3rd URSI AT-AP-RASC, Gran Canaria, 29 May – 3 June 2022

Astrophysically Informed Priors for Self-Attention Classification in Radio Astronomy

Micah Bowles*(1) and Anna Scaife(1)(2)
(1) Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester, M13 9PL, UK
(2) The Alan Turing Institute, Euston Road, London, NW1 2DB, UK

1 Extended Abstract

Automated object classification is essential in the era of big data radio astronomy in order to produce final science products from large-scale astronomical surveys in a timely fashion. However, in spite of the increasingly wide-spread adoption of deep learning for such tasks in other fields, the astronomical community still tends to be apprehensive of such solutions as they are often perceived to be "black box" methods, the results from which cannot be easily interpreted. In this work we demonstrate how deep-learning models that incorporate a *self-attention* mechanism can provide output meta-data showing how different structures within a radio image are used by the classification. Such outputs are commonly referred to as *attention maps*. Furthermore, we explore how incorporating informed priors on astrophysical structure related to equivariance over the isometries of 2-dimensional Euclidean space might further benefit such deep-learning classification models for radio astronomy.

We make use of the model and MiraBest radio galaxy data set introduced in [1]. This model is a common form of convolutional neural network designed for image classification, modified with custom feature-wise self-attention gates developed specifically for radio astronomy classification. To accommodate additional equivariant priors, we then extend this model to incorporate various forms of E(2)-equivariant convolution operations [2].

As well as maintaining the translational equivariance of standard convolution operations, the additional priors introduced in this work are equivariant to user determined sets of rotations and reflections. In our experiments we evaluate multiple orders of cyclic (rotation) and dihedral (rotation and reflection) equivariance, similar to [3]. Our experiments show minor classification performance improvements from using equivariant models over the standard convolutional baseline. However, by using the resultant attention maps to evaluate which features the models attend, we show that the attention maps produced by equivariant models more closely and clearly align with the features an expert radio astronomer would commonly use to classify the respective class of object. This is evaluated by visualising the mean pixel attention across a heavily augmented test set and comparing it to where class-specific features become dominant. This is possible due to the respective class definitions, various instrumental and data specific selection effects.

Our experiments show promise for the use of E(2)-equivariant operations when classifying radio astronomical objects. They also show the benefits of using priors informed by our astrophysical understanding of astronomical objects as a way to improve the relatability of self-attention maps as an explainable AI tool.

References

[1] Bowles, M., Scaife, A. M. M., Porter, F., Tang, H., and Bastien, D. J., “Attention-gating for improved radio galaxy classification”, *Monthly Notices of the Royal Astronomical Society*, vol. 501, pp. 4579–4595, 2021.

[2] Bowles, M., Bromley, M., Allen, M. and Scaife, A. M. M., “E(2) Equivariant Self-Attention for Radio Astronomy”, Fourth Workshop on Machine Learning and the Physical Sciences, *35th Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

[3] Scaife, A. M. M. and Porter, F., “Fanaroff-Riley classification of radio galaxies using group-equivariant convolutional neural networks”, *Monthly Notices of the Royal Astronomical Society*, vol. 503, pp. 2369–2379, 2021.

2 Acknowledgements

The authors gratefully acknowledge support from the UK Alan Turing Institute under grant reference EP/V030302/1. MB gratefully acknowledges support from the UK Science & Technology Facilities Council (STFC).