VisRec: A Semi-Supervised Approach to Visibility Data Reconstruction in Radio Astronomy

Ruoqi Wang[1], Haitao Wang[2], Qiong Luo[*1, 3], Feng Wang[4], Hejun Wu[2]

[1]The Hong Kong University of Science and Technology (Guangzhou)
[2]School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, China
[3]The Hong Kong University of Science and Technology
[4]Guangzhou University
rwang280@connect.hkust-gz.edu.cn, wanght39@mail2.sysu.edu.cn,
luo@ust.hk, fengwang@gzhu.edu.cn, wuhejun@mail.sysu.edu.cn

Abstract

Radio telescopes produce visibility data about celestial objects, but these data are sparse and noisy. As a result, images created on raw visibility data are of low quality. Recent studies have used deep learning models to reconstruct visibility data to get cleaner images. However, these methods rely on a substantial amount of labeled training data, which requires significant labeling effort from radio astronomers. Addressing this challenge, we propose VisRec, a model-agnostic semi-supervised learning approach to visibility data reconstruction in radio astronomy. Specifically, VisRec consists of both a supervised learning module and an unsupervised learning module. In the supervised learning module, we introduce a set of data augmentation functions to produce diverse visibility examples. In comparison, the unsupervised learning module in VisRec augments unlabeled data and uses reconstructions from non-augmented visibility as pseudo-labels for training. This hybrid approach allows VisRec to effectively leverage both labeled and unlabeled data. This way, VisRec performs well even when labeled data is scarce. Our evaluation results show that VisRec is applicable to various models, and outperforms all baseline methods in terms of reconstruction quality, robustness, and generalizability.

Code — https://github.com/RapidsAtHKUST/VisRec

Introduction

In radio astronomy, a radio interferometer consists of an array of antennas, all observing the same area of the sky, forming a single, unified telescope. The *visibility* refers to the signal obtained by cross-correlating pairs of antennas, which is in the frequency domain. Visibility data are then used to construct *images* of the celestial objects being observed (Dagli 2023). Due to the incompleteness and noise in the visibility data, the resultant images, known as *dirty images*, are often dominated by artifacts (Schmidt et al. 2022). As a result, radio interferometric data must be reconstructed before being used in scientific analysis. In this paper, we introduce a semi-supervised method to reconstruct the visibility.

*Qiong Luo is the corresponding author.

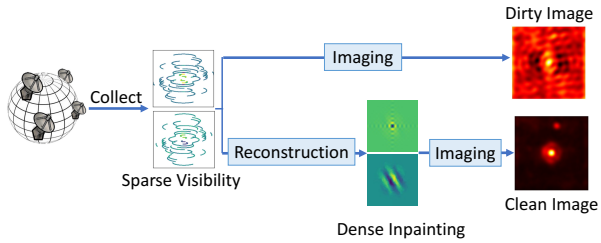


Figure 1: Illustration of radio interferometric data processing.

Traditional methods usually transfer the sparse visibility data into dirty images and then reconstruct the dirty images to *clean images* (Högbom 1974; Ables 1974; Bouman et al. 2016; Sun and Bouman 2021; Connor et al. 2022; Wang et al. 2023). In contrast, some recent deep-learning-based studies (Schmidt et al. 2022; Wu et al. 2022; Wang et al. 2024) have proposed to first do sparse-to-dense inpainting on the visibility data and then perform imaging to obtain the clean image, showing better performance. In this paper, we adopt the reconstruction-and-imaging processing flow and focus on the sparse-to-dense visibility reconstruction task. Figure 1 illustrates the workflow.

Existing visibility reconstruction research (Schmidt et al. 2022; Wu et al. 2022; Wang et al. 2024) use supervised learning methods in which the reconstruction error between the predicted results and the reference labels is minimized (Jia et al. 2022). However, these learning-based methods are highly dependent on a great amount of ground truth (labels) for model training, and obtaining ground truth data is difficult. In contrast to natural or medical image reconstruction, where long-time scanning can produce fully-sampled (labeled) training data, obtaining a fully-sampled sky measurement dataset via radio telescopes is impractical due to observational constraints. Furthermore, it takes expertise and human labor to get a ground truth clean image. As such, techniques that can leverage information from unlabeled datasets are indispensable where ground-truth images are scarce.

Another problem that often comes with the scarcity of labeled data is limited robustness and generalizability in

supervised-learning-based reconstruction methods, due to the lack of diversity in training data (Darestani, Chaudhari, and Heckel 2021; Desai et al. 2023). In particular, existing radio interferometry data reconstructing methods focus on enhancing reconstruction accuracy, and are sensitive to data distribution shift. We follow previous studies (Schmidt et al. 2022; Darestani, Chaudhari, and Heckel 2021) to generate test data with errors caused by noise perturbations during observation or offline antennas as well as data collected from different telescope configurations. Our results confirm the low robustness and generalizability of existing methods.

To address these problems, we propose a model-agnostic semi-supervised learning approach to radio interferometry data reconstruction, called VisRec. Specifically, VisRec consists of two modules: a supervised learning module and an unsupervised learning module. The supervised module integrates a variety of data augmentation functions, exposing the model to diverse training examples. The unsupervised module applies augmentations to unlabeled visibility data and employs reconstructions from non-augmented data as pseudo-labels. VisRec computes the consistency loss between the reconstruction of augmented visibilities and the pseudo-labels to lead the model to be more robust against perturbations. Our method effectively leverages both labeled and unlabeled data to improve model performance.

The main contributions of our work are: (1) We propose VisRec, the first model-agnostic semi-supervised framework for visibility sparse-to-dense reconstruction. VisRec reduces dependency on ground truth labels and successfully addresses the challenges of label scarcity in astronomical observation. (2) We empirically show that our method is robust to common observation perturbation, and has good generalizability when applied to data from different telescope configurations. (3) We introduce a variety of data augmentation methods for visibility data, including both label-invariant and label-variant augmentations. These augmentations improve the model performance when applied to VisRec as well as supervised learning methods.

Background and Related Work

Radio Interferometric Imaging

In radio astronomy, observing distant astronomical objects via radio interferometry demands large-aperture telescopes, as the angular resolution is inversely related to aperture size (Bouman et al. 2018). High angular resolution can be obtained using the technique of Very Long Baseline Interferometry (VLBI). VLBI uses a network of globally distributed radio telescopes to create a unified Earth-size telescope. Individual telescopes record radio waves from space, and these signals are then cross-correlated between antenna pairs to produce visibility data. The resulting visibility data is sparse due to the limited number of antennas in VLBI (Thompson, Moran, and Swenson 2017; Bouman et al. 2018).

Visibility data is in the form of complex values in the uv-plane, a geometric plane defined for interferometric observations. The imaging process is an inverse Fourier transformation, translating (u, v) coordinates in the frequency domain into (l, m) coordinates in the image domain (Wu et al. 2022):

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

Here, $V(u, v)$ denotes visibility data in the frequency domain, and $I(l, m)$ is the intensity distribution of the sky in the image domain.

Interferometric Data Reconstruction

The imaging results of the under-sampled sparse visibility data are dominated by artifacts so the data must be reconstructed before being used in scientific research (Schmidt et al. 2022). There are two kinds of process flows. Traditional methods, particularly the CLEAN method (Högbom 1974), have been widely used for this task. CLEAN iteratively refines dirty images to differentiate actual celestial structures from artifacts. However, the performance of CLEAN is limited due to its assumption of point-like sources and long inference time (Connor et al. 2022). Some learning-based methods also first transfer the sparse visibility data into dirty images through the imaging process and then reconstruct the dirty images into *clean images* (Bouman et al. 2016; Sun and Bouman 2021; Connor et al. 2022; Wang et al. 2023; Sortino et al. 2023; Drozdova et al. 2024). In contrast, some recent deep learning-based studies (Schmidt et al. 2022; Wu et al. 2022; Wang et al. 2024) have proposed to first do inpainting on the visibility data to reconstruct the sparse samples into dense coverage, and then perform imaging to obtain the clean image, showing better performance. Reconstructing the visibility data avoids the need for iteratively switching between Fourier and image spaces, enhances scalability, and facilitates the estimation of uncertainty in the output (Schmidt et al. 2022). Consequently, we mainly focus on the reconstruction-and-imaging process flow in this paper.

Our Method

Problem Formulation

Consider $x \in \mathbb{R}^n$ represents the brightness distribution of the sky, and its corresponding sampled visibility is $v \in \mathbb{R}^m$. They follow the observation model $v = \mathbf{P}_\Omega F(x) + \widetilde{\epsilon}$. In this model, F denotes the Fourier transformation, $\mathbf{P}_\Omega \in \mathbb{R}^{m \times n}$ is an under-sampling matrix where Ω is the telescope's sampling pattern, and $\widetilde{\epsilon}$ denotes noise with the same dimension as v. In this paper, our goal is to first estimate the full visibility data $F(x)$ from the observed sparse visibility data v, and then apply inverse Fourier transform F^{-1} to estimate the final brightness distribution of real sky x. Consider a dataset $\mathcal{D} = \mathcal{D}^{(s)} \cup \mathcal{D}^{(u)}$. $\mathcal{D}^{(s)}$ denotes the subset consisting of sparse visibility samples as well as corresponding ground truth images for supervised learning. $v_i^{(s)} \in \mathcal{D}^{(s)}$ is the sparse visibility measurement of the i^{th} example in $\mathcal{D}^{(s)}$. $x_i^{(s)} \in \mathcal{D}^{(s)}$ is the image ground truth corresponding to $v_i^{(s)}$. $\mathcal{D}^{(u)}$ is the subset including sparse visibility data lacking ground truth labels for unsupervised learning. $v_j^{(u)} \in \mathcal{D}^{(u)}$ is the j^{th} example of sparse visibility measurement in $\mathcal{D}^{(u)}$. The reconstruction model, denoted by f_θ, is parameterized

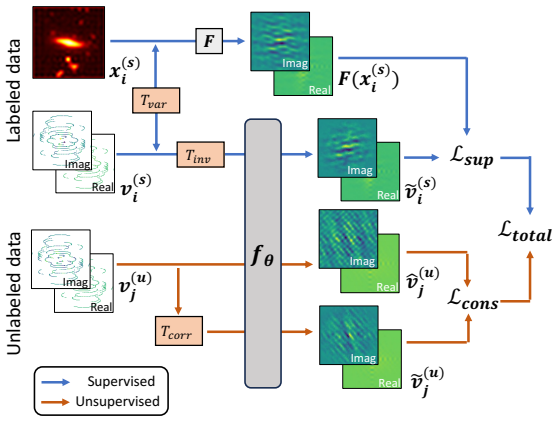


Figure 2: Overview of our method. The labeled data undergo supervised training and are augmented with label-variant and label-invariant augmentations T_{var} and T_{inv}. Then the neural network f_θ, processes augmented visibility data to produce reconstructions. These reconstructions are then compared against ground-truth references to compute supervised loss $\mathcal{L}_{sup}$. Unlabeled sparse visibility data are augmented by T_{corr}, and the same f_θ reconstructs both the non-augmented and augmented visibility data. The reconstruction from the non-augmented unlabeled data serves as a pseudo-label for that with augmentations to compute the consistency loss $\mathcal{L}_{cons}$. The overall loss combines supervised and consistency losses in a weighted sum.

by θ. $|\cdot|$ denotes the dataset size. In radio astronomical observation, the size of $|\mathcal{D}^{(s)}|$ is smaller than $|\mathcal{D}^{(u)}|$.

Label-Efficient Reconstruction of Visibility

We introduce VisRec, a semi-supervised learning framework that employs an arbitrary learnable reconstruction model f_θ, parameterized by weights θ, to reconstruct the visibility data into dense inpaintings. Then the inpaintings are used to produce high-fidelity clean images. VisRec consists of two modules, a supervised module and an unsupervised module.

Supervised Module In supervised radio interferometry data reconstruction, learning-based models are trained with datasets where reference clean images are available, denoted as $\mathcal{D}^{(s)}$. For each instance, the ground-truth clean image x is transformed into the reference visibility data $F(x)$ to be compared with sparse visibility $v_i^{(s)}$. Here, F is the transformation from the image domain to the visibility domain. The training objective of the model f_θ is to minimize the supervised loss function:

$$\min_\theta \frac{1}{|\mathcal{D}^{(s)}|} \sum_{i=0}^{|\mathcal{D}^{(s)}|} \mathcal{L}_{sup}\left(f_\theta\left(v_i^{(s)}\right), F\left(x_i^{(s)}\right)\right) \quad (2)$$

where $\mathcal{L}_{sup}$ represents the supervised loss function. The model f_θ refers to any parameterized learnable model.

To improve data diversity in label-scarce scenarios, supervised reconstruction techniques can employ data augmentation strategies. We define a series of label-invariant augmentation functions $T_{inv,l}, l = 1...m$ and label-variant augmentation functions $T_{var,k}, k = 1...n$, each applied to the visibility data with a certain probability:

$$T_{inv}(v_i^{(s)}) = T_{inv,1}^{(p_1)} \circ \ldots \circ T_{inv,m}^{(p_m)}(v_i^{(s)}) \quad (3)$$

$$T_{var}(v_i^{(s)}, x_i^{(s)}) = T_{var,1}^{(q_1)} \circ \ldots \circ T_{var,n}^{(q_n)}(v_i^{(s)}, x_i^{(s)}) \quad (4)$$

where p_j, q_k are the probabilities of applying each augmentation function.

The augmented loss for an instance with label-invariant visibility augmentation is:

$$\mathcal{L}_{sup}\left(f_\theta\left(T_{inv}\left(v_i^{(s)}\right)\right), F\left(x_i^{(s)}\right)\right). \quad (5)$$

The augmented loss for an instance with label-variant visibility augmentation is:

$$\mathcal{L}_{sup}\left(f_\theta\left(v_i'^{(s)}\right), F\left(x_i'^{(s)}\right)\right) \quad (6)$$

where $v_i'^{(s)}, x_i'^{(s)} = T_{var}(v_i^{(s)}, x_i^{(s)})$. This dual augmentation approach ensures that the learnable models are trained with a comprehensive range of variations. Suppose an instance of sparse visibility observation is in the form of $\{u_s, v_s, V(u_s, v_s)\}$, where u_s and v_s are coordinates at which a measurement is sampled, and $V(u_s, v_s)$ is the complex value of the sample.

For label-invariant visibility augmentation, we introduce the following methods:

- Position Offset: $(u_s', v_s') = (u_s, v_s) + \epsilon_{pos}$, where $\epsilon_{pos} \in \mathbb{R}^{\text{shape}((u_s,v_s))} \sim \mathcal{N}(0, \sigma_{\text{pos}})$ is Gaussian noise added to the position coordinates.
- Visibility Noise: $V'(u_s, v_s) = V(u_s, v_s) + \epsilon_{vis}$, where $\epsilon_{vis} \in \mathbb{C}^{\text{shape}(V(u_s,v_s))} \sim \mathcal{N}(0, \sigma_{vis})$ is complex Gaussian noise added to the visibility values.
- Global Position Offset: $(u_s', v_s') = (u_s, v_s) + (\Delta u, \Delta v)$, where $(\Delta u, \Delta v)$ is a uniform random offset applied to all position coordinates.
- Random Dropout: A random subset of visibility data is dropped out.
- Subset of Selective Frequency: The selected frequency subset is represented as: $\left\{(u_s, v_s, V(u_s, v_s)) \,\middle|\, d_{\text{min}} \leq \sqrt{u_s^2 + v_s^2} \leq d_{\text{max}}\right\}$ where d_{min} and d_{max} are the minimum and maximum frequencies of the selected range.

For label-variant visibility augmentation, we introduce some geometric transformations in the uv-plane. To augment sparsely sampled visibility data $\{u_s, v_s, V(u_s, v_s)\}$ while maintaining correspondence with the image ground-truth x, the augmentations must be consistent across both the visibility data and the image domain:

- Swap u_s and v_s coordinates for the visibility data while apply a 90-degree rotation to the clean image x.
- Negate the u_s or v_s coordinates in the visibility data. For the clean image x, apply a horizontal or vertical flip.
- Negate both u_s and v_s in the visibility data and apply a central inversion to the clean image x.

Unsupervised Module The unsupervised module of VisRec uses unlabeled visibility data $\mathcal{D}^{(u)}$ that miss ground-truth images. Our aim is to leverage these data to train the model f_θ to recover the original signal from a corruption-augmented measurement (Desai et al. 2023). Each unsupervised example $v_i^{(u)}$ is augmented by a predefined signal corruption model T_{corr}, which simulates expected corrupted patterns that occur in radio interferometry observations. The corruption model T_{corr} applied to each unsupervised instance simulates a variety of realistic observational impairments:

- Observation Noise: Introducing zero-mean Gaussian noise (Schmidt et al. 2022) to simulate electronic and atmospheric noise.
- Missing Data: Simulating the offline antennas by masking out some visibility points (Nasirudin et al. 2022).
- Antenna Offset: Introducing perturbations that represent errors in antenna positioning or alignment by adding noise to the position coordinates (Nasirudin et al. 2022).

The unsupervised training objective is to minimize the following loss function:

$$\min_{\theta} \frac{1}{|\mathcal{D}^{(u)}|} \sum_{j=0}^{|\mathcal{D}^{(u)}|} \mathcal{L}_{cons} \left(f_\theta \left(T_{corr}(v_j^{(u)}) \right), f_\theta \left(v_j^{(u)} \right) \right) \quad (7)$$

where T_{corr} is applied to the visibility data with a predefined probability. This approach trains the model to reconstruct the corrupted visibility by learning to reverse the effects of the corruption, which we refer to as *consistency training*.

Semi-Supervised Training VisRec consists of both supervised learning and unsupervised learning processes. The main framework of our method is described in Algorithm 1. In the supervised process, we sample a batch of $\{(v_i^{(s)}, x_i^{(s)})\}_{i=1}^{N_s}$ from $\mathcal{D}^{(s)}$. The visibility data $v_i^{(s)}$ and ground-truth image labels $x_i^{(s)}$ are augmented by label-variant augmentation function T_{var}. Then the visibility data is augmented by label-invariant augmentation function T_{inv}. Then the model f_θ generates the reconstruction of input visibility data. In the unsupervised process, we sample the batch $\{v_j^{(u)}\}_{j=1}^{N_u}$ from $\mathcal{D}^{(u)}$. Then each $v_j^{(u)}$ is augmented by corruption model T_{corr}. Finally, the same model f_θ reconstructs both augmented and original visibility. Finally, a supervised loss $\mathcal{L}_{sup}$ is computed using the mean squared error between the Fourier-transformed ground truth images and the reconstructed visibility data for the supervised subset. Meanwhile, a consistency loss $\mathcal{L}_{cons}$ is calculated for the unsupervised data by comparing the model's output on the original and augmented visibility samples. The total loss $\mathcal{L}_{total}$ is a weighted sum of these two losses, where λ is a hyperparameter that balances the two components. The model parameters θ are updated to minimize $\mathcal{L}_{total}$.

Algorithm 1: Main process of VisRec.

Input: Dataset $\mathcal{D} = \mathcal{D}^{(s)} \cup \mathcal{D}^{(u)}$
Output: Trained model f_θ

for each sampled batch $\{(v_i^{(s)}, x_i^{(s)})\}_{i=1}^{N_s}$ and $\{v_j^{(u)}\}_{j=1}^{N_u}$ **do**
  for all $i \in \{1, \ldots, N_s\}$ **do**
    $v_i^{\prime(s)}, x_i^{\prime(s)} \leftarrow T_{var}(v_i^{(s)}, x_i^{(s)})$
    $\tilde{v}_i^{(s)} \leftarrow f_\theta(T_{inv}(v_i^{\prime(s)}))$
  end for
  for all $j \in \{1, \ldots, N_u\}$ **do**
    $\tilde{v}_j^{(u)} \leftarrow f_\theta(v_j^{(u)})$
    $\hat{v}_j^{(u)} \leftarrow f_\theta(T_{corr}(v_j^{(u)}))$
  end for
  $\mathcal{L}_{sup} \leftarrow \frac{1}{N_s} \sum_{i=1}^{N_s} \mathcal{L}_{sup}(F(x_i^{\prime(s)}), \tilde{v}_i^{(s)})$
  $\mathcal{L}_{cons} \leftarrow \frac{\lambda}{N_u} \sum_{j=1}^{N_u} \mathcal{L}_{cons}(\hat{v}_j^{(u)}, \tilde{v}_j^{(u)})$
  $\mathcal{L}_{total} \leftarrow \mathcal{L}_{sup} + \lambda \mathcal{L}_{cons}$
  Update network f_θ to minimize $\mathcal{L}_{total}$
end for
return f_θ

Experiments

In this section, we evaluate how well VisRec (1) performs in comparison with existing state-of-the-art non-learning, supervised, and self-supervised methods in label-scarce scenarios and (2) improves the robustness and generalizability of reconstruction models.

Experimental Setting

Datasets: Following the latest astronomical data reconstruction studies (Wu et al. 2022; Schmidt et al. 2022; Wang et al. 2023; Geyer et al. 2023), we use two telescope configurations to sample the visibility observation from real astronomical images to build the visibility dataset. For reference images, we use the Galaxy10 DECals dataset (Henry 2021) which contains 17,736 images of various galaxies. The dataset is from the DESI Legacy Imaging Surveys (Dey et al. 2019), which merges data from the Beijing-Arizona Sky Survey (BASS) (Zou et al. 2017), the DECam Legacy Survey (DECaLS) (Blum et al. 2016), and the Mayall z-band Legacy Survey (Silva et al. 2016). Using these images as a reference, we employ the eht-imaging toolkit (Chael et al. 2019, 2018) to produce visibility data represented by $\{u_s, v_s, V(u_s, v_s)\}$. The parameters for observation are adjusted to mirror an 8-telescope Event Horizon Telescope (EHT) setup (Wu et al. 2022), with the EHT being one of the most prominent arrays leveraging VLBI techniques. In addition, following (Schmidt et al. 2022; Geyer et al. 2023), we apply the Very Long Baseline Array (VLBA) uv-coverage to simulate another dataset. In EHT dataset, each image has 1,660 visibility points sampled whereas in VLBA dataset, the number of visibility points is 2,298. The image dimensions are all set at 256×256 pixels. For each dataset, we randomly select 1,024 examples for testing and the remaining samples are for training. More details regarding the datasets are reported in the appendix. To simulate a label-scarce setting, we randomly split a small number of examples from the training set as the labeled dataset, and the remaining samples are all used as unlabeled data.

Platform: We conduct all experiments on a server with two AMD EPYC 7763 CPUs, 512GB main memory, and eight Nvidia RTX 4090 GPUs each with 24GB device memory. The server is equipped with two NVME 2TB SSD and

Method	EHT Dataset			VLBA Dataset		
	LFD (↓)	SSIM (↑)	PSNR (dB) (↑)	LFD (↓)	SSIM (↑)	PSNR (dB) (↑)
Dirty Image	N/A	0.685 (0.059)	10.58 (1.28)	N/A	0.745 (0.041)	12.24 (1.34)
CLEAN (Högbom 1974)	N/A	0.830 (0.032)	19.17 (2.61)	N/A	0.823 (0.031)	18.48 (2.52)
Noise2Astro (Zhang et al. 2022)	N/A	0.704 (0.055)	11.09 (1.36)	N/A	0.761 (0.034)	13.08 (1.37)
Neural Field (Wu et al. 2022)	1.083 (0.328)	0.875 (0.030)	21.36 (2.51)	1.152 (0.312)	0.874 (0.032)	20.26 (2.60)
NF Supervised + Aug	1.043 (0.343)	0.883 (0.028)	21.89 (2.47)	1.083 (0.334)	0.888 (0.030)	21.24 (2.75)
NF Self-Supervised	1.465 (0.234)	0.703 (0.056)	11.45 (1.45)	1.433 (0.239)	0.766 (0.018)	13.29 (0.90)
NF VisRec w/o Sup-Aug	1.029 (0.347)	0.877 (0.033)	22.53 (2.73)	1.070 (0.348)	0.880 (0.033)	22.22 (2.83)
NF VisRec	**1.029 (0.337)**	**0.886 (0.029)**	**22.63 (2.64)**	**1.067 (0.318)**	**0.889 (0.031)**	**22.87 (3.06)**
CNN (Schmidt et al. 2022)	1.061 (0.237)	0.864 (0.024)	22.22 (2.46)	0.949 (0.446)	0.875 (0.048)	22.15 (2.59)
CNN Supervised + Aug	0.991 (0.387)	0.870 (0.044)	22.32 (4.19)	0.950 (0.446)	0.875 (0.048)	23.21 (4.60)
CNN Self-Supervised	1.481 (0.226)	0.785 (0.020)	15.18 (1.52)	1.369 (0.222)	0.793 (0.022)	14.65 (1.32)
CNN VisRec w/o Sup-Aug	**0.897 (0.245)**	0.873 (0.029)	22.60 (3.20)	0.847 (0.251)	0.888 (0.029)	23.62 (3.13)
CNN VisRec	0.950 (0.292)	**0.892 (0.027)**	**23.37 (2.73)**	**0.793 (0.255)**	**0.899 (0.024)**	**23.94 (2.44)**

Table 1: Comparison of image reconstruction quality on EHT and VLBA datasets (mean and standard deviation). Rows with gray background denote our method and its variants, and the rows with white background represent methods from prior studies. All comparison methods have significant differences with our method ($p < 0.05$ in the Wilcoxon signed-rank test).

two 16TB SATA hard disks. The operating system is Ubuntu 20.04. Our model is implemented in PyTorch 1.8.1 (Paszke et al. 2019).

Evaluation Metrics: To evaluate frequency data differences, we use Log Frequency Distance (LFD) (Jiang et al. 2021), where lower values indicate better performance. To assess the quality of recovered images, we apply Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) computed via scikit-image package (Singh 2019). A higher PSNR and SSIM is better.

Baseline Methods

In this study, we compare our method with the most prevalent non-learning method CLEAN (Högbom 1974), the self-supervised denoising method for astronomical image Noise2Astro (Zhang et al. 2022), and two state-of-the-art models for visibility reconstruction (Wu et al. 2022; Schmidt et al. 2022). Radionet (Schmidt et al. 2022) uses a convolutional neural network (CNN) as the reconstruction model and Neural Interferometry (Wu et al. 2022) uses a neural field model (NF). We apply VisRec to these two models, and we also introduce three variants of our method as baseline approaches: (1) Supervised Learning with Data Augmentation: Adding the proposed augmentation methods to the supervised training of the NF and CNN model, denoted NF Supervised + Aug and CNN Supervised + Aug, respectively. Essentially, this approach represents the VisRec framework with its unsupervised consistency training component removed. (2) Self-Supervised Training: This is a variant of VisRec that is extended to a fully unsupervised setting. In this setup, the supervised training pathway is replaced with the self-supervised training setup from Noise2Noise (Lehtinen et al. 2018). (3) VisRec without Augmentation in Supervised Module: In this variant, we remove all augmentations in the supervised module to evaluate the impact of consistency learning separately, denoted VisRec w/o Sup-Aug.

Overall Comparison

To assess the performance of our method in scenarios with limited label availability, we conduct an overall comparison using only 1,024 labeled visibility examples ($|\mathcal{D}^{(s)}| = 1,024$) and a set of 15,668 unlabeled data ($|\mathcal{D}^{(u)}|$ = 15,668).

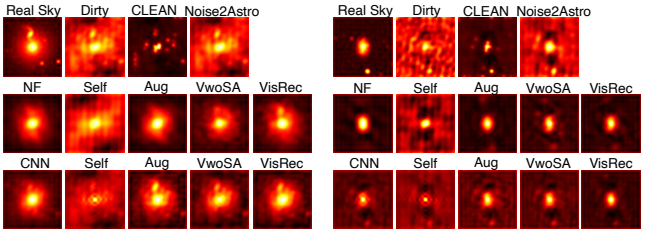


Figure 3: Visual Examples of overall comparison. VwoSA denotes VisRec w/o Sup-Aug.

We calculate the LFD, PSNR, and SSIM values for all test data reconstructed using our method and other baseline methods, presenting both mean values and standard deviations in Table 1. The results show that both NF model (Wu et al. 2022) and CNN model (Schmidt et al. 2022) trained by supervised methods outperform traditional non-learning method (CLEAN) and self-supervised Noise2Astro (Zhang et al. 2022). Our VisRec approach, applied to both neural field and CNN models, shows significant improvements in all metrics compared to the original supervised methods. Its variant model Supervised + Aug and VisRec w/o Sup-Aug also shows better reconstruction quality than the supervised versions, demonstrating that both the augmentation and the consistency training in VisRec are effective. We also present two sets of representative reconstructed images from the EHT dataset in Figure 3. More visual examples are reported in the appendix.

Effect of Labeled Data Size

We also evaluate the performance of our method and other baselines under various labeled data sizes. We apply VisRec, along with its two variants (Supervised + Aug and VisRec w/o Sup-Aug) to train the NF model (Wu et al. 2022) on the EHT dataset. Maintaining a constant total training data size $|\mathcal{D}|$ = 16,692, we vary the size of the labeled subset $|\mathcal{D}^{(s)}|$ from 64 to 2k (k = 1024) and use the remaining data as the unlabeled subset. We also evaluate the performance of the original supervised model trained with all 16,692 labeled data examples, denoted NF (Full data). As shown in Fig-

ure 4, our method and its variants consistently outperform the original supervised method. With only 64 labeled examples, our method outperforms the traditional non-learning method CLEAN. With only 2k labeled examples, VisRec performs on par with supervised methods trained with over 16k examples. The results show that the effectiveness of VisRec comes from the use of both supervised and unsupervised data, making it well-suited for limited-label scenarios. Moreover, incorporating data augmentation also improves model performance by enhancing training diversity. In addition, we report the effect of labeled data size on VisRec performance with the CNN model in the appendix.

Robustness Evaluation

Robustness against Noise. In radio astronomy, both instrumental and environmental factors can introduce noise (Schmidt et al. 2022), which may significantly degrade the quality of the images reconstructed by previous supervised learning methods. To evaluate the robustness of various methods to such noise, we conduct performance tests across different noise levels. Following Schmidt et al. (Schmidt et al. 2022), we add white noise that simulates potential measurement effects in the frequency domain by corrupting the visibilities with the noise of Gaussian distribution. This process can be represented as $V_{\text{noisy}}(u_s, v_s) = V(u_s, v_s) + g(x|\mu, \sigma)_{uv}$, where $g(x|\mu, \sigma)_{uv}$ is a random value generated for each frequency sample, with $\mu = 0$ and noise level $\sigma = 0$ to 10. Figure 6 shows that models applied VisRec consistently outperform other methods and have lower sensitivity to noise. The two variants of our method (Supervised + Aug and VisRec w/o Sup-Aug) also show better robustness than the original supervised methods. We also display some visual examples in Figure 5, showing that as the noise level increases, the reconstruction of supervised learning methods are severely corrupted, whereas the reconstruction results of VisRec still maintain high quality.

Robustness against Missing Visibility Points. Missing visibility points, often resulting from offline antennas in interferometric arrays (Nasirudin et al. 2022), also cause performance decrease of learning-based reconstruction methods. We vary the sample loss rate from 0 to 50% and evaluate the performance of our method and the baseline methods. To show the result more clearly, we list the value of the NF model and CNN model trained with VisRec and the supervised method respectively in Table 2. The rate of VisRec performance decrease is significantly lower than supervised learning when applied to both NF and CNN models, indicating that VisRec improves the robustness of deep models.

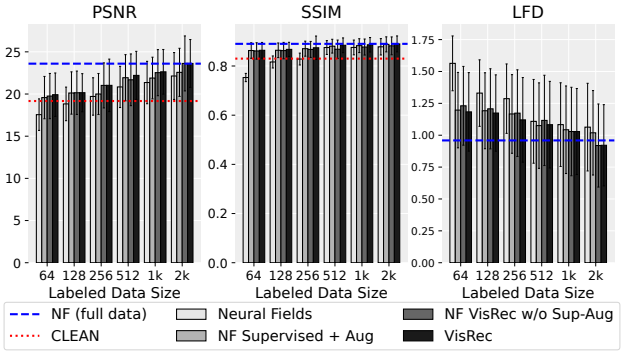


Figure 4: Effect of Labeled Data Size.

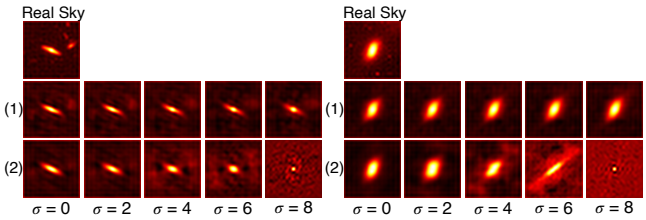


Figure 5: Visual examples of the effect of noise. Row (1) is the results of the NF_VisRec and row (2) is the results of supervised NF. The examples are all from the EHT dataset.

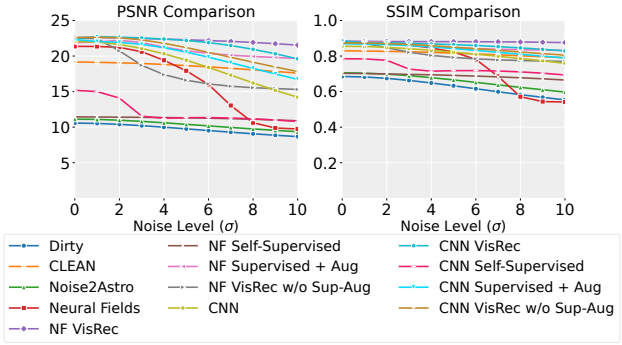


Figure 6: Performance of different methods across various noise levels. We only show PSNR and SSIM values because LFD is not applicable to Dirty, CLEAN, and Noise2Astro.

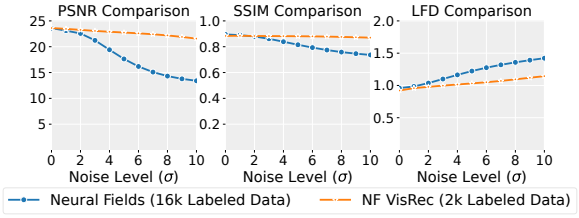
Comparison with Supervised Training using Full Data. Results in Figure 4 show that with only 2k labeled examples and remaining unlabeled data, Neural Fields trained with VisRec (NF VisRec) performs on par with supervised methods trained with over 16k examples (size = 16,692) in the full dataset when the testing data has no observation errors (ideal). Here, we further evaluate how the NF model with supervised training on 16k labeled data and NF VisRec with 2k data perform when the testing data are corrupted by observational perturbations. Figure 7 (a) and (b) show that NF VisRec trained with only 2k labels outperforms the supervised method when the noise level or the data missing rate increases. This means that VisRec not only maintains efficiency under data-limited conditions but can also handle robustness issues that are likely to be encountered in practical applications. Similarly, CNN VisRec trained with 1k labels performs on par with supervised CNN methods trained with over 16k examples when the testing data is ideal.

Generalization to Another Telescope

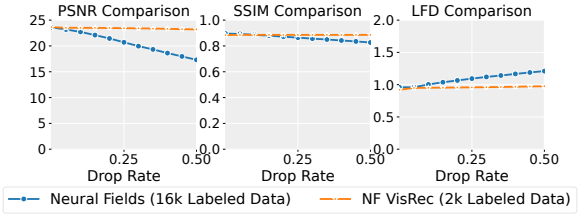
We evaluate the generalizability of different methods by training the models on the EHT dataset (the same train-

Loss Rate	NF Sup	NF VisRec	CNN Sup	CNN VisRec
0	21.36 / 0.8754	22.63 / 0.8855	22.22 / 0.8636	23.37 / 0.8920
50%	19.25 / 0.8506	22.40 / 0.8848	19.59 / 0.8372	22.51 / 0.8889
Decrease	-9.88% / -2.83%	-1.02% / -0.08%	-11.84% / -3.06%	-3.68% / -0.35%

Table 2: PSNR/SSIM values of methods at 0 and 50% sample loss rates.



(a) Comparison results in a noise settings.



(b) Comparison results in a missing samples settings.

Figure 7: Comparison between VisRec training in a label-scarce setting and supervised training with full data.

Model	Same Dataset			Different Dataset		
	LFD	SSIM	PSNR	LFD	SSIM	PSNR
NF	1.083	0.875	21.36	1.319	0.819	16.74
NF_v1	1.043	0.883	21.89	1.305	0.842	**18.95**
NF_v2	1.465	0.703	11.45	1.489	0.595	9.55
NF_v3	1.029	0.877	22.53	**1.242**	0.844	18.42
NF_ours	**1.029**	**0.886**	**22.63**	1.250	**0.855**	18.65
CNN	1.061	0.864	22.22	1.141	0.802	16.13
CNN_v1	0.991	0.870	22.32	1.238	0.823	18.28
CNN_v2	1.481	0.785	15.17	1.409	0.736	11.88
CNN_v3	**0.897**	0.873	22.60	1.044	0.822	16.99
CNN_ours	0.950	**0.892**	**23.37**	**1.001**	**0.866**	**20.65**

Table 3: Performance of training methods with the same or different datasets. v1, v2, v3 refer to Supervised + Aug, Self-Supervised and VisRec w/o Sup-Aug, respectively.

ing setting as in Section Overall Comparison) and testing their performance on both EHT and VLBA datasets. As shown in Table 3, the significant performance drop in all models when applied to different datasets underscores the challenge in model generalization. VisRec, VisRec w/o Sup-Aug, and Supervised+Aug methods outperform the traditional supervised method, especially on different datasets, indicating superior generalizability. While in some cases the variants of our method, e.g., VisRec w/o Sup-Aug, and Supervised+Aug show marginally better performance than VisRec in certain metrics, the overall performance of VisRec is superior. Particularly, the CNN model trained with VisRec shows the least performance degradation, suggesting its high generalizability in handling diverse data distributions.

Impact of Consistency Loss Weight

We evaluate the impact of consistency loss weight λ in VisRec on the reconstruction of visibilities with different noise corruption levels. The results are shown in Figure 8. The performance of VisRec does not change much for a large range of $\lambda \in [0.01, 0.8)$. Insensitivity to changes in λ may help eliminate the need for hyperparameter tuning, which can simplify network training (Desai et al. 2023).

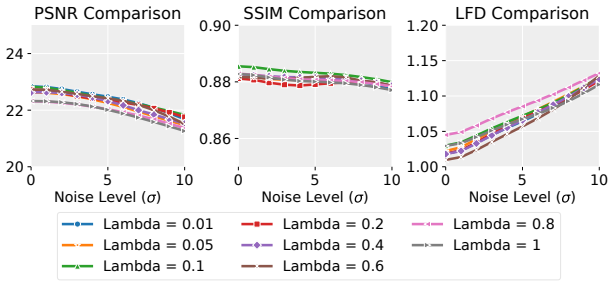


Figure 8: Impact of λ. The differences are small so we show only a selected range of the y-axis for close observation.

Conclusion and Future Work

In this study, we present VisRec, a semi-supervised learning framework for visibility data reconstruction in radio astronomy. VisRec effectively leverages both labeled and unlabeled data, significantly reducing dependence on extensive ground truth labels. Our results show that VisRec outperforms state-of-the-art methods in reconstruction quality, robustness to common observational perturbations, and generalizability across different telescope configurations. By improving the data reconstruction process, our work helps with subsequent scientific discoveries and a deeper understanding of the universe. While VisRec is designed for radio astronomical data reconstruction, we are considering extending this method to other domains such as magnetic resonance imaging, seismic imaging, and point cloud data analysis. The training approach and augmentations may not be directly applicable to these other domains, so further research is needed to adapt VisRec for use in different fields.

Acknowledgements

This work is supported by the China National SKA Programme (2020SKA0110300), the China National Science Foundation (NSF) (62372393, 12433012, 12373097), the Guangzhou Science and Technology Funds (2023A03J0016) and the Guangzhou Municipality Big Data Intelligence Key Lab (2023A03J0012).

References

Ables, J. 1974. Maximum entropy spectral analysis. *Astronomy and Astrophysics Supplement, Vol. 15, p. 383*, 15: 383.

Blum, R. D.; Burleigh, K.; Dey, A.; Schlegel, D. J.; Meisner, A. M.; Levi, M.; Myers, A. D.; Lang, D.; Moustakas, J.; Patej, A.; et al. 2016. The decam legacy survey. In *American Astronomical Society Meeting Abstracts# 228*, volume 228, 317–01.

Bouman, K. L.; Johnson, M. D.; Dalca, A. V.; Chael, A. A.; Roelofs, F.; Doeleman, S. S.; and Freeman, W. T. 2018. Reconstructing video of time-varying sources from radio interferometric measurements. *IEEE Transactions on Computational Imaging*, 4(4): 512–527.

Bouman, K. L.; Johnson, M. D.; Zoran, D.; Fish, V. L.; Doeleman, S. S.; and Freeman, W. T. 2016. Computational imaging for VLBI image reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 913–922.

Chael, A. A.; Bouman, K. L.; Johnson, M. D.; Narayan, R.; Doeleman, S. S.; Wardle, J. F.; Blackburn, L. L.; Akiyama, K.; Wielgus, M.; Chan, C.-k.; et al. 2019. ehtim: Imaging, analysis, and simulation software for radio interferometry. *Astrophysics Source Code Library*, ascl–1904.

Chael, A. A.; Johnson, M. D.; Bouman, K. L.; Blackburn, L. L.; Akiyama, K.; and Narayan, R. 2018. Interferometric imaging directly with closure phases and closure amplitudes. *The Astrophysical Journal*, 857(1): 23.

Connor, L.; Bouman, K. L.; Ravi, V.; and Hallinan, G. 2022. Deep radio-interferometric imaging with POLISH: DSA-2000 and weak lensing. *Monthly Notices of the Royal Astronomical Society*, 514(2): 2614–2626.

Dagli, R. 2023. Astroformer: More Data Might Not be All You Need for Classification. *arXiv preprint arXiv:2304.05350*.

Darestani, M. Z.; Chaudhari, A. S.; and Heckel, R. 2021. Measuring robustness in deep learning based compressive sensing. In *International Conference on Machine Learning*, 2433–2444. PMLR.

Desai, A. D.; Ozturkler, B. M.; Sandino, C. M.; Boutin, R.; Willis, M.; Vasanawala, S.; Hargreaves, B. A.; Ré, C.; Pauly, J. M.; and Chaudhari, A. S. 2023. Noise2Recon: Enabling SNR-robust MRI reconstruction with semi-supervised and self-supervised learning. *Magnetic Resonance in Medicine*, 90(5): 2052–2070.

Dey, A.; Schlegel, D. J.; Lang, D.; Blum, R.; Burleigh, K.; Fan, X.; Findlay, J. R.; Finkbeiner, D.; Herrera, D.; Juneau, S.; et al. 2019. Overview of the DESI legacy imaging surveys. *The Astronomical Journal*, 157(5): 168.

Drozdova, M.; Kinakh, V.; Bait, O.; Taran, O.; Lastufka, E.; Dessauges-Zavadsky, M.; Holotyak, T.; Schaerer, D.; and Voloshynovskiy, S. 2024. Radio-astronomical image reconstruction with a conditional denoising diffusion model. *Astronomy & Astrophysics*, 683: A105.

Geyer, F.; Schmidt, K.; Kummer, J.; Brüggen, M.; Edler, H.; Elsässer, D.; Griese, F.; Poggenpohl, A.; Rustige, L.; and Rhode, W. 2023. Deep-learning-based radiointerferometric imaging with GAN-aided training. *Astronomy & Astrophysics*, 677: A167.

Henry, L. 2021. Galaxy10 DECals Dataset. https://github.com/henrysky/Galaxy10.

Högbom, J. 1974. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 15: 417.

Jia, J.; Hong, M.; Zhang, Y.; Akcakaya, M.; and Liu, S. 2022. On the Robustness of deep learning-based MRI Reconstruction to image transformations. In *Workshop on Trustworthy and Socially Responsible Machine Learning, NeurIPS 2022*.

Jiang, L.; Dai, B.; Wu, W.; and Loy, C. C. 2021. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13919–13929.

Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; and Aila, T. 2018. Noise2Noise: Learning Image Restoration without Clean Data. In *International Conference on Machine Learning*, 2965–2974. PMLR.

Nasirudin, A.; Prelogovic, D.; Murray, S. G.; Mesinger, A.; and Bernardi, G. 2022. Characterizing beam errors for radio interferometric observations of reionization. *Monthly Notices of the Royal Astronomical Society*, 514(3): 4655–4668.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Schmidt, K.; Geyer, F.; Fröse, S.; Blomenkamp, P.-S.; Brüggen, M.; de Gasperin, F.; Elsässer, D.; and Rhode, W. 2022. Deep Learning-based Imaging in Radio Interferometry. *arXiv preprint arXiv:2203.11757*.

Silva, D. R.; Blum, R. D.; Allen, L.; Dey, A.; Schlegel, D. J.; Lang, D.; Moustakas, J.; Meisner, A. M.; Valdes, F.; Patej, A.; et al. 2016. The Mayall z-band Legacy Survey. In *American Astronomical Society Meeting Abstracts# 228*, volume 228, 317–02.

Singh, H. 2019. Basics of python and Scikit image. *Practical Machine Learning and Image Processing: For Facial Recognition, Object Detection, and Pattern Recognition Using Python*, 29–61.

Sortino, R.; Cecconello, T.; DeMarco, A.; Fiameni, G.; Pilzer, A.; Hopkins, A. M.; Magro, D.; Riggi, S.; Sciacca, E.; Ingallinera, A.; et al. 2023. RADiff: Controllable Diffusion Models for Radio Astronomical Maps Generation. *arXiv e-prints*, arXiv–2307.

Sun, H.; and Bouman, K. L. 2021. Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2628–2637.

Thompson, A. R.; Moran, J. M.; and Swenson, G. W. 2017. *Interferometry and synthesis in radio astronomy*. Springer Nature.

Wang, R.; Chen, Z.; Luo, Q.; and Wang, F. 2023. A Conditional Denoising Diffusion Probabilistic Model for Radio Interferometric Image Reconstruction. In *26th European Conference on Artificial Intelligence*, 2499–2506.

Wang, R.; Chen, Z.; Zhu, J.; Luo, Q.; and Wang, F. 2024. PolarRec: Improving Radio Interferometric Data Reconstruction Using Polar Coordinates. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12841–12850.

Wu, B.; Liu, C.; Eckart, B.; and Kautz, J. 2022. Neural Interferometry: Image Reconstruction from Astronomical Interferometers using Transformer-Conditioned Neural Fields. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Zhang, Y.; Nord, B.; Pagul, A.; and Lepori, M. 2022. Noise2Astro: Astronomical Image Denoising with Self-supervised Neural Networks. *Research Notes of the AAS*, 6(9): 187.

Zou, H.; Zhou, X.; Fan, X.; Zhang, T.; Zhou, Z.; Nie, J.; Peng, X.; McGreer, I.; Jiang, L.; Dey, A.; et al. 2017. Project overview of the beijing–arizona sky survey. *Publications of the Astronomical Society of the Pacific*, 129(976): 064101.