



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani

PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Centro Nazionale di Ricerca in HPC,  
Big Data and Quantum Computing

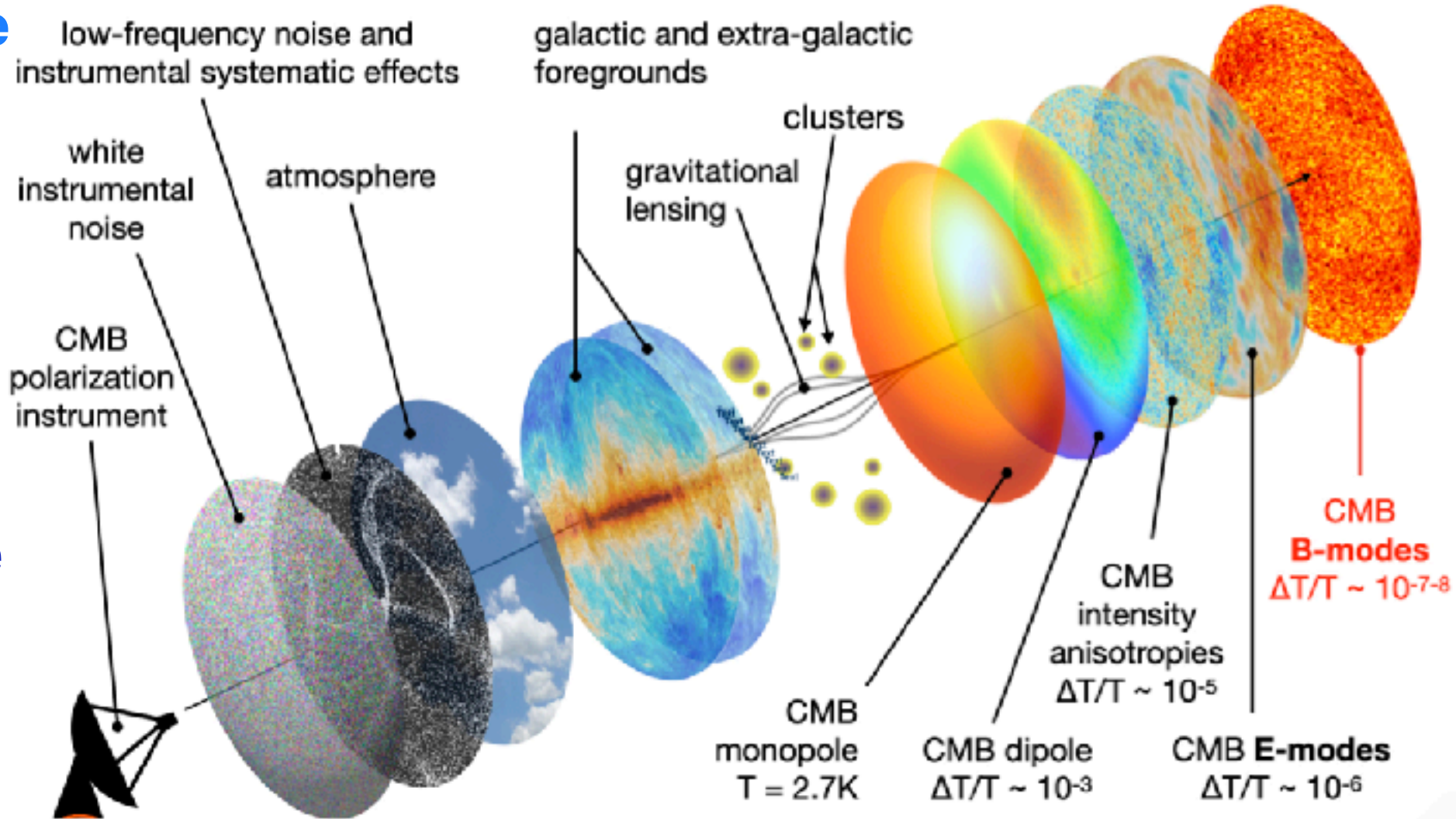
# *Learning Molecular Cloud emission with Neural nets*

*Giuseppe Puglisi (UniCT),  
Avinash Anand, Marina Migliaccio, Domenico Marinucci (UniToV)*

**Spoke 3 General Meeting, Elba May 5 / 9, 2024**

# Scientific Rationale

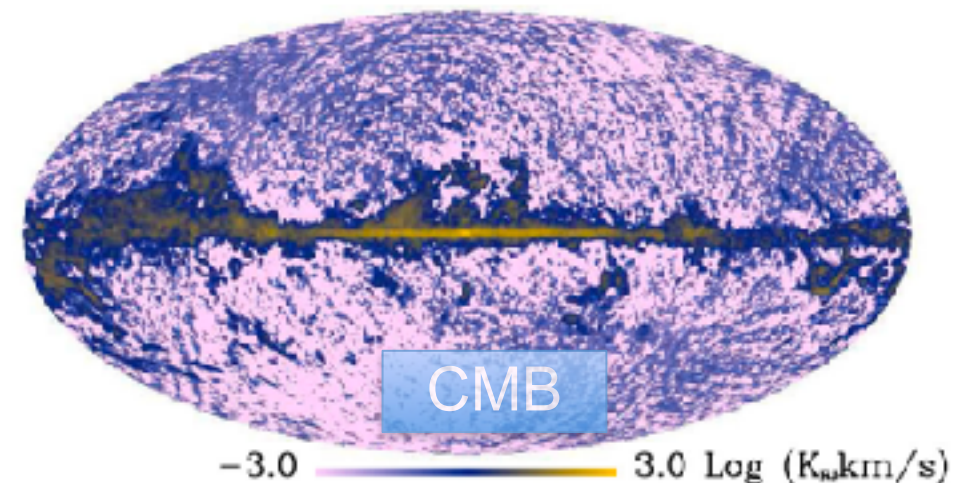
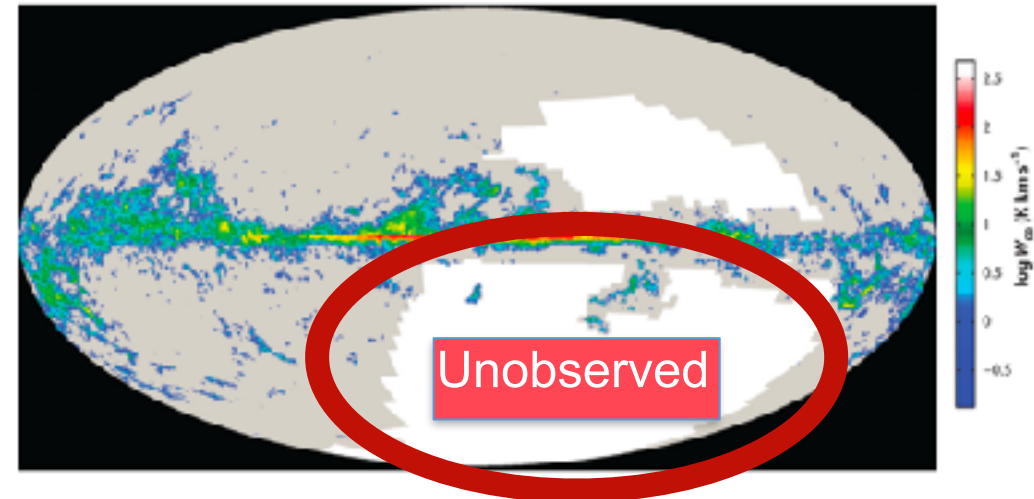
The Milky Way acts as a foreground wrt Cosmic Microwave Background (CMB)



Credit: J. Errard

## Scientific Rationale

- Full sky maps of Galactic emissions are needed for cosmological observations.
- There are regions that are not observed ... yet
- In the same area, CMB ground telescopes are observing...
- *Planck* data observed full sky, BUT also full of noise



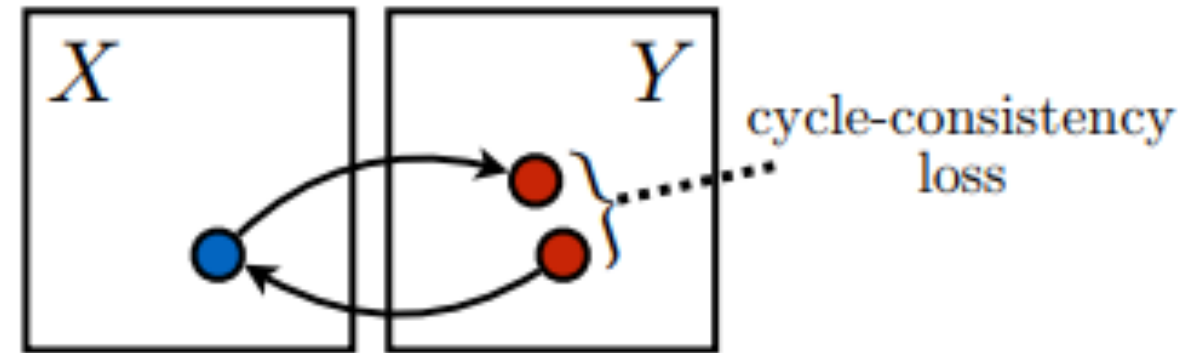
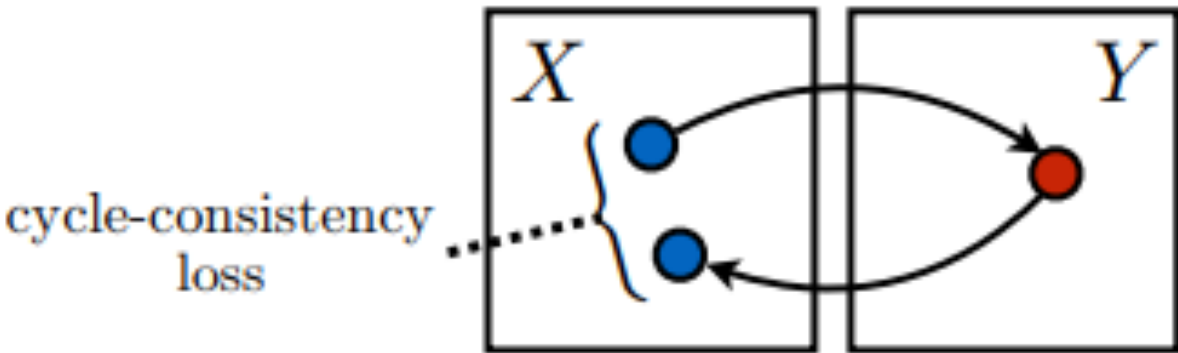
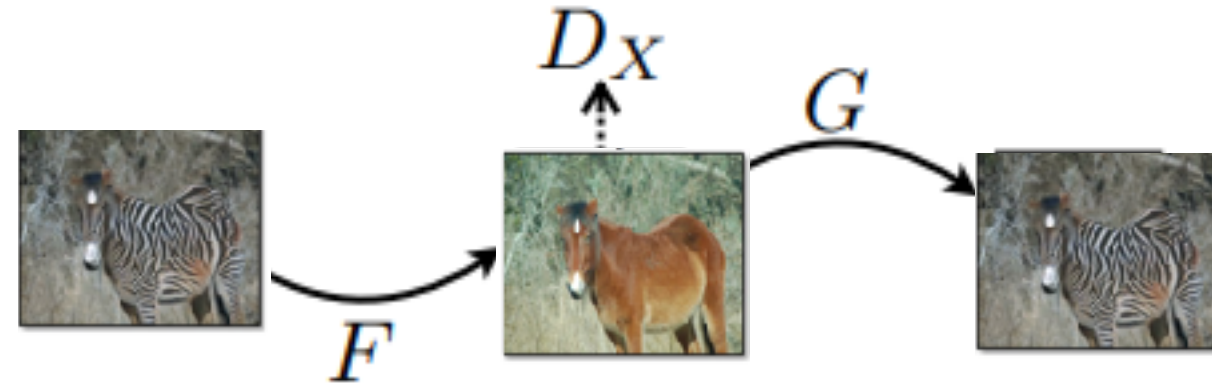
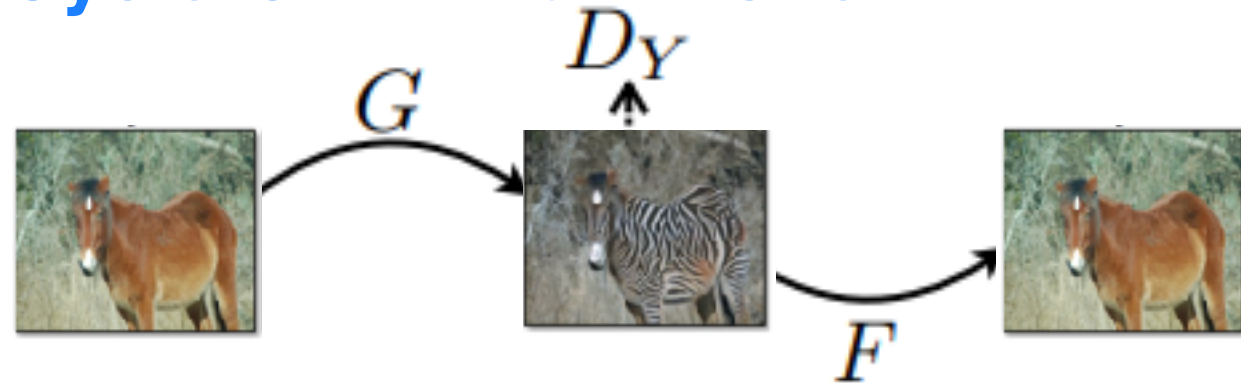
## Next Steps and Expected Results (by next checkpoint: April 2024)

- Reached a saturation in training phase with Res-UNet,
- Need to explore the latent space feature
- Jan 2024: trained data and first results from test-set
- Apr 2024: present stable results

## Next Steps and Expected Results (by next checkpoint: April 2024)

- Reached a saturation in training phase with Res-UNet > *de-noised dataset with non-local mean filter* (Quispe Pena & Frolov 2023)
- Need to explore the latent space feature > *change of architecture with Cycle-Consistent Adversarial Networks, Cycle-GAN*, Zhu J.Y. et al. 2017
- Jan 2024: trained data and first results from test-set
- Apr 2024: present stable results

# Cycle-GAN in a nutshell

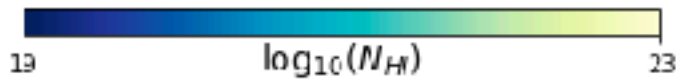
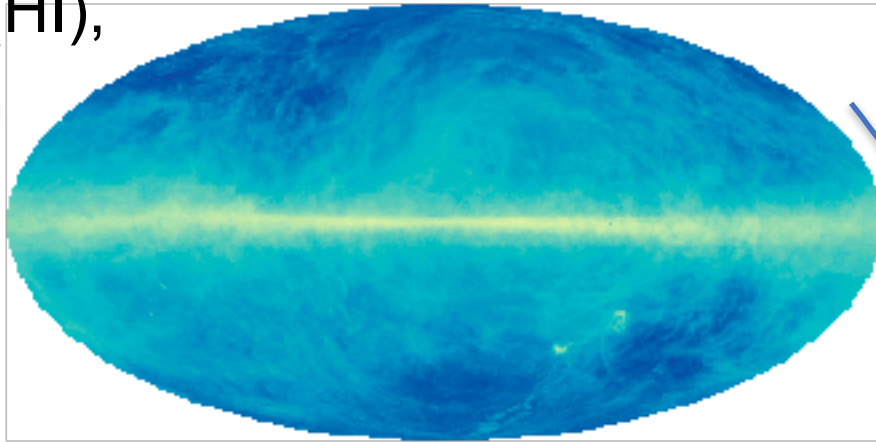


$$\mathcal{L}_G = \mathcal{L}_F = \text{binary Cross-Entropy}$$

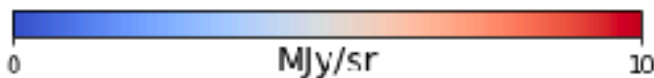
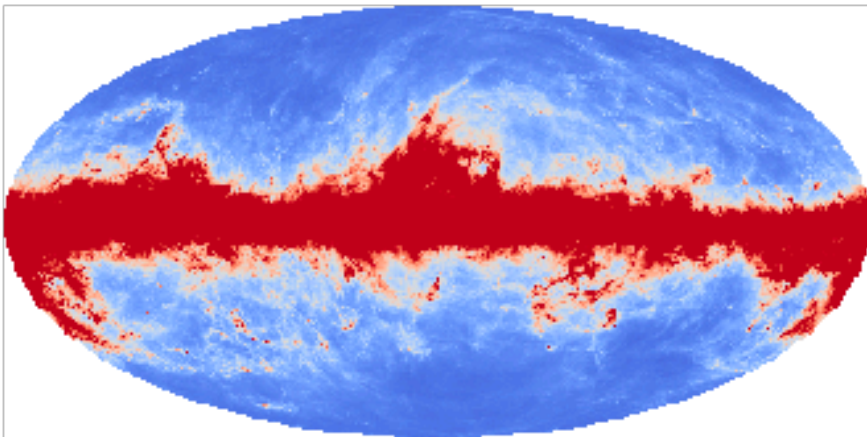
$$\mathcal{L}_{identity} = |Y - G(X)| + |X - F(Y)|$$

Input:  $N(\text{HI})$ , IR dust

HI4PI

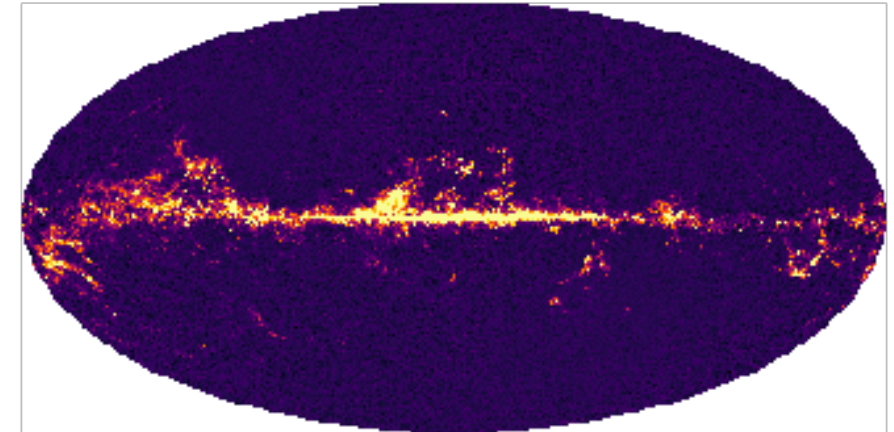


GNILC-857 GHz



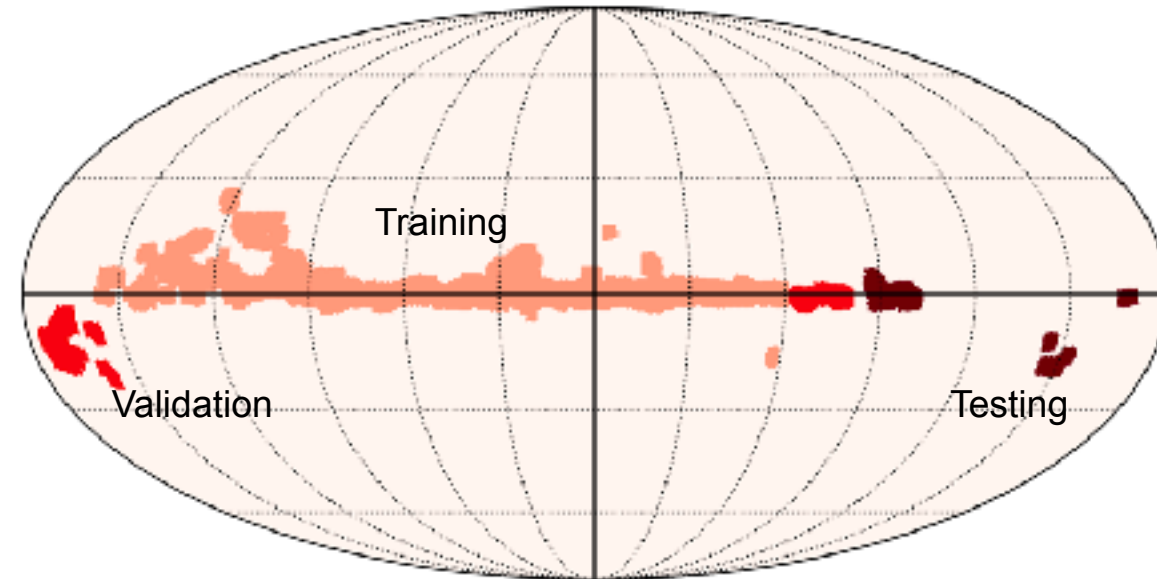
Target: CO map

CO:1-0 Planck (Type2)



# Technical Objectives, Methodologies and Solutions

- Build training set from available data (*Planck, HI4Pi*)
- Identify Galactic regions of bright emission, low noise contribution, -> high SNR >8
- Create the training set from those areas
- With augmentation patch:
  - ~~5650(training), 930(validat.), 2790 (test.)~~
- With de-noising:
  - 10,488 (training), 1166 (validat.) , 747(test.)

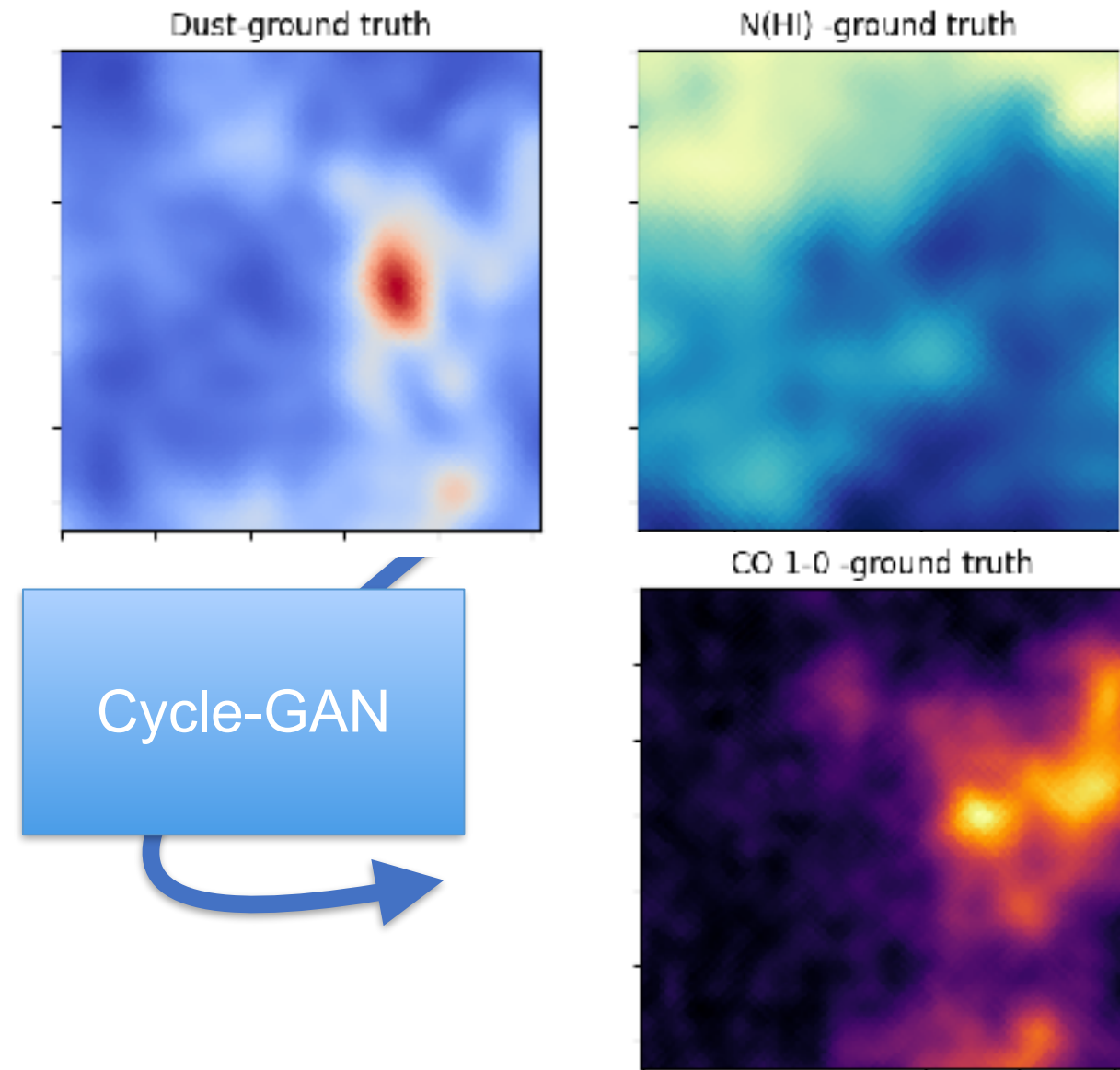




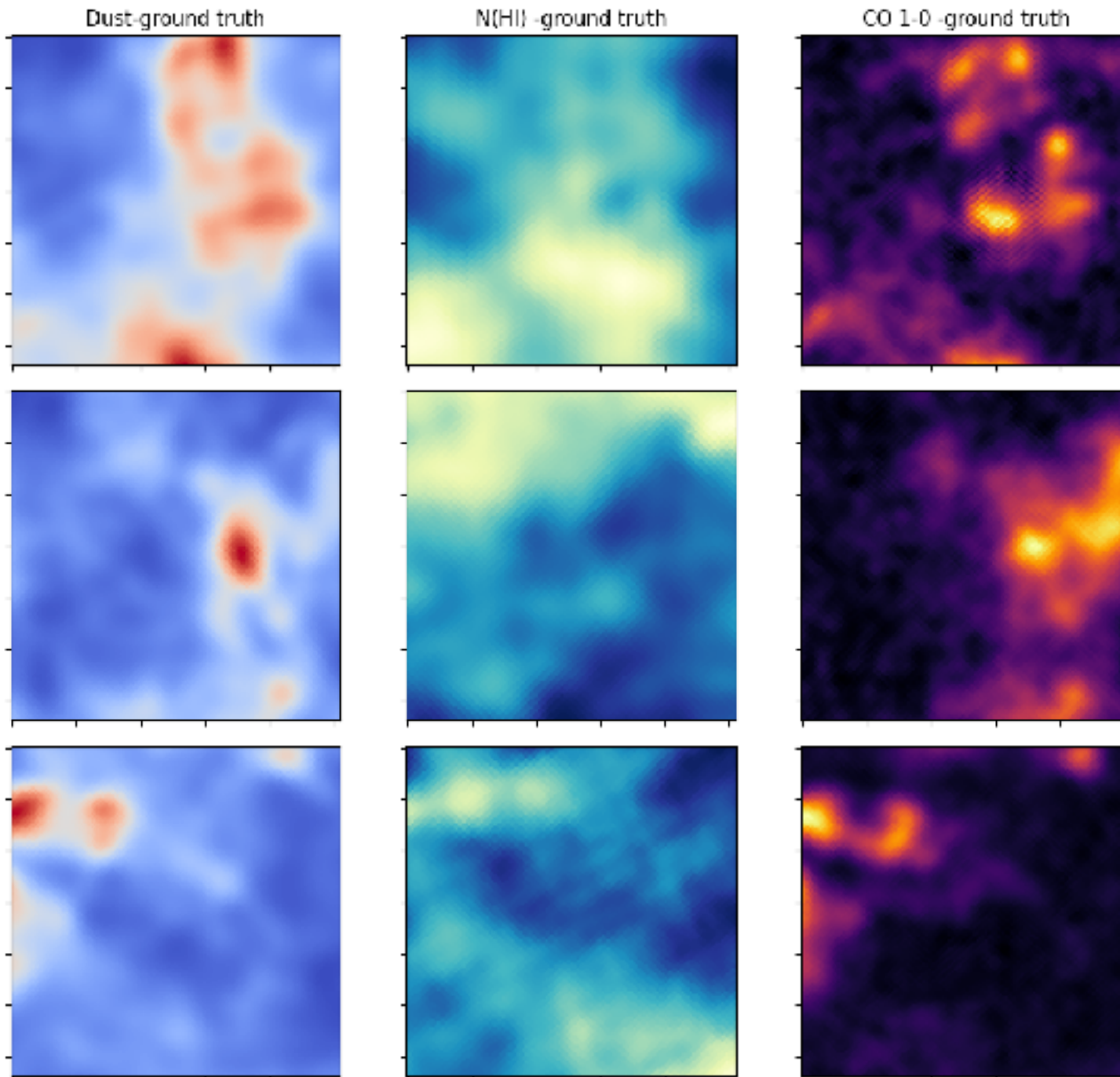
# Methodologies

## Training Cycle-GAN

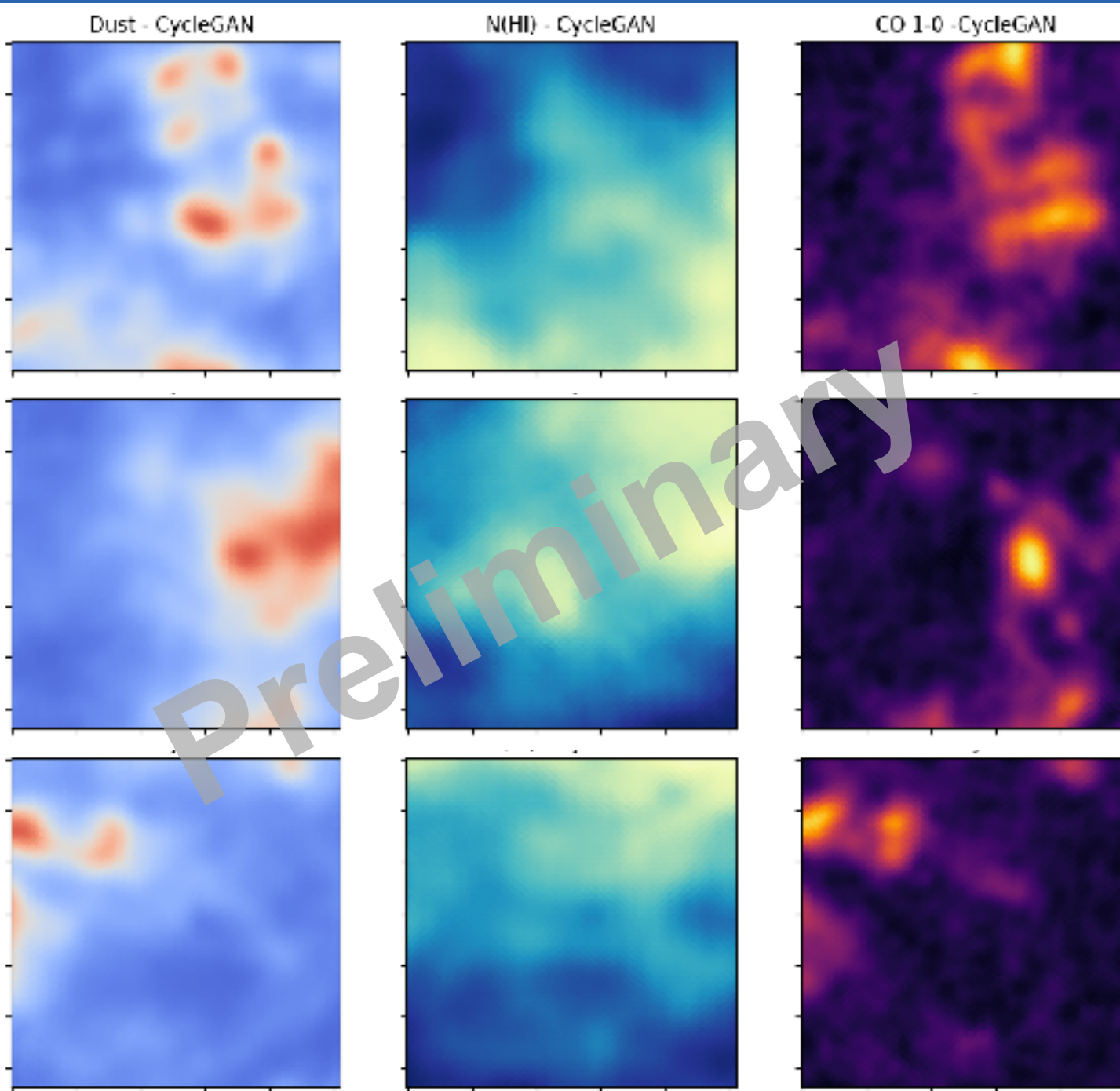
- batch size= 16
- training time performed on NVIDIA A100-SXM4-40GB (4GPUs @NERSC)
- 2000 epochs
- 80% accuracy
- 3x3 deg2 maps (128x128)



# Results on Test set



# Results on Test set



# Timescale, Milestones and KPIs

