Inferring subhalo effective density slopes from strong lensing observations with neural likelihood-ratio estimation

Gemma Zhang[1⋆], Siddharth Mishra-Sharma[1,2,3], and Cora Dvorkin[1]

[1]*Department of Physics, Harvard University, Cambridge, MA 02138, USA*
[2]*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[3]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

8 November 2022

ABSTRACT

Strong gravitational lensing has emerged as a promising approach for probing dark matter models on sub-galactic scales. Recent work has proposed the subhalo effective density slope as a more reliable observable than the commonly used subhalo mass function. The subhalo effective density slope is a measurement independent of assumptions about the underlying density profile and can be inferred for individual subhalos through traditional sampling methods. To go beyond individual subhalo measurements, we leverage recent advances in machine learning and introduce a neural likelihood-ratio estimator to infer an effective density slope for populations of subhalos. We demonstrate that our method is capable of harnessing the statistical power of multiple subhalos (within and across multiple images) to distinguish between characteristics of different subhalo populations. The computational efficiency warranted by the neural likelihood-ratio estimator over traditional sampling enables statistical studies of dark matter perturbers and is particularly useful as we expect an influx of strong lensing systems from upcoming surveys.

Key words: gravitational lensing: strong – dark matter

1 INTRODUCTION

Dark matter (DM) accounts for approximately 84% of the total matter budget of the Universe, yet its nature remains one of the biggest unanswered questions in fundamental physics (Planck Collaboration et al. 2020). The current standard ΛCDM cosmological model stipulates that dark matter is cold and collisionless and predicts a bottom-up hierarchical formation of structures. The ΛCDM model has been extensively tested by measurements at and above galactic scales, but evidence remains scarce on sub-galactic scales where many DM models differ in their predictions (Bode et al. 2001; Anderhalden et al. 2013; Bullock & Boylan-Kolchin 2017; Buckley & Peter 2018). Much of the challenge in testing the small-scale structure of dark matter lies in complications stemming from accurately modeling baryonic effects in galaxies. Low-mass subhalos ($\lesssim 10^9$ $M_\odot$) residing within a larger host dark matter halo provide a promising laboratory to test DM models because they are typically devoid of luminous matter and are thus less affected by baryonic processes (Fitts et al. 2017; Read et al. 2017; Kim et al. 2018).

Because of the lack of luminous matter in these subhalos, observations of their gravitational effects constitute one of the only ways of characterizing them. Within the Local Group, subhalos can be probed by analyzing stellar streams (Carlberg 2016; Bonaca et al. 2020; Banik et al. 2021), stellar wakes (Buschmann et al. 2018), and astrometric measurements (Feldmann & Spolyar 2015; Van Tilburg et al. 2018; Mishra-Sharma et al. 2020; Mondino et al. 2020). Beyond the immediate neighborhood of the Milky Way, strong gravitational lensing has been the most common method for constraining subhalo properties. Strong gravitational lensing occurs when a distant source, which can be an extended galaxy or a point-like quasar, is in close alignment with a foreground massive structure. In systems with quasar sources, flux ratio measurements are typically used to detect subhalos and constrain their properties (Mao & Schneider 1998; Metcalf & Madau 2001; Moustakas & Metcalf 2003; Fadely & Keeton 2012; MacLeod et al. 2013; Nierenberg et al. 2014, 2017; Gilman et al. 2020, 2021). In this work, we will focus on strong gravitational lensing systems where both the source and lens are galaxies. So far, $\mathcal{O}(100)$ such systems have been observed across existing datasets (Metcalf et al. 2019; Brownstein et al. 2012; Bolton et al. 2008; Jacobs et al. 2019; Storfer et al. 2022), and several systems contain claimed substructure detections (Vegetti et al. 2010, 2012; Hezaveh et al. 2016b). With forthcoming data from large-scale imaging surveys, the number of detected strong lensing systems is expected to increase by orders of magnitude (Laureijs et al. 2011; Collett 2015; McKean et al. 2015; Bechtol et al. 2019; Jacobs et al. 2019; Huang et al. 2021; Storfer et al. 2022), so it is timely

⋆ yzhang7@g.harvard.edu

to think about the prospects for using strong gravitational lensing for population-level studies of dark matter.

Previous analyses of substructure fall into two categories: direct detection in which individual subhalos are constrained through lens modeling (Vegetti et al. 2014; Ritondale et al. 2019), and statistical detection, which attempts to extract the collective effect of subhalos (Dalal & Kochanek 2002; Hezaveh et al. 2016a; Cyr-Racine et al. 2016; Díaz Rivero et al. 2018a,b; Birrer et al. 2017; Daylan et al. 2018; Brennan et al. 2019; Gilman et al. 2018; Cyr-Racine et al. 2019; Gilman et al. 2020; He et al. 2022). On the statistical detection front, machine learning has been an emerging approach for extracting population-level statistics of substructure (Brewer et al. 2016; Brehmer et al. 2019; Ostdiek et al. 2022b,a; Wagner-Carena et al. 2022). However, these works typically rely on constraining the subhalo mass function (SHMF), which depends on assumptions about the subhalo density profile in areas that are not directly probed. This can hinder our ability to reliably constrain the subhalo mass function. As a result, Minor et al. (2017) proposed a robust subhalo mass observable, called an effective subhalo lensing mass, and showed that this mass definition is independent of assumptions about the subhalo density profile. Moreover, relying on subhalo mass measurements overlooks information hidden in the subhalo central density profiles, which have been shown to be an important complementary probe of DM to the subhalo mass (Minor et al. 2021a,b; Amorisco et al. 2022). To address this, Şengül & Dvorkin (2022) proposed a new observable, which they call the "effective density slope," as a robust way of distinguishing between different classes of DM models. This effective density slope is defined as the slope of the density profile of a subhalo at scales where it has the largest observable effect. Şengül & Dvorkin (2022) showed that the effective slope of an individual subhalo can be accurately constrained using traditional sampling techniques. While characterizing individual subhalos is useful, it is often limited to leveraging the effect of more massive subhalos in a system. With the inclusion of more subhalos, the joint parameter space of individual subhalo properties increases significantly and sampling becomes prohibitive. This makes it particularly challenging to obtain enough measurements to make statistically robust inference. Moreover, characterizing individual subhalos does not account for the (expected) effect of a large population of unresolved lower-mass subhalos.

Modern machine learning developments have made it possible to overcome these challenges. Simulation-based inference using machine learning tools has been applied to an array of problems in cosmology (Perreault Levasseur et al. 2017; Alsing et al. 2018; Taylor et al. 2019; Alsing et al. 2019; Coogan et al. 2020; Legin et al. 2021; Wagner-Carena et al. 2021; Makinen et al. 2021; Dai & Seljak 2022; Cole et al. 2022). In this paper, we propose a machine learning approach that makes use of neural likelihood-ratio estimation (Cranmer et al. 2015; Baldi et al. 2016) in order to probe a population-level effective density slope. A crucial property of the method is that it can be used to efficiently implicitly marginalize over the large latent space in the lensing model without explicitly inferring parameters associated with individual subhalos. The advantage of this approach is two-fold: firstly, the number of subhalos is expected to increase as the mass decreases, so probing a population-level statistic takes into account the collective effect of low-mass subhalos; secondly, after investing in an initial training overhead, the machine learning model is computationally efficient at combining information across a large sample of lenses in a dataset.

This paper is organized as follows. In Sec. 2, we lay out the details of the lensing forward model used to generate the mock lensed images used for training as well as validation. In Sec. 3, we describe the neural likelihood-ratio estimation technique as well as model architecture used in this work. In Sec. 4, we demonstrate our model's ability to distinguish between subhalo populations with different density slopes and compare the model's predicted slopes with the analytically derived effective density slopes following Şengül & Dvorkin (2022). Finally, in Sec. 5, we conclude with a summary of our results and discuss potential improvements.

2 DATA GENERATION

To generate the strong lensing images for model training and testing, we use the software package LENSTRONOMY (Birrer et al. 2015; Birrer & Amara 2018). We generate (100×100) pixel^2 images with a resolution of $0.08''$ per pixel side. This corresponds to a field of view of $8'' \times 8''$ in an image. Generating an image requires several ingredients: a source galaxy, a main (host) lens galaxy, a population of subhalos, and a specification of the instrumental configuration. We discuss each of these components in detail below and the concrete values of parameters used for data generation are summarized in Table 1.

2.1 Source and main lens

The source and the main lens are the primary ingredients producing galaxy-galaxy lensing images. In each lensing system, light rays emitted by the source galaxy get gravitationally deflected by the main lens en route to the detector. In strong gravitational lensing, the bending of source light often results in characteristic lensed arcs in the observed images. In our datasets, we place the source at redshift $z_{\rm source} = 1.0$ and lens at redshift $z_{\rm lens} = 0.5$.

To simulate the source light in our images, we use source galaxy images provided by the PALTAS pipeline (Wagner-Carena et al. 2022). These sources are real galaxy images taken by the Hubble Space Telescope (HST) Cosmic Evolution Survey (COSMOS) (Scoville et al. 2007; Koekemoer et al. 2007). The PALTAS package takes a sub-sample of the HST COSMOS survey galaxies from the GREAT3 gravitational lensing challenge dataset (Mandelbaum et al. 2012, 2014). It then filters out 2,262 source candidates by putting constraints on parameters such as the pixel resolution and galaxy apparent magnitude and manually inspecting to keep only well-resolved images. Out of the 2,262 source galaxies available, we reserve 2,163 for modeling the training images, 70 for validation, and the remainder for testing and evaluation. This ensures that we can test the ability of our model to generalize to sources that it has not seen during training. To model the source light for a lensing image, we randomly draw a galaxy from the COSMOS catalog, apply a random rotation, and then draw the source position coordinates $x_{\rm source}, y_{\rm source}$ from a uniform interval of $[-0.1, 0.1]''$ with the center of the image being at $[0, 0]$.

We model the main lens using a singular isothermal ellipsoid (SIE) profile (Kormann et al. 1994), whose convergence is parameterized in LENSTRONOMY as

$$\kappa(x,y) = \frac{1}{2}\left(\frac{\theta_E}{\sqrt{qx_\phi^2 + y_\phi^2/q}}\right), \tag{1}$$

where θ_E is the Einstein radius, q is the minor/major axis ratio, and x_ϕ, y_ϕ are positions on the axes aligning with the major and minor axes of the lens. The angle of rotation ϕ is the angle between the major and minor axes and the fixed x, y axes of the image. The inputs into LENSTRONOMY are the ellipticity moduli instead of the major/minor axis ratio, and they are related as follows:

$$e_1 = \frac{1-q}{1+q}\cos(2\phi), \tag{2}$$

$$e_2 = \frac{1-q}{1+q}\sin(2\phi). \tag{3}$$

We also include an external shear parameterized by $\gamma_{\rm shear,1}$ and $\gamma_{\rm shear,2}$ as part of our lens model (Keeton et al. 1997). The shear parameters $\gamma_{\rm shear,1}$ and $\gamma_{\rm shear,2}$ are the diagonal and off-diagonal terms of the shear matrix, respectively. The external shear encodes the stretching of the source shape by nearby or line-of-sight massive structures. In modeling each image, we vary the center position, the Einstein radius, the shear parameters, and the ellipticity of the main lens as specified in Table 1. Our choice of the Einstein radius corresponds to an average main lens mass of approximately $1.6 \times 10^{13}\mathrm{M}_\odot$.

2.2 Subhalos

We refer to a lensed image resulting from a source and a main lens deflector as a "smooth model image." The presence of subhalos near the bright lensed arcs of a smooth model image can leave observable signatures in the lensed image. Thus, by learning to detect the effect of these perturbations in the lensed images, we can probe the properties of subhalos.

Only subhalos in the bright regions of a smooth model image have potentially observable effects because the lensing effects of subhalos elsewhere are degenerate with variations in the main lens mass. As such, in our datasets, we select pixels whose brightness is more than half of the maximum brightness in the smooth model image and then randomly place subhalos only in these areas. For our main results in Sec. 4, we draw subhalos with virial mass M_{200} between 10^7 $\mathrm{M}_\odot$ and 10^{10} $\mathrm{M}_\odot$ from a power-law subhalo mass function given by $dN_{\rm sub}/dM_{200} \propto M_{200}^{-1.9}$ where $N_{\rm sub}$ is the number of subhalos and M_{200} is the total mass of the subhalo within r_{200}, which is the radius within which the average mass density of the subhalo is 200 times the critical density of the Universe.

The number of subhalos we add to each image is uniformly distributed between 0 and 500. This range was chosen based on existing estimates of $f_{\rm sub}$, which denotes the fraction of substructure mass relative to the host galaxy mass. Concrete constraints on $f_{\rm sub}$ in existing literature are relatively scarce and can vary significantly between galaxies (Şengül et al. 2020). Using the SLACS galaxies, Vegetti et al. (2014) gives an upper bound on $f_{\rm sub}$ of approximately 0.007 for subhalos between $4 \times 10^6\mathrm{M}_\odot$ and $4 \times 10^9\mathrm{M}_\odot$ in a 32 kpc^2 region around the Einstein ring at $z_{\rm lens} = 0.2$; this translates to an approximate upper bound of 800 subhalos between $10^7\mathrm{M}_\odot$ and $10^{10}\mathrm{M}_\odot$ in the bright regions in our images at $z_{\rm lens} = 0.5$ for a main lens of $1.6\times10^{13}\mathrm{M}_\odot$. On the other hand, Ritondale et al. (2019) uses the BELLS GALLERY galaxies and gives an upper bound on $f_{\rm sub}$ of roughly 0.07 for subhalos between $10^5\mathrm{M}_\odot$ and $10^{11}\mathrm{M}_\odot$ in the whole image; this translates to an approximate upper bound of 250 subhalos between $10^7\mathrm{M}_\odot$ and $10^{10}\mathrm{M}_\odot$ in the bright regions in our images at $z_{\rm lens} = 0.5$ for a main lens of $1.6\times10^{13}\mathrm{M}_\odot$. For our datasets, we choose a rough average between these two estimates: an upper bound on $N_{\rm sub}$ of 500. In Sec. 4, for demonstrative and comparison purposes, we also show additional results for a few different mass ranges and numbers of subhalos, and we specify these ranges when applicable.

All of our mock images share the same aforementioned parameters. However, for practical purposes, we model the subhalo density profiles in the training/validation set and the test set differently and discuss them in detail below.

2.2.1 Training and validation sets

To produce the training and validation sets, we add subhalos that follow the elliptical power law (EPL) profile (Barkana 1998), whose convergence at position (x, y) on the lens plane is defined as follows:

$$\kappa(x,y) = \frac{3-\gamma}{2}\left(\frac{\theta_E}{\sqrt{qx_\phi^2 + y_\phi^2/q}}\right)^{\gamma-1}, \tag{4}$$

where $\theta_E, q, x_\phi, y_\phi$ are defined similarly to those in the SIE profile, and γ is an additional free parameter that controls the slope of the density profile of a subhalo. For $q = 1$, the average convergence $\bar{\kappa}$ within a given radius r from the center of an EPL subhalo is

$$\bar{\kappa} = \frac{2\pi\int_0^r r'\kappa(r')dr'}{\pi r^2} \propto r^{1-\gamma}, \tag{5}$$

so that its power-law slope is given as a function of the average convergence as

$$\gamma = 1 - \frac{d\ln\bar{\kappa}}{d\ln r}. \tag{6}$$

This particular profile choice allows us to conveniently vary the slope of the density profile and label each training image with its true underlying slope. Although EPL is not a sufficiently realistic profile for subhalos, this choice is deliberate because our model requires the true parameter value as an input during training, as will be discussed in Sec. 3.2.

The 3D density of EPL subhalos scales with the radial distance as $\propto r^{-\gamma}$. Using Eq. (4) to solve for the normalization of the 3D density, we see that, assuming $q = 1$ for no ellipticity, the radial density profile of an EPL subhalo is

$$\rho(r) = \rho_{\rm EPL} r^{-\gamma}, \tag{7}$$

where $\rho_{\rm EPL} = \frac{2\sqrt{\pi}}{3-\gamma}\frac{\Gamma(\frac{\gamma-1}{2})}{\Gamma(\frac{\gamma}{2})}$ is a constant with Γ being the Gamma function. Note that when $\gamma = 1$, the mass increases radially and diverges eventually, and when $\gamma = 3$, the density normalization $\rho_{\rm EPL}$ diverges. As a result, we constrain γ to values between 1 and 3 in order to ensure that the subhalo masses are physical. To model the subhalos in each image, we first draw a slope from the uniform distribution $\gamma \sim \mathcal{U}(1.1, 2.9)$ and then draw normally distributed slopes $\gamma_i \sim \mathcal{N}(\gamma, 0.1\gamma)$ for the ith subhalo. We truncate the

Parameter	Distribution
Source	
Source redshift	$z_{\rm source} = 1.0$
x-coordinate	$x_{\rm source} \sim \mathcal{U}(-0.1'', 0.1'')$
y-coordinate	$y_{\rm source} \sim \mathcal{U}(-0.1'', 0.1'')$
Main lens	
Lens redshift	$z_{\rm lens} = 0.5$
x-coordinate	$x_{\rm lens} \sim \mathcal{U}(-0.2'', 0.2'')$
y-coordinate	$y_{\rm lens} \sim \mathcal{U}(-0.2'', 0.2'')$
Einstein radius	$\theta_E \sim \mathcal{U}(2.7'', 3'')$
Ellipticities	$e_1 \sim \mathcal{U}(-0.2, 0.2)$ $e_2 \sim \mathcal{U}(-0.2, 0.2)$
External shear	$\gamma_{\rm shear,1} \sim \mathcal{U}(-0.1, 0.1)$ $\gamma_{\rm shear,2} \sim \mathcal{U}(-0.1, 0.1)$
Subhalos	
EPL ellipticities	$e_1 \sim \mathcal{U}(-0.2, 0.2)$ $e_2 \sim \mathcal{U}(-0.2, 0.2)$
EPL slope of density profile per lens system	$\gamma \sim \mathcal{U}(1.1, 2.9)$
EPL slope of density profile per subhalo	$\gamma_i \sim \mathcal{N}(\gamma, 0.1\gamma)$
Subhalo mass function power-law slope	-1.9
Subhalo mass	$M_{200} \in [10^7, 10^{10}]{\rm M}_\odot$; $[10^8, 10^{10}]{\rm M}_\odot$; $10^9{\rm M}_\odot$
Number of subhalos per image	$N_{\rm sub} \sim \mathcal{U}\{0, 500\}$; $\mathcal{U}\{0, 60\}$; $\mathcal{U}\{0, 100\}$

Table 1. Parameters of the main components of a strong gravitational lensing system and their respective training distribution in our mock images. The test set follows the same parameter distributions except for when the subhalos have NFW instead of EPL profiles, as discussed in Sec. 2.2.2.

$\mathcal{N}(\gamma, 0.1\gamma)$ distribution so that γ_i is constrained to $(1, 3)$. Drawing the slope of each subhalo from a normal distribution more realistically simulates the natural variations in slope measurements that are observed in simulations (Şengül & Dvorkin 2022).

2.2.2 Test set

In the test set, we model subhalos with profiles that are more realistic than an EPL profile. Here we use the Navarro–Frenk–White (NFW) profile (Navarro et al. 1997), whose radial density profile is defined as

$$\rho(r) = \frac{\rho_0}{\frac{r}{r_s}\left(1 + \frac{r}{r_s}\right)^2}, \tag{8}$$

where r is the radial distance from the center of the subhalo, and the normalization ρ_0 and the scale radius r_s are parameters that vary between subhalos. An NFW profile can alternatively be parameterized by M_{200} and concentration c_{200}. The concentration is defined as the ratio between r_{200} and the scale radius: $r_{200} = c_{200} r_s$. To produce our main test set, we draw masses from the subhalo mass function and set its c_{200} using a fixed mass-concentration relation extrapolated from Dutton & Macciò (2014), which we denote as $f_{\rm CDM}(M_{200})$. To simulate subhalo populations with different slopes, we vary the subhalo concentrations in the test set using an overall multiplicative factor a, such that the modified mass-concentration becomes $a \times f_{\rm CDM}(M_{200})$. Subhalo populations with $a < 1$ have concentrations lower than those given by $f_{\rm CDM}(M_{200})$ and have more cored inner density profiles while subhalo populations with $a > 1$ have more cuspy inner density profiles.

Adding subhalos in the lens plane increases the total mass that lenses the source light, thereby enlarging the Einstein radii in the images produced. In some of our experiments in Sec. 4.2, we add a large total subhalo mass in each lensing system, which risks producing an effective Einstein radius larger than the image size. To mitigate this, we add a sheet of constant negative convergence to our modeling so that the Einstein radius of an image including subhalos gets restored to its value in the smooth model image. To add the sheet of negative convergence to an image, we compute the total convergence of all the subhalos and add a constant negative convergence across all pixels in the image so that they cancel out the total subhalo convergence.

2.3 Instrumental details

To create the lensed images, we use the WFC3 F160W camera band configuration and a Gaussian point spread function (PSF) with a full width at half maximum of $0.08''$. We use the LENSTRONOMY HST configurations with a sky brightness of 22.3 and magnitude zero point of 25.96. In each image, we add 10 orbits of HST noise, with an exposure time of 5,400 seconds per orbit.

3 MODEL AND INFERENCE

For our task, we train a neural likelihood-ratio estimator using parameterized classifiers (Cranmer et al. 2015; Baldi et al. 2016; Hermans et al. 2019), a technique that has been previously proposed in the context of astrophysical dark matter searches for characterizing the subhalo mass function using galaxy-galaxy strong lenses (Brehmer et al. 2019; Anau Montel et al. 2022), astrometric datasets (Mishra-Sharma 2022), stellar streams (Hermans et al. 2021), and halo clustering statistics (Dimitriou et al. 2022). In this section, we will give

a brief overview of the technique and summarize the salient aspects of our machine learning model.

3.1 Inference method

Suppose θ denotes the parameters of interest. Then, we can generate samples using a simulator (forward model) conditioned on θ that implicitly define the likelihood, $\{x\} \sim p(x \mid \theta)$. We can learn an estimator for the likelihood ratio between this distribution and one corresponding to an alternate reference hypothesis, $r(x \mid \theta) = \frac{p(x|\theta)}{p_{\rm ref}(x|\theta)}$, by training a classifier parameterized through θ to distinguish between samples drawn according to the two hypotheses. An ideal classifier would learn the decision function $s(x,\theta) = \frac{p(x|\theta)}{p(x|\theta)+p_{\rm ref}(x|\theta)}$ which is one-to-one with the aforementioned likelihood ratio (Cranmer et al. 2015; Baldi et al. 2016; Mohamed & Lakshminarayanan 2016):

$$r(x \mid \theta) = \frac{s(x,\theta)}{1-s(x,\theta)}. \tag{9}$$

In neural likelihood-ratio estimation, the classifier $\hat{s}(x,\theta)$ is a neural network appropriate for the data modality in the given task and is trained through supervised learning. When the classifier decision is determined by applying a sigmoidal projection, the log-likelihood ratio estimates $\log \hat{r}(x \mid \theta)$ can be conveniently read off through the pre-sigmoidal outputs (logits) directly.

In our specific inference task, θ is a single parameter that characterizes the power law slope γ in the EPL profile and x is a lensed image generated with an underlying γ. Having generated an ensemble of lensed image-parameter pairs, we train a classifier with a dependence on θ to distinguish between samples from the joint data-parameter distribution $p(x,\theta)$ and those from the reference distribution, $p_{\rm ref}(x,\theta)$, taken to be the product of the marginal distributions of the data and slope parameter, $\{x,\theta\}_{\rm ref} \sim p(x)p(\theta)$ (Hermans et al. 2019). Samples following the reference hypothesis are generated by shuffling the γ labels in each batch so that they no longer align with the lensed images they generated. Intuitively, the classifier implicitly learns whether a γ value is likely to be the true underlying slope that generated a given image by learning to distinguish between data-parameter pairs from the two distributions: the higher the classification score a pair receives, the more likely it is to have been generated according to $p(x,\theta)$.

To compute the likelihood-ratio profile for a given lensed image as a function of our parameter of interest at test time, we obtain the classifier logits for a linearly-spaced array of input γ. The computational cost during test time therefore scales linearly with the number of test points in the parameter space. For our single parameter of interest γ, the inference is relatively efficient through direct computation; for a high-dimensional parameter, sampling techniques may be necessary (Hermans et al. 2019).

An aspect of likelihood-ratio estimation relevant to our application is that it allows us to conveniently combine information across multiple images with the same underlying γ and obtain stronger constraints (Brehmer et al. 2019), as desired when working with an ensemble of strong lensing observations. If we assume that the individual inputs are independently and identically distributed when conditioned on the parameter of interest γ, then the combined likelihood ratio for a set of images $\{x\}$ can be obtained through the product of the individual likelihood ratios,

$$\hat{r}(\{x\} \mid \gamma) = \prod_i \hat{r}(x_i \mid \gamma). \tag{10}$$

In contrast, when using simulation-based inference methods that directly learn a parameter posterior density on individual images, combining this information across images can be more challenging as it typically involves approximating integrals through sampling, which may introduce numerical inaccuracies that propagate when scaling to a large ensemble of lenses.

3.2 Model and training details

We use a modified version of the ResNet-18 convolutional neural network implemented in `PyTorch` (Paszke et al. 2019) as our classifier (He et al. 2016). We output the classification score $\hat{s}(x,\theta)$ by projecting through a sigmoid activation after the dense layers of the ResNet. At training time, we append the true γ to the latent vector after the convolutional layers, which is then passed into the dense layers. This ensures that the classifier output depends on both the image and the γ value. Our training objective for the classification task is the canonical binary cross-entropy loss.

The pixel values are normalized to zero mean and unit standard deviation across the training set. We also normalize the γ values to zero mean. At test time, we use the training mean and standard deviation to whiten our test data so that the inputs into the model are consistent from training to testing.

We performed a coarse grid hyperparameter search and chose an initial learning rate of 10^{-3} with the AdamW optimizer (Kingma & Ba 2014; Loshchilov & Hutter 2017) and batch size of 2000. We found that a larger batch size significantly increases the stability of training and produces better downstream results, consistent with previous findings (Hermans et al. 2021; Hoffer et al. 2017). During training, we follow a learning rate schedule that decays by an order of magnitude when the validation loss stagnates for 3 epochs, followed by a cool-down period of 2 epochs. We terminate our training when the validation loss plateaus for 6 epochs under a threshold of 10^{-3}. Our training and validation sets consist of 50,000 and 1,000 mock images respectively, generated using the forward model described in Sec. 2. The model was trained on NVIDIA V100 GPUs with typical training time of 8 to 10 hours for 25 to 30 epochs.

4 RESULTS

To infer the slope of subhalos in a given image x_i, we feed the image along with an array of test γ into the trained model to obtain the likelihood-ratio profile estimate $\hat{r}(x_i \mid \gamma)$. In Fig. 1, we show in the left panel the individual predictions compared to the true γ for 200 test images containing EPL subhalos. We show the maximum-likelihood estimates $\hat{r}_{\rm MLE}(x_i \mid \gamma)$ using the scatter points and the 1σ uncertainties using error bars, assuming a χ^2 distribution on the test-statistic $2\left[\log \hat{r}_{\rm MLE} - \log \hat{r}\right]$ (Wilks 1938). Predictions can be seen to follow the true γ values in trend, with reasonably large inferred uncertainties.

To leverage the statistical effect of multiple observations,

we can combine likelihood ratios from several images that share the same underlying γ, following Eq. (10). We show these results in the right panel of Fig. 1. Each prediction is obtained by combining the likelihood ratios (i.e., summing the log-likelihood ratios) of 50 images that share the same latent slope. As expected, combining information from a set of images significantly improves constraining power on the slope. Note that the predicted slope at the high-slope end has a small visible bias, likely due to the fact that we restrict the values of γ to $[1.1, 2.9]$ so that there are relatively few images corresponding to $\gamma \gtrsim 2.9$ in the training set. However, we expect realistic images to have subhalos whose slopes lie reasonably far from the edges of the $(1, 3)$ range, so this bias has a negligible effect on our overall conclusions.

4.1 Region of maximum observability

Our training set contains EPL subhalos because the EPL density profile allows us to conveniently label each image with its underlying slope. During test time, however, the EPL density profile is no longer sufficient for simulating realistic subhalos. With more realistic subhalo density profiles, the power-law slope is no longer constant at different radial distances from the center of a subhalo. For instance, the NFW profile scales as r^{-1} near its core but asymptotically transitions to r^{-3} past its scale radius. This necessitates a more concrete definition of "slope."

Şengül & Dvorkin (2022) developed a formalism for estimating a region of maximum observability (RMO) and defined an effective slope for each perturber in a lensing system. We will sketch out the core steps in the RMO calculation and refer readers to the reference for a more detailed discussion. The central step in the RMO calculation is approximating the deviations in the brightness of each image pixel caused by a subhalo. This is done by performing a perturbative expansion in the deflection angles, assuming the deflections due to the subhalo are sufficiently small. This allows us to obtain a linear mapping between the pixel brightness deviations and the subhalo's average convergence. The RMO is then defined as the region in radial distance from the subhalo center in which the subhalo's average convergence measurement has the smallest uncertainty. The RMO depends on various parameters of a lensing system, particularly the gradient of the source light and the noise in each pixel, so it needs to be calculated for each system individually. When we constrain an effective power-law slope of a subhalo, the slope we measure is approximately the mean slope of the average convergence in the RMO following Eq. (6). For the perturbative expansion in the RMO calculation to be valid, an important assumption is that the maximum angular deflection caused by the subhalo is much less than the pixel size of the image. In our case, the maximum angular deflection caused by an NFW subhalo with $M_{200} = 10^{10}$ $\mathrm{M}_\odot$ (which translates to $c_{200} \approx 10$) at $z_{\mathrm{lens}} = 0.5$ is around $0.009''$, which is much smaller than our pixel size of $0.08''$ and ensures the validity of our RMO approximation.

Following Şengül & Dvorkin (2022), we define our RMO to be the region where the relative error on the average convergence is less than twice the smallest relative error. As an illustrative example, we take one of our test images and compute the RMO of two NFW subhalos of different masses. The two subhalos come from the same source-lens system but are located at different positions in the lens plane. In Fig. 2, we show their average convergence $\bar{\kappa}$ as a function of radial distance r in the lens plane (in blue curves) and their RMOs (in shaded gray). The upper panel shows the RMO of a subhalo with $M_{200} \approx 1.5 \times 10^7 \mathrm{M}_\odot$, while the lower panel shows the RMO of a subhalo with $M_{200} \approx 1.6 \times 10^9 \mathrm{M}_\odot$. The lower limit on the radial distance is set by the resolution of the image. As expected, the lower-mass subhalo has much larger errors on its average convergence profile than the higher-mass subhalo because low-mass subhalos have a low signal-to-noise ratio. As a consequence, the power-law slope of a low-mass subhalo is much less well-constrained than its high-mass counterpart. In both cases, when we take into account the uncertainties, the NFW subhalos are roughly indistinguishable from EPL subhalos. We compute the average power-law slope in the RMO for the two subhalos using Eq. (6). The low-mass subhalo has an effective slope of approximately 2.55, while the high-mass subhalo has an effective slope of approximately 2.13. This agrees with the expectation that more massive subhalos have profiles that are farther extended radially and have flatter profiles than lighter subhalos when compared at the same radial distance.

The RMO approximation allows us to obtain an estimate of the power-law slope for a single subhalo at a time. Because the masses of subhalos in each of our test images span three orders of magnitude, we obtain a different effective slope for each subhalo. In contrast, when we carry out inference with our likelihood-ratio estimator, we get a unique slope prediction for each image. Attaching a sensible physical interpretation to our model output is challenging because the neural network implicitly marginalizes over the information of all the subhalos in an image. Moreover, when we combine a set of likelihoods from multiple images, we further marginalize over source and main lens parameters, both of which affect the RMO approximation. It is nonetheless useful to compare the set of effective slopes computed analytically from the RMO with our model prediction to gain some physical understanding of the neural network output.

To carry out this comparison, we focus on a single image and compare the RMO effective slopes of its subhalos with our model's prediction. In Fig. 3, we show the set of analytically approximated RMO effective slopes for 388 NFW subhalos in an image, indicated by histograms corresponding to three mass ranges. These NFW subhalos all have concentrations given by $f_{\mathrm{CDM}}(M_{200})$, based on their individual masses. For visualization purposes, we then take the estimated likelihood ratio $\hat{r}$ of the same image using the trained neural likelihood-ratio estimator and scale it up so that it fits roughly on the same scale as the histograms. We show the likelihood-ratio estimate from our neural network in black dashed line. The RMO effective slopes at the low and high mass ranges in Fig. 3 both roughly agree with the examples shown in Fig. 2. We see that $\hat{r}$ peaks at roughly the middle of the span of the histograms. It can be coarsely interpreted as a weighted average of the histogram where the more massive subhalos are given more weight than lighter subhalos. Intuitively, the more massive subhalos get more weight because subhalos in the $[10^7, 10^8]\mathrm{M}_\odot$ mass range have large errors in their RMOs so that their effective slope estimates also have large uncertainties. It may be tempting to conclude that the $\hat{r}$ likelihood is high-mass dominated and receives very little contribution from the low-mass subhalos. We will show later

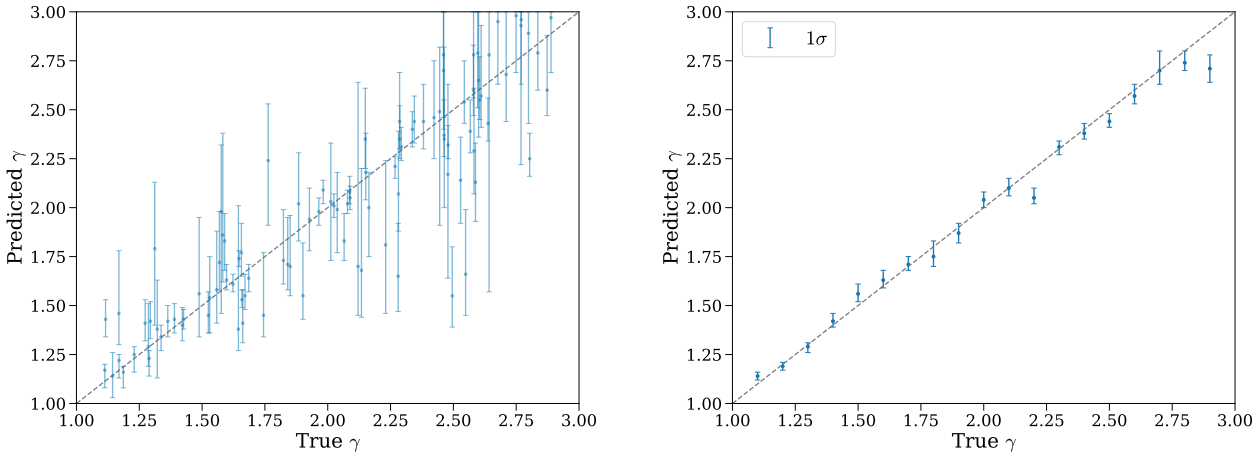

Figure 1. (*Left*) Scatter plot of the maximum-likelihood γ and associated 1-σ uncertainties predicted using the trained likelihood-ratio estimator compared with the true underlying γ of 100 test images with EPL subhalos. (*Right*) Maximum-likelihood γ and associated 1-σ uncertainties predicted from combined likelihoods of sets of 10 images containing EPL subhalos compared with the true γ underlying each set of images. For both panels, the model was trained on images containing EPL subhalos with $M_{200} \in [10^7, 10^{10}]\mathrm{M}_\odot$ and $N_{\rm sub} \sim \mathcal{U}\{0, 500\}$.

in Sec. 4.3 that this is in fact not the case. We will compare the $\hat{r}$ predictions from two different mass ranges and show that the model has significant sensitivity to the $[10^7, 10^8]\mathrm{M}_\odot$ subhalos.

4.2 Test on NFW subhalos with varying concentrations

As discussed in Sec. 4.1, realistic subhalo profiles are close to being indistinguishable from EPL subhalos. Therefore, even though our likelihood-ratio estimator is trained on images containing EPL subhalos, we expect it to produce sensible predictions on more realistic subhalos at test time.

To simulate subhalo populations with different slopes, we produce test sets with NFW subhalos while varying their mass-concentration relation, as discussed in Sec. 2.2.2. We expect that the predicted slopes of low-concentration multiplicative factor a should be smaller than those of high multiplicative factor a because the subhalo inner density profiles are flatter with lower concentrations. Because subhalos of different masses have different concentrations as well as different effective slopes, as a simplified toy experiment, we first work solely with $10^9\mathrm{M}_\odot$ subhalos. We train a model on mock images containing $N_{\rm sub} \sim \mathcal{U}\{0, 100\}$ subhalos of $M_{200} = 10^9\mathrm{M}_\odot$, and subsequently evaluate the model on test sets each with a different c_{200}. Note that the mass-concentration relation $f_{\rm CDM}(M_{200})$ gives $c_{200} = 12.2$ for a $M_{200} = 10^9\mathrm{M}_\odot$ subhalo. With these parameter choices, the total subhalo mass we add to the lensing systems is relatively large, so we add a negative mass sheet during modeling, as discussed in Sec. 2.2.2.

In Fig. 4, we show the model predictions by combining the likelihoods of 50 images in each test set. As expected, the model-predicted slope increases with the subhalo concentration. Remarkably, the model is able to effectively distinguish between different subhalo concentrations. Moreover, we see that the model predicts a slope of $\gamma \approx 2.1$ at approximately $c_{200} = 12.2$, which is consistent with the RMO slope estimate of the $1.6 \times 10^9\mathrm{M}_\odot$ subhalo in Fig. 2.

4.3 Effect of low-mass subhalos

One of the main reasons for harnessing population-level statistics of subhalos is that it allows us to extract information from the low-mass subhalos that are difficult, if not impossible, to be individually observed. It is therefore important to quantify the sensitivity of our model to low-mass subhalos. We can do so by comparing model predictions on two subhalo mass ranges: $[10^7, 10^{10}]\mathrm{M}_\odot$ and $[10^8, 10^{10}]\mathrm{M}_\odot$. The difference between the predictions on NFW subhalo populations in the two mass ranges will indicate the contributions from the low-mass subhalos.

To carry out this comparison, we generate new training and test sets with subhalo masses in the $[10^8, 10^{10}]\mathrm{M}_\odot$ range. These images are not meant to be realistic and, instead, serve the purpose of disentangling the effects of high- and low-mass subhalos on our model's slope predictions. In our $[10^7, 10^{10}]\mathrm{M}_\odot$ training set, we take $N_{\rm sub} \sim \mathcal{U}\{0, 500\}$, which translates to approximately $N_{\rm sub} \sim \mathcal{U}\{0, 60\}$ subhalos in the $[10^8, 10^{10}]\mathrm{M}_\odot$ range under the same assumed subhalo mass function. We keep all other lensing parameters the same between the two datasets. In Fig. 5, we show the maximum-likelihood model predictions of combining 50 images and their associated 1σ uncertainties on the low and high mass range datasets in blue and magenta, respectively. Because our datasets include subhalos of varying masses and thereby different c_{200} values, we show the concentration multiplicative factor a instead of c_{200} in the x-axis. In the figure, we see a clear separation between the predictions on the two sets of images, indicating a significant contribution of the low-mass subhalos to the slope estimates at all concentration scales. This demonstrates that our model is sensitive to low-mass subhalos and incorporates their information when making predictions. Moreover, we notice that subhalos in

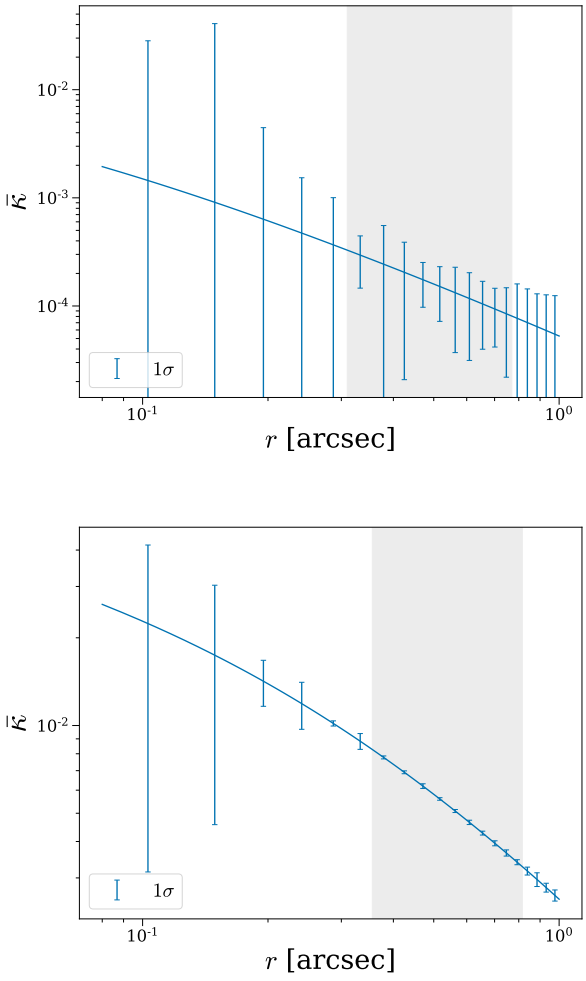

Figure 2. An example of the average convergence $\bar{\kappa}(r)$ within a radius r from the subhalo center, for NFW subhalos with masses $M_{200} \approx 1.5 \times 10^7 \mathrm{M}_\odot$ (*top*) and $M_{200} \approx 1.6 \times 10^9 \mathrm{M}_\odot$ (*bottom*). The error bars show the 1σ errors on $\bar{\kappa}$ and the shaded gray area shows the estimated RMO where the relative errors on $\bar{\kappa}$ are less than twice the smallest relative error.

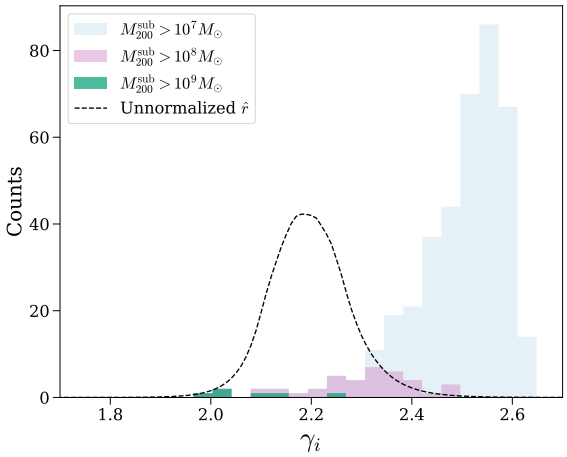

Figure 3. Histograms of the RMO effective power-law slope estimates of 388 NFW subhalos in a test set image separated by mass ranges of $[10^7, 10^{10}]\mathrm{M}_\odot$ (blue), $[10^8, 10^{10}]\mathrm{M}_\odot$ (magenta), and $[10^9, 10^{10}]\mathrm{M}_\odot$ (green). The likelihood-ratio estimate ($\hat{r}$) of the same image using the trained neural network model is scaled up for clarity and shown in black dashed line.

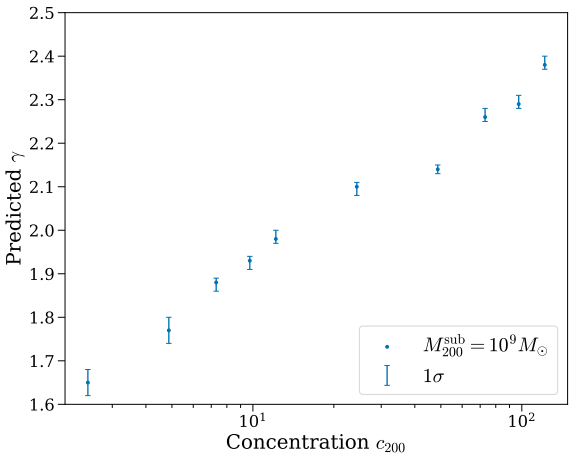

Figure 4. The maximum-likelihood predictions (scatter points) and 1σ uncertainties (error bars) obtained by combining 50 images with $N_{\mathrm{sub}} \sim \mathcal{U}\{0, 100\}$ of $10^9 \mathrm{M}_\odot$ NFW subhalos as a function of the concentration c_{200}. The model used for inference was trained on images with $N_{\mathrm{sub}} \sim \mathcal{U}\{0, 100\}$ and $10^9 \mathrm{M}_\odot$ EPL subhalos.

the $[10^8, 10^{10}]\mathrm{M}_\odot$ mass range produce lower predicted slopes than their lower mass counterparts at all concentration factors. This is consistent with the lower slope predictions for more massive subhalos through the RMO formalism shown in Fig. 2 and Fig. 3. As a robustness check, we also took our model trained on $[10^7, 10^{10}]\mathrm{M}_\odot$ population and tested it on the $[10^8, 10^{10}]\mathrm{M}_\odot$ population, and we observed a similar separation in the predicted slopes, as is shown in Fig. 5. This means that the model is sensitive to the low-mass subhalo contribution to the effective slopes regardless of the subhalo population during training. The $[10^8, 10^{10}]\mathrm{M}_\odot$ subhalo population can be thought of as having an extreme suppression in the low-mass region of the SHMF compared to the $[10^7, 10^{10}]\mathrm{M}_\odot$ population. Because our model is sensitive to this suppression, we expect it to be able to detect changes in the SHMF. Together with the results shown in Sec. 4.2, we see that two factors affect the model's effective density slope prediction: collective changes in the subhalo density profiles and deviations in the SHMF. CDM and alternative DM models differ in their predictions of these two factors, so our model's ability to predict effective density slopes of subhalo populations has the potential to help discern different classes of DM models.

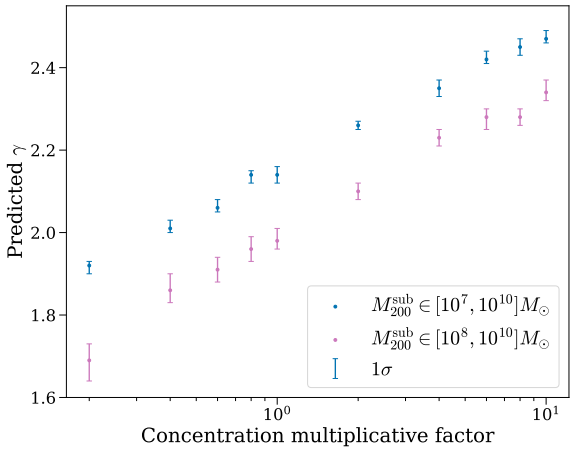

Figure 5. The maximum-likelihood predictions (scatter points) and 1σ uncertainties (error bars) obtained by combining 50 images with $N_{\rm sub} \sim \mathcal{U}\{0, 500\}$ and $M_{200} \in [10^7, 10^{10}]\mathrm{M}_\odot$ NFW subhalos as a function of the concentration c_{200} (blue), compared to those with $N_{\rm sub} \sim \mathcal{U}\{0, 60\}$ and $M_{200} \in [10^8, 10^{10}]\mathrm{M}_\odot$ NFW subhalos (magenta).

5 CONCLUSIONS AND OUTLOOK

Strong gravitational lensing provides a promising avenue for testing different classes of microphysical dark matter models on sub-galactic scales. Typically, previous works have focused on using individual subhalos or subhalo populations to constrain the subhalo mass function, a quantity that depends on the precise subhalo mass definition. An effective density slope of subhalos has been recently proposed as a reliable observable for probing dark matter properties. The advantage of measuring the effective subhalo density slope is that it makes no assumptions about its density profile, unlike the commonly used subhalo mass function. Şengül & Dvorkin (2022) demonstrated that measuring the effective density slope can be achieved on the most massive subhalo in an individual lensing system. However, two challenges lie ahead of going beyond the most massive subhalo in a single lensing system: first, adding more subhalos increases the parameter space dimension significantly and makes sampling inefficient; second, modeling each lensing system may take $\mathcal{O}$(days) at a time, so it quickly becomes computationally unfeasible to analyze a large number of lensing systems together.

Recent advances in machine learning provide promising tools for overcoming these challenges. Machine learning methods are capable of performing inference by implicitly marginalizing over a large parameter space without the need to characterize the effect of individual subhalos. In this work, we have introduced a neural likelihood-ratio estimator to extract a population-level effective density slope. This method allows us to make use of additional information contained in the lensing forward model, including the effect of low-mass subhalos. In addition, when compared to other neural simulation-based inference techniques such as those based on posterior density estimation, neural likelihood-ratio estimation can be used to more efficiently combine information across a large dataset.

Through testing our model on mock strong lensing images, we showed that the inferred effective density slope incorporates significant information from the low-mass subhalos and further demonstrated that our model is capable of distinguishing between subhalo populations with different density slopes. Different DM models predict effective density slopes of subhalo populations that deviate from the CDM prediction. For example, warm dark matter models predict a suppression in the number of low-mass subhalos and more cored subhalo density profiles (Bode et al. 2001; Lovell et al. 2012, 2014), and both of these effects will result in a smaller subhalo effective density slope prediction compared to that of the CDM model. Self-interacting dark matter models also predict more cored subhalo density profiles and thus will also produce effective density slope predictions smaller than that of the CDM model (Spergel & Steinhardt 2000; Vogelsberger et al. 2012; Rocha et al. 2013; Kahlhoefer et al. 2019). In addition, certain self-interacting dark matter models predict core-collapsed subhalos that have particularly steep density slopes (e.g. Lynden-Bell & Wood 1968; Kochanek & White 2000; Colín et al. 2002). This means that our likelihood-ratio estimator's ability to distinguish between subhalo populations with different effective density slopes has significant implications for discerning DM scenarios.

Several improvements to our analysis pipeline are necessary before being able to deploy it on real lensing data. First, our image modeling procedure does not include line-of-sight halos that also affect the lensing signature. Second, because of the need for well-resolved source galaxies, the COSMOS galaxies used by PALTAS have original redshifts lower than the $z_{\rm source} = 1$ used in modeling our lensing systems. There are differences in the morphology of galaxies at different redshifts, so to simulate more realistic images, galaxies with better resolutions and at higher redshifts could be used in the future. Third, in our mock images, we assume that the lens light has been subtracted, but to more realistically model the lensed systems we could include the lens light using real galaxy images in the forward model as done for the source galaxy. Finally, we demonstrated our model's sensitivity to the subhalo effective density slopes by testing on NFW subhalos with varying concentrations; in a realistic application, it would be necessary to understand the expected effect of different microphysical DM models as well. Despite these additional complexities, we are hopeful that the neural likelihood-ratio estimator will be a useful tool for characterizing the inner slopes of dark matter subhalos using the influx of strong lensing systems that will be discovered by upcoming imaging surveys.

ACKNOWLEDGEMENTS

We would like to thank Douglas Finkbeiner, Çağan Şengül and Arthur Tsang for valuable discussions. CD and GZ are partially supported by NSF grant AST-1813694. SM is partially supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics of U.S. Department of Energy under grant Contract Number DE-SC0012567. This work is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, http://iaifi.org/). The computations in this paper were

run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

DATA AVAILABILITY

The code used to produce the results shown in this paper is publicly available at `https://github.com/gemyxzhang/neural-subhalo-slope`.

REFERENCES

Alsing J., Wandelt B., Feeney S., 2018, MNRAS, 477, 2874–2885
Alsing J., Charnock T., Feeney S., Wandelt B., 2019, MNRAS, 488, 4440
Amorisco N. C., et al., 2022, MNRAS, 510, 2464
Anau Montel N., Coogan A., Correa C., Karchev K., Weniger C., 2022, arXiv e-prints, p. arXiv:2205.09126
Anderhalden D., Schneider A., Macciò A. V., Diemand J., Bertone G., 2013, J. Cosmology Astropart. Phys., 2013, 014
Baldi P., Cranmer K., Faucett T., Sadowski P., Whiteson D., 2016, European Physical Journal C, 76, 235
Banik N., Bovy J., Bertone G., Erkal D., de Boer T. J. L., 2021, J. Cosmology Astropart. Phys., 2021, 043
Barkana R., 1998, ApJ, 502, 531
Bechtol K., et al., 2019, BAAS, 51, 207
Birrer S., Amara A., 2018, Physics of the Dark Universe, 22, 189
Birrer S., Amara A., Refregier A., 2015, ApJ, 813, 102
Birrer S., Amara A., Refregier A., 2017, JCAP, 05, 037
Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93
Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Gavazzi R., Moustakas L. A., Wayth R., Schlegel D. J., 2008, ApJ, 682, 964
Bonaca A., et al., 2020, ApJ, 892, L37
Brehmer J., Mishra-Sharma S., Hermans J., Louppe G., Cranmer K., 2019, ApJ, 886, 49
Brennan S., Benson A. J., Cyr-Racine F.-Y., Keeton C. R., Moustakas L. A., Pullen A. R., 2019, Mon. Not. Roy. Astron. Soc., 488, 5085
Brewer B. J., Huijser D., Lewis G. F., 2016, MNRAS, 455, 1819
Brownstein J. R., et al., 2012, ApJ, 744, 41
Buckley M. R., Peter A. H. G., 2018, Phys. Rep., 761, 1
Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343
Buschmann M., Kopp J., Safdi B. R., Wu C.-L., 2018, Phys. Rev. Lett., 120, 211101
Carlberg R. G., 2016, ApJ, 820, 45
Cole A., Miller B. K., Witte S. J., Cai M. X., Grootes M. W., Nattino F., Weniger C., 2022, J. Cosmology Astropart. Phys., 2022, 004
Colín P., Avila-Reese V., Valenzuela O., Firmani C., 2002, ApJ, 581, 777
Collett T. E., 2015, Astrophys. J., 811, 20
Coogan A., Karchev K., Weniger C., 2020, in 34th Conference on Neural Information Processing Systems. (`arXiv:2010.07032`)
Cranmer K., Pavez J., Louppe G., 2015, arXiv e-prints, p. arXiv:1506.02169
Cyr-Racine F.-Y., Moustakas L. A., Keeton C. R., Sigurdson K., Gilman D. A., 2016, Phys. Rev. D, 94, 043505
Cyr-Racine F.-Y., Keeton C. R., Moustakas L. A., 2019, Phys. Rev. D, 100, 023013
Dai B., Seljak U., 2022, MNRAS, 516, 2363
Dalal N., Kochanek C. S., 2002, Astrophys. J., 572, 25
Daylan T., Cyr-Racine F.-Y., Diaz Rivero A., Dvorkin C., Finkbeiner D. P., 2018, ApJ, 854, 141
Díaz Rivero A., Cyr-Racine F.-Y., Dvorkin C., 2018a, Phys. Rev. D, 97, 023001
Díaz Rivero A., Dvorkin C., Cyr-Racine F.-Y., Zavala J., Vogelsberger M., 2018b, Phys. Rev. D, 98, 103517
Dimitriou A., Weniger C., Correa C. A., 2022, arXiv e-prints
Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359
Fadely R., Keeton C. R., 2012, MNRAS, 419, 936
Feldmann R., Spolyar D., 2015, MNRAS, 446, 1000
Fitts A., et al., 2017, MNRAS, 471, 3547
Gilman D., Birrer S., Treu T., Keeton C. R., Nierenberg A., 2018, MNRAS, 481, 819
Gilman D., Birrer S., Nierenberg A., Treu T., Du X., Benson A., 2020, MNRAS, 491, 6077
Gilman D., Bovy J., Treu T., Nierenberg A., Birrer S., Benson A., Sameie O., 2021, MNRAS, 507, 2432
He K., Zhang X., Ren S., Sun J., 2016, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp 770–778, doi:10.1109/CVPR.2016.90
He Q., et al., 2022, MNRAS, 511, 3046
Hermans J., Begy V., Louppe G., 2019, arXiv e-prints, p. arXiv:1903.04057
Hermans J., Banik N., Weniger C., Bertone G., Louppe G., 2021, MNRAS, 507, 1999
Hezaveh Y., Dalal N., Holder G., Kisner T., Kuhlen M., Perreault Levasseur L., 2016a, JCAP, 11, 048
Hezaveh Y. D., et al., 2016b, ApJ, 823, 37
Hoffer E., Hubara I., Soudry D., 2017, Advances in neural information processing systems, 30
Huang X., et al., 2021, ApJ, 909, 27
Jacobs C., et al., 2019, ApJS, 243, 17
Kahlhoefer F., Kaplinghat M., Slatyer T. R., Wu C.-L., 2019, J. Cosmology Astropart. Phys., 2019, 010
Keeton C. R., Kochanek C. S., Seljak U., 1997, Astrophys. J., 482, 604
Kim S. Y., Peter A. H. G., Hargis J. R., 2018, Phys. Rev. Lett., 121, 211302
Kingma D. P., Ba J., 2014, arXiv preprint arXiv:1412.6980
Kochanek C. S., White M., 2000, ApJ, 543, 514
Koekemoer A. M., et al., 2007, ApJS, 172, 196
Kormann R., Schneider P., Bartelmann M., 1994, A&A, 284, 285
Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193
Legin R., Hezaveh Y., Levasseur L. P., Wandelt B., 2021, arXiv
Loshchilov I., Hutter F., 2017, arXiv e-prints, p. arXiv:1711.05101
Lovell M. R., et al., 2012, MNRAS, 420, 2318
Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, MNRAS, 439, 300
Lynden-Bell D., Wood R., 1968, MNRAS, 138, 495
MacLeod C. L., Jones R., Agol E., Kochanek C. S., 2013, ApJ, 773, 35
Makinen T. L., Charnock T., Alsing J., Wandelt B. D., 2021, JCAP, 11, 049
Mandelbaum R., Hirata C. M., Leauthaud A., Massey R. J., Rhodes J., 2012, MNRAS, 420, 1518
Mandelbaum R., et al., 2014, ApJS, 212, 5
Mao S., Schneider P., 1998, MNRAS, 295, 587
McKean J., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 84 (`arXiv:1502.03362`)
Metcalf R. B., Madau P., 2001, ApJ, 563, 9
Metcalf R. B., et al., 2019, A&A, 625, A119
Minor Q. E., Kaplinghat M., Li N., 2017, ApJ, 845, 118
Minor Q., Kaplinghat M., Chan T. H., Simon E., 2021a, MNRAS, 507, 1202
Minor Q., Gad-Nasr S., Kaplinghat M., Vegetti S., 2021b, MNRAS, 507, 1662
Mishra-Sharma S., 2022, Machine Learning: Science and Technology, 3, 01LT03
Mishra-Sharma S., Van Tilburg K., Weiner N., 2020, Phys. Rev. D, 102, 023026

Mohamed S., Lakshminarayanan B., 2016, arXiv e-prints, p. arXiv:1610.03483
Mondino C., Taki A.-M., Van Tilburg K., Weiner N., 2020, Phys. Rev. Lett., 125, 111101
Moustakas L. A., Metcalf R. B., 2003, MNRAS, 339, 607
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Nierenberg A. M., Treu T., Wright S. A., Fassnacht C. D., Auger M. W., 2014, MNRAS, 442, 2434
Nierenberg A. M., et al., 2017, MNRAS, 471, 2224
Ostdiek B., Diaz Rivero A., Dvorkin C., 2022a, Astron. Astrophys., 657, L14
Ostdiek B., Diaz Rivero A., Dvorkin C., 2022b, Astrophys. J., 927, 83
Paszke A., et al., 2019, in Wallach H., Larochelle H., Beygelzimer A., d'Alché-Buc F., Fox E., Garnett R., eds, , Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp 8024–8035, http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
Perreault Levasseur L., Hezaveh Y. D., Wechsler R. H., 2017, Astrophys. J. Lett., 850, L7
Planck Collaboration et al., 2020, A&A, 641, A1
Read J. I., Iorio G., Agertz O., Fraternali F., 2017, MNRAS, 467, 2019
Ritondale E., Vegetti S., Despali G., Auger M. W., Koopmans L. V. E., McKean J. P., 2019, MNRAS, 485, 2179
Rocha M., Peter A. H. G., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Oñorbe J., Moustakas L. A., 2013, MNRAS, 430, 81
Scoville N., et al., 2007, ApJS, 172, 1
Spergel D. N., Steinhardt P. J., 2000, Phys. Rev. Lett., 84, 3760
Storfer C., et al., 2022, arXiv e-prints, p. arXiv:2206.02764
Taylor P. L., Kitching T. D., Alsing J., Wandelt B. D., Feeney S. M., McEwen J. D., 2019, Physical Review D, 100
Van Tilburg K., Taki A.-M., Weiner N., 2018, J. Cosmology Astropart. Phys., 2018, 041
Vegetti S., Koopmans L. V. E., Bolton A., Treu T., Gavazzi R., 2010, MNRAS, 408, 1969
Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., 2012, Nature, 481, 341
Vegetti S., Koopmans L. V. E., Auger M. W., Treu T., Bolton A. S., 2014, MNRAS, 442, 2017
Vogelsberger M., Zavala J., Loeb A., 2012, MNRAS, 423, 3740
Wagner-Carena S., Park J. W., Birrer S., Marshall P. J., Roodman A., Wechsler R. H., 2021, Astrophys. J., 909, 187
Wagner-Carena S., Aalbers J., Birrer S., Nadler E. O., Darragh-Ford E., Marshall P. J., Wechsler R. H., 2022, arXiv e-prints, p. arXiv:2203.00690
Wilks S. S., 1938, Annals Math. Statist., 9, 60
Şengül A. Ç., Dvorkin C., 2022, MNRAS, 516, 336
Şengül A. C., Tsang A., Diaz Rivero A., Dvorkin C., Zhu H.-M., Seljak U., 2020, Phys. Rev. D, 102, 063502