IL NUOVO CIMENTO **47 C** (2024) 304
DOI 10.1393/ncc/i2024-24304-8

COMMUNICATIONS: SIF Congress 2023

Deep Focus —A meta-learner solving inverse problems based on Deep Learning: Applications to Radio Interferometry

M. DELLI VENERI(*) on behalf of the BRAIN STUDY COLLABORATION
INFN Section of Naples (INFN) - Via Cintia, 80126, Naples, Italy

received 28 January 2024

Summary. — We present Deep Focus (DF) results on the ALMA image deconvolution problem for point-like sources. We compare DF reconstruction capabilities and execution times with those of CLEAN and other Deep Learning algorithms, showing improvements in reconstruction capabilities with respect to both, and an average speed-up factor of 280 with respect to CLEAN.

1. – Introduction

The detection of sources within data cubes produced by Radio Interferometers, requires the resolution of an ill-posed inverse problem. This problem entails recovering the underlying signal $G(l, m)$ from the observed sky $D(l, m)$ in sky coordinates and frequencies. The relationship between the two is described by the Van Cittert-Zernike theorem,

$$D(l, m) = P(l, m) \otimes T(l, m) \otimes G(l, m) + N(l, m), \tag{1}$$

where $\otimes$ is the convolution operator, while $P(l, m)$ and $T(l, m)$ are the primary and dirty beams, two operators describing the response function of each antenna and the interferometric response function of the array, respectively), $N(l, m)$ represents noise, and l and m are the coordinates in the image plane. Traditionally, this problem is solved through the CLEAN algorithm [1], which works by iteratively subtracting a fraction of the signal convolved by the dirty beam. At the end of the process, this signal is restored by a "clean" Gaussian beam, hereby removing the sidelobe pattern from the data. The planned upgrades to ALMA will results in an increase of data collection by roughly a factor of 100 [2]. Deep Learning (DL) autoencoder-like models (with heads constructed as ConvNet, ResNet, U-Net, VGG-Net, DenseNet) have been applied to the resolution of the ALMA deconvolution problem showing comparable performances with CLEAN

(*) E-mail: `delliven@na.infn.it`

and other traditionally used algorithms [3-7] in case of Gaussian-like simulations. In this work we build upon the work outlined in [7], and we present Deep Focus (DF) a meta-learner aimed at optimising the resolution of the ALMA deconvolution problem by exploring the space of DL models architectures (Architecture Hyperparameter Space, or AHS). We measure DF Sky Model reconstruction capabilities on simulated ALMA cubes produced through our simulation pipeline ALMASim [8] showing that it surpasses the performances of other DL models and CLEAN, and we measure the average processing time on 29×10^3 archived real ALMA cubes showing a substantial increase in processing speed over CLEAN. The article is structured as follows: in sect. **2** we describe DF, in sect. **3** we discuss ALMASim and in sect. **4** we report our experimental results.

2. – Deep Focus

Deep Learning models need to be trained, *i.e.*, presented with examples of Observed and true Sky Models pairs, in order to learn how to approximate the unknown mapping between them. Usually, training a model involves the optimization of the so-called training hyperparameters, *i.e.*, a set of parameters that significantly impact the model's learning dynamics, convergence, and generalisation performances. In the context of DF, we define a point in the AHS as the set of parameters that are needed to fully describe the architecture of a DL model: number of convolutional layers, number of channels, kernel size, presence or absence of Skip Connections, Residual Connections, Bottleneck blocks, Multi-Head self-attention, drop-out, type of pooling strategy. DF takes architectural hyperparameters as input and builds the corresponding DL architecture. Through Bayesian optimization, DF samples the architectural parameter space in order to minimize the problem loss function. Bayesian optimization is a probabilistic model-based approach for optimizing expensive, black-box functions such as DL models, employing surrogate models: *i.e.*, the best set of parameters for a given objective is efficiently found while reducing the number of expensive model evaluations. The DL model is approximated by by a Gaussian Processes (GPs): $y = GP(\mu, \Sigma)$ where $\mu(x)$ and $\Sigma(x, x')$ are the mean and covariance functions, respectively. The Matérn kernel is chosen as covariance function. The Expected Improvement (EI) has been used as acquisition function: $EI(x) = E[\max(f(x) - f(x'), 0)]$ where $f(x)$ and $f(x')$ are, respectively, the predicted mean of the objective function at point x, based on the GP, and the best value observed so far in the optimization process. EI prefers models for which $\mu > f(x')$ (exploitation) or where the standard deviation $\sigma(x)$ is high (exploration). To speed-up convergence, we employ parallelisation on multiple GPUs. DF constructs architectures in parallel by drawing, at each iteration, n samples from the architectural parameter space (where n is the number of available GPUs). The 2 points maximising μ and $\sigma(x)$ are always chosen, the remaining $n - 2$ are randomly drawn.

3. – ALMASim simulations

ALMASim simulator is built upon the CASA Simulator [9], and allows the creation of realistic ALMA mock observations of high and low redshift point-like, extended and diffuse sources. The pipeline works by first generating the Sky Model and scaling it to the desired brightness. The Sky Model is fed to the simobserve task of CASA. The task

TABLE I. – *Mean Residual Scores (MRS) obtained on the 1000 Test Set cubes.*

	DF	CLEAN	U-Net	ResNet-50	DenseNet
mRS	0.0002	0.047	0.012	0.0074	0.0059

simulates an ALMA observation of the Sky Model returning the Measurement Set (MS) (whose Fourier inversion is the Observed Sky Model). Before inverting it, to simulate atmospheric and gain induced errors, we modify the MS adding random Phase and Amplitude errors components. The resulting MS is then Fourier inverted to get the Observed model and scaled to the Sky Model total brightness. ALMASim was used to generate 10000 pairs of cubes containing always a bright central point-like source (central Calibrator) and between 1 and 5 serendipitous sources. Simulations were performed using Cycle 9 C-3 antenna configuration, band 6.

4. – Results and conclusions

We divided the 10000 cubes in Train, Validation and Test sets (80%, 10%, 10%), and let DF search for the best performing architecture. To measure the reconstruction quality we employed the Mean Residual Score (MRS), while to train the architectures generated by DF, we employed a weighted combination of L_1 norm and mean Structural Similarity Index (mSSI). CLEAN is run for 10000 iterations, while hyperparameter tuning is performed on all DL models. DF is trained on 4 Tesla V100 GPUs. Table I shows the MRSs obtained by DF's best architecture, CLEAN, U-Net, ResNet-50 and DenseNet, while fig. 1 shows a direct comparison between the Observed Sky, the true Sky Model, CLEAN prediction of the Sky model, and that of DF. As can be seen, both CLEAN and DF can distinguish between noise components, PSF's side lobes, and the true signal, but DF is the only one also recovering the true source morphologies. This reflects in the lower MRS. The computational time measurements are performed running DF inference on a single V100 GPU, while CLEAN on two Intel Xeon E5-2680. CLEAN takes, on average, 364 minutes to process a data cube, while DF takes only 1.3 minutes (a speed up in processing times by a factor 280 on average).

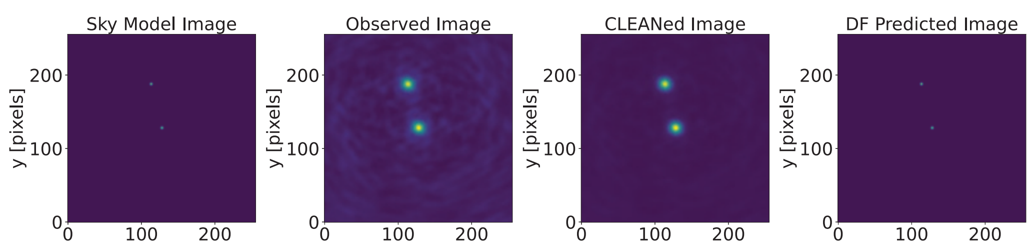

Fig. 1. – From left to right: Sky Model Image, Observed Image, CLEAN image, DF Prediction of the Sky Model.

REFERENCES

[1] Högbom J. A., *Astron. Astrophys. Suppl.*, **15** (1974) 417.

[2] Ossenkopf-Okada Volker *et al.*, *Physics and chemistry of star formation: The dynamical ISM across time and spatial scales*, in *Proceedings of the 7th Chile-Cologne-Bonn-Symposium* (University of Cologne Librari) 2023, https://doi.org/10.18716/KUPS/64624, https://kups.ub.uni-koeln.de/64624/1/final.pdf.

[3] Lukic Vesna, de Gasperin Francesco and Brüggen Marcus, *Galaxies*, **8** (2019) 3.

[4] Connor Liam *et al.*, *Mon. Not. R. Astron. Soc.*, **514** (2022) 2614.

[5] Schmidt K. *et al.*, *Astron. Astrophys.*, **664** (2022) A134.

[6] Rezaei S. *et al.*, *Mon. Not. R. Astron. Soc.*, **510** (2021) 5891.

[7] Delli Veneri Michele *et al.*, *Mon. Not. R. Astron. Soc.*, **518** (2022) 3407.

[8] Delli Veneri Michele *et al.*, *ALMASim: An ALMA Simulation Package to enable comparative testing of novel deconvolution and de-noising models* (2024), https://doi.org/10.5281/zenodo.12763887.

[9] The CASA Team, *Pub. Astron. Soc. Pac.*, **134** (2022) 537.