stochastic information processing

UNIVERSITÉ DE GENÈVE

Conditional Denoising Diffusion Reconstruction of Radio Astronomical Images

M. Drozdova, V. Kinakh, O. Bait, O. Taran, E. Lastufka, M. Dessauges-Zavadsky, T. Holotyak, D. Schaerer, S. Voloshynovskiy

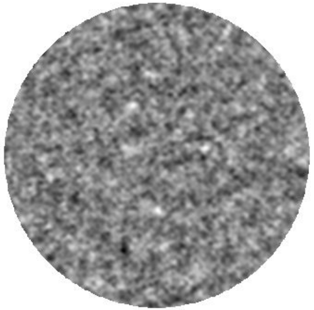
Presented by Mariia Drozdova

Problem Formulation

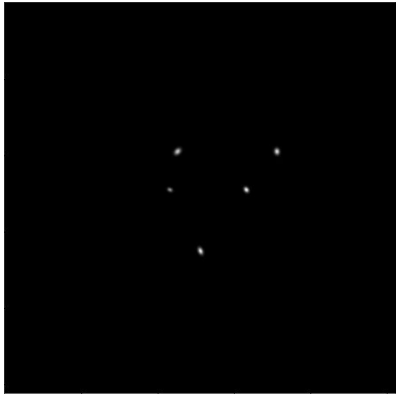
Problem: Localization and characterization of radio sources

Solution: A model which reconstructs sky model images from dirty noisy images

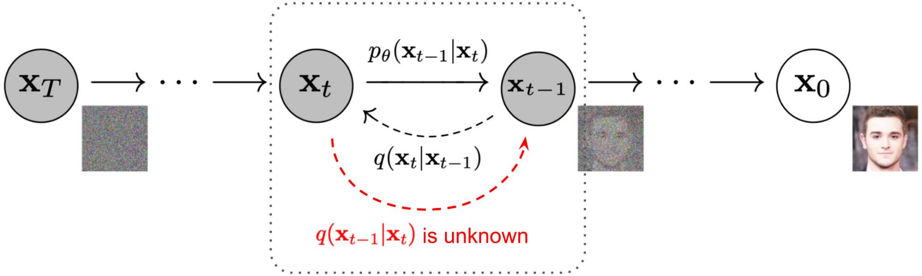
Tool: Denoising Diffusion Probabilistic Models (DDPMs)



Dirty noisy image



Sky model image



DDPM*

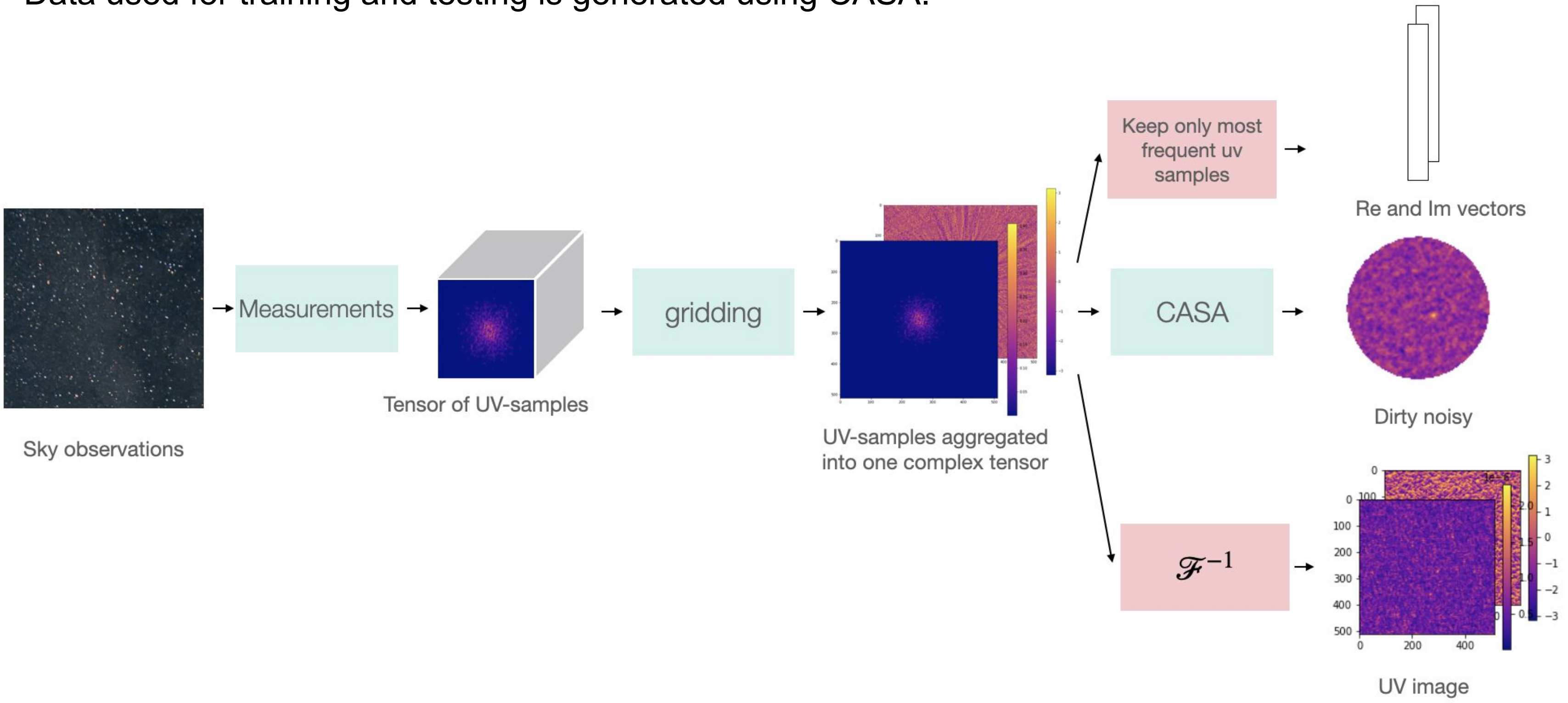
*scheme from Lil'Log

Overview

1. Data Acquisition
2. The Need for Normalization
3. Schemes: Inference and Training
4. Results - Multiple Estimates and Aggregation
5. Flux Estimation
6. Conclusion and Future Work

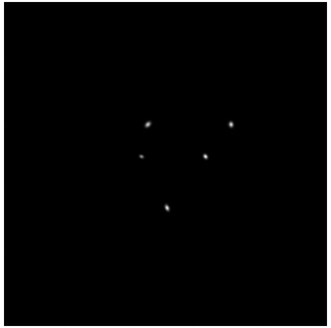
Data Acquisition

Data used for training and testing is generated using CASA.



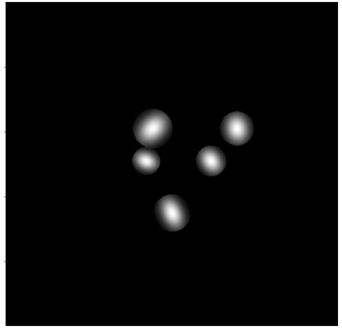
The Need for Normalization

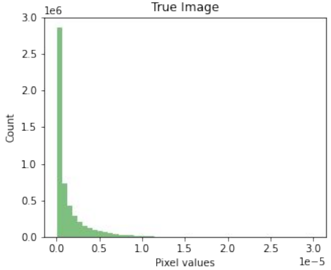
The Sky Model images are very sparse -> normalization is needed for MSE loss to work

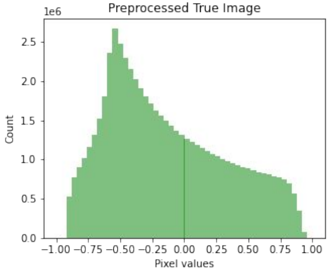


$$I_{\text{true, normalized}} = \left(\frac{I_{\text{true}}}{C}\right)^{\frac{1}{30}}$$

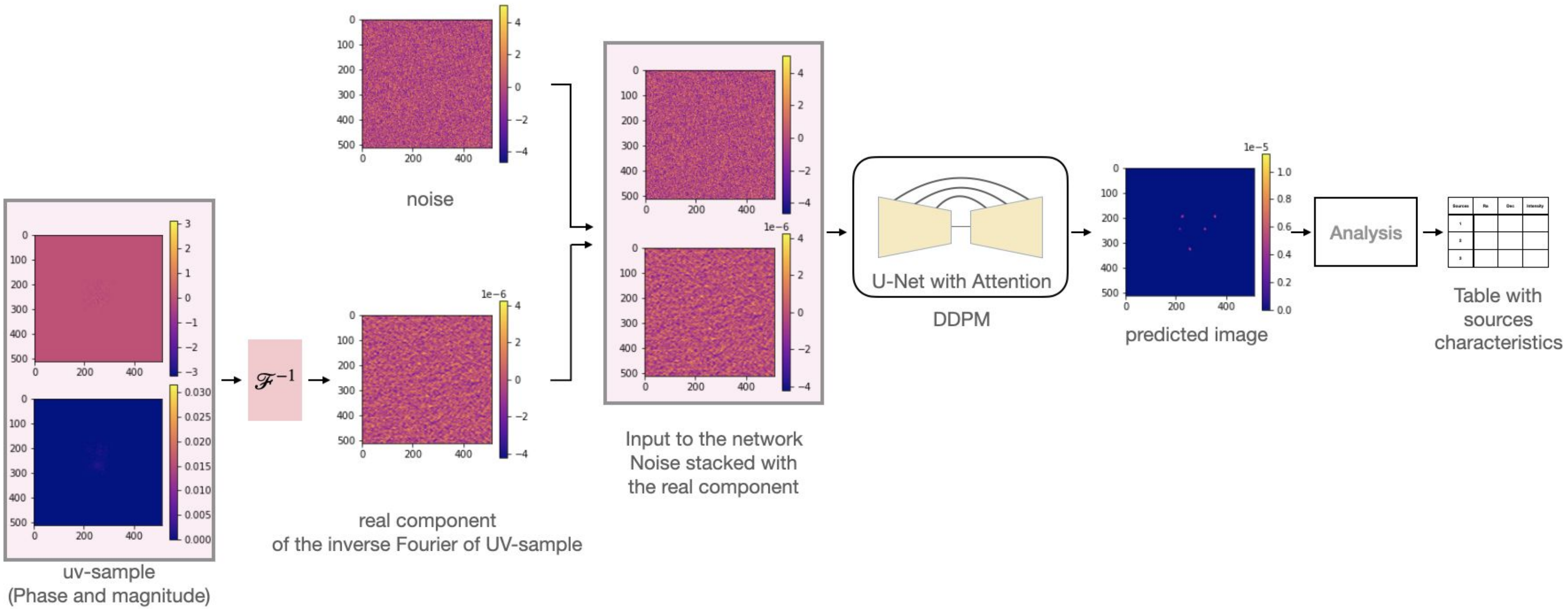
$$I_{\text{true, scaled}} = \frac{I_{\text{true, normalized}} - 0.5}{0.5}$$





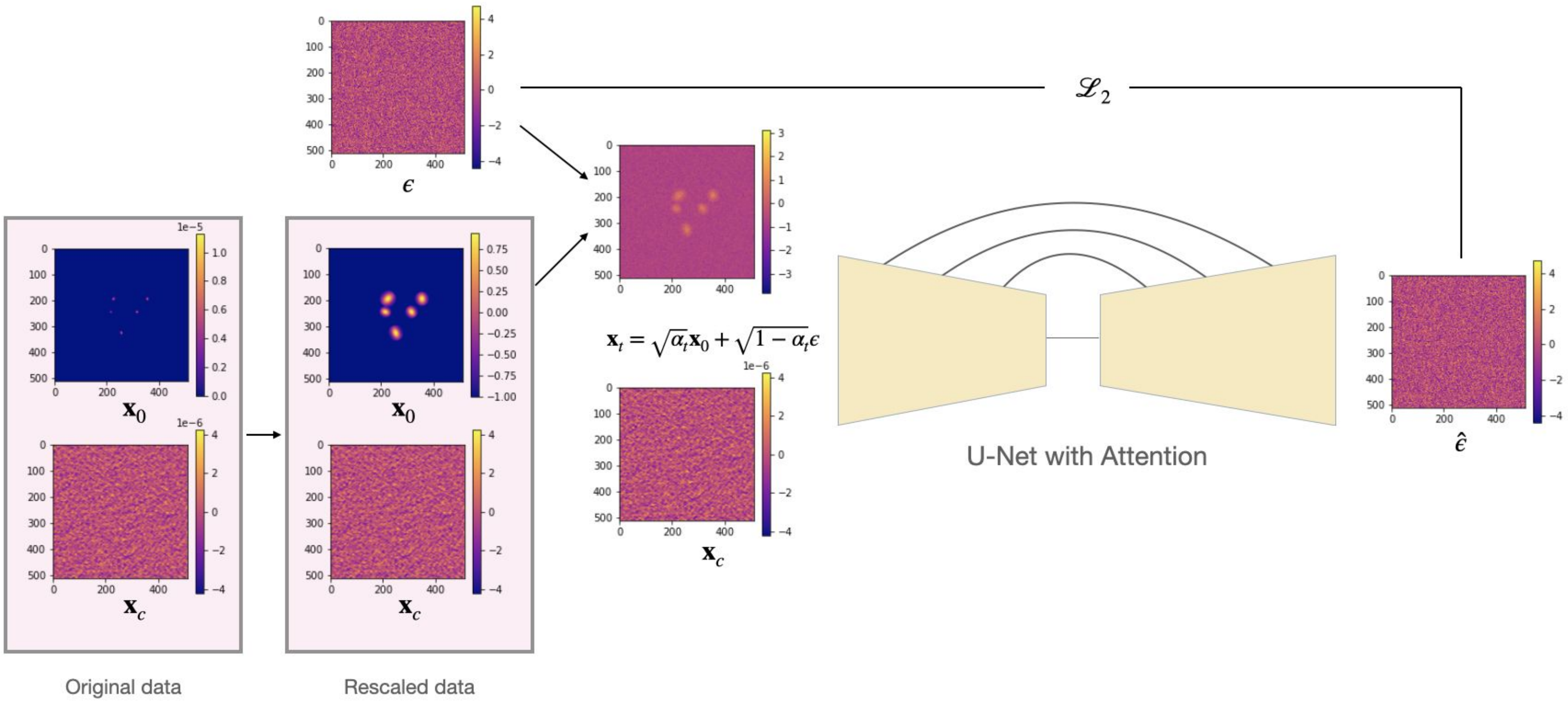


Scheme: Inference



[Palette: Image-to-Image Diffusion Models, C. Saharia et al., 2022]

Scheme: Training The Model



Evaluation Metrics

Reconstruction:

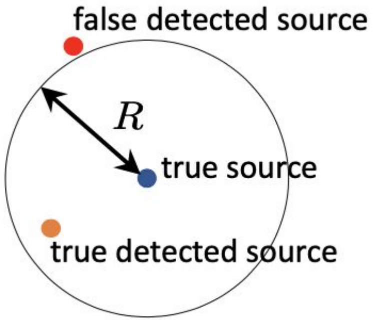
- MSE = mean square error
- PSNR = ratio between the maximal possible power of a signal to the power of the distortion in a logarithmic scale
- SSIM = structural similarity index

Sources Localization:

- Purity = fraction of true sources among detected sources
- Completeness = fraction of true sources which are detected

Flux estimation:

- Correlation Coefficient
- R-squared



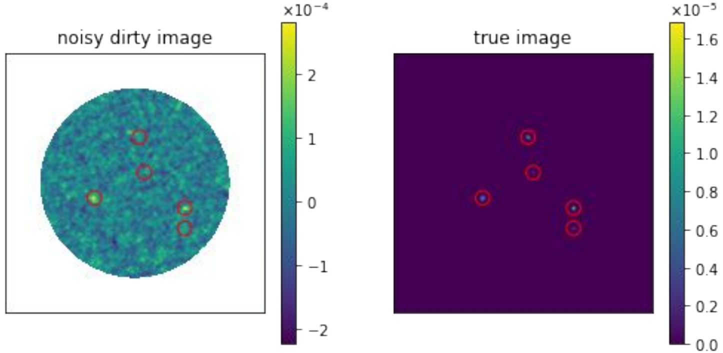
$$\text{Purity} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

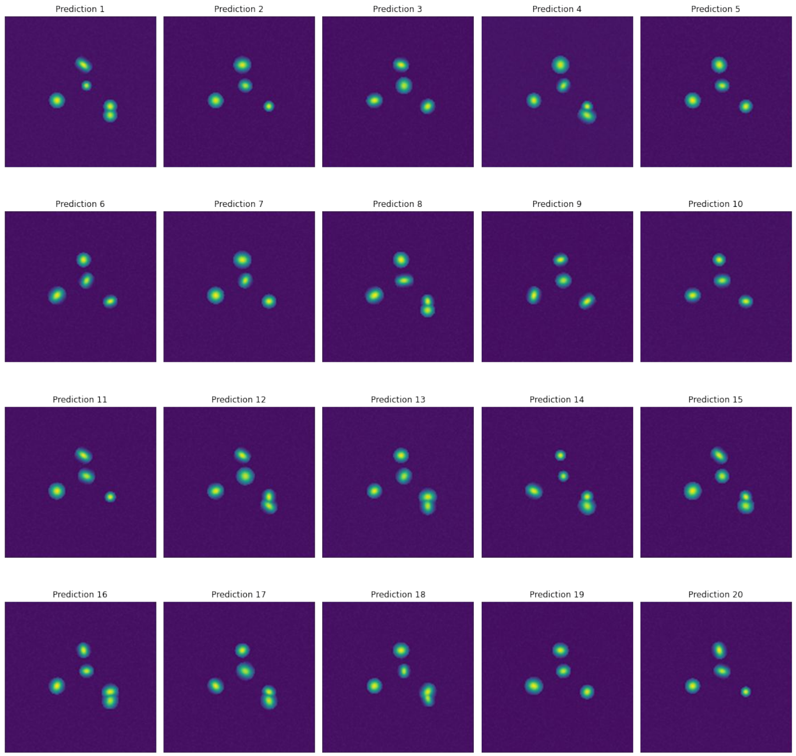
$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Visualization from [O.Taran paper]

Results: Reconstruction

Due to stochasticity of DDPM we can have multiple reconstruction for the same dirty noisy image:

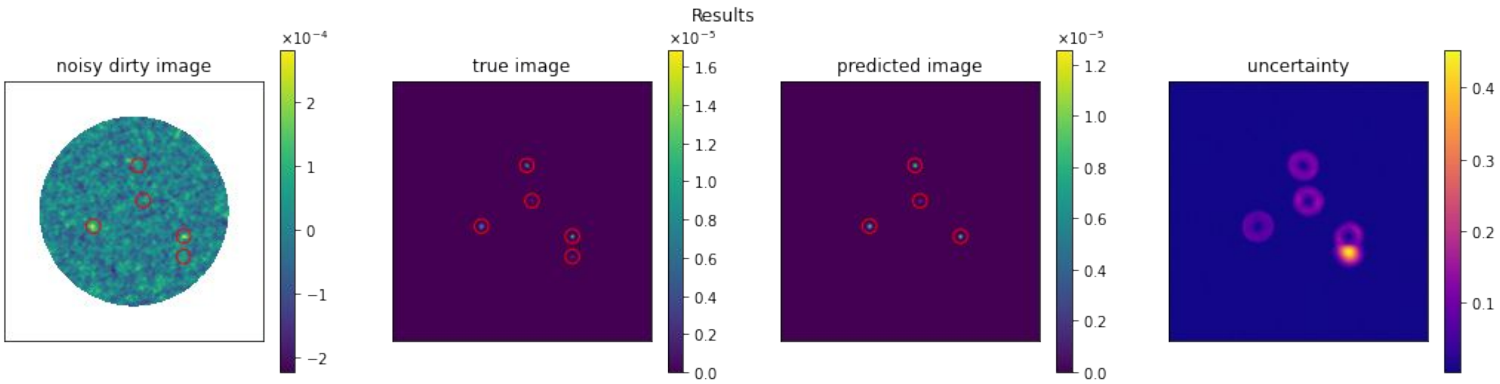




Results: Uncertainty estimation

The predicted image is an aggregation of the 20 outputs using the median function
The uncertainty image is the standard deviation among 20 outputs.

This uncertainty can be used to estimate the robustness of the predictions, as well as identifying possible missed sources.



Results: Multiple Estimates and Aggregation

Table 1: Reconstruction Metrics

Conditioning	**L2 Distance** ($\times 10^{-5}$) ± Error ($\times 10^{-5}$)	**PSNR** ± Error	**SSIM** ± Error
Inverse Fourier of UV samples	3.262 ± 1.618	32.040 ± 4.806	0.9842 ± 0.0087
Inverse Fourier of UV samples, multiple runs, mean	**1.973 ± 0.893**	**33.575 ± 4.916**	**0.9862 ± 0.0081**
Inverse Fourier of UV samples, multiple runs, median	1.984 ± 0.877	33.339 ± 5.164	0.9862 ± 0.0082
CASA dirty noisy images	2.618 ± 1.342	33.245 ± 4.836	**0.9867 ± 0.0082**
CASA dirty noisy images, multiple runs, mean	**1.724 ± 0.816**	**33.673 ± 4.839**	0.9863 ± 0.0079
CASA dirty noisy images, multiple runs, median	1.728 ± 0.829	33.568 ± 4.888	0.9864 ± 0.0079

Table 2: Comparison of Diffusion Model Purity

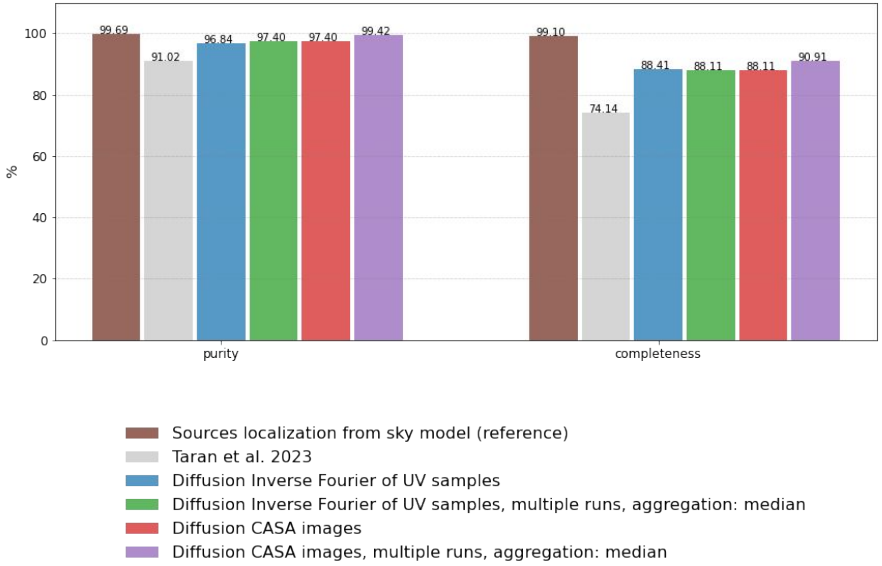
Conditioning	**One Run**	**Multiple Runs**	
		Mean	**Median**
Inverse Fourier of UV-samples	96.84	**98.16**	**98.16**
CASA dirty noisy images	98.25	99.25	**99.42**

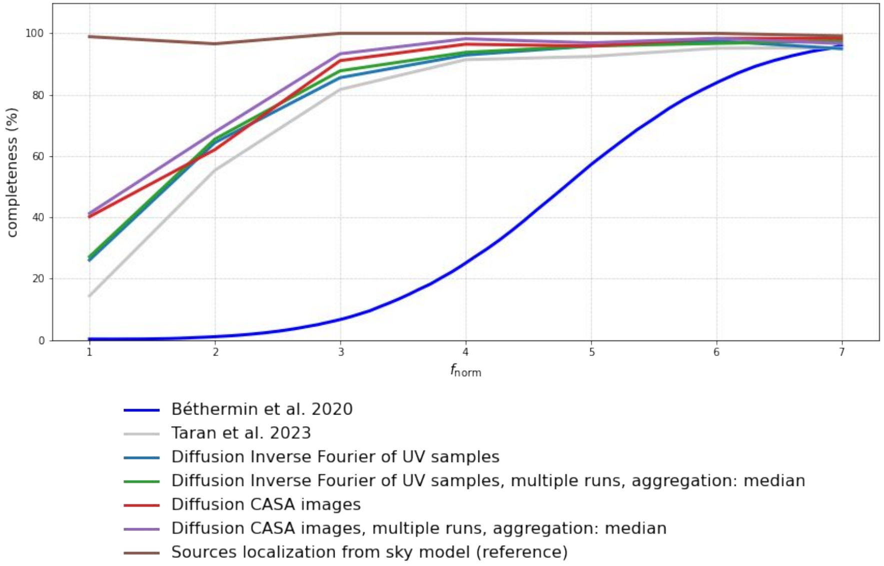
Table 3: Comparison of Diffusion Model Completeness

Conditioning	**One Run**	**Multiple Runs**	
		Mean	**Median**
Inverse Fourier of UV-samples	88.41	88.11	**89.02**
CASA dirty noisy images	89.02	90.23	**90.91**

Comparative Results

To estimate sources positions and other properties we use photutils algorithm, the brown column/line is its performance on true sky images.



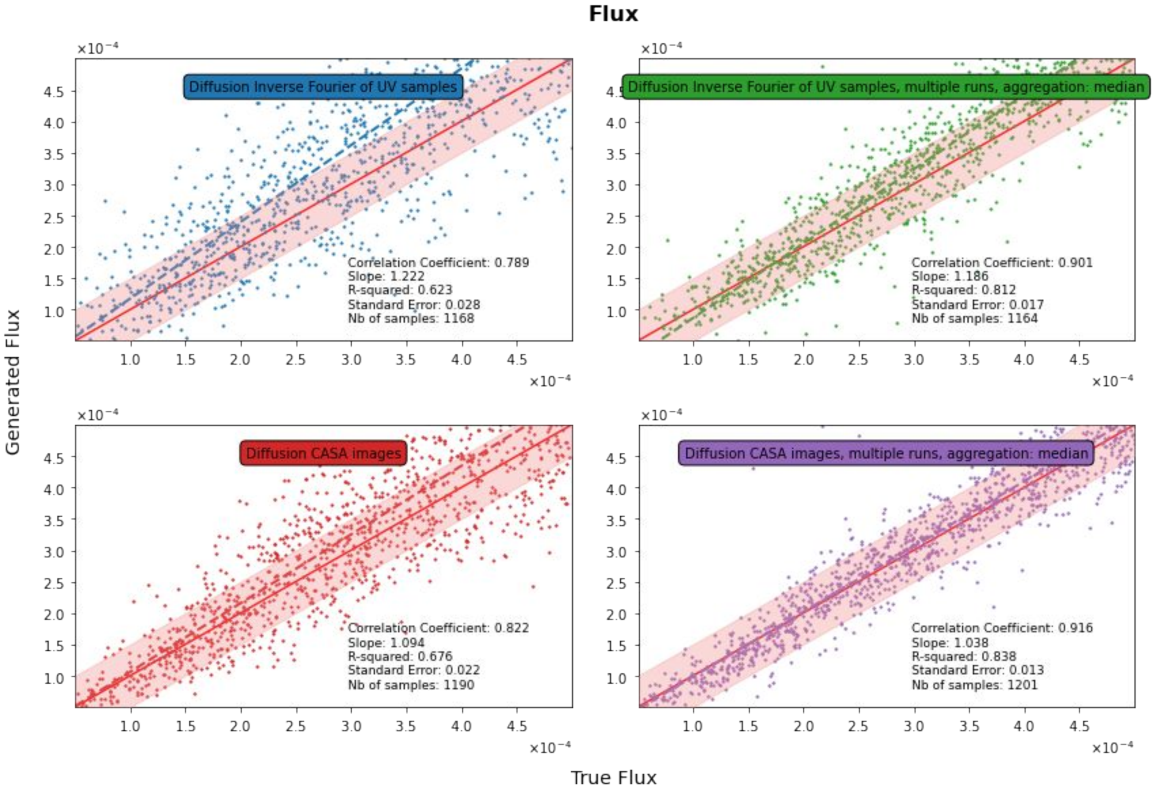


Results: Flux Estimation

For each model we visualized:
- Predicted Flux on y axis
- Real Flux on x axis

The red line is the error which is expected due to the simulation.

The aggregating from multiple estimates gives a significant improvement over one run.



Conclusions and Future Work

In this study we presented:

- application of DDPMs for directly reconstructing Sky Model images
- a normalization technique for Sky Model Images
- improved results for sources localization
- flux estimation

Future Work:

- test on a real data
- investigate the shape estimation
- optimization of the inference time

Thank you!

Sources

1) https://lilianweng.github.io/posts/2021-07-11-diffusion-models/
2) Palette: Image-to-Image Diffusion Models, C. Saharia et al., 2022
3) Challenging interferometric imaging: Machine learning-based source localization from uv-plane observations, O.Taran et al., 2023