RADiff: Controllable Diffusion Models for Generating Radio Maps

Renato Sortino, Thomas Cecconello, Andrea DeMarco, Giuseppe Fiameni, Andrew M. Hopkins, Daniel Magro, Simone Riggi, Eva Sciacca, Concetto Spampinato

Section 1: Introduction

- **Identifying objects** in radio maps is crucial in radio astronomy
  - **Deep learning** models have Been successfully employed to tackle this challenge.
  - **Labeling data** to train deep models is **time-consuming** and requires domain expertise.

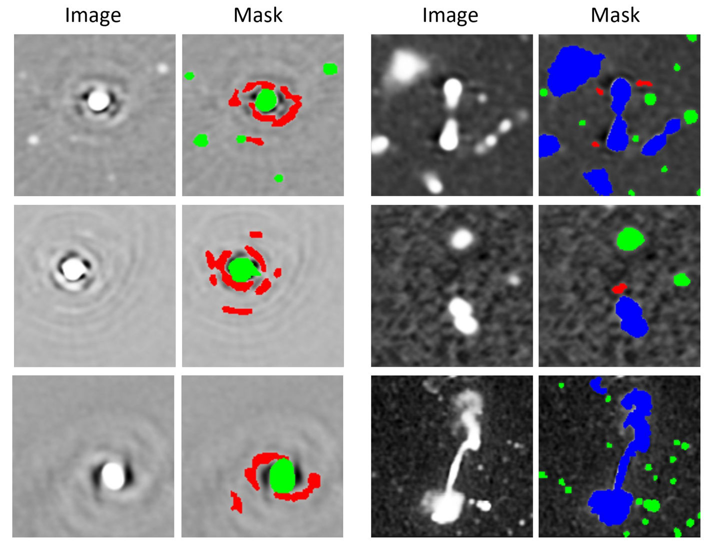

Figure 1:Samples of images and masks used in our dataset. We indicate compact sources in green, extended sources in blue, and spurious sources in red.

- Other methods generate radio images but the generation process is not controllable, especially with multiple conditions.
  - Scully et al. [1] and Kummer et al. [2] propose unconditional Generative Adversarial Networks to augment astronomical datasets for object detection.
- We propose **RADiff**, a controllable generative approach for **augmenting** small-sized labeled datasets with **fully-synthetic** samples.
- We efficiently employ the model to generate high-quality samples and increase the performance of semantic segmentation models

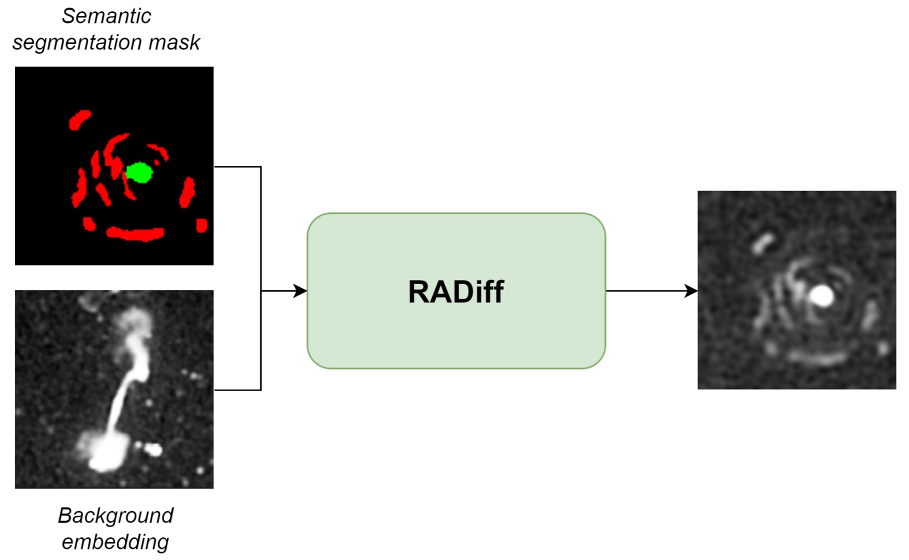

Figure 2: Presentation of the proposed idea. Our model takes as input a segmentation map and an image and generates a synthetic image following the structure defined in the semantic map and with the background features of the image

[1] - J. Scully, R. Flynn, E. Carley, P. Gallagher, M. Daly, Simulating solar radio bursts using generative adversarial networks, Solar Physics 298 (1) (2023) 1–16
[2] - J. Kummer, L. Rustige, F. Griese, K. Borras, M. Br¨uggen, P. L. Connor, F. Gaede, G. Kasieczka, P. Schleper, Radio galaxy classification with wgan-supported augmentation, arXiv preprint arXiv:2206.15131 (2022).

Section 2: Method

- **Latent Diffusion Model (LDM)** [3] architecture
  - *Autoencoder* to project the data to and from a latent space
  - *Diffusion model* to generate synthetic data
  - *Condition Encoder $\mathcal{E}_c$* to inject conditioning information to guide the generation process
    - **Segmentation mask**, describing the shape and location of the objects within the image (concatenated to noise vector)
    - **Image embedding** to control the background pattern (cross-attention between embedding and latent vector).
- **MSE loss** between the noise added in the forward diffusion process, ϵ, and the one predicted in the backward process, ϵ_θ, at each step t, as done in [3]

$$\mathcal{L}_{\text{diff}} = E_{t,x_0,\epsilon,c}\left[\|\epsilon_t - \epsilon_\theta(x_t, t, c)\|^2\right]$$

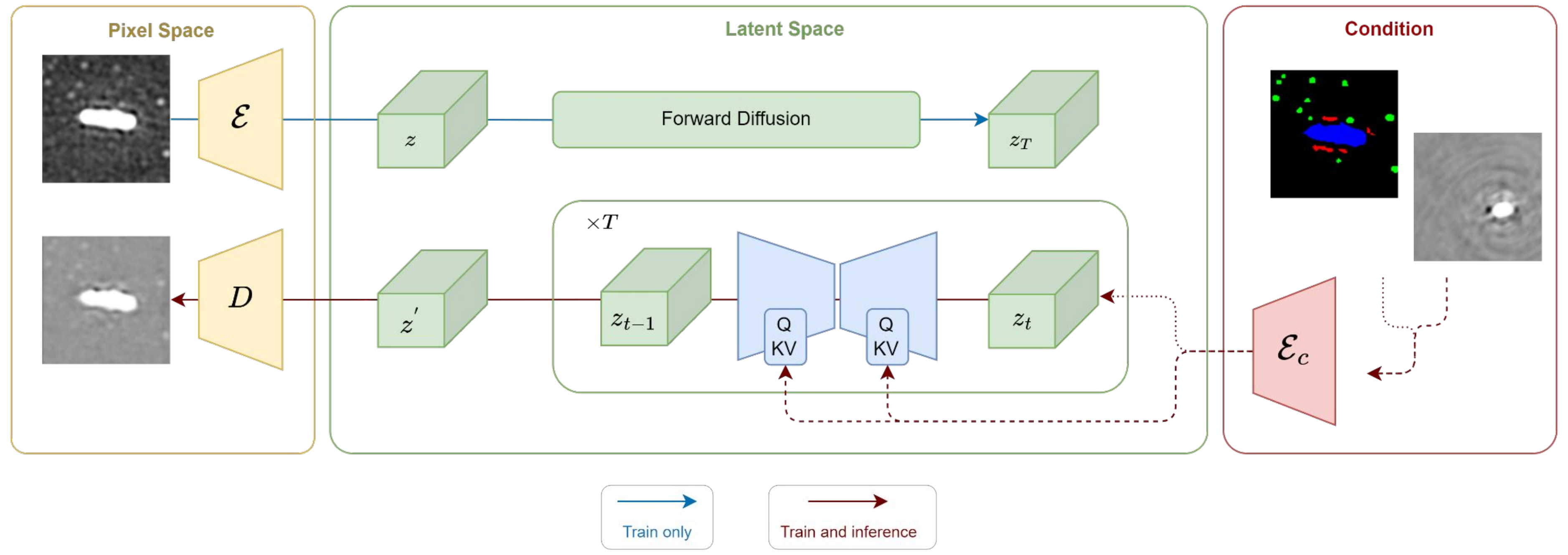

Figure 3: Overview of RADiff, our proposed approach based on diffusion models. The autoencoder projects the data from the pixel space into the latent space z and vice versa. The diffusion model progressively adds noise to the latent vector for a number T of timesteps and then learns to revert this process, estimating z'. The condition encoder $\mathcal{E}_c$ projects the semantic mask and the image information into a latent embedding to be used by the diffusion model at each timestep t. The first one is concatenated to the latent noise vector, while the second one is used in a cross-attention operation with the output of the U-Net intermediate layers.

[3] - R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10684–10695.

Section 3: Results

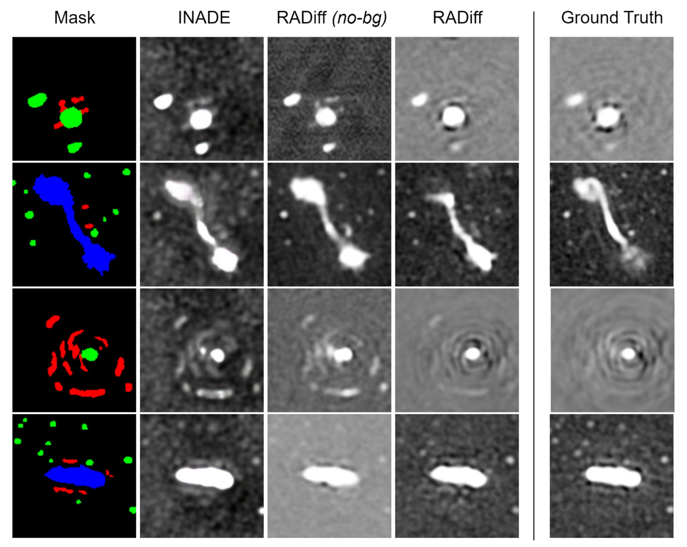

Figure 4: Visual quality of the generated samples. *no-bg* represents our model without the conditioning on the background.

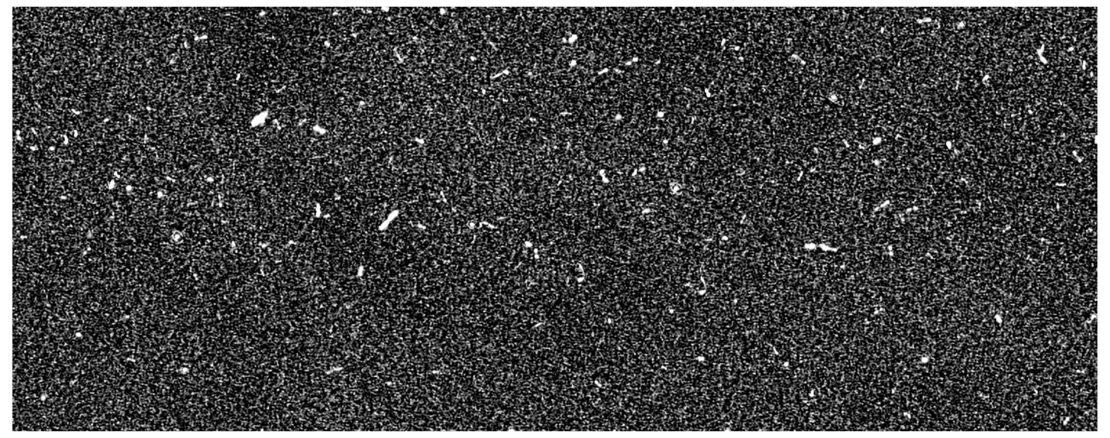

Figure 5: Populating a large-scale radio background noise map (~12K × 15K pixels) with synthetic objects. We filter out the background from the generated images, extract the single objects, scale their value to the physical range of the background and place them randomly on the image, without overlap.

Section 4: Evaluation

Table 1: Semantic segmentation model (Tiramisu [5]) trained on the real SC dataset augmented with synthetic image-mask pairs. We first generate 5,000 masks using DDPM [4] and then feed these masks to RADiff to generate the associated images. The first column represents the objects included in the synthetic masks. All model instances are evaluated on real data.

Augmentation	All classes	Extended	Compact
Compact	72.29%	71.87%	**72.87%**
Extended	**74.85%**	**76.40%**	71.63%
Both	72.89%	74.51%	71.25%
None	71.67%	72.14%	70.72%

Table 2: We evaluate the impact of extending the dataset with synthetic images using real masks. We remove 30\% of image-mask pairs from the training set **SC**, obtaining its reduced version $\mathbf{SC_R}$, and use these masks to generate a dataset of synthetic images $\mathbf{SC_S}$. Best results are highlighted in bold.

Model	All classes	Extended	Compact
SC_R	63.93%	53.02%	58.25%
$+ SC_S$	**70.65%**	**71.64%**	**69.52%**
Synthetic	67.40%	68.68%	58.13%
SC	71.67%	72.14%	70.72%

[4] - J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Advances in Neural Information Processing Systems 33 (2020) 6840–6851.
[5] - C. Pino, et al., Semantic segmentation of radio-astronomical images, in: International Workshop on Artificial Intelligence and Pattern Recognition, Springer, 2021, pp. 393–403.

Section 5: Conclusions

- We employed LDMs to generate **synthetic radio maps** with specific objects **guided** by semantic segmentation masks
- We improved the performance **up to 18%** of a segmentation model by **augmenting** the dataset with synthetic samples
- We showed how the generated objects can be used to populate a real noise background to **accurately simulate** large radio maps.