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

Radio source-component association for large-scale sky surveys with Region-based Convolutional Neural Networks

Rafaël I.J. Mostert*(1)(2)(3), Huub J.A. Röttgering(1), Kenneth J. Duncan(1)(4) and Raffaella Morganti(2)(5)
(1) Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, The Netherlands
(2) ASTRON, the Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands
(3) Leiden Institute of Advanced Computer Science, Niels Bohrweg 1, Leiden, The Netherlands
(4) SUPA, Institute for Astronomy, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK
(5) Kapteyn Astronomical Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

Radio loud active galactic nuclei (RLAGN) are often morphologically complex objects that can consist of multiple, spatially separated, components. Only when spatially separated radio components are correctly grouped together can we start to look for the corresponding optical host-galaxy and infer physical parameters like the size and luminosity of the radio object. Current radio detection software is not designed to group these spatially separated components together, but this task of radio component association is easily solved through visual inspection. However, applying this manual process to all the hundreds of thousands of well-resolved RLAGN that appear in the images from the Low Frequency Array (LOFAR) [1] Two-metre Sky Survey (LoTSS) [2] at 150 MHz or future Square Kilometre Array surveys, is a daunting, time-consuming process, even when delegated to many humans.

Using a machine learning approach, we aim to automate the radio component association of large ($>$ 15 arcsec) radio components. We cast the association problem into a classification problem and train an adapted Fast region-based convolutional neural network (Fast R-CNN) [3] to mimic the expert annotations from the first LoTSS data release. We implement rotational data augmentation to reduce over-fitting and simplify the association process by removing unresolved radio sources that are likely unrelated to the considered large radio components using an XGBoosted decision tree.

Qualitatively, the predictions are able to correctly associate the radio components of a morphologically wide range of sources: nearby starforming galaxies, typical edge-brightened double-lobed RLAGN (even those where the emission connecting the lobes fell below the noise), edge-darkened RLAGN, all of different sizes and different signal-to-noise ratios. Quantitatively, our model predicts the same associations as visual inspection for $85.9\% \pm 0.6$ of the large and bright ($>$ 10 mJy) radio components in the LoTSS first data release. This accuracy goes down to roughly 80% for the large and faint radio components, with an uncertainty of a few percent as visual inspection is increasingly difficult for fainter sources. We show that 5% of the accuracy of our automated approach can be attributed to the combined effect of the rotational data augmentation and unresolved radio source removal.

Although there is room for improvement, our automated approach already performs component associations with an accuracy that is comparable to that attained by public crowd-source efforts. Our method can replace manual radio component association and serve as a basis for either automated radio morphology classification [4] or automated optical host-identification. This work opens up an avenue to study the completeness and reliability of samples of radio sources with extended complex morphologies.

References

[1] van Haarlem, M. P., et al., "LOFAR: The LOw-Frequency ARray", *Astronomy and Astrophysics*, vol. **556**, 2013. doi:10.1051/0004-6361/201220873.

[2] Shimwell, T. W., et al., "The LOFAR Two-metre Sky Survey. I. Survey description and preliminary data release", *Astronomy and Astrophysics*, vol. **598**, 2017. doi:10.1051/0004-6361/201629313.

[3] R. Girshick, "Fast R-CNN", *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[4] Mingo, B., "Revisiting the Fanaroff-Riley dichotomy and radio-galaxy morphology with the LOFAR Two-Metre Sky Survey (LoTSS)", *MNRAS*, vol. **488**, no. 2, pp. 2701–2721, 2019. doi:10.1093/mnras/stz1901.