Detection of point sources in CMB intensity maps using machine learning techniques

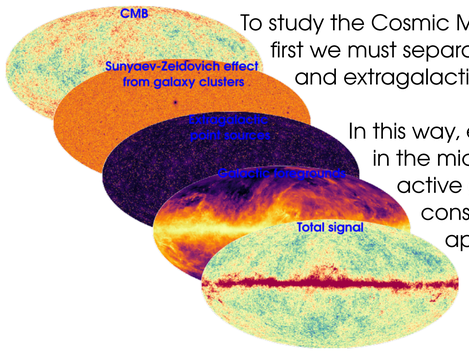
P. Diego-Palazuelos[1,2], R. B. Barreiro[1], P. Vielva[1], D. Balbás[3], M. López-Caniego[4], D. Herranz[1,2], B. Casaponsa[1]

[1] Instituto de Física de Cantabria (CSIC-Universidad de Cantabria), Santander, Spain
[2] Departamento de Física Moderna, Universidad de Cantabria, Santander, Spain
[3] IMDEA Software Institute, Madrid, Spain
[4] Aurora Technology for the European Space Agency, European Space Astronomy Centre, Madrid, Spain

E-mail: **diegop@ifca.unican.es**

MACHINE LEARNING FOR ASTROPHYSICS
CATANIA, 30 MAY - 1 JUNE, 2022

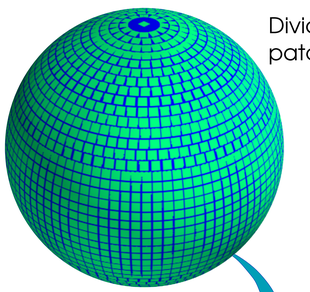
Science case



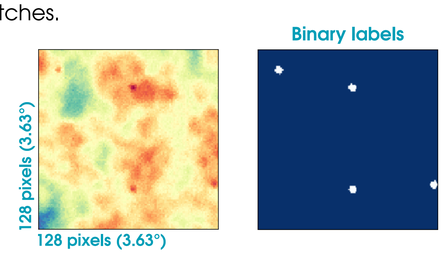
To study the Cosmic Microwave Background (CMB), first we must separate it from the rest of Galactic and extragalactic emissions.

In this way, extragalactic sources emitting in the microwave range (*e.g.*, radio-loud active galactic nuclei or dusty galaxies) constitute a contaminant that often appears in the form of point-like objects in CMB observations.

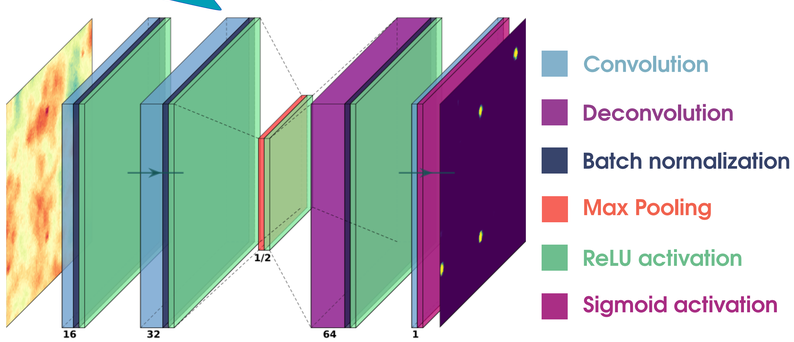
Machine learning approach



Divide the sphere into 2762 non-overlapping patches.

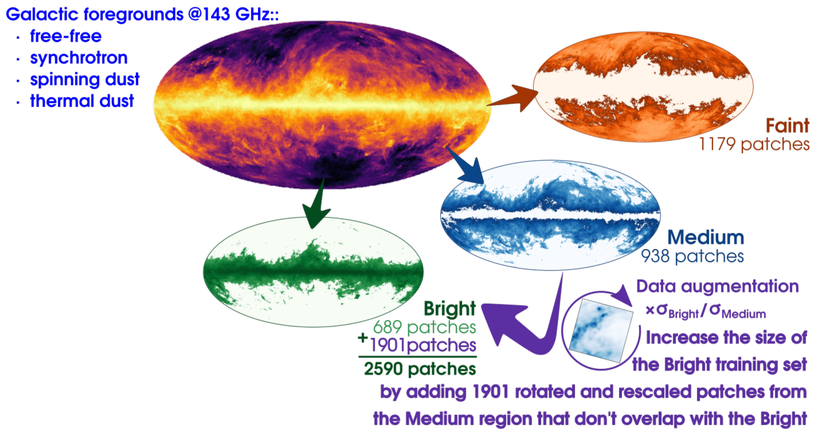


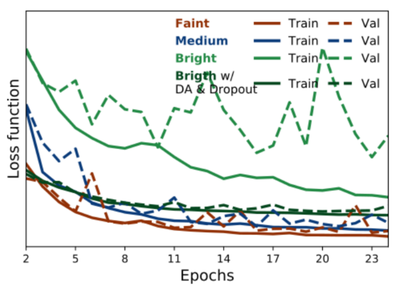
Train a convolutional neural network (CNN) that solves source detection as an image segmentation problem.



Point source detection in realistic simulations

We prepare the CNN for its application to data by training with simulations of Galactic foregrounds, CMB, and noise. We address this high level of complexity by dividing the sky into separate regions of progressively increasing foreground intensity and independently training specialized CNNs for them.



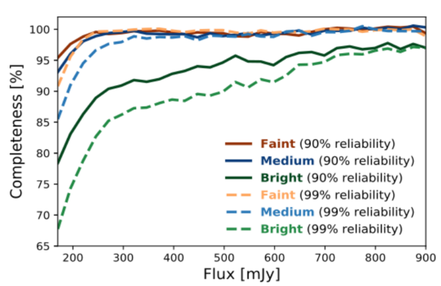


This partition of the sky reduces the volume of data available for training in each region and can limit the CNN's generalization ability, leading to overfitting.

Preliminary results show that this problem can be solved through data augmentation (DA) and dropout techniques.

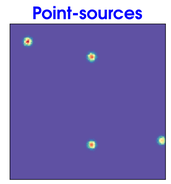
We achieve a promising completeness in all regions of the sky.

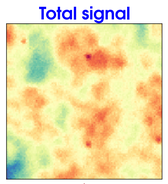
For reference, the *Second Planck Catalog of Compact Sources* (base reliability of 80%) reaches a 90% completeness at 177 mJy in the **Faint** region.



Binary segmentation vs flux estimation

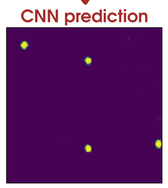
By treating source detection as a binary segmentation operation, we decouple the process of localization from flux estimation.





Machine learning approach

Source detection as binary segmentation. The sources' position is our only concern. We don't have sensitivity to the sources' flux.

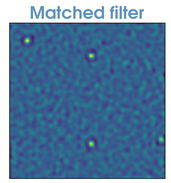


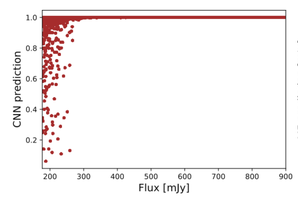
Conventional approach

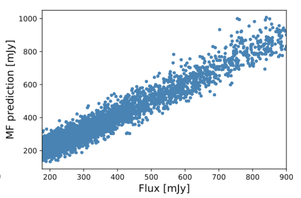
Source detection through the combination of a matched filter (MF) and a detection criterion.

By definition, the MF is the optimal linear filter for maximizing the signal-to-noise ratio in the presence of a stochastic background.

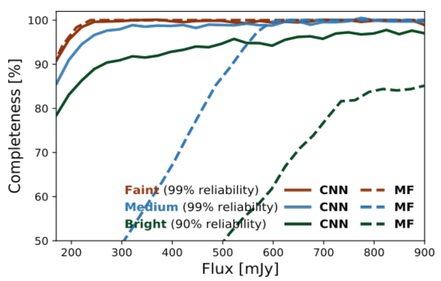
It provides an unbiased estimate of the source's flux.







Once we move to regions of the sky dominated by Galactic foregrounds, the MF is no longer able to correctly characterize the statistics of the background, but the CNN is still able to provide high levels of completeness at low fluxes.



Going forward

Once overfitting is fully under control, we will be ready to extend the training to other frequencies and, eventually, apply the CNN to real data and produce deeper and more complete extragalactic source catalogs.