6^{th} *RFI Conference, Bariloche, Argentina, 14-18 October 2024*

Hierarchical vision transformers for RFI mitigation in radio astronomy

Xiaowei Ouyang(1), Henk Dreuning(1,2), Michael Mesarcik*(2), and Rob V. van Nieuwpoort(3)
(1) Vrije Universiteit Amsterdam, Amsterdam, The Netherlands,
(2) University of Amsterdam, Amsterdam, The Netherlands,
(3) Leiden University, Leiden, The Netherlands

1 Introduction

Radio Frequency Interference (RFI) is an enormous challenge for radio astronomy due to the increasing scale and sensitivity of modern radio telescopes. Furthermore, the growing number of electronic devices emitting RF signals severely hampers the reliability of radio observations, necessitating advanced methods for RFI mitigation. While machine learning methods have shown promise in this area, most existing approaches rely on Convolutional Neural Networks (CNNs), which carry inherent biases in their assumptions about the inference structure and morphology.

In this paper, we explore an alternative approach using hierarchical vision transformers, specifically the Swin-UnetR [1] architecture, to tackle RFI detection in radio astronomy. Unlike CNNs, transformers do not assume any specific feature morphologies, allowing them to potentially identify RFI more flexibly. This being said, training large transformer models is computationally expensive; often requiring significant resources. To overcome this, we employ pipeline parallelism with CAPSlog [2] as the partitioning strategy, to enable more efficient use of both computational resources and memory to facilitate scalability for future studies as larger datasets become available.

This is the first known application of transformer-based architectures for RFI detection in radio astronomy, representing a shift from traditional methods. Our experiments demonstrate that the Swin-UnetR architecture achieves state-of-the-art performance, outperforming existing methods on three key metrics using data from the LOFAR [3] radio telescope. This novel approach not only advances RFI mitigation but also sets the stage for broader adoption of transformer-based models in radio astronomy.

2 Shifted window vision transformers

The *S*hifted *Win*dow (SWIN) Transformer architecture [1], has gained significant attention for its approach to handling spatial hierarchies in vision tasks. Unlike traditional transformer architectures that maintain a fixed global view, the Swin Transformer operates by gradually merging patches at multiple scales, allowing it to capture both local and global information.

The Swin-UnetR model, which we apply to RFI detection in radio astronomy, combines the down-sampling mechanism of Swin Transformer with the up-sampling architecture of U-Net. The down-sampling stage uses the Swin Transformer's patch-merging technique to reduce the spatial dimensions while extracting essential features. The up-sampling stage, inspired by U-Net, allows the model to generate high-resolution output, which is necessary for accurately identifying RFI patterns within the radio spectrum.

3 Evaluation

The models in this paper are trained and evaluated on the spectrograms from a publicly available dataset generated by the LOFAR telescope [4]. This dataset contains 7,500 training samples labelled by AOFlagger [5] and 109 test samples labelled by human experts. To manage the dataset's size, we down-sample the input spectrograms, reducing them to approximately 10 GB, and then crop them into 512×512 pixel images.

We use three configurations of the Swin-UNETR model, each varying a different hyper-parameter of the Swin Transformer architecture. We train each model for 100 epochs using the AdamW optimiser with early stopping, additionally we use 20 cosine warming-up epochs and a learning rate of 10^{-5} and set the decay rate at 0.001. Table 1 presents the number of parameters and the computational resources required for training these configurations.

Model Configuration	# Parameters	Feature size	# Transformer blocks per stage	TFLOPs
Swin-6M	6.5 M	24	(2, 2, 18, 2)	1.105
Swin-100M	102.1 M	96	(2, 2, 18, 2)	12.334
Swin-400M	408.1 M	192	(2, 2, 18, 2)	67.025

Table 1. Parameters and TFLOPs per iteration for each Swin-UnetR configuration

We compare the RFI detection performance of the Swin-UnetR model with existing deep learning solutions, specifically NLN [4] and U-Net [6] as well as AOFlagger [5]. In Table 2 we show the model's detection performance using three metrics: the AUPRC (Area Under Precision–Recall Curve) score, the AUROC (Area Under Receiver Operating Characteristic Curve) score, and the F1-score.

It can be seen that the transformer-based models offers significant improvements across all metrics relative to the other models. Interestingly, the Swin-400M model exhibits higher detection performance compared to both Swin-100M and Swin-6M, suggesting that model size correlates with improved accuracy. Overall, these findings provide insights into the performance dynamics among Swin-UnetR variants and their efficacy against other models.

Metric	AOFlagger[5]	U-Net [6]	NLN [4]	Swin-6M	Swin-100M	Swin-400M
AUROC	0.7883	0.8017	0.8622	0.9711	0.9743	**0.9771**
AUPRC	0.5716	0.5920	0.6216	0.6783	0.6831	**0.6938**
F1-Score	0.5698	0.5876	0.5114	0.6302	0.6281	**0.6401**

Table 2. RFI detection performance in AUROC, AUPRC and F1 score for each model, where bold is best.

4 Conclusion

In this paper we have demonstrated that transformer-based architectures, such as Swin-UnetR, demonstrate notable improvements over traditional architectures for applications in radio astronomy RFI detection. Our findings indicate that model performance tends to improve with the number of parameters in the Swin-UnetR architecture. Larger Swin-UnetR models consistently achieve better performance metrics compared to their smaller counterparts, supporting the notion that increased model capacity allows for the capture of more complex patterns. Additionally, by utilising a parallel method to train these larger architectures, we effectively facilitated more efficient handling of large-scale datasets and potentially reducing training times, which will be further demonstrated in the full paper. This approach could be valuable in broader contexts within radio astronomy, especially for tasks requiring many computational resources. These results not only show the potential of transformer-based models but also open pathways for further research into scalable training techniques and their applicability to RFI detection and other related fields.

References

[1] A. Hatamizadeh, V. Nath, Y. Tang, *et al.*, "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images," pp. 272–284, 2022.

[2] H. Dreuning, A. B. Liokouras, X. Ouyang, H. E. Bal, and R. V. Van Nieuwpoort, "CAPSlog: Scalable Memory-Centric Partitioning for Pipeline Parallelism," in *2024 32nd Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, pp. 17–25, IEEE, 3 2024.

[3] M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, *et al.*, "LOFAR: The LOw-Frequency ARray," *Astronomy & Astrophysics*, vol. 556, p. A2, 8 2013.

[4] M. Mesarcik, A.-J. Boonstra, E. Ranguelova, and R. V. van Nieuwpoort, "Learning to detect radio frequency interference in radio astronomy without seeing it," *Monthly Notices of the Royal Astronomical Society*, vol. 516, pp. 5367–5378, 9 2022.

[5] A. R. Offringa, J. J. Van De Gronde, and J. B. Roerdink, "A morphological algorithm for improving radio-frequency interference detection," *Astronomy and Astrophysics*, vol. 539, 2012.

[6] J. Akeret, C. Chang, A. Lucchi, and A. Refregier, "Radio frequency interference mitigation using deep convolutional neural networks," *Astronomy and Computing*, vol. 18, pp. 35–39, 2017.