

photo cred: giphy



Deep learning of astrophysical features with big data - Maggie Lieu

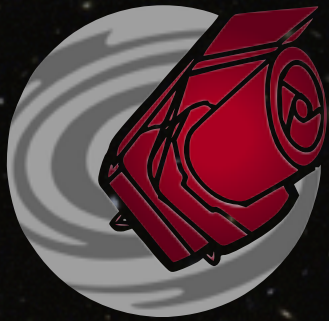
Bruno Merin, Deborah Baines, Fabrizio Giordano, Christophe Arviset
Bruno Altieri, Luca Conversi, Benoit Carry

 @space_mog





MACHINE LEARNING



euclid

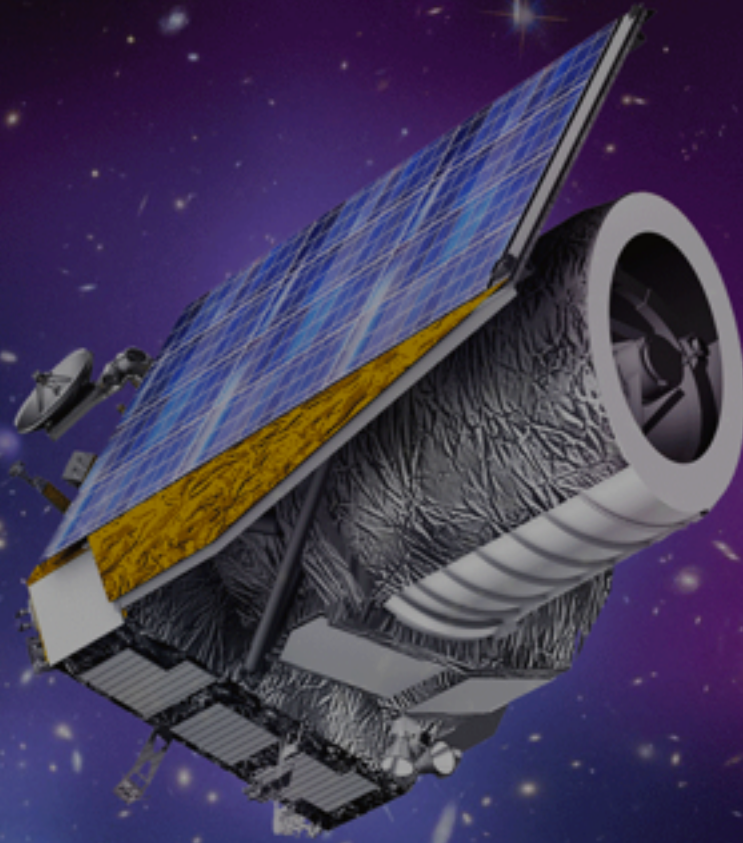


photo cred: LoCuSS

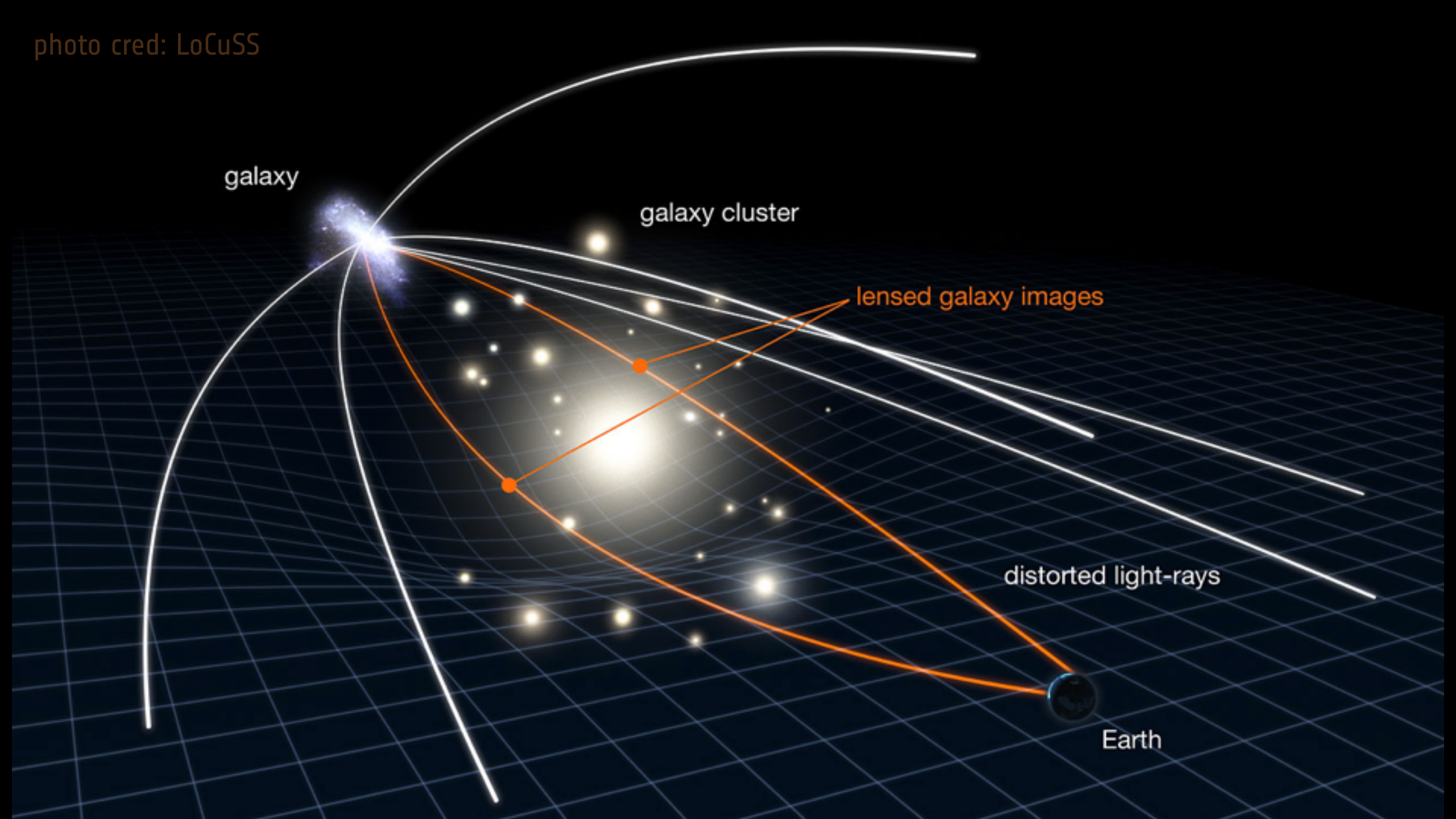
galaxy

galaxy cluster

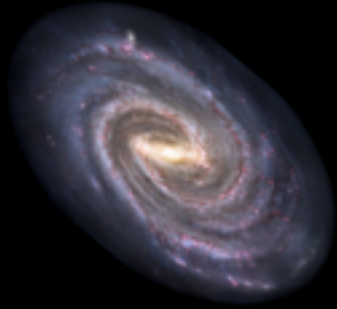
lensed galaxy images

distorted light-rays

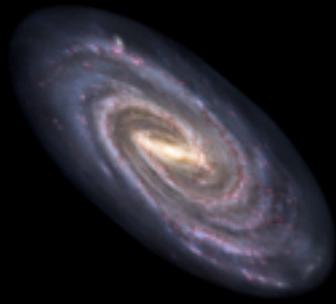
Earth



real galaxy



shear



atmosphere &
telescope
blur



pixelised by
detectors



noise



photo cred: CFHT

Raw Euclid **data is nasty!** -
noise, cosmic rays, ghosts, CTI,
stars, etc.

Asteroids are **big contaminants**
of weak lensing signal

Euclid will be producing ~800GB
of data a day. **Traditional**
methods not efficient!

See Lieu+18

What is machine learning?



photo cred: pinktentacle

ML:
automated model building to
map an input to an output.

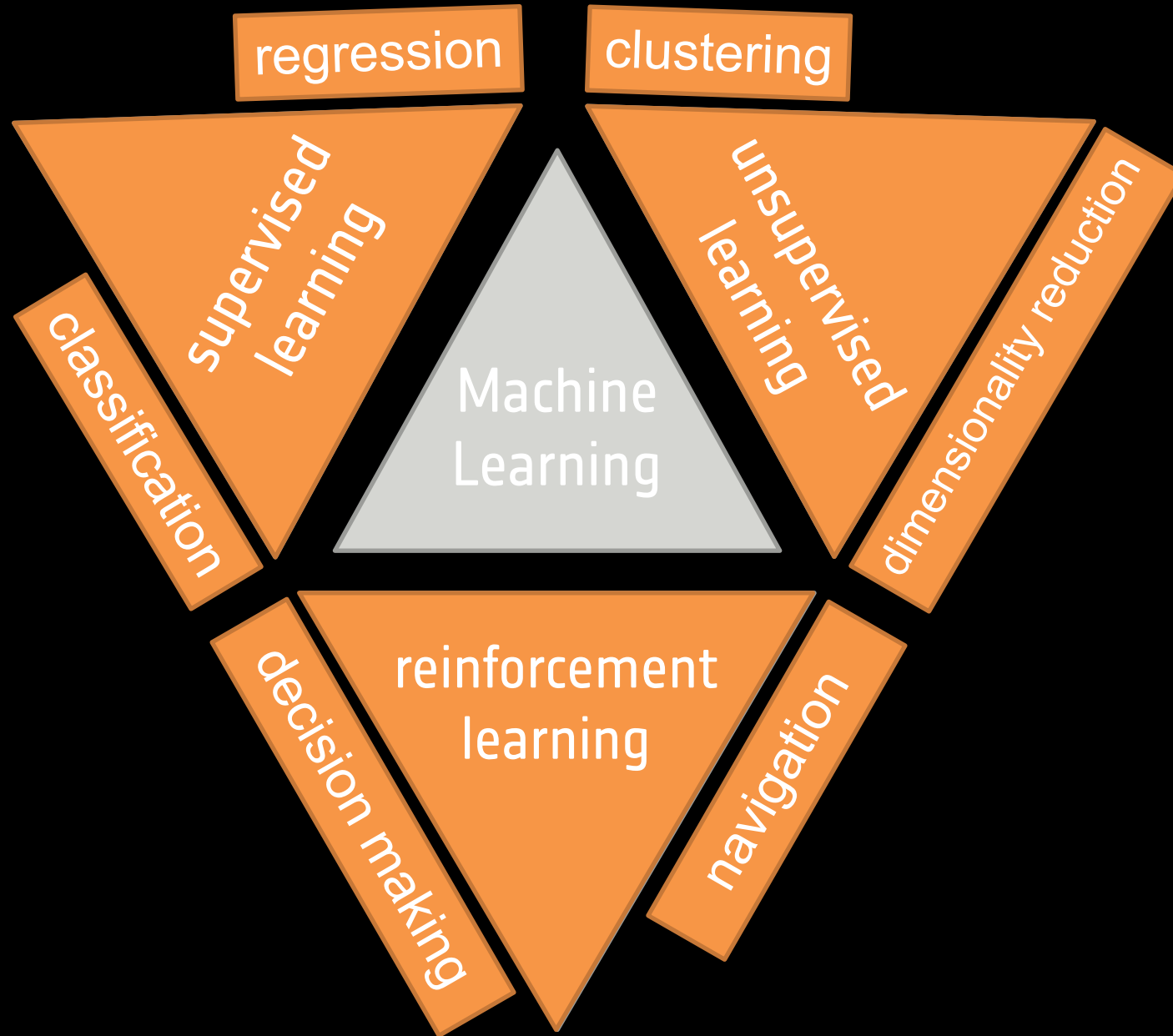
Humans:

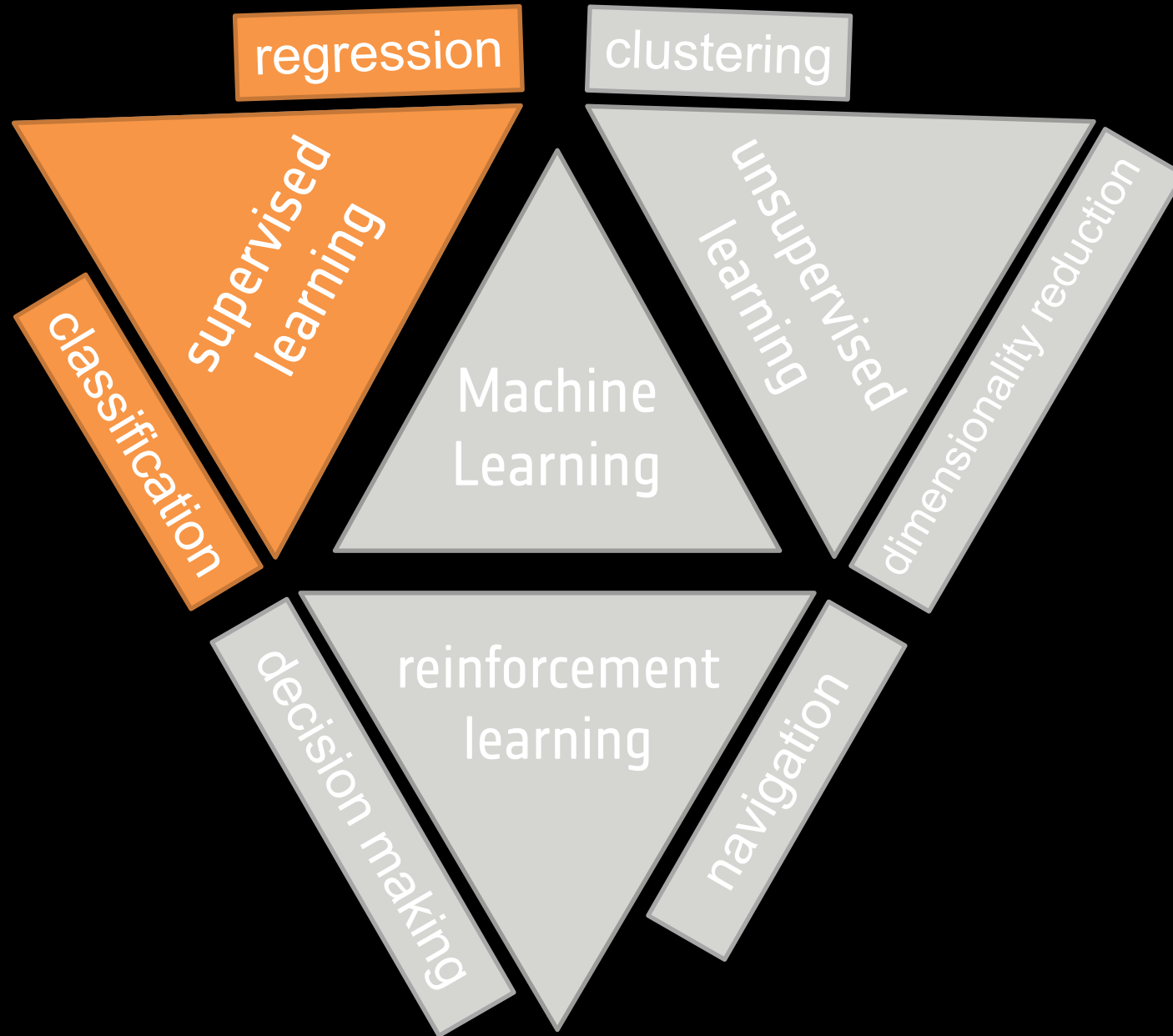
- slow
- inconsistent
- unreliable

Machines:

- improve productivity
- cheap
- logical

Machine learning is your friend!





Intro to machine learning

Automatic model building to map input to output

1
2
3
4
5

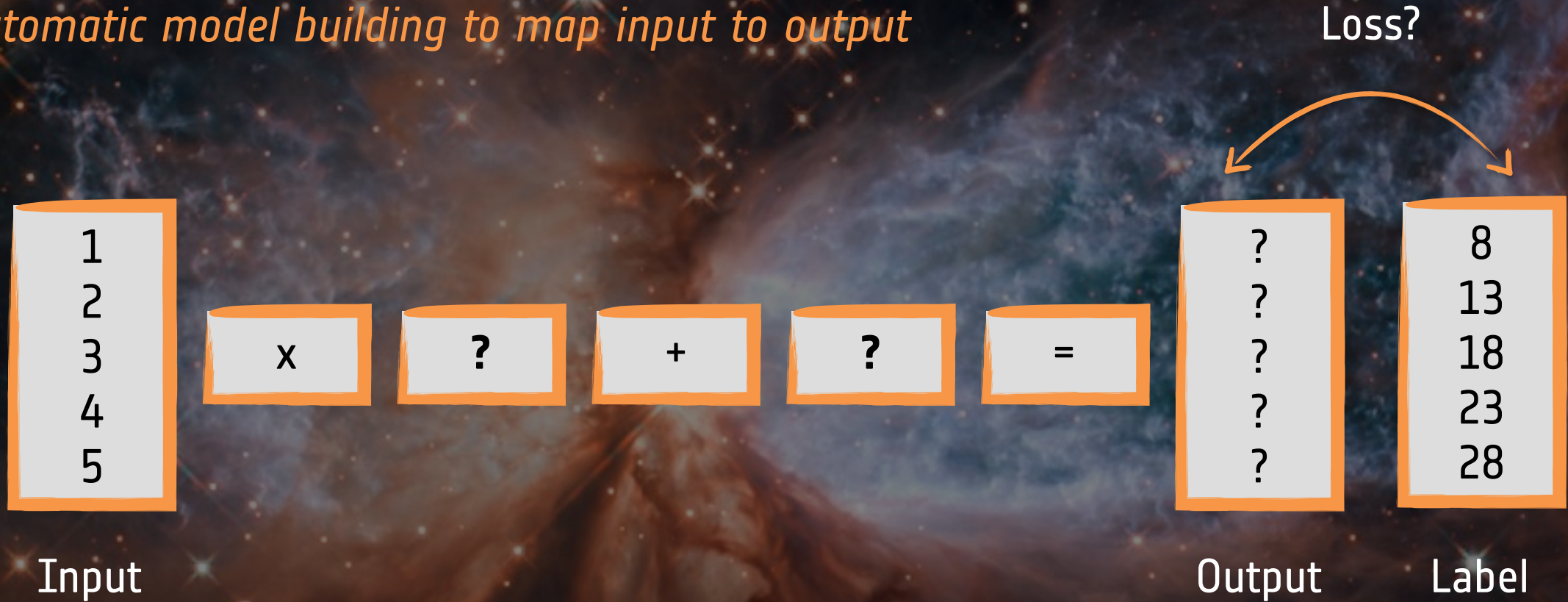
Input

8
13
18
23
28

Label

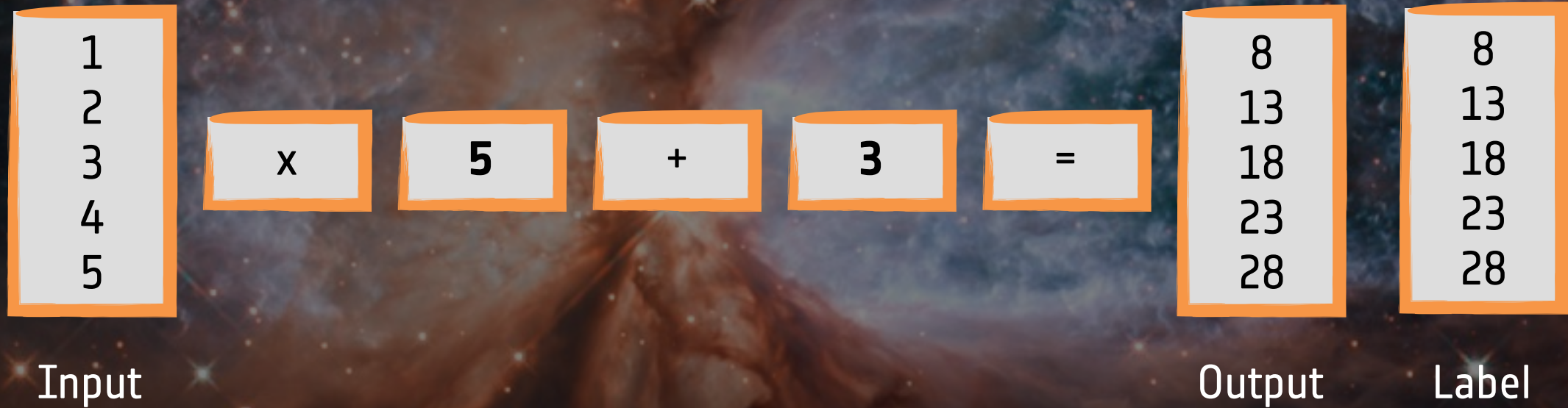
Intro to machine learning

Automatic model building to map input to output

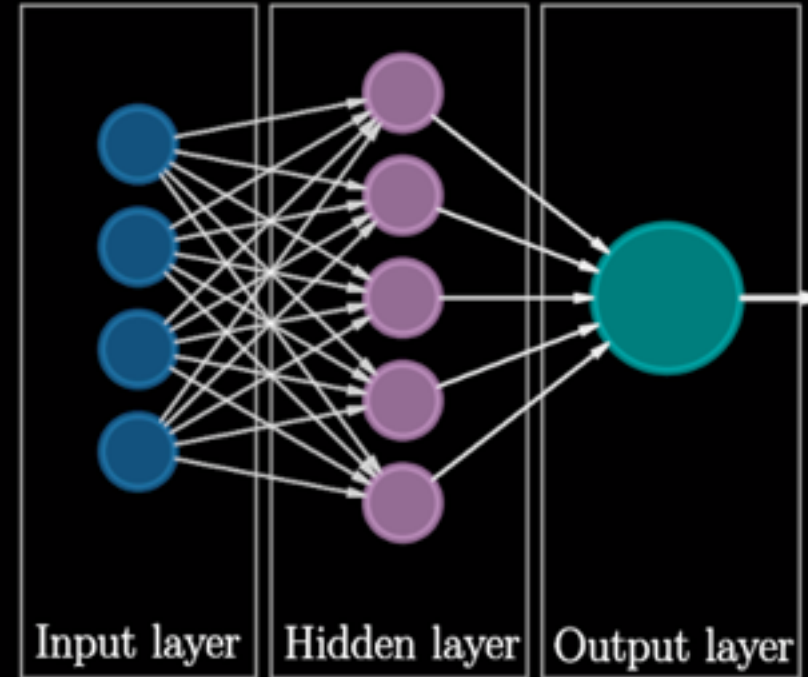
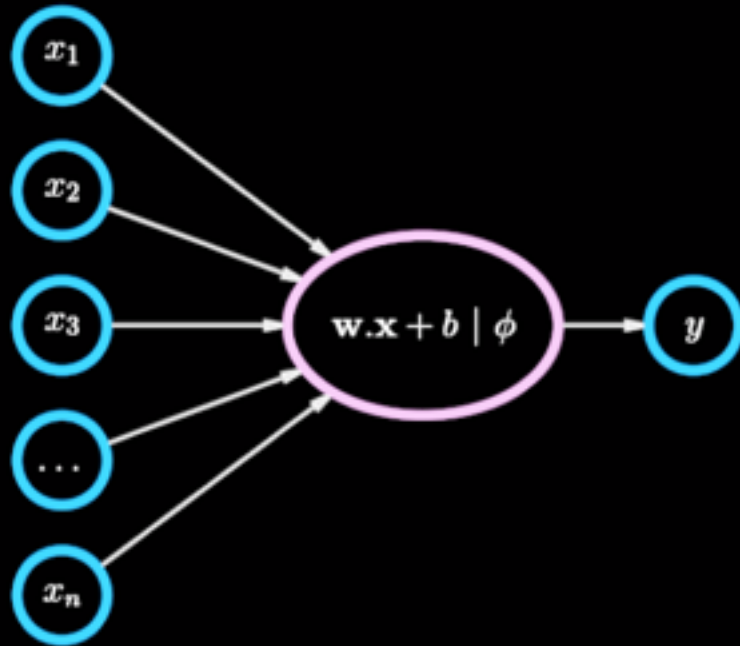


Intro to machine learning

Automatic model building to map input to output

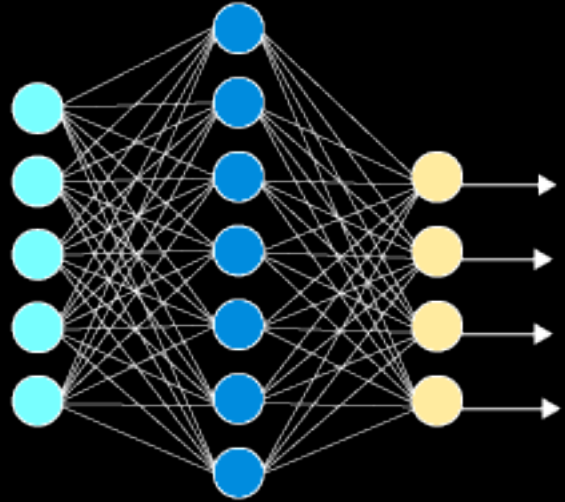


Artificial neuron

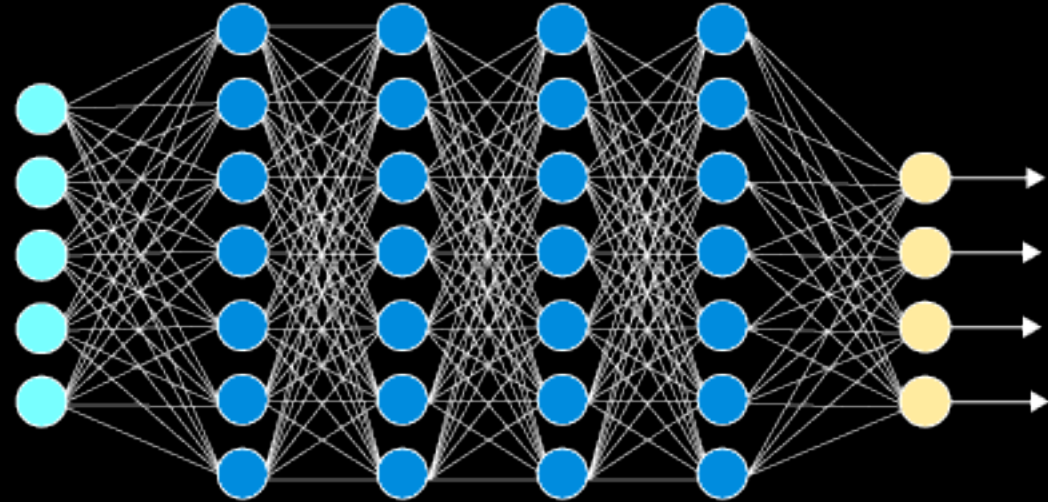


Deep learning

Simple Neural Network



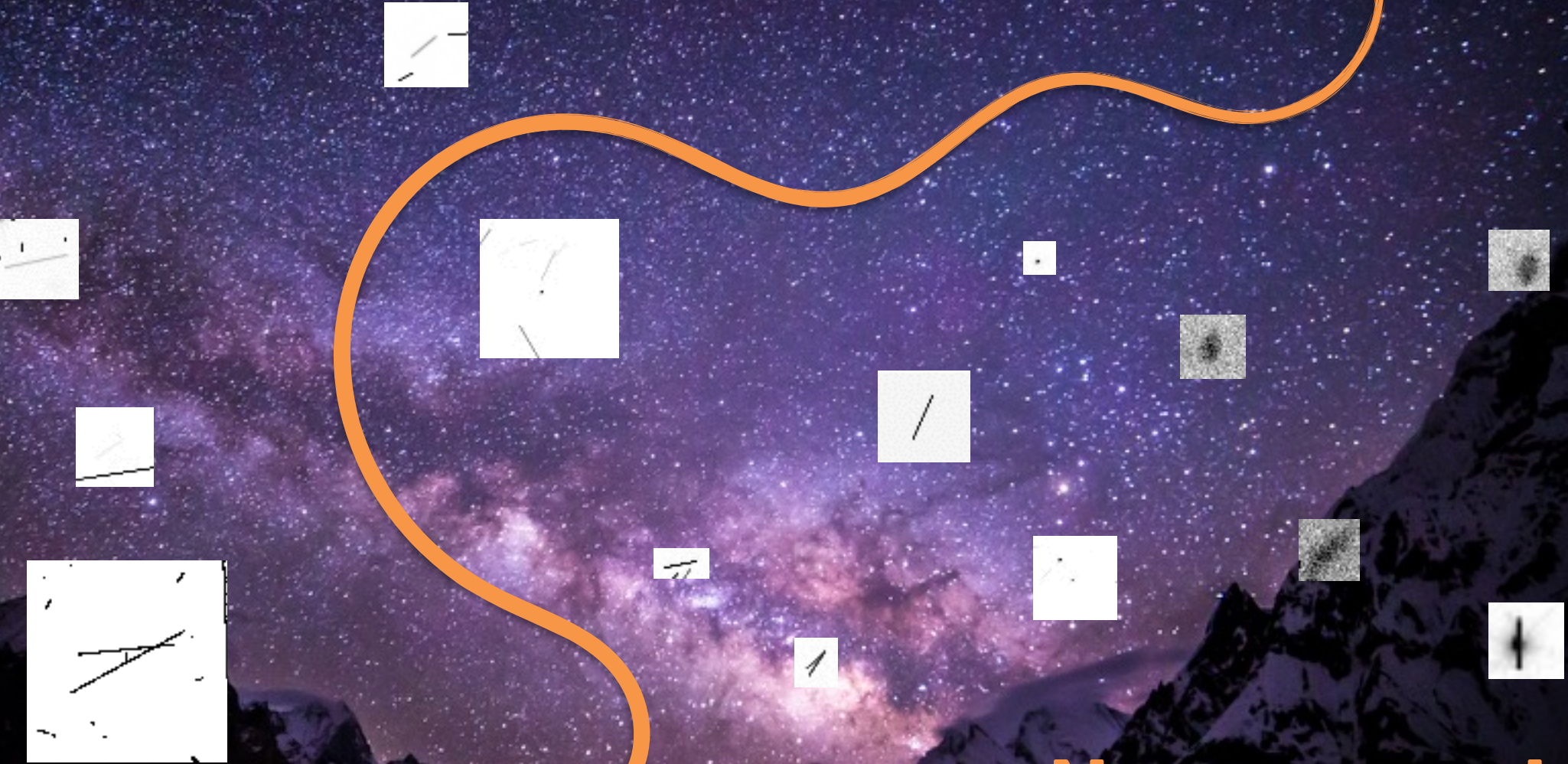
Deep Learning Neural Network



● Input Layer ● Hidden Layer ● Output Layer

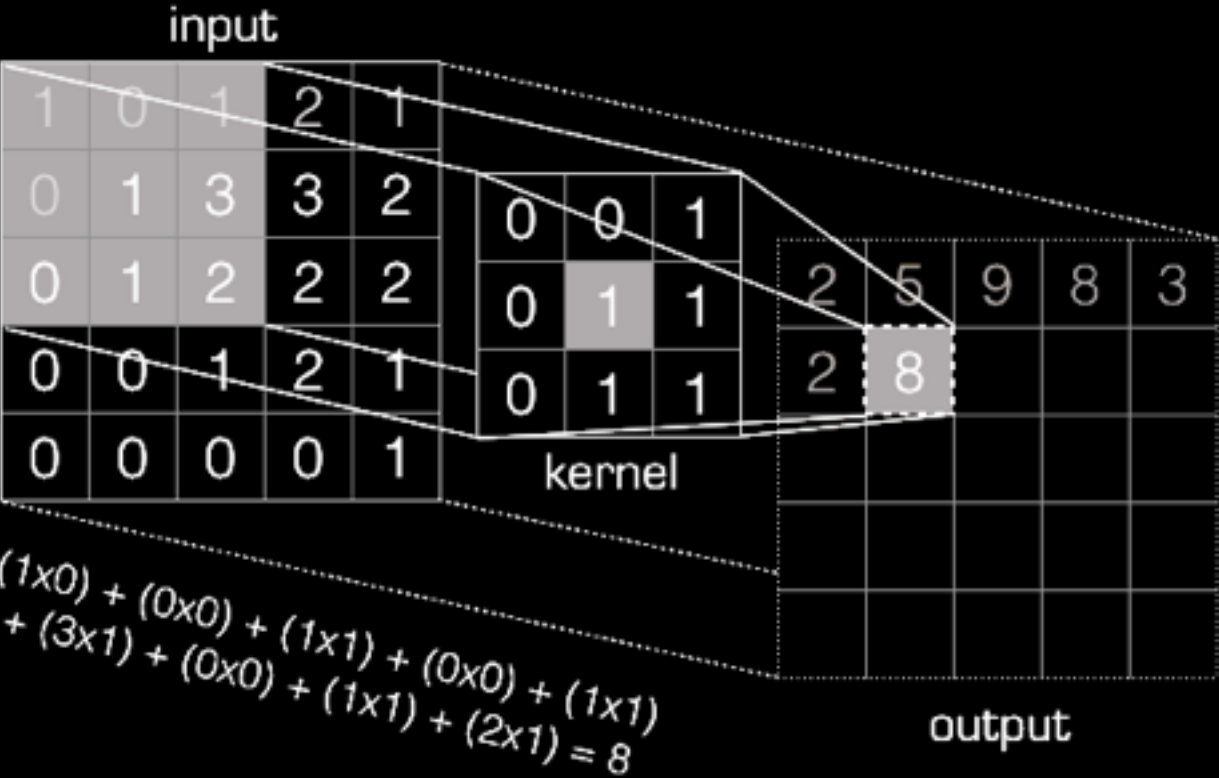
Multiple hidden layers

Asteroids

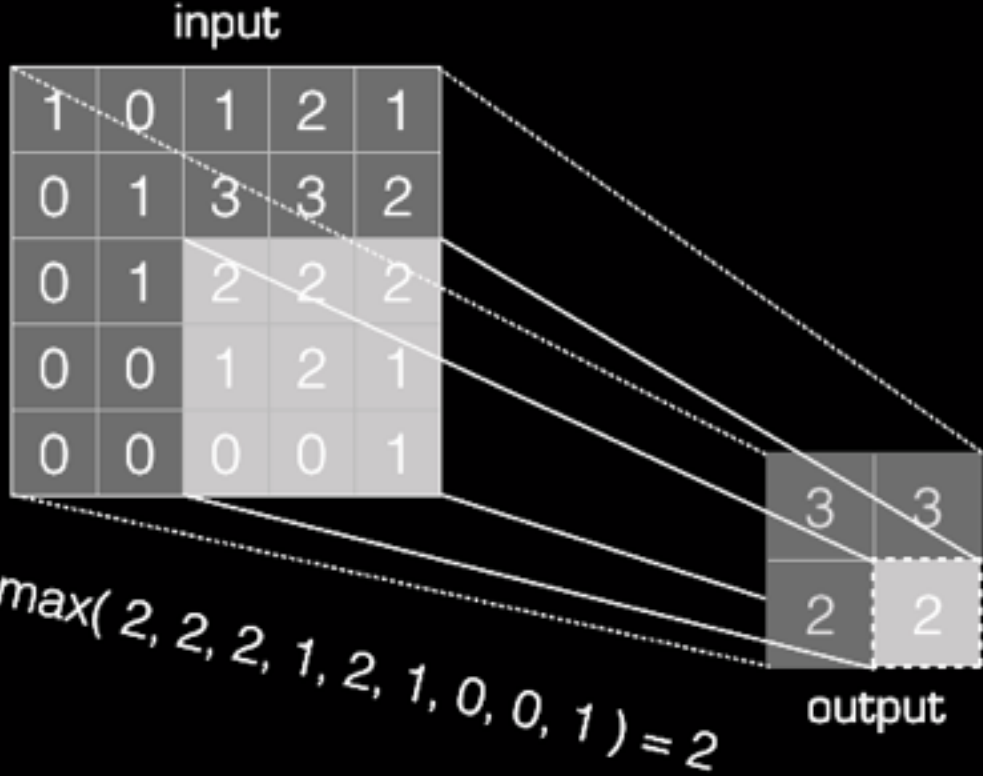


Non-asteroids

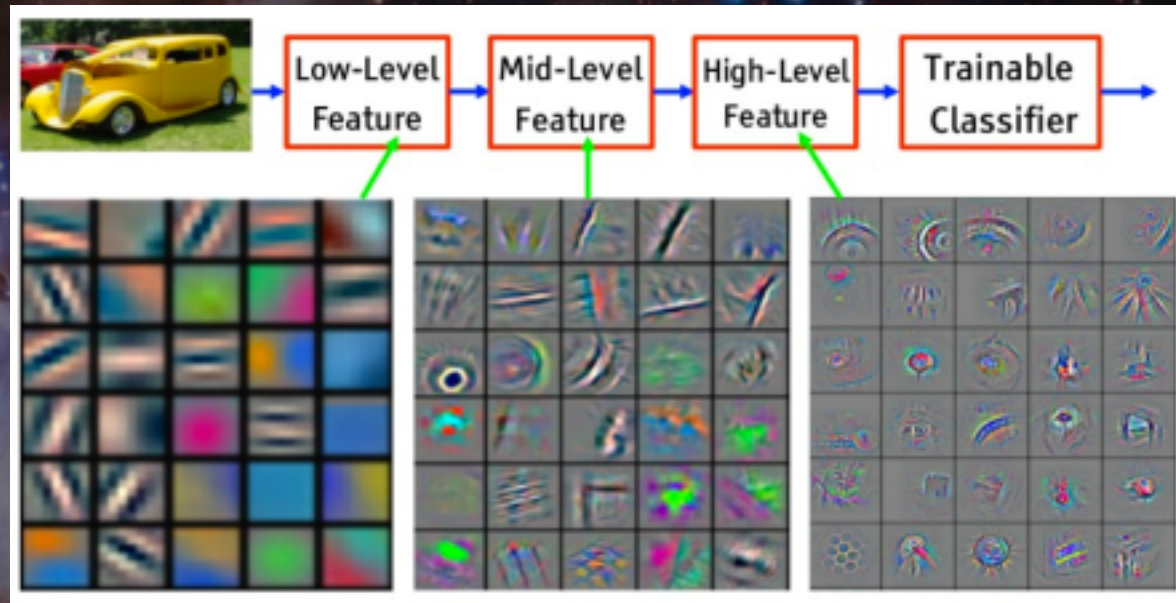
Convolutional Neural Nets



Convolutional layer



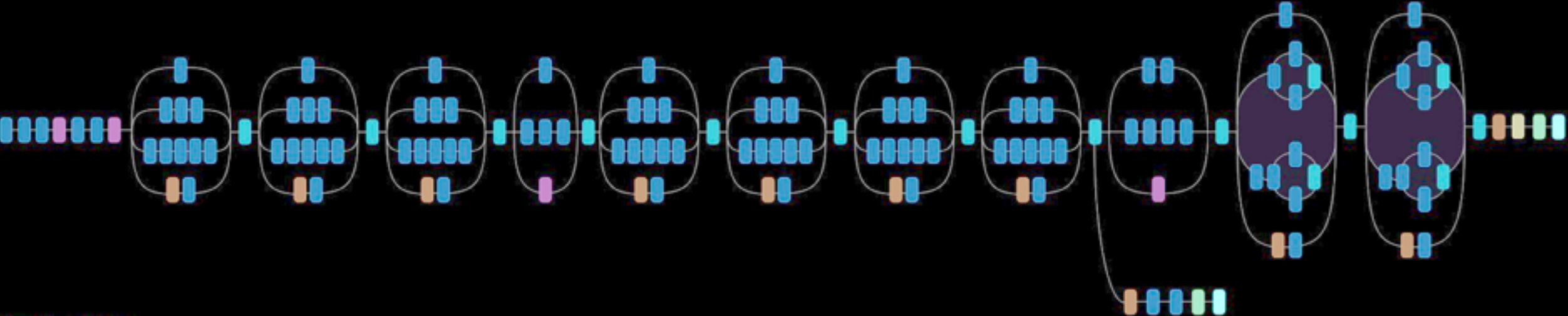
Pooling layer



0	1	0	0	0	1	0
0	1	1	0	1	1	0
0	1	1	1	1	1	0
0	1	2	1	2	1	0
0	1	1	2	1	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0



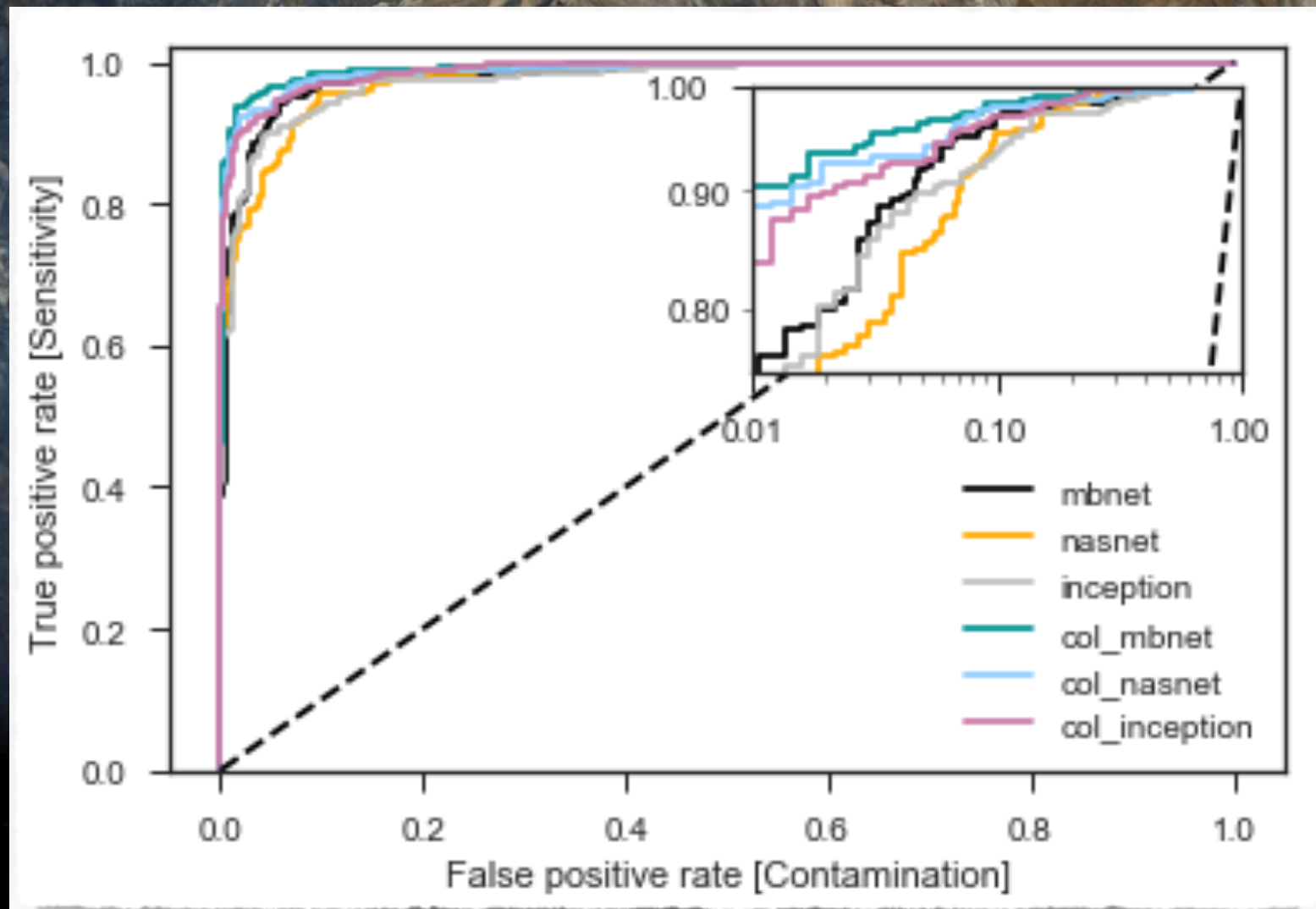
Transfer learning



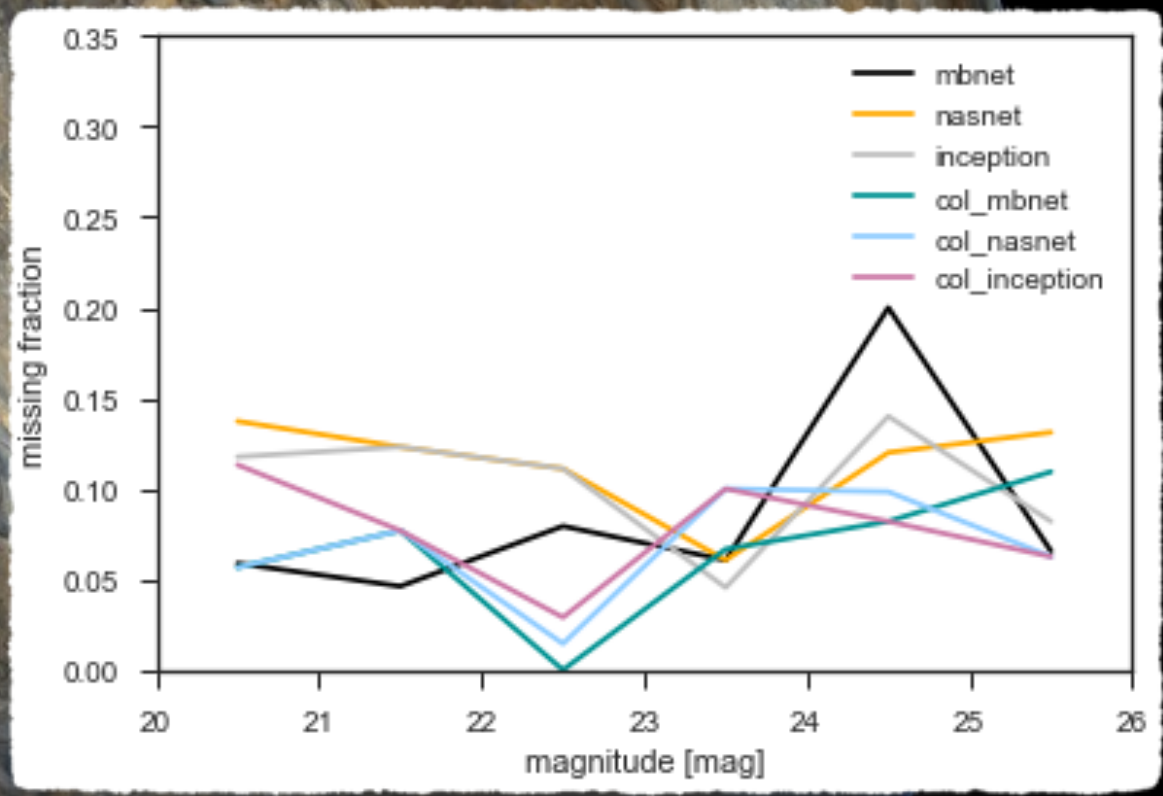
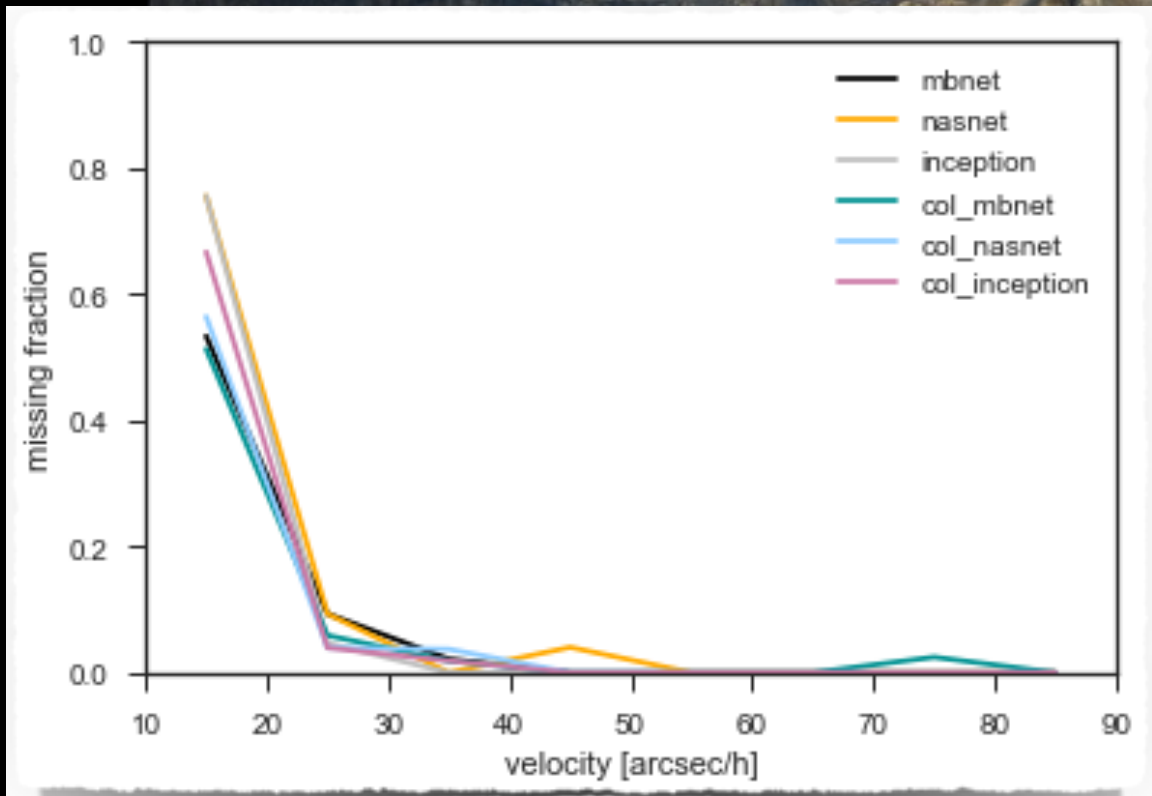
- Convolution
- AvgPool
- MaxPool
- Concat
- Dropout
- Fully connected
- Softmax



ROC curves



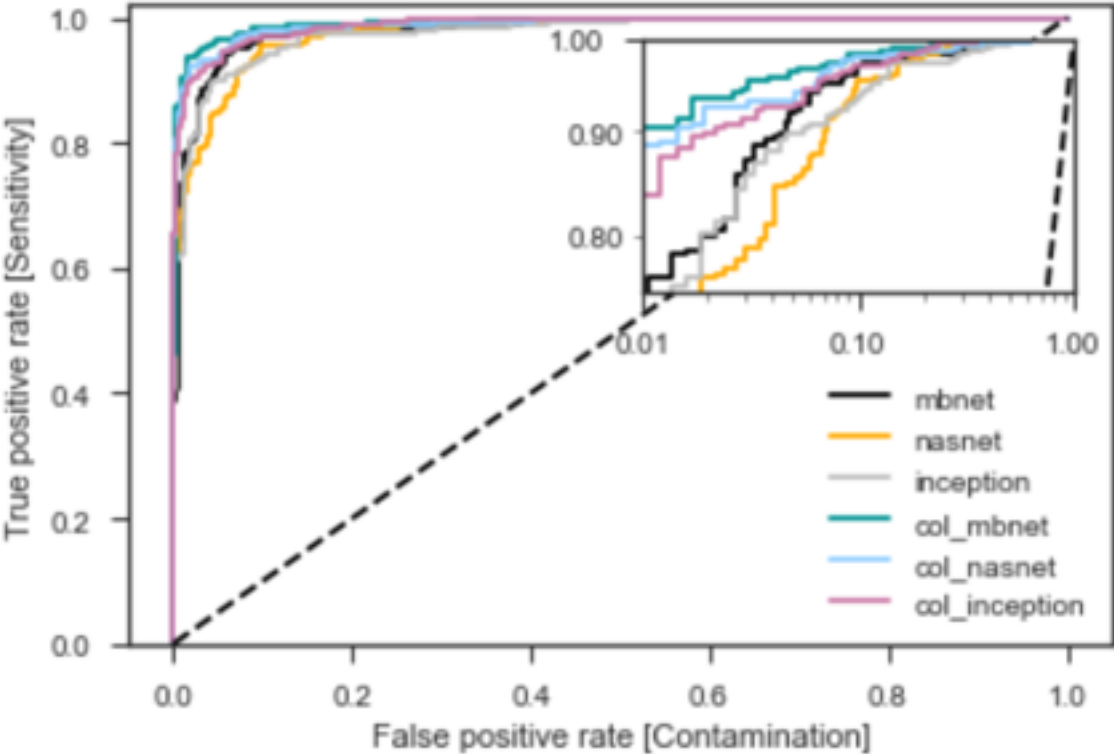
Completeness



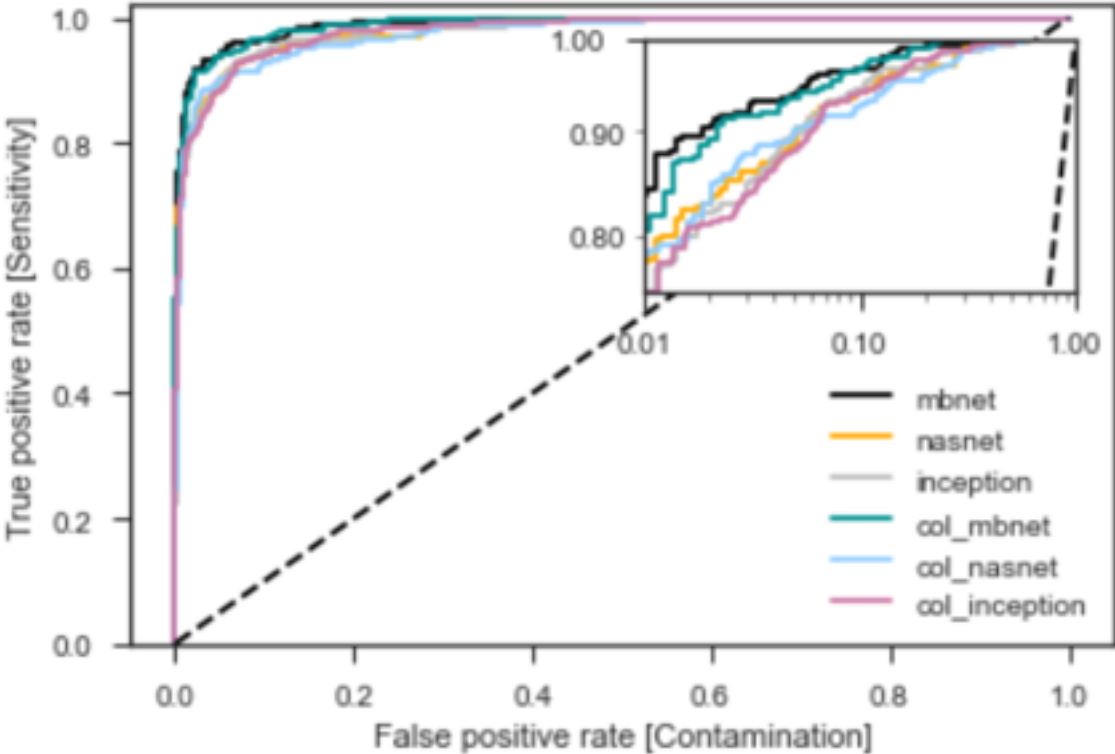
Low bias, high accuracy and super fast!

Not limited to binary classification

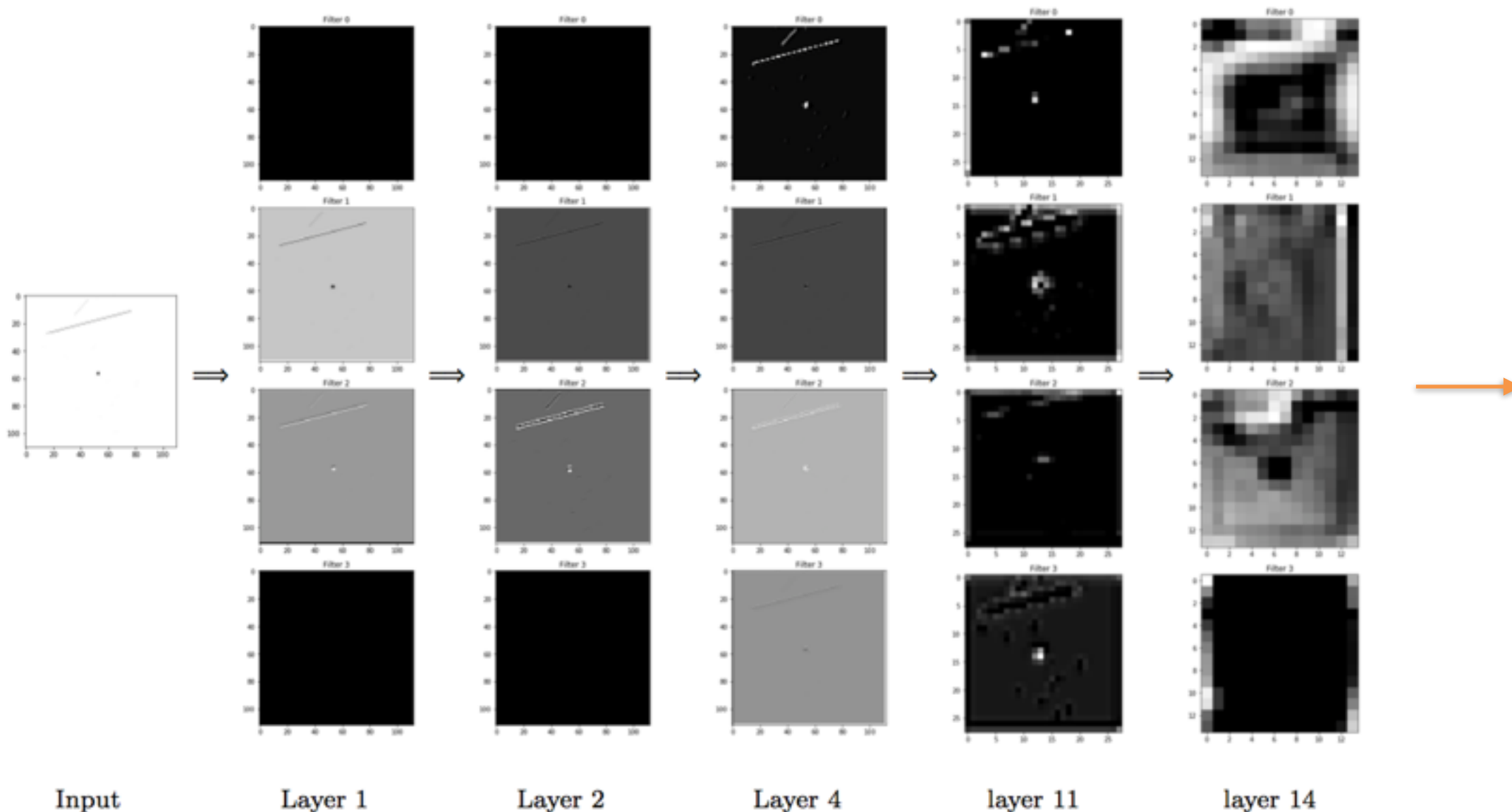
binary classification



4 classes: asteroids, stars, galaxies, CRs



Activation layers



Machine-learning shopping list:



lots of data

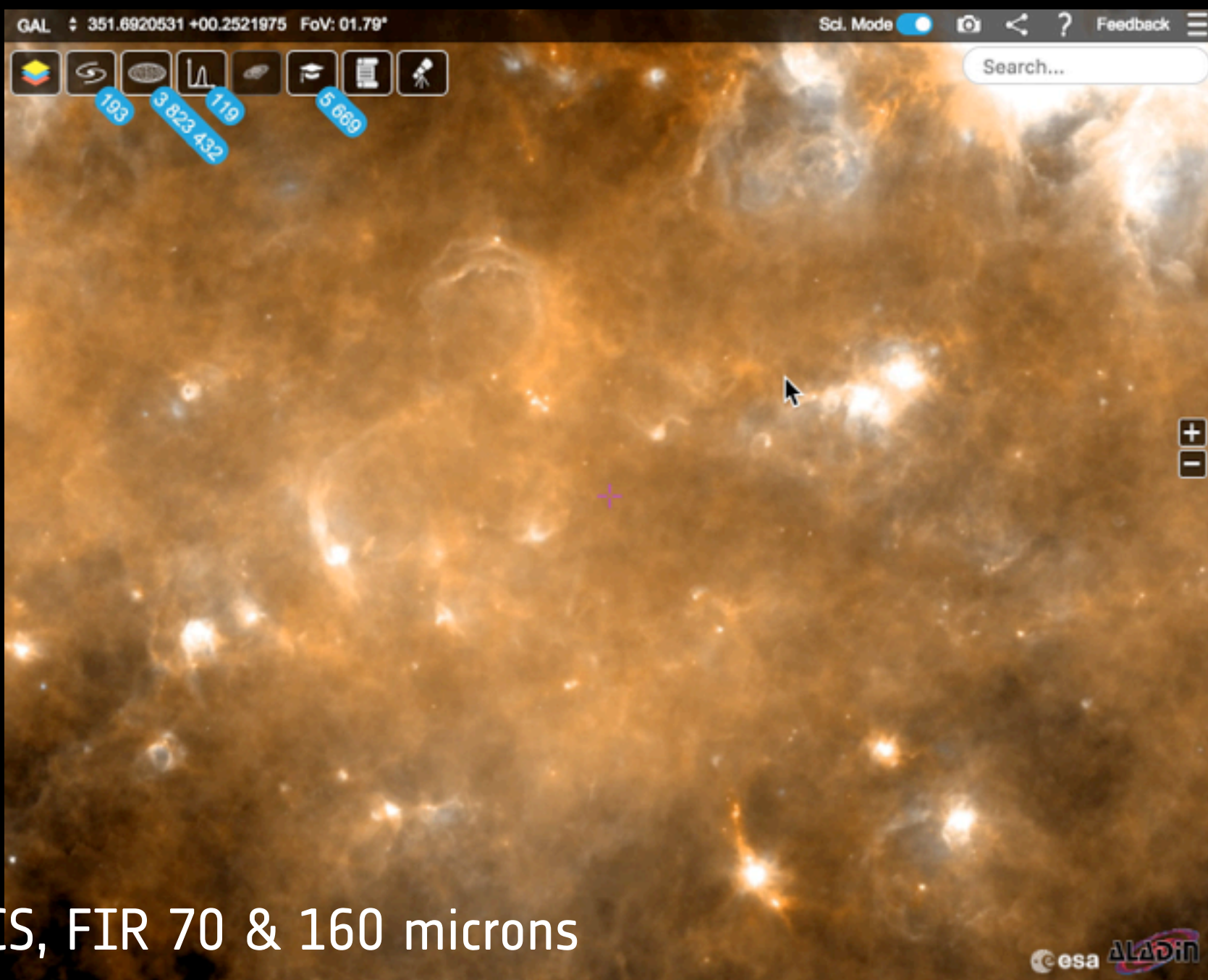


truth labels



computational
resources

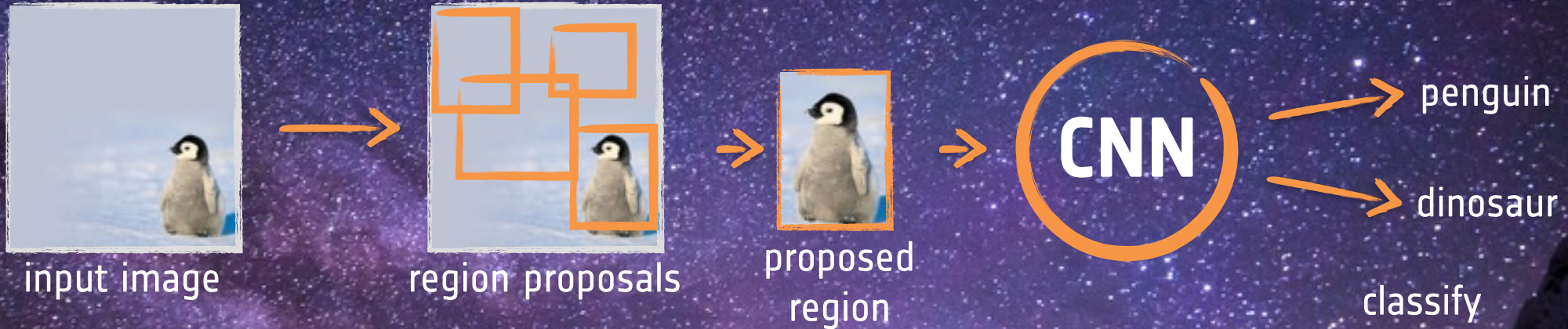
ESAC SCIENCE DATA CENTRE



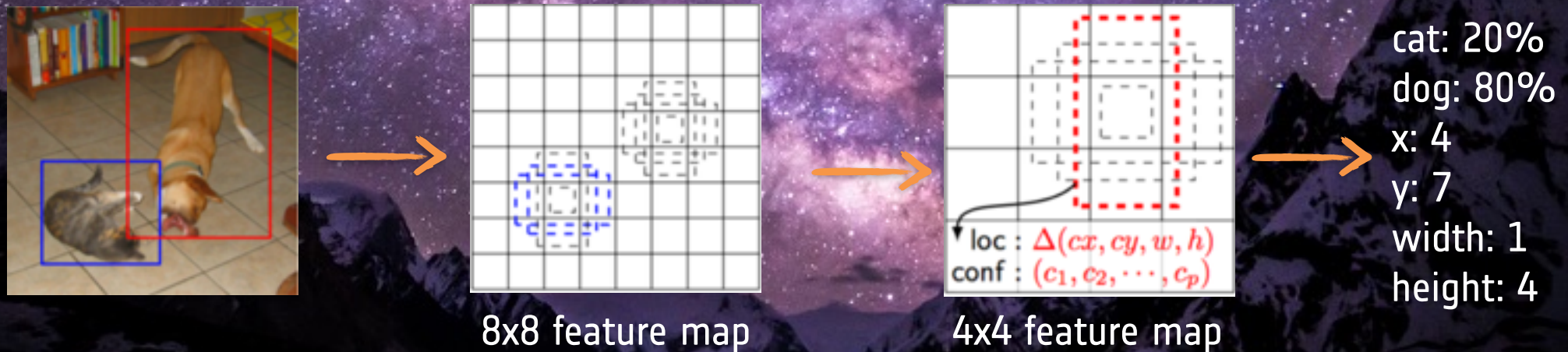
Herschel PACS, FIR 70 & 160 microns

Multi-object detection

- Classification and regression
- R-CNNs



- **Single shot detectors**



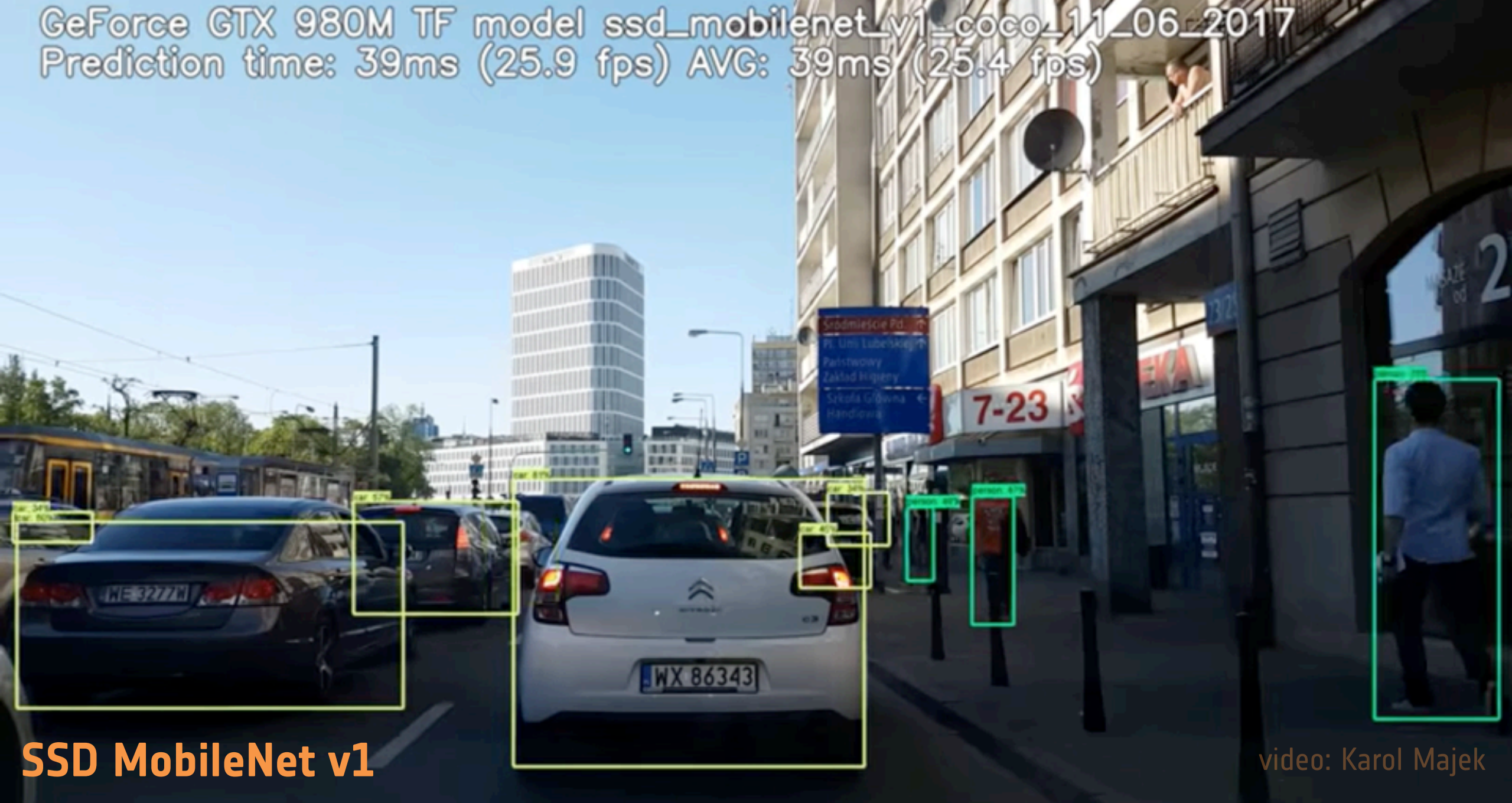


Microsoft coco dataset (Lin+2015):
91 classes
2.5Mn labels
300k images

Dataset examples



GeForce GTX 980M TF model ssd_mobilenet_v1_coco_11_06_2017
Prediction time: 39ms (25.9 fps) AVG: 39ms (25.4 fps)



SSD MobileNet v1

video: Karol Majek





April 23 ·  

Allowed on timeline 

Add a description With Brian May.

 Tag Photo  Add Location  Edit



 Like  Comment  Share



Write a comment...



People You May Know

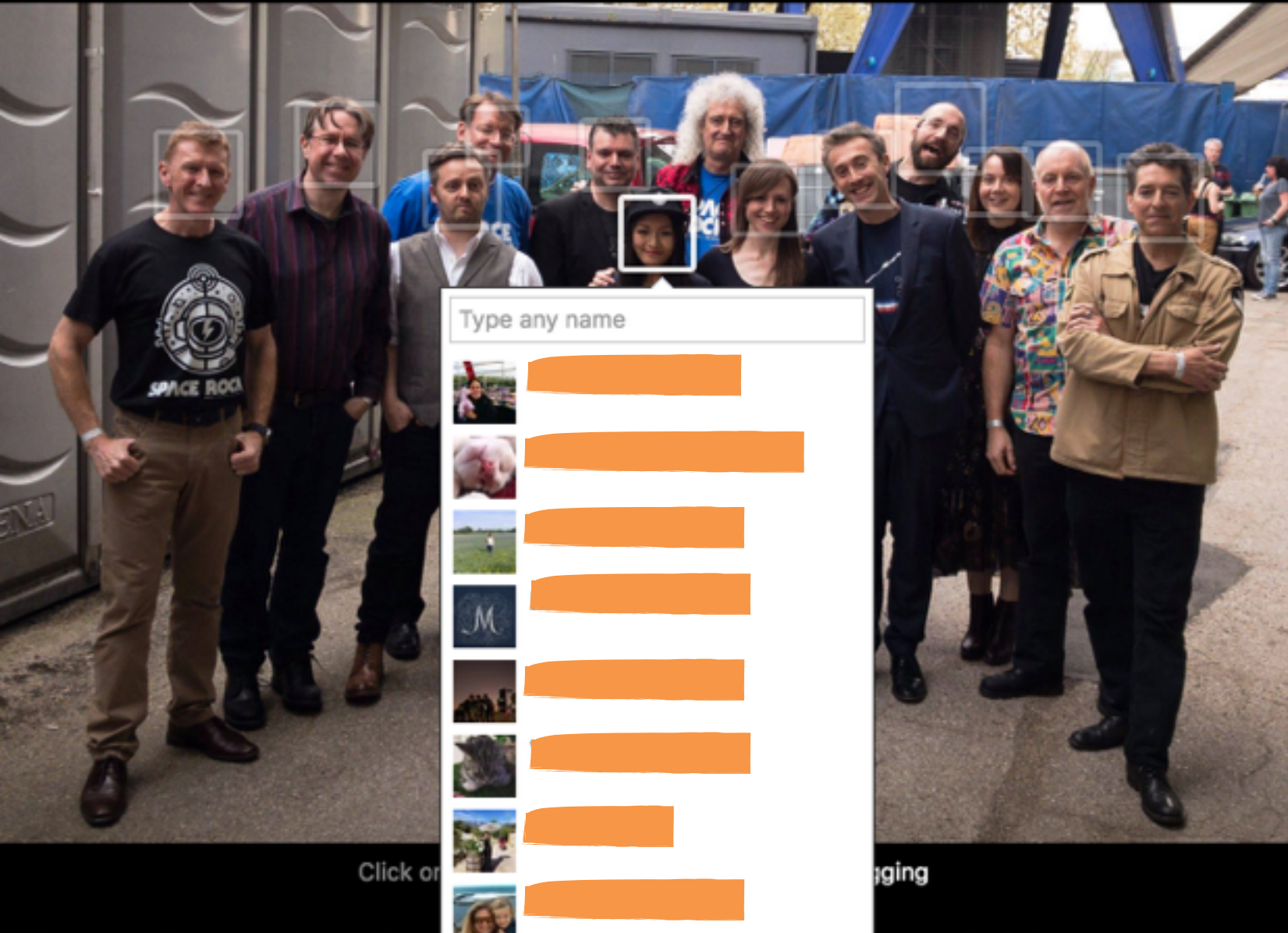
[See All](#)



Tara Marie

5 mutual friends

 Add Friend



[Redacted name]



April 23 · [Redacted privacy]

Allowed on timeline

Add a description With Brian May.

Done Tagging Add Location Edit



Like Comment Share



Write a comment...

People You May Know See All



Tara Marie
5 mutual friends



Space bubbles

HII regions incl. compact star forming regions

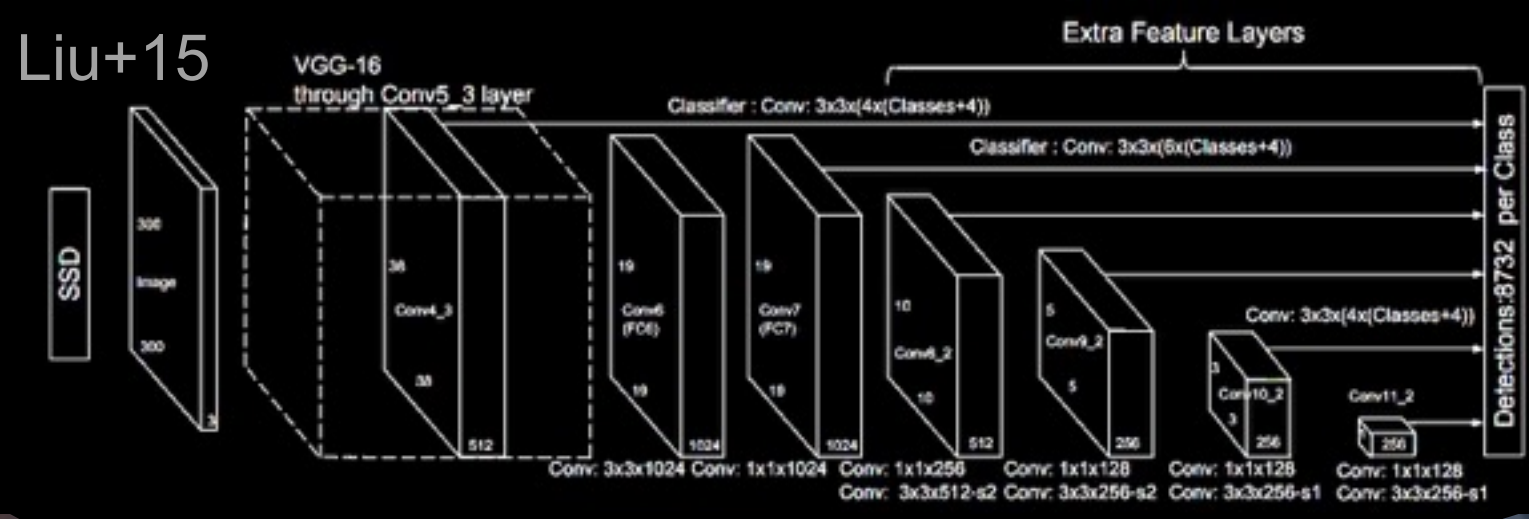


**Milky way project's Spitzer GLIMSE/
MIPSGAL images (Simpson+2012): boxes
& classifications**



Spitzer MIR, IRAC

Liu+15



Labels

MWP boxes & classes



Input

Herschel PACS images



trained on MSCOCO

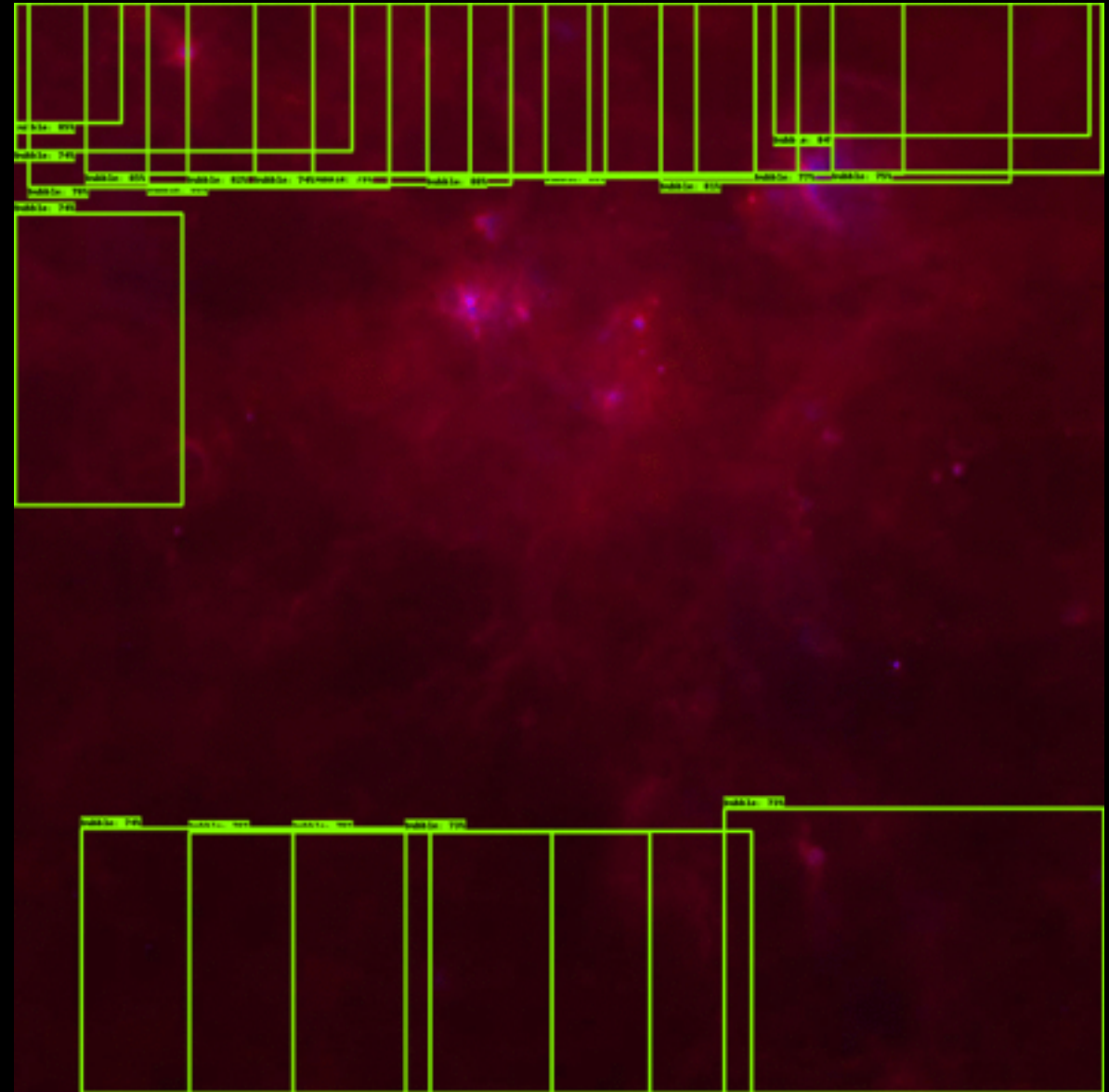
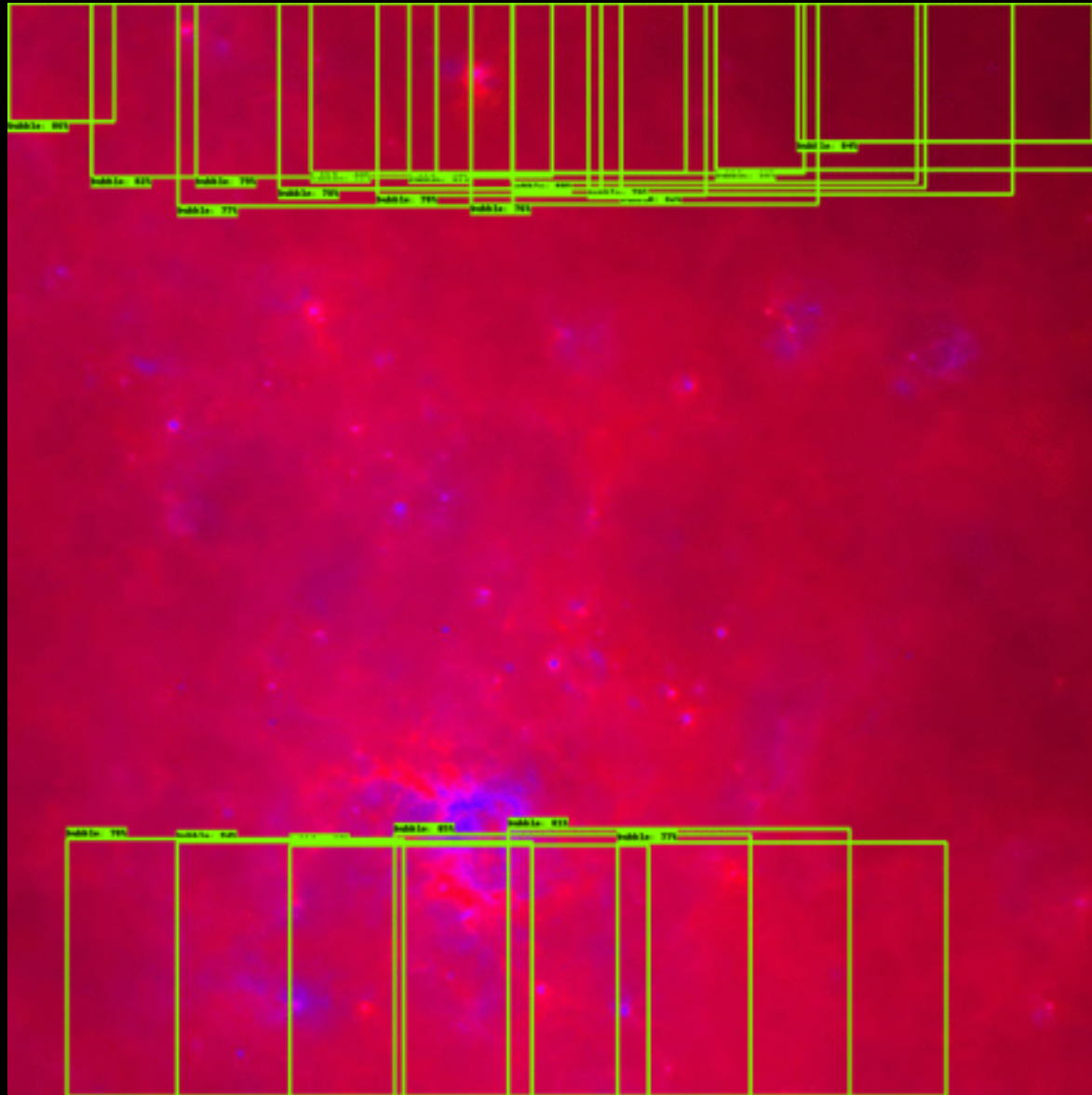


Output

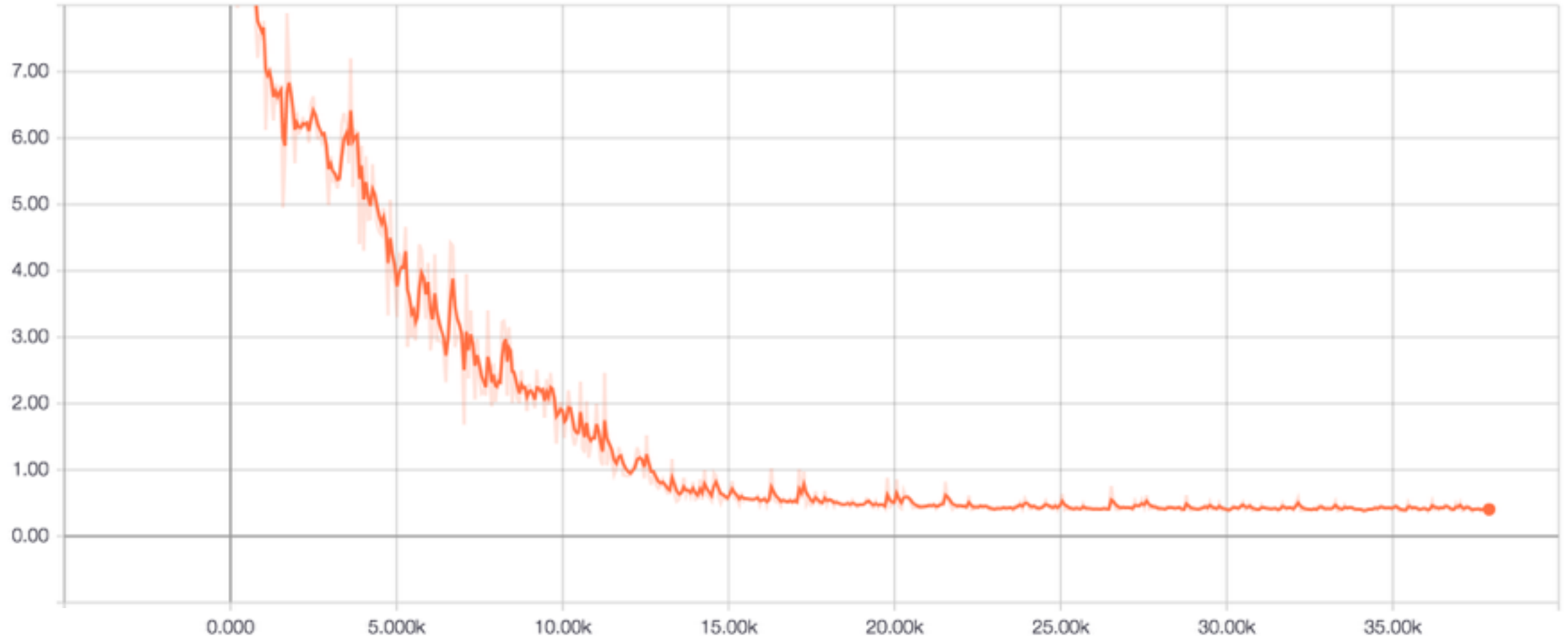


loss

Training in action as seen on validation data

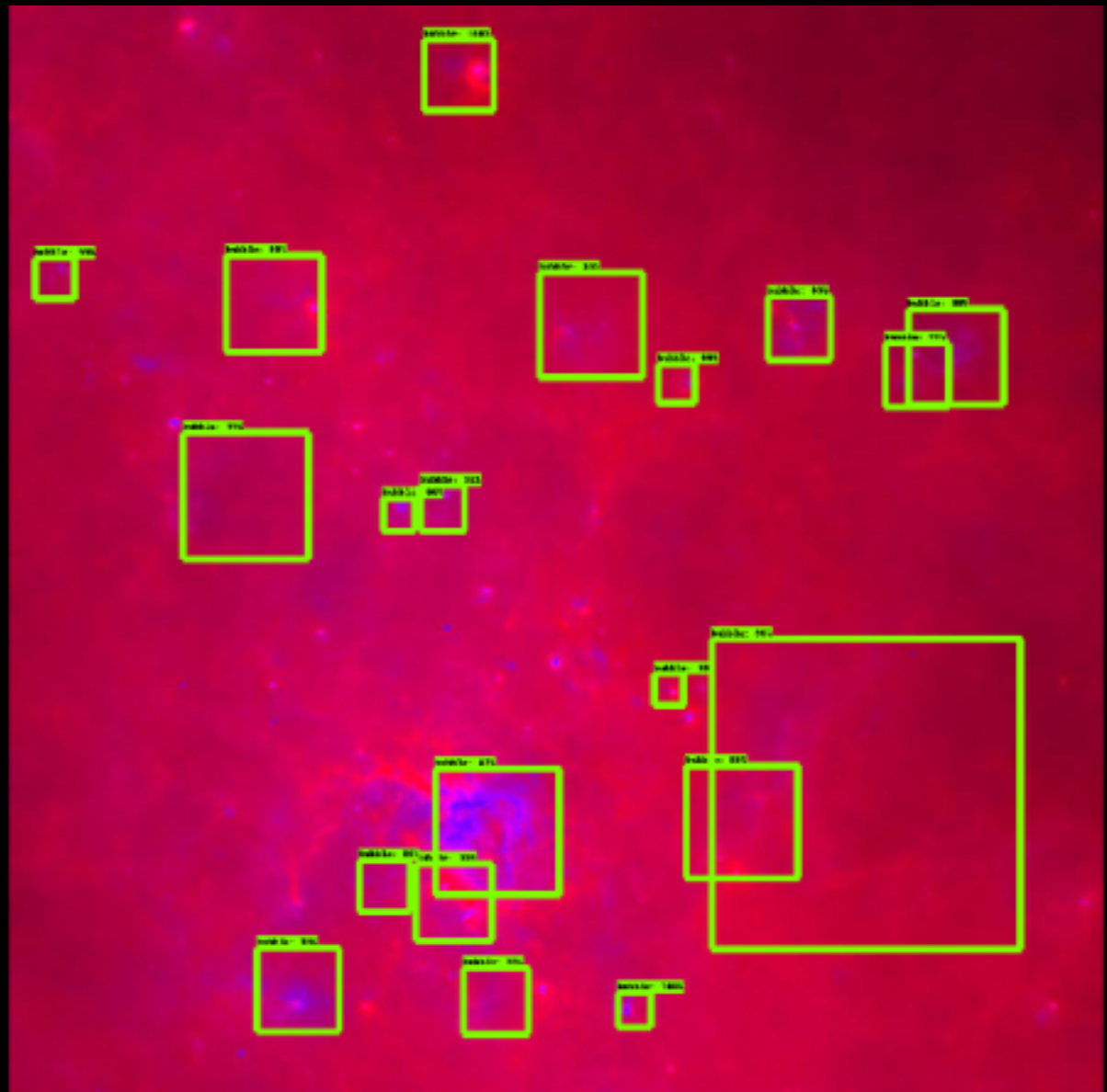


TotalLoss

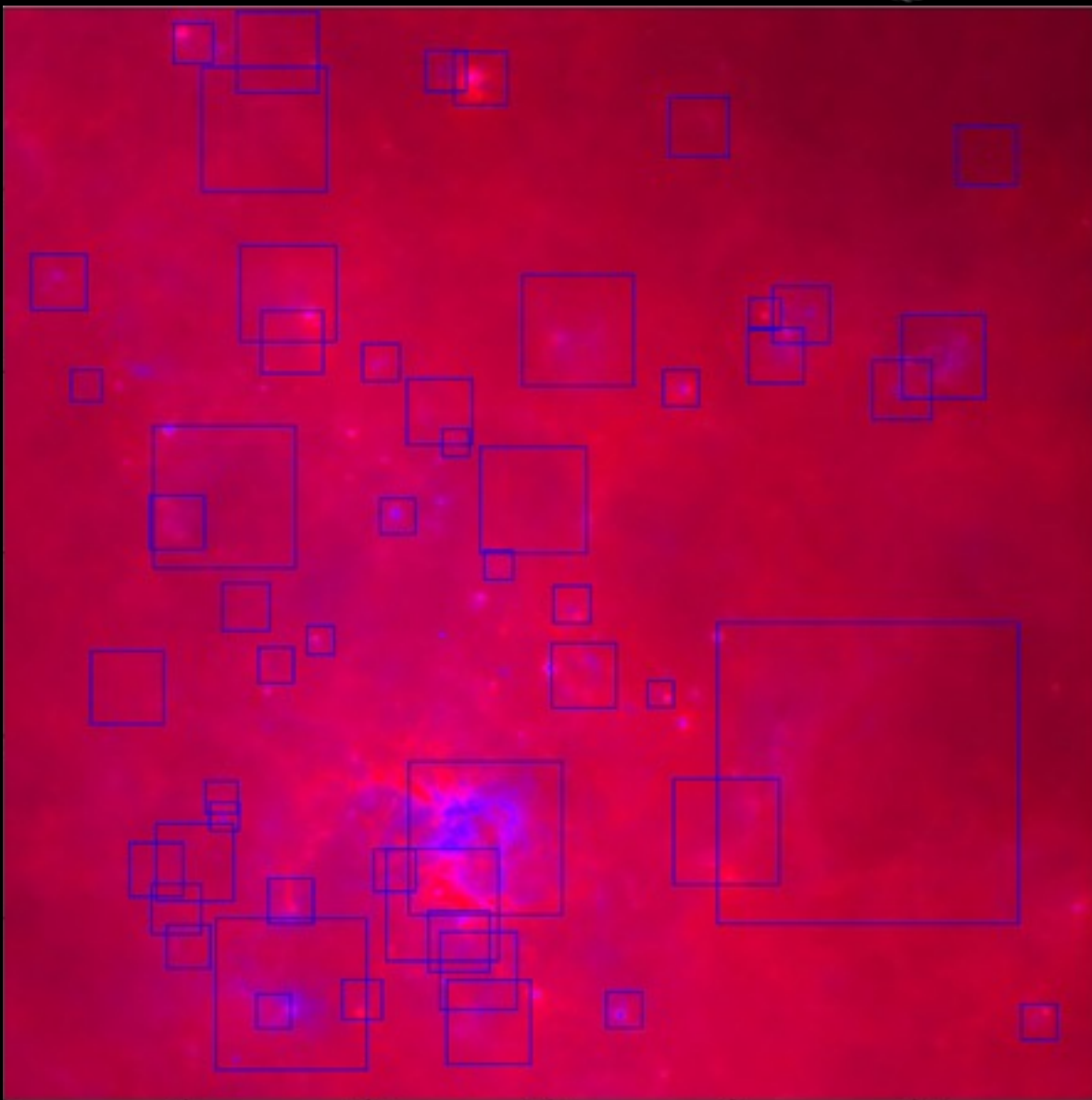


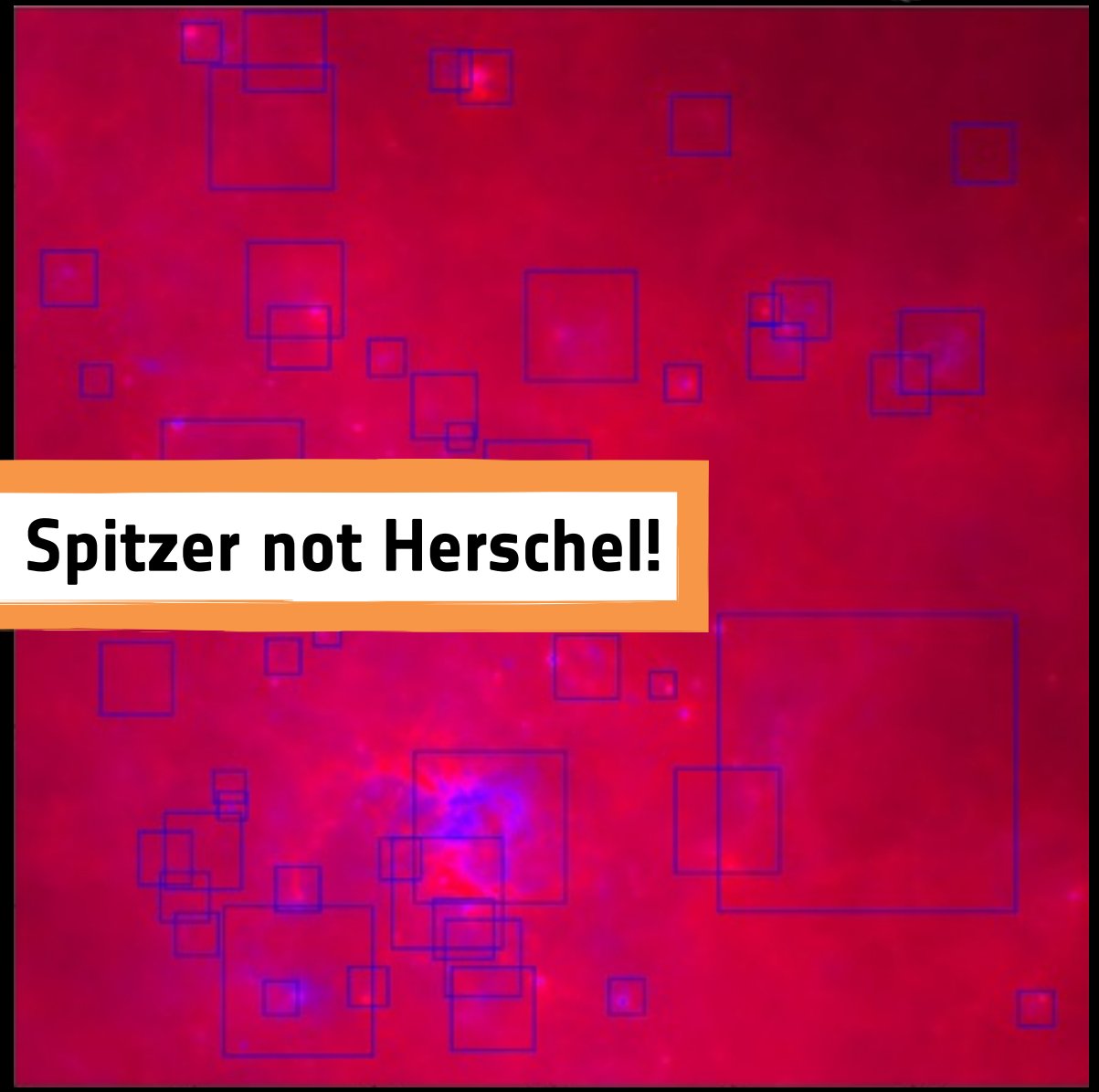
~600 images , ~4000 objects, 10% validation

Machine learning



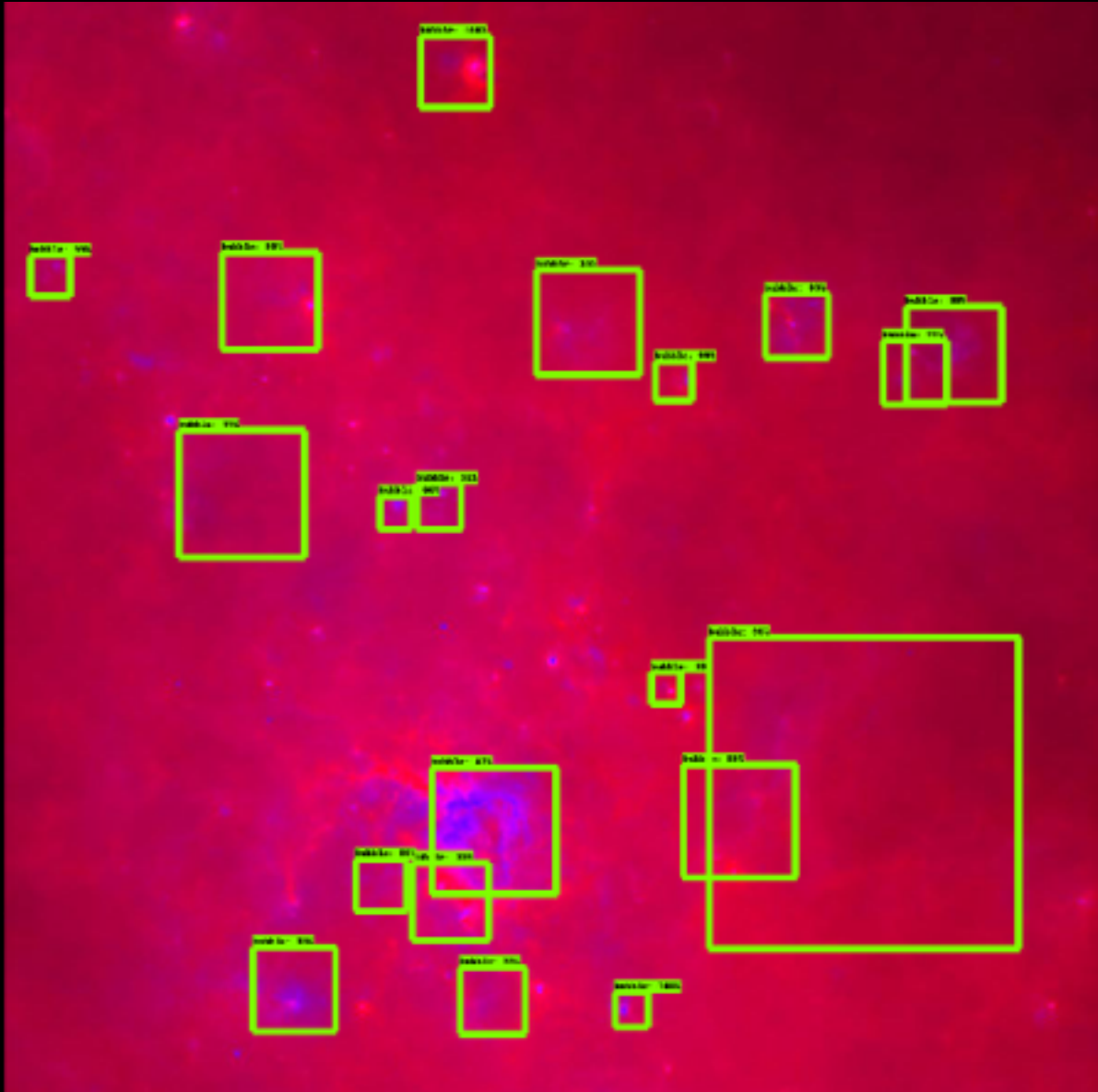
Citizen science



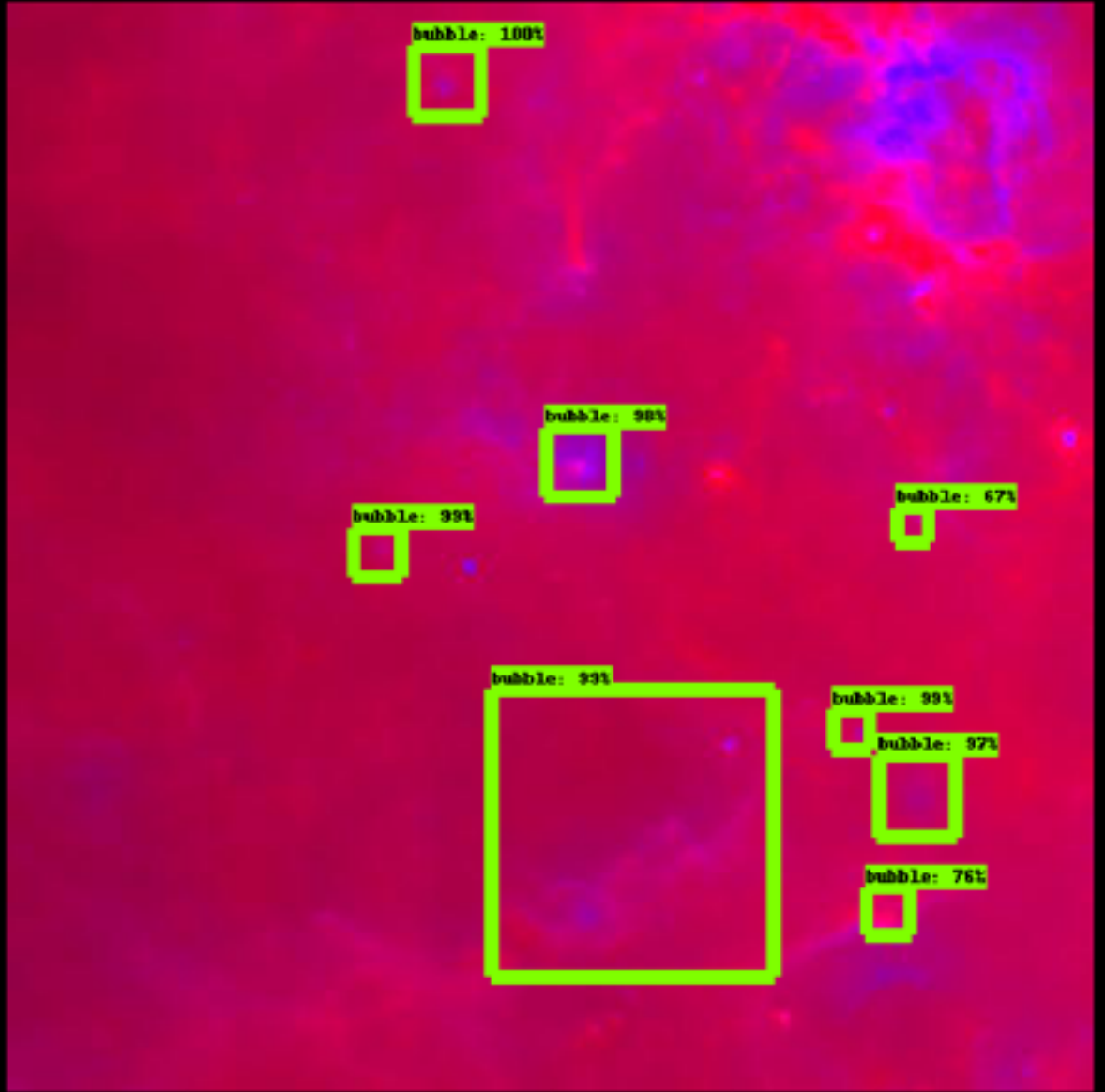


The labels come from Spitzer not Herschel!

applied to 1200x1200



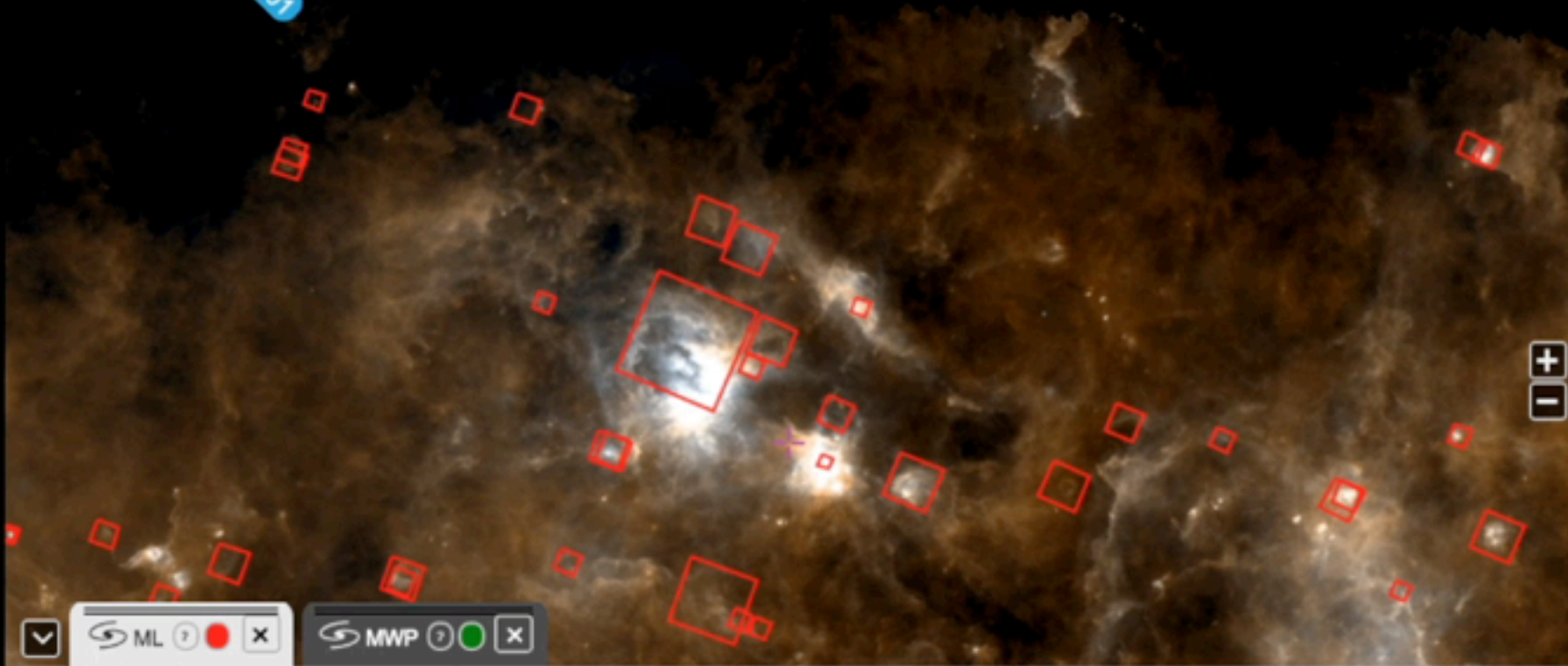
applied to 600x600



🏠
🔄
📊
📈
🎓
📄
🚀

3 767
12 981 201
771
27 548

Search...



📄 ML ? 🔴 ✕
📄 MWP ? 🟢 ✕

<input type="checkbox"/>	<input type="checkbox"/>	Name	id	wid	hei	score
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	2965	2965	0.0469026696833339	0.0452581568469874	0.562308788299561
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	2966	2966	0.0295482488183723	0.0285908031789202	0.999996662139893
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	2967	2967	0.0176068844646977	0.0170042056114692	0.996415853500366
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	2968	2968	0.0248743126943793	0.0237259600980728	0.932490587234497
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	2969	2969	0.0234461656885685	0.0225580393023321	0.911825120449066
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	2970	2970	0.0554963909035564	0.053729375506177	0.99999761581421
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	2971	2971	0.038421063194		0.99999761581421

3574 Observation(s) in the chosen area

Take home message:

“Transfer learning, allows low level features of existing datasets to be repurposed for new problems”

“Science data archives are a gold mine for ML. By using citizen science combined with the multi-wavelength interface of ESASky we can unlock new science opportunities.”



European Space Agency