ВАК 2024 MODERN ASTRONOMY: FROM THE EARLY UNIVERSE TO EXOPLANETS AND BLACK HOLES
300 years Russian Academy of Sciences

Classification of jets in active galactic nuclei using machine learning

D. Zagorulia[1] and M. Lisakov[2,3]

[1] Moscow Institute of Physics and Technology, Dolgoprudny, 141701 Russia
[2] Institute of Physics, Pontifical Catholic University of Valparaíso, Casilla, Valparaíso, 4059 Chile
[3] Astro Space Center of P. N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow, 117997 Russia

Abstract. This work explores the application of machine learning techniques to classifying jetted active galactic nuclei (AGN) based on Very-Long-Baseline Interferometry (VLBI) observations at frequencies 1–90 GHz. Building upon previous work by Fanaroff and Riley, who classified relativistic jets in radio galaxies on kiloparsec scales, we extend this classification to parsec scales, closer to the central supermassive black hole. This approach enables detailed study of jet spatial structures and can help enhancing accuracy in global positioning systems. We define four morphological classes: single Gaussian source, double Gaussian source, and sources with single or double-sided jets. Synthetic models of AGN jets were generated to create a training dataset for a convolutional neural network (CNN). The CNN was trained on these synthetic data and subsequently applied to classify 130 thousand AGN jet images from the Astrogeo database. The distribution of images into designated classes, predicted by CNN, qualitatively matches the expected outcome.

Keywords: galaxies: active; quasars: general; techniques: high angular resolution; methods: machine learning

DOI: 10.26119/VAK2024.018

1 Introduction

About 10 % of all normal galaxies host an active nucleus, a bright central part that emits more radiation than the stars within it due to the accretion of matter onto a supermassive black hole. The total estimated power of an active galactic nucleus (AGN) can reach 10^{48} erg/s. About 10 % of these active nuclei eject narrowly focused ultrarelativistic streams of plasma (jets). All the energy powering these jets is generated in an incredibly compact region less than one parsec in size. The relativistic plasma ejections themselves can propagate far beyond the host galaxy, reaching up to hundreds of kiloparsecs (Blandford et al. 2019).

The kiloparsec structure of radio-loud galaxies was studied earlier by Fanaroff and Riley (Fanaroff & Riley 1974), who distinguished two classes of jetted radio sources, whose luminosity decreases with increasing distance along the jet from the central galaxy (FR-I) and whose luminosity, on the contrary, increases toward the edges of the jet (FR-II). It turned out that this difference in morphology correlates with the total luminosity of these objects.

But with the advent of Very-Long-Baseline Interferometry (VLBI), images on parsec scales appeared (Lister et al. 2018), and what is the most interesting is that the direction of jets on small and large scales do not always coincide. Parsec and kiloparsec scales are separated in both space and time. Thus the changes in the jet morphology could serve as an indication of the evolution of the intrinsic direction of the jet and possibly the accretion regime.

2 Data preparation

In the Astrogeo database[1] (Astrogeo VLBI FITS image database), more than 130 thousand interferometric measurements and images of 20 thousand AGNs have been collected, which were mainly observed for astrometry and geodesy purposes. However, image processing was performed automatically and we discovered that some images were defective. To exclude these images from the further processing, the criterion of a defective image was formulated, which consisted of two parts:

1. Mismatch between the coordinates of the map centre and the coordinates of the maximum intensity of the image;
2. Signal-to-noise ratio less than 10.

Such properties are typical in the presence of amplitude and phase calibration errors of individual telescopes, incorrect pointing or poor automatic processing. [2]

[1] http://astrogeo.org/vlbi_images
[2] This simple criterion immediately filtered out 2.3 % of images.

3 Morphological classification

3.1 Synthetic images

For image classification, convolutional neural networks (CNNs) (LeCun et al. 1989) have proven their efficiency. Due to the fact that one of the desired results is to determine the probability of the morphological classes for each image of AGN, a synthetic image dataset was developed to supervise learning of the CNN. Real images are a spatial convolution of the brightness distribution in the sky with the corresponding point spread function (PSF).

Therefore, from the filtered sample of AGN images, only those with known PSF parameters (major and minor axes, position angle) were selected. Next, clustering was performed in this parameter space using the K-MEANS (Forgy 1965; Lloyd 1982) algorithm to identify 10 characteristic values. Two of which were eventually discarded due to the fact that the major axis of the PSF was 2–3 times larger than the other clusters. The centroids of remaining clusters were used as Gaussian filters to convolve with 4 model sources. The model sources (Fig. 1, top row) included: a Gaussian source, double Gaussian source, a single- and double-sided jet sources (Gaussian with a line segment).

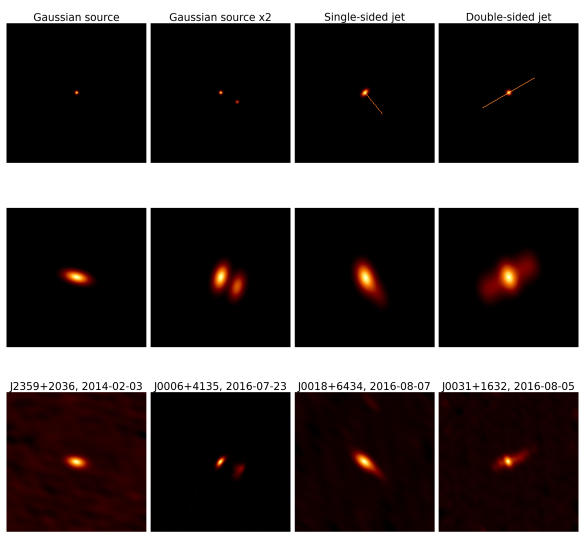

Fig. 1. Top row – model sources, middle row – model sources after convolution with PSF, bottom row – examples of real objects.

These morphological classes were chosen for their utility in both astrometry and astrophysics. In astrometric applications, Gaussian sources can improve the accuracy of global positioning systems. They also allow for the precise localization of a source behind the core, rather than being affected by the local brightening of the jet in double Gaussian objects. For astrophysical research, sources with single and double-sided jets are especially valuable. They help in studying core shifts at different frequencies and investigating the correlation between source morphology and total luminosity.

3.2 Image predictions

After training developed CNN[3], we used it for classification of 130 thousand images from Astrogeo[4]. We got the following distribution: 58.6 % – Gaussian sources, 11.5 % – double Gaussian sources, 27.5 % and 2.4 % for single- and double-sided jet sources respectively. For comparison we have manually classified a little over 500 random images and obtained the following result: 41.2 % – Gaussian sources, 11.8 % – double Gaussian sources, 42.6 % and 4.4 % for single- and double-sided jet sources.

4 Summary

We have developed a CNN model for distinguishing images of AGNs by their spatial structures and obtained predictions for Astrogeo database. Currently, we are working on improving obtained predictions.

In the near future, we plan to investigate correlation of luminosity and AGN morphology on parsec scales and morphological dependencies between VLBI images (Petrov 2016, 2021) and VLA (Very Large Array) images. For the latter, we will use data from the first two rounds of VLASS observations (Lacy et al. 2020), which are already available for public use. Developed software will be used to analyse the morphology at kiloparsec scales.

References

Blandford R., Meier D., and Readhead A., 2019, Annu. Rev. Astron. Astrophys, 57, p. 467
Fanaroff B. and Riley J., 1974, MNRAS, 167, p. 31P
Forgy E., 1965, Biometrics, 21, p. 768
Lacy M., Baum S., Chandler C., et al., 2020, PASP, 132, id. 035001
LeCun Y., Boser B., Denker J., et al., 1989, Neural Comput., 1, p. 541
Lister M., Aller M., Aller H., et al., 2018, ApJ Supp. Serires, 234, p. 12
Lloyd S., 1982, IEEE Trans. Inf. Theory, 28, p. 129
Petrov L., 2016, The 13th EVN Symposium and Users Meeting Proceedings, arXiv:1610.04951
Petrov L. 2021, AJ, 161, p. 14

[3] https://github.com/dimicorn/astrogeo_image_classifier
[4] We classified images of AGNs and different images of the same AGN could be classified differently.