When Visual Foundation Models meet Astronomical Data

Matthieu Le Lain Sébastien Lefèvre

IRISA, UMR CNRS 6074, UBS, Vannes, France

Introduction

In the last decade, Artificial Intelligence has achieved breakthroughs in various fields like Computer Vision and Natural Language Processing. Astronomy has also benefited, with Deep Neural Networks, particularly CNNs, being used to classify light curves for exo-planet detection using real and synthetic data [1] and to identify low-surface-brightness galaxies in Dark Energy Survey images [3], among other applications.

Motivations

A new paradigm shift is currently occurring in AI, with the rise of **Foundation Models**. These models, which are characterized by their training on extremely large amounts of (unlabeled) data, are able to learn more generic features, allowing them to be adapted to a variety of downstream tasks (Figure 1).

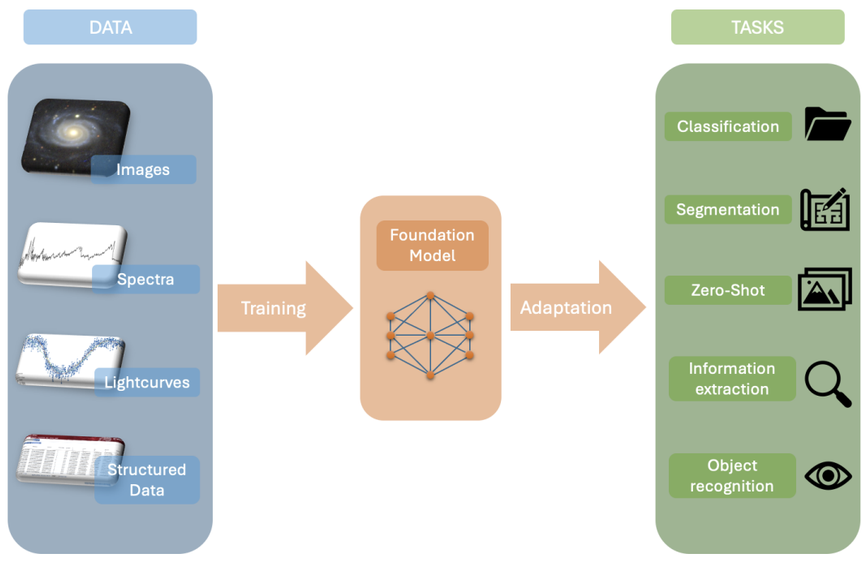

Figure 1. Diagram of Foundation models. Images from ESA Gaia DR3 website and Galaxy10 DECals [6].

It is anticipated that these foundation models will also represent a new AI breakthrough for the analysis of astronomical data. Indeed, the existing datasets generated by telescopes and ongoing projects require tools for ever more efficient analysis, given the complexity and ever-increasing volume of data involved in modern astronomical research [5]. The needs appear in line with the capabilities of Foundation Models.

Method

We aim to evaluate the performance of existing visual foundation models, and compare them to more classical architectures. We adopt a common methodology based on fine-tuning to assess the relevance of adapting pre-trained models for astronomical tasks.

We report various metrics based on datasets used in several studies, specifically for classification tasks. In particular, we consider the classification of galaxy images using the Galaxy10Decals dataset (Figure 2), from GalaxyZoo.

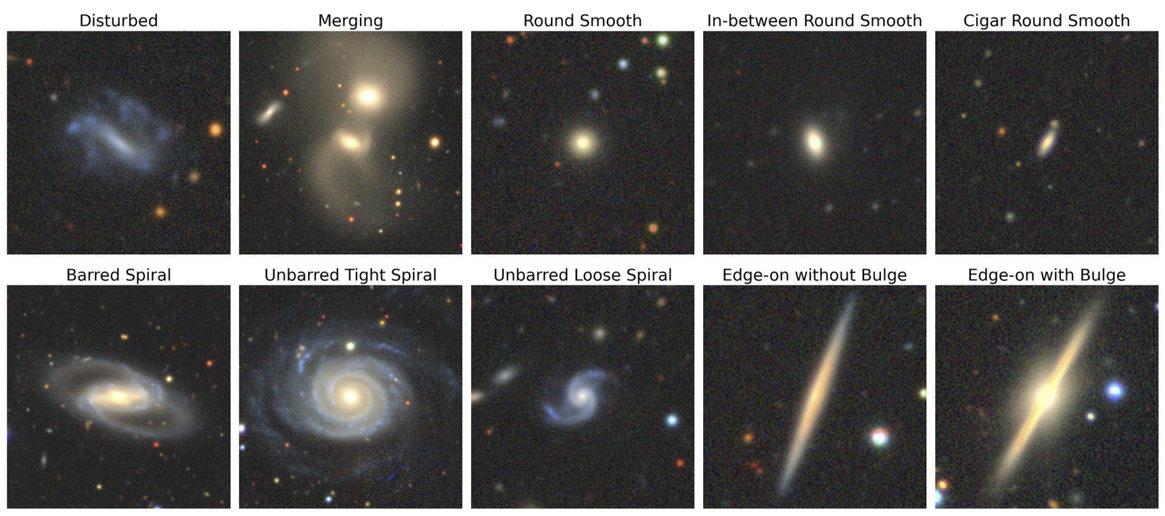

Figure 2. Galaxy10Decals from GalaxyZoo data, Henry Leung/Jo Bovy 2021.

Results

- **Transformer model's efficacy**: Popular models such as ViT, Swin Transformer V2, DINOv2, and BEiT show no significant advantage on small datasets.
- **Best performance**: ConvNeXtV2, a classical architecture enhanced with masked autoencoder, outperforms these models despite being less complex.
- **Impact of model complexity**: Training time and carbon footprint generally increase with model complexity.
- **Energy vs. compactness**: ConvNeXtV2, despite having three times fewer parameters than ViT, reduces the carbon footprint by only 50% (see Table 1).
- **Accuracy upper bound**: None of the tested architectures exceeded 87% accuracy, suggesting that the task remains too complex for generic foundation models.
- Most errors occur between similar classes, e.g. Unbarred Spirals (Figure 3), calling for more **fine-grained characterization** ability and including **astronomical priors**.

Accuracy vs Model complexity

Models	Accuracy (%)	Params (millions)	CO2 (kg eq.)	Time (h)
Swinv2 Base	85.96	86.9	0.022	2h28
DINOv2 Base	85.63	86.6	0.018	1h54
ViT Base	86.02	86.1	0.024	2h41
BEiT Base	86.02	85.7	0.015	1h31
ConvNeXtV2 Nano	86.87	**15.0**	**0.010**	**1h02**
ConvNeXtV2 Tiny	**87.31**	27.9	0.011	1h09
EfficientNet B7	85.74	66.0	0.022	2h29
Resnet101	85.12	48.0	0.012	1h32
VGG16	84.05	138.0	**0.010**	1h03

Table 1. Best accuracy per models pretrained on ImageNet and finetuned with GZ10 DECals.

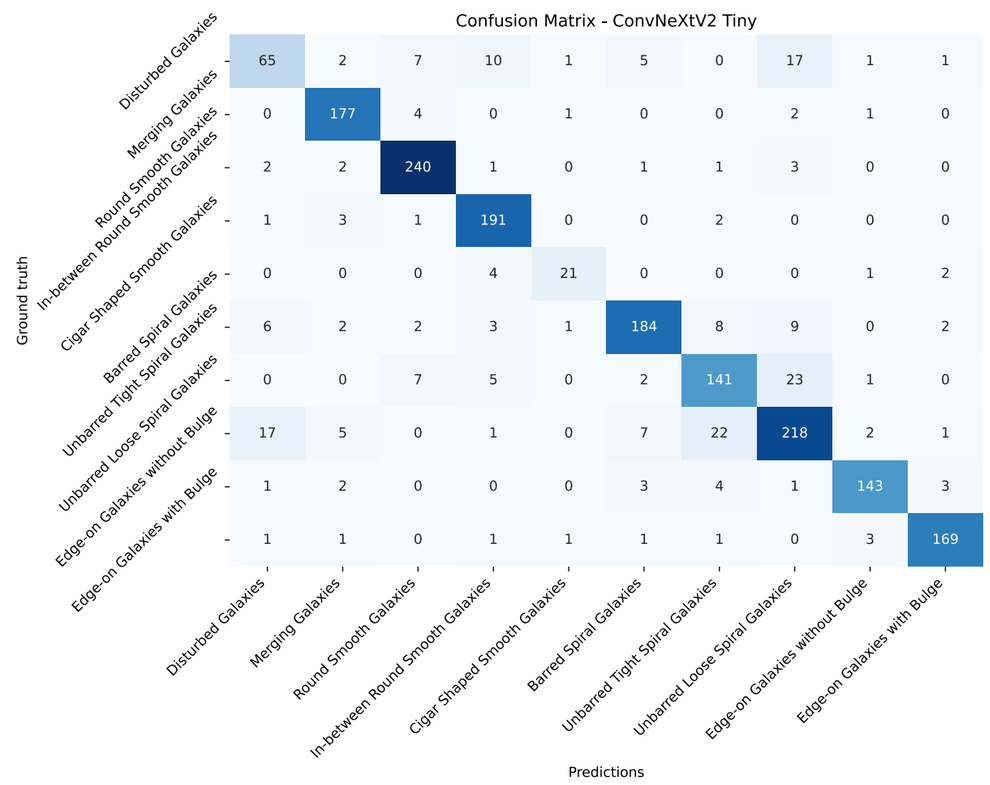

Figure 3. Confusion matrix for model ConvNeXtV2 Tiny finetuned on GalaxyZoo10Decals dataset.

Influence of pretraining strategy

We also compared models pretrained with galaxy images (~840k) from Zoobot models [4] against the same models pretrained with ImageNet 1K and without pretraining.

Models	No Pretrain	Pretrain ImageNet1K	Pretrain Zoobot
MaxVit Base	76.04 (-11.67)	82.70 (-5.01)	**87.71**
ConvNeXt Nano	65.54 (-20.65)	85.57 (-0.62)	86.19
EfficientNet B0	74.41 (-12.23)	80.16 (-6.48)	86.64

Table 2. Accuracy per models for various pretraining strategies and finetuned on GZ10 DECals.

- Models without pre-training have some difficulty in generalising effectively with a limited number of galaxy images (~17.7k).
- Zoobot pretraining achieves systematic superior performance. But this might be due to the possible use of the finetuned dataset in the pretraining step.

Conclusion

We find that more complex models does not lead to better performance.

- Vision foundation models finetuned on a modest-size dataset of galaxies do not provide better performance than more traditional architectures such as CNNs.
- Modern CNN-type architectures relying on advanced concepts (MAEs in ConvNeXtV2), outperform other models despite their relatively low complexity.
- Solving astronomical tasks with foundation models, whatever the pretraining strategy, is still an open issue, since the best accuracy reported here is below 88%: **Astronomical-driven AI models are needed.**
- Our study, **with no pretrain and only fine-tuning**, emitted only 0.766kg eq. CO2 (40 models, 81h GPU), equivalent to 3.07km driven by an average ICE car. [2]

References

[1] Cuéllar et al. Deep learning exoplanets detection by combining real and syntheticdata. Plos one,17(5):e0268199, 2022. [2] Lacoste Alexandre et al. Quantifying the carbon emissions of machine learning. arXiv preprint arXiv:1910.09700, 2019. [3] Tanoglidis et al. Deepshadows: Separating low surface brightness galaxies from artifacts using deep learning. Astronomy and Computing, 35:100469, 2021. [4] Walmsley et al. Scaling laws for galaxy images. arXiv preprint arXiv:2404.02973, 2024. [5] Muhammad Faaique. Overview of big data analytics in modern astronomy. International Journal of Mathematics, Statistics, and Computer Science, 2:96–113, 2023. [6] Henry W Leung and Jo Bovy. Deep learning of multi-element abundances from high-resolution spectroscopic data. Monthly Notices of the Royal Astronomical Society, 483(3):3255–3277.