Fast kinematics modeling for conjunction with lens image modeling

Matthew R. Gomer
STAR Institute
Université de Liège
4000 Liège, Belgium
mgomer@uliege.be

Luca Biggio
Eidgenössische Technische Hochschule Zürich
CH-8092 Zürich, Switzerland
luca.biggio@inf.ethz.ch

Sebastian Ertl
Max-Planck-Institut für Astrophysik
Technische Universität München
85748 Garching, Germany
ertlseb@mpa-garching.mpg.de

Han Wang
Max-Planck-Institut für Astrophysik
Technische Universität München
85748 Garching, Germany
wanghan@mpa-garching.mpg.de

Aymeric Galan
Institute of Physics, Laboratory of Astrophysics
Ecole Polytechnique Fédérale de Lausanne (EPFL)
1290 Versoix, Switzerland
aymeric.galan@epfl.ch

Lyne Van de Vyvere
STAR Institute
Université de Liège
4000 Liège, Belgium
lyne.vandevyvere@uliege.be

Dominique Sluse
STAR Institute
Université de Liège
4000 Liège, Belgium
dsluse@uliege.be

Georgios Vernardos
Institute of Physics, Laboratory of Astrophysics
Ecole Polytechnique Fédérale de Lausanne (EPFL)
1290 Versoix, Switzerland
georgios.vernardos@epfl.ch

Sherry Suyu
Max-Planck-Institut für Astrophysik
Technische Universität München
85748 Garching, Germany
Academia Sinica Institute of Astronomy and Astrophysics (ASIAA)
Taipei 10617, Taiwan
suyu@mpa-garching.mpg.de

Abstract

Galaxy kinematics modeling is currently the computational bottleneck for a joint gravitational lensing+kinematics modeling procedure. We present as a proof of concept the Stellar Kinematics Neural Network (SKiNN), which emulates kinematics calculations for the context of gravitational lens modeling. After a one-time upfront training cost, SKiNN creates velocity dispersion images which are accurate to $\lesssim 1\%$ within the region of interest at a speed $\mathcal{O}(10^2 - 10^3)$ times faster than existing kinematics modeling methods. This speedup makes it feasible to jointly model lensing data with spatially resolved kinematic data, which corrects for the largest source of uncertainty in the determination of the Hubble constant.

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

1 Introduction

Gravitational lensing offers a distance-ladder-independent method to directly measure the Hubble constant, H_0, by measuring the differences in arrival time between different multiple images of the same distant source. Comparison of these time delays to those predicted from a model for the mass distribution of the lens (typically a galaxy) gives a direct measure of distance and therefore H_0. Unfortunately, lensing degeneracies exist which allow different mass distributions to reproduce the same resulting image. To overcome these degeneracies and recover a precise H_0, kinematic constraints are required to provide an independent measure of the lens mass distribution.

Until recently, the best available kinematic data were slit aperture constraints, which give a single velocity dispersion (second moment of the velocity distribution function) for the whole system. Recently, spatially resolved kinematics of lensing galaxies have become available through telescopes like the James Webb Space Telescope (JWST). Lens modelers have traditionally used the spherical Jeans model to describe single-slit kinematics [e.g., Suyu et al., 2010, Sonnenfeld et al., 2012, Wong et al., 2017, Birrer et al., 2019, Rusu et al., 2020], which in addition to lacking self-consistency with the elliptical lens mass model, is too simplistic for spatially resolved kinematic data. The next generalization to make from the spherical case is the axisymmetric case, for which software such as Jeans Anisotropic MGE [JAM; Cappellari, 2008] can model galaxy kinematics (using the Multi-Gaussian Expansion method [Emsellem et al., 1994, Cappellari, 2002]), albeit at significant computational cost.

Existing frameworks to calculate kinematics in conjunction with lens modeling come with limitations. van de Ven et al. [2010] first compared an axisymmetric kinematics model self-consistently with a lens model, but these models were fit separately. Barnabè et al. [2012] introduced a joint modeling code designed for galaxy structure research but not used for H_0 determination. Recently Yıldırım et al. [2020, 2021] expanded the framework using JAM to be compatible with cosmological parameter determination. However, the combined modeling procedure is still computationally expensive, with the bottleneck being that JAM must be called for each evaluation of the likelihood, which itself is often sampled via a Markov Chain Monte Carlo (MCMC) to estimate uncertainties. To reduce costs, modelers often first optimize the lens model, then combine the kinematics to the likelihood post-processing. To efficiently explore the full parameter space of both models simultaneously requires a significant speed improvement for kinematic modeling.

Our insight is to emulate JAM using a neural network (NN), which we call SKiNN (Stellar Kinematics Neural Network). Once trained, SKiNN is able to create the same output as JAM in the short time it takes to call the network. This factor of 300 speed increase makes it possible to include kinematic modeling within a joint framework with lens modeling. We note that SKiNN is trained using JAM-generated data, and as such is designed only to emulate the JAM modeling procedure rather than being applied directly to the observations. In this way, confining the machine learning aspect to JAM emulation exploits the speed and versatility of NNs while retaining the physics of the overall model. Figure 1 gives a summary of SKiNN's role in a joint modeling framework.

In this paper, we focus on the machine learning aspect of the modeling framework. In an upcoming work, we will present and discuss the full modeling procedure for joint lensing+kinematics using our NN-based technique.

2 Related Works

The general method to speed up modeling code by replacing the expensive piece with a NN has been applied before in adjacent fields. NNs have been used to replace expensive solver operations in cosmological applications [Albers et al., 2019, Bonici et al., 2022], stellar population synthesis in spectral modeling [Alsing et al., 2020], and even to speed up gravitational lens modeling itself [Hezaveh et al., 2017, Perreault Levasseur et al., 2017, Pearson et al., 2019, Schuldt et al., 2021, 2022, Park et al., 2021]. This work applies this strategy for the first time to the kinematic modeling aspect within the framework of lens modeling.

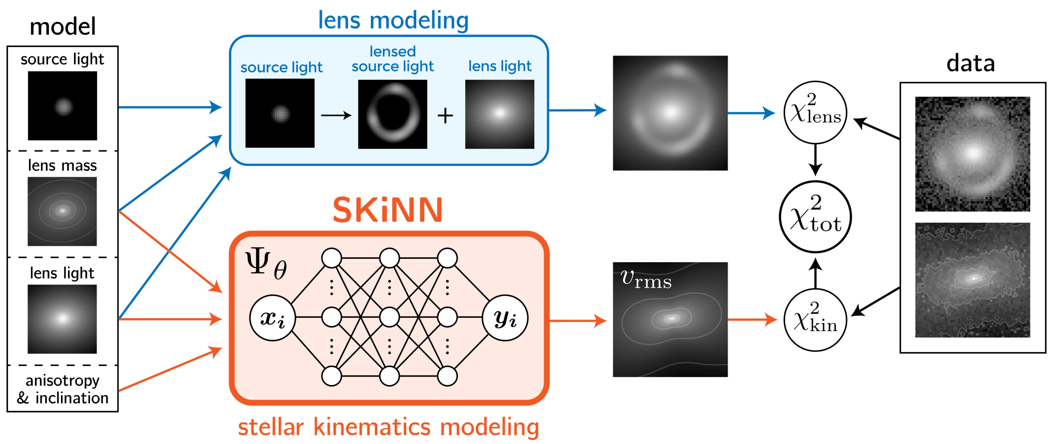

Figure 1: SKiNN's role in a joint modeling framework. Traditional lens modeling (blue) uses a model for a source, lens mass, and lens light to construct an image which is compared against the imaging data (top). SKiNN (orange) inputs lens mass, lens light, and kinematics parameters to produce a $v_{\rm rms}$ image which is compared against the kinematic data (bottom). The combined $\chi^2_{\rm tot}$ can be minimized to jointly optimize the combined model.

3 Methods

In essence, SKiNN precalculates JAM for use in lensing applications. Like JAM, the input is a vector of parameters which describes the lensing galaxy, and the output is a $v_{\rm rms}$ image, the quadratic sum of velocity dispersion and rotational velocity. More concretely, SKiNN can be seen as a function $\Psi_\theta : \mathbb{R}^8 \to \mathbb{R}^{d\times d}$, mapping an 8-dimensional vector of galaxy parameters into a $d \times d$ image of $v_{\rm rms}$. Given a training dataset $\mathcal{D} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$ where $\mathbf{x} \in \mathbb{R}^8$, $\mathbf{y} \in \mathbb{R}^{d\times d}$ and N is the size of the dataset, the training process consists in finding an optimal set of parameters θ^*, such that a loss function $\mathcal{L}$, measuring the performance of Ψ on $\mathcal{D}$, is minimized. We use the standard mean-squared-error as the loss function, which we minimize using the Adam optimizer [Kingma and Ba, 2014][1]. The design of the architecture of SKiNN consists of 5 blocks each comprising two 2-dimensional convolutional layers followed by an upsampling layer and a ReLU nonlinearity.

Data set construction Our data set is constructed using JAM, which SKiNN is intended to mimic. JAM [Cappellari, 2008] works by decomposing the sky-projected (2D) mass and light distributions of a galaxy into MGEs, for which deprojection into 3D is calculable given an inclination of the galaxy, i. With an anisotropy, β, which describes the mixture of orbits aligned with the main ellipsoidal axis relative to those in other directions, the axisymmetric Jeans equations are solved in cylindrical coordinates, and the velocity is projected into the sky plane. For this proof of concept, we restrict ourselves to a single (but widely used) class of profiles for the lens mass and light, using Power law Elliptical Mass Distributions [PEMD Barkana, 1998] and elliptical Sérsic light profiles [Sérsic, 1963]. In total, these distributions are described with 6 parameters (see Appendix A), which, together with i and β, constitute the 8 parameters of the $\mathbf{x}$ vector.

From these parameters, we use the GLEE software [Gravitational Lens Efficient Explorer, Suyu et al., 2010, 2012] to emulate the light and mass profiles which are then fed into JAM, and hence create an injective mapping from the $\mathbf{x}$ vector to the $v_{\rm rms}$ image. The ranges of parameters we probed are physically motivated (see Appendix A). We create $v_{\rm rms}$ images of 551×551 pixels at a larger size and higher resolution than real data. Image creation through JAM is quite slow, taking ~ 15 seconds per image.

We construct our full data set with a total of 5000 input-output pairs. Our training set is a random subsample of $N = 4000$ of these, while we keep 500 instances for the validation set and 500 for

[1] Overall our final model counts 7,065,451 trainable parameters. The pre-trained weights of our best model are available upon reasonable request, with our intention being to make the final version of SKiNN open-source.

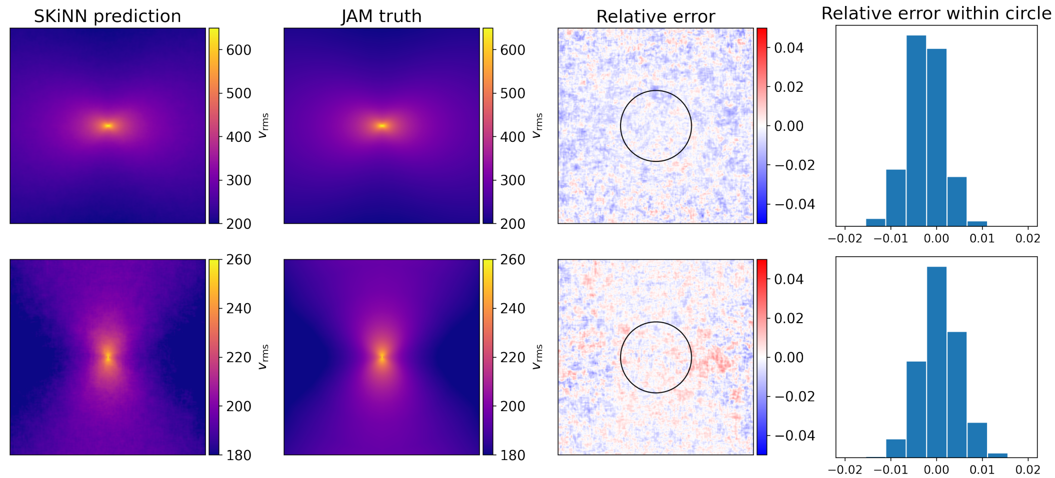

Figure 2: Two example $v_{\rm rms}$ images created by SKiNN and compared to the JAM ground truth. The black circle indicates the innermost two arcseconds, where real data is most constraining. Within this region, the error in the emulation is less than 1%, as shown by the histograms on the right.

the test set. As the training proceeds we constantly monitor the performance of the model on the validation set and save its weights whenever an improvement in the validation accuracy is observed.

Computational details The network is implemented in Pytorch [Paszke et al., 2019] and is trained on 5 Tesla P100 GPUs with 16GB memory each. For efficient multi-GPU training we wrap our model into Pytorch Lightning [Falcon et al., 2019], a recent Pytorch-compatible framework for large scale GPU-based training of NNs.

4 Results

Speed Training our model in the above setting requires about one day, the stopping criterion being the validation loss not decreasing for a certain number of consecutive epochs. Once trained, the model can be run either on CPU or on a single GPU (the latter being roughly 10 times faster than the former). When run on the GPU, the model infers the velocity image associated with an input parameter vector in about 50 ms, which is roughly 300 times faster than JAM. Clearly, this advantage comes at the price of a rather expensive training phase which, however, has to be performed only once.

Emulator accuracy We assess the performance of our model by computing the relative error for each pixel for the 500 input-output pairs in the test set. Two example generated $v_{\rm rms}$ images with residuals are shown in Fig. 2. Parts of the image can have error of $\sim$few percent for some images, but the particularly important region to match is the innermost region of the image, where real data is able to constrain. Approximated as the black circle in Fig. 2, the error within this region is within $\pm 1\%$ for almost all pixels. Across the 500 test images, the median absolute error averaged over an entire image is around 0.47% while the 90th percentile is approximately 1.1%. This is lower than the typical systematic uncertainty of $\gtrsim 2\%$ in the velocity measurements [e.g., Collett et al., 2018]. The model is thus capable of capturing fine details of the ground-truth velocity images while generating them much faster for applications to real observations of galaxies.

Limitations and outlook This is only a preliminary version of SKiNN, which we intend to update with other physically motivated mass models. We are also considering possible ways to speed up the training of the network by exploiting the symmetry of the $v_{\rm rms}$ images.

In this current iteration, our result is limited to the specifics of our training set, that is, we can only use this NN to quickly calculate kinematics for axisymmetric systems with a PEMD mass profile and Sérsic light profile. In the short term, this technique can be expanded to more general mass and light models; one only needs to construct new data sets using JAM and retrain the NN. In the longer term, this general method can be reapplied using other kinematics calculation software to go beyond the axisymmetric assumption inherent in JAM.

Joining SKiNN to the lens modeling procedure is currently undergoing implementation in the `lenstronomy` [Birrer and Amara, 2018, Birrer et al., 2021], GLEE [Suyu and Halkola, 2010, Suyu et al., 2012] and GLaD[2] [Chirivì et al., 2020, Yıldırım et al., 2020] software packages, consistent with the framework of Yıldırım et al. [2020]. This joint modeling framework will efficiently tackle the strongest source of bias in lensing studies thanks to the constraining power of spatially resolved kinematics [Birrer et al., 2020, Yıldırım et al., 2021]. The resulting determination of cosmological parameters will then be more accurate.

A joint lensing+SKiNN framework will represent a methodology that likely has analogous applications in other fields. For any modeling context in which a piece of the model is calculable, but too computationally expensive to implement and therefore excluded, a similar technique to this work could be applied: the slow piece can be replaced by a trained NN while retaining the existing physical aspects of the overall model.

5 Conclusions

We present SKiNN, a NN created to replace the slow process of galaxy kinematics emulation which can currently be done through JAM. The process of creating a training set and fitting the NN is a significant computational investment, but the resulting trained network is able to create similar outputs to JAM at greatly increased speed. The accuracy of the emulator is $< 1\%$ in the region comparable with real data, with median error $\sim 1\%$ across entire images. The roughly million-fold increase in speed makes it possible to model kinematics jointly with existing lens modeling, for which we are currently building infrastructure, with the goal being to perform cosmological analysis accurately by using lensing and kinematic data together to break degeneracies. We believe this method can be expanded with more general training sets, and the general strategy likely has applications in other fields.

Impact statement

The general methodology used in this work has had applications for related fields, as discussed in Sect. 2, and therefore likely can be applied to additional fields. Where applicable, the net effect would be to speed up existing modeling, or perhaps add pieces to existing models which were previously computationally unfeasible. While we do not foresee any ethical concerns with these applications, it is possible that the general methodology could be used to speed up hypothetical modeling procedures which are themselves unethical.

References

J. Albers, C. Fidler, J. Lesgourgues, N. Schöneberg, and J. Torrado. CosmicNet. Part I. Physics-driven implementation of neural networks within Einstein-Boltzmann Solvers. J. Cosmology Astropart. Phys., 2019(9):028, Sept. 2019. doi: 10.1088/1475-7516/2019/09/028.

J. Alsing, H. Peiris, J. Leja, C. Hahn, R. Tojeiro, D. Mortlock, B. Leistedt, B. D. Johnson, and C. Conroy. SPECULATOR: Emulating Stellar Population Synthesis for Fast and Accurate Galaxy Spectra and Photometry. ApJS, 249(1):5, July 2020. doi: 10.3847/1538-4365/ab917f.

R. Barkana. Fast Calculation of a Family of Elliptical Mass Gravitational Lens Models. ApJ, 502(2): 531–537, Aug. 1998. doi: 10.1086/305950.

M. Barnabè, A. A. Dutton, P. J. Marshall, M. W. Auger, B. J. Brewer, T. Treu, A. S. Bolton, D. C. Koo, and L. V. E. Koopmans. The SWELLS survey - IV. Precision measurements of the stellar and dark matter distributions in a spiral lens galaxy. MNRAS, 423(2):1073–1088, June 2012. doi: 10.1111/j.1365-2966.2012.20934.x.

S. Birrer and A. Amara. lenstronomy: Multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, 22:189–201, Dec. 2018. doi: 10.1016/j.dark.2018.11.002.

[2] Gravitational Lensing and Dynamics

S. Birrer, T. Treu, C. E. Rusu, V. Bonvin, C. D. Fassnacht, J. H. H. Chan, A. Agnello, A. J. Shajib, G. C. F. Chen, M. Auger, F. Courbin, S. Hilbert, D. Sluse, S. H. Suyu, K. C. Wong, P. Marshall, B. C. Lemaux, and G. Meylan. H0LiCOW - IX. Cosmographic analysis of the doubly imaged quasar SDSS 1206+4332 and a new measurement of the Hubble constant. MNRAS, 484(4): 4726–4753, Apr. 2019. doi: 10.1093/mnras/stz200.

S. Birrer, A. J. Shajib, A. Galan, M. Millon, T. Treu, A. Agnello, M. Auger, G. C. F. Chen, L. Christensen, T. Collett, F. Courbin, C. D. Fassnacht, L. V. E. Koopmans, P. J. Marshall, J. W. Park, C. E. Rusu, D. Sluse, C. Spiniello, S. H. Suyu, S. Wagner-Carena, K. C. Wong, M. Barnabè, A. S. Bolton, O. Czoske, X. Ding, J. A. Frieman, and L. Van de Vyvere. TDCOSMO. IV. Hierarchical time-delay cosmography - joint inference of the Hubble constant and galaxy density profiles. A&A, 643:A165, Nov. 2020. doi: 10.1051/0004-6361/202038861.

S. Birrer, A. Shajib, D. Gilman, A. Galan, J. Aalbers, M. Millon, R. Morgan, G. Pagano, J. Park, L. Teodori, N. Tessore, M. Ueland, L. Van de Vyvere, S. Wagner-Carena, E. Wempe, L. Yang, X. Ding, T. Schmidt, D. Sluse, M. Zhang, and A. Amara. lenstronomy II: A gravitational lensing software ecosystem. *The Journal of Open Source Software*, 6(62):3283, June 2021. doi: 10.21105/joss.03283.

M. Bonici, L. Biggio, C. Carbone, and L. Guzzo. Fast emulation of two-point angular statistics for photometric galaxy surveys. *arXiv e-prints*, art. arXiv:2206.14208, June 2022.

M. Cappellari. Efficient multi-Gaussian expansion of galaxies. MNRAS, 333(2):400–410, June 2002. doi: 10.1046/j.1365-8711.2002.05412.x.

M. Cappellari. Measuring the inclination and mass-to-light ratio of axisymmetric galaxies via anisotropic Jeans models of stellar kinematics. MNRAS, 390(1):71–86, Oct. 2008. doi: 10.1111/j.1365-2966.2008.13754.x.

M. Cappellari, E. Emsellem, R. Bacon, M. Bureau, R. L. Davies, P. T. de Zeeuw, J. Falcón-Barroso, D. Krajnović, H. Kuntschner, R. M. McDermid, R. F. Peletier, M. Sarzi, R. C. E. van den Bosch, and G. van de Ven. The SAURON project - X. The orbital anisotropy of elliptical and lenticular galaxies: revisiting the (V/σ, ϵ) diagram with integral-field stellar kinematics. MNRAS, 379(2): 418–444, Aug. 2007. doi: 10.1111/j.1365-2966.2007.11963.x.

G. Chirivì, A. Yıldırım, S. H. Suyu, and A. Halkola. Gravitational Lensing and Dynamics (GLaD): combined analysis to unveil properties of high-redshift galaxies. A&A, 643:A135, Nov. 2020. doi: 10.1051/0004-6361/202037929.

T. E. Collett, L. J. Oldham, R. J. Smith, M. W. Auger, K. B. Westfall, D. Bacon, R. C. Nichol, K. L. Masters, K. Koyama, and R. van den Bosch. A precise extragalactic test of General Relativity. *Science*, 360(6395):1342–1346, June 2018. doi: 10.1126/science.aao2469.

X. Ding, T. Treu, S. Birrer, G. C. F. Chen, J. Coles, P. Denzel, M. Frigo, A. Galan, P. J. Marshall, M. Millon, A. More, A. J. Shajib, D. Sluse, H. Tak, D. Xu, M. W. Auger, V. Bonvin, H. Chand, F. Courbin, G. Despali, C. D. Fassnacht, D. Gilman, S. Hilbert, S. R. Kumar, J. Y. Y. Lin, J. W. Park, P. Saha, S. Vegetti, L. Van de Vyvere, and L. L. R. Williams. Time delay lens modelling challenge. MNRAS, 503(1):1096–1123, May 2021. doi: 10.1093/mnras/stab484.

E. Emsellem, G. Monnet, and R. Bacon. The multi-gaussian expansion method: a tool for building realistic photometric and kinematical models of stellar systems I. The formalism. A&A, 285: 723–738, May 1994.

W. Falcon et al. Pytorch lightning. 2019. URL `https://www.pytorchlightning.ai`.

M. R. Gomer, D. Sluse, L. van de Vyvere, S. Birrer, and F. Courbin. TDCOSMO X: A key test of systematics in the hierarchical method of time-delay cosmography. *arXiv e-prints*, art. arXiv:2209.02076, Sept. 2022.

Y. D. Hezaveh, L. Perreault Levasseur, and P. J. Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. Nature, 548(7669):555–557, Aug. 2017. doi: 10.1038/nature23463.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, Dec. 2014.

J. W. Park, S. Wagner-Carena, S. Birrer, P. J. Marshall, J. Y.-Y. Lin, A. Roodman, and LSST Dark Energy Science Collaboration. Large-scale Gravitational Lens Modeling with Bayesian Neural Networks for Accurate and Precise Inference of the Hubble Constant. ApJ, 910(1):39, Mar. 2021. doi: 10.3847/1538-4357/abdfc4.

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

J. Pearson, N. Li, and S. Dye. The use of convolutional neural networks for modelling large optically-selected strong galaxy-lens samples. MNRAS, 488(1):991–1004, Sept. 2019. doi: 10.1093/mnras/stz1750.

L. Perreault Levasseur, Y. D. Hezaveh, and R. H. Wechsler. Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing. ApJ, 850(1):L7, Nov. 2017. doi: 10.3847/2041-8213/aa9704.

C. E. Rusu, K. C. Wong, V. Bonvin, D. Sluse, S. H. Suyu, C. D. Fassnacht, J. H. H. Chan, S. Hilbert, M. W. Auger, A. Sonnenfeld, S. Birrer, F. Courbin, T. Treu, G. C. F. Chen, A. Halkola, L. V. E. Koopmans, P. J. Marshall, and A. J. Shajib. H0LiCOW XII. Lens mass model of WFI2033-4723 and blind measurement of its time-delay distance and H_0. MNRAS, 498(1):1440–1468, Oct. 2020. doi: 10.1093/mnras/stz3451.

S. Schuldt, S. H. Suyu, T. Meinhardt, L. Leal-Taixé, R. Cañameras, S. Taubenberger, and A. Halkola. HOLISMOKES. IV. Efficient mass modeling of strong lenses through deep learning. A&A, 646: A126, Feb. 2021. doi: 10.1051/0004-6361/202039574.

S. Schuldt, R. Cañameras, Y. Shu, S. H. Suyu, S. Taubenberger, T. Meinhardt, and L. Leal-Taixé. HOLISMOKES – IX. Neural network inference of strong-lens parameters and uncertainties from ground-based images. *arXiv e-prints*, art. arXiv:2206.11279, June 2022.

J. L. Sérsic. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina*, 6:41–43, Feb. 1963.

A. Sonnenfeld, T. Treu, R. Gavazzi, P. J. Marshall, M. W. Auger, S. H. Suyu, L. V. E. Koopmans, and A. S. Bolton. Evidence for Dark Matter Contraction and a Salpeter Initial Mass Function in a Massive Early-type Galaxy. ApJ, 752(2):163, June 2012. doi: 10.1088/0004-637X/752/2/163.

S. H. Suyu and A. Halkola. The halos of satellite galaxies: the companion of the massive elliptical lens SL2S J08544-0121. A&A, 524:A94, Dec. 2010. doi: 10.1051/0004-6361/201015481.

S. H. Suyu, P. J. Marshall, M. W. Auger, S. Hilbert, R. D. Blandford, L. V. E. Koopmans, C. D. Fassnacht, and T. Treu. Dissecting the Gravitational lens B1608+656. II. Precision Measurements of the Hubble Constant, Spatial Curvature, and the Dark Energy Equation of State. ApJ, 711(1): 201–221, Mar. 2010. doi: 10.1088/0004-637X/711/1/201.

S. H. Suyu, S. W. Hensel, J. P. McKean, C. D. Fassnacht, T. Treu, A. Halkola, M. Norbury, N. Jackson, P. Schneider, D. Thompson, M. W. Auger, L. V. E. Koopmans, and K. Matthews. Disentangling Baryons and Dark Matter in the Spiral Gravitational Lens B1933+503. ApJ, 750(1):10, May 2012. doi: 10.1088/0004-637X/750/1/10.

G. van de Ven, J. Falcón-Barroso, R. M. McDermid, M. Cappellari, B. W. Miller, and P. T. de Zeeuw. The Einstein Cross: Constraint on Dark Matter from Stellar Dynamics and Gravitational Lensing. ApJ, 719(2):1481–1496, Aug. 2010. doi: 10.1088/0004-637X/719/2/1481.

K. C. Wong, S. H. Suyu, M. W. Auger, V. Bonvin, F. Courbin, C. D. Fassnacht, A. Halkola, C. E. Rusu, D. Sluse, A. Sonnenfeld, T. Treu, T. E. Collett, S. Hilbert, L. V. E. Koopmans, P. J. Marshall, and N. Rumbaugh. H0LiCOW - IV. Lens mass model of HE 0435-1223 and blind measurement of its time-delay distance for cosmology. MNRAS, 465(4):4895–4913, Mar. 2017. doi: 10.1093/mnras/stw3077.

A. Yıldırım, S. H. Suyu, and A. Halkola. Time-delay cosmographic forecasts with strong lensing and JWST stellar kinematics. MNRAS, 493(4):4783–4807, Apr. 2020. doi: 10.1093/mnras/staa498.

A. Yıldırım, S. H. Suyu, G. C. F. Chen, and E. Komatsu. TDCOSMO VIII: Cosmological distance measurements in light of the mass-sheet degeneracy – forecasts from strong lensing and IFU stellar kinematics. *arXiv e-prints*, art. arXiv:2109.14615, Sept. 2021.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] We intend to make the final version open source. The current version is available upon reasonable request
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] Section 3
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Specifically, errors with respect to emulation of images, Section 4
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Section 3
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] JAM and GLEE modeling codes
   (b) Did you mention the license of the assets? [N/A] JAM and GLEE are open-source
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Appendix

Lens profile parameters Here we describe the profiles used for the mass and light distributions in this work. The mass is a PEMD, with convergence (i.e., dimensionless surface density) described as

$$\kappa_{\mathrm{PEMD}}(x,y) = \theta_{\mathrm{E}}^{\gamma_{\mathrm{PL}}-1}\left(\frac{3-\gamma_{\mathrm{PL}}}{1+q_m}\right)\left(x^2+\frac{y^2}{q_m^2}\right)^{\frac{1-\gamma_{\mathrm{PL}}}{2}}, \tag{1}$$

with Einstein radius θ_{E}, axis ratio q_m, and slope of the 3D mass density γ_{PL}. For the light, we use an elliptical Sérsic profile,

$$I(x,y) = A\exp\left[-k\left\{\left(\frac{\sqrt{x^2+(\frac{y}{q_L^2})^2}}{r_{\mathrm{eff}}}\right)^{1/n}-1\right\}\right], \tag{2}$$

with effective radius r_{eff}, Sérsic index n, and axis ratio q_L (the normalization A is irrelevant for the kinematics calculation and constant k is defined uniquely through n). In total, six input parameters describe the mass and light distributions: θ_{E}, γ_{PL}, q_m, q_L, r_{eff}, and n.

These parameters together with β and i constitute the input vector $\mathbf{x}$. The ranges from which these parameters are drawn are physically motivated. Each realization is drawn uniformly from the range indicated: For $\theta_{\mathrm{E}} \in [0.5", 2"]$ and $\gamma_{\mathrm{PL}} \in [1.5, 2.5]$, we select ranges which cover the observed SLACS and TDCOSMO lens systems ((Gomer et al., 2022)); ranges for $r_{\mathrm{eff}} \in [0.5\theta_E, \theta_E]$ and $n \in [2, 4]$ and were chosen based on the priors for the Time Delay Lens Modeling Challenge ((TDLMC Ding et al., 2021)); for $q_m \in [0.6, 1.0]$ and $q_L \in [0.6, 1.0]$, we have selected a wider range than the TDLMC priors to better generalize to more elliptical galaxies; anisotropy $\beta \in [-0.4, 0.4]$ is motivated by the prior used by Yıldırım et al. ((2020)), but slightly widened as motivated by JAM models of SAURON observations of early-type galaxies ((Cappellari et al., 2007)); Inclination $i \in [\arccos(q_{\min}), \pi/2]$ has a lower bound set by the flattest Gaussian component of the MGE decomposition (with axis ratio $q_{\min}$), below which the deprojection becomes nonphysical.