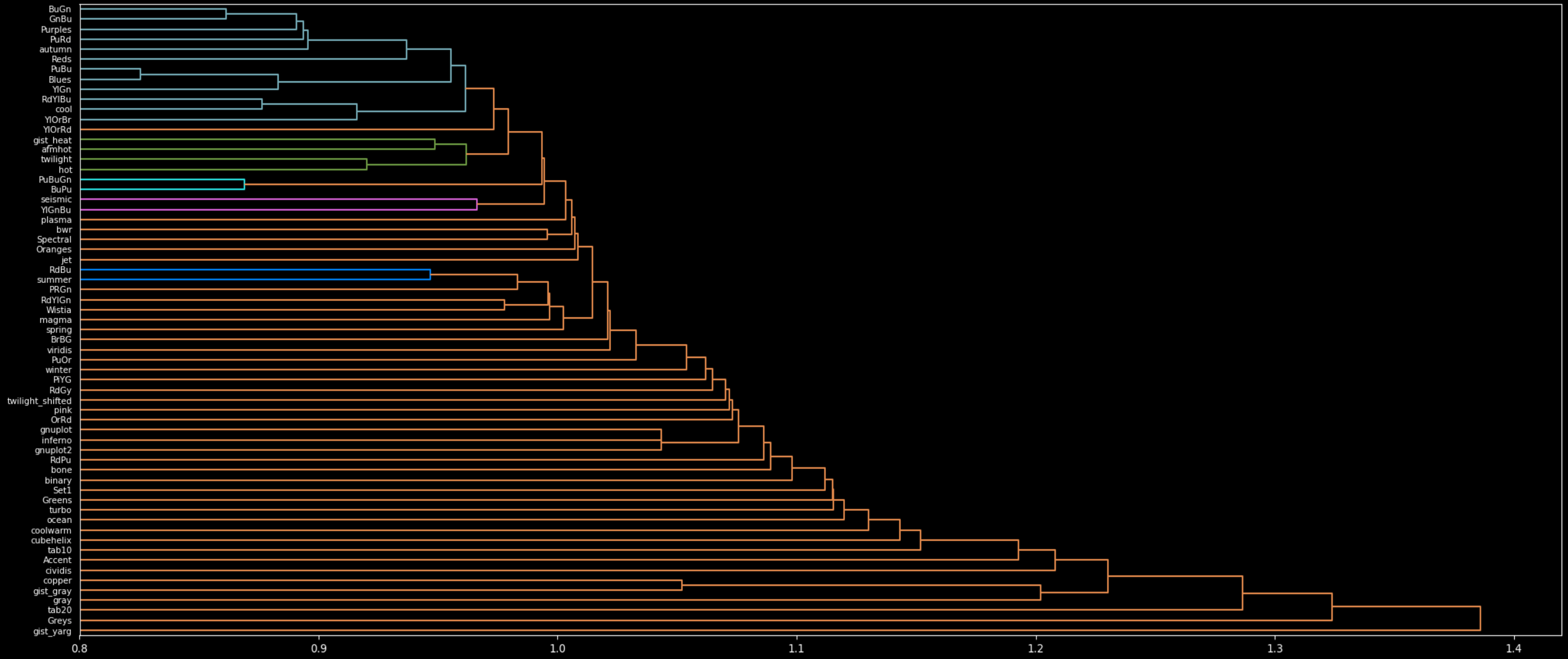
- Jet is not the best one !



Colormaps

Dendogram (L2 metric) : Clustering of inferences (not training) with all matplotlib colormaps :

Greys is bad too because it is a simple truncation of data.



Colormaps

MAISON DE LA SIMULATION

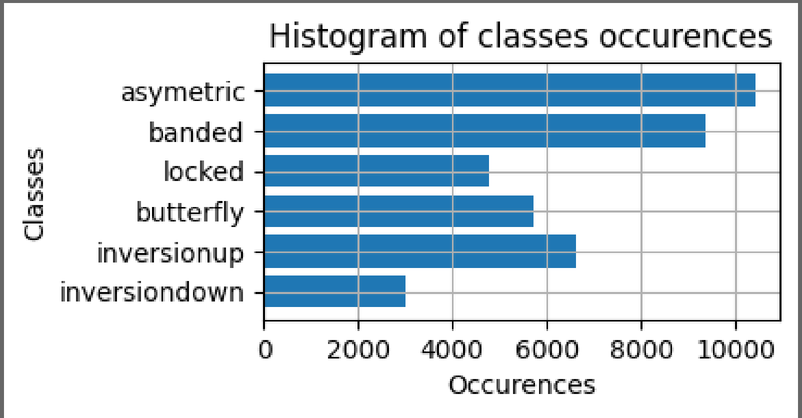
We are looking for an optimization algorithm to answer « what is the best colormap for this AI ».

In progress : compute PNSR with RNSE / ACP + KNN

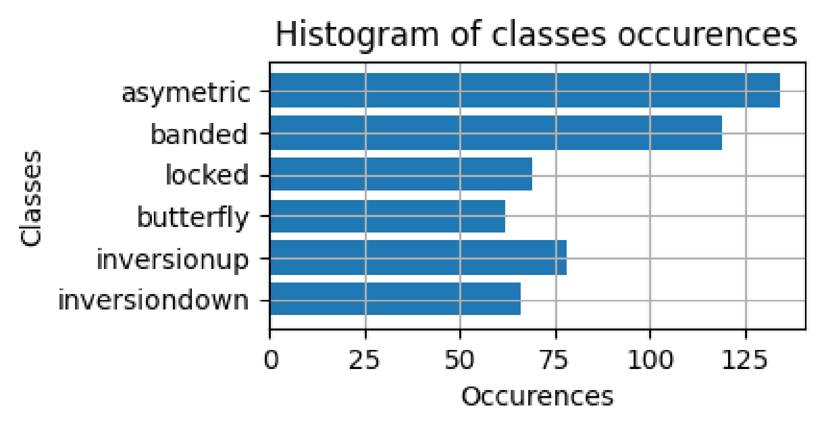
Ideas ?

AI on simulation datas

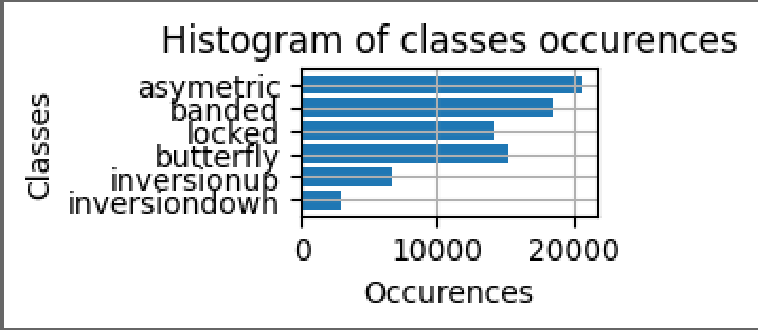
- Datas with pictures versus direct mean of simulations.
- Less points with our data sets 180x360x3 => 90x180x1



Overlapping on all labels on pictures.



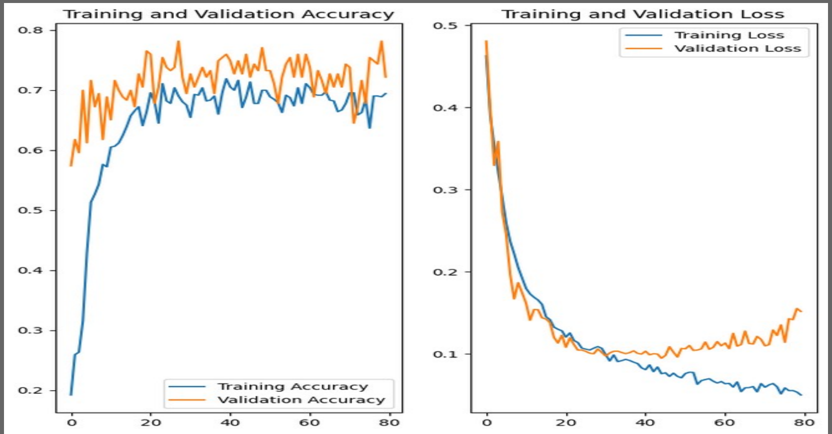
Only inversion overlapping on pictures.



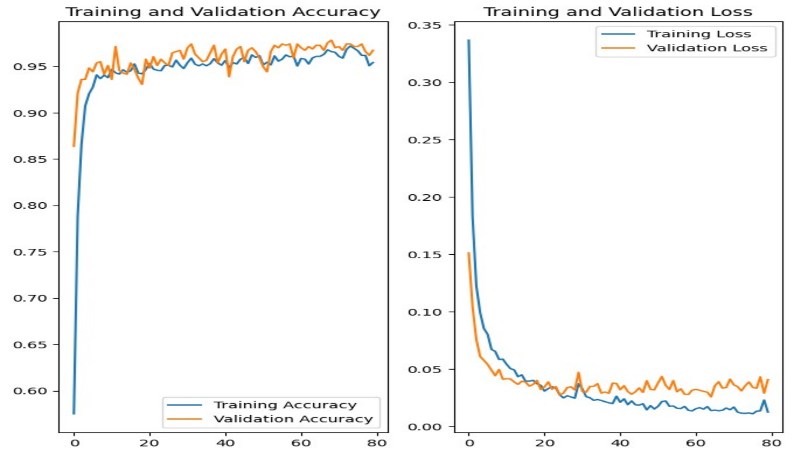
Overlapping on all labels on datas.

IA on simulation datas

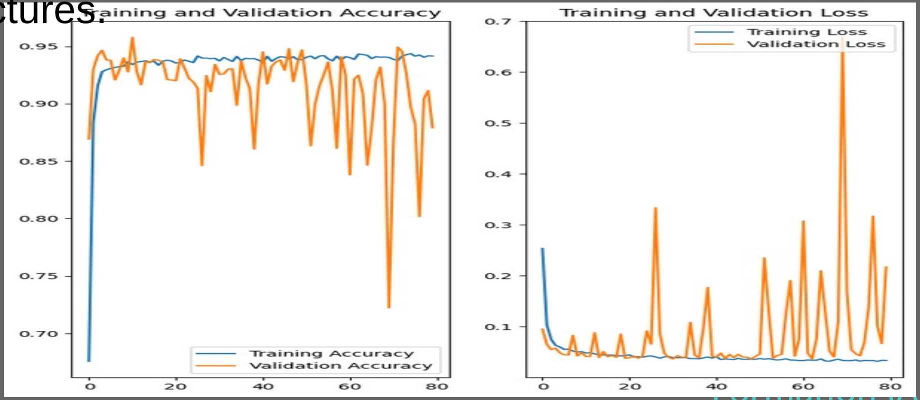
- Instabilities on datas vs pictures.



Loss with inversion overlapping only on pictures.



Loss on global overlapping with pictures.



Instabilities on overlapping on all labels on datas.

Detection of position

- Work of Hiba.
- Object detection algorithms : we study usability of Faster RCNN, SSD and YOLO algorithms and compare the results on inferances.
- First test with YOLO (You Only Look Once) : Unified, Real-Time Object Detection https://arxiv.org/pdf/1506.02640.pdf
- Link code YOLOv5: https://github.com/ultralytics/yolov5

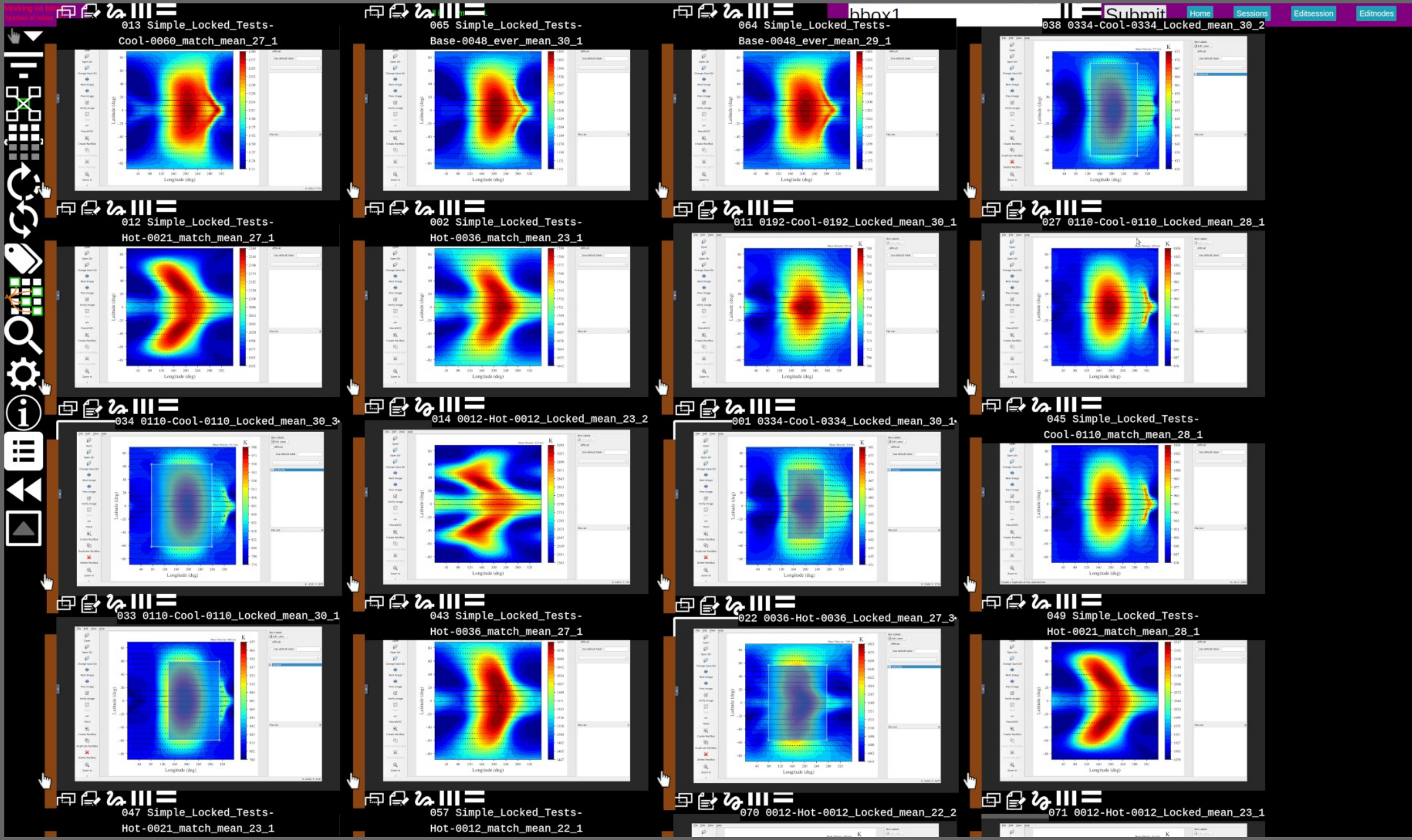
Detection of position

MAISON DE LA SIMULATION

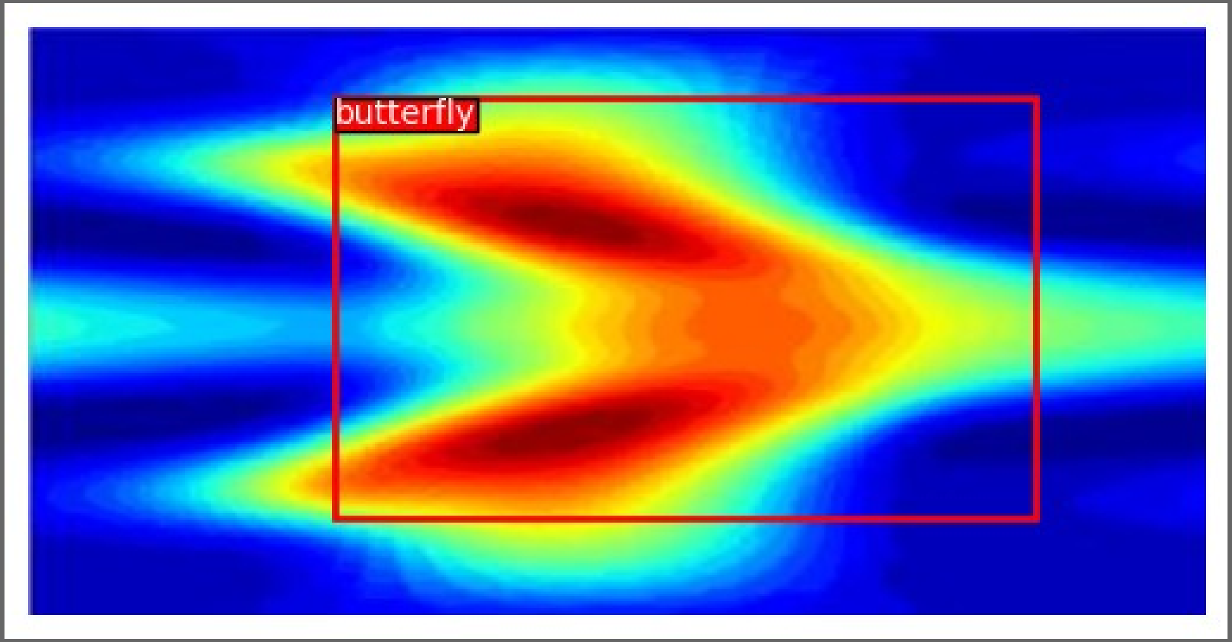
IA on bounding boxes ! https://github.com/heartexlabs/labelImg

Supervision :
« TiledLocalize »
TiledViz case with
LabelImg package

(in Label Studio suite)



Detection of position



Detection of position

- « Inversion » is not an event « locked » and musn’t be detected as such.

  Discussion :

- Those methods are built for usual object or animal detections
  but we have physical results.

  Sometimes we known more on properties of our plots :

- Here , the 0 longitude is the position of the sun.
- « Locked » is a bubble in front of the sun.

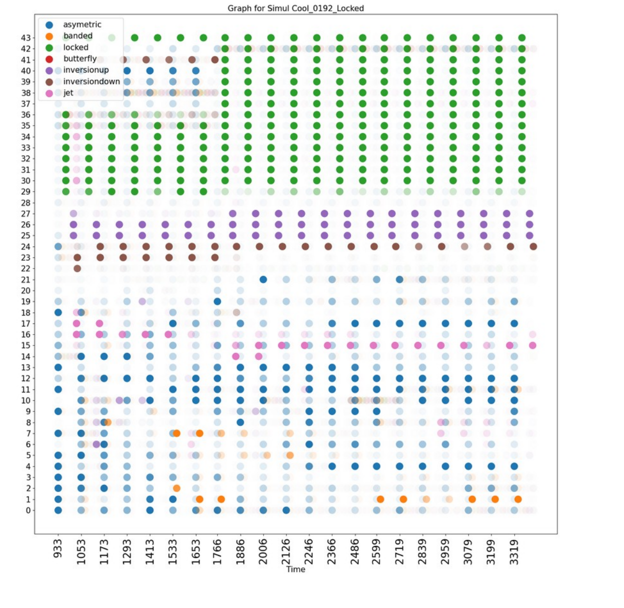
  Oversampling with simple interpolation may be better
  to conserve physical properties than those inside YoloV5.

Detection of position

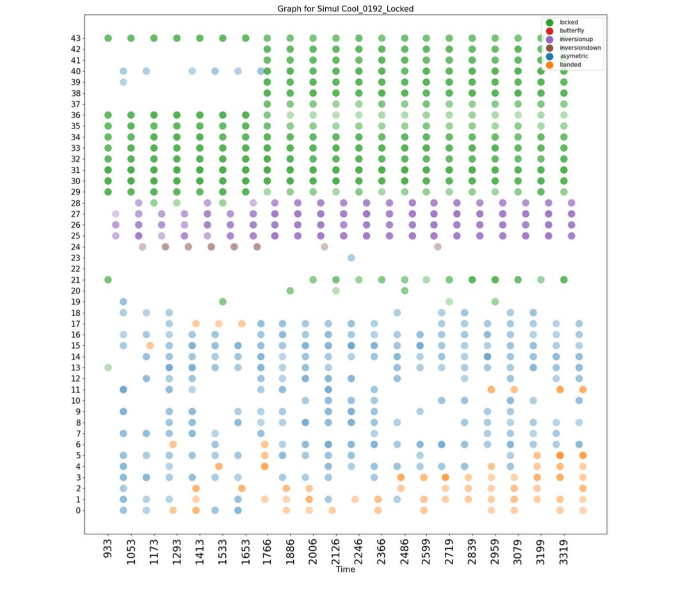
- Fine tune a pre-trained YOLOv5 model:
  YOLOv5s summary: 157 layers, 7015519 parameters
  mAP : 0.985

- Good results for CNN part :

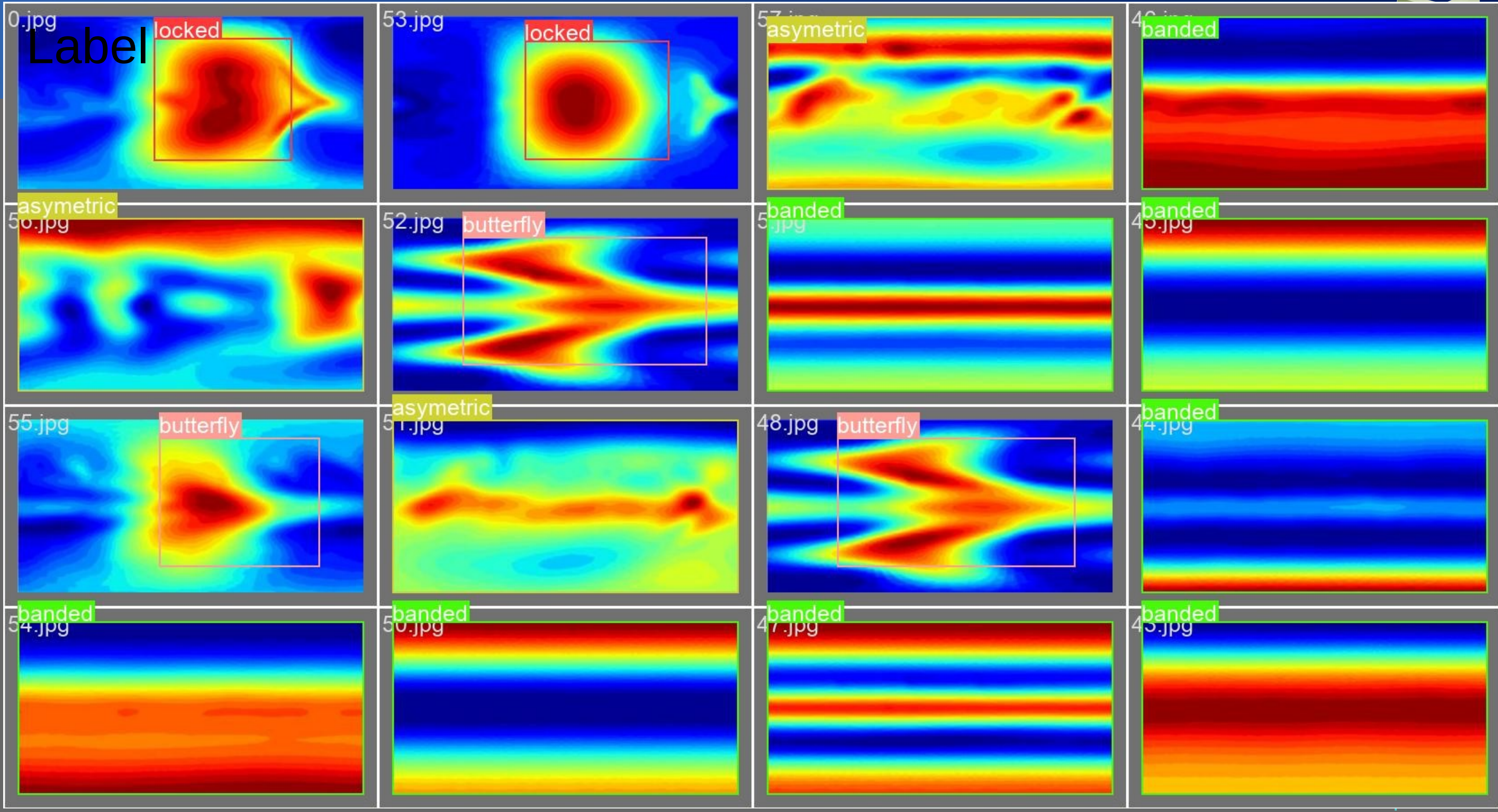
Simple CNN



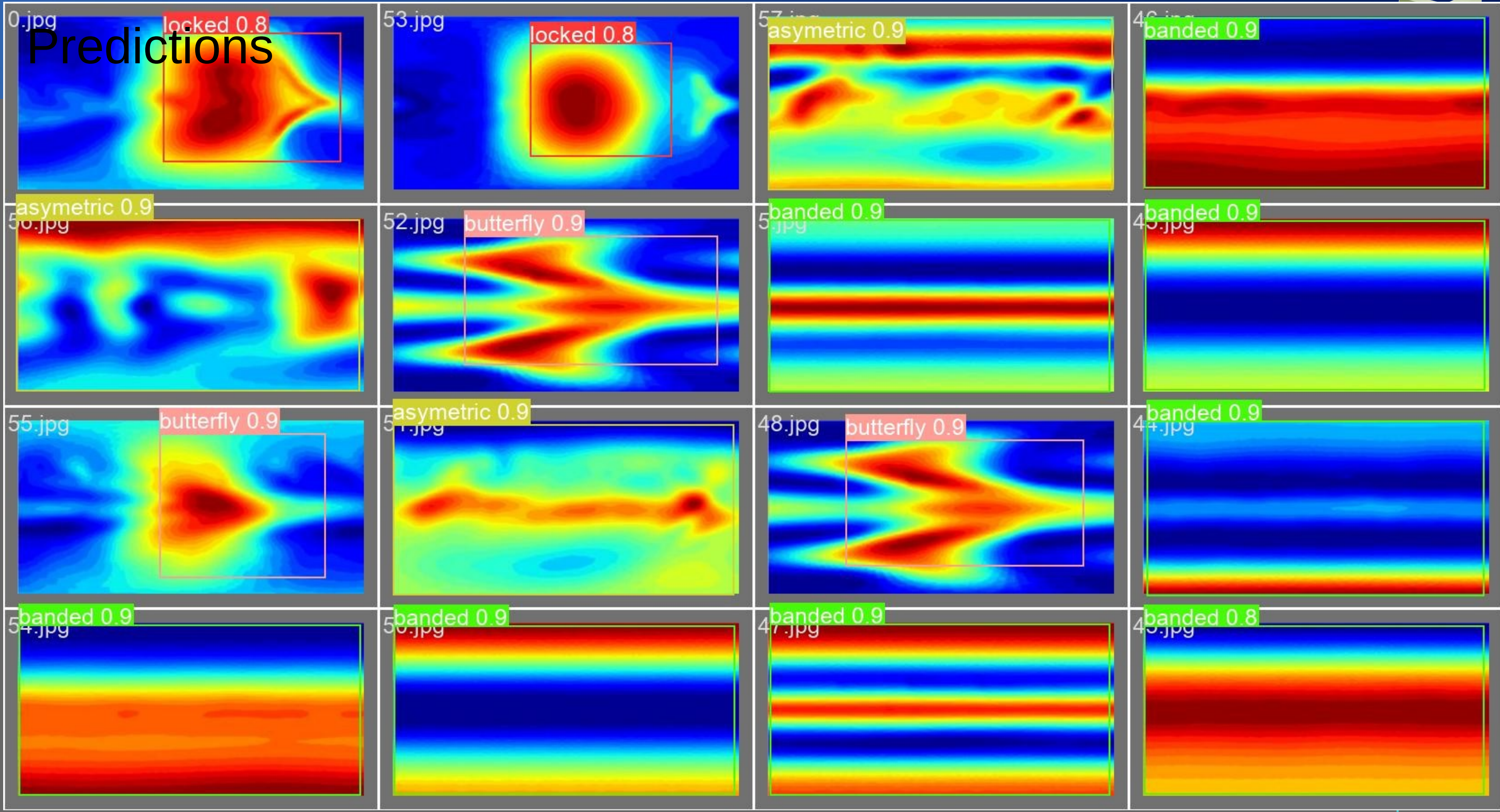
Yolo v5 CNN



Label


0.jpg
locked
53.jpg
locked
asymetric
banded
asymetric
52.jpg
butterfly
banded
banded
55.jpg
butterfly
asymetric
48.jpg
butterfly
banded
banded
banded
banded
banded

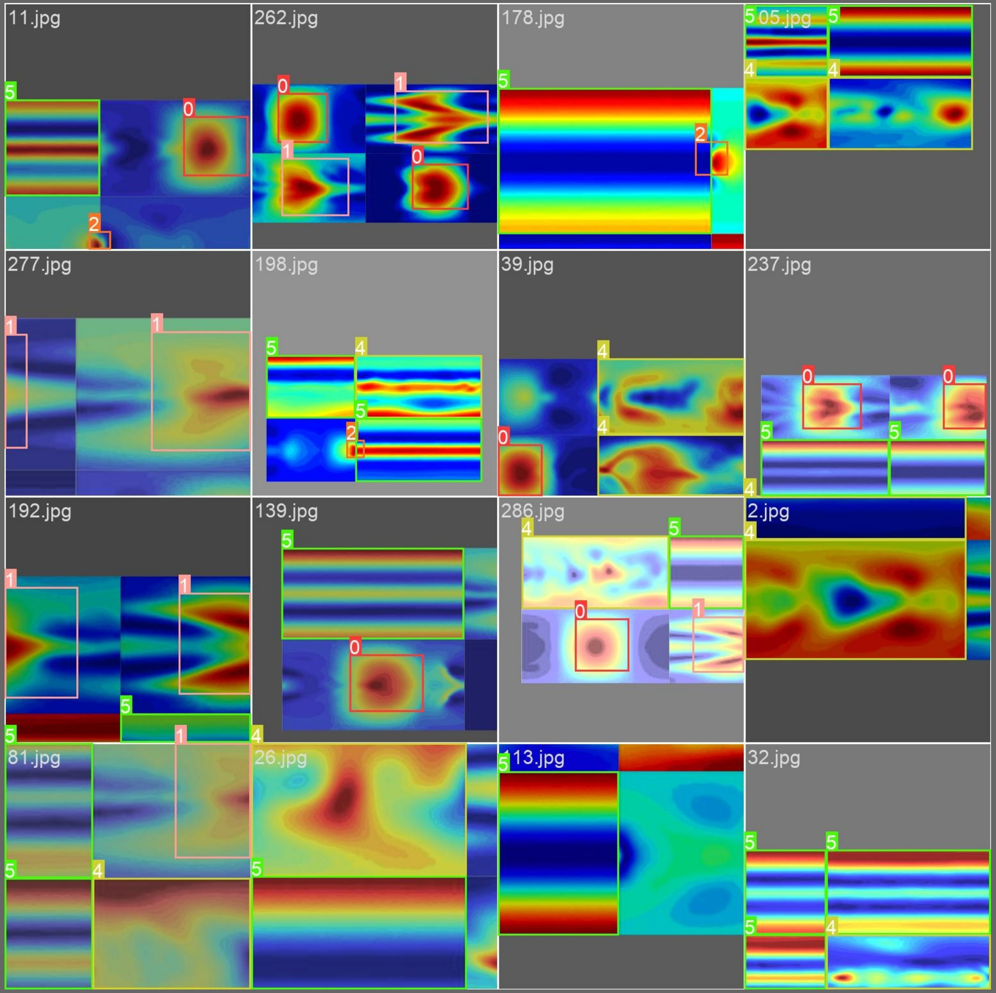
Predictions



Yolo internals

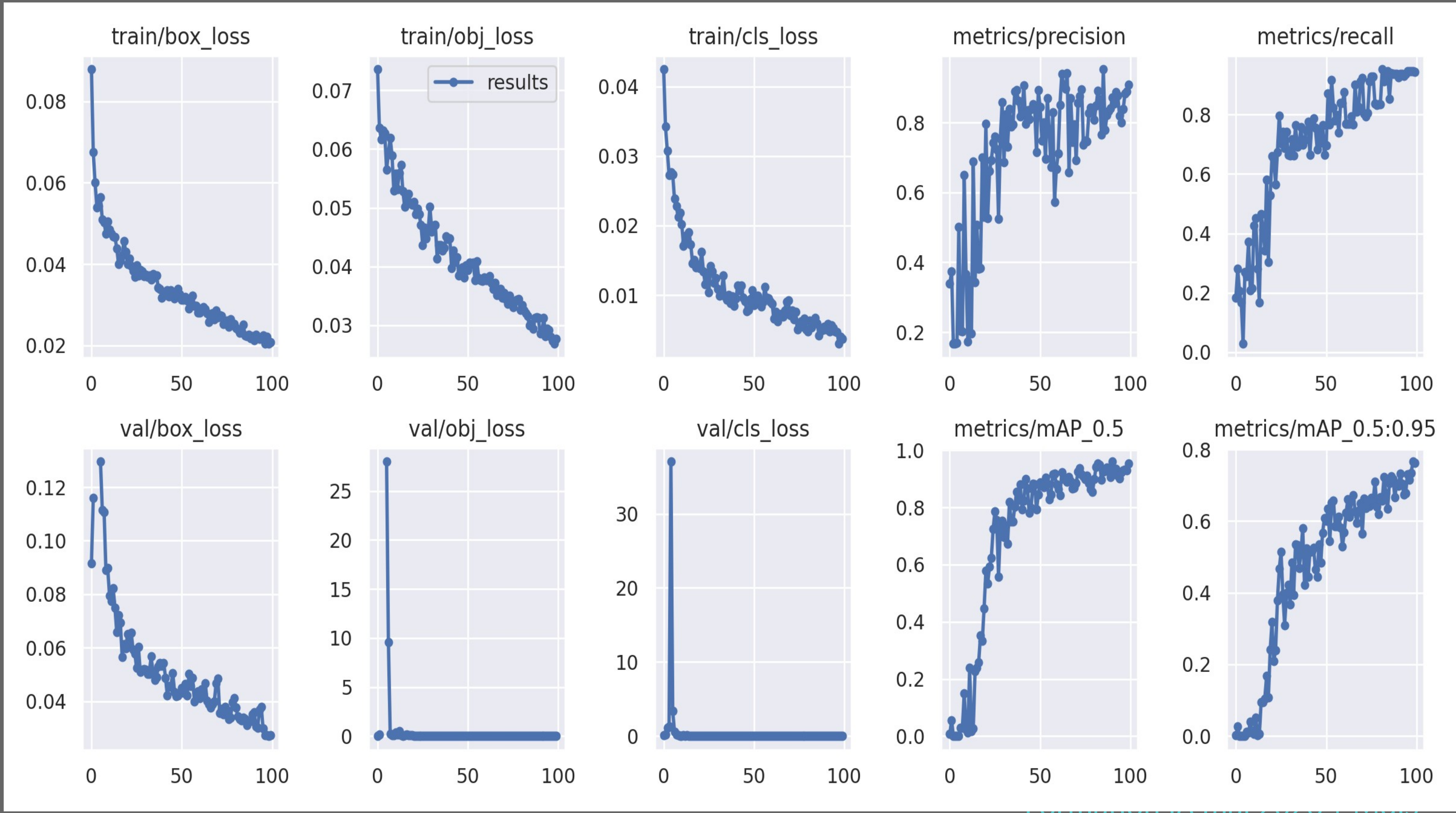
Internal oversampling for Yolo algo

A way to multi-resolution learning : usability with HPC simulations and parallelism ?



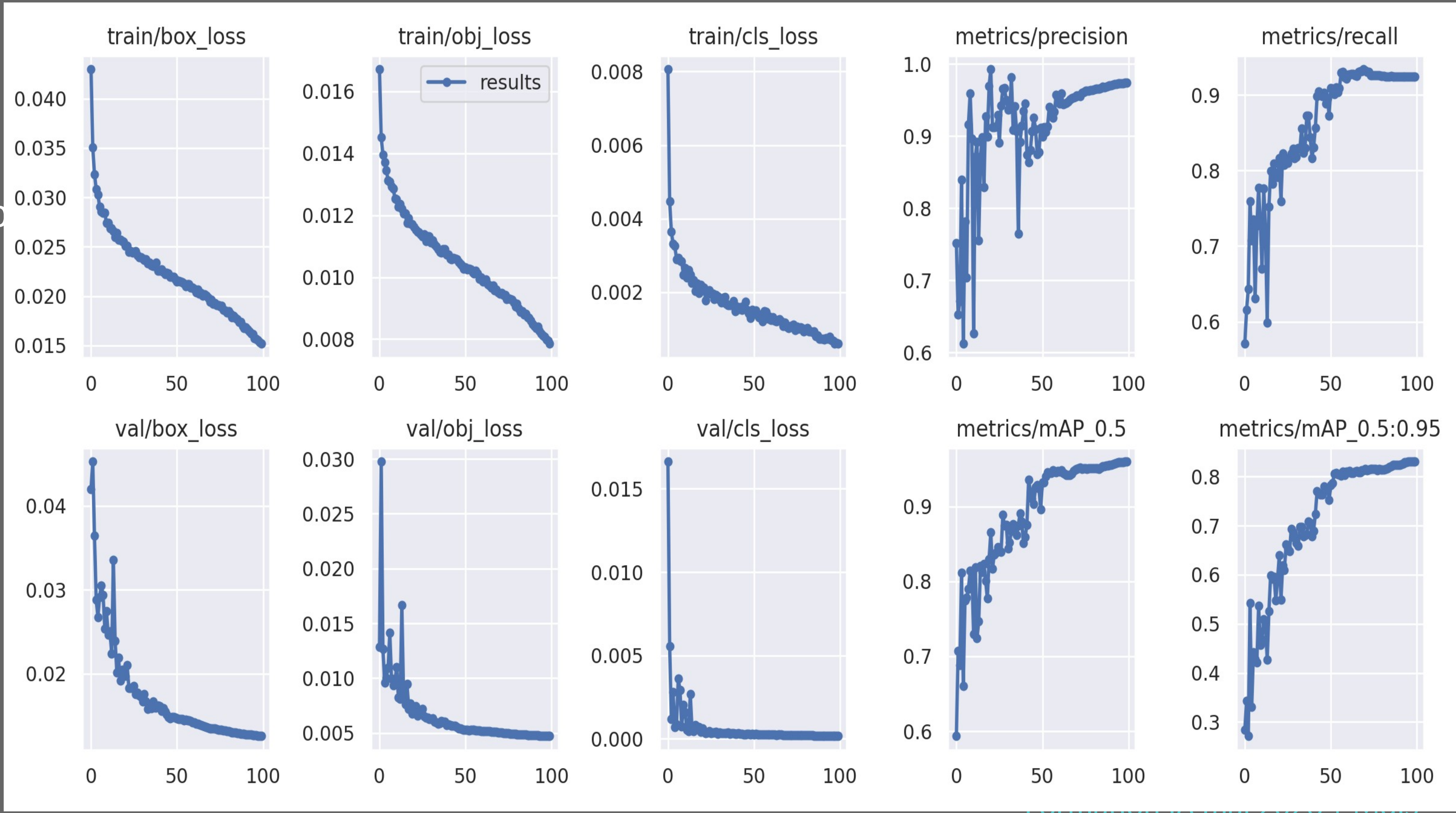
Yolo convergence

Internal oversampling for Yolo algo



Yolo convergence

With first « physical » oversampling before Yolo algo



Conclusion

AI with computer vision technics gives good results in HPDA.

- We have produced maps of events for ensemble of simulations.
- We have build many AI and validate with visualizations.
- The algorithm is abble to help to detect new physical features with big data.
- We are able to give bounding box prediction with Yolo v5.

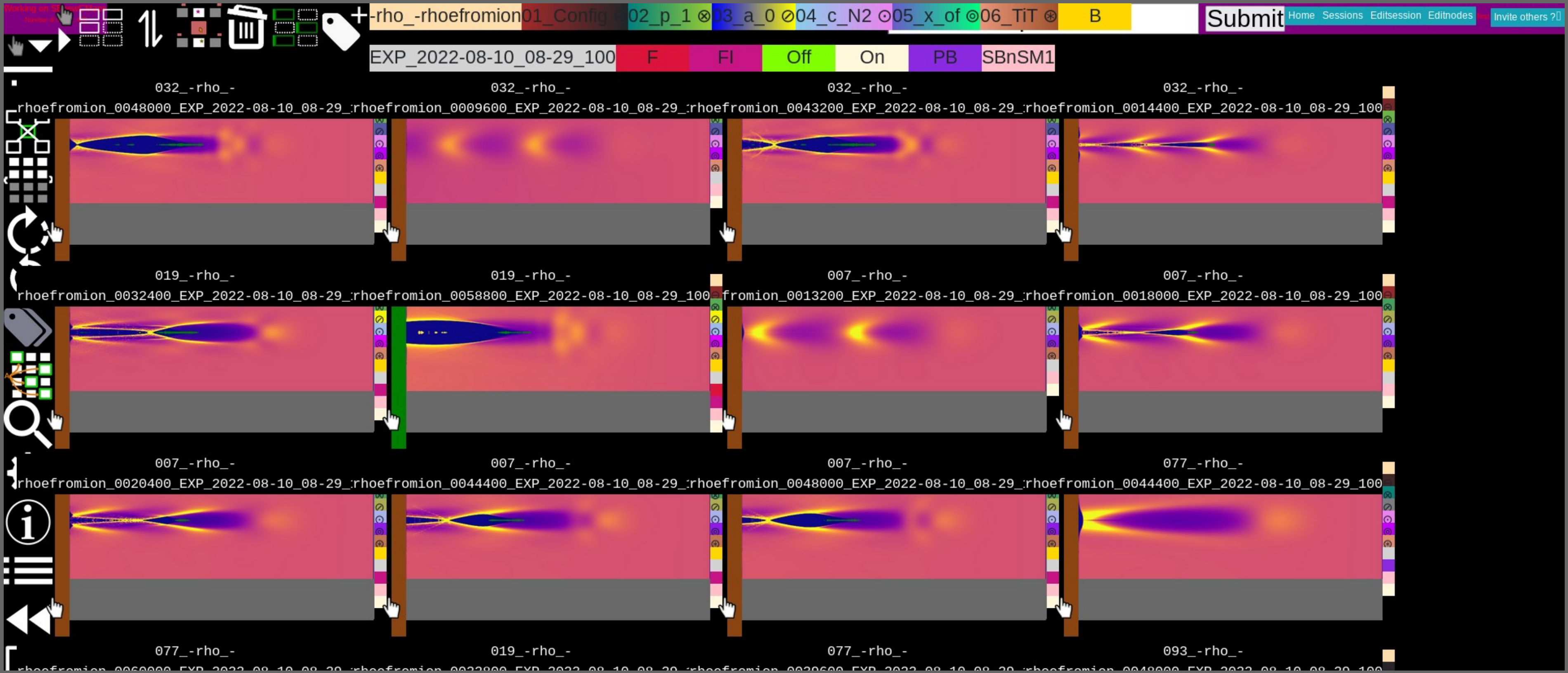
Conclusion

Improvement : New projects ?

- Optimization of IA with colormaps :
  detect best one in progress.
- In 3D ? 3D rendering ? Learn on slices only ?
- In-situ AI-BB with Paraview-Catalyst / Cinema
  to monitor HPC simulations asynchronously :
  - Add new post-treatements in-situ after detections.
  - Add segmentation to localize event and force high frequency outputs in bounding boxes.

Smilei New case : AIDELESS project

First tests with beam and injection and one bubble type.



Smilei New case : AIDELESS project

MAISON DE LA SIMULATION

Laser-plasma PIC simulations : localizations