Research in Astronomy and Astrophysics, 25:104001 (10pp), 2025 October
https://doi.org/10.1088/1674-4527/adf710
CSTR: 32081.14.RAA.adf710
© 2025. National Astronomical Observatories, CAS and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

A U-Net Based Method for Radio Astronomical Image Deconvolution

Xinghui Zhou[1], Qianyun Yun[2], Hui Deng[1], Yangfan Xie[1], Yijun Xu[1], Feng Wang[1], and Ying Mei[1]
[1] Center For Astrophysics and Great Bay Center of National Astronomical Data Center, Guangzhou University, Guangzhou 510006, China; denghui@gzhu.edu.cn, fengwang@gzhu.edu.cn
[2] Shanghai Jiao Tong University, Shanghai 200240, China
Received 2025 February 4; revised 2025 May 13; accepted 2025 May 14; published 2025 September 12

Abstract

Deconvolution in radio interferometry faces challenges due to incomplete sampling of the visibilities in the spatial frequency domain caused by a limited number of antenna baselines, resulting in an ill-posed inverse problem. Reconstructing dirty images into clean ones is crucial for subsequent scientific analysis. To address these challenges, we propose a U-Net based method that extracts high-level information from the dirty image and reconstructs a clean image by effectively reducing artifacts and sidelobes. The U-Net architecture, consisting of an encoder-decoder structure and skip connections, facilitates the flow of information and preserves spatial details. Using simulated data of radio galaxies, we train our model and evaluate its performance on the testing set. Compared with the CLEAN method and the visibility and image conditioned denoising diffusion probabilistic model, our proposed model can effectively reconstruct both extended sources and faint point sources with higher values in the structural similarity index measure and the peak signal-to-noise ratio. Furthermore, we investigate the impact of noise on the model performance, demonstrating its robustness under varying noise levels.

Key words: techniques: image processing – techniques: interferometric – radio continuum: galaxies – methods: data analysis

1. Introduction

With the advent and rapid development of the techniques of radio interferometry (Ryle & Hewish 1960) and very long baseline interferometry (Broten et al. 1967), virtual telescopes with baselines spanning vast distances were created, resulting in extremely high angular resolutions and new opportunities for human understanding of the universe. Current radio-interferometric techniques rival or exceed optical imaging in angular resolution, despite operating at wavelengths thousands of times longer (Asaki et al. 2023). Modern radio interferometers have led to significant scientific breakthroughs, such as capturing the first image of the M87 supermassive black hole with the Event Horizon Telescope (EHT, The Event Horizon Telescope Collaboration et al. 2019) and the precise location of extragalactic transients (Hallinan et al. 2017).

Radio interferometers capture celestial signals, yielding visibility data related to sky objects through a Fourier transform, as articulated by the van Cittert–Zernike theorem (Thompson et al. 2017). Despite this theoretical framework, the finite number of antennas leads to incomplete sampling in the spatial frequency domain, commonly resulting in imaging artifacts (Schmidt et al. 2022). Theoretically, the deconvolution problem of recovering the underlying sky brightness from sparse visibility is a very difficult and challenging problem (Wang et al. 2023). This long-lasting issue has been the focus of computational and astronomical experts, driving continuous innovation in image reconstruction techniques over the past several decades (Connor et al. 2022).

Many methods have been proposed to address this ill-posed inverse problem. Classical methods such as the CLEAN algorithm have long been the cornerstone of radio interferometric imaging. CLEAN, introduced by Högbom (Högbom 1974) and further refined by Clark (Clark 1980), iteratively models the brightest points in the dirty image as point sources, subtracting a scaled response from the data until no significant sources remain. Then, the Multi-Scale CLEAN algorithm was later developed as a variant of the CLEAN that can better deal with extended sources (Bhatnagar & Cornwell 2004; Cornwell 2008).

In addition to CLEAN-based algorithms, some other methods have been developed. For example, the Regularized Maximum Likelihood Method (Frieden 1972; Müller et al. 2023; Zawadzki et al. 2023; Mus et al. 2024) is a Bayesian-inspired method that solves for an image by minimizing an objective function (Connor et al. 2022; Zawadzki et al. 2023). Maximum Entropy Methods (Frieden 1972; Cornwell & Evans 1985; Ramesh & Nityananda 1986; Bontekoe et al. 1994; Mus & Martí-Vidal 2024) aim to find the image that fits the observed data while maximizing the entropy. Bayesian Methods (Molina et al. 2001; Broderick et al. 2020; Arras et al. 2022) provide a statistical framework that uses prior knowledge of the image and observed data to infer the probable true image (Molina et al. 2001). Compressed sensing methods (Wiaux et al. 2009; Wenger et al. 2010; McEwen & Wiaux 2011; Carrillo et al. 2012, 2014) can recover the signal or image

from sparse measurements. The sparse modeling methods (Honma et al. 2014; Akiyama et al. 2017) are capable of finding a sparse solution to produce super-resolution images.

In recent years, with rapid growth in data volume and complexity, learning-based methods have been widely used in astronomy on various tasks (Baron 2019). For example, Medeiros et al. (2023) proposed a novel technique, Principal Component Interferometric Modeling (PRIMO), which leverages Principal Component Analysis (PCA) to reconstruct images of black holes from data obtained by EHT. Feng et al. (2024) introduced a framework for reconstructing images of the supermassive black hole M87* using deep generative image priors. Lai et al. (2025) proposed a novel deep learning method, Deep Learning Image Reconstruction with Closure Terms (DIReCT), which can reconstruct the image from closure invariants directly.

Especially, Schmidt et al. (2022) developed Radionets to recover the missing value in the sparsely sampled visibility and then convert the full sampled visibility into a clean image, despite still needing a large amount of unnecessary computation (Wang et al. 2024). Another work by Wu et al. (2022) also tried to recover the sparse visibility measurement in a dense spectral reconstruction using the Transformer (Vaswani et al. 2017) and FiLM conditioning (Perez et al. 2018). Then, Wang et al. (2023) further developed a Visibility and Image Conditioned Denoising Diffusion Probabilistic Model (VIC-DDPM) that utilizes both visibility and image data to clean a dirty image. However, they overlooked the challenge of accurately reconstructing weak signals, such as faint point sources.

In this work, we employ the U-Net architecture (Ronneberger et al. 2015) as a backbone to address the limitations of existing methods and improve the performance of deconvolution through machine learning. The rest of the paper is organized as follows. In Section 2, we describe the data preparation process, including the simulation of dirty images and the preprocessing steps for the neural network. In Section 3, we present the architecture of our model and the loss function. In Section 4, we show the training details of our model. In Section 5, we evaluate our model's performance and compare it with other methods in terms of image quality and computing efficiency. Finally, in Section 6, we present our conclusions and future work.

2. Data Preparation

In our study, we use the data set provided by Wu et al. (2022). This data set is also used in VIC-DDPM (Wang et al. 2023). This data set contains a total of 17,736 galaxy images obtained from Galaxy10 DECals.[3] and visibility data generated by the EHT-imaging (Chael et al. 2019) toolkit with an 8-telescope EHT configuration for the observation of M87* from 2017 (Wu et al. 2022). Then we use the imaging method in VIC-DDPM (Wang et al. 2023) to generate the dirty image as input to our model.

Our ground truth (GT) image with dimensions of 128×128 pixels is the downsample version of their original higher resolution galaxy image (256×256 pixels). Our model's target image is generated by convolving the GT image with the clean beam. This process yields the maximum resolution image that can be observed by the telescope, free of artifacts (Bouman et al. 2016). These can be illustrated in Figure 1.

In fact, these data are typically affected by noise and artifacts due to sparse UV coverage, atmospheric disturbances, and instrumental limitations. Gain errors and calibration errors will also have the potential risk of generating artifacts. In this study, we focus solely on the influence of sparse UV coverage and random Gaussian noise on the sampled visibility.

To be consistent with real observations, we added random Gaussian noise with various standard deviations of σ to the visibility at the sampled points (u, v). Then we use these noisy visibility data to generate the dirty image. The dirty image and the target image can be described as:

$$x = \mathrm{DFT}(V + \mathcal{N}(0, \sigma^2)), \tag{1}$$

$$y = \mathrm{GT} \otimes \text{clean beam}, \tag{2}$$

where V is the sampled visibility without noise generated from simulation observation, $\mathcal{N}(0, \sigma^2)$ is the random Gaussian noise with zero mean and various variance σ^2. We use the discrete Fourier transform (DFT) to generate the dirty image, which is the same as Wang et al. (2023). y is the target image obtained by convolving the GT image with the clean beam.

We then normalize the dirty images with noise and the target images within the range between 0 and 1 using Min–Max normalization. The Min–Max normalization is:

$$x_{\mathrm{normalize}} = \frac{x - \min(x)}{\max(x) - \min(x)}, \tag{3}$$

$$y_{\mathrm{normalize}} = \frac{y - \min(y)}{\max(y) - \min(y)}. \tag{4}$$

This normalization step accelerates the convergence of the model and stabilizes the training process. After the above data preparation, we obtain the data set for training and testing.

3. Model and Loss Function

3.1. Model Architecture

Our model aims to remove artifacts in the dirty image caused by sparse UV coverage. To achieve this, we design a deep learning model based on the U-Net architecture (Ronneberger et al. 2015), which was originally developed for biomedical image segmentation. U-Net employs skip connections between symmetrically paired downsampling and upsampling blocks, enabling accurate pixel-wise intensity

[3] https://github.com/henrysky/Galaxy10

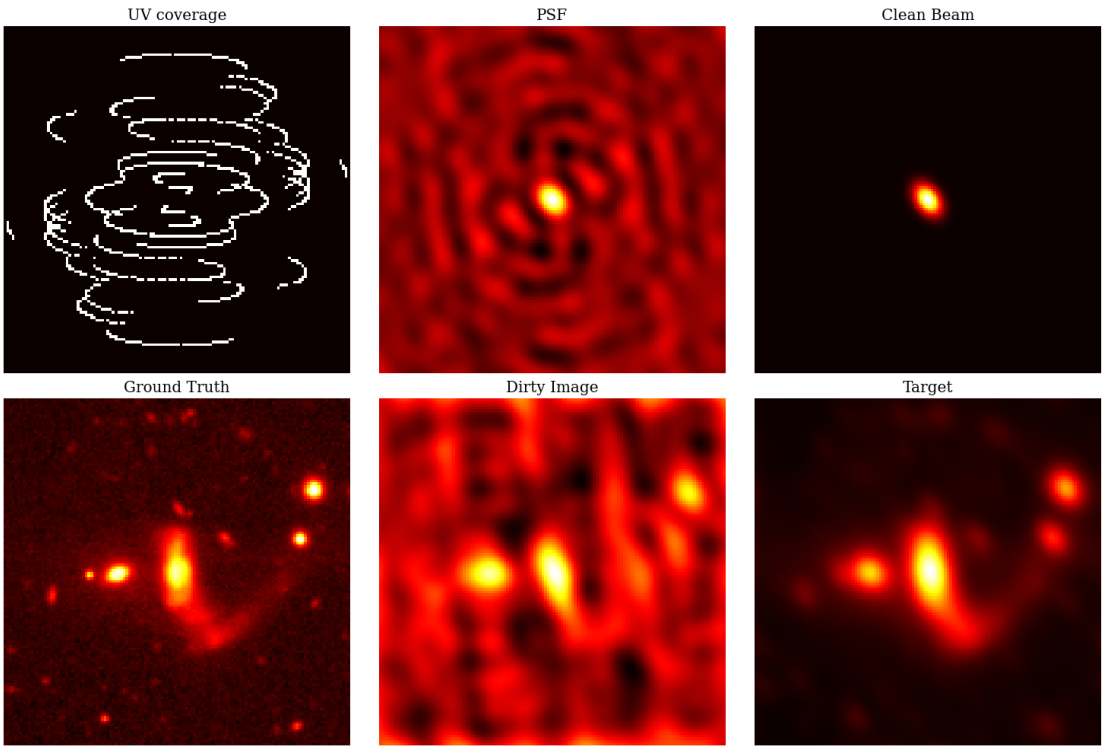


Figure 1. Data preparation. (a) UV-coverage of the 8-telescope EHT array from VIC-DDPM; (b) Point-Spread Function (PSF); (c) Clean beam; (d) Ground truth image downsampled from 256 × 256 to 128 × 128 pixels; (e) Dirty image generated using the method in VIC-DDPM; (f) Target image created by convolving the ground truth with the clean beam, representing the maximum artifact-free resolution observable by the telescope.

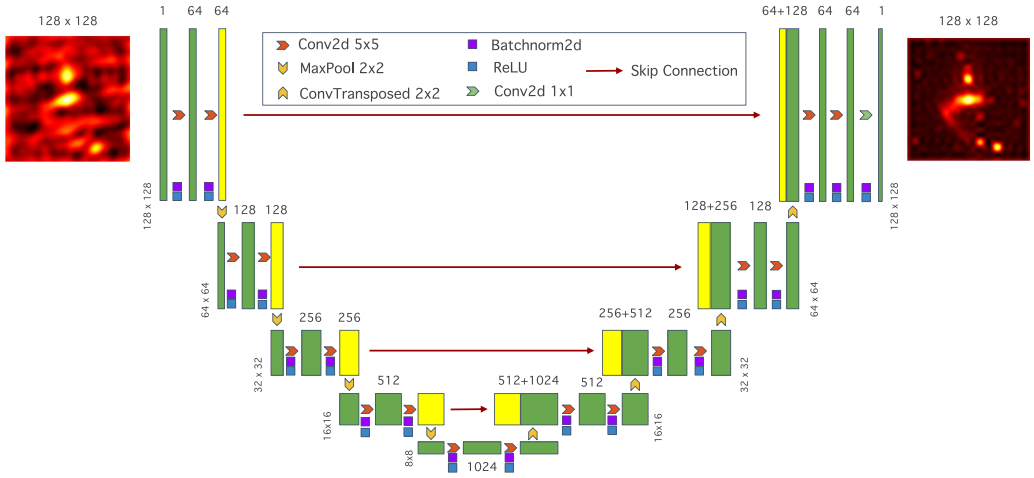


Figure 2. Comprehensive representation of our U-Net based model. The model consists of two main components: the compressing path on the left and the expanding path on the right. The dirty image is inputted into the InConv block, and the reconstructed image is outputted from the OutConv block.

prediction. Unlike the method that fills the missing visibility values (Wu et al. 2022) and the method that uses both visibility data and image data to predict a clean image (Wang et al. 2023), our approach specifically targets the removal of artifacts present in the dirty image due to the sparse UV coverage. Consequently, the input of our model is a "dirty image" and the output is a "reconstructed image."

Figure 2 provides a comprehensive illustration of our model. The model is composed of two main components: the contracting path and the expanding path. The contracting path (left half) consists of multiple blocks designed to identify and eliminate artifacts in the dirty image, reducing it to a more compact representation. The expanding path (right half) then reconstructs an artifact-free image. Using a large data set, the

Table 1
Overview of the U-Net Architecture

Stage	Layer	Channel In	Channel Out	Kernel Size	Stride	Padding	Image Size
InConv	DoubleConv	1	64	5	1	2	128 × 128
Down 1	MaxPool2d	64	64	2	2	0	64 × 64
	DoubleConv	64	128	5	1	2	64 × 64
Down 2	MaxPool2d	128	128	2	2	0	32 × 32
	DoubleConv	128	256	5	1	2	32 × 32
Down 3	MaxPool2d	256	256	2	2	0	16 × 16
	DoubleConv	256	512	5	1	2	16 × 16
Down 4	MaxPool2d	512	512	2	2	0	8 × 8
	DoubleConv	512	1024	5	1	2	8 × 8
Up 1	ConvTranspose2d	1024	512	2	2	0	16 × 16
	DoubleConv	1024	512	5	1	2	16 × 16
Up 2	ConvTranspose2d	512	256	2	2	0	32 × 32
	DoubleConv	512	256	5	1	2	32 × 32
Up 3	ConvTranspose2d	256	128	2	2	0	64 × 64
	DoubleConv	256	128	5	1	2	64 × 64
Up 4	ConvTranspose2d	128	64	2	2	0	128 × 128
	DoubleConv	128	64	5	1	2	128 × 128
OutConv	Conv2d	64	1	1	1	0	128 × 128

Note. For each layer, the number of input and output channels, kernel size, stride, padding, and image size changes are given.

model effectively identifies and removes artifacts introduced by the Point-Spread Function (PSF), resulting in an estimate of the intensity of the true sky $I_{\rm sky}$ or $I_{\rm target}$. Table 1 provides a detailed description of the model's architecture. The code of our model is available at https://github.com/urahyou/AI_CLEAN.

3.2. Loss Function

We define our model abstractly as M. The output of the model, denoted as $\hat{y}$, represents the best estimation. It is defined as:

$$\hat{y} = M(x), \tag{5}$$

where x is the normalized input data (noisy dirty image) as described in Section 2.

To achieve this goal, we design a custom loss function using the pixel-wise Mean Square Error (MSE) between the prediction and the target image as the main loss term. In addition, we also add an SSIM loss term and an MAE (Mean Absolute Error) loss term. So the loss function is defined as:

$$L_{\rm Total} = L_{\rm MSE}(\hat{y}, y) + \alpha L_{\rm SSIM}(\hat{y}, y) + \beta L_{\rm MAE}(\hat{y}, y). \tag{6}$$

The MSE is the Mean Square Error between the model prediction $\hat{y}$ and the target y on the pixel level, which is:

$$L_{\rm MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2. \tag{7}$$

The SSIM Loss term $L_{\rm SSIM}(\hat{y}, y)$ is:

$$L_{\rm SSIM}(\hat{y}, y) = 1 - {\rm SSIM}(\hat{y}, y). \tag{8}$$

The MAE Loss term $L_{\rm MAE}(\hat{y}, y)$ is:

$$L_{\rm MAE}(\hat{y}, y) = \frac{1}{N}\sum_{i=1}^{n}|\hat{y}_i - y_i|. \tag{9}$$

The SSIM loss is often used in super-resolution and image processing. It captures changes in structural information in the image by comparing local regions instead of using a global approach (Hwang et al. 2020; Nay 2021). With two images x and y, SSIM(x, y) can be written as (Nay 2021):

$${\rm SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \tag{10}$$

Then we set α to 0.01 and β to 0.5. The SSIM we used in the loss function is implemented in the TorchMetrics (Detlefsen et al. 2022) package. In their implementation, the constants C_1

and C_2 are defined as

$$C_1 = (K_1 \cdot R)^2, \tag{11}$$

$$C_2 = (K_2 \cdot R)^2, \tag{12}$$

where R is the dynamic range of the image data, typically the difference between the maximum and minimum pixel values. The constants K_1 and K_2 are set to 0.01 and 0.03, respectively (Wang et al. 2004). In short, our model is designed to learn and produce an estimation, denoted by $I_{\rm estimate}$, which is the reconstructed image of the dirty image $I_{\rm dirty}$. The goal is to determine the optimal parameters θ that minimize the loss function $L_{\rm total}$, which quantifies the difference between the model output and the target image. This optimization problem can be mathematically expressed by:

$$\hat{\theta} = \underset{\theta}{\rm argmin}(L_{\rm Total}(M(I_{\rm dirty}), I_{\rm target})|\theta). \tag{13}$$

4. Model Training

To effectively train and test our model, we strategically split the data set into a ratio of 7:3 for training and testing, respectively. This partition leads to a training set consisting of 12415 pairs of samples ($I_{\rm dirty}$, $I_{\rm target}$) and a testing set containing 5321 pairs. The incorporation of noise during the training phase, coupled with our meticulously designed loss function, plays a pivotal role in achieving this balanced data set division. This approach not only enhances the model's robustness to real-world conditions but also ensures a reliable evaluation on unseen data.

Instead of using the stochastic gradient descent (SGD) optimizer, we use the adaptive moment estimation (ADAM) optimizer (Kingma & Ba 2014), which is known for its efficiency in handling sparse gradients and adapting the learning rate. The learning rate is set to 0.0005. The model is trained for 150 epochs. Figure 3 depicts the loss curve throughout the training process. The close alignment between the training loss and the testing loss shows that our model does not suffer from overfitting. This indicates that our model is capable of performing robustly on both the training data and the unseen testing data.

5. Model Evaluation

In this section, we evaluate the ability of our model to deconvolve a dirty image in the image domain. We apply the trained model described above to the test data set used in VIC-DDPM (Wang et al. 2023). We then compare the reconstruction intensity on the pixel level with the target image and show the deviation between them. Then we compare our deconvolution ability with the Högbom CLEAN (Högbom 1974) and the deep learning-based method: VIC-DDPM (Wang et al. 2023).

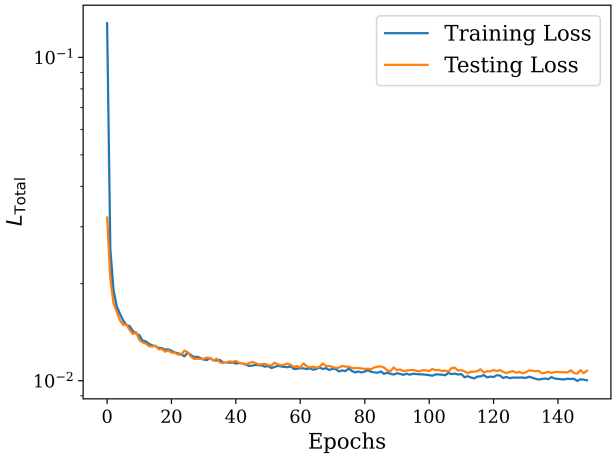


Figure 3. Loss curves for training and testing over epochs. The loss values are plotted as a function of epochs. The consistent reduction in both training and testing loss indicates that the model is generalizing well and is not overfitting to the training data. To more clearly depict the trend of loss values, the y-axis of the loss function plot employs a logarithmic scale.

5.1. Evaluation Metrics

To evaluate the performance of our model, we adopt the traditional image quality metrics: Peak Signal-to-Noise Ratio (PSNR; Hore & Ziou 2010) and SSIM (Wang et al. 2004), which are two well-known image quality metrics used in computer vision. Given a target image y and a model output image $\hat{y}$, the PSNR can be described as:

$$\mathrm{PSNR}(y, \hat{y}) = 10\log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}(y, \hat{y})}\right), \tag{14}$$

where MAX_I^2 is the maximum possible pixel value of the image. The MSE is calculated as:

$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[y(i,j) - \hat{y}(i,j)]^2. \tag{15}$$

It is important to note that the PSNR value approaches infinity as the MSE value approaches zero. This relationship indicates that a higher PSNR value corresponds to a better image reconstruction quality, and vice versa (Setiadi 2021).

On the other hand, the SSIM metric, developed by Wang et al. (2004), is well known for its alignment with the human visual system (Hore & Ziou 2010). SSIM is based on three key components: luminance comparison $l(x, y)$, contrast comparison $c(x, y)$, and structure comparison $s(x, y)$.

The term $l(x, y)$ is a luminance comparison function. Comparison of the average pixel intensity values of the two images ensures similar brightness levels. $I(x, y)$ will reach the maximum value 1 only when $\mu_x = \mu_y$ (Hore & Ziou 2010). In the context of our study, this factor measures the closeness of the mean intensity between the output image and the target

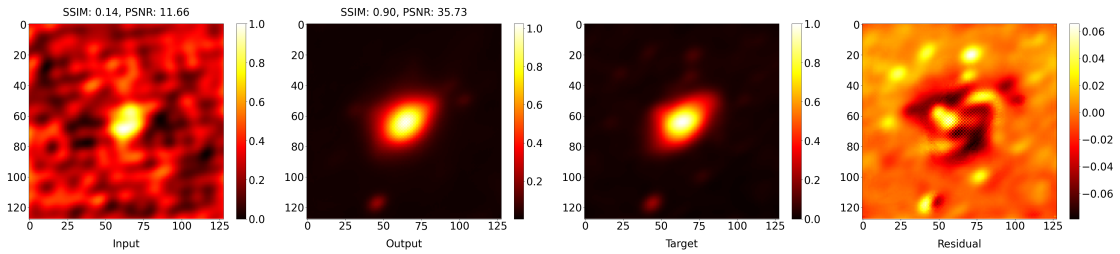


Figure 4. Exemplary reconstruction of brightness distributions. This figure demonstrates the capabilities of our model in the deconvolution problem. Arranged from left to right, the figure is compartmentalized into four distinct panels, representing the input data, the model's output, the target image, and the residual between output and target. We normalize the input, output, and target images to the range [0, 1] when calculating the SSIM and PSNR.

image. The $l(x, y)$ is defined as:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}. \tag{16}$$

The term $c(x, y)$ represents the contrast comparison function. It evaluates the similarity in the dynamic range of pixel values, which corresponds to the contrast of the images. The contrast is measured by the standard deviations σ_x and σ_y. The value of $c(x, y)$ reaches its maximum of 1 when $\sigma_x = \sigma_y$ (Hore & Ziou 2010). The $c(x, y)$ is defined as:

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \tag{17}$$

The term $s(x, y)$ is the Structure Comparison function that measures the similarity in the patterns of the pixel values, which relates to the spatial distribution of features in the images. In particular, σ_{xy} is the covariance between x and y (Hore & Ziou 2010), which quantifies the correlation between the output image and the target image. The $s(x, y)$ is defined as:

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}. \tag{18}$$

Combining the three components, the SSIM can be described as:

$$\mathrm{SSIM}(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \tag{19}$$

where α, β and γ are the weights of each term. In our study, they are all set to 1. C_1 and C_2 have been defined in Equations (11) and (12), and C_3 is defined as $C_3 = C_2/2$ (Wang et al. 2004). In our evaluation step, we use the PSNR and SSIM implemented in the scikit-image package (Van der Walt et al. 2014).

Notably, when calculating the SSIM and PSNR, we first normalize the pixel values of the two images to the range of [0, 1].

5.2. Reconstruction of Brightness Distributions

Our deep-learning model effectively reconstructs a dirty image into a clean one by eliminating artifacts. In Figure 4, we present an example showing the input image, the reconstructed image (output), the target image, and the residual from left to right.

The first panel on the left shows the input image, representing the dirty image we assume to receive from radio telescopes. These data are typically affected by noise and artifacts due to sparse UV coverage, atmospheric disturbances, and instrumental limitations. In this work, we focus solely on the influence of sparse UV coverage and noise.

Moving to the second panel, the output from our model showcases the results of our cleaning process. In this figure, we observe a significant improvement in the clarity and resolution of the image. Our model has effectively removed noise and artifacts, providing a much cleaner and more detailed representation of the celestial radio sources.

The third panel is the target image, which represents the ideal or target image that our model aims to reconstruct. The target image acts as a benchmark for assessing our model's performance. The close alignment between the model output and the target image in the figure highlights the exceptional accuracy and efficiency of our model in reconstructing high-fidelity astronomical images.

Finally, the fourth panel represents the residual between the model output and the target image. As shown in the figure, the residual fluctuates between -0.07 and 0.07, with the most significant fluctuations concentrated around the source.

In fact, our model does not reconstruct the absolute distribution of celestial intensity but rather yields a clean image with relative intensity distribution. If we record the range of the input dirty image, then we could rescale the intensity distribution back to match the input scale, or through the calibration in the post-processing step.

5.3. Comparison with Other Methods

5.3.1. Reconstruction Quality

To evaluate the performance of our model, we compare the reconstruction quality with other methods, including CLEAN and VIC-DDPM. In fairness, we used the test data set that contains 512 images from VIC-DDPM. The visual performance of our model and the other different methods can be

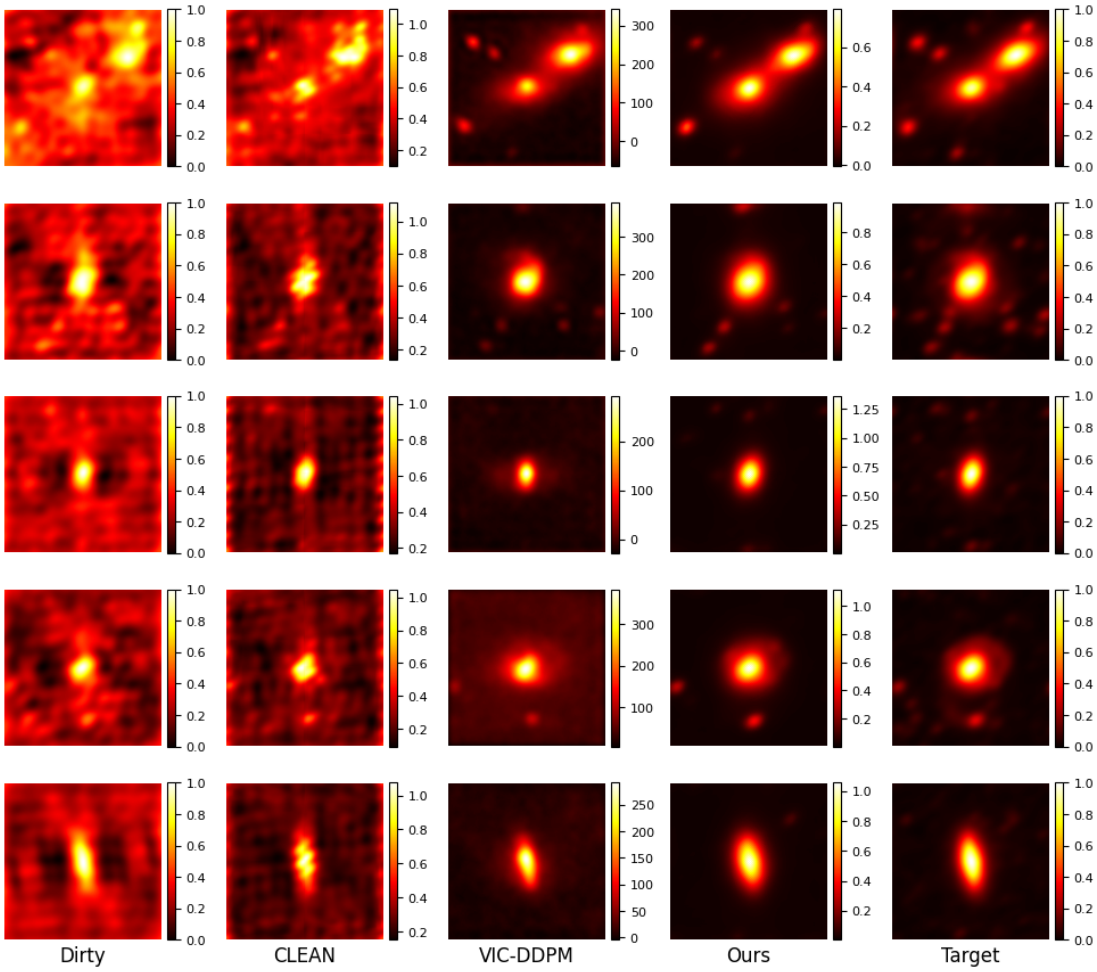


Figure 5. Visual comparisons of different models' performance. This figure contains five columns, from left to right respectively present the (a) Dirty, (b) CLEAN, (c) VIC-DDPM, (d) Ours, and (e) Target. Each row represents a different sample.

seen in Figure 5. The first column displays the dirty images that require deconvolution, followed by clean results from the Högbom CLEAN method in the second column, the VIC-DDPM model in the third column, and our model in the fourth column. The final column is the target images. The CLEAN algorithm that we utilized in this study is referenced in Fundamentals of Radio Interferometry.[4] Then we set the gain to 0.01, the number of iterations to 1500 and the stop threshold to 0.000001, without boxing.

All three methods are capable of reconstructing signals from dirty images and producing clean outputs. However, in terms of image quality, our model outperforms the others. Although the CLEAN method is effective at identifying point sources, it still struggles to recover extended sources. We recognize that this limitation may be partly attributed to the chosen parameters in the iterative process of the CLEAN algorithm, and optimizing these parameters could potentially yield better results. This highlights the fact that the CLEAN algorithm heavily relies on the operator's expertise and demands careful tuning for each image.

The VIC-DDPM model effectively reduces artifacts and preserves details while recovering central extended sources from dirty images. However, it struggles with detecting faint and small sources. In contrast, our model, as demonstrated in the fourth column, excels not only in artifact reduction and detail preservation but also in recovering faint sources that other methods miss.

We further analyze and compare our reconstruction results using the histograms presented in Figure 6. Our model (depicted in red) consistently achieves higher SSIM and PSNR values compared to the other methods.

The Högbom CLEAN method shows some improvements by mitigating the negative effects caused by the PSF, resulting in an SSIM of approximately 0.30 and a PSNR of

[4] https://github.com/astronomical-data-processing/fundamentals_of_interferometry_chinese

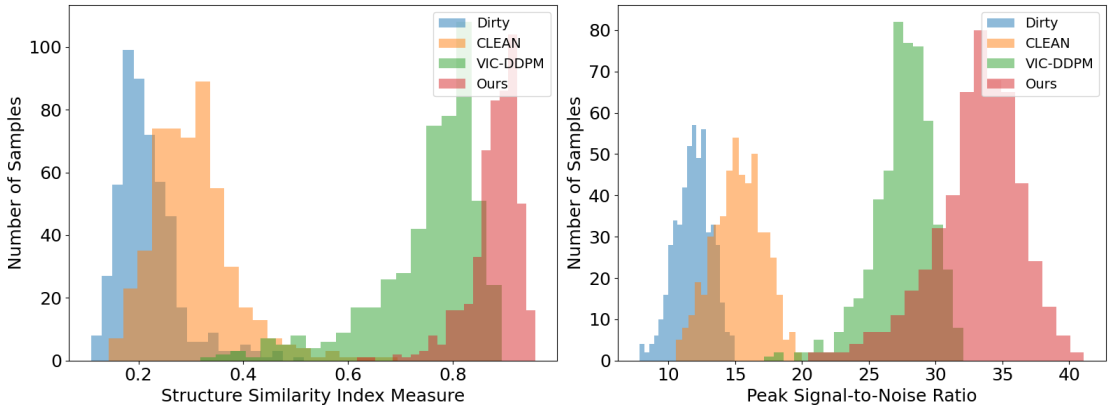


Figure 6. Comparison of SSIM and PSNR across different methods. The dirty input images (blue) predominantly have SSIM values below 0.5 and PSNR values mostly below 15, indicating low structural similarity and image quality. The CLEAN method (orange) shows moderate improvement in both metrics, with SSIM values varying around 0.2 and 0.5, and PSNR values concentrated around 10 and 20. The VIC-DDPM model (green) performs better, with SSIM values clustered around 0.6–0.85 and PSNR around 25–30. Our model (red) achieves the highest SSIM values, with the majority exceeding 0.8, and most PSNR values exceeding 30, indicating better image reconstruction quality.

approximately 15.24 on average. The VIC-DDPM method performs better in reducing artifacts, leading to an SSIM of approximately 0.75 and a PSNR of about 27.51 on average.

However, our model outperforms all the others, achieving a PSNR of 32.71 and an SSIM of 0.87. These results underscore the effectiveness of our model in radio astronomical deconvolution.

The result shows that our model can reconstruct not only the extended sources but also the point sources in the dirty images. Furthermore, our model has better performance in reconstructing faint sources compared with VIC-DDPM and CLEAN.

5.3.2. Reconstruction on Different Noise Levels

To comprehensively evaluate the performance of different methods under various noise conditions, we add different levels of noise ($\sigma = 0.001$, 0.01, 0.1, 0.3) to the visibility data. Then we compare the reconstruction performance using different methods including the CLEAN method, VIC-DDPM, and our proposed U-Net based method.

As shown in Figure 7, we add different levels of noise to the visibility data. At low noise levels ($\sigma = 0.001$ and $\sigma = 0.01$), all methods can restore the image fairly well, but our method can retain more details closer to the target image. When the noise level reaches 0.1, our method can still restore the image quite well, while other methods may start to lose some details and introduce many fake components. When the noise level reaches 0.3, VIC-DDPM and CLEAN methods could hardly recover the central extended sources in the image and tend to introduce a large number of spurious sources. Our method also introduces some spurious sources, but remains effective in recovering the extended source and preserving background clarity to a reasonable extent.

5.3.3. Computational Efficiency and Resource Usage

In this section, we compare the computational efficiency and the usage of GPU memory of different methods (models).

As shown in Table 2, the CLEAN method, executed on an Intel® Xeon® Platinum 8260 CPU with the same parameter settings detailed in Section 5.3.1, takes approximately 0.104 s to process a single image. Our model runs on an NVIDIA 4090 GPU, requiring 0.01 s to process one image. The VIC-DDPM method, which is executed on the same 4090 GPU, takes about 137.118 s for the same task. This comparison highlights the efficiency of our model, demonstrating a significant reduction in processing time over both the traditional CLEAN method and the VIC-DDPM method.

Our model's relatively low GPU memory usage of 0.56 GB with 81.2 million parameters makes it more suitable for deployment on GPU architectures. In contrast, the VIC-DDPM model requires a substantial 1.69 GB of GPU memory with 167.10 million parameters.

Overall, the results indicate that our model is superior in both resource usage and computational efficiency.

6. Conclusion and Future Works

In this work, we introduce the U-Net architecture and demonstrate that the U-Net based method can effectively tackle the complex challenge of deconvolution in interferometric imaging, achieving a higher SSIM of 0.87 and a PSNR of 32.71 on average. Moreover, we find that the incorporation of the SSIM loss and the MAE loss can enhance the performance of our model. SSIM loss may improve the performance of the model in recovering the complex extended structure, and MAE loss will make residual smoother. In our current study, we train and test the model using simulated data, and we demonstrate the model's ability to deconvolve dirty

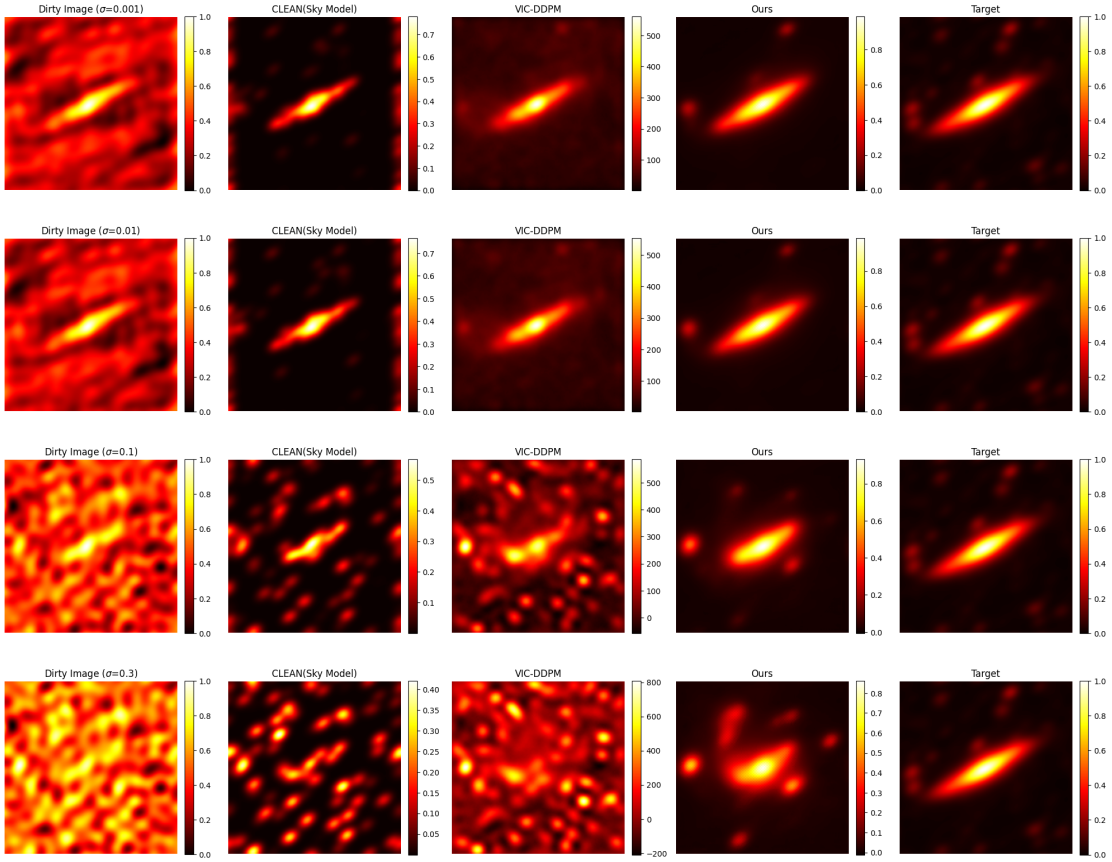


Figure 7. Reconstruction results at noise levels $\sigma = 0.001$, 0.01, 0.1, and 0.3 on visibilities. Rows correspond to noise levels; columns, from left to right, show the dirty image, CLEAN (Sky Model), VIC-DDPM, our method, and the target image. At $\sigma = 0.001$ and 0.01, all methods recover well, but our method better preserves extended sources. At $\sigma = 0.1$ and 0.3, our method maintains effective reconstruction while others will introduce some artifacts. This visual comparison suggests that our method demonstrates a comparatively better resilience to noise perturbations in the sparse visibility.

Table 2
Comparison of Parameters, GPU Memory Usage, and Computing Time, Using an NVIDIA 4090 GPU and a CPU (Intel® Xeon® Platinum 8260 @ 2.40 GHz)

Model	Image Size	Parameters (million)	GPU Memory Usage (GB)	Computing Time (s)
CLEAN	(128, 128)	···	···	0.104
VIC-DDPM	(128, 128)	167.10	1.69	137.118
U-Net(Ours)	(128, 128)	81.20	0.56	0.010

images into clean ones, emphasizing the effectiveness and efficiency of our approach.

We also compare our model with the classical CLEAN algorithm, as well as the more recent approach, VIC-DDPM. Performance metrics, including SSIM and PSNR, indicate that our model achieves the most accurate reconstruction performance, closely matching the target ones by preserving the extended sources and finding faint point sources. These results underscore the potential of our model for interferometric imaging, highlighting it as a valuable method for astronomers and researchers in the field. Meanwhile, to evaluate the robustness of our algorithm, we introduced various levels of noise for comparative analysis. The result implies that our model is capable of effectively reconstructing noisy dirty images, even at different noise levels, thus maintaining a reasonable level of accuracy and reliability in the reconstruction process.

Currently, we focus on the impact of incomplete UV coverage at different noise levels, and are capable of reconstructing a dirty image into a clean one with a relative intensity distribution. In addition, we have investigated the impact of noise in different methods, demonstrating the robustness of our model under varying noise levels. In the future, we plan to develop a more robust model based on different kinds of telescope configurations such as ALMA (Partnership et al. 2015) and SKA (Dewdney et al. 2009; Xie et al. 2022) with different UV coverage. We will investigate how to calibrate the relative intensity distribution to the absolute intensity distribution. This will enable us to process

more complex radio interferometric data and meet the need for higher resolution and sensitivity, as well as real-time data processing capabilities.

Acknowledgments

This work is supported by the National SKA Program of China (2020SKA0110300, 2020SKA0110201), the National Natural Science Foundation of China (NSFC, grant Nos. 12433012 and 12373097), the Guangdong Province Project of the Basic and Applied Basic Research Foundation (2024A1515011503), the Guangzhou Science and Technology Funds (2023A03J0016).

References

Akiyama, K., Ikeda, S., Pleau, M., et al. 2017, AJ, 153, 159
Arras, P., Frank, P., Haim, P., et al. 2022, NatAs, 6, 259
Asaki, Y., Alcalde Pampliega, B., Edwards, P. G., Iguchi, S., & Murphy, E. J. 2023, NaRMP, 3, 89
Baron, D. 2019, arXiv:1904.07248
Bhatnagar, S., & Cornwell, T. 2004, A&A, 426, 747
Bontekoe, T. R., Koper, E., & Kester, D. 1994, A&A, 284, 1037
Bouman, K. L., Johnson, M. D., Zoran, D., et al. 2016, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (Las Vegas, NV, USA: IEEE), 913
Broderick, A. E., Pesce, D. W., Tiede, P., Pu, H.-Y., & Gold, R. 2020, ApJ, 898, 9
Broten, N., Legg, T., Locke, J., et al. 1967, Sci, 156, 1592
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, MNRAS, 426, 1223
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2014, MNRAS, 439, 3591
Chael, A. A., Bouman, K. L., Johnson, M. D., et al., 2019 Astrophysics Source Code Library, record, ascl:1904.004
Clark, B. 1980, A&A, 89, 377
Connor, L., Bouman, K. L., Ravi, V., & Hallinan, G. 2022, MNRAS, 514, 2614
Cornwell, T. J. 2008, ISTSP, 2, 793
Cornwell, T. J., & Evans, K. 1985, A&A, 143, 77
Detlefsen, N. S., Borovec, J., Schock, J., et al. 2022, JOSS, 7, 4101
Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. 2009, IEEEP, 97, 1482
Feng, B. T., Bouman, K. L., & Freeman, W. T. 2024, ApJ, 975, 201
Frieden, B. R. 1972, JOSA, 62, 511
Hallinan, G., Corsi, A., Mooley, K., et al. 2017, Sci, 358, 1579
Högbom, J. 1974, A&AS, 15, 417
Honma, M., Akiyama, K., Uemura, M., & Ikeda, S. 2014, PASJ, 66, 95
Hore, A., & Ziou, D. 2010, in 2010 20th Int. Conf. on Pattern Recognition (Piscataway, NJ: IEEE), 2366
Hwang, J., Yu, C., & Shin, Y. 2020, in Int. Conf. Information and Communication Technology Convergence (ICTC) (Piscataway, NJ: IEEE), 191
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Lai, S., Thyagarajan, N., Wong, O. I., Diakogiannis, F., & Hoefs, L. 2025, MNRAS, 536, 446
McEwen, J. D., & Wiaux, Y. 2011, MNRAS, 413, 1318
Medeiros, L., Psaltis, D., Lauer, T. R., & Özel, F. 2023, ApJ, 943, 144
Molina, R., Núñez, J., Cortijo, F. J., & Mateos, J. 2001, ISPM, 18, 11
Müller, H., Mus, A., & Lobanov, A. 2023, A&A, 675, A60
Mus, A., & Martí-Vidal, I. 2024, MNRAS, 528, 5537
Mus, A., Müller, H., Martí-Vidal, I., & Lobanov, A. 2024, A&A, 684, A55
Nay, D. J. B. 2021, Single Image Super Resolution using ESPCN – With SSIM Loss Available at SSRN: https://ssrn.com/abstract=3898913
Partnership, A., Fomalont, E., Vlahakis, C., et al. 2015, ApJL, 808, 1
Perez, E., Strub, F., De Vries, H., Dumoulin, V., & Courville, A. 2018, in Proc. AAAI Conf. Artificial Intelligence (New Orleans, LA, USA: AAAI), 3942
Ramesh, N., & Nityananda, R. 1986, ARA&A, 24, 127
Ronneberger, O., Fischer, P., & Brox, T. 2015, in Medical Image Computing and Computer-assisted Intervention–MICCAI 2015: 18th Int. Conf., Proc., Part III 18 (Berlin: Springer), 234
Ryle, M., & Hewish, A. 1960, MNRAS, 120, 220
Schmidt, K., Geyer, F., Fröse, S., et al. 2022, A&A, 664, A134
Setiadi, D. R. I. M. 2021, Multimedia Tools and Applications, 80, 8423
The Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019, ApJL, 875, L1
Thompson, A. R., Moran, J. M., Swenson, G. W., et al. 2017, Interferometry and Synthesis in Radio Astronomy (3rd ed.; Springer), 59
Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., et al. 2014, PeerJ, 2, e453
Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in Advances in Neural Information Processing Systems, ed. I. Guyon et al., Vol. 30 (Curran Associates, Inc.)
Wang, R., Chen, Z., Luo, Q., & Wang, F. 2023, in ECAI (Kraków, Poland: IOS Press), 2499
Wang, R., Chen, Z., Zhu, J., Luo, Q., & Wang, F. 2024, in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (Seattle, WA, USA: IEEE), 12841
Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. 2004, ITIP, 13, 600
Wenger, S., Magnor, M., Pihlström, Y., Bhatnagar, S., & Rau, U. 2010, PASP, 122, 1367
Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M., & Vandergheynst, P. 2009, MNRAS, 395, 1733
Wu, B., Liu, C., Eckart, B., & Kautz, J. 2022, in Proc. AAAI Conf. Artificial Intelligence (Virtual: AAAI), 2685
Xie, Y.-F., Wang, F., Deng, H., et al. 2022, MNRAS, 515, 1973
Zawadzki, B., Czekala, I., Loomis, R. A., et al. 2023, PASP, 135, 064503