ECAI 2023
K. Gal et al. (Eds.)

doi:10.3233/FAIA230554

A Conditional Denoising Diffusion Probabilistic Model for Radio Interferometric Image Reconstruction

Ruoqi Wang[a], Zhuoyang Chen[a], Qiong Luo[a;*] and Feng Wang[b]

[a]The Hong Kong University of Science and Technology (Guangzhou)
{rwang280, zchen190}@connect.hkust-gz.edu.cn, luo@ust.hk
[b]Guangzhou University
fengwang@gzhu.edu.cn

Abstract.

In radio astronomy, signals from radio telescopes are transformed into images of observed celestial objects, or sources. However, these images, called dirty images, contain real sources as well as artifacts due to signal sparsity and other factors. Therefore, radio interferometric image reconstruction is performed on dirty images, aiming to produce clean images in which artifacts are reduced and real sources are recovered. So far, existing methods have limited success on recovering faint sources, preserving detailed structures, and eliminating artifacts. In this paper, we present VIC-DDPM, a Visibility and Image Conditioned Denoising Diffusion Probabilistic Model. Our main idea is to use both the original visibility data in the spectral domain and dirty images in the spatial domain to guide the image generation process with DDPM. This way, we can leverage DDPM to generate fine details and eliminate noise, while utilizing visibility data to separate signals from noise and retaining spatial information in dirty images. We have conducted experiments in comparison with both traditional methods and recent deep learning based approaches. Our results show that our method significantly improves the resulting images by reducing artifacts, preserving fine details, and recovering dim sources. This advancement further facilitates radio astronomical data analysis tasks on celestial phenomena. Our code is available at https://github.com/RapidsAtHKUST/VIC-DDPM.

1 Introduction

Radio interferometers, or radio telescopes, receive radio waves from astronomical sources and produce interferometric measurement data, called *visibility*. These visibility data are in the spectral domain, and can be sampled and converted through inverse Fourier transforms into images for subsequent analysis. However, due to data sparsity and other factors, the result images are often dominated by artifacts [23]. As such, these *dirty images* must be reconstructed before they can be used in scientific analysis. Figure 1 illustrates the workflow of imaging and reconstruction. In this paper, we propose an image reconstruction method, aiming to effectively recover real sources and eliminate artifacts in dirty images.

Reconstruction of interferometric images is a challenging problem. Its effectiveness heavily depends on suitable priors to select from a vast array of potentially valid images in an infinite solution space [33]. For example, the classic CLEAN method [14] uses human knowledge as priors and assumes that the sky is composed of point-like sources [7]. As real celestial objects can be either point or extended sources, this assumption limits the quality of reconstructed images [33]. Therefore, recent deep-learning based approaches [18, 19, 33, 23] learn data-driven priors from training datasets. These methods successfully predict general structures of main objects in observation areas, but they still struggle in recovering faint sources, preserving details, and eliminating artifacts.

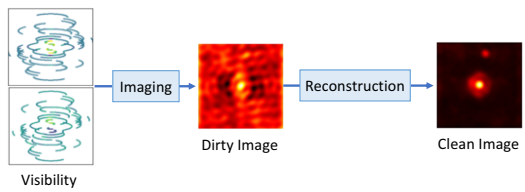

Figure 1. Imaging and Reconstruction

To perform high-quality radio interferometry reconstruction, we propose VIC-DDPM, a Visibility and Image Conditioned Denoising Diffusion Probabilistic Model (DDPM). We adopt a DDPM [13, 26] because it can generate images with fine details. We use both the original visibility data in the spectral domain and dirty images in the spatial domain to guide the clean image generation with DDPM. This way, our model can leverage the complementary strengths of both domains: Visibility data offers frequency information where noise can be effectively separated from signals [3], aiding in identifying artifacts; dirty images provide spatial information, helping recover dim sources and object structures. Consequently, our model removes artifacts as well as restores fine details, including details in main structures and surrounding dim sources. To the best of our knowledge, VIC-DDPM is the first work that adapts diffusion models to radio interferometric image reconstruction.

In short, our main contributions are as follows:

- We present a DDPM based generative model with two conditioning mechanisms, for dirty image and visibility data, respectively, in radio interferometric image reconstruction.
- We show that data-driven priors in interferometric datasets can be

* Qiong Luo is the corresponding author. Her other affiliation is the Hong Kong University of Science and Technology.

effectively learned by incorporating both visibility and image data.

We have conducted experiments to evaluate VIC-DDPM in comparison with both popular traditional methods and a few recent state-of-the-art deep-learning based models. Our results show that our method significantly improves the output images. In particular, VIC-DDPM removes artifacts more thoroughly, reconstructs structural details better, and recovers surrounding dim sources. This advancement facilitates a more comprehensive and accurate analysis of radio astronomical data, enabling researchers to better understand complex celestial phenomena.

The remainder of this paper is organized as follows. In section 2, we describe the background of radio interferometric imaging and image reconstruction, related reconstruction algorithms, and basic denoising diffusion probabilistic models. In section 3, we present the details of VIC-DDPM, including the two conditioning factors – dirty image and visibility. In section 4, we describe experimental settings and present the results with analysis.

2 Background and Related Work

2.1 Interferometric Imaging

In radio astronomy, observational data are gathered from radio telescopes, or *interferometers*, which detect radio signals from remote celestial objects. These interferometers utilize aperture synthesis [4] to process their captured signals, generating *visibility* data through Fourier transforms of the sky's brightness distribution [16]. Then, the *imaging* process constructs images of the observed sky from the visibility data for subsequent analysis [31] .

Specifically, as the complex value of the visibility of an astronomical object is equivalent to the Fourier transform of the object's intensity distribution function, we can apply an inverse Fourier transform to convert the (u, v) coordinates in the Fourier domain into (l, m) in the image domain [33]:

$$I(l,m) \quad = \quad \int_u \int_v e^{2\pi i(ul+vm)} V(u,v) du dv \qquad (1)$$

where $V(u, v)$ denotes the complex visibility in the Fourier space, and $I(l, m)$ denotes the intensity distribution in the image.

2.2 Interferometric Image Reconstruction

Due to the limited number of radio telescopes and their non-uniform distribution on the ground, visibility data is often sparsely and irregularly sampled, leading to an incomplete representation of the true sky. When the inverse Fourier transform is applied to this sparse data, the resulting image, called a *dirty image*, is dominated by artifacts [23]. For better applicability in scientific analysis, dirty images must go through further reconstruction to enhance the clarity of sources and reduce artifacts.

Interferometric image reconstruction is an ill-posed problem because there are infinite solution images that fit the observed visibility [28]. Traditional reconstruction methods [1, 14] utilize human prior knowledge to reduce the solution space. In particular, the maximum entropy method (MEM) [1] mathematically determines the solution that best agrees with the available sparse measurement, whereas the CLEAN algorithm [14] iteratively operates on the dirty image to distinguish real structures from sidelobe disturbances. For the past decades, CLEAN [14] has served as the main-stream radio interferometric reconstruction method; however, its resulting images have limitations [33] due to its assumption of all point-like sources in the sky [7].

Recently, a flurry of deep-learning based data processing methods have been proposed in radio astronomy and computational imaging. Morningstar et al. [18, 19] used neural networks in radio interferometry and can generate reproducible results [23]. Sun et al. presented a variational depth probability imaging method for quantifying reconstruction uncertainty, which can efficiently sample posteriori distributions [28]. Schmidt et al. [23] used radionets with a convolutional neural network structure to generate reproducible clean radio images on simulated data. Wu et al. [33] proposed to perform sparse-to-dense inpainting in the spectral domain with coordinate-based neural fields, outperforming other established interferometry techniques.

We experimented several of these state-of-the art deep learning based methods [33, 23, 22] on galaxy images each of which contains a number of astronomical sources, and found that these methods reconstructed the main structure of the most obvious object in the image. However, some details and surrounding dim sources were missing, and considerable amounts of noise remained in the reconstructed image. Therefore, we turn to diffusion models, which could potentially address these limitations, since they are able to generate diverse high-resolution images with high semantic fidelity of conditioning [29]. To the best of our knowledge, no prior studies have used diffusion models to reconstruct dirty images in radio interferometry.

2.3 DDPM

Denoising Diffusion Probabilistic Models (DDPMs) are a class of deep generative models that learn to generate realistic samples from a given data distribution. The core idea behind DDPMs, introduced by Sohl-Dickstein et al. [26] and further developed by Ho et al. [13], is to reverse a denoising process that transforms the data distribution into Gaussian noise. The diffusion process can be thought of as a series of noise-corrupted versions of the original data, where each step progressively adds more noise to the data until the result becomes Gaussian noise [26]. The model then learns to denoise the data by predicting the preceding, less noisy version of the data at each step, given the current noisy version. To generate new samples, the model starts from Gaussian noise and simulates the reverse diffusion process by iteratively denoising the data at each step [13]. As the model has learned to denoise data at every step of the process, the generated samples become progressively more realistic and structured.

Diffusion models show great generative capabilities to produce fine levels of details [8]. Current diffusion architectures are well designed to work with image data in the spatial domain. When their underlying neural backbone is implemented as a U-Net, the generative capacity of these models shifts from the natural fitting of inductive biases to image-like data [10, 13, 22, 27, 21]. Latent diffusion models further introduce cross-attention layers into the U-Net model architecture, making diffusion models more powerful and flexible with general conditioning input [29]. In our work, we propose to apply cross-attention with visibility data and dirty image for radio interferometric image reconstruction.

3 Our Method

In this section, we present our Visibility and Dirty Image Conditional DDPM (VIC-DDPM) for radio interferometric image reconstruction. We first give an overview of our method and formulate the VIC-DDPM. Then we present the details of visibility and dirty image conditioning, respectively.

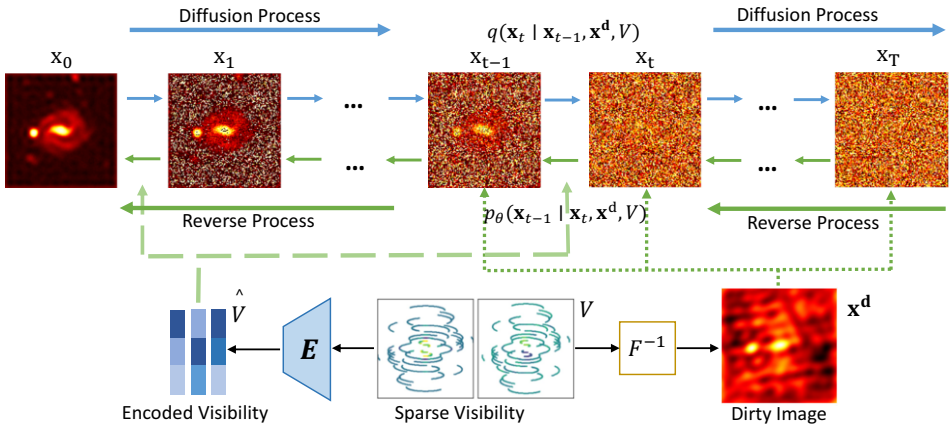

Figure 2. Overview of the Proposed VIC-DDPM. The diffusion process starts from $\mathbf{x_0}$ and Gaussian noise is gradually added to $\mathbf{x_{t-1}}$ to obtain $\mathbf{x_t}$ until $t = T$. The reverse process starts from random noise and generates $\mathbf{x_{t-1}}$ from $\mathbf{x_t}$. Visibility data and dirty images are added to the reverse process to condition the clean image generation. Details are described in Section 3.1.

3.1 Formulation of VIC-DDPM

Suppose $\mathbf{x} \in \mathbb{R}^n$ is an image of real sky's brightness distribution and $V \in \mathbb{R}^m$ is the sampled visibility. They follow the forward model:

$$V = \mathbf{P}_\Omega F(\mathbf{x}) + \epsilon' \tag{2}$$

where F refers to the Fourier transformation, $\mathbf{P}_\Omega \in \mathbb{R}^{m \times n}$ is an under-sampling matrix with Ω denoting the sample pattern produced by the sampling function of the telescope, and ϵ' denotes noise.

The task of radio interferometric image reconstruction can be simplified to estimating the posterior distribution $q(\mathbf{x} \mid V)$. Since $\mathbf{x}^{\mathbf{d}} = F^{-1}(V)$, the task can then be converted to estimate $q(\mathbf{x} \mid \mathbf{x}^{\mathbf{d}}, V)$, leveraging the complementary advantages of visibility observations and dirty images. Then we define our diffusion model in the following form:

$$p_\theta\left(\mathbf{x}_0 \mid \mathbf{x}^{\mathbf{d}}, V\right) := \int p_\theta\left(\mathbf{x}_{0:T} \mid \mathbf{x}^{\mathbf{d}}, V\right) \mathrm{d}\mathbf{x}_{1:T}. \tag{3}$$

The reverse process $p_\theta\left(\mathbf{x}_{0:T} \mid \mathbf{x}^{\mathbf{d}}, V\right)$ is defined as:

$$p_\theta\left(\mathbf{x}_{0:T} \mid \mathbf{x}^{\mathbf{d}}, V\right) = p\left(\mathbf{x}_T\right) \prod_{t=1}^{T} p_\theta\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}^{\mathbf{d}}, V\right), \tag{4}$$

$$p_\theta\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}^{\mathbf{d}}, V\right) = \mathcal{N}\left(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right), \sigma_t^2 \mathbf{I}\right). \tag{5}$$

Similar to the original DDPM [13], we fix the diffusion process to a Markov chain, defined as:

$$q\left(\mathbf{x}_{1:T} \mid \mathbf{x}_0, \mathbf{x}^{\mathbf{d}}, V\right) := \prod_{t=1}^{T} q\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}^{\mathbf{d}}, V\right), \tag{6}$$

$$q\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}^{\mathbf{d}}, V\right) := \mathcal{N}\left(\mathbf{x}_t; \alpha_t \mathbf{x}_{t-1}, \beta_t^2 \mathbf{I}\right). \tag{7}$$

The diffusion process gradually adds Gaussian noise to data based on scheduled $\alpha_1, ..., \alpha_t$ and $\beta_1, ..., \beta_t$.

Different from the original DDPM, we follow most recent work [10, 34] to set the α_t and β_t schedule.

Let $\bar{\alpha}_t = \prod_{i=i}^{t} \alpha_i, \bar{\beta}_t^2 = \sum_{i=1}^{t} \frac{\bar{\alpha}_t^2}{\bar{\alpha}_i^2} \beta_i^2, \bar{\alpha}_0 = 1, \bar{\beta}_0 = 0$:

$$q\left(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}^{\mathbf{d}}, V\right) = \mathcal{N}\left(\mathbf{x}_t; \bar{\alpha}_t \mathbf{x}_0, \bar{\beta}_t^2 \mathbf{I}\right), \tag{8}$$

$$q\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0, \mathbf{x}^{\mathbf{d}}, V\right) = \mathcal{N}\left(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}_t\left(\mathbf{x}_t, \mathbf{x}_0\right), \tilde{\beta}_t^2 \mathbf{I}\right), \tag{9}$$

where

$$\tilde{\boldsymbol{\mu}}_t\left(\mathbf{x}_t, \mathbf{x}_0\right) = \frac{\alpha_t \bar{\beta}_{t-1}^2}{\bar{\beta}_t^2} \mathbf{x}_t + \frac{\bar{\alpha}_{t-1} \beta_t^2}{\bar{\beta}_t^2} \mathbf{x}_0, \text{and } \tilde{\beta}_t = \frac{\beta_t \bar{\beta}_{t-1}}{\bar{\beta}_t}. \tag{10}$$

With our schedule setting of α_t, we have $\bar{\alpha}_T \approx 0$.

Then, we perform the training of $p_\theta\left(\mathbf{x}_0 \mid \mathbf{x}_{t-1}, \mathbf{x}^{\mathbf{d}}, V\right)$ by optimizing the variational bound on the negative log likelihood, similar to DDPM [13]:

$$\begin{aligned}
&\mathbb{E}\left[-\log p_\theta\left(\mathbf{x}_0 \mid \mathbf{x}^{\mathbf{d}}, V\right)\right] \\
&\leq \mathbb{E}_q\left[-\log \frac{p_\theta\left(\mathbf{x}_{0:T} \mid \mathbf{x}^{\mathbf{d}}, V\right)}{q\left(\mathbf{x}_{1:T} \mid \mathbf{x}_0, \mathbf{x}^{\mathbf{d}}, V\right)}\right] \\
&= \mathbb{E}_q\Bigg[-\log p\left(\mathbf{x}_T \mid \mathbf{x}^{\mathbf{d}}, V\right) \\
&\qquad - \sum_{t \geq 1} \log \frac{p_\theta\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}^{\mathbf{d}}, V\right)}{q\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}^{\mathbf{d}}, V\right)}\Bigg] \\
&=: L.
\end{aligned} \tag{11}$$

We choose the parameterization:

$$\boldsymbol{\mu}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right) = \frac{1}{\alpha_t}\left(\mathbf{x}_t - \frac{\beta_t^2}{\bar{\beta}_t} \boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right)\right) \tag{12}$$

where $\boldsymbol{\epsilon}_\theta$ is a function approximator for predicting $\boldsymbol{\epsilon}$ from $\mathbf{x}_t$. If $t = 0$, we set $\mathbf{z}_t = 0$; otherwise, we sample $\mathbf{x} \sim p_\theta\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}^{\mathbf{d}}, V\right)$ by computing $\mathbf{x}_{t-1} = \frac{1}{\alpha_t}\left(\mathbf{x}_t - \frac{\beta_t^2}{\bar{\beta}_t} \boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right)\right) + \sigma_t \mathbf{z}_t$, where $\mathbf{z}_t \sim \mathcal{N}(0, \mathbf{I})$.

Then, based on (Eq. 9), and with the parameterization (Eq. 12), we simplify L to:

$$L_{simple}(\theta) = \mathbb{E}_{\mathbf{x}_0, t, \boldsymbol{\epsilon}}\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\bar{\alpha}_t \mathbf{x}_0 + \bar{\beta}_t \boldsymbol{\epsilon}, t, \mathbf{x}^{\mathbf{d}}, V\right)\right\|_2^2 \tag{13}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $t \sim \text{Uniform}(0, ..., T)$.

In summary, Figure 2 illustrates the diffusion process and reverse process in VIC-DDPM. Algorithm 1 presents the diffusion process using the loss function (Eq. 13). Algorithm 2 shows the sampling procedure.

Algorithm 1 Training

repeat
  $\{\mathbf{x}\} \sim q(\mathbf{x})$, obtain $V, \mathbf{x}^{\mathbf{d}} = F^{-1}(V)$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$
  $t \sim \text{Uniform}(0, ..., T)$
  $\mathbf{x}_t = \bar{\alpha}_t \mathbf{x}_0 + \bar{\beta}_t \boldsymbol{\epsilon}$
  Compute the loss $L(\theta) = \|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right)\|_2^2$
  Update θ with gradient descent on $L(\theta)$
until convergence

Algorithm 2 Sampling

Given $V, \mathbf{x}^{\mathbf{d}} = F^{-1}(V)$
for $t = T, ..., 1,$ **do**
  $\mathbf{z}_t \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $\mathbf{z}_t = 0$
  $\mathbf{x}_{t-1} = \frac{1}{\alpha_t}\left(\mathbf{x}_t - \frac{\beta_t^2}{\bar{\beta}_t}\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right)\right) + \sigma_t \mathbf{z}_t$
return $\mathbf{x}_0$

3.2 Details of Conditioning

We use the U-Net [22] architecture with ResNet [11] blocks and attention layers [32] as our backbone neural network to represent $\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right)$. Figure 3 gives an overview of the network.

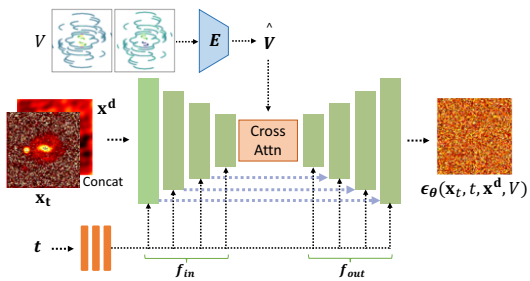

Figure 3. The U-Net architecture in our method to represent $\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right)$.

3.2.1 Dirty Image Conditioning

As shown on the left in Figure 3, we concatenate the dirty image x_d with the input x_t of the U-Net, to leverage the information provided by the dirty image. The concatenation is done by each channel of the image. This way, we enable the U-Net to extract features from both the input data and the dirty image, thereby incorporating prior information from the dirty image into the reconstruction process.

3.2.2 Sparse Visibility Encoding

Visibility data is sparse, so we apply encoding to increase the information density and then incorporate the encoded visibility $\hat{V}$ into our network, as shown in the top left of Figure 3.

Figure 4 illustrates our encoding method. As shown on the left in Figure 4, the sparsely sampled visibility data is in the form of $\{u_s, v_s, V(u_s, v_s)\}$, where (u_s, v_s) are the coordinates at which a measurement is sampled and $V(u_s, v_s)$ is the complex value of the sample.

To integrate each visibility value and its corresponding position, we first encode each sample individually using positional embedding (PE(u_s, v_s) in Figure 4). Specifically, we first encode the positional information of a sample point using a NeRF-style [17] embedding method (axis-aligned, power-of-two frequency sinusoidal encoding) [33]. It facilitates the following Multi-Layer-Perception (MLP) to learn high frequency information [30, 33]. After that, the positional embedding PE(u_s, v_s) and the complex value of visibility $V(u_s, v_s)$ are concatenated to form visibility tokens V'.

Denote the sparse visibility tokens as $V' = [v_1'; v_2'; v_3'; \dots v_n']$, and $V' \in \mathbb{R}^{n \times d}$, where n is the number of sampled measurement points and d is the number of dimensions of visibility tokens. Furthermore, the sampled points are divided into m groups by their rotation angles. Then, we apply a linear mapping layer M and a mean pooling for encoding, as shown on the right of Figure 4. The encoding result is $\hat{V} = [\hat{v}_1; \hat{v}_2; \hat{v}_3; \dots \hat{v}_m]$, where $\hat{V} \in \mathbb{R}^{m \times d}$:

$$\hat{v}_i = \frac{m}{n} \sum_{j \in Group_i} M(v_j),\ i = 1 \text{ to } m. \tag{14}$$

This encoded visibility is then used as the input of the cross-attention module in the U-Net.

3.2.3 Cross-Attention Fusion

Next, we employ a cross-attention mechanism [32, 21] to fuse the encoded visibility information into the U-Net. Cross-attention can effectively learn relationships between different data domains. Specifically, we use the concatenation of both image features and visibility features to form *Key* and *Value*:

$$CrossAttention = \text{softmax}\left(\frac{Query \cdot Key^T}{\sqrt{d}}\right) \cdot Value, \tag{15}$$

where

$$Query = W_{Query} \cdot \varphi\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}\right), \tag{16}$$

$$Key = W_{Key} \cdot \text{Concat}(\varphi\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}\right), \text{E}(V)), \tag{17}$$

$$Value = W_{Value} \cdot \text{Concat}(\varphi\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}\right), \text{E}(V)). \tag{18}$$

Here, $\varphi\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}\right)$ denotes the intermediate latent representation of the U-Net that implements the predictor ϵ_θ. $W_{query}, W_{Key}, W_{Value}$ are learnable weights. E(V) refers to the encoded visibility tokens.

The aim of this cross-attention mechanism is to weigh the contributions of features from both the dirty image and the sparse visibility, allowing the U-Net to selectively attend to the most relevant information when generating the reconstructed image.

In summary, we design the $\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right)$ as follows: $\boldsymbol{\epsilon}_\theta\left(\mathbf{x}_t, t, \mathbf{x}^{\mathbf{d}}, V\right) = f_{out}(\text{Cross-Attn}(f_{in}(g(\mathbf{x}_t, \mathbf{x}^{\mathbf{d}}), t), \text{E}(V)), t)$

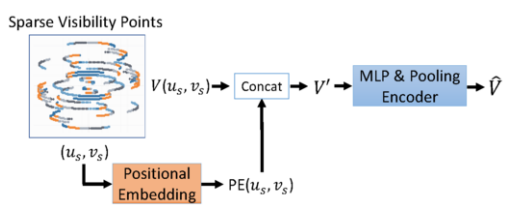

Figure 4. Sparse Visibility Encoding.

where f_{in} refers to the input blocks , and f_{out} the output blocks after the cross-attention module in the U-Net structure, as illustrated in Figure 3. $\mathrm{E}(V)$ refers to the encoded visibility tokens, and $g(\cdot,\cdot)$ the concatenation of images by channel.

4 Experiments

In this section, we perform a comprehensive evaluation on our method in comparison with several classic and recent state-of-the-art methods to demonstrate the overall improvement achieved by our approach. We also conduct ablation experiments to study the effects of individual techniques in our method. Additionally, we vary the number of sample noise used to obtain a result clean image, the number of diffusion steps to perform towards each noise image, and the total number of training images to examine our model performance.

4.1 Experimental Setup

Platform. We conduct all experiments on a server with two AMD EPYC 7302 CPUs, 128GB main memory, and eight Nvidia RTX 3090 GPUs each with 24GB device memory. The server is equipped with an NVME 2TB SSD and four 1TB SATA hard disks. The operating system is Ubuntu 20.04. Our model is implemented in PyTorch 1.8.1 [20].

Datasets. We use the Galaxy10 DECals dataset [12] in our experiments, as in the most recent work on astronomical interferometric image reconstruction [33]. This dataset contains a total of 17,736 galaxy images, derived from the DESI Legacy Imaging Surveys [9], which combine data from the Beijing-Arizona Sky Survey (BASS) [35], the DECam Legacy Survey (DECaLS) [2], and the Mayall z-band Legacy Survey [24]. In our experiments, all images are of a size 128×128 in pixels. With these images as the ground truth, we then use the eht-imaging toolkit [5, 6] to generate visibility data in the form of $\{u_s, v_s, V(u_s, v_s)\}$, with the observation parameters set to match an 8-telescope Event Horizon Telescope (EHT) array [33]. The number of sampled visibility points on each image is 1660. Also following Wu et al. [33], we use the discrete Fourier transform (DFT) method to generate dirty images from the visibility data. Finally, we randomly split the 17,736 images into a training set of 17,224 images and a testing set of 512 images following Wu et al. [33].

Implementation Details. Our experimental model architectures are based on U-Net [22] and augmented with timestep embedding and channel-wise self-attention block following DDPM [13] and Guided-Diffusion [10]. In visibility encoding, the number of elements in a positional embedding vector is set to 22 and the number of points in each point group is 20. In addition, we add spatial-wise attention layers and cross-attention layers in U-Net. Finally, the training batch size in VIC-DDPM is set to 32 images, and the total number of training steps is 25k.

Methods under Comparison. We compare VIC-DDPM with three other methods for radio interferometric image reconstruction, including the classic method CLEAN [14] and recent deep learning-based approaches radionets [23] and Transformer-Conditioned Neural Fields [33]. We use the original code of each of these methods and follow the parameter setting in the original code for the best performance. In addition, we test our U-Net in both visibility domain only, U-Net(Vis), and image domain only, U-Net(Img), as supplementary baselines.

Evaluation Metrics. We use two common metrics to evaluate image quality: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). They quantify the overall image quality of reconstructed images and the perceptual similarity to the ground truth images, respectively. We compute these two metrics using the scikit-image package [25], which follow the formulas presented by Hore et al. [15].

4.2 Overall Comparison

First, we present in Figure 5 five representative reconstructed images on the Galaxy10(DECals) dataset from all methods under comparison as well as the dirty images and the ground truth images. Comparing the dirty images in Figure 5 (a) and the ground truth in Figure 5 (h), we see that dirty images contain many artifacts and that actual objects are often distorted.

As shown in Figure 5 (b), the CLEAN [14] method reduces most artifacts, but misses the original structures of the sources. Due to the assumptions in the CLEAN [14] algorithm, the emission distribution cannot be smoothly reconstructed. In comparison, the U-Net approach in the visibility domain (Figure 5 (c)) leaves a lot of artifacts. The spatial domain U-Net (Figure 5 (d)) achieves relatively better reconstruction, but does not fix the distortion and overlooks some faint sources or generates non-existent faint sources. Radionets [23] (Figure 5 (e)) performs well in denoising and recovering from distortion; however, it has problems distinguishing separate sources that are close to one another. The Transformer-Conditioned Neural Fields method (denoted TransNF) [33] (Figure 5 (f)) faithfully recovers the main objects in the observation area except when the edges of the objects are blurry or when the surrounding objects are small and dim.

The results of VIC-DDPM (Figure 5 (g)) show that our method effectively removes artifacts and reconstructs details of the true objects, including relatively small or faint sources and detailed structures. These results demonstrate the effectiveness of our DDPM conditioned with both dirty images and visibility data.

Next, we compute the PSNR and SSIM scores, with mean and standard deviation, for all 512 reconstructed images produced by our method and the methods under comparison. The results are shown in Table 1. The greater the PSNR and SSIM scores, the better.

The experimental results demonstrate that our method consistently achieves better performance in both PSNR and SSIM, indicating the effectiveness of our approach in providing accurate reconstructions.

Table 1. Performance of Different Methods.

Models	Domain		Metrics	
	Img	Vis	PSNR↑	SSIM↑
Dirty Image			10.406 ± 1.101	0.6321 ± 0.0590
CLEAN	✓		17.366 ± 2.406	0.7130 ± 0.0405
U-Net (Img)	✓		21.414 ± 1.654	0.8028 ± 0.0194
U-Net (Vis)		✓	13.015 ± 1.103	0.7079 ± 0.0230
Radionets		✓	21.433 ± 2.105	0.7940 ± 0.0237
Transformer_NF		✓	20.707 ± 2.970	0.8084 ± 0.0404
VIC-DDPM	✓	✓	$\mathbf{22.615 \pm 1.531}$	$\mathbf{0.8242 \pm 0.0224}$

4.3 Ablation Experiments

In the ablation experiments, we test the following variants of our method by selectively removing one or more components to inves-

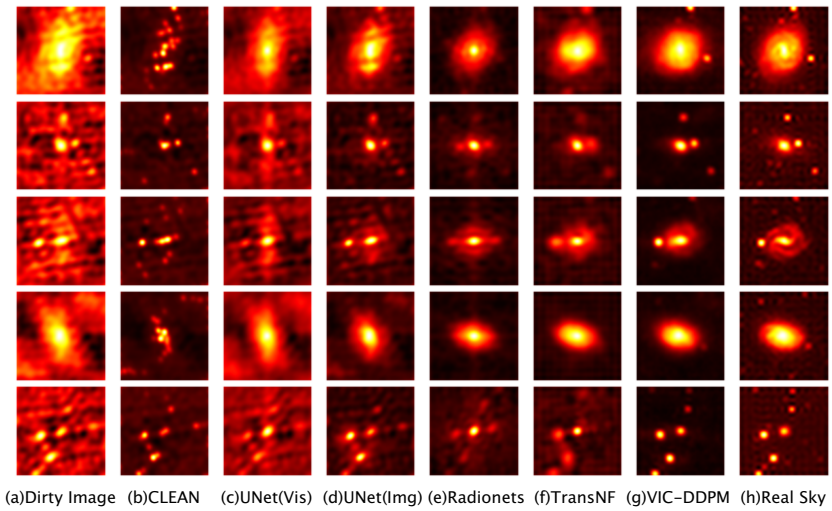

Figure 5. Visual Results of Overall Comparison.

tigate the individual contributions of the components to the performance: (1) without any conditioning; (2) without the visibility data conditioning; and (3) without the dirty image conditioning. We train these variants by add no conditioning, replacing the visibility data with zeros, and by replacing the dirty image with zeros, respectively. We compare these three variants with the original dirty images and the reconstructed images from the full VIC-DDPM.

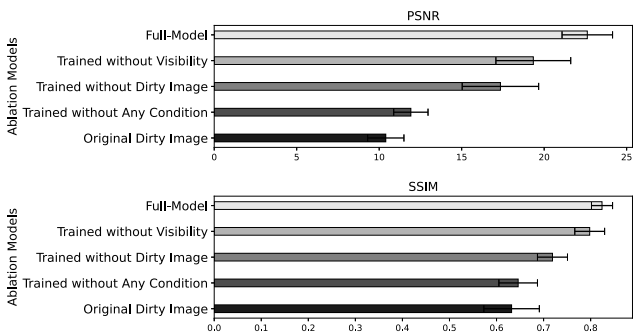

Figure 6. Ablation Results.

We evaluate the performance of each ablation model using the same dataset and evaluation metrics (PSNR and SSIM) as in the previous experiment. The results in Figure 6 show that our full VIC-DDPM model consistently achieves the best performance, demonstrating the effectiveness of combining dirty image conditioning and visibility conditioning. The dirty image conditioning seems more effective than the visibility data conditioning, suggesting that spatial information plays a more important role than signal-to-noise separation in DDPM. Some visual examples of ablation results are shown in Figure 7. These results confirm that the full model (column e) is best, and that dirty image conditioning (column d) is more effective

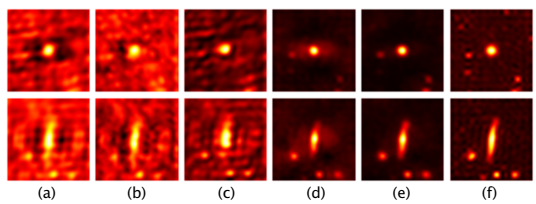

Figure 7. Visual Examples of Ablation Results. Columns from let to right are (a) original dirty image, (b) unconditional DDPM, (c) our model trained without dirty image conditioning, (d) our model trained without visibility conditioning, (e) full model of VIC-DDPM, and (f) real sky, respectively.

than visibility conditioning (column c). Without any conditioning, column (b) is the least effective.

4.4 *Effect of Parameter Setting*

We also conduct experiments to investigate the effect of two important parameters in our model – number of random noise sample images to test for each dirty image and number of diffusion steps. We report the performance on PSNR and SSIM in Figure 8. Figure 8(a) shows that as the number of random noise sample images used increases for each test dirty image, the mean quality of the result images also increases; however, when the number of sample images further increases, the mean quality remains stable or slightly decreases. In comparison, Figure 8(b) shows that the result quality is almost constant regardless of number of diffusion steps. These results indicate that in our experimental setting, selecting 5 samples for each test dirty image and running each test for 1000 steps would be a good choice for best result quality.

4.5 *Effect of Training Data Size*

Model performance may vary with the training dataset size. We therefore use different sizes of training datasets to evaluate our model

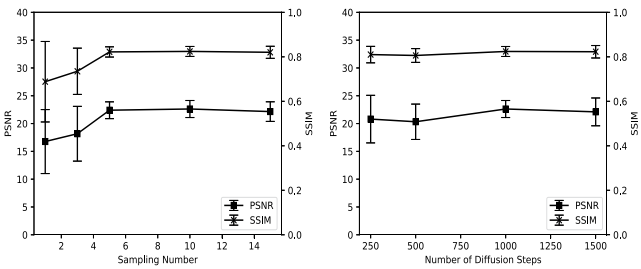

Figure 8. Effect of Parameters: (a) number of random noise images to test for each dirty image; (b) number of diffusion steps.

Table 2. Performance (PSNR/SSIM) with training dataset size varied

Data	Methods		
	RadioNet	Transformer_NF	VIC-DDPM
0.5k	19.401 ± 1.639 0.7586 ± 0.0210	18.235 ± 2.802 0.7610 ± 0.0634	17.149 ± 6.856 0.7417 ± 0.1220
1k	20.267 ± 1.822 0.770 ± 0.0222	18.935 ± 2.900 0.7781 ± 0.0675	22.420 ± 3.765 0.8236 ± 0.0354
2k	20.610 ± 1.926 0.7765 ± 0.0241	18.565 ± 2.360 0.7735 ± 0.0578	22.170 ± 1.986 0.8137 ± 0.0211
4k	20.657 ± 1.996 0.7787 ± 0.0269	20.069 ± 2.611 0.8028 ± 0.0518	22.403 ± 2.172 0.8210 ± 0.0328
17k	21.433 ± 2.105 0.7940 ± 0.023	20.707 ± 2.970 0.8084 ± 0.040	22.615 ± 1.531 0.8242 ± 0.022

performance in PSNR/SSIM in comparison with two latest deep-learning methods – Radionets [23] and Transformer-Conditioned Neural Fields (denoted Transformer_NF) [33]. To create training sets of different sizes, we randomly choose 500, 1000, 2000, and 4000 of the 17k training images in the Galaxy10 DECals dataset [12]. As shown in Table 2, when the dataset size increases from 0.5k to 17k, the PSNR/SSIM performance increases by 10.5%/4.7%, 13.6%/6.2% and 31.9%/11.1% on three models, respectively. Compared with the other two methods, the performance of VIC-DDPM significantly improves when the dataset size increases from 500 to 1000. With a training set of 1000 images, VIC-DDPM performance is already considerably higher than the other two methods, and further increase of training dataset size shows little effect. In contrast, both radionets and Transformer_NF continue to improve as the training dataset size increases, but the performance improvement is not as significant as VIC-DDPM. As a result, VIC-DDPM outperforms the other two on data sizes from 1K to 17K. This result suggests that VIC-DDPM requires a much smaller training set than the other two to obtain high-quality reconstruction. We plan to find larger training datasets to give more room to the other two methods for performance improvement.

To further investigate the performance differences, we examine the visual effects of reconstructed images and provide a few examples with different training dataset sizes in Figure 9. Subfigures on each row are from a single method, and those on each column are on the same training dataset size. We find that, even though our model is slightly lower than the other two methods on the numeric performance metrics when the training set size is 500 in Table 2, our method's reconstruction of the structures, in the bottom left of each

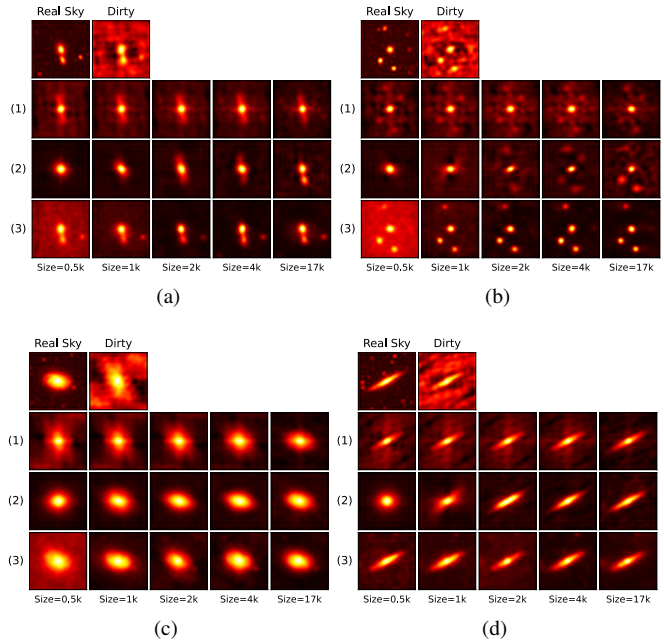

Figure 9. Visual results on four example images (a)-(d) from three models including (1) radionets, (2) Transformer-Conditioned Neural Fields, and (3) VIC-DDPM trained with datasets of different sizes.

example (a) through (d) is noticeably superior to the other two methods when visually inspected by the human eye, as shown in Figure 9. These visual results suggest that the lower metric values of our method at 500 training images are possibly because the background noise has not been effectively reduced by then.

As observed in Figure 9 (a) , (c) and (d), VIC-DDPM is capable of reconstructing the basic contours of astronomical sources after learning only 500 samples. In contrast, other methods require learning from 1000 to 4000 samples to reconstruct the basic contours. On reconstruction of images containing multiple astronomical sources, as shown in Figure 9 (a) and (b), our method can discern the positions and structures of multiple astronomical sources after learning from 500 training samples, while other methods cannot.

5 Conclusion and Future Work

In conclusion, we have presented VIC-DDPM, a Visibility and Image Conditioned Denoising Diffusion Probabilistic Model, for radio interferometric image reconstruction. Inheriting the capability of fine detail generation from DDPM as well as incorporating the complementary strengths of data in spectral and spatial domains, VIC-DDPM improves the accuracy of reconstructed images with effective removal of artifacts, and recovery of structural details and dim sources. Experimental results demonstrate the superiority of VIC-DDPM over both traditional and deep-learning based approaches on the Galaxy10 DECals dataset.

In the future, we will apply our methods on both real and synthetic datasets based on different kinds of radio telescope arrays and generalize our methods to images in other domains.

Acknowledgements

This work was partially supported by Grant 2020SKA0110300 from the National SKA Program of China.

References

[1] JG Ables, 'Maximum entropy spectral analysis', *Astronomy and Astrophysics Supplement, Vol. 15, p. 383*, **15**, 383, (1974).

[2] Robert D Blum, Kaylan Burleigh, Arjun Dey, David J Schlegel, Aaron M Meisner, Michael Levi, Adam D Myers, Dustin Lang, John Moustakas, Anna Patej, et al., 'The decam legacy survey', in *American Astronomical Society Meeting Abstracts# 228*, volume 228, pp. 317–01, (2016).

[3] Alexandre Boulch and Renaud Marlet, 'Deep learning for robust normal estimation in unstructured point clouds', in *Computer Graphics Forum*, volume 35, pp. 281–290. Wiley Online Library, (2016).

[4] Bernard F Burke, Francis Graham-Smith, and Peter N Wilkinson, *An introduction to radio astronomy*, Cambridge University Press, 2019.

[5] Andrew A Chael, Katherine L Bouman, Michael D Johnson, Ramesh Narayan, Sheperd S Doeleman, John FC Wardle, Lindy L Blackburn, Kazunori Akiyama, Maciek Wielgus, Chi-kwan Chan, et al., 'ehtim: Imaging, analysis, and simulation software for radio interferometry', *Astrophysics Source Code Library*, ascl–1904, (2019).

[6] Andrew A Chael, Michael D Johnson, Katherine L Bouman, Lindy L Blackburn, Kazunori Akiyama, and Ramesh Narayan, 'Interferometric imaging directly with closure phases and closure amplitudes', *The Astrophysical Journal*, **857**(1), 23, (2018).

[7] Liam Connor, Katherine L Bouman, Vikram Ravi, and Gregg Hallinan, 'Deep radio-interferometric imaging with polish: Dsa-2000 and weak lensing', *Monthly Notices of the Royal Astronomical Society*, **514**(2), 2614–2626, (2022).

[8] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah, 'Diffusion models in vision: A survey', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2023).

[9] Arjun Dey, David J Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, et al., 'Overview of the desi legacy imaging surveys', *The Astronomical Journal*, **157**(5), 168, (2019).

[10] Prafulla Dhariwal and Alexander Nichol, 'Diffusion models beat gans on image synthesis', *Advances in Neural Information Processing Systems*, **34**, 8780–8794, (2021).

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, 'Deep residual learning for image recognition', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, (2016).

[12] Leung Henry. Galaxy10 decals dataset. https://github.com/henrysky/Galaxy10, 2021.

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel, 'Denoising diffusion probabilistic models', *Advances in Neural Information Processing Systems*, **33**, 6840–6851, (2020).

[14] JA Högbom, 'Aperture synthesis with a non-regular distribution of interferometer baselines', *Astronomy and Astrophysics Supplement Series*, **15**, 417, (1974).

[15] Alain Hore and Djemel Ziou, 'Image quality metrics: Psnr vs. ssim', in *2010 20th international conference on pattern recognition*, pp. 2366–2369. IEEE, (2010).

[16] Honghao Liu, Qiong Luo, and Feng Wang, 'Efficient radio interferometric imaging on the gpu', in *2022 IEEE 18th International Conference on e-Science (e-Science)*, pp. 95–104. IEEE, (2022).

[17] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng, 'Nerf: Representing scenes as neural radiance fields for view synthesis', *Communications of the ACM*, **65**(1), 99–106, (2021).

[18] Warren R Morningstar, Yashar D Hezaveh, Laurence Perreault Levasseur, Roger D Blandford, Philip J Marshall, Patrick Putzky, and Risa H Wechsler, 'Analyzing interferometric observations of strong gravitational lenses with recurrent and convolutional neural networks', *arXiv preprint arXiv:1808.00011*, (2018).

[19] Warren R Morningstar, Laurence Perreault Levasseur, Yashar D Hezaveh, Roger Blandford, Phil Marshall, Patrick Putzky, Thomas D Rueter, Risa Wechsler, and Max Welling, 'Data-driven reconstruction of gravitationally lensed galaxies using recurrent inference machines', *The Astrophysical Journal*, **883**(1), 14, (2019).

[20] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al., 'Pytorch: An imperative style, high-performance deep learning library', *Advances in neural information processing systems*, **32**, (2019).

[21] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer, 'High-resolution image synthesis with latent diffusion models', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, (2022).

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, 'U-net: Convolutional networks for biomedical image segmentation', in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241. Springer, (2015).

[23] Kevin Schmidt, Felix Geyer, Stefan Fröse, Paul-Simon Blomenkamp, Marcus Brüggen, Francesco de Gasperin, Dominik Elsässer, and Wolfgang Rhode, 'Deep learning-based imaging in radio interferometry', *arXiv preprint arXiv:2203.11757*, (2022).

[24] David R Silva, Robert D Blum, Lori Allen, Arjun Dey, David J Schlegel, Dustin Lang, John Moustakas, Aaron M Meisner, Francisco Valdes, Anna Patej, et al., 'The mayall z-band legacy survey', in *American Astronomical Society Meeting Abstracts# 228*, volume 228, pp. 317–02, (2016).

[25] Himanshu Singh and Himanshu Singh, 'Basics of python and scikit image', *Practical Machine Learning and Image Processing: For Facial Recognition, Object Detection, and Pattern Recognition Using Python*, 29–61, (2019).

[26] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli, 'Deep unsupervised learning using nonequilibrium thermodynamics', in *International Conference on Machine Learning*, pp. 2256–2265. PMLR, (2015).

[27] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, 'Score-based generative modeling through stochastic differential equations', *arXiv preprint arXiv:2011.13456*, (2020).

[28] He Sun and Katherine L Bouman, 'Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging', in *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 2628–2637, (2021).

[29] Yu Takagi and Shinji Nishimoto, 'High-resolution image reconstruction with latent diffusion models from human brain activity', *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (In Press)*, (2023).

[30] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng, 'Fourier features let networks learn high frequency functions in low dimensional domains', *Advances in Neural Information Processing Systems*, **33**, 7537–7547, (2020).

[31] A Richard Thompson, James M Moran, and George W Swenson, *Interferometry and synthesis in radio astronomy*, Springer Nature, 2017.

[32] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, 'Attention is all you need', *Advances in neural information processing systems*, **30**, (2017).

[33] Benjamin Wu, Chao Liu, Benjamin Eckart, and Jan Kautz, 'Neural interferometry: Image reconstruction from astronomical interferometers using transformer-conditioned neural fields', in *Proceedings of the AAAI Conference on Artificial Intelligence*, (2022).

[34] Yutong Xie and Quanzheng Li, 'Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction', in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pp. 655–664. Springer, (2022).

[35] Hu Zou, Xu Zhou, Xiaohui Fan, Tianmeng Zhang, Zhimin Zhou, Jundan Nie, Xiyan Peng, Ian McGreer, Linhua Jiang, Arjun Dey, et al., 'Project overview of the beijing–arizona sky survey', *Publications of the Astronomical Society of the Pacific*, **129**(976), 064101, (2017).