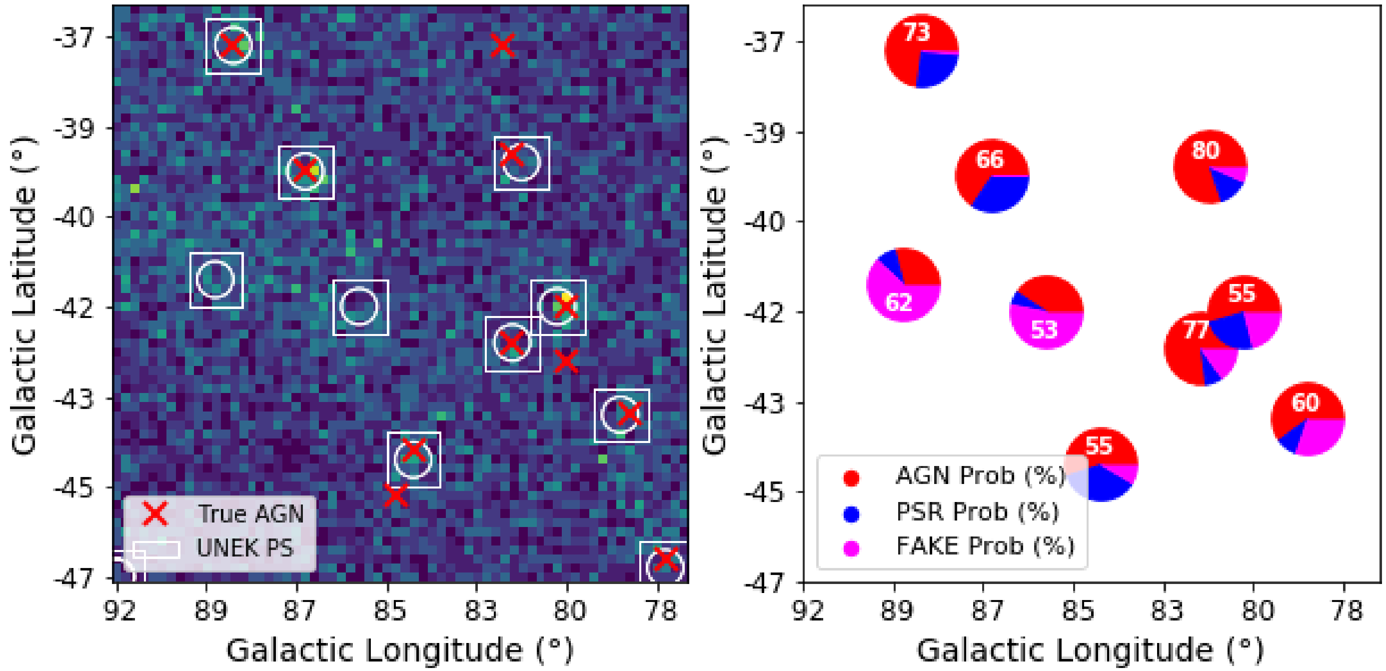
Identification of point sources in gamma rays using U-shaped convolutional neural networks and a data challenge



arXiv: atro-ph/2103.11068, **GitHub**: Gamma-Ray-Point-Source-Detector

Authors: **Boris Panes**, Sascha Caron, Klaas Dijkstra, Christopher Eckner, Luc Hendriks, Gudlaugur Johannesson, Roberto Ruiz de Austri, Gabrijela Zaharijas

DARK
MACHINES
COMMUNITY

Motivation

- Detection and classification of (faint) sources is an important task (e.g. Fermi-LAT collaboration arXiv:1902.10045). Traditional approaches face some challenges

  - Interstellar background dependance
  - Multi-wavelength based classification

- Machine learning algorithms successfully applied to LAT catalogues for source classification (e.g. Saz Parkinson arXiv:1602.00385), but never at the raw LAT images

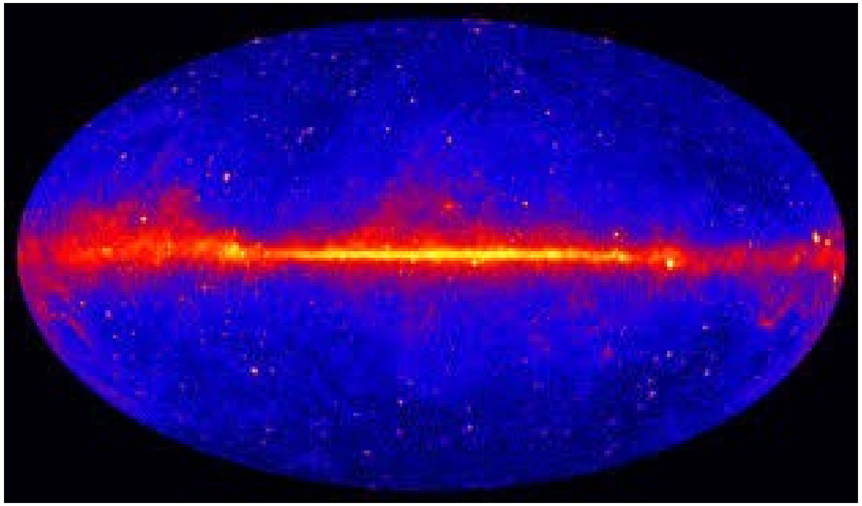
- In radio, deep learning algorithms successfully developed for source detection in images (e.g. DeepSource arXiv:1807.02701)

- **Our goal:** develop a deep learning based pipeline to detect and classify sources based on LAT sky images:

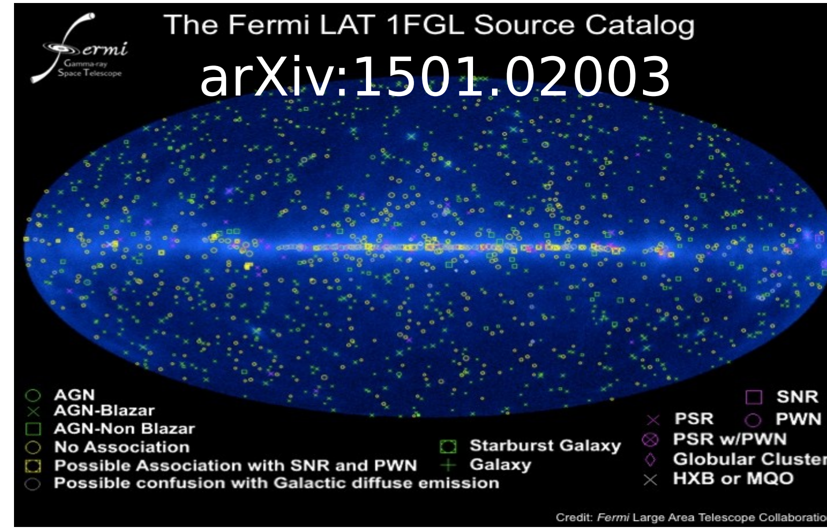
  - PROs: DL is a powerful tool, (in principle) easy to generalise to other wavelengths
  - CONs: training-data dependent and significant amount of training data required + no ‘proper’ statistical framework (in terms of e.g. TS)

Note: no application to real data so far, proof of principle work.

Definition of the problem



Likelihood Analysis



Gamma-ray map in the region 0.5 GeV to > 20 GeV

Input Format:

(400,800,5) => (lat, lon, E)

In each pixel and energy bin we know the value of the photon counting

Machine Learning

Catalog of point sources

Output Format:

(N, 3) => (id, lat, lon, class)

This output includes the point source positions and the class of each source in the catalog

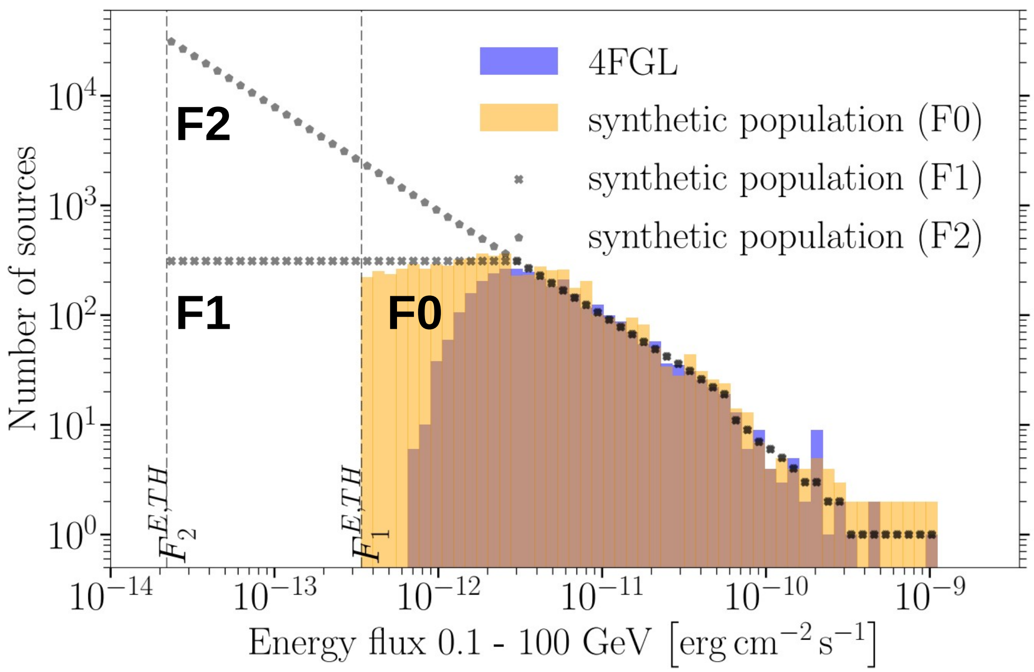
Synthetic data setup: app 9.5 yrs Fermi-LAT

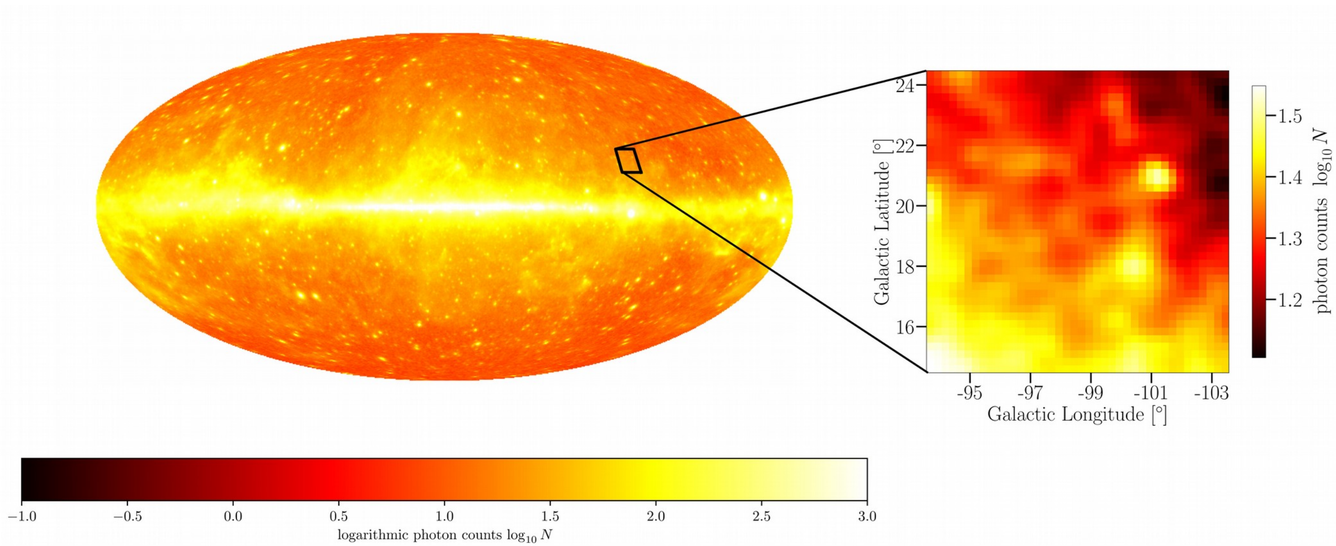
We consider only Active Galactic Nucleai (**AGN**) and Pulsars (**PSR**)

Three different sets of **point source populations** (F0, F1, F2)

Each of these populations is combined with three **versions of the IEM**

B1: gll_iem_v07.fits
B2: SNR CR, z=4, TS=150K
B3: gll_iem_v06.fits
+ Isotropic emission





We train and evaluate ML models in the basis of small sky patches

- 10°x10° PS localisation
- 1°x1° PS classification

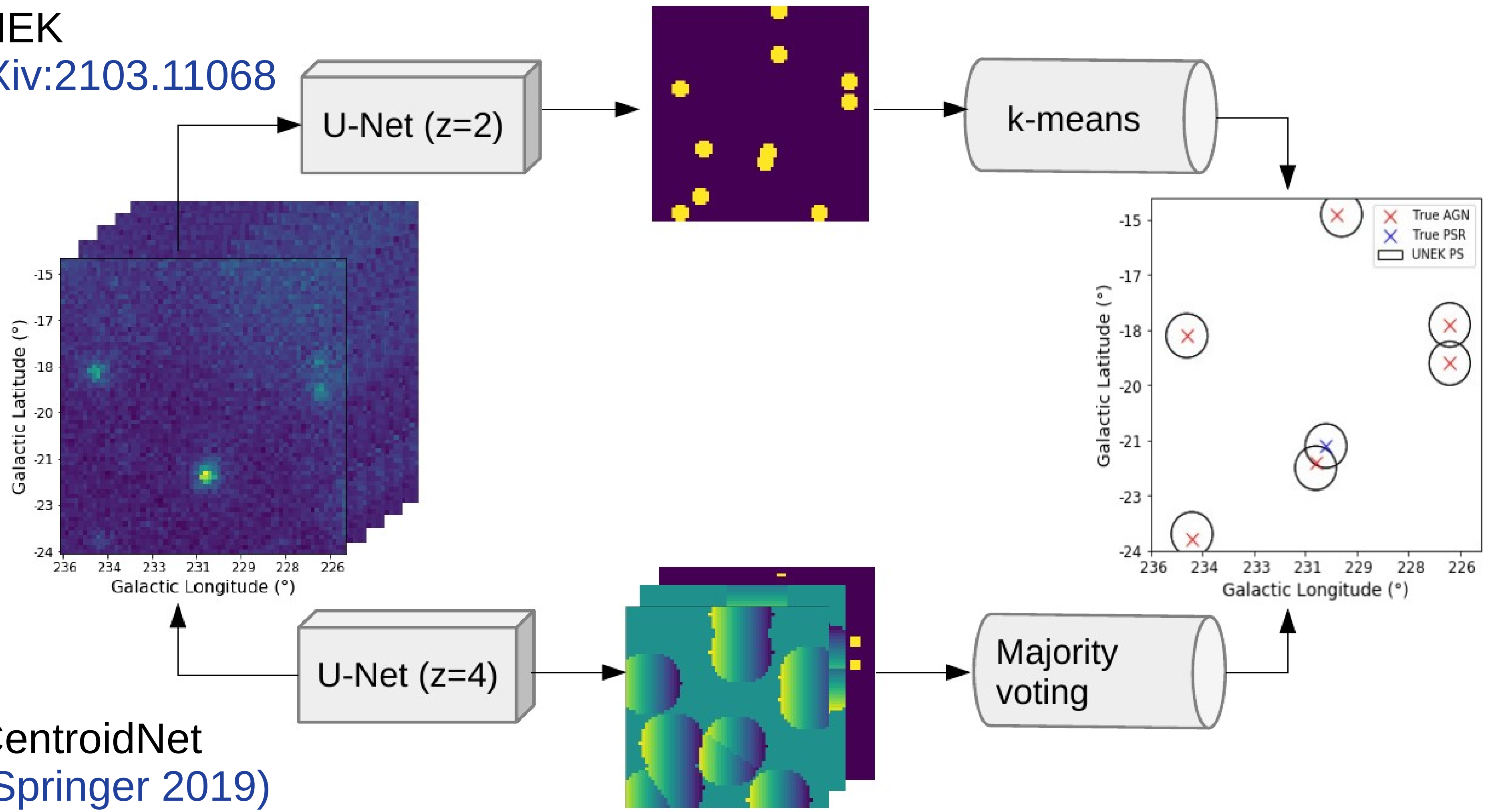
Localisation pipeline

Tha localisation algorithm includes two main parts, **segmentation and clustering**

In order to cross-validate our algorithms we use **two different approaches**

UNEK
arXiv:2103.11068

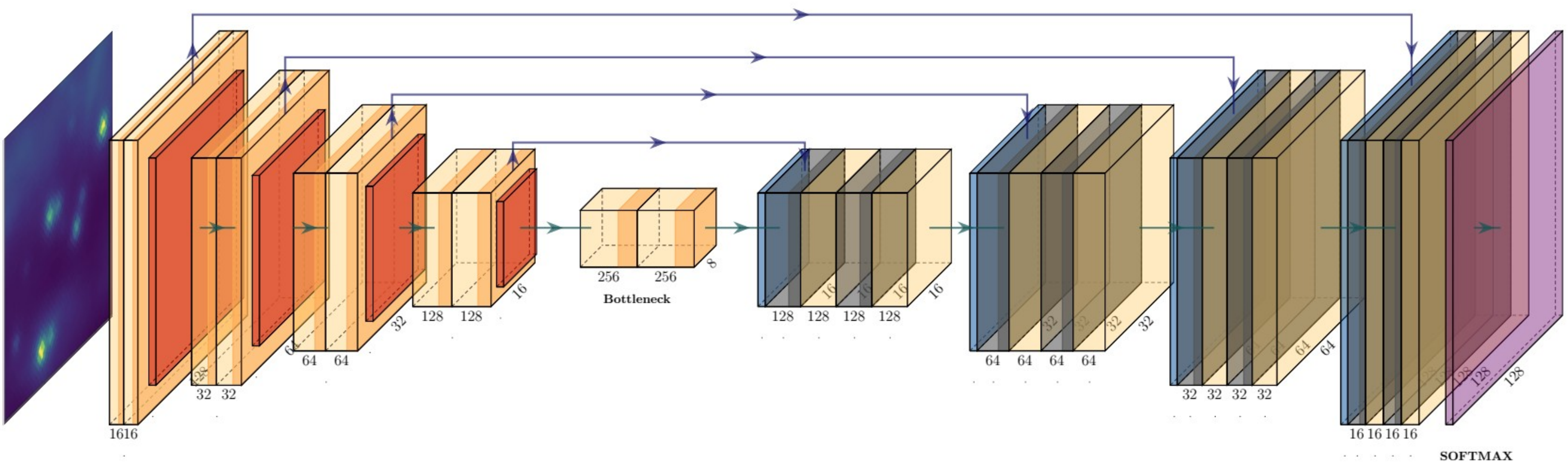
CentroidNet
(Springer 2019)



U-Net hypothesis

For segmentation we consider the architecture called **U-Net** (arXiv:1505.04597), which is a **fully convolutional neural network** that can generate outputs with the same 2D shape as the input

The U-Net is normally used to **learn how to classify each pixel** of a given input image in different classes. This can be useful to distinguish between point sources and background



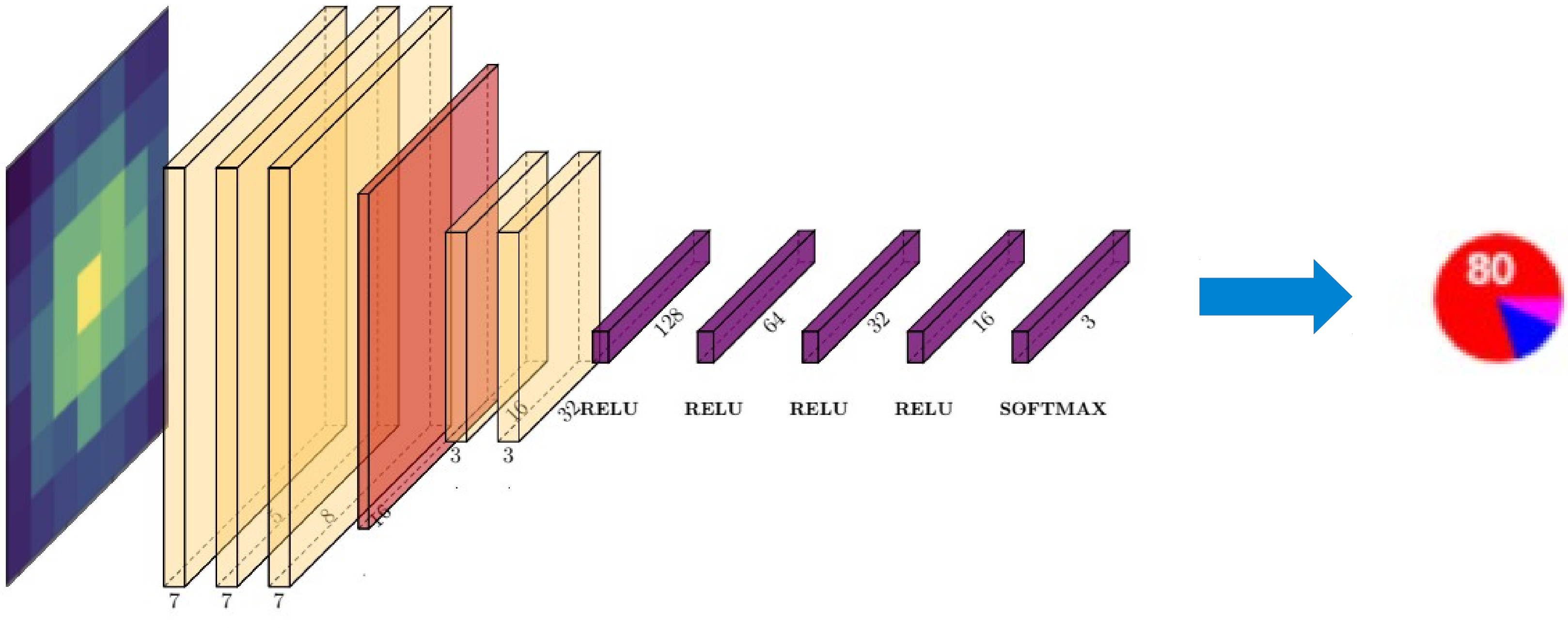
From a very pragmatic point of view, the U-Net is a map that goes from R(n,m) to R(n,m). The free parameters of the function are fixed through supervised training

Classification pipeline

From the positions obtained from the localisation algorithm we are able to produce a dataset of point source patches that consider realistic effects, such as off-set

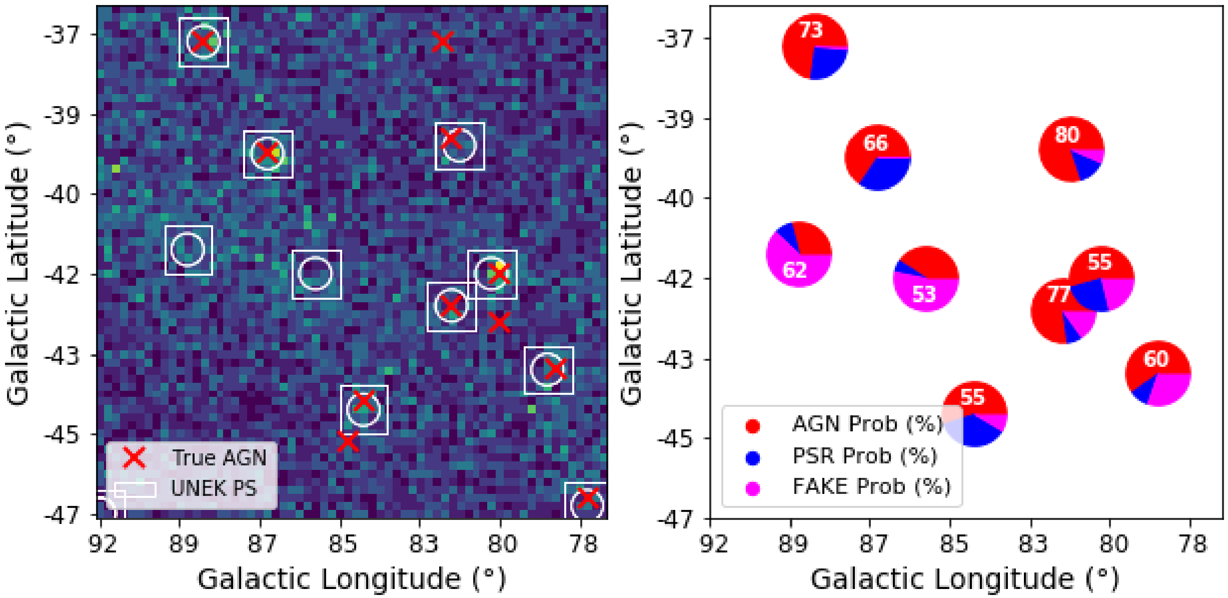
For each point source patch we know the true class **(AGN, PSR, FAKE)**. With this information we can train a supervised CNN model to classify sources

The FAKE class is obtained when the prediction of the localisation algorithm finds a point source which is not matche to any true source. Thus, this class allows us to identify wrongly predicted sources from the localisation algorithm



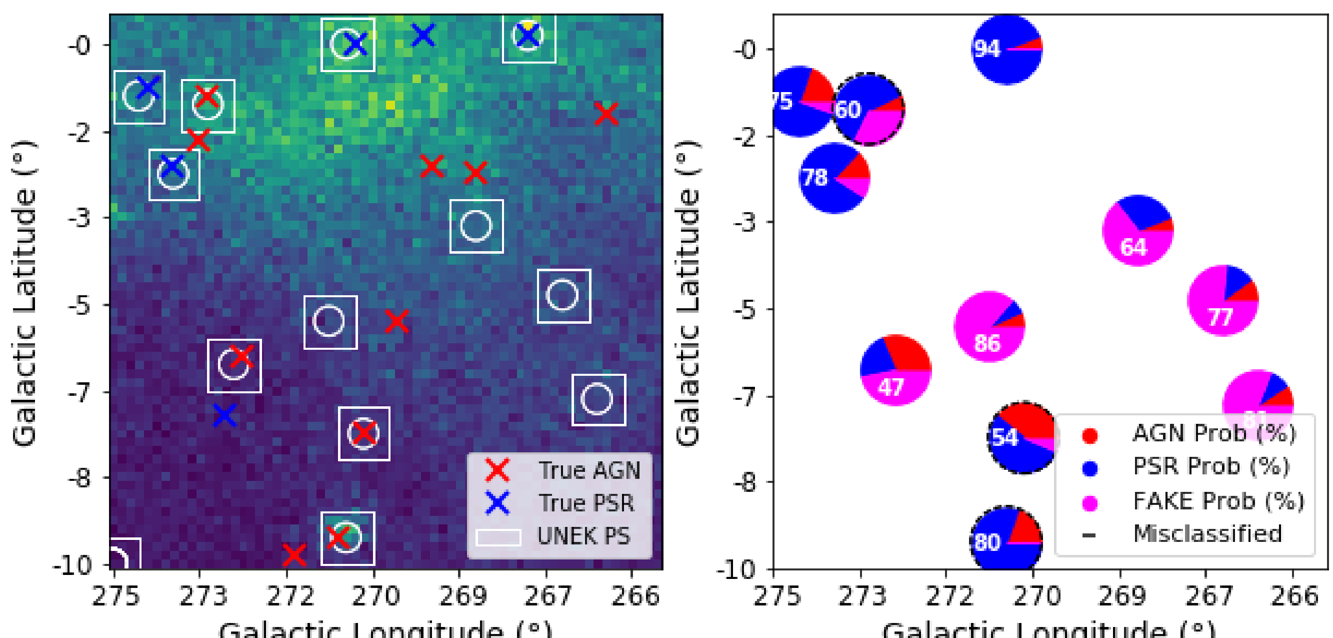
In order to improve the performance of the network we must balance the original distribution of sources

Results and performance: general

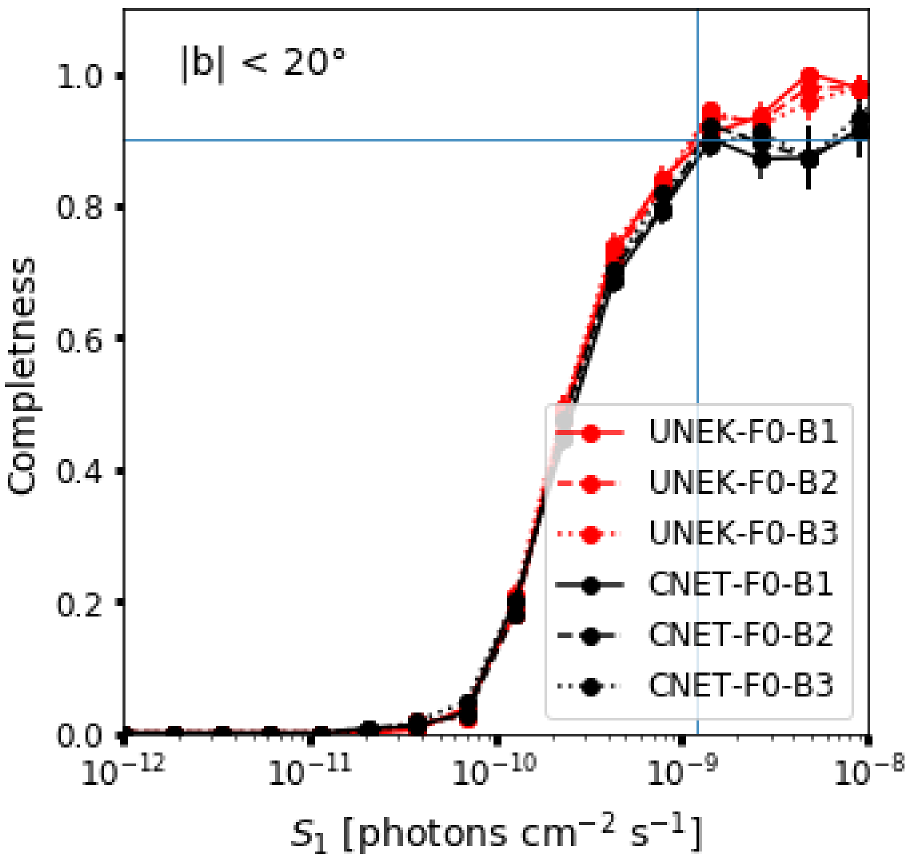


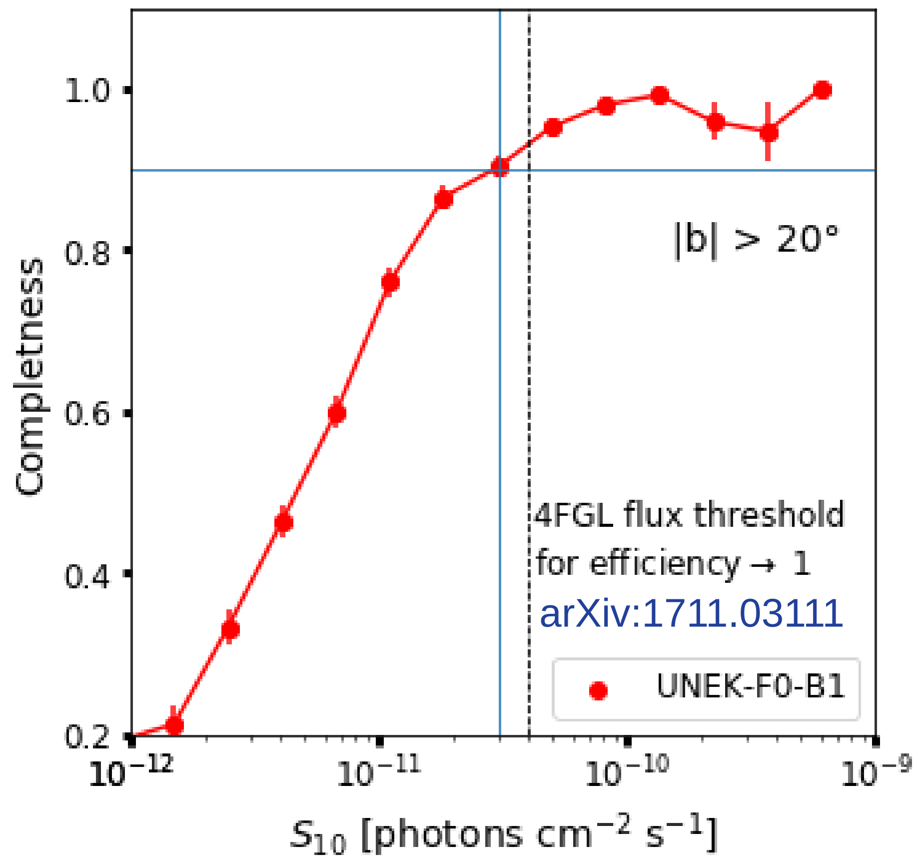
At high latitude the signal to noise ratio is favorable, so we obain high accuracy on localisation and classification

At low latitude the sky is IEM dominated, which results in a decrease of the performance



Results and performance: localisation



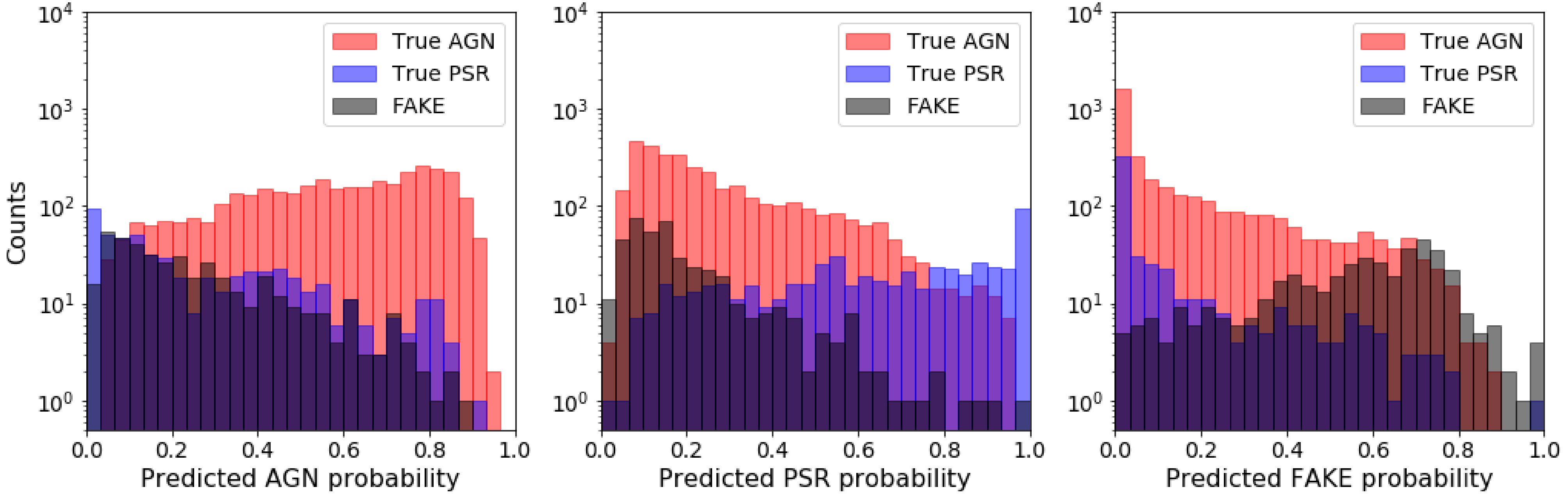


$$Completeness = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

In order to compare results we may focus our attention on the 90% completeness threshold

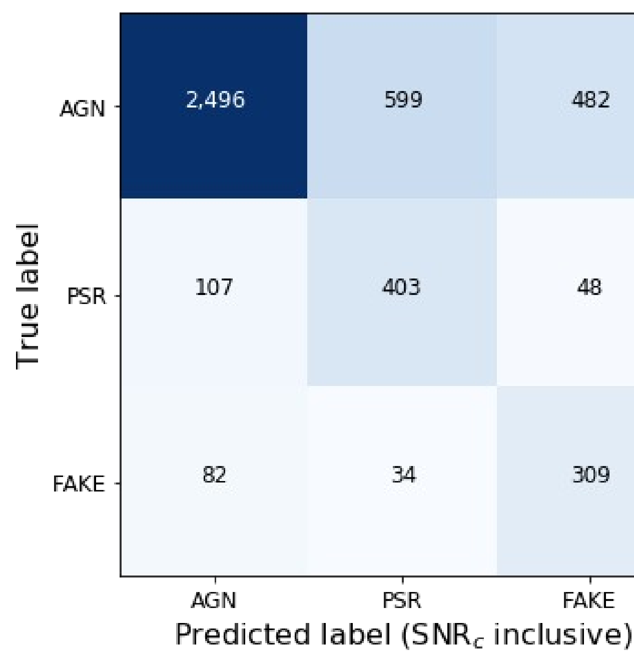
- Largely independent on **IEM**
- Comparable to **4FGL** results
- Compatibility between **both ML algorithms**

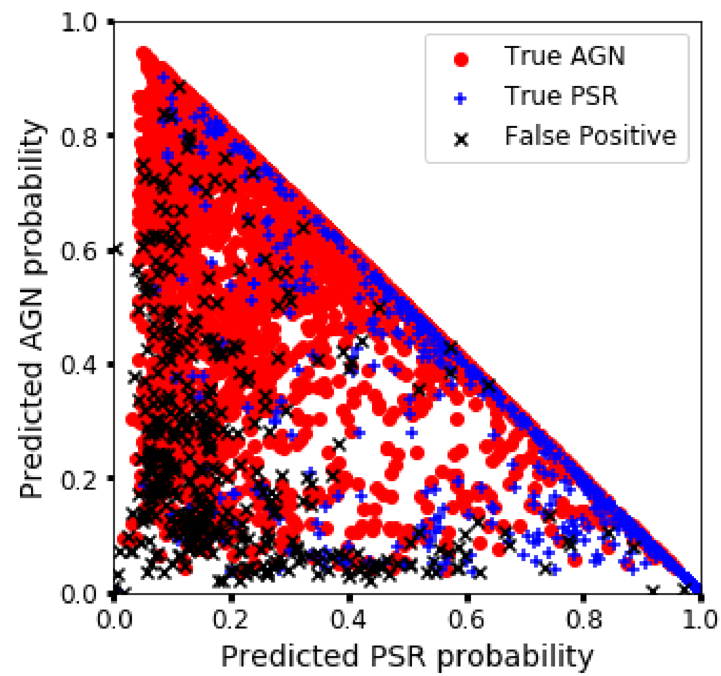
Results and performance: classification



The global accuracy of our current algorithm is around 70%

Possitive correlation between predicted probability outputs and true information





Summary and prospects

- Setup an automatic deep learning image segmentation pipeline which localizes and classifies gamma-ray point sources starting from the raw Fermi LAT data

- In terms of PS detection, performance seems comparable to 4FGL + robust to IEM

- Classification, based only on gamma-ray spectral properties successfully distinguishes PSRs and AGNs (accuracy around 70%)

- **Future**: more complex training data + real data applications

Public Github repository to allow the reproducibility of research results

https://github.com/bapanes/Gamma-Ray-Point-Source-Detector

Public data realese and challenge call to support the systematic improvement of performance

https://zenodo.org/record/4587205#.YFOKBSPhD_Q