HEAL-PINN: Physics-Informed Swin Transformer for Dark Matter Studies for Sparse Lensing Data

Dhruv Srivastava*
Department of Physics,
University of Illinois at Urbana-Champaign,
Champaign, IL - 61820
dhruv12@illinois.edu

Pranath Reddy Kumbam
University of Florida,
Gainesville, FL 32611, USA
kumbam.pranath@ufl.edu

Michael W. Toomey
Center for Theoretical Physics,
Massachusetts Institute of Technology,
Cambridge, MA 02139, USA
mtoomey@mit.edu

Sergei Gleyzer
Department of Physics
Astronomy,
University of Alabama,
Tuscaloosa, AL 35401, USA
sgleyzer@ua.edu

Abstract

Strong gravitational lensing offers a powerful probe of the nature of dark matter from the morphology of its substructure. While expected to change in the next few years, current available data is sparse, making analyses of lensing systems for extraction of dark-matter properties difficult. In this work we propose a physics-informed Swin Transformer model, including a novel HEAL-Swin variant with the gravitational lensing equation embedded in the architecture, to classify between different models for dark matter from simulated lensing systems. We test the classification performance of various Swin Transformer models on a small dataset of these simulated lensing images, mimicking the current availability of lenses. The architectures include a base Swin Transformer model, a Swin Transformer model with the lensing equation baked into its architecture, a HEAL-Swin model, and HEAL-Swin with a physics-informed architecture. We then evaluate the models based on the evolution of Receiver Operating Characteristic Area Under the Curve (ROC AUC) and demonstrate that physics-informed HEAL-Swin evolves ROC AUC the fastest among all tested models. [2]

1 Introduction

The microphysical nature of dark matter still remains a mystery despite overwhelming evidence for its existence. A corner stone of the concordance model of cosmology, cold dark matter (CDM) dominates modern cosmology, and accurately describes the Universe on large scales, especially with observations like the Cosmic Microwave Background (CMB) and galaxy clusters [8]. However, on highly non-linear scales the CDM paradigm faces challenges on sub-galactic scales. Consequently there is a myriad of dark matter models that have been proposed that predicted differing behavior on these scales, while retaining the well constrained behavior on large-scales [6].

The formation of dark matter halos, which are host to (baryonic) galaxies, is famously achieved through hierarchical structure formation. Under the standard assumptions for CDM, one expects

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

[2]Code: https://anonymous.4open.science/r/HEALSwin-PINN-58EA

a population of low-mass dark matter subhalos still present in the host halo. Other models predict similar behavior, but deviate on smaller scalers. Famously, axions and other ultra-light dark matter represent promising alternate dark matter theories. In particular, ultra-light axions can have unique properties like the formation of vortices and exhibiting wave-like properties on astrophysical scales. This can also manifest in a suppression of low-mass structure. Other axion models can form dense, low-mass clusters [15]. This discrepancy allows us to develop a key observing discriminant for testing the two theories.

Historically, gravitational lensing substructure analysis has been done with direct detection methods [14], which rely on having to model the smooth component of the lens galaxy before inferring properties. Bayesian methods have also been used for this purpose which can determine if lensing images are consistent with a given dark matter model [5]. These models are complex and computationally intensive, which may be unfeasible to use for large sky survey data.

Deep Learning techniques have also been used to classify dark matter substructures [8][6][1][17][16]. Convolutional Neural Networks (CNNs) have proven to be quite effective at this task, in part thanks to their translational invariance and their optimization for extracting correlation in image data (see for example, [8][6][1]). Some other examples, like [17][16] have also shown the effectiveness of a Vision Transformer (ViT) in lens classification.

More recently, the application of Physics-Informed Neural Network (PINN) (for an example of PINNs, see [10]) has seen boosts in performance over standard architectures. The governing equation being used in [17] is the lensing equation, and using the following ansatz:

$$\Psi(x_i, y_i) = k(x_i, y_i) \cdot \Psi_{\text{SIS}}(x_i, y_i) \tag{1}$$

Previous implementations have a computational problem, however. Vision Transformers have a computational complexity that scales quadratically with image size [7], and become unfeasible to use with limited computing resources or high-definition images. This would be a problem in high-quality surveys such as the LSST, which stream a high volume of dense image data every timestep [11]. Thus, we propose using Swin Transformers [12] for fast inference. Swin Transformers compute self-attention locally instead of globally, allowing them to be linear with respect to image size.

HEAL-Swin [4] Transformers accomodate the inherent artifacts that may emerge when a celestial plane is projected onto a 2-D plane for imaging, using HEALPix [9], which may otherwise be misinterpreted as real signals by standard deep learning models. Thus, it is hypothesized that HEAL-Swin would perform better than a regular Swin Transformer model.

We will introduce Swin Transformers and HEAL-Swin, both with and without Physics-informed blocks, having a novel application of HEAL-Swin to astrophysical data, and developing a physics-informed Swin Transformer, along with a physics-informed HEAL-Swin, and test their performance against an incredibly small simulated lensing dataset, which will mirror a real-world scenario where labeled training data from sky surveys are limited. The primary method of evaluation will be the evolution of ROC AUC throughout epochs.

2 Dataset

We evaluate our architectures using simulated galaxy-galaxy strong lensing images from a mock HST-like instrument, generated with `lenstronomy` [2]. The 150 × 150 pixel, single-channel images include a Gaussian point-spread function and a signal-to-noise-ration ~ 25 consistant with expectations. We model the background galaxies with a Sérsic light profile, while the lens's main dark matter halo follows a singular isothermal profile. We simulate three substructure classes: (1) CDM with subhalos; (2) axion dark matter with mass (m$\sim 10^{-23}$ eV) with suppressed substructure and vortices; and (3) a no-substructure baseline. The default dataset contains 30,000 images per class (90,000 total), but are pruned to 500 per class (1,500 total) for this analysis. We further include data augmentation: random flips and rotations up to 90 degrees. We employ an 85:15 training/validation split with a batch size of 32.

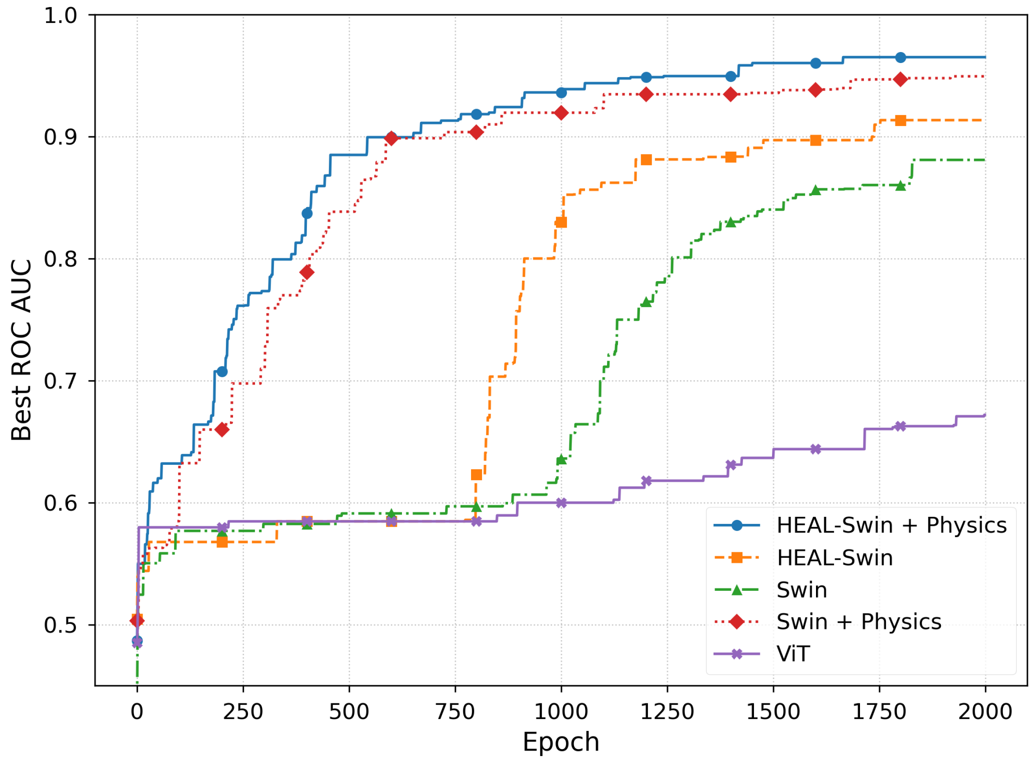

Figure 1: ROC AUC evolution in training across all tested models

3 Architectures

All architectures here are, or are based on, the Swin Transformer [12], to maximize training efficiency even in a small dataset such as the one constructed above. The Physics-Informed encoding uses the ansatz as in the introduction, and Lensiformer, in [17].

The base Swin Transformer model being used is adapted from a Pytorch implementation of a Swin Transformer, from [3]. The Swin Transformer has built-in and toggle-able physics encoding, which changes the architecture of the model from the initialization and forward pass. When the toggle for Physics encoding is enabled, the input image is passed onto the relativistic encoder [17], along with the original patch weights, instead, transforming it into a physics-informed Swin Transformer. The encoder uses an internal transformer to apply inverse transformation, producing a physically corrected image. The new image is then passed onto the standard Swin Transformer workflow.

HEAL-Swin is adapted as per the implementation from [4]. As with the original implementation, the base Swin Transformer is injected with a modified HEALPix-conscious window shifting and reverse, along with other changes as per the paper. The inputs were changed to support the testing dataset, as the original paper dealt with fisheye images.

The Physics-informed encoding is added to HEAL-Swin as well, albeit more embedded within the model and not toggle-able as the other one. Just like the base Swin Transformer, the input image is passed onto the relativistic encoder, along with the original patch weights, the encoder applies inverse transformation, producing a physically corrected image. Thus, this model actually corrects two times; first, it reconstructs based on the lensing equation, and then, it accounts for the spherical deformations in sky surveys.

A standard Vision Transformer, based on the original architecture in [7] is used as a null test, slightly adapted to accept the input tensors from this dataset.

4 Setup

To assess the performance of the models, a benchmarking study was done using the models. The dataset created with the methods discussed in Section 2 were used, NVIDIA A100 GPUs were the primary computing resource . The loss criterion was cross entropy, and AdamW was used as the optimizer [13] . All transformer models have an embedding dimension of 48 and depth of 2. The models were evaluated over 2000 epochs with a starting learning rate of 1e-5, and a learning rate scheduler that is a hybrid of a discrete lambda function, and cosine annealing.

Table 1: Micro-average and macro-average ROC curves for all tested models at their best saved states.

Model	Micro-Average ROC	Macro-Average ROC
ViT	0.68	0.68
Swin	0.88	0.87
Swin + Physics	0.96	0.96
HEAL-Swin	0.91	0.89
HEAL-Swin + Physics	0.97	0.96

5 Results and Discussion

From the table and the model evolution, it is clear that the physics-informed models perform the best, with HEAL-Swin + Physics coming on top. As for the base models, HEAL-Swin slightly edges out the default Swin Transformer, while the basic Vision Transformer struggles in this dataset. The performance gap between the baseline Swin Transformer (0.88 AUC) and the fully adapted HEAL-Swin + Physics model (0.97 AUC) represents a substantial 9-point improvement.

The raw performance gain from adding the physics encoding to the Swin Transformer is large, allowing for a (+0.08) point improvement in ROC AUC. However, while the final performance of the HEAL-Swin + Physics model is only slightly higher, the ROC AUC evolution is faster than the Swin Transformer with physics. This shows that the HEAL-Swin architecture provides a cleaner and more literal representation of the lensed features through HEALPix lens shifting. The physics encoding then leverages this representation more effectively. This suggests that establishing physics-informed structure at the architectural level serves as a crucial foundation for physics-based classification.

As for the evolution of the models, HEAL-Swin + Physics exhibits an ideal training trajectory. It learns exceptionally rapidly, surpassing an ROC AUC of 0.9 before 500 epochs have passed. Its convergence is smooth and monotonic. Swin + Physics also learns quickly and converges to a high-performance plateau. While its initial learning rate is slightly slower than its HEAL-Swin counterpart, it still demonstrates a highly efficient and stable training process. HEAL-Swin shows an improvement in learning speed compared to base Swin Transformer. However, it learns substantially slower than the two physics-informed models. The baseline Swin model displays the most problematic training dynamic. It learns extremely slowly. Its performance increases in a distinct “staircase” pattern, characterized by long periods of stagnation followed by abrupt jumps. This behavior is symptomatic of a model struggling to find useful patterns, likely getting trapped for extended periods. While on the other hand, ViT struggles both in evolution and final best ROC AUC, highlighting its ineffectiveness in small datasets.

This study demonstrates conclusively that for the task of classifying dark matter substructure from gravitational lensing images, particularly in a relative lack of data, a generic deep learning approach is insufficient. The baseline Swin Transformer struggles to learn efficiently and achieves only modest performance. Results are only realized through the combination of two distinct forms of inductive bias: a geometry-aware architecture (HEAL-Swin) that respects the spherical nature of the data, and domain-specific feature engineering (relativistic physics encoding) that injects prior knowledge of the underlying physics. The resulting HEAL-Swin + Physics model not only achieves the highest classification accuracy (0.97 ROC AUC) but also exhibits superior evolution speed and training stability.

References

[1] S. Alexander, S. Gleyzer, E. McDonough, M. W. Toomey, and E. Usai. Deep Learning the Morphology of Dark Matter Substructure. *The Astrophysical Journal*, 893(1):15, 2020.

[2] Simon Birrer and Adam Amara. lenstronomy: Multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, 22:189–201, December 2018.

[3] Hu Cao. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation, 2021.

[4] Olof Carlsson, Jens E. Gerken, Hampus Linander, Hedda Spieß, Frida Ohlsson, Christoffer Petersson, and Daniel Persson. HEAL-SWIN: A Vision Transformer On The Sphere, 2024.

[5] T. Daylan, F.-Y. Cyr-Racine, A. Diaz Rivero, C. Dvorkin, and D. P. Finkbeiner. Probing the Small-scale Structure in Strongly Lensed Systems via Transdimensional Inference. *The Astrophysical Journal*, 854(2):141, 2018.

[6] A. Diaz Rivero and C. Dvorkin. Direct detection of dark matter substructure in strong lens images with convolutional neural networks. *Physical Review D*, 101(2):023515, 2020.

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2021.

[8] C. Gannon, A. Nierenberg, A. Benson, R. Keeley, X. Du, and D. Gilman. Dark matter substructure: A lensing perspective. *Physical Review D*, 112(2):023532, 2025.

[9] Krzysztof M. Gorski, Eric Hivon, A. J. Banday, Benjamin D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622:759–771, 2004.

[10] Zhongkai Hao, Jiachen Yao, Chang Su, Hang Su, Ziao Wang, Fanzhi Lu, Zeyu Xia, Yichi Zhang, Songming Liu, Lu Lu, and Jun Zhu. Pinnacle: A comprehensive benchmark of physics-informed neural networks for solving pdes, 2023.

[11] Željko Ivezić et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal*, 873(2):111, 2019.

[12] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, 2021.

[13] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.

[14] Shude Mao and Peter Schneider. Evidence for substructure in lens galaxies? *Monthly Notices of the Royal Astronomical Society*, 295(3):587–594, 1998.

[15] D. J. E. Marsh and J. Silk. A model for halo formation with axion mixed dark matter. *Monthly Notices of the Royal Astronomical Society*, 437(3):2652–2663, 2013.

[16] Ashutosh Ojha, Sergei Gleyzer, Michael W. Toomey, and Pranath Reddy Kumbam. LensPINN: Physics Informed Neural Network for Learning Dark Matter Morphology in Lensing. In *NeurIPS ML4PS 2024 Workshop*, 2024.

[17] Lucas J. Velôso, Sergei Gleyzer, and Michael W. Toomey. Lensformer: A physics-informed vision transformer for gravitational lensing. In *NeurIPS ML4PS 2023 Workshop*, 2023.