

Residual Learning for Astronomical Images: A PSF-Blind Approach to Deconvolution

Ali SaraerToosi, David Bromley, and Nhan Luong

Abstract—Reconstructing high-fidelity astronomical images from radio interferometric data is a critical challenge due to sparse UV plane sampling, noise, and calibration errors. Traditional methods such as CLEAN and recent deep learning approaches like POLISH rely on Point Spread Function (PSF) priors, which limits their adaptability to real-world scenarios where PSFs may be imprecisely calibrated or highly variable. In this paper, we propose a novel modification to the POLISH pipeline that leverages residual learning to address these challenges. By reframing the task as the prediction of residuals between the dirty and true sky images, the proposed approach effectively decouples reconstruction from PSF priors, simplifying the learning problem and enabling robust deconvolution under noisy and undersampled conditions. We evaluate our method, against the original POLISH model on simulated datasets with and without PSF priors. Experimental results demonstrate that our model outperforms its counterpart in scenarios without PSF priors while maintaining comparable performance when PSFs are available. Qualitative analysis reveals that the residual-based approach mitigates visual artifacts, and improves the reconstruction of faint sources in noisy backgrounds. Our findings suggest that PSF-blind residual learning is a promising direction for real-time interferometric imaging, especially in the context of next-generation instruments like the DSA-2000. Code is available at: <https://github.com/lennemo09/strong-lensing-polish-torch>

Index Terms—Astronomy, Interferometry, Deconvolution, Denoising, Deep Learning, Residual Learning

1 INTRODUCTION

Radio interferometry has revolutionized our ability to observe the universe, offering high-resolution imaging of celestial phenomena by synthesizing signals from distributed arrays of antennas. This technique is a cornerstone of modern astronomy, enabling the study of faint and distant objects, as well as transient events across the electromagnetic spectrum. However, the imaging process presents unique challenges due to the inherent sparsity of UV plane sampling, noise, calibration errors, and the computational demands of processing large datasets. These limitations are becoming more pronounced with next-generation interferometers such as the DSA-2000, which operates at an unprecedented scale, producing 10 TB of data per second. The need for real-time, high-fidelity image reconstruction has thus emerged as a critical bottleneck. Furthermore, the wide dynamic range of astronomical images, as shown in Fig. 1, poses additional difficulties, as faint sources must be distinguished from bright ones without loss of fidelity.

Traditional algorithms, such as CLEAN [1], [2], iteratively deconvolve the instrumental Point Spread Function (PSF) to reconstruct sky images. While effective, these methods are computationally intensive and sensitive to PSF inaccuracies, especially in wideband and dynamic observational regimes. Furthermore, CLEAN and its derivatives often fail to fully recover faint sources embedded in noise or artifacts caused by incomplete UV coverage. Recent advances in machine learning have introduced deep learning-based solutions, such as POLISH, which leverage convolutional neural networks (CNNs) to achieve fast, feed-forward image reconstruction. By incorporating the PSF as a prior, POLISH sig-

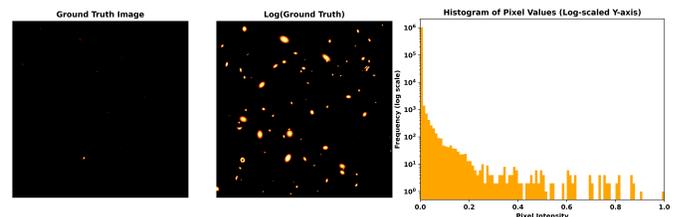


Fig. 1: An example of the pixel distribution of a simulated true sky image without noise. The pixel distribution presents a problem of very high dynamic range, where both very faint sources and very bright sources can be present in the image.

nificantly reduces computational costs and offers improved angular resolution. However, its reliance on pre-calibrated PSFs limits its robustness to real-world conditions, where the PSF may be distorted or poorly characterized.

To address these limitations, we propose a novel modification to the POLISH pipeline [3] by incorporating residual learning. Unlike direct sky reconstruction methods, residual learning shifts the focus to predicting the residual—the difference between the dirty and true sky images. This reframing simplifies the learning task and reduces dependence on explicit PSF priors. The proposed method inherently accounts for PSF variations and noise effects without requiring detailed PSF calibration, making it robust to instrumental errors and environmental disturbances.

This work introduces our modification to POLISH, detailing its theoretical foundation, implementation, and performance evaluation. We present a comprehensive analysis of its performance compared to the original POLISH model, using synthetic datasets with and without PSF priors. Ex-

• A. SaraerToosi, D. Bromley, and N. Luong are with the Toronto Computational Imaging Group, Department of Computer Science, University of Toronto, Canada.

perimental results demonstrate that our model not only matches but can also exceed the performance of the original model, especially in PSF-blind scenarios.

2 BACKGROUND

2.1 Radio Interferometry

Radio interferometry combines signals from arrays of antennas to observe celestial objects at radio wavelengths, achieving resolutions equivalent to a telescope as large as the array's maximum baseline. This involves measuring spatial frequency data, known as visibilities ($V(u, v)$), which represent the Fourier transform of the sky brightness distribution $I(l, m)$:

$$V(u, v) = \iint I(l, m) e^{-2\pi i(ul+vm)} dl dm. \quad (1)$$

The goal is to reconstruct $I(l, m)$ by inverting the Fourier measurements, but challenges arise due to sparse sampling of the UV plane (determined by the baselines between antennas), noise, calibration errors, and the computational demands of processing massive data volumes. An illustration of this relationship between the Fourier and image domain can be found in the Supplementary section.

Modern interferometric pipelines also leverage distributed computing and GPUs to handle the computational demands of calibration and imaging. Recent innovations, such as neural interferometry, explore the direct reconstruction of visibility data in the Fourier domain using machine learning, bypassing the need for explicit PSF deconvolution. These developments highlight the growing role of data-driven techniques in radio astronomy, promising greater scalability and accuracy for next-generation instruments.

One such instrument system, DSA-2000, with 2000 fixed antennas spread over an area of 15 km in diameter, addresses these challenges with its dense baseline coverage, offering a spatial resolution of 3.5 arcseconds and a wide field of view. However, gaps in UV coverage still introduce artifacts in images, and wideband observations complicate reconstruction due to frequency-dependent effects. Advanced algorithms, such as CLEAN and neural fields, are used to mitigate these issues by interpolating the missing UV data and improving image fidelity. Additionally, the computational demands of processing large datasets require distributed computing, GPUs, and machine learning for real-time calibration and imaging. These innovations will enable DSA-2000 to produce high-fidelity images of galaxies, transient events, and black hole environments.

2.2 Synthetic Sky Model and Data Generation

The image model and generation procedure is established by Connor et al. for POLISH [3], which produces realistic input-output pairs comprising high-resolution *true sky* images and degraded, low-resolution *dirty* images that emulate the measurements of an interferometric array. We also explored generating an equivalent data set in the Fourier domain by artificially generating our own UV baselines and calculating the PSF, however due to limited publicly available information we were unable to perfectly reproduce the DSA-2000 full band PSF used by POLISH, thus we opted

to include this process in the supplementary material. In our work, the resolution disparity is not taken into account, thus the image pair have the same resolution. We also introduced the residual for training, which we argue simplifies the learning problem which help improve the signal reconstruction. This section details the key models, intuition behind the data modelling, and the implementation specifics used to generate a dataset of 800 training triplets, 100 validation triplets, and 200 test triplets.

2.2.1 Synthetic True Sky Images

The true images represent the intrinsic sky brightness distribution, consisting of synthetic galaxy sources whose properties are modeled from observational statistics. These images serve as the ground truth for the model. For our work, the true sky images generated are 1024×1024 pixels in size.

Spatial and Flux Distribution

The number of sources per image is sampled from a Poisson distribution, with the expected mean calculated as:

$$N_{src} = \rho_{src} \times \frac{A_{image}}{3600^2},$$

where ρ_{src} is the source density in sources per square degree, and A_{image} is the image area in square arcminutes.

Flux densities S of individual sources are drawn from a broken power-law distribution:

$$P(S) \propto \begin{cases} S^{-\frac{2}{3}}, & \text{if } S > S_{\text{threshold}} \\ S^{-1}, & \text{if } S \leq S_{\text{threshold}} \end{cases}$$

where $S_{\text{threshold}}$ defines the flux value at which the power-law behavior transitions between the two regimes. This distribution accounts for the abundance of faint sources while modeling the sparser distribution of bright sources.

Source Shapes

Galaxy shapes are modeled as two-dimensional Gaussian ellipsoids, with their parameters derived from empirical distributions:

Semi-major axis (σ_x): The semi-major axis is sampled from a gamma distribution:

$$\sigma_x \sim \Gamma(k, \theta),$$

where k is the shape parameter and θ is the scale parameter of the gamma distribution¹.

Ellipticity (e): The ellipticity, defined as

$$e = \frac{a - b}{a + b},$$

is drawn from a beta distribution. Here, a and b represent the semi-major and semi-minor axes of the ellipse, respectively.

Semi-minor axis (σ_y): The semi-minor axis is derived from the semi-major axis and ellipticity using:

$$\sigma_y = \sigma_x \sqrt{\frac{1 - e}{1 + e}}.$$

Orientation angle (θ): The orientation angle of the ellipse is uniformly distributed between -90° and 90° .

1. https://en.wikipedia.org/wiki/Gamma_distribution

Gaussian brightness profile: The brightness distribution of a galaxy is modeled as:

$$G(x, y) = A \exp\left(-\frac{1}{2} \begin{bmatrix} x' & y' \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x' \\ y' \end{bmatrix}\right),$$

where $\begin{bmatrix} x' \\ y' \end{bmatrix}$ are coordinates rotated by the angle θ , and Σ is the covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}.$$

The parameter ρ represents the correlation coefficient, which is set to zero in these simulations, assuming no correlation between axes.

Gravitational lensing

A subset of galaxies is modeled as gravitationally lensed sources to simulate the distortions caused by massive foreground objects. A more thorough description of the lensing model is present in the Supplementary section.

The addition of gravitational lensing provides in the training data aims to provide a realistic representation of gravitationally distorted sources, which are essential for the neural network to be able to deconvolve in practice with real sky images.

2.2.2 Synthetic Dirty Images

In principle, the true sky images are degraded by the instrumental response of the radio interferometer to generate dirty images. This degradation process includes convolution with the PSF and subsequent downsampling to match the resolution of the synthesized beam.

Point Spread Function (PSF): The PSF is derived from the Fourier transform of the telescope’s sampling function $S(u, v)$, where (u, v) are spatial frequency coordinates sampled by the array:

$$\text{PSF}(x, y) = \mathcal{F}^{-1}\{S(u, v)\}.$$

In POLISH, variations in the PSF are introduced to account for errors in instrumental calibration and measurement:

- **Ideal PSF:** Derived directly from the array layout and observing parameters.
- **Distorted PSF:** Elastic transformations are applied to the PSF to simulate calibration errors, ionospheric effects, or asymmetries due to sparse sampling.

For our purpose, we only account for the ideal PSF from the DSA-2000 using fullband 1.3GHz, which is readily provided with the POLISH code². Distortion to the PSF is planned for further work.

Noise: Gaussian noise is added to the true sky image to simulate thermal noise and other measurement errors:

$$I_{\text{noisy}}(x, y) = I_{\text{sky}}(x, y) + \eta(x, y),$$

where $\eta(x, y) \sim \mathcal{N}(0, \sigma^2)$.

The noise level σ is set based on the expected signal-to-noise ratio (SNR) of the instrument. In POLISH and in our work, $\sigma = 5$ is chosen for the noise.

2. <https://github.com/liamconnor/polish-torch/blob/main/psf/dsa-2000-fullband-psf.fits>

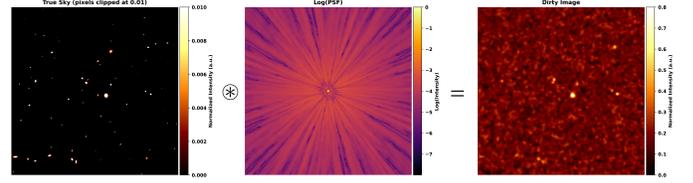


Fig. 2: An example of a dirty image generated using the true sky convolved with the provided fullband DSA-2000 PSF. The true sky image and dirty image pixel values are clipped and the PSF is displayed in log scale for visualization only.

Degradation: Each noisy sky image $I_{\text{noisy}}(x, y)$ is convolved with the PSF $k(x, y)$:

$$I_{\text{dirty}}(x, y) = I_{\text{noisy}}(x, y) \otimes k(x, y), \quad (2)$$

where \otimes denotes the convolution operation.

In practice, the convolution is performed in the Fourier domain:

$$\mathcal{F}(I_{\text{dirty}}) = \mathcal{F}(I_{\text{noisy}}) \cdot \mathcal{F}(k).$$

Downsampling: In POLISH, downsampling is applied to the convolved image to form a low-resolution dirty image. After convolution, the dirty image is downsampled by a factor r to match the effective resolution of the synthesized beam:

$$I_{\text{downsampled}}(x, y) = I_{\text{dirty}}(r \cdot x, r \cdot y).$$

Bilinear interpolation is used to avoid aliasing during the downsampling process.

In our work, we are focused on reconstructing the residual which contains the convolved noise, thus upsampling the network output is not a desirable operation. Hence, we opted to not utilize the upsampling functionality of the POLISH network and select $r = 1$ for dirty images generation. An example of the dirty image generation is shown in Fig. 2.

2.2.3 Residual Images

The residual image captures the difference between the true sky and the dirty image, providing a measure of the artifacts and distortions introduced by the instrument’s PSF and noise. The residual is computed as part of the data generation process and serves as an alternative to compute the true sky reconstruction. It is important to note that the residual is computed from the normalized true-dirty pair, the pixel range of the residual image is $[-1, 1]$, which is not normalized during training and validation.

Definition of the Residual: The residual image $R(x, y)$ is defined as the pixel-wise difference between the PSF-convolved true sky image (dirty image) $I_{\text{dirty}}(x, y)$ and the original high-resolution sky image $I_{\text{sky}}(x, y)$:

$$R(x, y) = I_{\text{dirty}}(x, y) - I_{\text{sky}}(x, y). \quad (3)$$

3 RELATED WORK

Connor et al. introduced POLISH [3], a deep learning-based method for deconvolution and super-resolution of DSA-2000 measurements. POLISH utilizes the WDSR architecture

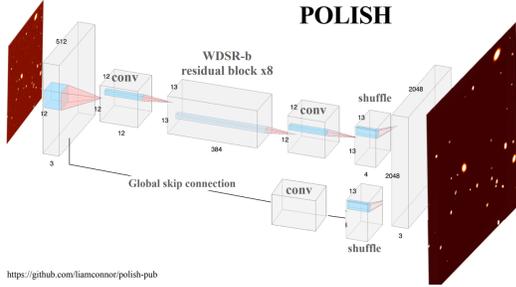


Fig. 3: The POLISH architecture proposed by Connor et al. based on the WDSR architecture by Fan et al. [4].

[4] to reconstruct high-resolution sky images from low-quality dirty images. By leveraging a pre-calibrated PSF as a physics-informed prior, it achieves high angular resolution and effective super-resolution, distinguishing sources below the scale of the PSF’s corruption. However, this dependence on pre-calibrated PSFs limits its adaptability to real-world scenarios where the PSF may be imperfectly known or exhibit significant variations.

Denoising CNNs (DnCNN) have also demonstrated strong performance in residual learning for noise reduction [5], [6]. These methods operate by learning a mapping from noisy images to their noise components, effectively separating noise from the underlying signal. While successful in other domains such as medical imaging and microscopy, their potential for interferometric image reconstruction—particularly in deconvolution tasks with unknown PSFs—has not been fully explored. Zhang et al. [5] highlight the flexibility of residual learning in general denoising tasks but do not address its application to interferometric image recovery.

Building on these foundations, our work introduces residual learning to interferometric image reconstruction. While POLISH directly reconstructs the sky image utilizing a known PSF prior, or a distribution of PSFs as a prior, we reformulate the task as the prediction of residuals between the dirty image and the true sky. This reframing simplifies the learning process, reduces reliance on explicit PSF priors, and improves robustness to calibration errors.

By incorporating the strengths of both POLISH and DnCNN, our approach maintains the speed of feed-forward methods while addressing the challenges posed by unknown or inaccurate PSFs. This combination allows for accurate reconstructions in scenarios where PSF priors are unavailable or unreliable, making our method better suited for practical, real-world interferometric data.

4 PROPOSED METHOD

The proposed method introduces a modification to the POLISH pipeline by transitioning from direct sky reconstruction to residual learning which mitigates the need for a PSF prior as an input. Here, we describe the intuition, mathematical framework, network architecture, and implementation details of the proposed method in detail.

In essence, the residual image quantifies the extent to which the PSF and noise degrade the original signal, highlighting errors and artifacts introduced during observation.

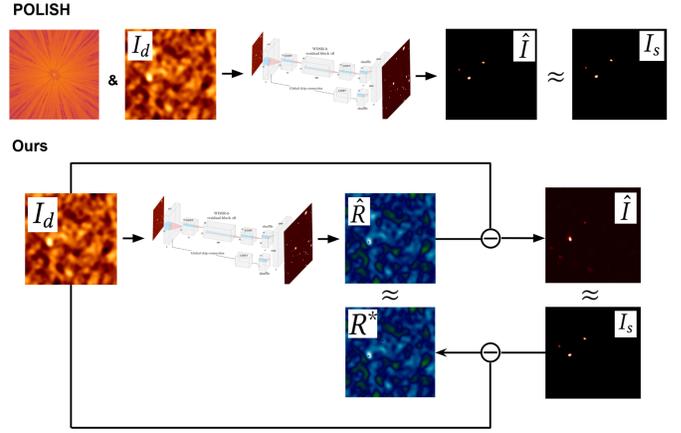


Fig. 4: Top: Standard POLISH procedure: Given as input a dirty image and PSF(s), the network outputs an deconvolved true sky image. Bottom: Our proposed alteration to the POLISH training procedure: the network learns to map the dirty image I_d to a predicted residual \hat{R} which should approximate the ground truth residual R^* which can be used to reconstruct the approximated sky image \hat{I} .

Based on the definition of the dirty image from Eq. 2, we can rearrange the terms to get:

$$R(x, y) = I_{\text{sky}}(x, y) \otimes k(x, y) - I_{\text{sky}}(x, y) + \eta(x, y) \otimes k(x, y). \quad (4)$$

This shows that by making the network learn to reconstruct the residual, the model also learns a deep representation of the true sky. Thus, we hypothesize that this learning process is a simpler task than producing the true sky at the last layer, which can help reduce the reliance on a given PSF prior. Our modification to POLISH is illustrated in Fig. 4.

Intuitively, this is because the distance between the residual and the dirty image is much closer than the distance between the true sky image and the dirty image, as the convolved noise along with the sampling noise from the PSF make up the majority of the pixel differences between the true sky and the dirty image. Fig. 5 shows that denoising task is simple for the model to learn as it only need to construct a near-identical version the dirty image. The harder task is to recover the galaxies, which is equivalent to learning to place the *divots* (circled in red) accurately in the constructed noise background.

A natural derivation of this can also be seen from the mean squared error (MSE) loss between the true sky I_s and the reconstructed sky:

$$\mathcal{L}_I = \frac{1}{N} \sum_i \left(I_s^{(i)} - \hat{I}^{(i)} \right)^2$$

This is equivalent to the MSE loss between the reconstructed residual and the ground truth residual if we use

TABLE 1: Performance comparison of POLISH and POLISH_Residual models with and without PSF. The numbers reported are the mean \pm standard deviation over the test set of 200 samples.

Metric	POLISH (PSF)	POLISH_Residual (PSF)	POLISH (No PSF)	POLISH_Residual (No PSF)
SSIM	0.31661 \pm 0.26165	0.42376 \pm 0.24384	0.26295 \pm 0.21049	0.44731 \pm 0.25399
MSE	0.00090 \pm 0.00137	0.00034 \pm 0.00051	0.00155 \pm 0.00332	0.00030 \pm 0.00050
PSNR	34.87310 \pm 6.70634	37.92690 \pm 5.39336	33.24559 \pm 6.52640	38.56120 \pm 5.57084

$$\hat{I} = I_d - \hat{R}:$$

$$\begin{aligned} \mathcal{L}_R &= \frac{1}{N} \sum_i \left(R^{*(i)} - \hat{R}^{(i)} \right)^2 \\ &= \frac{1}{N} \sum_i \left(I_s^{(i)} - (I_d^{(i)} - \hat{R}^{(i)}) \right)^2 = \mathcal{L}_{R'} \end{aligned}$$

Then, we have the new loss functions that can be used to direct the network towards learning to compute the residual image from the dirty image.

In practice, while both loss functions would guide the model to learn the residuals, using \mathcal{L}_R as the loss function only requires loading the residual and dirty images onto memory, while using \mathcal{L}_I as the loss function requires loading the residual, dirty, and the true sky images onto memory which requires the use of a smaller batch size for training.

Another key difference is while POLISH originally found MAE loss to give faster convergence during training, our experiments showed that using MSE loss leads to more stable training faster convergence with our setup for both methods using our setup and training configurations, thus we decided to use MSE loss for all of our training runs.

5 EXPERIMENTAL RESULTS

5.1 Setup and Implementation Details

Our implementation uses a fork of the `polish-torch` repo³. The resulting dataset used for training consists of true-dirty-residual images triplets, where the training set contains 800 triplets, the validation set contains 100 triplets, and the test set contains 200 triplets.

For experiments, we trained the POLISH model with and without the given PSF prior for both residual reconstruction task and sky reconstruction task. For comparison purposes, we used $r = 1$ for no super-resolution in the upsampling layer. The input images are 1024×1024 pixels with a `RandomCrop` size of 512×512 to avoid overfitting. For the optimization algorithm, we used Adam which is a standard choice also used by POLISH, but we changed the learning rate scheduler to `ReduceLROnPlateau` which shown benefit to later epochs.

All experiments were conducted on an NVIDIA RTX TITAN GPU with 24 GB VRAM and trained for 160 epochs. While we would have preferred to extend the training period, time constraints prevented us from doing so. Nevertheless, we observed that the models reached a plateau after 80 epochs, suggesting that additional training would likely result in only marginal gains.

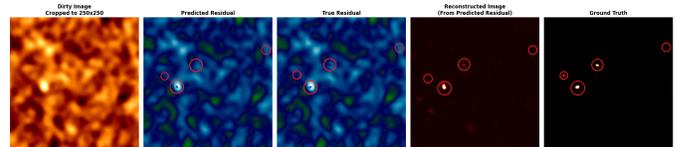


Fig. 5: An example of a reconstructed true sky using the predicted residual. Highlighted are some of the galaxies of interest, most which the model manages to recover. Note that the model also produces very faint artifacts from small differences between the residuals, which may be counted as false positive galaxies. This occasionally occurs in regions with signals very close to, or under the noise floor.

5.2 Results and Discussion

Comparison with Standard POLISH

Table 1 shows the quantitative performance of the models on 3 metrics also used by POLISH: Structural Similarity Index Measure (SSIM), Mean Squared Error (MSE), and Peak Signal-to-Noise Ratio (PSNR). Both methods are evaluated with and without the PSF as an additional input channel, providing insight into the role of residual learning and the necessity of PSF priors.

The results shows that vanilla POLISH performs significantly better given a PSF prior across all metrics. When using residual learning, the performance is slightly better without a given PSF prior with a small margin. The distribution of the metrics results are provided in the Supplementary section.

In addition to quantitative metrics, the image outputs from the models provide insights into the deconvolution artifacts and differences in reconstruction quality. For a more detailed comparison, we examined the output images from all models under both PSF configurations. Representative samples are illustrated in Fig. 6.

A recurring artifact, resembling an airy disk, is observed in the reconstructions from models incorporating PSF priors. This consistent artifact appears as a faint concentric circular structure centered within the image field. The feature is more pronounced in the standard POLISH model and appears as a subtle "darkened" region in the residual-based POLISH_Residual model. These artifacts are less noticeable when the PSF prior is removed (Fig. 6.b and Fig. 6.c).

Due to the consistent appearance and position, we hypothesize that this artifact potentially arises as a side effect of how the PSF is modeled and applied in the WDSR block, and from the global skip connection of the POLISH architecture. While faint, this artifact could interfere with the detection of low-flux sources located at the center of the field, especially when interpreting images with a faint

3. <https://github.com/liamconnor/polish-torch/>

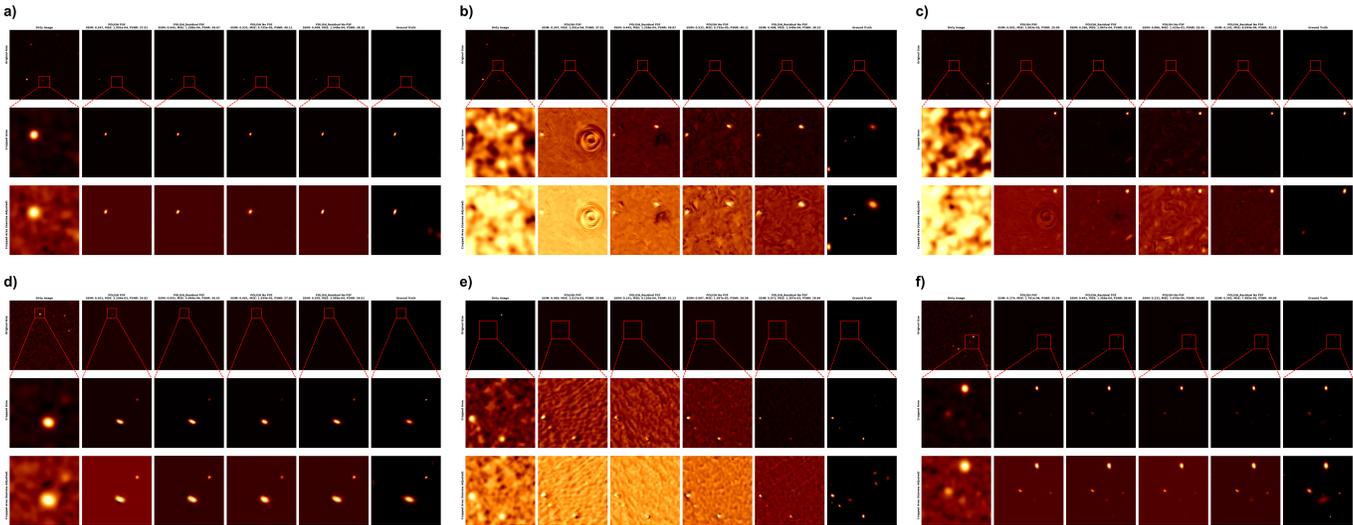


Fig. 6: Comparison of image reconstruction quality from the models tested. For each subplot: First row: the entire image plane; Second row: the cropped region highlighted and the color scale corresponds to the normalized cropped region; Third row: the cropped region with gamma correction $\gamma = 2.2$ for visualization. For each subplot, the columns from left to right are: Dirty image, POLISH PSF, POLISH_Residual PSF, POLISH No PSF, POLISH_Residual No PSF, Ground Truth.

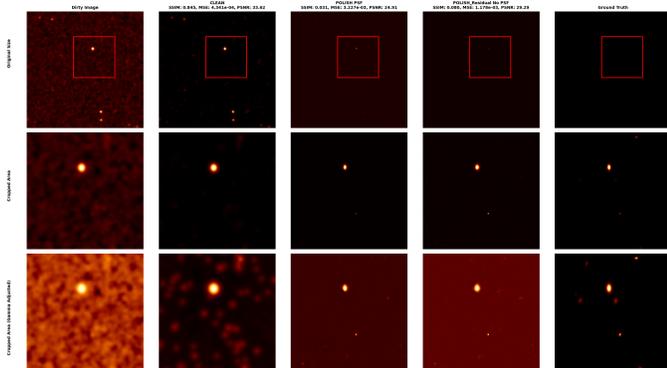


Fig. 7: Comparison of CLEAN versus learned methods. First row: The entire image plane. Second row: the cropped region highlighted and the color scale corresponds to the normalized cropped region. Third row: cropped region with gamma correction $\gamma = 2.2$ for visualization. The columns from left to right are: Dirty image, CLEAN, POLISH PSF, POLISH_Residual No PSF, Ground Truth.

galaxy near the center (Fig. 6.b). A closer investigation to understand the cause of this artifact is desirable.

Across the samples, models trained with a PSF prior exhibit similar qualitative outputs. However, POLISH_Residual demonstrates a more robust handling of noisy regions, avoiding over-amplification of residual noise. Interestingly, in configurations without the PSF prior, POLISH_Residual produces visually cleaner results compared to POLISH_Residual with the PSF. This is evident in Fig. 6.e, where faint sources and structural details are recovered with minimal interference from background noise.

The absence of PSF priors in the residual-based model leads to outputs that closely align with those from PSF-enabled models but without the central artifact. This suggests that the residual-based approach is able to implicitly

captures essential image corrections without requiring explicit knowledge of the PSF due to the simpler learning task.

Additionally, we would also want to use the domain-standard procedure of detecting galaxies in an image using SNR threshold of 5σ in a given subregion of the image. However, this process is non-trivial when the galaxy distribution is non-uniformly spaced apart, and the size of each galaxy can vary greatly. This causes a challenge in finding the correct window size for detection with SNR threshold. We proposed a procedure to adapt this procedure to our dataset in the Supplementary section.

Comparison with CLEAN

CLEAN is an useful baseline to compare against as it is the domain-standard method for interferometry image reconstruction. POLISH have been shown to outperform CLEAN in deconvolution tasks [3]. Here, we want to briefly compare against CLEAN performance, especially in terms of speed as well. While there is no off-the-shelf implementation for CLEAN, the algorithm is straightforward enough for us to re-implement. The details of which are discussed in the Supplementary section.

Table 2 shows that quantitatively, CLEAN is significantly better at reconstruction the true sky image. However, this is due to the specific caveat of these metrics for our data. Because the true sky image is noise-free, with a lot of very faint signals very close to zero, and most pixels are zero, the metrics are biased towards images with low noise background. Fig. 7 shows that even though the quantitative metrics favor CLEAN’s output, we can observe that the reconstructed image from CLEAN is not desirable as it struggles with removing the noise entirely, even though the average noise background is lower. More comparison examples are included in the Supplementary section.

Regarding the time taken to process an image, CLEAN takes on average 2.77 seconds while the feed-forward meth-

TABLE 2: Performance of CLEAN algorithm on the test set with maximum 1000 iterations for each sample.

Metric	CLEAN
SSIM	0.94339 ± 0.06141
MSE	0.00022 ± 0.00016
PSNR	37.30026 ± 2.36678
Time (s)	2.77214 ± 0.28830

ods take on average 0.006 seconds. This results in a typical speed-up of 500 times on the same system configuration.

An additional downside of CLEAN is the need to manually adjust the parameters of each image to achieve the best results. A specific threshold value for one image is not suitable for another. To apply CLEAN to a batch of images, where some contain only near-zero signals, an unsuitable threshold value would result in the algorithm stopping early and returning a faulty output image. But since many of the galaxies are near noise level, automatically determining a good threshold value is not straightforward.

6 CONCLUSION

We introduced a PSF-blind residual learning framework for reconstructing astronomical images from interferometric data. By reframing the task from direct image reconstruction to residual prediction, we demonstrated improved robustness to calibration errors and reduced dependence on explicit PSF priors. Our approach simplifies the complex task of deconvolution into a more manageable residual learning problem, enabling faster convergence during training and improved reconstruction fidelity in challenging conditions.

Experimental results confirm that our residual-based method outperforms traditional POLISH, particularly in scenarios without a known PSF. Moreover, we showed that our framework generalizes to varying levels of noise and image complexity as well as standard POLISH, producing high-fidelity reconstructions with fewer artifacts.

From a practical standpoint, the lack of reliance on a PSF prior and the absence of the central artifact demonstrates the ability to perform better deconvolution with the POLISH_Residual model. The model’s ability to decompose the residual signal without direct dependence on the PSF demonstrates the viability of our proposed approach for real-world noisy and undersampled interferometric data.

Future work will explore decoupling the deblurring and denoising tasks into separate networks⁴, potentially introducing a third stage network for super-resolution to match the capability presented by POLISH. Models such as [6] present a potential avenue for an improved architecture capable facilitating this modularity. This modular approach could further enhance the robustness and accuracy of the reconstruction pipeline, paving the way for scalable, real-time processing in next-generation instruments like DSA-2000.

ACKNOWLEDGMENTS

The authors would like to thank Liam Connor for providing the code and instruction for POLISH, and general guidance,

and Aviad Levis for their guidance and feedback. We would also like to acknowledge both for the inspiration and motivation for the project.

REFERENCES

- [1] J. A. Hogbom, “Aperture synthesis with a non-regular distribution of interferometer baselines,” *Astronomy and Astrophysics Supplement Series*, vol. 15, p. 417, Jun. 1974.
- [2] A. R. Thompson, J. M. Moran, and J. Swenson, George W., *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*. Springer Cham, 2017.
- [3] L. Connor, K. L. Bouman, V. Ravi, and G. Hallinan, “Deep radio-interferometric imaging with polish: Dsa-2000 and weak lensing,” *Monthly Notices of the Royal Astronomical Society*, vol. 514, no. 2, p. 2614–2626, May 2022. [Online]. Available: <http://dx.doi.org/10.1093/mnras/stac1329>
- [4] Y. Fan, J. Yu, and T. S. Huang, “Wide-activated deep residual networks based restoration for bpg-compressed images,” in *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018*. Computer Vision Foundation / IEEE Computer Society, 2018, pp. 2621–2624. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018_workshops/w50/html/Fan_Wide-activated_Deep_Residual_CVPR_2018_paper.html
- [5] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [6] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual Dense Network for Image Restoration,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 43, no. 07, pp. 2480–2495, Jul. 2021. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/TPAMI.2020.2968521>

4. Motivated by Eq. 4.

SUPPLEMENTARY

Illustration of Radio Interferometry Image-UV Relationship

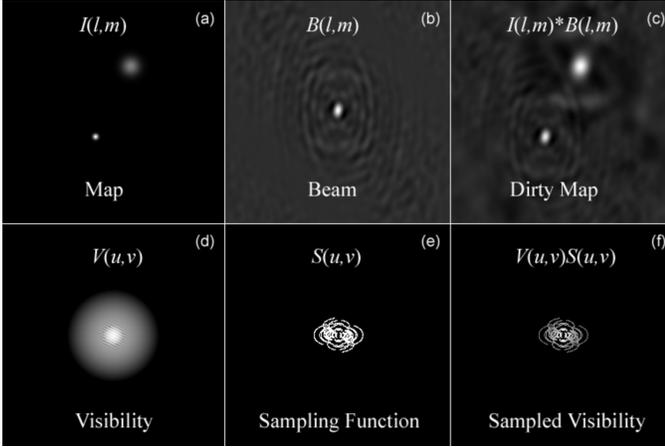


Fig. 8: Figure from Radio Astronomy: Lecture 6 by Prof. Dale E. Gary (Link). a) An example (model) sky map. d) The corresponding visibilities (Fourier Transform of the map). c) The synthesized beam, or point-spread-function, of a model antenna array. e) The sampling function of the array, whose Fourier Transform gives the beam in (b). f) The product of panels (d) and (e), representing the sampled visibilities. These are the actual measurements from the array. c) The dirty map that results from the Fourier Transform of the sampled visibilities. This is the same as the convolution of the map in (a) and the synthesized beam in (b).

PSF Generation From Simulated UV Baselines

Here, we describe the methodology for generating synthetic PSFs from simulated UV baselines, which had been omitted from the main writing due to missing key information to perfectly accurately recreate the DSA-2000 UV mask and PSF.

UV Baseline Simulation

The UV plane represents the spatial frequency coverage of an interferometric array. The baselines between pairs of antennas define the sampled spatial frequencies. For an array with N antennas, the baselines \mathbf{b}_{ij} are calculated as:

$$\mathbf{b}_{ij} = \mathbf{r}_j - \mathbf{r}_i = (x_j - x_i, y_j - y_i),$$

where $\mathbf{r}_i = (x_i, y_i)$ is the position of the i -th antenna. The corresponding UV coordinates, scaled by the observing wavelength λ , are:

$$(u, v) = \frac{\mathbf{b}_{ij}}{\lambda}.$$

UV Mask Generation

The UV coverage is discretized onto a Fourier grid of size $N_x \times N_y$, corresponding to the desired image resolution. A binary UV mask $M(u, v)$ is created by marking grid points sampled by the baselines:

$$M(u, v) = \begin{cases} 1, & \text{if } (u, v) \text{ within the sampled UV coordinates,} \\ 0, & \text{otherwise.} \end{cases}$$

Conjugate symmetry is enforced by adding points $(-u, -v)$ for each (u, v) :

$$M(-u, -v) = M(u, v).$$

PSF Computation

The PSF is derived as the inverse Fourier transform of the UV mask:

$$\text{PSF}(x, y) = \mathcal{F}^{-1}(M(u, v)),$$

where \mathcal{F}^{-1} denotes the 2D inverse Fourier transform. Zero-padding is applied to $M(u, v)$ to increase the spatial resolution of the resulting PSF.

Optimizations

As we needed to compute the sampled FFT of hundreds of images given thousands of baselines, a faster method to achieve this is implemented. Here, we describe the optimization steps that we implemented to generate the dirty images.

Precomputed UV masks: Instead of recalculating the UV coordinates for each image, a binary UV mask $M(u, v)$ is precomputed and stored. Sampling the Fourier transform of an image $F(I)$ can then be performed as:

$$F_{\text{sampled}}(u, v) = F(I)(u, v) \cdot M(u, v),$$

where $M(u, v)$ is a binary mask. This approach eliminates the need for point-wise mapping of many thousands of UV coordinates onto the Fourier grid for each image.

Binary mask: Using a binary UV mask $M(u, v) \in \{0, 1\}$ is computationally advantageous compared to a weighted mask $W(u, v)$. In the binary case, the non-zero indices of $M(u, v)$ are extracted:

$$\mathcal{I} = \{(u, v) \mid M(u, v) = 1\}.$$

Sampling reduces to indexing the Fourier-transformed image:

$$F_{\text{sampled}}(u, v) = F(I)(u, v), \quad \forall (u, v) \in \mathcal{I}.$$

This avoids element-wise multiplications across the full grid, which would be required for weighted masks.

Vectorized baseline calculations: The baselines \mathbf{b}_{ij} are computed for all antenna pairs using vectorized operations:

$$\mathbf{b} = \text{vec}(\mathbf{r}) \otimes \text{vec}(\mathbf{r}),$$

where $\text{vec}(\mathbf{r})$ represents the antenna positions as a vector. This helps scale efficiently with large arrays.

Precomputed mask indices: For repeated FFT operations, the UV mask indices \mathcal{I} are precomputed once. Applying the mask becomes:

$$F_{\text{masked}}(u, v) = \begin{cases} F(I)(u, v), & (u, v) \in \mathcal{I}, \\ 0, & \text{otherwise.} \end{cases}$$

Batch processing: Batch processing of images is implemented using parallelization. For B batches, the sampled Fourier transforms are computed concurrently:

$$F_{\text{sampled}}^{(b)}(u, v) = F^{(b)}(I)(u, v) \cdot M(u, v), \quad \forall b \in \{1, \dots, B\}.$$

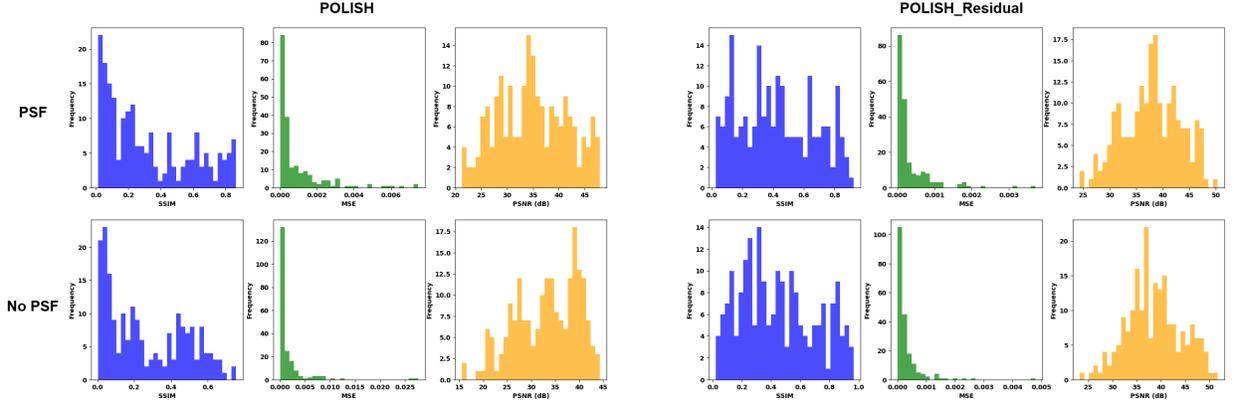


Fig. 9: Distribution of metrics results between methods on the validation set.

Gravitational Lensing

Deflection angles (α_x, α_y): The deflection angles induced by the lens are calculated as:

$$\alpha_x = \frac{\theta_E x}{\sqrt{x^2 + y^2}}, \quad \alpha_y = \frac{\theta_E y}{\sqrt{x^2 + y^2}},$$

where x and y are the coordinates of the source relative to the lens center, and θ_E is the Einstein radius, defined as:

$$\theta_E = \sqrt{\frac{4GM}{c^2} \frac{D_{ls}}{D_l D_s}}.$$

Here, M is the lens mass, G is the gravitational constant, c is the speed of light, and D_l , D_s , and D_{ls} are the angular diameter distances to the lens, the source, and between the lens and source, respectively.

Lensed image coordinates: The lens equation relates the apparent (lensed) position of the source (x, y) to its true (unlensed) position (β_x, β_y) through the deflection angles:

$$\beta_x = x - \alpha_x, \quad \beta_y = y - \alpha_y.$$

Lensed brightness profile: The brightness distribution of the lensed source is mapped from the intrinsic source profile by applying the lens equation to distort the coordinates. For a Gaussian source:

$$G_{\text{lensed}}(x, y) = A \exp\left(-\frac{1}{2} \begin{bmatrix} x' & y' \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x' \\ y' \end{bmatrix}\right),$$

where x' and y' are transformed according to the lens equation.

Einstein ring formation: For symmetric configurations where the source lies directly behind the lens, the lensed image forms a complete Einstein ring with a radius approximately equal to θ_E . In asymmetric cases, partial arcs or multiple lensed images are produced. Einstein radii (θ_E) are drawn from a uniform distribution in a physically reasonable range, e.g., $[0.025, 0.3]$ arcseconds. Source positions (β_x, β_y) relative to the lens center are randomly sampled within the image field.

Domain-Standard Detection Process Using SNR Thresholds in Astronomy

In astronomical imaging, detecting faint galaxies and other celestial sources often relies on identifying regions with

a signal-to-noise ratio (SNR) above a specific threshold, typically set at 5σ . This threshold ensures that detections are statistically significant and minimizes false positives due to noise fluctuations. Below, we outline the steps for implementing this detection process, including determining the appropriate window size for galaxy detection and challenges associated with the method.

Definition of σ for a Patch

The σ value for a given patch is the standard deviation of the noise in that region. To compute σ :

- 1) **Identify noise-only regions:** Choose patches of the image that are unlikely to contain any sources (e.g., background regions away from bright sources).
- 2) **Measure standard deviation:** Compute the standard deviation (σ) of pixel intensities within these regions. This σ represents the background noise level for the image.

For a patch P of size $N \times N$ pixels:

$$\sigma_P = \sqrt{\frac{\sum_{i,j \in P} (I_{i,j} - \bar{I}_P)^2}{N^2}},$$

where $I_{i,j}$ is the intensity of the pixel at (i, j) in the patch, and \bar{I}_P is the mean intensity of the patch.

Calculating SNR for a Patch

The SNR for a patch is the ratio of the signal intensity to the noise standard deviation. For a patch P :

$$\text{SNR}_P = \frac{\text{Signal}_P}{\sigma_P}.$$

- **Signal (Signal_P):** Defined as the sum or maximum of the pixel intensities within the patch, depending on the detection strategy.
- **Noise (σ_P):** Standard deviation of the noise in the patch, as calculated above.

For significant detections, the patch must satisfy:

$$\text{SNR}_P \geq 5.$$

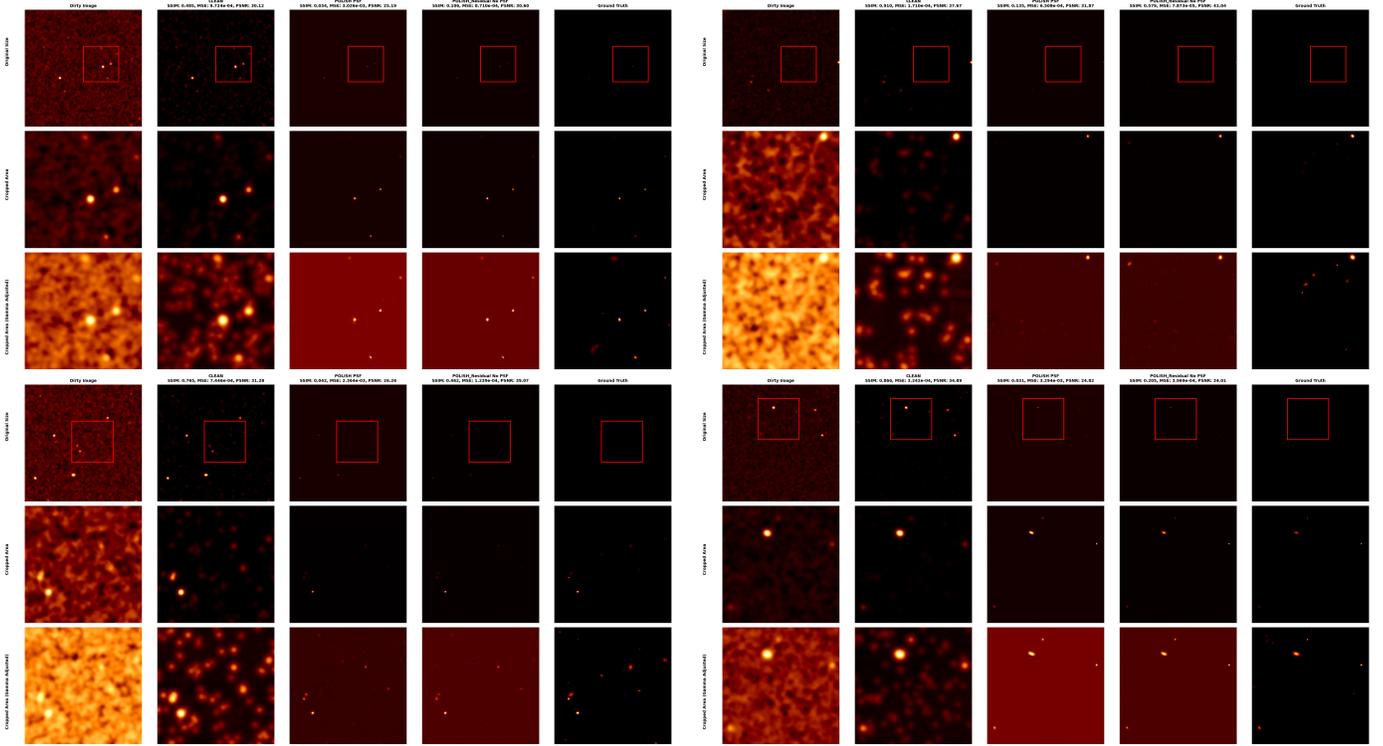


Fig. 10: Comparison of CLEAN versus learned methods. First row: The entire image plane. Second row: the cropped region highlighted and the color scale corresponds to the normalized cropped region. Third row: cropped region with gamma correction $\gamma = 2.2$ for visualization. The columns from left to right are: Dirty image, CLEAN, POLISH PSF, POLISH_Residual No PSF, Ground Truth.

Challenges in Choosing the Patch Size

A critical aspect of this detection process is selecting an appropriate patch (or window) size. If the window size is too small, it may capture only part of a galaxy, reducing the SNR. Conversely, if the window size is too large, it may include noise or multiple sources, complicating detection.

Estimate the number of sources: Use the Poisson distribution to sample the number of sources in the image:

$$n_{\text{src}} \sim \text{Poisson}(\rho_{\text{src}} \times A_{\text{image}}),$$

where:

- ρ_{src} is the source density in sources per square arcminute,
- A_{image} is the image area in square arcminutes.

textbfWindow size:

- Start with a window size that is expected to contain, on average, one galaxy based on the source density and image resolution.
- For a pixel size of PIXEL_SIZE arcminutes and an $N_x \times N_y$ pixel grid:

$$A_{\text{patch}} = N_{\text{patch}}^2 \times \text{PIXEL_SIZE}^2,$$

where N_{patch} is the side length of the patch in pixels.

Noise distribution: Slide the window across the image, compute the standard deviation (σ) for each patch, and examine the histogram of σ values.

For a well-chosen window size, patches containing noise should cluster around a single σ value, while those containing galaxies should form a distinct higher-SNR tail in the histogram.

Optimizing Window Size Through Simulations

Given the source density and the hypothesis that galaxies are randomly distributed.

Calculate expected noise-only patches: The number of patches containing only noise can be estimated by:

$$n_{\text{noise}} = n_{\text{patches}} - n_{\text{src}},$$

where n_{patches} is the total number of patches for the chosen window size.

Analyze distribution of σ : Generate a histogram of the standard deviations (σ) over all patches.

- Patches with only noise will contribute to a peak around the background noise level.
- Patches containing galaxies will form a tail or secondary peak at higher σ values.

By tweaking the window size, the goal is to maximize the separation between noise-only patches and those containing galaxies in the histogram. The optimal window size balances detection sensitivity and computational efficiency.

Once an optimal window size is determined:

- 1) Slide the window across the entire image and calculate the SNR for each patch.

- 2) Identify patches where $\text{SNR} \geq 5$ as potential galaxy detections.
- 3) Validate detections by visually inspecting candidate regions or cross-referencing with known catalog data.

This approach ensures robust galaxy detection while minimizing false positives and balancing computational demands.

Distribution of Metrics Results Across the Test Set

See Fig. 9.

CLEAN Algorithm

The CLEAN algorithm [1], [2] is the standard iterative method for deconvolution in radio interferometry, designed to reconstruct high-fidelity astronomical images from sparse UV plane data. It addresses the artifacts and distortions introduced by the incomplete sampling of the UV plane, modeled through the instrumental PSF.

Steps in CLEAN

In this section we briefly describe the key steps in CLEAN:

Identify the peak: Locate the pixel with the highest intensity in the dirty image, representing the strongest source.

Model the source: Represent this peak as a point source and scale it by a gain factor (typically between 0.1 and 1.0) to prevent over-subtraction.

Subtract the PSF: Subtract the scaled PSF, centered at the peak location, from the dirty image. This removes the influence of the detected source and its associated artifacts.

Iterative refinement: Repeat the identification and subtraction process until a stopping criterion is met, such as reaching a noise threshold or completing a fixed number of iterations.

Construct the CLEAN image: Combine all identified sources (CLEAN components) into a model image and convolve them with a smoothed PSF to restore resolution.

Add residuals: Add the residual image (which contains noise and unresolved sources) back to the CLEAN model to produce the final image.

CLEAN provides a simple and robust framework for interferometric imaging, making it the gold standard in radio astronomy for decades. However, it can be computationally intensive, particularly for large datasets. CLEAN is also sensitive to calibration errors and struggles with complex source structures or noise-dominated regions. It assumes that the PSF is well-characterized, which may not hold true in all observational scenarios, especially for wideband or wide-field imaging.

Additional Comparison Examples with CLEAN and POLISH Methods

See Fig. 10.