Simulation-to-Real Learning for VLBI Jet Morphology Classification

Dmitrii Zagorulia [1 2] **Mikhail Lisakov** [3]

Abstract

We investigate simulation-to-real learning for morphological classification of parsec-scale active galactic nucleus (AGNs) jets observed with Very Long Baseline Interferometry (VLBI) across frequencies from 1 to 90 GHz. To overcome the scarcity of labelled VLBI data, we combine real observations with physically motivated synthetic images generated from the Blandford–Königl jet model via relativistic ray tracing and realistic interferometric reconstruction. We compare supervised learning, self-supervised pretraining, and domain adaptation strategies based on a ResNet-18 backbone, and show that explicit domain adaptation substantially improves transfer to real VLBI observations, achieving an F_1-score of 0.86 on manually labelled Astrogeo sources. The trained model is subsequently applied to more than 130 000 unlabelled AGN images, enabling one of the largest parsec-scale morphology studies to date. The inferred populations reveal statistically significant differences in radio brightness between compact and extended sources, suggesting a connection between parsec-scale jet structure and observed radio emission properties.

1. Introduction

About 10% of all normal galaxies host an active galactic nucleus (AGN), a compact region powered by accretion onto a supermassive black hole. A fraction of AGNs, commonly referred to as radio-loud AGNs, launch narrowly collimated ultrarelativistic plasma outflows (jets) whose total power can reach $10^{48}\,\mathrm{erg/s}$. Although the energy powering these jets is generated within a region smaller than one parsec, the jets themselves can propagate far beyond the host galaxy, extending to hundreds of kiloparsecs (Blandford et al., 2019).

The large-scale morphology of radio galaxies was studied by Fanaroff & Riley (1974), who introduced the FR-I and FR-II classes based on the distribution of radio emission along the jets. FR-I sources are brightest near the central galaxy and fade outward, while FR-II sources exhibit bright terminal hotspots. This morphological distinction was later shown to correlate with the total radio luminosity of the sources (Singal & Rajpurohit, 2014).

On parsec scales, however, the morphology of AGN jets remains significantly less explored, largely due to the limited availability of labelled Very Long Baseline Interferometry (VLBI) data. Modern VLBI surveys contain tens of thousands of radio images across various frequencies, making manual morphological classification impractical. This creates a natural machine learning challenge: how to learn robust morphological classifiers from largely unlabelled astronomical datasets.

The choice of binary classification is motivated both by the physical properties of parsec-scale radio sources and by the limitations of current VLBI observations. At parsec scales, the dominant observable distinction is whether a source appears compact or exhibits a resolved jet structure. More detailed morphological taxonomies are difficult to define robustly because many apparent differences arise from instrumental effects such as limited angular resolution, sensitivity, sparse uv-coverage, and reconstruction artifacts, rather than from intrinsic physical diversity. By reducing the problem to two broad classes — compact sources and sources with extended jets — we minimize ambiguities associated with subjective visual labelling and improve the robustness of the classification. At the same time, this binary separation remains astrophysically meaningful, since it reflects differences in geometry, Doppler boosting, and potentially the physical conditions within the relativistic jets. Such a simplified formulation also makes it possible to train reliable machine learning models on synthetic data and transfer them to real observational datasets despite the substantial domain gap between simulations and VLBI images.

In this work, we investigate the morphological classification

[1]Phystech School of Applied Mathematics and Computer Science, Moscow Institute of Physics and Technology, Dolgoprudny, Russia [2]Laboratory of Heavy Quarks and Leptons, Lebedev Physical Institute of Russian Academy of Sciences, Moscow, Russia [3]Institute of Physics, Pontifical Catholic University of Valparaíso, Valparaíso, Chile. Correspondence to: Dmitrii Zagorulia <zagorulia.ds@phystech.edu>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

of parsec-scale AGN jets in VLBI images using deep learning. We define two classes: compact sources and sources with extended jets. To address the scarcity of labelled observational data, we combine simulated jet images generated from the Blandford–Königl model with real VLBI observations and investigate supervised learning, self-supervised learning, and domain adaptation strategies for simulation-to-real (*sim2real*) transfer. We then apply the best-performing model to approximately 130 000 images from the Astrogeo VLBI database and study the relationship between parsec-scale morphology and radio brightness.

The main contributions of this work are threefold. First, we construct a large synthetic dataset of parsec-scale VLBI jet images generated from physically motivated Blandford–Königl simulations with realistic interferometric reconstruction, providing labelled source-domain data for morphology classification. Second, we perform a systematic comparison of supervised learning, self-supervised learning, and domain adaptation approaches for sim2real transfer in VLBI imaging, demonstrating the limitations of standard pretraining strategies and the benefits of explicit domain alignment. Third, we show that Sinkhorn-based domain adaptation (SIDDA) substantially improves generalization from synthetic to real VLBI observations. Finally, we apply the resulting classifier to previously unlabelled Astrogeo images, enabling large-scale statistical studies of parsec-scale AGN morphology and its association with observed radio properties.

2. Data

2.1. Source (Blandford-Königl simulation)

We use the analytical Blandford-Königl (BK) model (1979) to generate synthetic images of radio sources. The BK model describes a steady conical relativistic jet emitting self-absorbed synchrotron radiation. Different radio frequencies become optically thin at different distances from the black hole, producing a flat radio spectrum and a compact "radio core".

Synthetic images were produced via relativistic ray tracing of the BK model emission, followed by realistic sampling of the interferometric uv-plane and image reconstruction using the CLEAN algorithm (Högbom, 1974), thereby closely reproducing the characteristics of real VLBI observations. This approach (Pashchenko & Plavin, 2019) enables the generation of images with known morphological classes, eliminating the need for manual labelling and reducing the impact of human bias and error.

The final source dataset contains 110 833 compact sources and 110 277 sources with extended jets, for details of AGN jet generation see Appendix A.

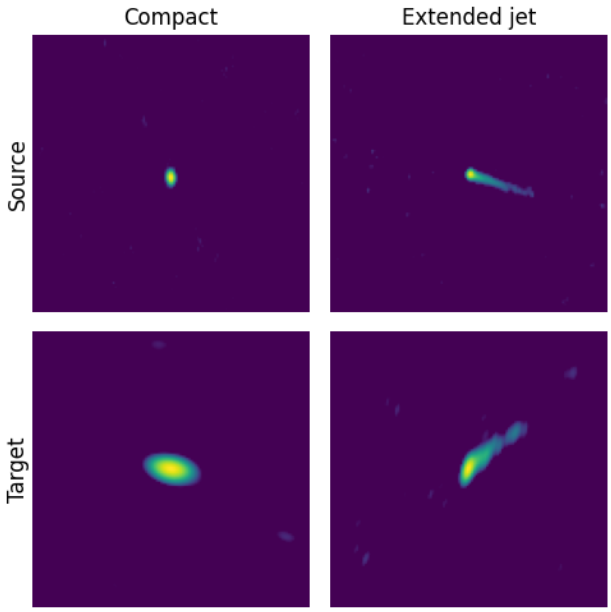

Figure 1. **Example source and target data.** The source domain consists of synthetic VLBI images generated from the Blandford-Königl jet model, while the target domain contains real VLBI observations from the Astrogeo database.

2.2. Target (Astrogeo)

The target set comprises 134 046 images from the Astrogeo VLBI FITS image database[1] (Petrov & Kovalev, 2025), which corresponds to 20 494 AGNs. During preprocessing, we identified a subset of defective images produced by failures in automatic calibration or image reconstruction. To exclude these images from further processing, a two-part criterion for identifying defective images was introduced, following the approach of (Zagorulia & Lisakov, 2024):

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.

These properties are typically associated with amplitude or phase calibration errors, incorrect telescope pointing, or failures in automatic image processing. This simple criterion immediately filtered out 2 877 ($\sim 2\%$) of images, which left us with 131 169 images for 19 817 AGNs as final target set.

One of the main challenges of this work is that the target dataset is largely unlabelled. One possible approach would

[1] http://astrogeo.org/vlbi_images

be to obtain annotations through volunteer-based labelling, as in the Galaxy Zoo project (Lintott et al., 2008), reducing the problem to a standard supervised learning task. Instead, we adopt a more challenging setting: a small subset of 472 images was manually labelled and reserved exclusively for testing, while the remaining target images were kept unlabelled. Training therefore relies primarily on the synthetic source dataset with *a priori* known labels. This choice was motivated by the long-term goal of studying correlations between the physical parameters of synthetic jets and the observed morphology of real VLBI sources.

The labelled test set consists of 472 randomly selected Astrogeo images. All images were manually inspected and labelled by the authors according to the binary morphology scheme adopted in this work. Sources were assigned to either the compact class (224 images) or the extended-jet class (248 images). To reduce subjectivity in the labels, sources with ambiguous morphology or insufficient visual evidence for a confident classification were excluded from the labelled sample. This labelled subset was reserved exclusively for evaluation and was not used during model training, hyperparameter tuning, or domain-adaptation training.

3. Methods

3.1. Architecture

All experiments in this work use the ResNet-18 architecture (He et al., 2016). ResNet-18 was selected because it provides a good balance between simplicity, computational efficiency, and classification performance. Although more recent architectures such as ConvNeXt (Liu et al., 2022b), EfficientNet (Tan & Le, 2019), Swin (Liu et al., 2021; 2022a), and ViT (Kolesnikov et al., 2021) can achieve higher performance on large-scale image classification benchmarks, exploring alternative architectures is outside the scope of this work.

Since VLBI images are single-channel intensity maps, the first convolutional layer of ResNet-18 was modified to accept one-channel input, and the final MLP layer was replaced with a single-output layer for binary classification.

Since the target Astrogeo dataset is largely unlabelled, this work focuses on exploring different training strategies, including supervised learning, self-supervised learning, and domain adaptation.

3.2. Supervised Learning

As a baseline, a ResNet-18 model was trained from randomly initialized weights on synthetic images (SCRATCH), under the assumption that the synthetic images adequately approximate the morphology of real VLBI observations and that the synthetic and real data are drawn from the same underlying distribution. The trained model was subsequently evaluated on the held-out test set.

In addition to the baseline model, two ResNet-18 models pretrained on the ImageNet dataset (Russakovsky et al., 2015) were fine-tuned. In the first configuration (FT1), the first two convolutional blocks were frozen during training. In the second configuration (FT2), all layers were unfrozen and updated during training.

3.3. Self-supervised Learning

In this work, self-supervised learning (SSL) is motivated by the need to incorporate unlabelled Astrogeo images directly into the training of the ResNet-18 model. Although the supervised baseline results are reasonably strong, they rely on synthetic training data and therefore remain affected by the domain gap between simulated and real radio observations. Recent evidence further supports this choice: Cecconello et al. (2024) showed that SSL-based models consistently outperform ImageNet-pretrained models in radio-galaxy morphology classification, with SimCLR (Chen et al., 2020), BYOL (Grill et al., 2020), and All4One (Estepa et al., 2023) generally among the best-performing methods. Based on these results, SimCLR was selected as the SSL approach used in this work.

After the SSL pretraining stage, an MLP classifier was attached to the encoder and fine-tuned on the labelled source dataset. As in the supervised setting, two fine-tuning configurations were explored, freezing the first two convolutional blocks in the first configuration (SSL1), and fully unfrozen in the second configuration (SSL2).

3.4. Domain Adaptation

Although self-supervised pretraining on unlabelled target images enables the encoder to learn representations informed by real observational data, it does not explicitly align the source and target distributions. Contrastive objectives such as SimCLR encourage invariance to augmentations within a single domain, but do not guarantee overlap between latent representations of simulated and real observations. As a result, the encoder may still retain domain-specific features rather than morphology-specific ones.

In sim2real settings, domain adaptation (DA) is the standard approach for addressing this distribution mismatch. By explicitly minimizing a statistical discrepancy between source and target latent distributions, DA encourages the encoder to learn features invariant to systematic differences between simulated and real VLBI data, including thermal noise, residual calibration artifacts, and incomplete uv-coverage that are not fully reproduced by the BK simulations.

We adopt SIDDA (Pandya et al., 2025), which replaces the kernel-based maximum mean discrepancy (MMD) (Gret-

ton et al., 2008; Gretton et al., 2012) objective with the Sinkhorn divergence (Altschuler et al., 2017), a regularized optimal-transport distance. SIDDA additionally learns adaptive weights for the classification and alignment losses during training, reducing the need for manual hyperparameter tuning.

SIDDA optimizes a joint objective consisting of a supervised classification term and a domain-alignment term based on the Sinkhorn divergence. Let $f_\theta(\cdot)$ denote the feature extractor and $g_\phi(\cdot)$ the classifier. Given labelled source-domain samples (x_s, y_s) and unlabelled target-domain samples x_t, the latent representations are defined as

$$z = f_\theta(x_s), \qquad z^* = f_\theta(x_t). \tag{1}$$

The training objective is

$$\mathcal{L} = \frac{1}{2\eta_1^2}\mathcal{L}_{\mathrm{CE}} + \frac{1}{2\eta_2^2}\mathcal{L}_{\mathrm{DA}} + \log(|\eta_1\eta_2|), \tag{2}$$

where $\mathcal{L}_{\mathrm{CE}}$ is the source-domain cross-entropy loss and $\mathcal{L}_{\mathrm{DA}}$ is the Sinkhorn divergence between the source and target latent distributions. The parameters η_1 and η_2 are trainable scalars that are jointly optimized with the network weights and automatically balance the relative contributions of the classification and alignment objectives throughout training. The logarithmic regularization term prevents trivial solutions in which either loss weight collapses toward zero.

3.5. Training procedures

All images underwent a unified preprocessing procedure before being used as model input. Each FITS image was resized to 128×128 pixels, normalized by its 99th-percentile absolute intensity, and transformed using logarithmic normalization. The lower bound of the logarithmic scale was set to three times the estimated background noise level, computed as the median standard deviation of four corner regions, each covering 10% of the image size.

For SCRATCH, FT1, FT2, and the fine-tuning stages of SSL1 and SSL2, the models were trained for up to 100 epochs using a batch size of 128 and early stopping based on the validation loss (patience of 5 epochs). Optimization was performed using AdamW (Loshchilov & Hutter, 2017) with a weight decay of 10^{-3}. A learning rate of 10^{-2} was used for fully unfrozen models and for the MLP classifier in models where the first two convolutional blocks were frozen. The learning rates for the third and fourth convolutional blocks were set to 3×10^{-4} and 10^{-3}, respectively. For SCRATCH, FT1, and FT2, a ReduceLROnPlateau scheduler monitoring the validation loss was used, with a decay factor of 0.1 and a patience of 3 epochs. For SSL1 and SSL2, a CosineAnnealingLR scheduler was employed.

During the SimCLR stages of SSL1 and SSL2, each FITS image was independently augmented twice to form a positive pair, in addition to the universal preprocessing pipeline. The augmentation pipeline included random resized cropping, horizontal and vertical flips, rotations, brightness and contrast jitter, Gaussian blur, and normalization. The model was trained using the NT-Xent loss with temperature parameter $T = 0.2$ for 100 epochs using the LARS optimizer (You et al., 2017), a batch size of 128, and a weight decay of 10^{-6}. The base learning rate was scaled according to the batch size following the SimCLR formulation, resulting in a learning rate of $0.075\sqrt{128}$. The learning-rate schedule consisted of 10 warm-up epochs followed by cosine decay.

The SIDDA model was trained for 50 epochs using a batch size of 128. The first 10 epochs served as a warm-up phase during which only the supervised cross-entropy loss was applied, allowing the model to learn an initial feature representation before introducing the domain-alignment losses. Optimization was performed using AdamW with a weight decay of 10^{-3} and a learning rate of 10^{-2}. A MultiStepLR scheduler with milestones at epochs 25 and 40 and a decay factor of 0.1 was used.

Table 1. Model performance on validation (subsample of source) and test (labelled subsample of target) sets.

MODEL	VAL ACC	VAL F_1	TEST ACC	TEST F_1
SCRATCH	0.991	0.991	0.799	0.800
FT1	0.987	0.987	0.790	0.782
FT2	0.991	0.991	0.782	0.779
SSL1	0.986	0.986	0.792	0.786
SSL2	0.991	0.991	0.807	0.803
SIDDA	**0.996**	**0.996**	**0.864**	**0.864**

4. Results

The results reveal a clear sim2real domain gap between synthetic and observational VLBI data and demonstrate that explicit DA substantially improves generalization to real observations.

From Table 1, all methods achieve extremely high performance on the source-domain validation set, with accuracies ranging from 0.986 to 0.996. This indicates that the classification problem on synthetic BK-generated images is relatively easy for modern convolutional neural networks and that all training strategies successfully learn the simulated distribution. However, performance on the labelled target-domain test set is significantly lower, revealing a substantial domain shift between simulated and real VLBI observations.

The purely supervised baselines (SCRATCH, FT1, and FT2) achieve test accuracies of approximately 0.78–0.80. Interestingly, the model trained from scratch slightly outperforms both ImageNet fine-tuning variants, suggesting that natural-image pretraining transfers only weakly to VLBI

morphology classification. This is expected because radio interferometric images differ strongly from natural images.

Self-supervised pretraining (SSL1 and SSL2) provides a modest improvement over the supervised baselines, with SSL2 achieving the best non-domain-adaptation performance (0.807 accuracy and 0.803 F_1 score). This suggests that contrastive learning on large amounts of unlabelled target-domain data helps the encoder learn more robust morphology-sensitive representations. Nevertheless, the improvement remains relatively limited, indicating that self-supervised invariance alone is insufficient to fully bridge the domain gap between synthetic and observational VLBI data.

In contrast, SIDDA yields a substantial improvement, reaching a test accuracy and F_1 score of 0.864. This corresponds to an absolute improvement of approximately 6 percentage points over the best self-supervised baseline and roughly 8 percentage points over the standard supervised approaches. These results strongly support the hypothesis that explicit alignment between source and target latent distributions is critical for successful sim2real transfer in VLBI imaging tasks.

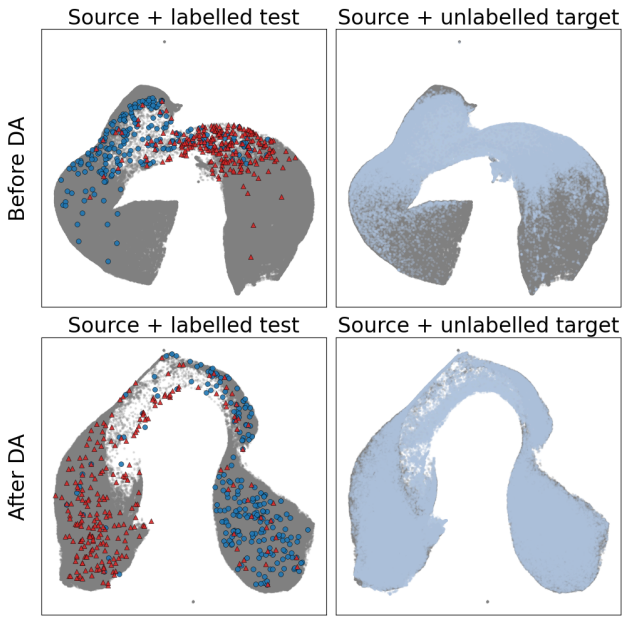

Figure 2. **UMAP visualization of latent representations.** The source-domain samples are shown in gray, while the unlabelled target-domain samples are shown in light blue. Test-set samples with ground-truth labels are overlaid using class-specific markers: blue circles denote compact sources and red squares denote sources with extended jets. Improved overlap between source and target distributions (AUC 0.97 before DA, AUC 0.57 after DA) indicates successful DA and morphology-driven feature alignment.

The UMAP (McInnes et al., 2018) visualizations in Figure 2 qualitatively support this interpretation. Before DA, the source and target latent representations occupy partially separated regions of feature space, indicating that the encoder retains strong domain-specific information. After applying SIDDA, the two domains become substantially more mixed and aligned while preserving the overall structure of the latent manifold. This suggests that the model learns representations that are more invariant to systematic differences between simulated and real interferometric observations.

To quantify domain alignment beyond qualitative UMAP visualizations, we trained a linear classifier to distinguish source and target latent embeddings. Before adaptation, the domains were almost perfectly separable (ROC AUC = 0.97), indicating a substantial sim2real domain gap. After applying SIDDA, the separability decreased dramatically (ROC AUC = 0.57), suggesting that the learned representations became largely domain-invariant while preserving morphology-relevant information, as evidenced by the retained classification performance.

Importantly, the adapted latent space does not collapse into a single undifferentiated cluster, implying that the learned representation still preserves morphology-relevant information necessary for classification. This observation suggests that SIDDA achieves domain alignment without destroying discriminative structure in the latent space.

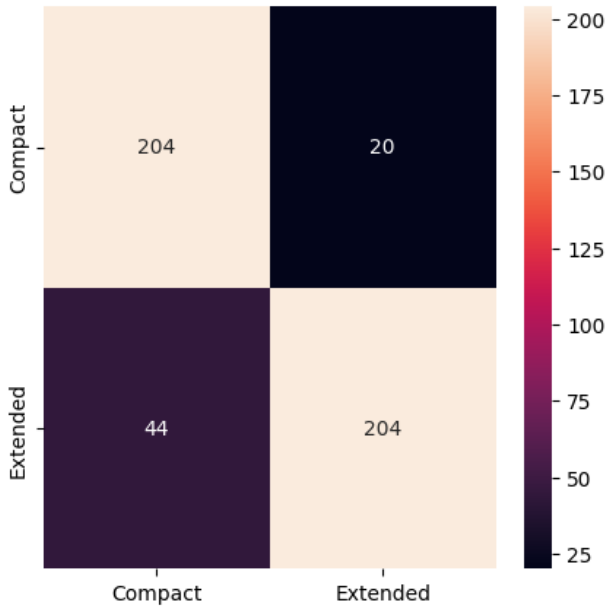

Figure 3. **Target-domain classification performance.** Confusion matrix evaluated on the labelled target-domain test set. Rows correspond to ground-truth classes and columns correspond to model predictions.

The confusion matrix in Figure 3 shows relatively balanced performance across the two morphological classes. The compact class achieves a slightly higher recall ($204/224 \approx 91\%$) than the extended-jet class ($204/248 \approx 82\%$). Most errors correspond to extended jets being classified as compact sources. This asymmetry is physically plausible, since weak or short one-sided jets may become difficult to distinguish from compact cores in noisy VLBI observations, particularly at low dynamic range or limited angular resolution. In contrast, genuinely compact sources contain fewer extended structures that could mimic jet-like morphology.

Overall, these results indicate that the primary challenge is not learning the morphology itself, but transferring that knowledge from simulation to real observations. While SSL partially alleviates this issue, explicit DA proves substantially more effective. These findings suggest that DA may become an essential component of future machine-learning pipelines for next-generation radio surveys, where realistic simulations are abundant but labelled observational data remain limited.

5. Astrophysical Insights

We applied the trained SIDDA model to the unlabelled Astrogeo target set, enabling a population-level comparison of parsec-scale morphologies. The predicted morphology classes exhibit a statistically significant separation in radio flux density (Figure 4). Sources classified as extended jets tend to have higher observed flux densities than sources classified as compact. A Mann-Whitney U test rejects the null hypothesis that the two flux-density distributions are identical ($p < 0.01$). The corresponding dominance probability is approximately 64%, indicating that a randomly selected source classified as extended is more likely to have a higher flux density than a randomly selected source classified as compact. These results demonstrate a statistically significant association between the predicted morphology classes and observed radio brightness. However, the present analysis does not distinguish between intrinsic physical differences and potential contributions from selection effects, source orientation, or other observational biases.

Radio luminosities were estimated from the observed flux density as (Singal et al., 2011)

$$L_\nu = \frac{4\pi D_L^2(z) S_\nu}{(1+z)^{1+\alpha}}, \tag{3}$$

where D_L is the luminosity distance, S_ν is the observed flux density, z is the redshift, and α is the spectral index. A flat ΛCDM cosmology with $H_0 = 71\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_m = 0.27$ was adopted.

To obtain redshift measurements, the classified Astrogeo sources were cross-matched with the Milliquas catalogue

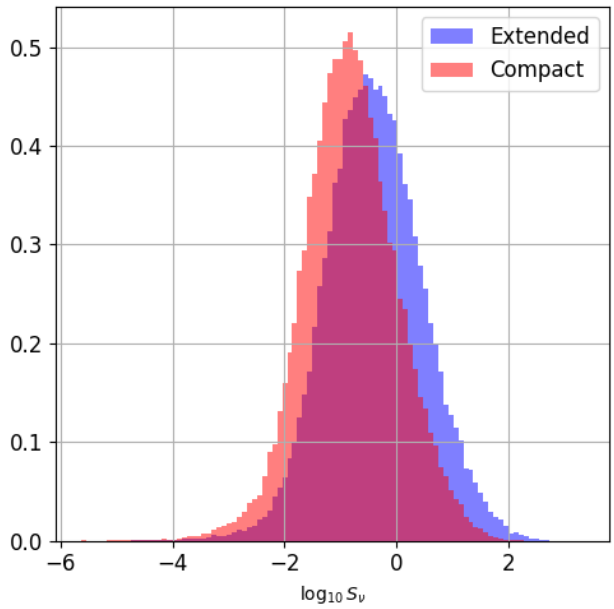

Figure 4. **Radio flux density distributions.** Radio flux density distributions for sources predicted as compact and extended jets in the target Astrogeo sample.

(Flesch, 2023). RFC 2025d was used to obtain VLBI source coordinates, while only X-band Astrogeo observations were retained to reduce frequency-dependent effects. Sources with multiple observations were averaged in morphology probability and flux density. The Astrogeo/RFC sources were then matched to Milliquas using nearest-neighbour search in celestial coordinates.

Compact sources and extended jets were compared using their radio luminosity distributions with a Mann-Whitney U test. Similar to the flux-density analysis, the luminosity distributions of the two predicted morphology classes show a statistically significant separation for a fixed spectral index of $\alpha = 0.5$. The estimated dominance probability is approximately 61%. This result indicates that the observed differences between the two populations persist after converting flux densities to luminosities under the adopted assumptions. However, the present analysis remains exploratory, and additional studies will be required to assess the impact of redshift incompleteness, sample-selection effects, source orientation, and other observational biases.

6. Discussion & Next Steps

This work demonstrates that DA is a critical component for sim2real transfer in VLBI morphology classification. Although all methods achieved near-perfect performance on synthetic validation data, their behaviour on real Astro-

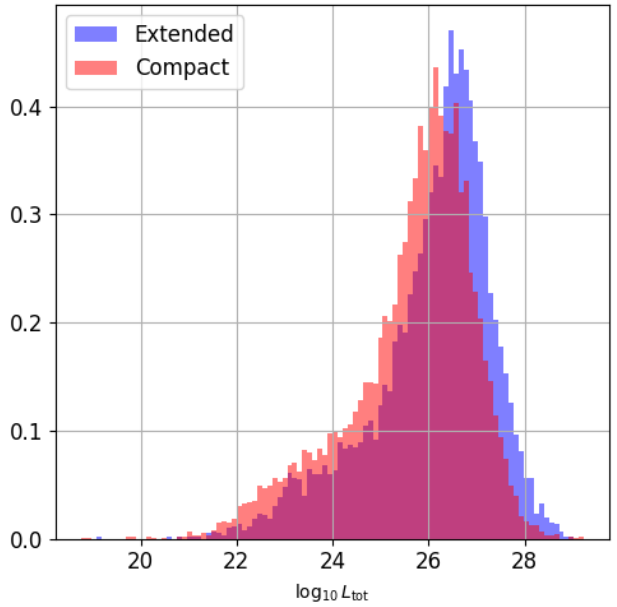

Figure 5. **Total radio luminosity distributions.** Distributions of total radio luminosity for X-band Astrogeo sources predicted as compact or extended jets by SIDDA.

geo observations differed substantially, indicating that the dominant challenge is not morphology recognition itself, but robustness to the domain gap between simulations and observational data.

The relatively weak transfer performance of ImageNet pretraining is consistent with the fundamentally different statistical properties of radio interferometric images compared to natural images. As a result, features learned from natural-image datasets appear only partially transferable to radio astronomy.

SSL improved performance only modestly. Although SimCLR pretraining exposed the encoder to large amounts of unlabelled target-domain data, contrastive objectives primarily learn invariance to augmentations within a single domain and do not explicitly enforce alignment between simulated and observational distributions. Consequently, the learned latent representations still retained significant domain-specific structure. This result suggests that SSL alone may be insufficient for sim2real transfer in interferometric imaging tasks where the domain shift is dominated by instrumental and reconstruction effects rather than by limited labelled data.

Several alternative domain-adaptation approaches were also explored but did not achieve competitive performance. In particular, classical kernel-based MMD alignment produced unstable results and weaker target-domain generalization. This may be related to the sensitivity of MMD to kernel choice and bandwidth selection in high-dimensional latent spaces. Similarly, adversarial DA using Domain-Adversarial Neural Networks (Ganin et al., 2015) did not improve over the supervised or SSL baselines. In practice, adversarial training frequently became unstable and tended to reduce class separability in the latent space while attempting to enforce domain invariance.

In contrast, SIDDA consistently produced substantially better target-domain performance and more coherent latent-space alignment. One possible explanation is that the Sinkhorn divergence provides a geometrically meaningful optimal-transport distance between latent distributions while remaining more stable than adversarial objectives. Additionally, the adaptive balancing of classification and alignment losses reduces the need for careful manual tuning, which is often a major practical difficulty in domain-adaptation pipelines.

The large-scale application of the classifier reveals statistically significant differences in both radio flux density and radio luminosity between the predicted compact and extended populations. These findings indicate that the learned morphology classes are associated with measurable differences in observed source properties. However, the present analysis remains exploratory. The observed trends may be influenced by a combination of intrinsic physical differences, orientation effects, sample-selection biases, redshift incompleteness, and observational systematics. A more detailed astrophysical investigation will be required to determine the relative contribution of these factors.

Several limitations remain. First, the target-domain test set is relatively small and manually labelled, introducing unavoidable subjectivity near the boundary between compact and weakly extended jets. Second, the binary classification scheme intentionally simplifies the diversity of parsec-scale morphologies in order to reduce ambiguities caused by observational limitations. Extending the approach to richer morphological taxonomies will likely require larger labelled datasets.

Future work could explore more advanced architectures, including ViTs and modern convolutional networks, together with more realistic jet-emission simulations that account for complex structures such as jet bending. Another promising direction is the incorporation of interferometric information directly through the uv-coverage, which may help the model better account for observational biases and imaging artefacts. Finally, the resulting morphology catalogue for more than 130 000 Astrogeo images opens the possibility of large-scale statistical studies linking parsec-scale jet morphology with variability, jet kinematics, and multiwavelength AGN properties.

Acknowledgements

This work is supported in the framework of the State project "Science" by the Ministry of Science and Higher Education of the Russian Federation under the contract 075-15-2024-541.

This research has made use of NASA's Astrophysics Data System.

This research made use of *Astropy*, a community-developed core Python package for Astronomy (Astropy Collaboration et al., 2013), *NumPy* (van der Walt et al., 2011), *SciPy* (Virtanen et al., 2020), *FINUFFT* (Barnett et al., 2019), *PyTorch* (Ansel et al., 2024), *OpenCV* (Bradski, 2000), *Matplotlib* (Hunter, 2007) Python package was used for generating all plots in this paper.

Authors thank the anonymous reviewers for their useful comments. D.Z. would especially like to thank Aleksandra Ćiprijanović and Nabeel Rehemtulla for valuable discussions and ideas.

Impact Statement

This paper presents work whose primary goal is to advance machine learning for scientific discovery in astronomy. The expected societal impact is mainly indirect: improving tools for large-scale analysis of astronomical surveys and supporting reproducible scientific workflows. We do not identify any specific direct societal risks that require separate discussion.

References

Altschuler, J., Weed, J., and Rigollet, P. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. *arXiv e-prints*, art. arXiv:1705.09634, May 2017. doi: 10.48550/arXiv.1705.09634.

Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Ellison, E., Feng, W., Gong, J., Gschwind, M., Hirsh, B., Huang, S., Kalambarkar, K., Kirsch, L., Lazos, M., Lezcano, M., Liang, Y., Liang, J., Lu, Y., Luk, C. K., Maher, B., Pan, Y., Puhrsch, C., Reso, M., Saroufim, M., Siraichi, M. Y., Suk, H., Zhang, S., Suo, M., Tillet, P., Zhao, X., Wang, E., Zhou, K., Zou, R., Wang, X., Mathews, A., Wen, W., Chanan, G., Wu, P., and Chintala, S. Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *Proc. 29th ACM Int. Conf. Archit. Support Program. Lang. Oper. Syst.*, volume 2, pp. 929–947. ACM, 2024.

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., and Streicher, O. Astropy: A community Python package for astronomy. *A&A*, 558:A33, 2013.

Barnett, A. H., Magland, J., and af Klinteberg, L. A Parallel Nonuniform Fast Fourier Transform Library Based on an "Exponential of Semicircle" Kernel. *SIAM J. Sci. Comput.*, 41(5):C479–C504, 2019.

Blandford, R., Meier, D., and Readhead, A. Relativistic Jets from Active Galactic Nuclei. *Annual Review of Astronomy and Astrophysics*, 57:467–509, August 2019. doi: 10.1146/annurev-astro-081817-051948.

Blandford, R. D. and Königl, A. Relativistic jets as compact radio sources. *The Astrophysical Journal*, 232:34–48, August 1979. doi: 10.1086/157262.

Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools*, 2000.

Cecconello, T., Riggi, S., Becciani, U., Vitello, F., Hopkins, A. M., Vizzari, G., Spampinato, C., and Palazzo, S. Self-Supervised Learning for Radio-Astronomy Source Classification: A Benchmark. In *Proceedings of International Conference on Pattern Recognition*, pp. 424–439, 2024.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of International Conference on Machine Learning*, 2020.

Clausen-Brown, E., Savolainen, T., Pushkarev, A. B., Kovalev, Y. Y., and Zensus, J. A. Causal connection in parsec-scale relativistic jets: results from the MOJAVE VLBI survey. *Astronomy and Astrophysics*, 558:A144, October 2013. doi: 10.1051/0004-6361/201322203.

Estepa, I. G., Sarasua, I., Nagarajan, B., and Radeva, P. All4One: Symbiotic Neighbour Contrastive Learning via Self-Attention and Redundancy Reduction. In *Proceedings of IEEE International Conference on Computer Vision*, pp. 16243–16253, 2023.

Fanaroff, B. L. and Riley, J. M. The morphology of extragalactic radio sources of high and low luminosity. *Monthly Notices of the Royal Astronomical Society*, 167: 31P–36P, May 1974. doi: 10.1093/mnras/167.1.31P.

Flesch, E. W. The Million Quasars (Milliquas) Catalogue, v8. *The Open Journal of Astrophysics*, 6:49, December 2023. doi: 10.21105/astro.2308.01505.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. Domain-Adversarial Training of Neural Networks. *arXiv e-prints*, art. arXiv:1505.07818, May 2015. doi: 10.48550/arXiv.1505.07818.

Gretton, A., Borgwardt, K., Rasch, M. J., Scholkopf, B., and Smola, A. J. A Kernel Method for the Two-Sample Problem. *arXiv e-prints*, art. arXiv:0805.2368, May 2008. doi: 10.48550/arXiv.0805.2368.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

Grill, J. B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Pires, B. A., Guo, Z. D., Azar, M. G., Piot, B., Kavukcuoglu, K., Munos, R., and Valko, M. Bootstrap your own latent: A new approach to self-supervised Learning. In *Proceedings of Conference on Neural Information Processing Systems*, 2020.

He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Högbom, J. A. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy and Astrophysics Supplement Series*, 15:417, June 1974.

Hunter, J. D. Matplotlib: A 2d graphics environment. *Comput. Sci. Eng.*, 9(3):90–95, 2007.

Kolesnikov, A., Dosovitskiy, A., Weissenborn, D., Heigold, G., Uszkoreit, J., Beyer, L., Minderer, M., Dehghani, M., Houlsby, N., Gelly, S., Unterthiner, T., and Zhai, X. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *Proceedings of International Conference on Learning Representations*, 2021.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P., and Vandenberg, J. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189, September 2008. doi: 10.1111/j.1365-2966.2008.13689.x.

Lister, M. L., Homan, D. C., Hovatta, T., Kellermann, K. I., Kiehlmann, S., Kovalev, Y. Y., Max-Moerbeck, W., Pushkarev, A. B., Readhead, A. C. S., Ros, E., and Savolainen, T. MOJAVE. XVII. Jet Kinematics and Parent Population Properties of Relativistically Beamed Radio-loud Blazars. *The Astrophysical Journal*, 874(1):43, March 2019. doi: 10.3847/1538-4357/ab08ee.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of IEEE International Conference on Computer Vision*, 2021.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., and Guo, B. Swin Transformer V2: Scaling Up Capacity and Resolution. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2022a.

Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., and Xie, S. A ConvNet for the 2020s. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2022b.

Loshchilov, I. and Hutter, F. Decoupled Weight Decay Regularization. *arXiv e-prints*, art. arXiv:1711.05101, November 2017. doi: 10.48550/arXiv.1711.05101.

McInnes, L., Healy, J., and Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*, February 2018.

Pandya, S., Patel, P., Nord, B. D., Walmsley, M., and Ciprijanovic, A. SIDDA: SInkhorn Dynamic Domain Adaptation for image classification with equivariant neural networks. *Machine Learning: Science and Technology*, August 2025. doi: 10.1088/2632-2153/adf701.

Pashchenko, I. N. and Plavin, A. V. Inferring the jet parameters of active galactic nuclei using Bayesian analysis of VLBI data with a non-uniform jet model. *Monthly Notices of the Royal Astronomical Society*, 488(1):939–953, September 2019. doi: 10.1093/mnras/stz1677.

Petrov, L. Y. and Kovalev, Y. Y. The Radio Fundamental Catalog. I. Astrometry. *The Astrophysical Journal Supplement Series*, 276(2):38, February 2025. doi: 10.3847/1538-4365/ad8c36.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

Singal, A. K. and Rajpurohit, K. Dependence of Fanaroff-Riley break of radio galaxies on luminosity and redshift. *Astrophysics and Space Science*, 353(1):233–240, September 2014. doi: 10.1007/s10509-014-2011-y.

Singal, J., Petrosian, V., Lawrence, A., and Stawarz, Ł. On the Radio and Optical Luminosity Evolution of Quasars. *The Astrophysical Journal*, 743(2):104, December 2011. doi: 10.1088/0004-637X/743/2/104.

Sobol', I. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.

Tan, M. and Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of International Conference on Machine Learning*, 2019.

van der Walt, S., Colbert, S. C., and Varoquaux, G. The numpy array: A structure for efficient numerical computation. *Comput. Sci. Eng.*, 13(2):22–30, 2011.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods*, 17:261–272, 2020.

You, Y., Gitman, I., and Ginsburg, B. Large Batch Training of Convolutional Networks. *ArXiv e-prints*, 2017.

Zagorulia, D. and Lisakov, M. Classification of Jets in Active Galactic Nuclei using Machine Learning. In *Modern astronomy: from the Early Universe to exoplanets and black holes*, pp. 113–116, 2024.

A. Physical Parameters for Synthetic Sources

For compact sources, we used a circular Gaussian model parametrised by three quantities: $\varphi \in [0, 2\pi]$ – the image-plane rotation angle in radians, $r \in [0.1, 1]$ – the angular offset from the image centre in milliarcseconds, and $S \in [0.5, 20]$ – the total radio flux density in janskys.

For sources with resolved spatial structure, we used the BK model, which describes a conical relativistic jet with bulk Lorentz factor Γ and half-opening angle ϕ, observed at a line-of-sight angle θ. The magnetic field strength and the normalization of the particle energy distribution vary with distance from the jet apex as

$$K(r) = K_1 \left(\frac{r}{r_1}\right)^{-n}, \qquad B(r) = B_1 \left(\frac{r}{r_1}\right)^{-m}, \tag{4}$$

where $r_1 = 1\,\mathrm{pc}$, and K_1 and B_1 are the corresponding values at r_1. The number density of particles with Lorentz factor γ is assumed to follow a power-law distribution,

$$N(\gamma) = K(r)\gamma^{-s}. \tag{5}$$

The exponent s is related to the optically thin spectral index α through $\alpha = (s-1)/2$. We set $s = 2$, corresponding to $\alpha = 0.5$, and fixed $n = 2$ and $m = 1$. The jet half-opening angle and viewing angle were set to $\phi = \min(0.13/\Gamma, \theta/2)$ following (Clausen-Brown et al., 2013), and $\theta = 0.7/\Gamma$ following (Lister et al., 2019). We further assumed equipartition between the particle and magnetic-field energy densities,

$$K_1 = \frac{B_1^2}{8\pi m_e c^2 \ln(\gamma_{\max}/\gamma_{\min})}, \tag{6}$$

with $\ln(\gamma_{\max}/\gamma_{\min}) \approx 10$. This reduces the model to three free parameters: the image-plane rotation angle $\varphi \in [0, 2\pi]$, the bulk Lorentz factor $\Gamma \in [1, 50]$, and the magnetic-field strength at 1 pc, $B_1 \in [0.01, 2]\,\mathrm{gauss}$.

All physical parameters were sampled using a Sobol' sequence (1967) to achieve a more uniform coverage of the parameter space.