Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Machine Learning Based Cosmological Radio Source Detection

(aka…kill two birds with one stone)

Nicoletta Sanvitale (INAF-IRA), Chiara Stuardi (INAF-IRA), Claudio Gheller (INAF-IRA), Andrea Botteon (INAF-IRA) and Franco Vazza (UniBo)

HaMMon (Hazard Mapping and vulnerability Monitoring) : WP 4

WP4 objective:

Development of vulnerability curves for flood, windstorm, hail and seismic risk, based on the characteristics of the buildings automatically identifiable by the algorithms.

Automatic assessment of Italy’s buildings patrimony through Machine Learning

INAF focus on WP4

AI solutions for automatic extraction of building features

Building Footprint

Roof type

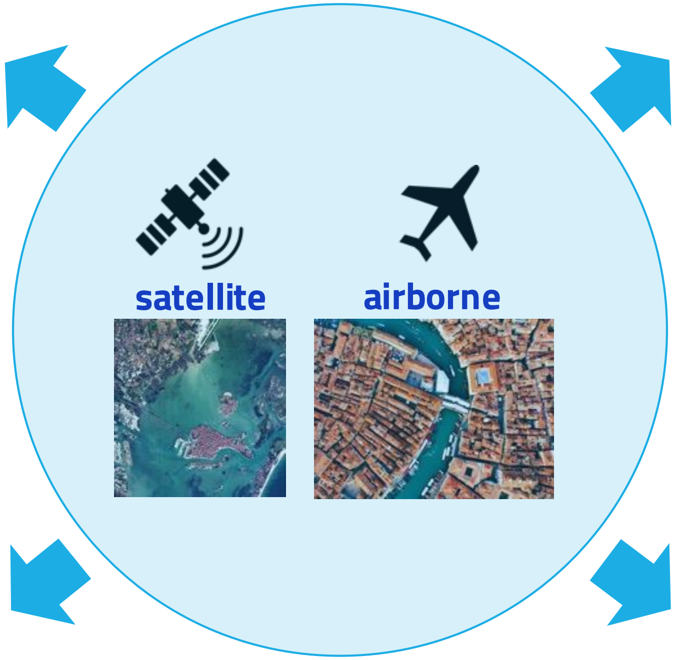
simple gable

simple cross gable

complex cross gable

simple hip

unknown



Roof material

Gravel tiles

Red clay tiles

Metal sheets

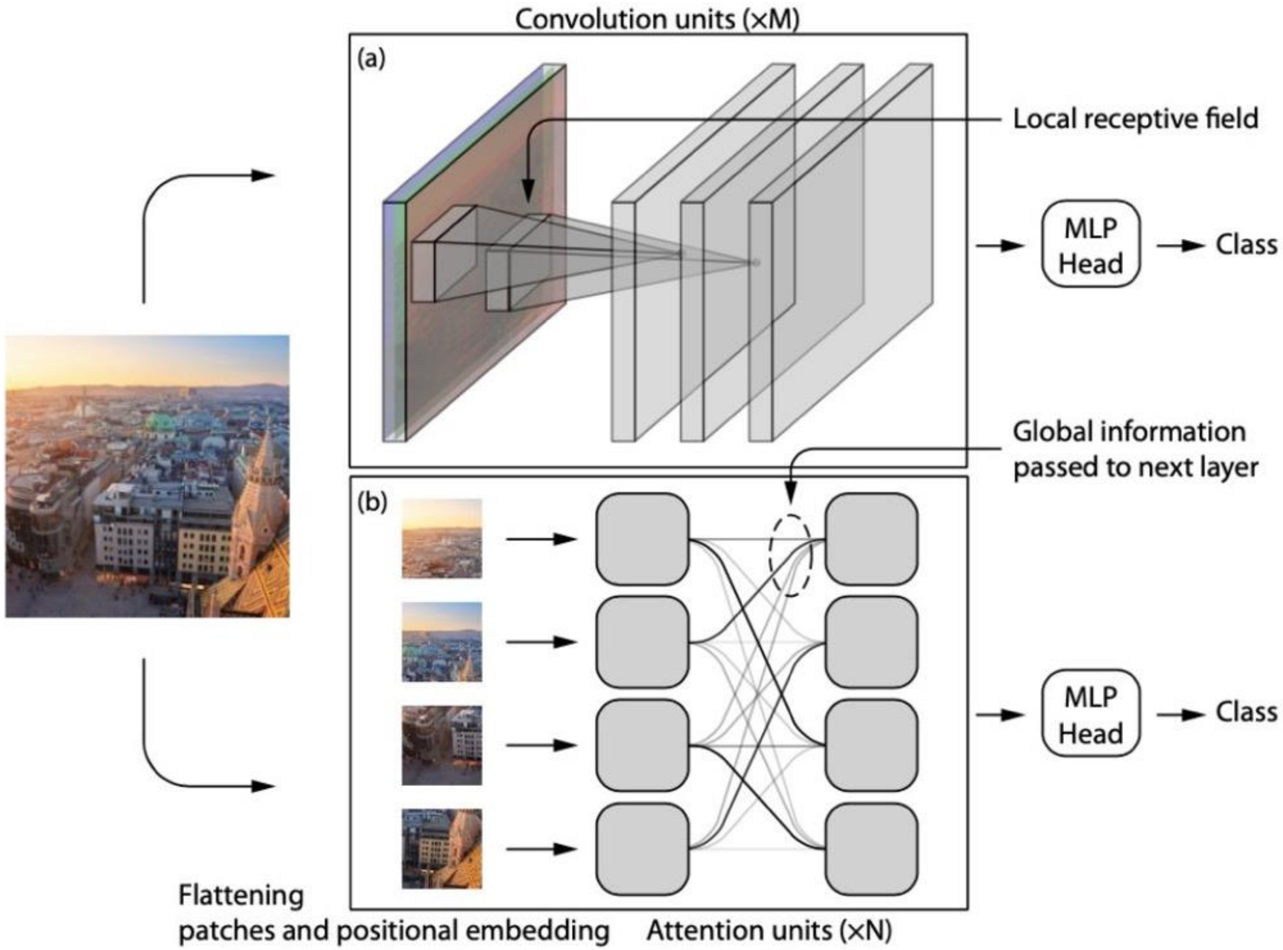
Loose gravel

Sustainable Roof

Solar Rooftop

Green Rooftop

Approches : CNN and Vision Transformers (ViT) architectures



(Tuli, Shikhar et al. 2021 "Are Convolutional Neural Networks or Transformers more like human vision?" *ArXiv* abs/2105.071972021)

Convolution uses a kernel to apply sliding multiplication and addition to local data points:

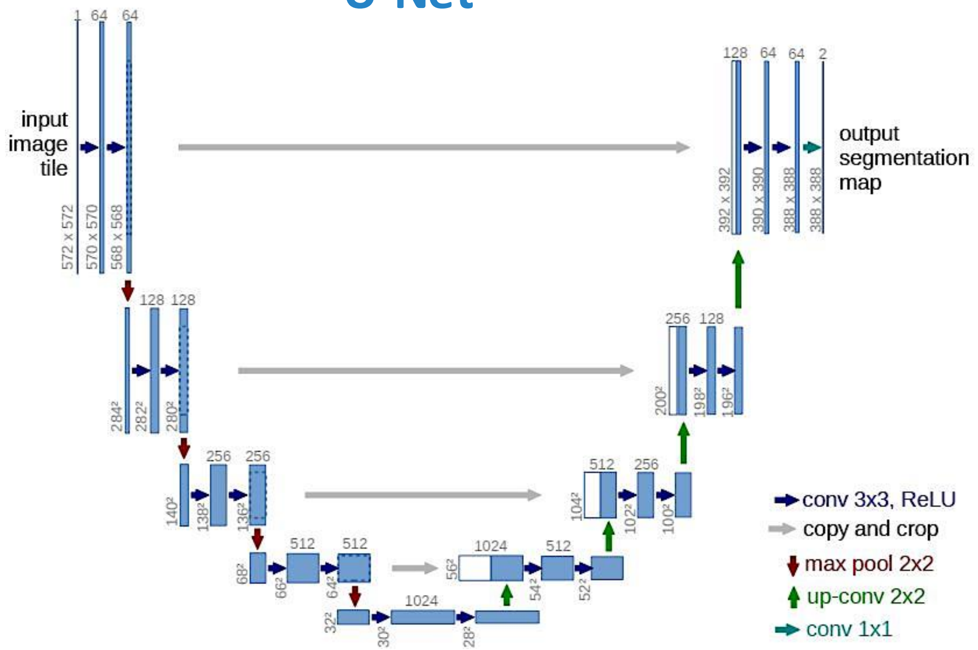
- Strong ability to model local properties
- Less data for training
- Scaling linearly with the datasize O(N)

Vision Transformers treat images as sequences of image patches. The Transformer encoder is a stack of **Attention layers** that allow the model to learn the relationships between the patches:

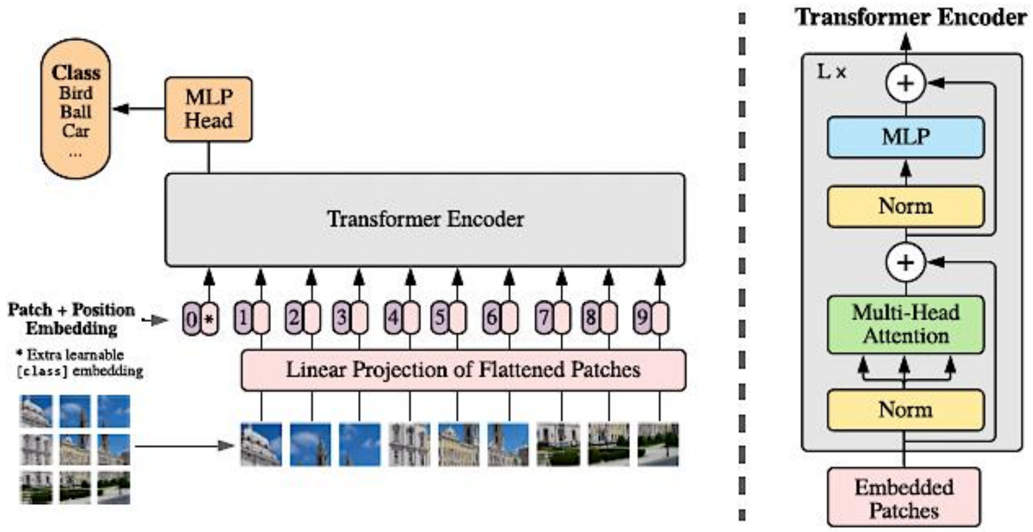
- Strong ability to model long-range dependencies
- More data for training
- Scaling quadratically with the datasize $O(N^2)$

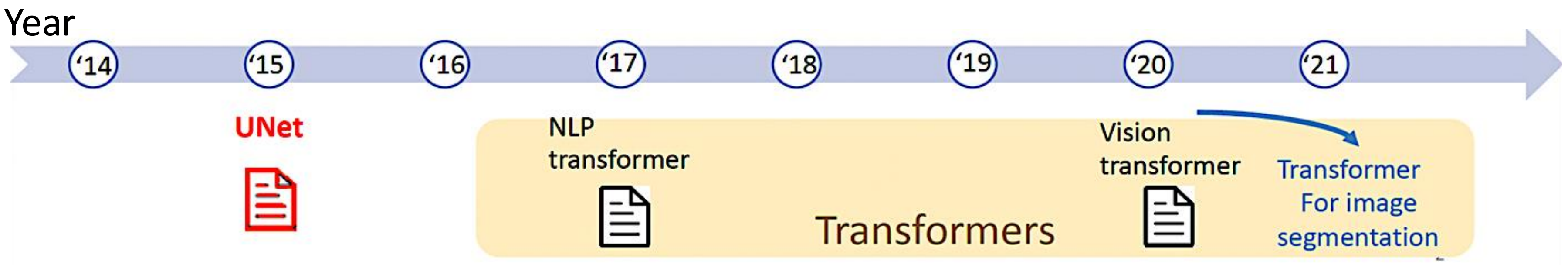
Hybrid ViTs with Convolutions

U-Net

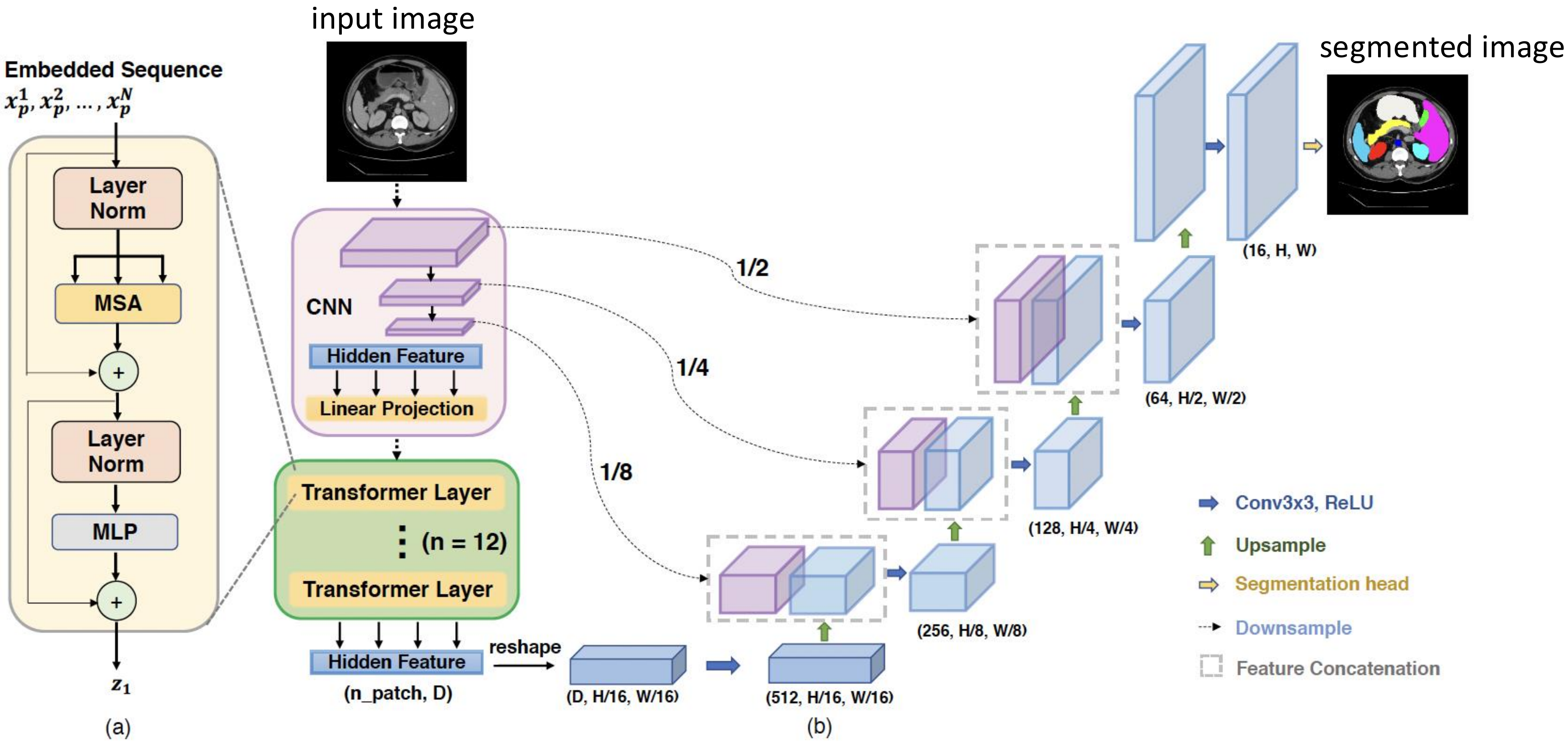


Vision Transformer





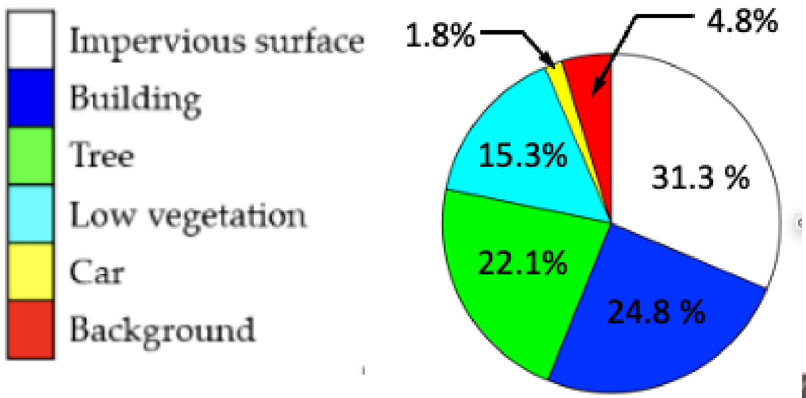
TransUnet

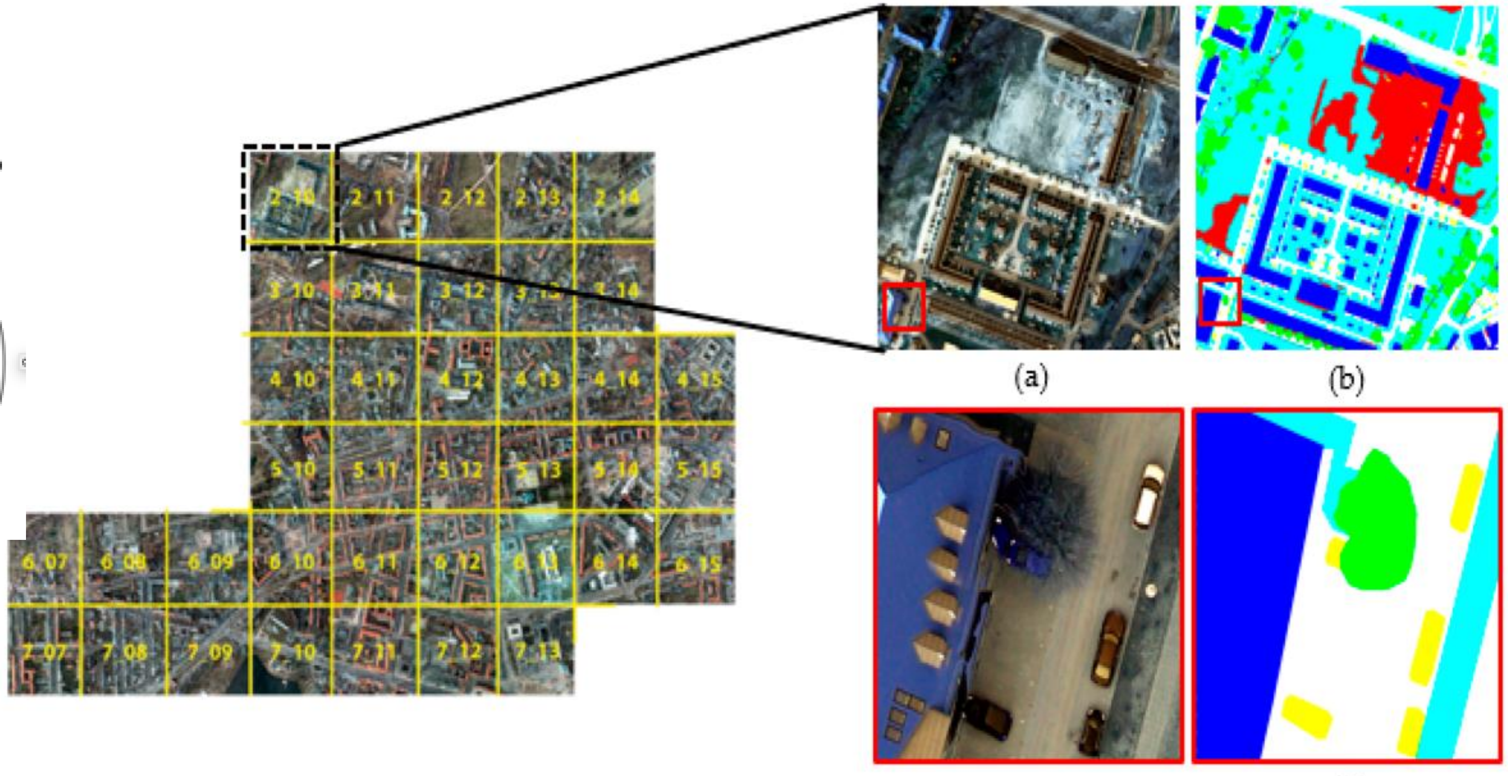


Chen, Jieneng, et al. (2021) "Transunet: Transformers make strong encoders for medical image segmentation." *arXiv preprint arXiv:2102.04306*

Open benchmark Potsdam/Germany dataset

- 38 high-resolution airborne images (6000 ×6000 pixels)
- training dataset 24 images, testing dataset 14 images
- spatial resolution of 5 cm.
- Near-infrared (NIR), RGB, orthorectified imagery
- ground-truth comprising 6 classes

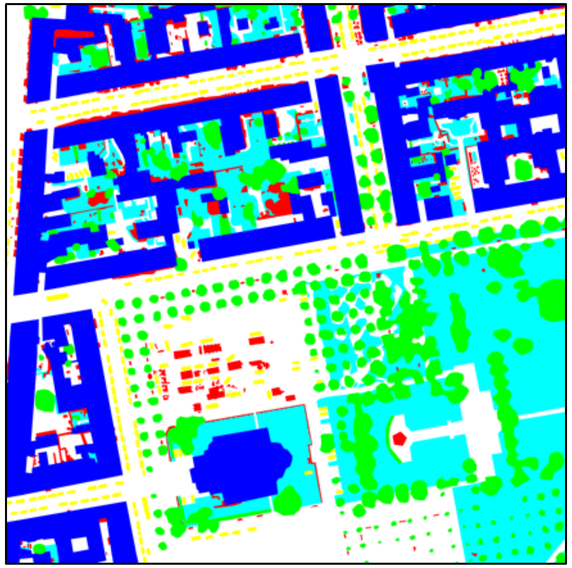




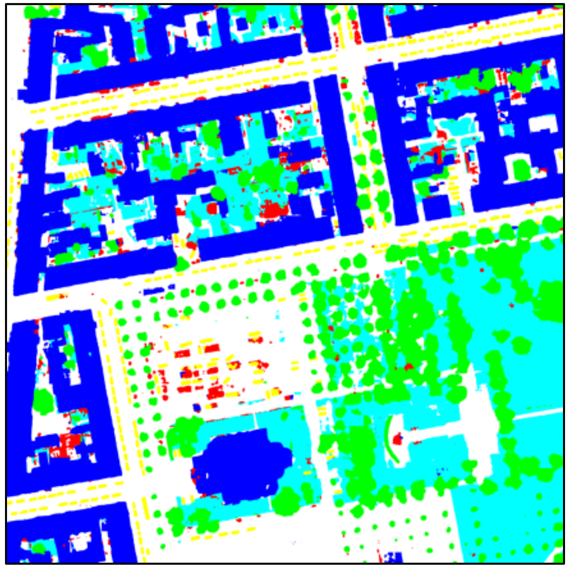
Comparison of predictions across Resolutions

TransUnet: lr 0.001, epoch 20, batch size 24

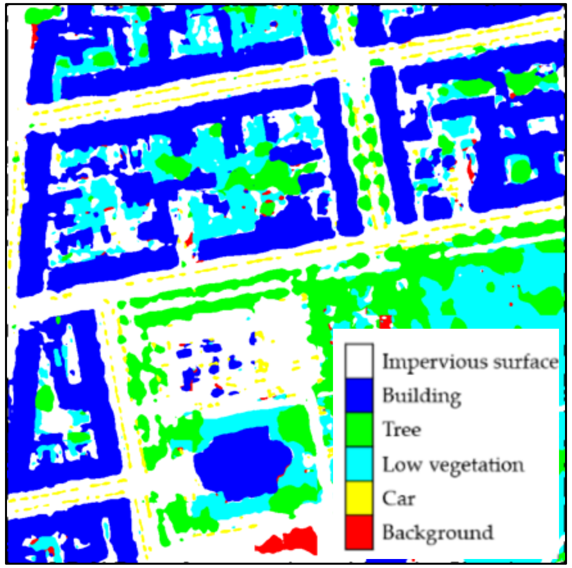
ground truth

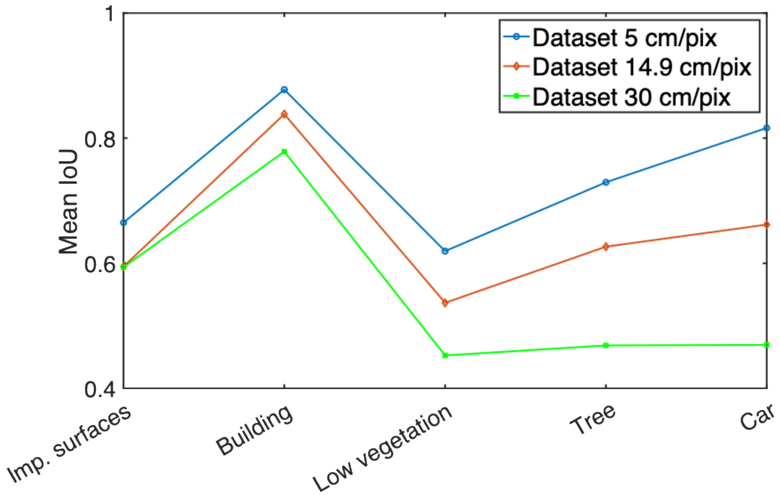


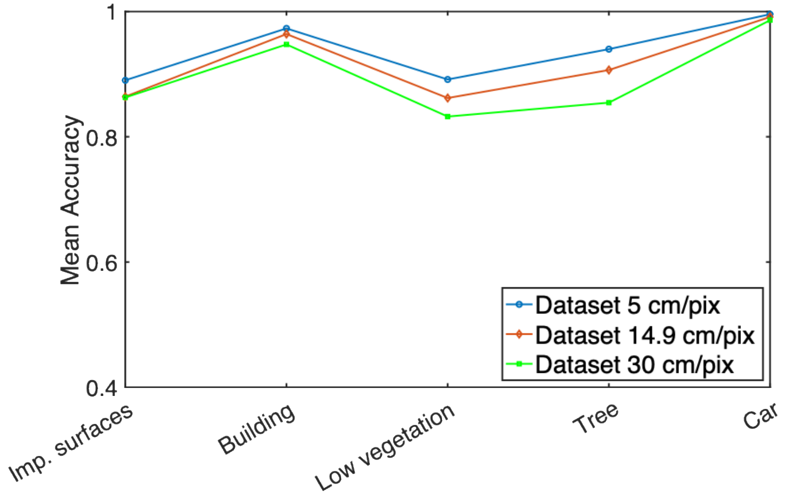
prediction 5.5 cm/pix

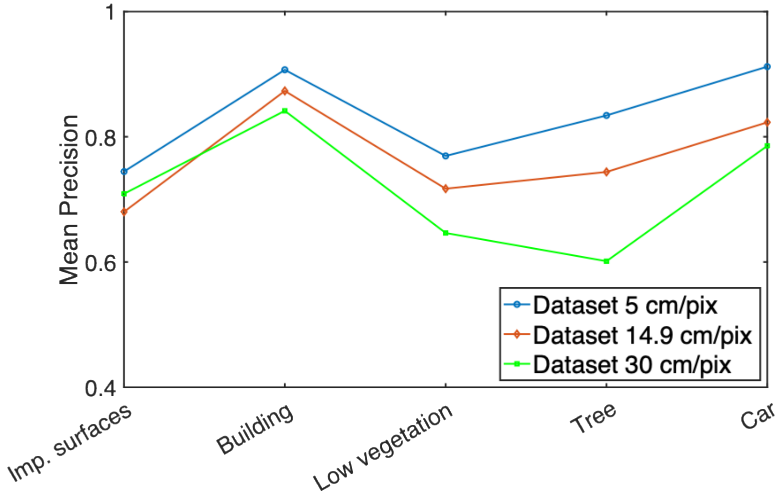


prediction 30 cm/pix









Radio Source Detection

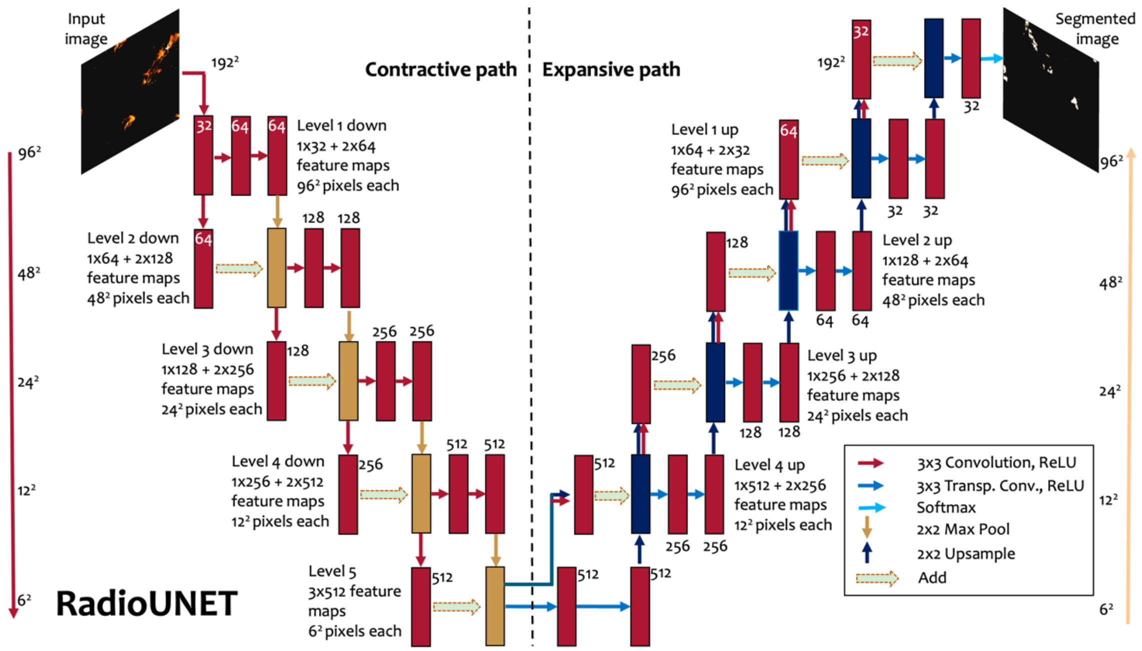
Target: detection of radio sources in data collected from radio telescopes (galaxies, cluster of galaxies, cosmic web, etc.).

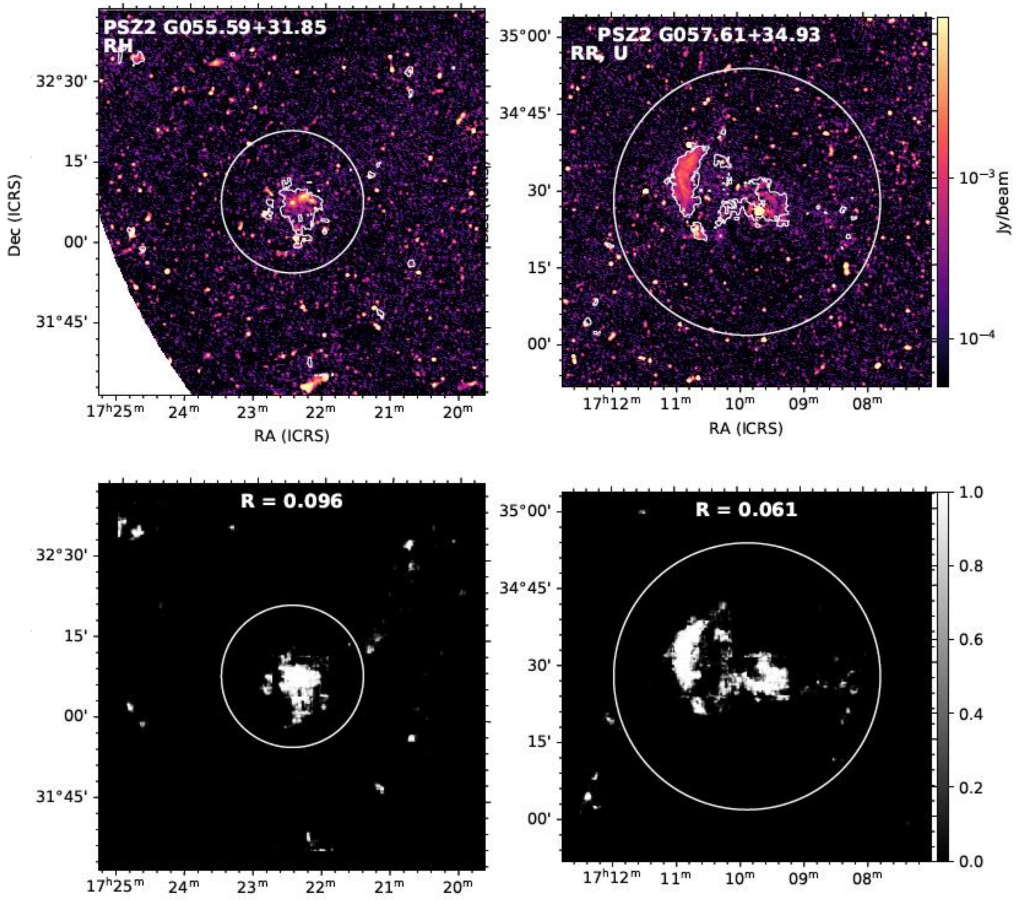
Due to the size and complexity of the data produced by current and upcoming astrophysical observatories (LOFAR, MeerKAT, SKA, etc), its processing needs to be:

- automated
- accurate
- fast

Deep learning approach: RadioUnet

C Stuardi, C Gheller, F Vazza, A Botteon, Radio U-Net: a convolutional neural network to detect diffuse radio sources in galaxy clusters and beyond, *Monthly Notices of the Royal Astronomical Society*, Volume 533, Issue 3, September 2024

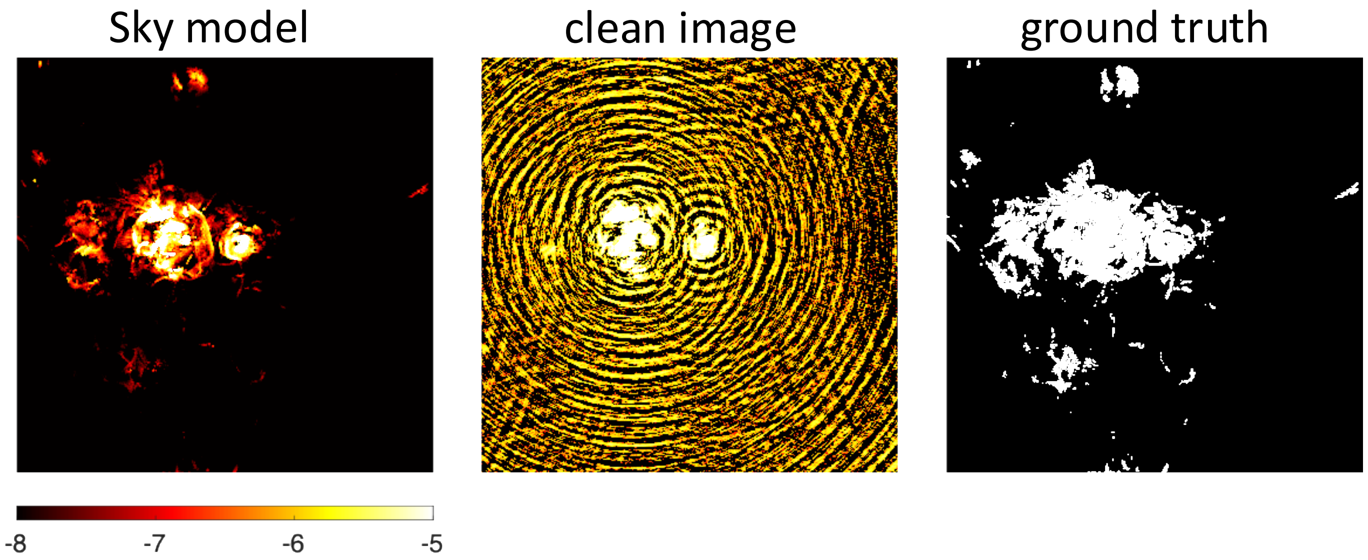




Examples of LOFAR LoTSS-DR2/PSZ2 galaxy clusters processed by Radio U-Net and successfully classified

Exploring Vision Transformers for Radio Source Detection

Comparison of RadioUnet vs TransUnet (TU) based on the simulated dataset



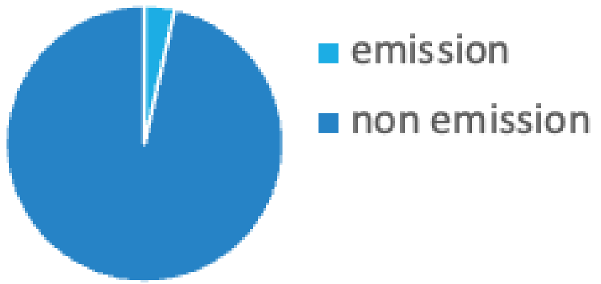
the ground truth mask is calculated as the pixels of the sky image with flux > 10^{-8} Jy

Mock observation:

- Light cones from cosmological simulations (sky model)
- Observations simulated with WSClean (https://gitlab.com/aroffringa/wsclean) + custom tools to reproduce noise and artifacts **such as**
    - ✓ **Random instrumental noise** (thermal, electronics…)
    - ✓ **Incomplete sampling of visibility space** (PSF or dirty beam)
- LOFAR HBA – 8 hrs observation

Comparison of Evaluation Metrics

- Dataset with 140 mock images (2000 ×2000 px)
- training 100 images, testing 40 images
- nominal spatial resolution 2 arcsec
- single frequency 150 MHz
- ground-truth comprising 2 classes:
  - ✓ class 1: emission (3.5%)
  - ✓ class 0: non emission (96.5%)

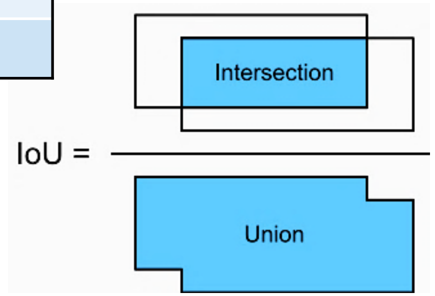


Class 1 Statistics: Emission

RadioUnet: 200 epochs, lr 10^{-4}, batch size 50, tile size 192 px		
metric	value	std
Dice	0.4559	± 0.0074
IoU	0.3054	± 0.0055
Precision	0.6372	± 0.0096
Recall	0.3916	± 0.0078

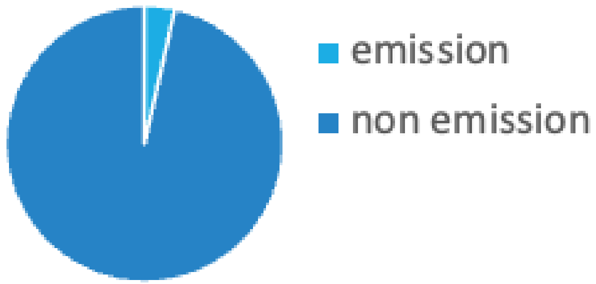
TransUnet: 200 epochs, lr 0.005, batch size 24, tile size 512 px		
metric	value	std
Dice	0.5377	± 0.0096
IoU	0.3766	± 0.0090
Precision	0.6121	± 0.0184
Recall	0.4987	± 0.0120

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \qquad \text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \qquad \text{Dice} = \frac{2 \cdot \text{IoU}}{1 + \text{IoU}}$$



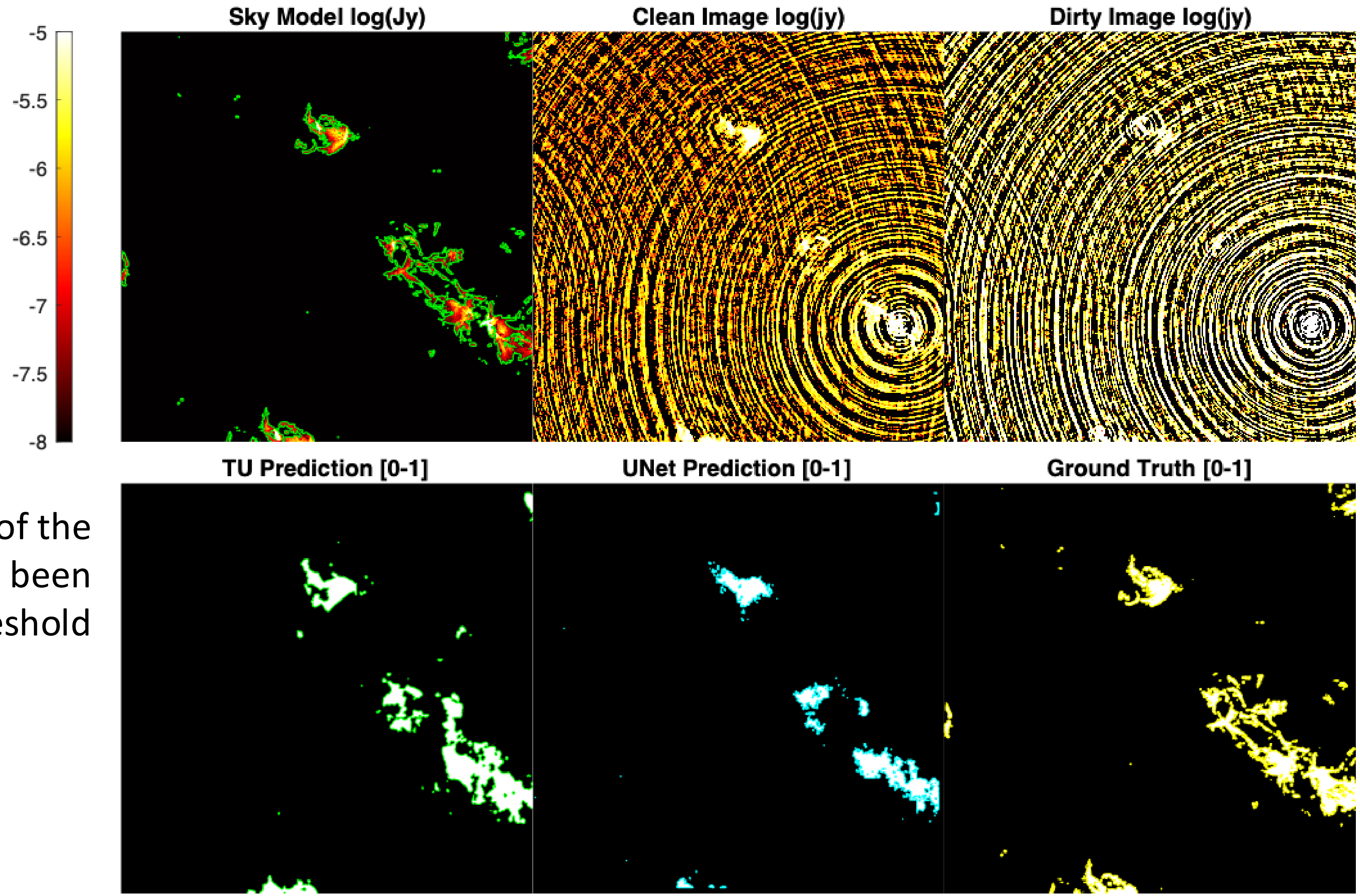
Comparison of Evaluation Metrics

- Dataset with 140 mock images (2000 ×2000 px)
- training 100 images, testing 40 images
- nominal spatial resolution 2 arcsec
- single frequency 150 MHz
- ground-truth comprising 2 classes:
  - ✓ class 1: emission (3.5%)
  - ✓ class 0: non emission (96.5%)



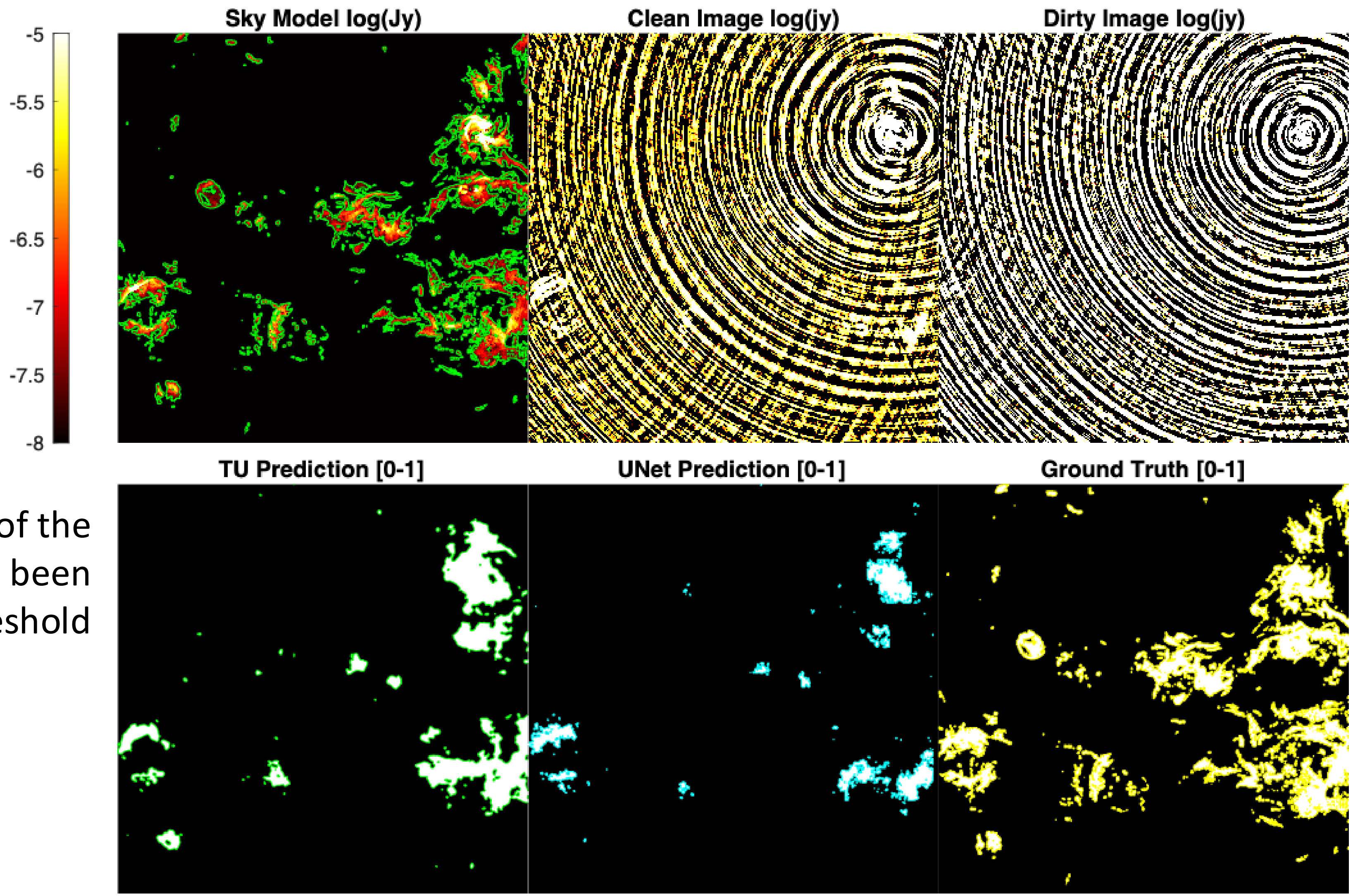
Computational Perfomance (NVIDIA A100)	TransUNet	UNet
Inference time (sec/image)	2.9	0.36
Training time (sec/epoch)	88.5	13.5
Image resolution	2000x2000	2000x2000

Results



The mask (ground truth) of the network prediction has been set with probability threshold > 0.5

Results

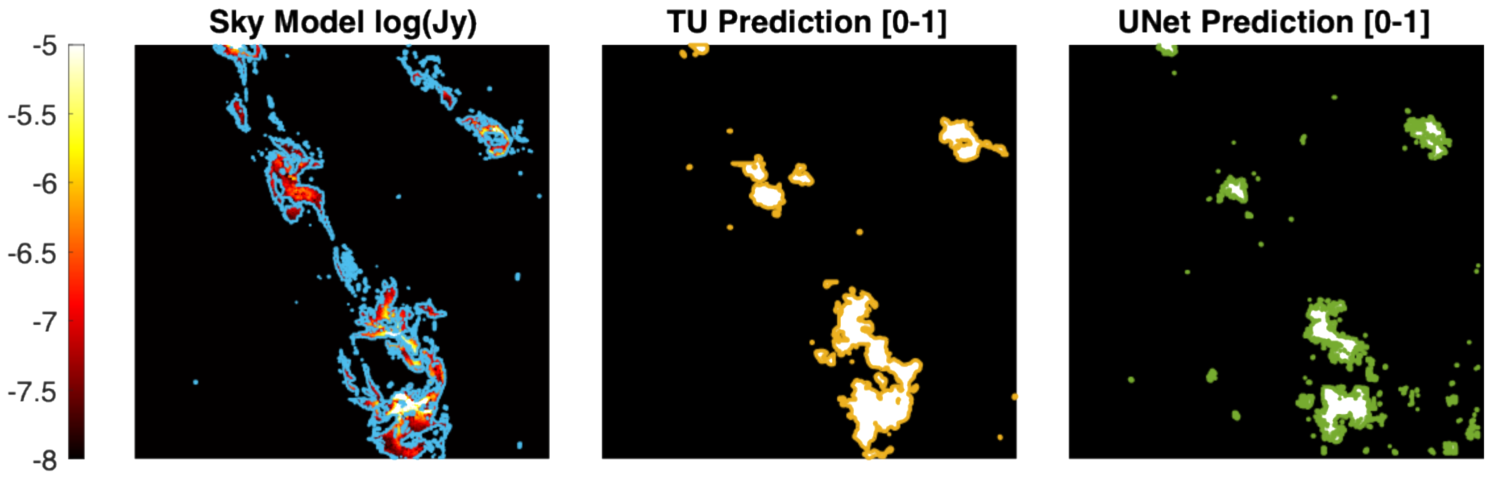


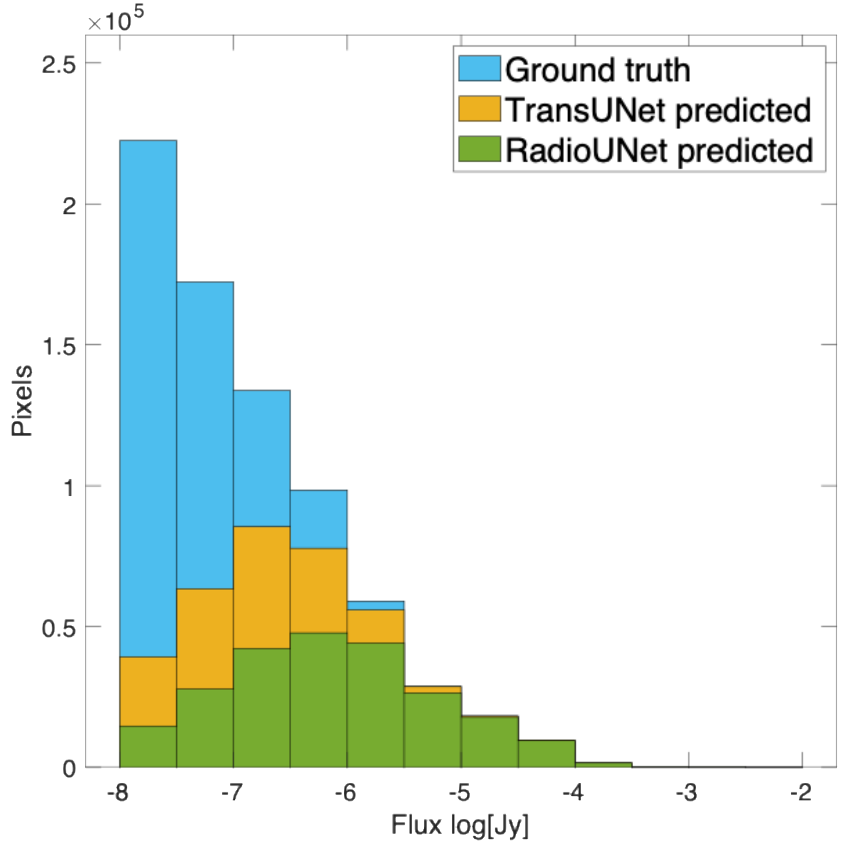
The mask (ground truth) of the network prediction has been set with probability threshold > 0.5

Results: Flux Distribution of the test dataset

Flux distribution of:

- all pixels of the sky images with flux > 10^{-8} Jy, i.e. ground truth
- all pixels of the sky images overlapping the TransUNet predictions
- all pixels of the sky images overlapping the RadioUnet predictions





Take Home Messages

- **The same network** can be used successfully on **completely different data and fields**
- Deep Learning can be **effectively used to process and analyze large and complex astronomical images**
- **Vision transformers** can be effective in enhancing segmentation performance of traditional convolutional approach but application on real astronomical images is to be investigated

THANK YOU
FOR YOUR ATTENTION