A BRAIN study to tackle imaging with artificial intelligence in the ALMA2030 era

Guglielmetti, F.; Delli Veneri, M.; Baronchelli, I.; Johnson, V.; Arras, P.; Blanco, C.; ... ; Villard, E.

Citation

Guglielmetti, F., Delli Veneri, M., Baronchelli, I., Johnson, V., Arras, P., Blanco, C., ... Villard, E. (2023). *A BRAIN study to tackle imaging with artificial intelligence in the ALMA2030 era*. doi:10.5281/zenodo.10221271

A BRAIN study to tackle imaging with artificial intelligence in the ALMA2030 era

Guglielmetti Fabrizia[1*], Delli Veneri M.[2], Baronchelli I.[3], Johnson V.[4], Arras P.[2], Blanco C.[2], Brescia M.[5], Dosi A.[5], Enßlin T.[2], Longo G.[5], Roth J.[2], Stoehr F.[1], Tychoniec Ł.[1], Villard E.[1]

(1) European Southern Observatory, Garching, Germany (2) National Institute for Nuclear Physics, Naples, Italy (3) INAF IRA, Bologna, Italy; Italian ALMA Regional Centre, Bologna, Italy (4) Max Planck Institute for Astrophysics, Garching, Germany (5) University of Naples Federico II, Italy

* fguglie@eso.org


Credit https://almaobservatory.org

① INTRODUCTION

ALMA, the Atacama Large Millimeter/submillimeter Array, with the planned electronic upgrades will deliver an unprecedented amount of deep and high resolution observations [1]. Sparse sampling, large variety of celestial sources' morphology and their intensities, instrumental responses, pervasive presence of noise increase complexities to the demanding task of image reconstruction. Wider fields of view are possible with the consequential cost of computing time. Alternatives to commonly used applications in image processing have to be sought and tested [2,3].

Astrostatistics and astroinformatics offer interdisciplinary approaches at the intersection of observational astronomy, statistics, algorithm development, and data science [4]. In this study, we provide evidence of the benefits in employing these approaches to ALMA image analysis for operational and scientific purposes. We show the potentials of two techniques, RESOLVE [5,6,9] and DeepFocus [7], applied to ALMA calibrated science visibilities. Significant advantages are provided with the potential to improve the quality and completeness of the data products and overall processing time. Both approaches evidence the logical pathway to address the incoming revolution in data analysis dictated by ALMA2030 [1]. Moreover, we bring to the community additional products through a new package (ALMASim) to promote advancements in these fields, providing a refined ALMA simulator usable by a large community for training and/or testing new algorithms [8].

② METHODS

RESOLVE:

Input: ALMA calibrated measurement set in the uv plane

$$d = Re^s + n.$$

Data *d* modelled as a combination of celestial signal *s* corrupted by the dirty beam R and by the noise *n* (systematic and random errors)

The process of synthesizing an image involves estimating the posterior probability density function (pdf) of potential true sky signal configurations using variational inference:

$$P(s|d) = \frac{e^{-H(s,d)}}{Z(d)}$$

Output: Reconstructed sky image, uncertainty map and power spectrum

DeepFocus:

Input: ALMA dirty cubes

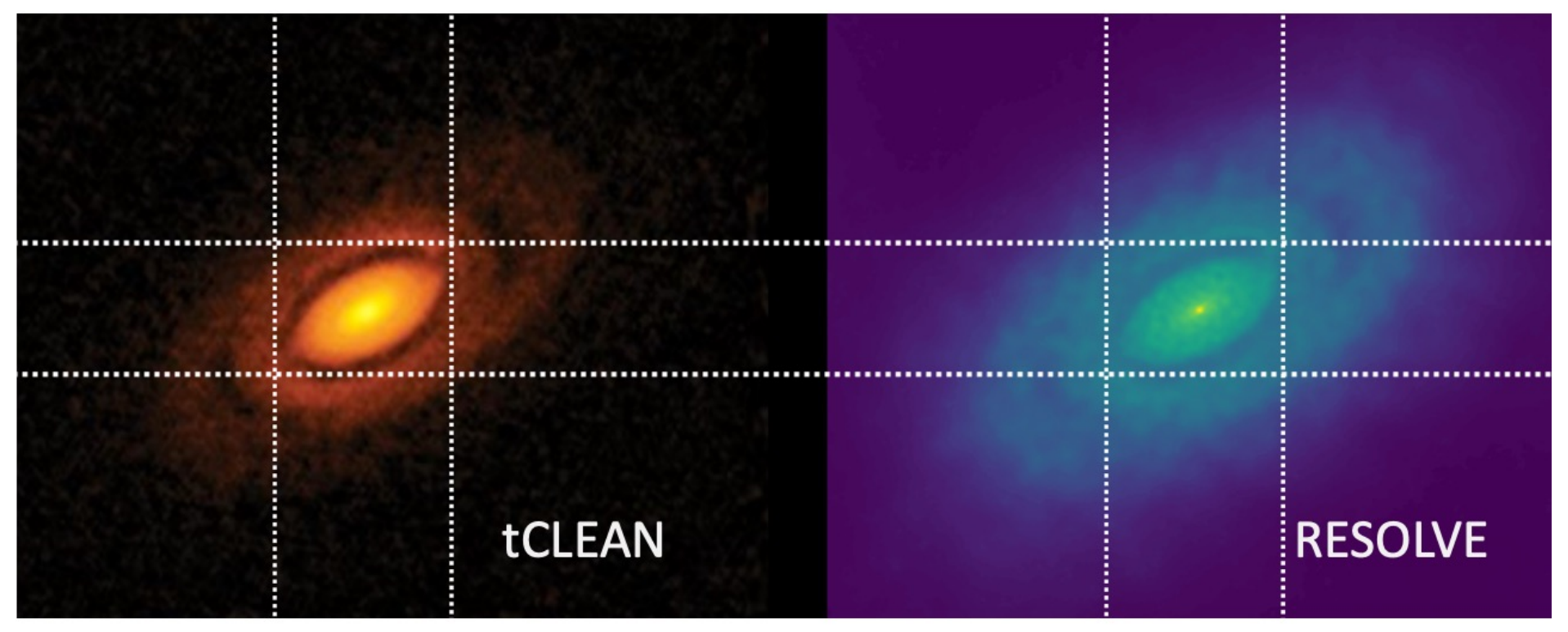
Deep Learning Networks and Autoencoders used to solve for I from ALMA dirty cubes corrupted by noise and convolved with the dirty beam:

$$I^D = I_{db} * I + n$$

Bayesian parameter optimization is enabled to search for the most performing architecture given a task and a set of Interferometric data.
Training/Testing/ Validation are executed through this supervised machine learning technique.

Output: Reconstructed sky cubes (deconvolution)
Source detection and characterization

③ RESULTS



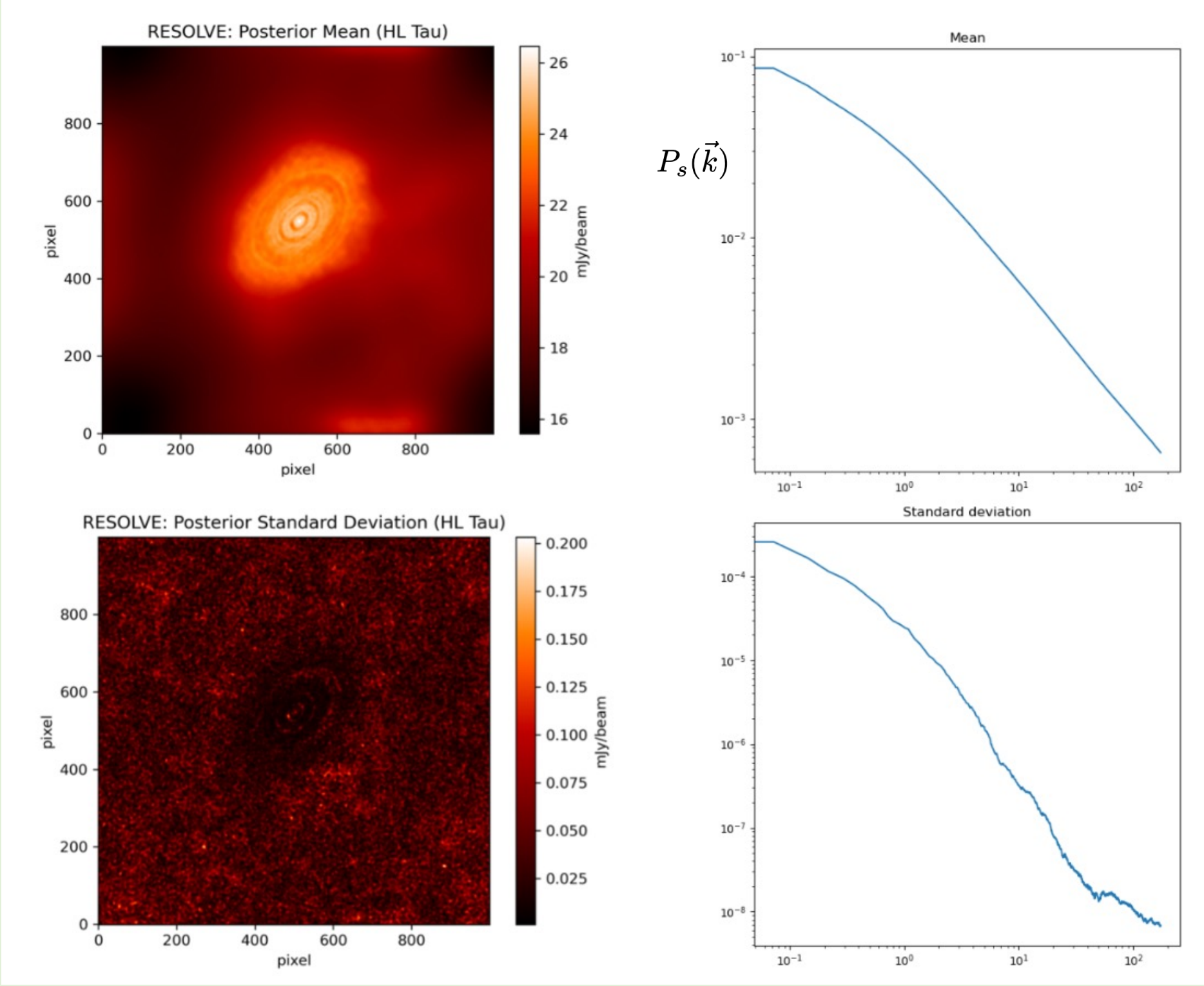
Elias 27 is a young star located in the Orion Nebula. Structures around the young star are detected with both techniques. This high-resolution observation is better described by RESOLVE in the detailed detection of both the protoplanetary disk and both the central part.

*Benchmarking **RESOLVE** on real data:*

Application to Elias 27 from the DSHARP ALMA project at 240 GHz (1.25 mm) continuum.

On the left, fiducial image as given by the DSHARP team and using the CLEAN algorithm https://almascience.eso.org/almadata/lp/DSHARP/

On the right, RESOLVE sky map.

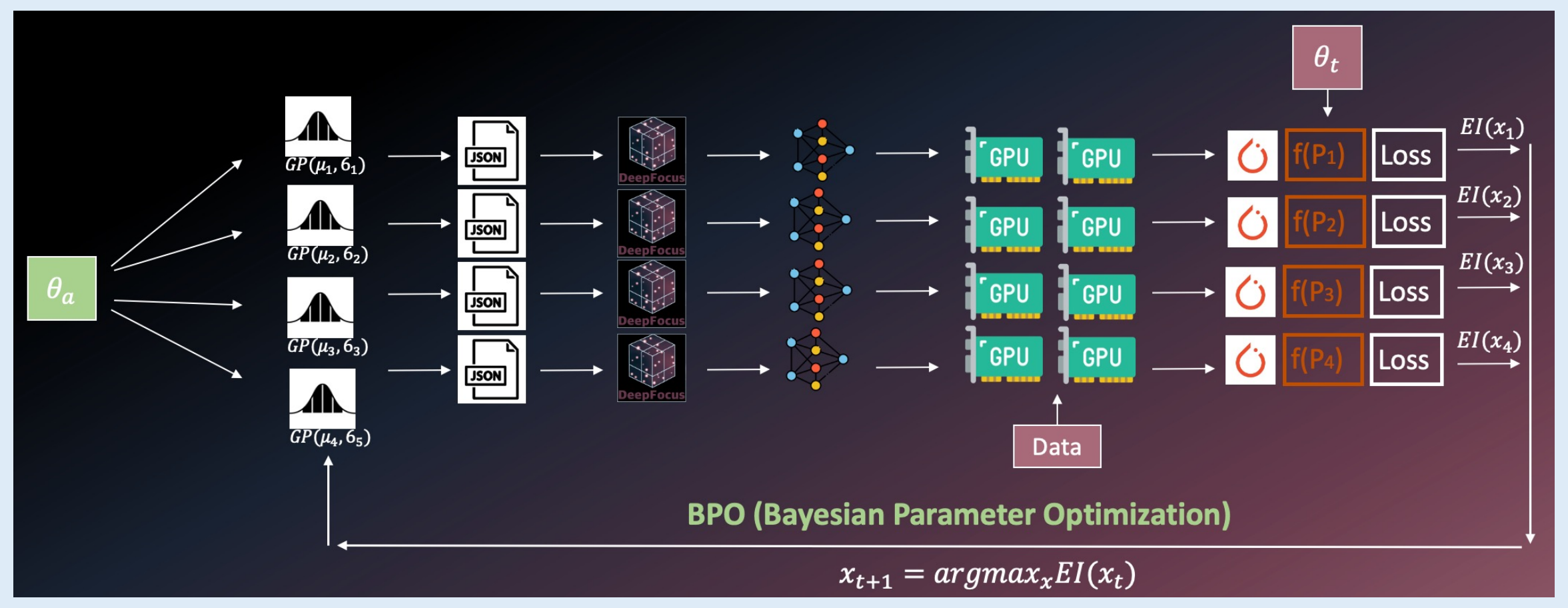


Example of products provided by RESOLVE:

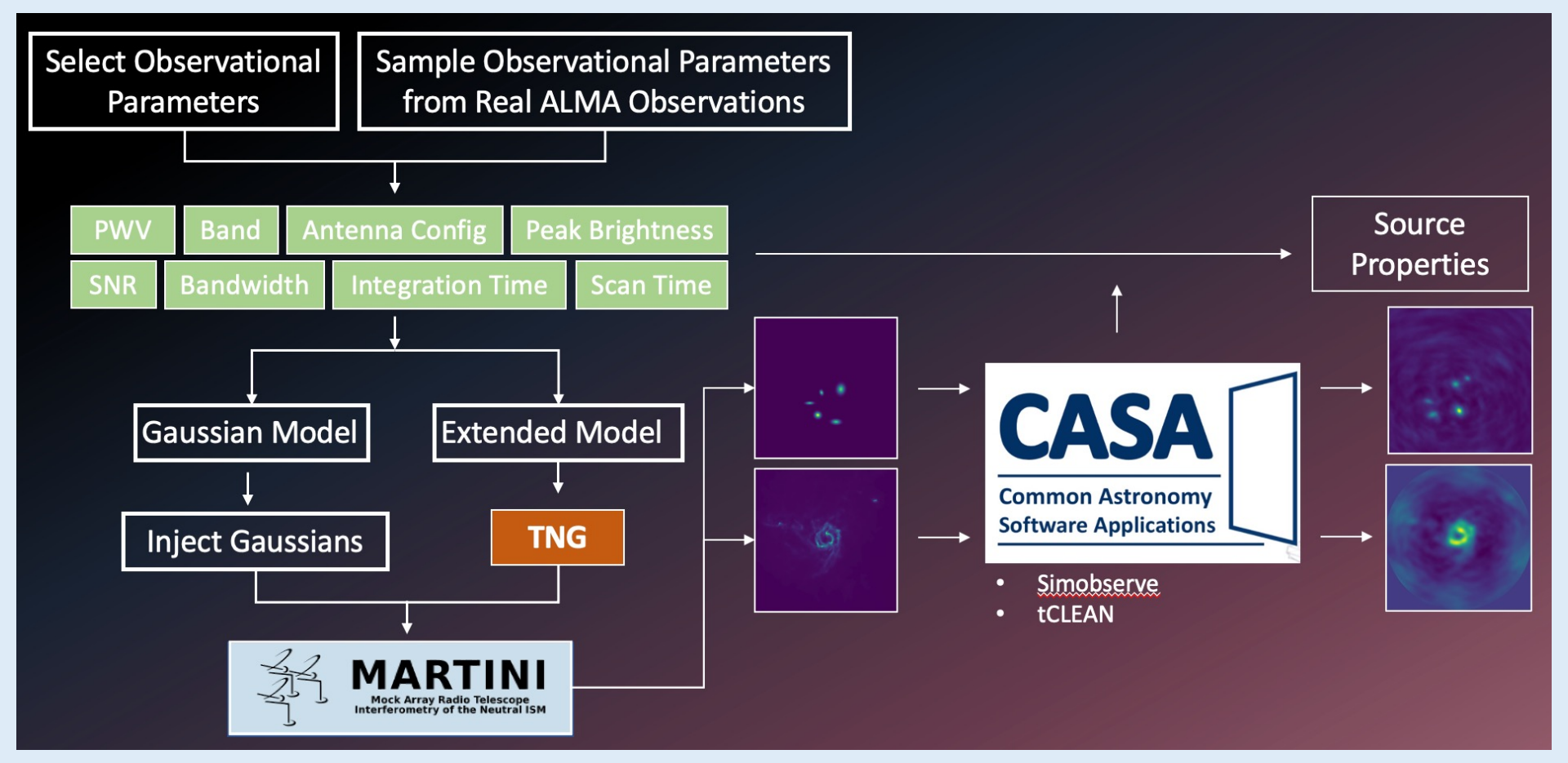
On the left, the mean posterior pdf (upper) and its uncertainty map (lower) of the estimated celestial signal for the HL-Tau data in band 6 (Brogan, C.L. et al. 2015, ApJ, 808(1): L3)

On the right, power spectrum estimation corresponding to the estimated posterior mean (upper) and its uncertainty (lower).

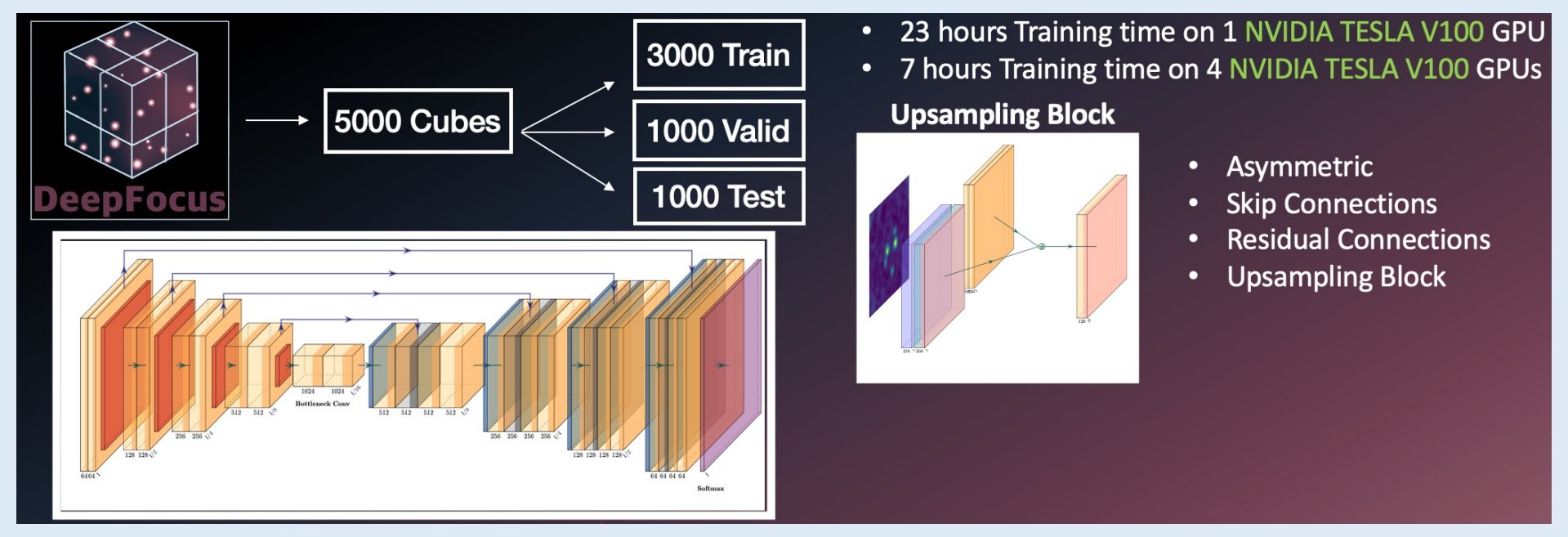
The signal's power spectrum Ps(k) describes how the signal's variance is distributed over the different frequencies of the signal.



DeepFocus is a Meta-Learner designed to search for the most effective architecture based on a given task and a set of interferometric data. A Bayesian optimization algorithm is used for model selection.



DeepFocus is a supervised machine learning algorithm. To Train/Test and Validate, the **ALMASim** package has been developed: generates mock observations of high redshift point sources and low redshift galaxies and extended structures, both continuum and emission lines in ALMA cubes.

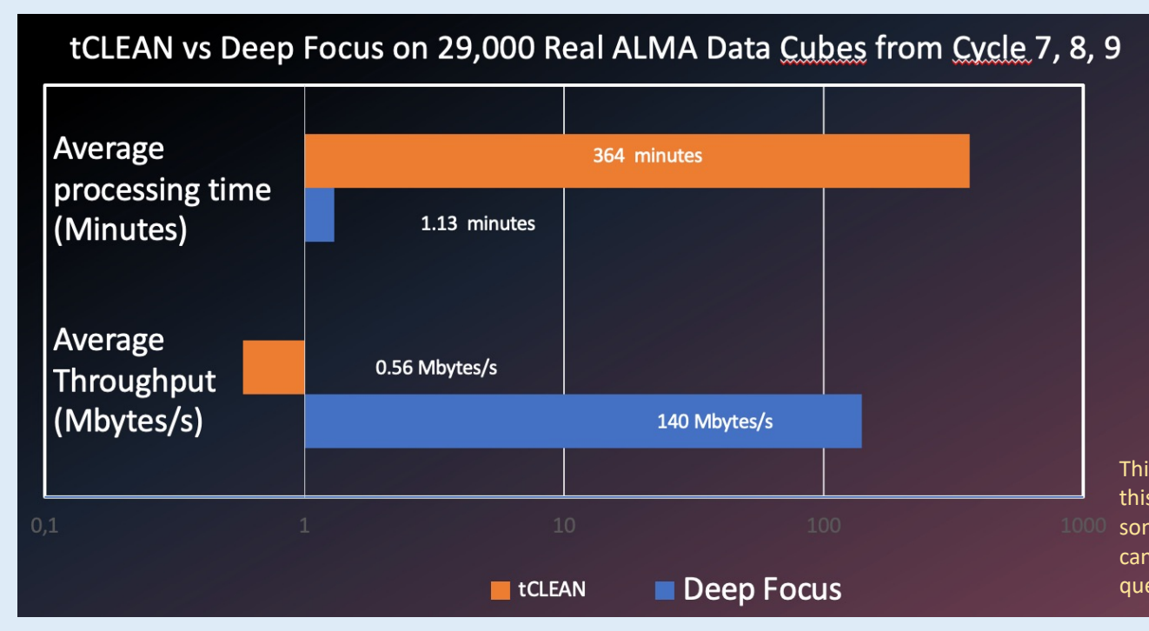


- 23 hours Training time on 1 NVIDIA TESLA V100 GPU
- 7 hours Training time on 4 NVIDIA TESLA V100 GPUs

Upsampling Block

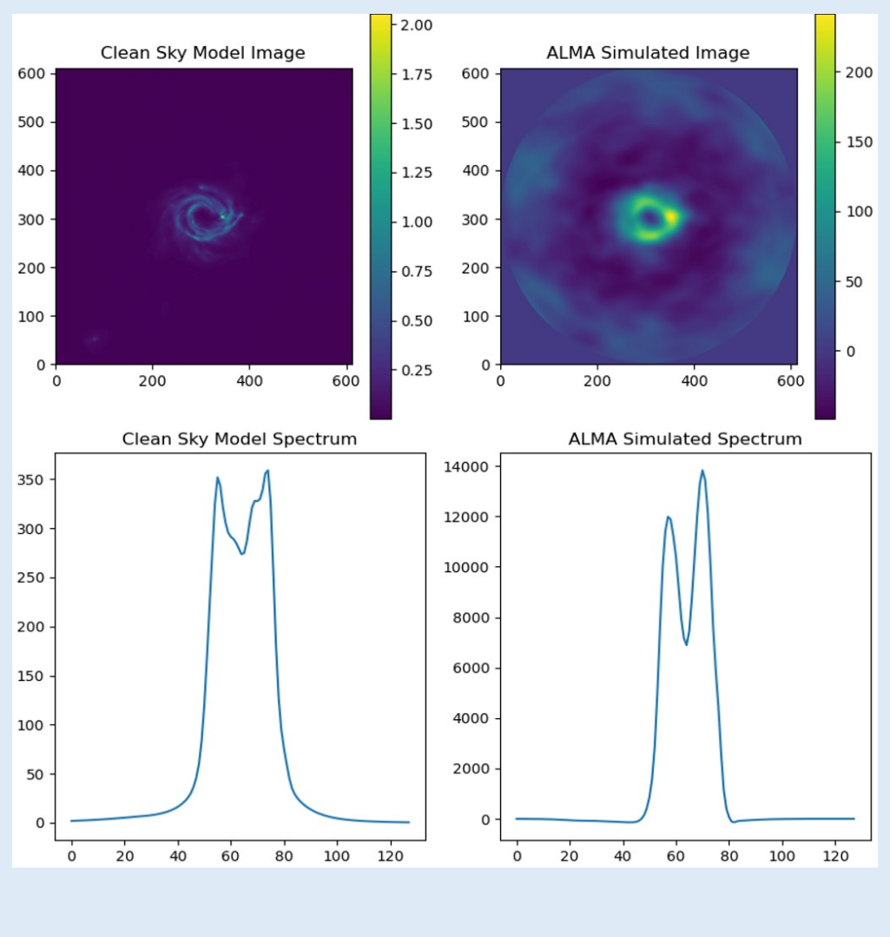
- Asymmetric
- Skip Connections
- Residual Connections
- Upsampling Block

tCLEAN vs Deep Focus on 29,000 Real ALMA Data Cubes from Cycle 7, 8, 9

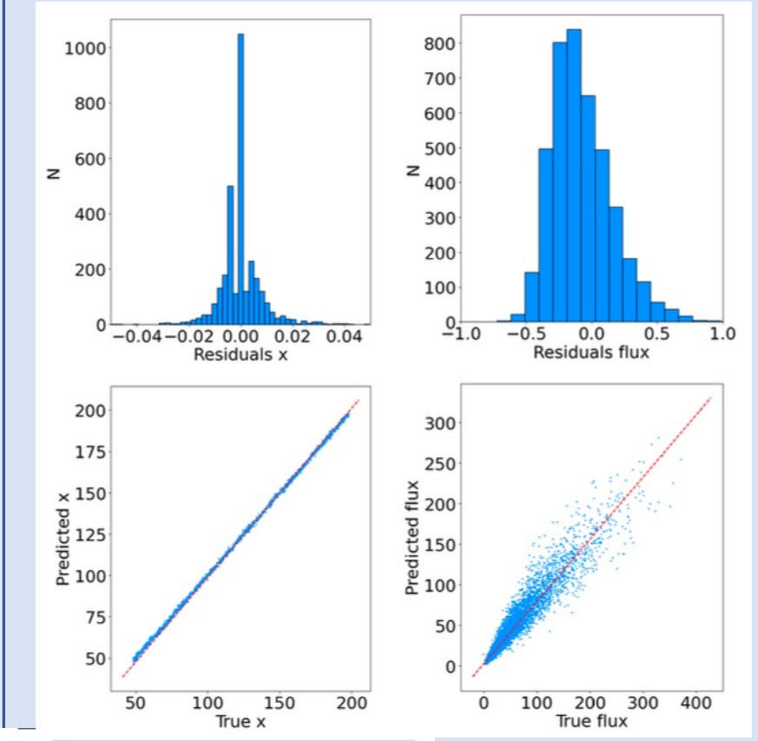


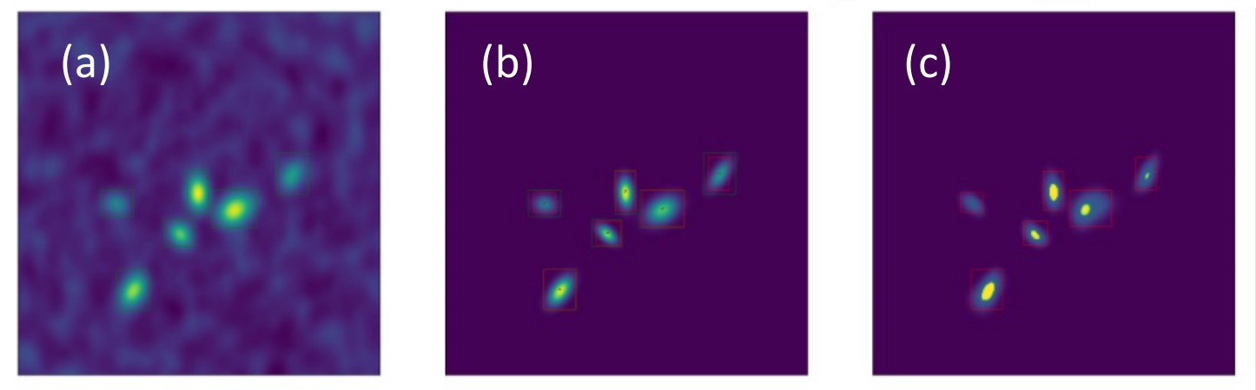
- On the same hardware, depending on the cube, we register a gain in speed of a factor between x280 and x5500
- Computations which took days are handled in minutes by DeepFocus

This result provides a preliminary estimate since, at this stage of development, a definite robust comparison is not feasible. However, it underscores the significant advantage posed by artificial intelligence techniques in ALMA 2030 operations.



DeepFocus comparison on ALMA Simulated Cubes: Scatter plots of true against predicted parameters and the corresponding residuals histogram on 5000 simulated line emitters.



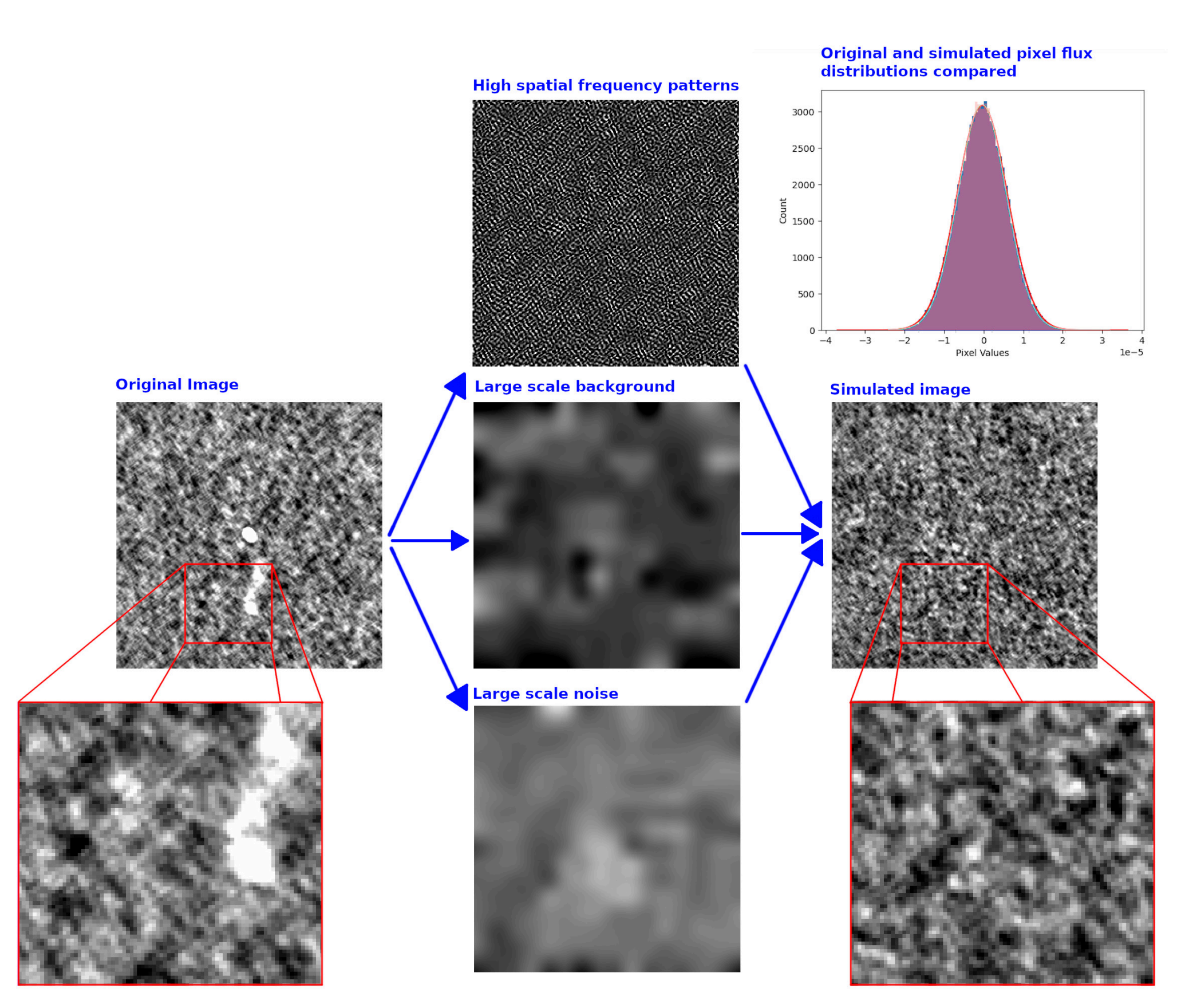


(a) Input Dirty Cube (b) Simulated Sky model (c) DeepFocus prediction

④ Empirical Noise

Rather than relying solely on artificial noise added to a simulated "real sky", an empirical approach to noise modelling is used and applied to **ALMASim** to encompass a broader range of noise components.

Noise components are characterized at larger scales and at scales similar to and smaller than the ALMA beam size from observed data. By extracting high spatial-frequency noise patterns present in the image, we try to capture the intricate details of the noise "fabric", reproducing the complexity of the noise environment.



When replicated to create the new simulated image, we take care to preserve the statistical properties of the flux distribution on individual pixels, matching those observed in the original image (except in regions originally occupied by the astronomical sources of interest).

In summary, preserving the statistical properties of pixel distributions of real images and replicating the diverse noise components and patterns in simulated images, we can aim for more realistic ALMA simulations by capturing intricate noise characteristics observed in ALMA data.
ALMASim opens doors for the creation of tailored AI imaging algorithms designed to address unique scientific cases.

⑤ CONCLUSION

Novel imaging techniques applicable on large data volume and tailored in view of the challenges of ALMA 2030 are explored to enable efficiency improvements in data processing while requiring the least amount of human intervention.

RESOLVE is a robust algorithm founded on a principled method. It is designed for the detection of diffuse emission. Complex structures in the celestial signal and point-like sources are well detected. The reconstructed images provide a reliable solution with no need of extra human intervention. Although RESOLVE is computing-expensive, the algorithm delivers, in addition to the reconstructed image, other informative products (such as the uncertainty map, the power spectrum and its uncertainty). These products have the potential to lay the foundations for designing a fully automated pipeline.

DeepFocus demonstrated high image fidelity and high-performance computing for image reconstruction on ALMA data cubes. The technique is applied on ALMA dirty cubes, and it learns the celestial sources, the noise, and the instrumental point-spread function from the image. It allows for extreme data compression by leveraging both spatial and frequency information. This algorithm has the potential to revolutionize data management in science archives. DeepFocus may allow one to create images on user demand through a web-interface.

REFERENCES

[1] John Carpenter, Daisuke Iono and Francisca Kemper et al. , 2020, "The ALMA Development Program: Roadmap to 2030" arXiv:2001.11076
[2] Fabrizia Guglielmetti et al., 2022, "Bayesian and Machine Learning Methods in the Big Data Era for Astronomical Imaging'', *Phys. Sci. Forum,* 5(1), 50
[3] Fabrizia Guglielmetti et al., 2023, "A BRAIN study to tackle image analysis with artificial intelligence in the ALMA 2030 era'', arXiv:2311.10657
[4] Aneta Siemiginovska et al., 2019, "Astro2020 Science White Paper: The Next Decade of Astroinformatics and Astrostatistics'', arXiv:1903.06796
[5] Henrik Junklewitz et al., 2016, "RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy'', A&A, 586, A76
[6] Łukasz Tychoniec et al., 2022, "Bayesian Statistics Approach to Imaging of Aperture Synthesis Data: RESOLVE Meets ALMA'', *Phys. Sci. Forum*, 5(1), 52
[7] Michele Delli Veneri et al., 2023, "3D detection and characterization of ALMA sources through deep learning'', MNRAS, 518, 3
[8] https://pypi.org/project/ALMASim/
[9] Jakob Roth et al., 2023, "Bayesian Radio Interferometric imaging with direction –dependent calibration'', A&A, 678, A177
RESOLVE: https://gitlab.mpcdf.mpg.de/ift/resolve
DeepFocus: https://github.com/MicheleDelliVeneri/DeepFocus

"Without community there is no liberation..." Audre Lorde