Lensformer: A Physics-Informed Vision Transformer for Gravitational Lensing

Lucas J. Velôso
École CentraleSupélec, Paris-Saclay University
Gif-sur-Yvette, 91190, France
lucasjose.velosodesouza@student-cs.fr

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

We introduce Lensformer, a state-of-the-art transformer architecture that incorporates the lens equations directly into the architecture for the purpose of studying dark matter in the context of strong gravitational lensing. This architecture combines the strengths of Transformer models from natural language processing with the analytical rigor of Physics-Informed Neural Networks (PINNs). By putting the lensing equation into the design of the architecture, Lensformer is able to approximate the gravitational potential of the lensing galaxy. The physics-based features are then integrated into a Vision Transformer (ViT) neural network, which helps to provide a nuanced understanding when applied to various problems related to strong lensing. In this work, we consider an example of classifying between simulations of different models of dark matter. To validate the model, we benchmark Lensformer against other leading architectures and demonstrate that it exhibits superior performance.

1 Introduction

Despite being studied for over a century, it feels as if we are not much closer to knowing the underlying microphysical description of dark matter than we were at the time of its discovery. Indeed, countless experimental programs have embarked on searches for promising dark matter candidates, but to date, all efforts have come up empty [1, 2, 3, 4, 5, 6]. Another avenue to constrain and understand dark matter, beyond leveraging its (possible) couplings to the Standard Model, is to study its gravitational imprints. Indeed, a rampant area of research has been the application of data from strong galaxy-galaxy lensing images to study and constrain dark matter. This is a promising avenue to constrain dark matter since the extended lensing phenomenon is very sensitive to perturbations induced in the lensing potential due to the presence of substructure in the main halo, see for example [7, 8, 9, 10, 11, 12, 13]. Characteristic differences in the distribution and morphology of dark matter substructure can then be leveraged to potentially identify different models. This is, of course, easier

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

said than done. The inherent complexity makes extracting information about the sub-dominant lensing signal tricky.

Precisely because of the difficulties associated with using lensing data, machine learning algorithms have been studied extensively in this context. However, currently, there is only a small sample of high-quality lensing data available for training, but thankfully next-generation experiments such as the Vera Rubin Observatory (VRO) [14] and Euclid [15] are expected to provide thousands of images, heightening the need for refined analytical methods. However, in the meantime, simulations have become crucial in studying the interface of machine learning and strong lensing. Indeed, many studies have shown that machine learning is likely to be a powerful method to extract information about dark matter from lensing data, see for example [16, 17, 18, 19, 20]. In particular, these studies have consistently shown that convolutional neural networks (CNN) are ideally suited for strong lensing. This result isn't surprising since a CNN has two important features: 1) it leverages correlations in the image data and 2) it exhibits translation invariance. The major benefit here is that the CNN is able to integrate out irrelevant (i.e. redundant) "degrees of freedom" for free – precisely because CNNs are translationally invariant by construction. Indeed, building known redundancies in a given data set into an architecture by hand is known to increase model performance; in the context of lensing this has been demonstrated with equivariant neural networks in the context of domain adaptation for simulated strong lensing data [21].

At its heart, the success of CNNs and equivariant neural networks can be attributed to their ability to instill known structure from the data directly into the architecture. Along this same line of thought, it has recently been shown that one can also instill physical laws directly into one's architecture [22] to solve challenging problems; these are so-called physics-informed neural networks. In this context, we now introduce `Lensformer`, a neural network model inspired by the Vision Transformer (ViT) architecture [23] and with the *lensing equation* directly embedded into the architecture's structure. To demonstrate `Lensformer`'s superior performance, we benchmark it against other state-of-the-art architecture by classifying between mock Hubble observations of galaxy-galaxy strong lensing for different realizations of dark matter.

2 Dataset

To test our proposed architecture, we make simulations of galaxy-galaxy strong lensing data, which are meant to resemble data from observations with an HST-like survey. Our data sets are constructed with the publicly available code `lenstronomy`[24] where we make images sized 64×64 which are single channel. For modeling background galaxies, which we will lens, we implement a Sersic light profile. For our dark matter models, we have generated three different classes. The first class realizes the standard cold dark matter (CDM) scenario where the main dark matter halo, which is modeled with a spherical isothermal profile, contains subhalos drawn from a standard subhalo mass distribution (see [16] for further details). For our second model, we instead model the lensing signature of very light axion dark matter. Concretely, the axion simulations in this work correspond to dark matter with a particle mass $\sim 10^{-23}$ eV. In this range, the formation of substructure is highly suppressed, and the primary substructure observable comes from topological defects in the halo in the form of vortices [16]. As a final model for dark matter, we construct a realization of dark matter which contains no substructure. While not a realistic model, data clearly rule this out, this data set is useful since we are working in the realm of simulations as it serves as a useful scenario to compare to the other substructure classes. Our work generates a total of 400 simulations, split evenly between training and testing and between our three types of dark matter.

3 `Lensformer` Architecture

Extracting information about dark matter, via its substructure, from strong lensing images is a very high-dimensional problem. To help reduce this problem, we propose the `Lensformer` architecture as a novel approach to the study of dark matter from strong lensing data. Specifically, our architecture leverages two things: 1) a transformer architecture and 2) `Lensformer` directly incorporates the physics of strong lensing into the architecture through the `lens equation`. By endowing the architecture with the basic governing equation of gravitational lensing, the hope is that it will reduce the complexity (dimensionality) of the problem that has to be solved by the architecture. This is achieved through a two-pronged approach consisting of a specialized encoder and decoder.

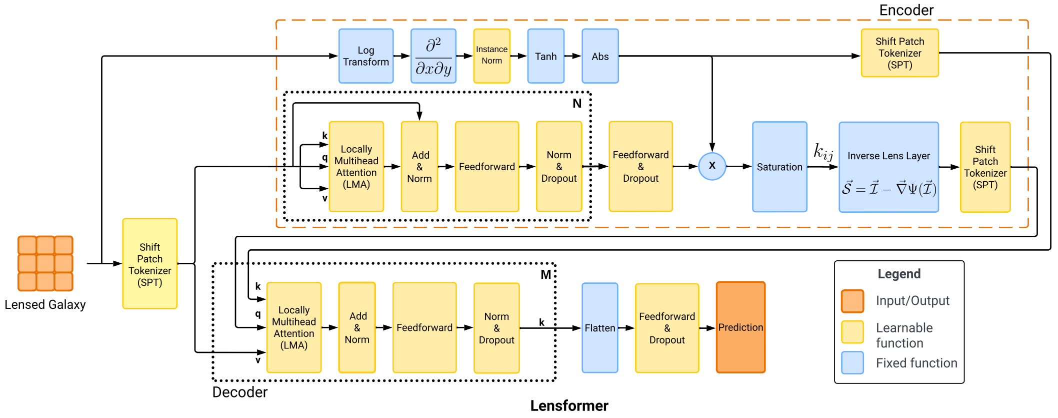

Figure 1: `Lensformer` Architecture.

3.1 Relativistic Physics-Informed Encoder

As mentioned earlier, the encoder leverages the governing equation of strong lensing. That is, it employs the *lens equation* (see [25] for more details) where we have additionally included the analytic form of the potential for a spherical isothermal profile as an ansatz [26]. The intent of this is to use this ansatz as a first-order approximation to the potential of the dark matter halo. This extra information can then be leveraged to reconstruct the source galaxy that is being lensed, but also enhance model performance.

In the Inverse Lens Layer, see Fig. 1, we use the *lens equation*, in its dimensionless form, which is given by the following relation: $\vec{\mathcal{S}} = \vec{\mathcal{I}} - \vec{\nabla}\Psi(\vec{\mathcal{I}})$[27]. In this equation, $\vec{\mathcal{S}} = (x_s, y_s)$ represents the dimensionless source vector position in the source plane, which corresponds to the position of the source galaxy. On the other hand, $\vec{\mathcal{I}} = (x_i, y_i)$ represents the dimensionless image vector position in the image plane, which corresponds to the image we observe. Finally, $\vec{\nabla}\Psi(\vec{\mathcal{I}}) = \big(\Psi_x(x_i, y_i), \Psi_y(x_i, y_i)\big)$ represents the gradient of the dimensionless gravitational potential produced by the lens, which in our case, includes both the lensing galaxy and the possible dark matter.

This equation involves three unknowns: the source position $\vec{\mathcal{S}}$, the image position $\vec{\mathcal{I}}$, and the gravitational potential of the system $\Psi(x_i, y_i)$. Yet, we only have knowledge of the produced image $\vec{\mathcal{I}}$. To estimate the position of the source galaxy $\vec{\mathcal{S}}$, we need to make assumptions about the potential of the system (i.e., the lensing galaxy plus dark matter). Here, we adopted the following *ansatz*:

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

Where $\Psi_{SIS}(x_i, y_i)$ is the potential of the Singular Isothermal Spherical (SIS) profile [26] and, then, $\Psi_{SIS}(x_i, y_i) = \sqrt{x_i^2 + y_i^2}$[28]; and the function $k(x_i, y_i)$ is a learnable function from the data. To predict the values of $k(x_i, y_i)$, we use the image of the lensed source galaxy and employ the architecture of a ViTSD (Vision Transformer for Small Datasets)[29]. This is done in conjunction with image transformations specifically designed to detect minute variations in the image gradient, potentially attributable to the presence of dark matter. We predict a corresponding k_{ij} value for each pixel (i, j), as is shown in Figure 1, in the encoder.

With the obtained values k_{ij}, we can estimate the potential $\Psi(x_i, y_i)$. Alongside the input image, this enables us to solve the gravitational lens equation and generate an estimated image of the source galaxy. This image then undergoes *Shifted Patch Tokenization (SPT)*[29] and serves as the *query* input for the decoder's transformer blocks, while the original image transformations enter the decoder as the *key* input.

Table 1: Performance Metrics of Different Models

Model Name	Accuracy	AUC	F1 Score		
			No Subs.	CDM	Axion
Resnet	0.511	0.690	0.564	0.375	0.596
Inception	0.404	0.583	0.435	0.331	0.453
CvT	0.637	0.812	0.677	0.444	0.755
ViT	0.573	0.740	0.608	0.485	0.652
CaiT	0.556	0.745	0.642	0.376	0.642
ViTSD	0.729	0.873	0.780	0.609	0.799
Lensformer	**0.903**	**0.966**	**0.916**	**0.936**	**0.857**

3.2 Decoder

For the decoder part of the architecture, Lensformer adopts techniques like *Shifted Patch Tokenization (SPT)*[29] and *Locality Self-Attention (LSA)*[29]. By using the original lensed image as the *query* with the image transformation as the *key*, the decoder can generate more accurate and physics-consistent representations of lensing phenomena. These techniques not only improve the model's effectiveness but also increase its adaptability to different types of data.

4 Experiment

To rigorously assess the effectiveness of Lensformer, a benchmarking study was conducted against other top vision transformer models using the created dataset and NVIDIA A100 GPUs. All models were standardized to have around 11 million parameters and were trained for 300 epochs with an initial learning rate of 10^{-3}. The AdamW [30] optimization algorithm was used, and a dynamic learning rate adjustment strategy was employed, reducing the learning rate by half if no performance improvement was observed over 10 consecutive epochs. Performance was evaluated using a Cross-Entropy-based loss function, a standard metric for classification tasks. The roster of models selected for this study is diverse, including Vision Transformer (ViT)[23]; Vision Transformers for Small Datasets (ViTSD)[29]; Convolutional Transformer (CvT)[31]; Class-Attention in Image Transformers (CaiT)[32]; Residual Networks (ResNet)[33]; and Inception[34].

5 Results and Discussion

In a comprehensive evaluation using the created dataset, Lensformer outperformed other state-of-the-art models in key metrics for the toy example of classifying between strong lensing simulations with different underlying dark matter physics, i.e. ΛCDM, ultra-light axions, and no substructure classes. We have compiled the accuracy, AUC, and F1 score for these studies in Table 1. Furthermore, we show the evolution of the test set accuracy and training loss together with ROC curves in Fig. 2. With an accuracy of 90.3%, a ROC-AUC score of 0.966, and F1 scores of 0.916, 0.936, and 0.857 for the no substructure, CDM, and ultra-light axion classes, respectively, Lensformer demonstrates significantly better performance than both transformer-based and CNN-based models. The second-best performer on the data set was ViTSD which had a significantly degraded performance by comparison with an accuracy of 72.9% and ROC-AUC score of 0.873.

The superior performance of Lensformer suggests that the model's physics-informed architecture is able to appreciably decrease the dimensionality of the effective task it is asked to solve. While this will require more testing to fully understand, this increase in performance can likely be compared to the successes seen with applications of both CNNs and equivariant neural networks that improve model performance precisely because they are better structured for the problem at hand. Indeed, this is particularly useful when handling sparse or limited datasets, suggesting that tools like Lensformer could be useful for other specialized tasks in cosmology and astrophysics in which some previous domain knowledge can be leveraged to increase model performance.

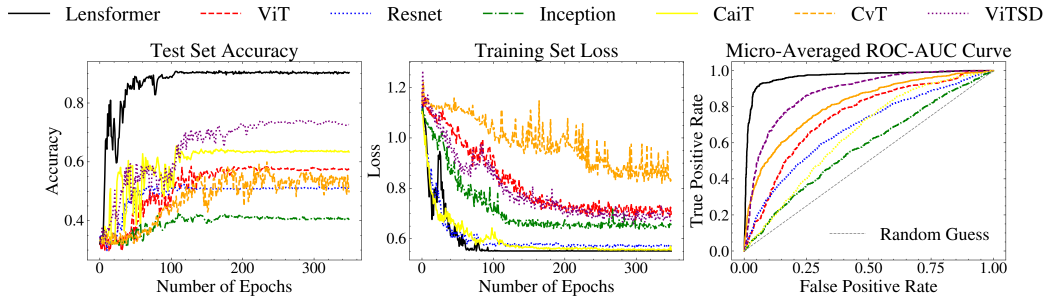

Figure 2: Model Performance Metrics. The left panel shows test accuracy, the middle panel shows training loss, and the right panel displays the Micro-Averaged ROC-AUC curves for various models. A gray 'Random Guess' line provides a baseline on the ROC-AUC plot.

6 Acknowledgments

L.J. was a participant in the ML4SCI 2023 Google Summer of Code program. L.J. was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001. S.G. was supported in part by U.S. National Science Foundation award No. 2108645. M.W.T. was partially supported in part by U.S. National Science Foundation award No. 2108866. Portions of this work were conducted in MIT's Center for Theoretical Physics and partially supported by the U.S. Department of Energy under grant Contract Number DE-SC0012567.

References

[1] A. K. Drukier, Katherine Freese, and D. N. Spergel. Detecting Cold Dark Matter Candidates. *Phys. Rev.*, D33:3495–3508, 1986.

[2] Mark W. Goodman and Edward Witten. Detectability of Certain Dark Matter Candidates. *Phys. Rev.*, D31:3059, 1985. [,325(1984)].

[3] D. S. Akerib et al. Results from a search for dark matter in the complete LUX exposure. *Phys. Rev. Lett.*, 118(2):021303, 2017.

[4] Xiangyi Cui et al. Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment. *Phys. Rev. Lett.*, 119(18):181302, 2017.

[5] E. Aprile et al. Dark Matter Search Results from a One Ton-Year Exposure of XENON1T. *Phys. Rev. Lett.*, 121(11):111302, 2018.

[6] Morad Aaboud et al. Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{s} = 13$ TeV pp collision data collected by the ATLAS detector. *JHEP*, 05:142, 2019.

[7] S. Mao and P. Schneider. Evidence for Substructure in lens galaxies. *MNRAS*, 295:587–594, 1998. arXiv.

[8] J.W. Hsueh et al. SHARP - IV. An apparent flux ratio anomaly resolved by the edge-on disc in B0712+472. *MNRAS*, 469(3):3713–3721, 2017. arXiv.

[9] N. Dalal and C.S. Kochanek. Direct Detection of CDM Substructure. *ApJ*, 572:25–33, 2002. arXiv.

[10] Y.D. Hezaveh et al. Detection of Lensing Substructure Using ALMA Observations of the Dusty Galaxy SDP.81. *ApJ*, 823(1):37–56, 2016. arXiv.

[11] S. Vegetti and L.V.E. Koopmans. Bayesian strong gravitational-lens modelling on adaptive grids: objective detection of mass substructure in Galaxies. *MNRAS*, 392(3):945–963, 2009. arXiv.

[12] L.V.E. Koopmans. Gravitational imaging of cold dark matter substructures. *MNRAS*, 363(4):1136–1144, 2005. Oxford Journals.

[13] S. Vegetti and L.V.E. Koopmans. Statistics of mass substructure from strong gravitational lensing: quantifying the mass fraction and mass function. *MNRAS*, 400:1583–1592, 2009. arXiv.

[14] Aprajita Verma, Thomas Collett, Graham P. Smith, Strong Lensing Science Collaboration, and the DESC Strong Lensing Science Working Group. Strong lensing considerations for the lsst observing strategy, 2019.

[15] Masamune Oguri and Philip J. Marshall. Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys. *Monthly Notices of the Royal Astronomical Society*, pages no–no, apr 2010.

[16] Stephon Alexander, Sergei Gleyzer, Evan McDonough, Michael W. Toomey, and Emanuele Usai. Deep learning the morphology of dark matter substructure. *The Astrophysical Journal*, 893(1):15, apr 2020.

[17] Siddharth Mishra-Sharma and Ge Yang. Strong Lensing Source Reconstruction Using Continuous Neural Fields. In *39th International Conference on Machine Learning Conference*, 6 2022.

[18] Gemma Zhang, Siddharth Mishra-Sharma, and Cora Dvorkin. Inferring subhalo effective density slopes from strong lensing observations with neural likelihood-ratio estimation. *Monthly Notices of the Royal Astronomical Society*, 517(3):4317–4326, 10 2022.

[19] Thuruthipilly, Hareesh, Zadrozny, Adam, Pollo, Agnieszka, and Biesiada, Marek. Finding strong gravitational lenses through self-attention - study based on the bologna lens challenge. *A&A*, 664:A4, 2022.

[20] Ana Diaz Rivero and Cora Dvorkin. Direct Detection of Dark Matter Substructure in Strong Lens Images with Convolutional Neural Networks. *Phys. Rev. D*, 101(2):023515, 2020.

[21] Stephon Alexander, Sergei Gleyzer, Hanna Parul, Pranath Reddy, Marcos Tidball, and Michael W. Toomey. Domain Adaptation for Simulation-based Dark Matter Searches with Strong Gravitational Lensing. *Astrophys. J.*, 954(1):28, 2023.

[22] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[23] 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.

[24] Simon Birrer, Adam Amara, and Alexandre Refregier. Gravitational lens modeling with basis sets. *The Astrophysical Journal*, 813(2):102, nov 2015.

[25] Ramesh Narayan and Matthias Bartelmann. Lectures on Gravitational Lensing. 1997.

[26] Charles R. Keeton. A catalog of mass models for gravitational lensing, 2002.

[27] Ramesh Narayan and Matthias Bartelmann. Lectures on gravitational lensing, 1997.

[28] R. Kormann, P. Schneider, and M. Bartelmann. Isothermal elliptical gravitational lens models. *Astronomy and Astrophysics*, 284:285–299, 1994. ISSN 0004-6361.

[29] Seung Hoon Lee, Seunghyun Lee, and Byung Cheol Song. Vision transformer for small-size datasets, 2021.

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

[31] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. Cvt: Introducing convolutions to vision transformers, 2021.

[32] Hugo Touvron, Matthieu Cord, Alexandre Sablayrolles, Gabriel Synnaeve, and Hervé Jégou. Going deeper with image transformers, 2021.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[34] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.