

# Detecting Subhalos in Strong Gravitational Lens Images with Image Segmentation

Bryan Ostdiek,<sup>\*</sup> Ana Diaz Rivero,<sup>†</sup> and Cora Dvorkin<sup>‡</sup>  
*Department of Physics, Harvard University, Cambridge, MA 02138, USA*  
 (Dated: September 16, 2020)

We develop a machine learning model to detect dark substructure (subhalos) within simulated images of strongly lensed galaxies. Using the technique of image segmentation, we turn the task of identifying subhalos into a classification problem where we label each pixel in an image as coming from the main lens, a subhalo within a binned mass range, or neither. Our network is only trained on images with a single smooth lens and either zero or one subhalo near the Einstein ring. On a test set of noiseless simulated images with a single subhalo, the network is able to locate subhalos with a mass of  $10^8 M_\odot$  and place them in the correct or adjacent mass bin, effectively detecting them 97% of the time. For this test set, the network detects subhalos down to masses of  $10^6 M_\odot$  at 61% accuracy. However, noise limits the sensitivity to light subhalo masses. With 1% noise (with this level of noise, the distribution of signal-to-noise in the image pixels approximates that of images from the *Hubble Space Telescope* for sources with magnitude  $< 20$ ), a subhalo with mass  $10^{8.5} M_\odot$  is detected 86% of the time, while subhalos with masses of  $10^8 M_\odot$  are only detected 38% of the time. Furthermore, the model is able to generalize to new contexts it has not been trained on, such as locating multiple subhalos with varying masses, subhalos far from the Einstein ring, or more than one large smooth lens.

Strong gravitational lensing occurs when light from a distant source (for example, a galaxy or quasar) is distorted and magnified due to the gravitational influence of a large foreground mass (typically a galaxy or a cluster), which acts as a lens. The dark matter halo surrounding the foreground galaxy accounts for most of the mass of the lens, and smaller dark matter structures within the lens can cause small perturbations to it.<sup>1</sup> Galaxy-galaxy lenses are particularly interesting because these perturbations are caused by very low-mass halos, which, if resolved, can give us a window into the smallest scales of structure formation and test so-far untested predictions of the standard Lambda Cold Dark Matter ( $\Lambda$ CDM) paradigm. Exotic dark matter models, such as warm dark matter, self-interacting dark matter and ultralight bosonic dark matter, all cause significant deviations from the standard model of CDM, for example, introducing a low-mass cutoff of the halo mass function, meaning that overdensities below a certain threshold do not collapse to form bound structures, modifying the dark matter distribution within halos. Because of this, there has been an increased body of work that aims to detect substructure in strong gravitational lenses.

Traditional techniques to directly find substructure from strong lens images rely on modeling the smooth component of the lensing galaxy and reconstructing the source, subsequently ray-tracing the source through the gravitational potential of the smooth lens, and looking for residuals between the generated and observed images that could be explained by local over-

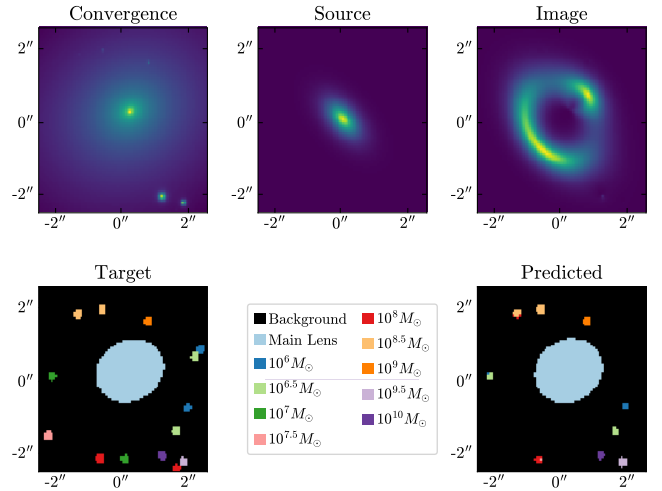


FIG. 1: Example of finding subhalos with image segmentation. The simulated convergence field of lens halos (upper-left panel) is used to define target labels (lower-left).

We define our subhalo labels to be a circle with radius 2 pixels centered on the subhalo. Even though this is not representative of relative convergences of the different masses, it was found to lead to more stable results. A light source (upper-middle) is placed behind the lens, resulting in an observed image (upper-right) after ray-tracing the light through the gravitational potential of the lens. A neural network takes in only the observed image and predicts the label for each pixel (lower-right).

densities [6–10]. Two systems with evidence for substructure have been found using one such direct detection method, gravitational imaging, one with a mass of  $(3.51 \pm 0.15) \times 10^9 M_\odot$  [11] and one with a mass of  $(1.9 \pm 0.1) \times 10^8 M_\odot$  [12], although in theory this method should be sensitive to masses as low as  $10^7 M_\odot$  for a subhalo on the Einstein ring [9]. The downside to this method is that the smooth component and source have

<sup>1</sup> Technically perturbations can also be caused by field halos along the line of sight [1–5], and while here we concern ourselves with substructure, our method is fully generalizable to images with line-of-sight perturbers.

to be modelled and/or reconstructed, and inaccuracies in the lens model can lead to extra residuals and false positives [13]. Furthermore, the effect of substructure can be partially reabsorbed by changing the source model and, to a lesser extent, the smooth model. Finally, the modeling pipeline can take weeks, and alternative modeling approaches can lead to significantly different results for the same system [14].

These so-called *direct detection* efforts have a further drawback, which is that they usually hope to find a single (sometimes two) perturber(s) in an image. Due to the  $\Lambda$ CDM expectation of a steep low-mass end of the halo mass function, we actually expect a large population of such halos, which might individually be undetectable. *Statistical detection* efforts therefore take a different approach: instead of attempting to individually resolve and localize  $\mathcal{O}(1)$  clumps above some detection threshold, they characterize the statistical properties of the population of subhalos whose collective effect can be detected by looking at statistical perturbations in images. Despite their benefits, most of these still require either modeling the main lens or simultaneously inferring both the main lens and substructure parameters [15–22].

Despite the difficulty of characterizing the substructure content of a lensing galaxy, the fact that the data comes in the form of images makes this problem an ideal candidate for cutting-edge machine learning (ML) techniques that have excelled at image recognition tasks across many domains. In Ref. [23], a CNN was used for binary classification to determine if a strong lens image contained substructure beyond the main lens or not. Refs. [24] and [25] built networks that assume the presence of a population of substructure and infer its properties, such as the fraction of the overall galaxy mass contained in substructure along with the slope of the subhalo mass function power law, or the low-mass cutoff of the subhalo mass function, respectively. Finally, Ref. [26] used CNNs to classify and distinguish between different types of dark matter substructure.

In this letter, we take an entirely different approach to substructure detection. We train a network to do image segmentation (defined as labeling every pixel within an image) to find subhalos in strong lens images, as shown in Fig. 1. To segment the images, we use a U-Net architecture [27] consisting of multiple convolutional layers along both contracting (course-graining the image) and expanding (up sampling the course-grained image) paths.<sup>2</sup> With the U-Net, we can find substructure within a fraction of a second directly from the image instead of minimizing residuals from the modeled system. Specific details about the network and our simulated data can be

found in our companion paper [31], in which we count the number of pixels predicted for each subhalo mass to determine the subhalo mass function slope.

## SUBHALO DETECTION ACCURACY

Our U-Net is trained on images with either no substructure or exactly one subhalo in a bright pixel near the Einstein ring. We define this to be any pixel that is at least 20% as bright as the brightest pixel. The training set is comprised of  $9 \times 10^4$  images, with 1/10 having only a smooth lens of mass  $\mathcal{O}(10^{13} M_\odot)$  and the other 9/10 additionally containing one subhalo that falls into one of nine mass bins. The center of the mass bins are logarithmically spaced as  $\{10^6, 10^{6.5}, 10^7, 10^{7.5}, 10^8, 10^{8.5}, 10^9, 10^{9.5}, 10^{10}\} M_\odot$ . The task of the network is to predict the class for each pixel within the image into one of 11 categories (the main lens, a subhalo within one of the nine mass bins, or neither). Each simulated image contains a unique lens and source. While at face value this is a classification problem, we note that it simultaneously locates and estimates the mass of the subhalos.

The network is tested against an independent set of  $10^4$  images, which are drawn from the same population. In Fig. 2, we show the confusion matrix for the individual pixels for all the images in the test set. The columns correspond to the true class (target) of a given pixel, while the rows show what the model classifies it as (predicted). We normalize the columns such that they sum to unity: with this choice, the values in each column show the probability that a pixel with a given true label was predicted to be in each of the eleven possible classes. The left panel corresponds to a model trained and tested on images with no noise, while the right panel corresponds to a model trained and tested on images with Gaussian noise with a standard deviation of 1% of the mean brightness of the pixels near the Einstein ring. All images (with or without noise) are convolved with a Gaussian point spread function (PSF) with a full-width half-max of  $0.07''$ , which roughly corresponds to the PSF size of the Hubble Space Telescope.

We can see from the left panel of Fig. 2 that the matrix is mostly diagonal, implying remarkable accuracy overall. The two dark squares in the lower-left corner show that background and main lens pixels are almost always classified correctly.

For the subhalo classes, we see that there is always a non-zero probability of pixels getting assigned to the two adjacent classes by the model. When the network misclassifies a pixel, it is often still locating a subhalo but getting a slightly higher or lower mass estimate. This is not surprising since the subhalo masses can lie anywhere within their class and can therefore lie near the boundaries of adjacent classes. It is remarkable that even for extremely low-mass subhalos  $\lesssim 10^7 M_\odot$ , the highest

<sup>2</sup> Details of our PYTORCH [28] implementation of the U-Net, along with specifics of our image simulation within LENSTRONOMY [29, 30] can be found in the supplementary material.

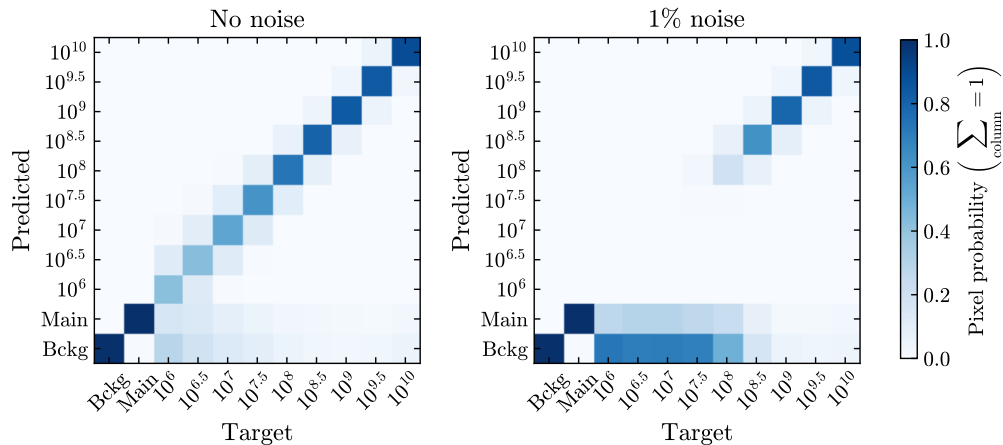


FIG. 2: Confusion matrix for the pixels of the  $10^4$  test images. Each image contains  $80 \times 80$  pixels which are then placed in the matrix according to their truth target label and the label predicted by the network. The matrix is then normalized so the sum of the columns is unity. Most of the data lies along the diagonal, showing good accuracy. The network makes some errors by correctly locating the subhalo, but predicting the neighboring mass bin. When including noise, subhalos with  $m \lesssim 10^8 M_\odot$  are likely to be missed altogether and get predicted as background or the main lens.

	Prediction Category	$10^6 M_\odot$	$10^{6.5} M_\odot$	$10^7 M_\odot$	$10^{7.5} M_\odot$	$10^8 M_\odot$	$10^{8.5} M_\odot$	$10^9 M_\odot$	$10^{9.5} M_\odot$	$10^{10} M_\odot$
No noise	Correct mass bin %	48.0	47.5	58.5	65.9	78.0	84.1	87.4	87.9	96.2
	+ adjacent mass bins %	60.7	72.1	83.1	90.2	97.0	97.9	99.9	99.9	99.9
1% noise	Correct mass bin %	0	0	0	0.4	29.9	71.6	84.1	89.2	95.6
	+ adjacent mass bin %	0	0	0.1	3.5	38.0	86.0	98.7	99.6	99.8

TABLE I: If most of the true pixels for subhalo are predicted by the network to be in a mass bin, we label the subhalo with the corresponding mass bin. The numbers represent what percent of the time a subhalo of a given mass is detected by the network. We also display when the network detects the subhalo but possibly assigns it to an adjacent mass bin.

probability lies along the diagonal, although there is also a non-negligible probability of getting classified as main or background, as expected. Impressively, the network is able to identify these light subhalos despite their perturbations to the brightness being less than a 0.1% effect.

While the method in principle has an exceptionally good sensitivity, it invariably degrades with noise. When including noise at the level of 1% of the mean image brightness, the lower mass reach of the model is around  $10^8 M_\odot$ . This 1% noise is found to be a realistic approximation for sources brighter than magnitude 20 [31]. Above this mass, the matrix is again nearly diagonal. Below this mass, nearly all of the pixels are predicted as main or background because the effect induced by the substructure is less than the noise. One important conclusion we draw is that, despite losing sensitivity, the network does not add spurious subhalos randomly across the image when noise is added.

The confusion matrices discussed so far were computed for individual pixels in the images of the test set. However, we are actually more interested in the subhalo detection accuracy as opposed to the pixel accuracy. For instance, the network could miss the edge of a subhalo,

but still get 75% of the subhalo pixels correct. We would then say that the network found and identified the subhalo (by contrast, in the per-pixel accuracy view, this would correspond to 75% accuracy). To assess this, we examined the pixels that were supposed to be marked as subhalo pixels with a given mass and analyzed the resulting predictions. The prediction with the most counts determines our label for the overall subhalo mass. The per-subhalo results are summarized in Table I, both for getting the correct mass bin, or detecting the substructure but possibly assigning it to an adjacent mass bin. This shows that without noise, the network finds 61% of the subhalos in the  $10^6 M_\odot$  bin and gets the mass correct for nearly 80% of these. With 1% noise, the network identifies subhalos in the  $10^9 M_\odot$  bin and above more than 98% of the time with more than 85% of these in the correct mass bin. Image segmentation is able to both locate and get the mass of subhalos in strongly lensed images.

## DOMAIN ADAPTATION

Machine learning models work by discovering correlations in training data. It is challenging to get them to work when new data is outside the realm of what they have been trained on, and a whole subfield is dedicated to the problem of adapting to new domains.

Both in the training and in the testing presented, each image had either no substructure or exactly one subhalo. Due to the steep slope of the subhalo mass function, we expect that many subhalos should be present in strong lensing images [32]. We therefore assess whether the network is capable of generalizing to other lensing situations after this training regimen.

In Fig. 1 we show an example with no noise, where we included many subhalos both near the Einstein ring and closer to the edge. Near the edge, the network is able to detect many of the heavier subhalos, but does not capture the lighter ones. The fact that training on a single subhalo generalizes to many subhalos is remarkable. It opens the possibility of using the results of the network to infer the subhalo mass function using many fewer images than if we could only detect a single subhalo at a time. This is the subject of the companion paper [31].

In panels A and B of Fig. 3, we examine the effect of two subhalos being close or overlapping. We choose two subhalos in the  $10^{9.5}M_\odot$  bin because they are easy to identify by eye. The source light and main galaxy lens are kept constant throughout these images to see only the effect due to subhalos, and we include 1% Gaussian noise. In A, the two subhalos are far enough away from each other that the network is able to resolve them separately. In panels B, the two subhalos are close enough that the true pixels are touching each other. The network does not correctly identify two individual subhalos in this case, but it does classify the pixels as belonging to a single, higher mass subhalo,  $10^{10}M_\odot$ . An animation of the subhalo traversing the image and its effect on the network output can be found at [this link](#).

Finally, in panels C and D of Fig. 3 a second large halo is included in the image with a mass on the same order of the main lens, meaning that there are two main lens halos rather than a single one. The combined lenses result in much larger distortions of the light, as shown in the input images in the left columns. The Einstein ring in the training images was always around  $1''$  in radius, while in these images it is closer to  $2''$ . This is why the network predicts the central lens as around 30% too large. While the network is capable of identifying the presence and location of a second main lens, its shape is not captured well. In fact, we can see in panel D, how it adds a heavy subhalo to the center of the second lens. An animation showing the different positions of the second lens, and its effect on pixel classification, is available at [this link](#).

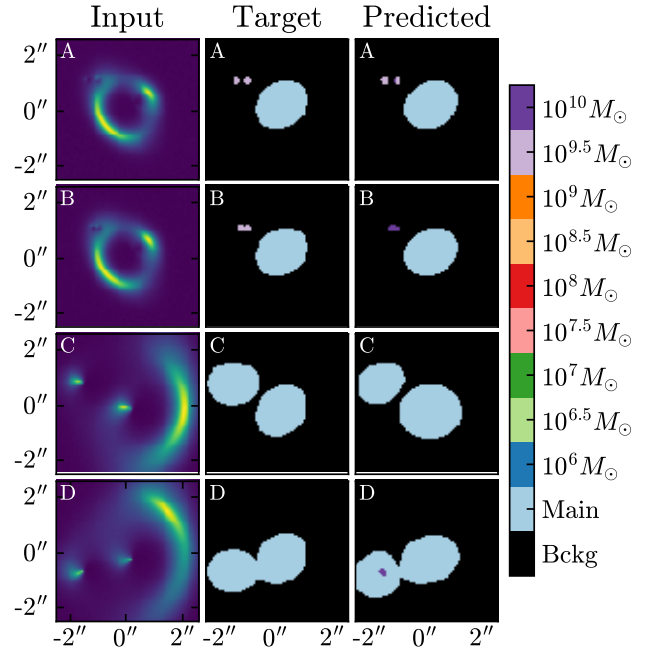


FIG. 3: The left column displays the strongly lensed images used as input to the neural network. The middle column contains the pixel-by-pixel labels we want the network to reproduce. The right column shows the network predictions.

## DISCUSSION AND OUTLOOK

This letter presents a method to determine both the mass and location of dark matter substructure within the halos of strong gravitational lenses using image segmentation on simulated images. When tasked with classifying pixels among eleven different classes, the neural network can identify the collection of pixels corresponding to subhalos of mass  $10^6M_\odot$  with 48% accuracy and identifies subhalos with mass  $m \geq 10^8M_\odot$  with more than 78% accuracy for noiseless images. This is quite impressive since the fractional change in intensity for a single subhalo of this mass is  $\mathcal{O}(\sim 10^{-4})$  [ $\mathcal{O}(\sim 10^{-2})$ ] for subhalos with mass  $10^6M_\odot$  [ $10^8M_\odot$ ]. When including 1% noise, the sensitivity is reduced to 72% for  $m \geq 10^{8.5}M_\odot$ . It drops to 30% for  $10^8M_\odot$  subhalos, because the standard deviation of the Gaussian noise is of the same order as the pixel brightness changes induced by  $10^8M_\odot$  subhalos, and lower mass subhalos are missed altogether. We note that the network does not add spurious subhalos at a significant rate, even when noise is included in the images.

While the network was trained only on images with a single strong lens and a single low-mass perturber near the Einstein ring, it is able to detect subhalos that are far from the lensed images, as well as the mass and location of additional substructure. When subhalos overlap, the network labels the sum of their effects. Furthermore, we showed that the network is also capable of identifying

two separate main lenses. These results suggest that the network has learned something fundamental about the lensing of light in General Relativity. Domain adaptation is a notoriously difficult task for machine learning models, but by learning the fundamental physics giving rise to these images, the network is able to generalize very well to images that look very different from the training samples but are nevertheless governed by the same principles.

The success of our image segmentation technique is very encouraging for various science applications of strong lens images. Our primary goal here has been the identification of dark substructure. By showing that it is able to locate and assign mass accurately to even very low mass halos paves the way for a new method of extracting the subhalo mass function from strong lens images (which we do for simulated images in our companion paper [31]). The subhalo mass function is a key target for dark matter science as we can use it to diagnose deviations from the cold dark matter scenario.

In sum, when including realistic levels of noise, we reach good accuracy for the same range of masses detected by traditional methods, but while these take  $\mathcal{O}(\text{weeks})$  to analyze each system, we do it in a fraction of a second. Furthermore, the success of the network in generalizing to multiple main lenses can also be seen as an indication that using image segmentation could help in the fundamental task of mass modeling for strong gravitational lenses, which is a very difficult and time-consuming process (inferring the presence due to substructure often requires mass-modeling as a first step). Of course here we simply used the two-lens images as an example of the out-of-sample adaptability of our network, but one could train an image segmentation model exclusively on images with more complicated main lens mass distributions.

We thank Simon Birrer, A. Cagan Sengul, and Arthur Tsang for help with the code LENSTRONOMY. We also thank Simon Birrer, Johann Brehmer, Tim Cohen, and Siddharth Mishra-Sharma for helpful comments on a previous version of this manuscript. BO was supported in part by the U.S. Department of Energy under contract DE-SC0013607. CD was partially supported by the Department of Energy (DOE) Grant No. DE-SC0020223.

# Detecting Subhalos in Strong Gravitational Lensing with Image Segmentation

## *Supplementary Material*

Bryan Ostdiek, Ana Diaz Rivero, and Cora Dvorkin

### DATA GENERATION

The main text of this letter presented results of a machine learning technique to detect subhalos in strongly lensed images. A large dataset of images was used to both train the neural network as well as test its capabilities. To acquire this dataset, we simulated strong gravitational lens images using the software package LENSTRONOMY [29, 30]. We study images which are  $80 \times 80$  pixels covering an area of  $5'' \times 5''$ , giving a resolution of  $0.06''$ . Distance are computed using the *Planck* 2015 results in Ref. [33]. The **main lens** is chosen to have a singular isothermal ellipsoid (SIE) [34] profile with an Einstein radius around  $1''$ . We allow for some ellipticity and place the lens near the center, such that the lensed images are contained in the window. The lens is placed at redshift  $z_{\text{lens}} = 0.2$ . **Subhalos** are added to the lens, modeled by truncated NFW profiles [35] with a concentration parameter  $c = 15$ . The truncation occurs at five times the scale radius. With the full lens (main galaxy plus subhalos) the **target labels** are defined. All of the pixels within the Einstein radius of the SIE are marked as the main lens. Pixels within a circle of radius 2 pixels of a subhalo are labeled according to the subhalo mass bin. The **source light** is placed near the center of the image at a redshift of  $z_{\text{source}} = 0.6$ . We allow for up to four clumps of light, each modeled as a Srsic ellipse, with a radius between 0.1 and 1.0 kpc. The ellipticity of each clump is constrained to be less than 0.4 and each of the clumps must lie near each other. In this way, the resulting image resembles galaxy-galaxy strong lens images, rather than multiple far-away clumps of light lensed into multiple different images. All of our images are convolved with a Gaussian point spread function (PSF) kernel with a full-width half-maximum of  $0.07''$ , similar to that of the *Hubble Space Telescope* (HST). When **noise** is added to an image, it is done by drawing from a Gaussian with zero mean and standard deviation which is 1% of the mean pixel brightness of the image. For further discussion of the data generation, see our companion paper [31].

### NETWORK ARCHITECTURE AND TRAINING

To perform the image segmentation task, we implement a U-Net model [27] with PYTORCH [28]. The input to the network is an  $80 \times 80$  pixel image with a single channel (a grayscale image). We normalize the image by dividing by the maximum value. The U-Net consists of two halves, first a contracting path (where the image dimension shrinks), which is then followed by an expanding path (where the image dimension is dilated again).

The U-Net is built by repeating a series of three operations. These operations are convolution, batch normalization, and ReLU activation, which are then repeated a second time. This combination of operations will be referred to as a *block*. During the first block, the convolutional layers have height and width of  $3 \times 3$  and 64 filters (resulting in 64 channels). The stride and padding on the convolution block are set to keep the height and width unchanged throughout the block, while allowing for more channels.

The resulting image is then down-sampled to  $40 \times 40$  pixels using a max pooling layer. The down-sampling allows the next layers to effectively apply to a larger area of the image, giving the network the ability to learn features on different scales. From here, another block with 128 filters is applied. The result is again down-sampled to  $20 \times 20$  pixels. Next is a block with 256 filters. The final contraction is down-sampling to  $10 \times 10$  pixels. Additional blocks are used, with filter sizes of 512 and 256 in the two convolutional layers, respectively.

In the expansion path, lower resolution information is up-sampled to higher resolution. This is done with a large number of feature channels, which allows the network to propagate information to the high resolution layers. The upsamplings are done with a transposed convolutional layer with a height and width of  $3 \times 3$  in PYTORCH.

The output of the last layer in the contracting path is a  $10 \times 10 \times 256$  array. In the first step of the expanding path the first two dimensions are up-sampled to  $20 \times 20$  and then concatenated with the last layer in the contracting path with the same resolution. Following the concatenation, the data has dimensions of  $20 \times 20 \times 512$ . This is followed by a block with 256 and 128 filters. The data is again up-sampled, now to  $40 \times 40$  and concatenated with the corresponding layer of the contracting path. A blocks is again applied, with 128 and 64 filters.

A final up-sampling brings the data back to the original resolution. This is then concatenated with the last layer before down-sampling. A block is applied, with 64 and 11 filters. These final 11 filters correspond to the 11 different classes the pixels can belong to (main lens, one of nine subhalo mass bins, or background). The softmax activation

function is applied along the channel dimension such that the sum of the 11 features for each pixel is 1. Thus, the pixel channels correspond to the probability of a pixel belonging to each class.

The cross entropy loss function is computed for each pixel and the network parameters are tuned using the Adam optimizer [36] with the standard  $\beta$  values to minimize the loss. We start with a learning rate of  $10^{-3}$  and iterate through the data in batches of 100 images. The loss is evaluated on an independent validation set after each epoch. When the validation loss has not improved for 5 epochs, the learning rate is lowered by a factor of 10; we do not allow the learning rate to drop below  $10^{-6}$ . The training ends when the validation loss has not improved for 15 epochs. Further discussion of the model architecture and training methodology can be found in our companion paper [31].

---

\* bostdiek@g.harvard.edu

† adiazrivero@g.harvard.edu

‡ cdvorkin@g.harvard.edu

- [1] A. DAloisio and P. Natarajan, Monthly Notices of the Royal Astronomical Society **411**, 16281640 (2010).
- [2] R. Li, C. S. Frenk, S. Cole, Q. Wang, and L. Gao, Mon. Not. Roy. Astron. Soc. **468**, 1426 (2017), arXiv:1612.06227 [astro-ph.CO].
- [3] C. McCully, C. R. Keeton, K. C. Wong, and A. I. Zabludoff, Astrophys. J. **836**, 141 (2017), arXiv:1601.05417 [astro-ph.CO].
- [4] G. Despali, S. Vegetti, S. D. M. White, C. Giocoli, and F. C. van den Bosch, Mon. Not. Roy. Astron. Soc. **475**, 5424 (2018), arXiv:1710.05029 [astro-ph.CO].
- [5] A. c. Şengül, A. Tsang, A. Diaz Rivero, C. Dvorkin, H.-M. Zhu, and U. Seljak, Phys. Rev. D **102**, 063502 (2020), arXiv:2006.07383 [astro-ph.CO].
- [6] S. Mao and P. Schneider, Mon. Not. R. Astron. Soc. **295**, 587 (1998), arXiv:astro-ph/9707187 [astro-ph].
- [7] L. A. Moustakas and R. B. Metcalf, Mon. Not. R. Astron. Soc. **339**, 607 (2003), arXiv:astro-ph/0206176 [astro-ph].
- [8] L. V. E. Koopmans, Mon. Not. R. Astron. Soc. **363**, 1136 (2005), arXiv:astro-ph/0501324 [astro-ph].
- [9] S. Vegetti and L. V. E. Koopmans, Mon. Not. R. Astron. Soc. **392**, 945 (2009), arXiv:0805.0201 [astro-ph].
- [10] Y. Hezaveh, N. Dalal, G. Holder, M. Kuhlen, D. Marrone, N. Murray, and J. Vieira, ApJ **767**, 9 (2013), arXiv:1210.4562 [astro-ph.CO].
- [11] S. Vegetti, L. V. E. Koopmans, A. Bolton, T. Treu, and R. Gavazzi, Mon. Not. R. Astron. Soc. **408**, 1969 (2010), arXiv:0910.0760 [astro-ph.CO].
- [12] S. Vegetti, D. J. Lagattuta, J. P. McKean, M. W. Auger, C. D. Fassnacht, and L. V. E. Koopmans, Nature **481**, 341 (2012), arXiv:1201.3643 [astro-ph.CO].
- [13] E. Ritondale, S. Vegetti, G. Despali, M. W. Auger, L. V. E. Koopmans, and J. P. McKean, Mon. Not. R. Astron. Soc. **485**, 2179 (2019), arXiv:1811.03627 [astro-ph.CO].
- [14] S. Vegetti, L. V. E. Koopmans, M. W. Auger, T. Treu, and A. S. Bolton, Mon. Not. R. Astron. Soc. **442**, 2017 (2014), arXiv:1405.3666 [astro-ph.GA].
- [15] Y. Hezaveh, N. Dalal, G. Holder, T. Kisner, M. Kuhlen, and L. Perreault Levasseur, JCAP **2016**, 048 (2016), arXiv:1403.2720 [astro-ph.CO].
- [16] B. J. Brewer, D. Huijser, and G. F. Lewis, Mon. Not. R. Astron. Soc. **455**, 1819 (2016), arXiv:1508.00662 [astro-ph.IM].
- [17] F.-Y. Cyr-Racine, L. A. Moustakas, C. R. Keeton, K. Sigurdson, and D. A. Gilman, Phys. Rev. D **94**, 043505 (2016), arXiv:1506.01724 [astro-ph.CO].
- [18] S. Birrer, A. Amara, and A. Refregier, JCAP **2017**, 037 (2017), arXiv:1702.00009 [astro-ph.CO].
- [19] T. Daylan, F.-Y. Cyr-Racine, A. Diaz Rivero, C. Dvorkin, and D. P. Finkbeiner, ApJ **854**, 141 (2018), arXiv:1706.06111 [astro-ph.CO].
- [20] A. Diaz Rivero, F.-Y. Cyr-Racine, and C. Dvorkin, Phys. Rev. D **97**, 023001 (2018), arXiv:1707.04590 [astro-ph.CO].
- [21] A. Diaz Rivero, C. Dvorkin, F.-Y. Cyr-Racine, J. Zavala, and M. Vogelsberger, Phys. Rev. D **98**, 103517 (2018), arXiv:1809.00004 [astro-ph.CO].
- [22] S. Brennan, A. J. Benson, F.-Y. Cyr-Racine, C. R. Keeton, L. A. Moustakas, and A. R. Pullen, Mon. Not. R. Astron. Soc. **488**, 5085 (2019), arXiv:1808.03501 [astro-ph.GA].
- [23] A. Diaz Rivero and C. Dvorkin, Phys. Rev. D **101**, 023515 (2020), arXiv:1910.00015 [astro-ph.CO].
- [24] J. Brehmer, S. Mishra-Sharma, J. Hermans, G. Louppe, and K. Cranmer, ApJ **886**, 49 (2019), arXiv:1909.02005 [astro-ph.CO].
- [25] S. Varma, M. Fairbairn, and J. Figueroa, arXiv e-prints, arXiv:2005.05353 (2020), arXiv:2005.05353 [astro-ph.CO].
- [26] S. Alexander, S. Gleyzer, E. McDonough, M. W. Toomey, and E. Usai, ApJ **893**, 15 (2020), arXiv:1909.07346 [astro-ph.CO].
- [27] O. Ronneberger, P. Fischer, and T. Brox, arXiv e-prints, arXiv:1505.04597 (2015), arXiv:1505.04597 [cs.CV].
- [28] 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, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alche-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019) pp. 8024–8035.
- [29] S. Birrer and A. Amara, Physics of the Dark Universe **22**, 189 (2018), arXiv:1803.09746 [astro-ph.CO].

- [30] S. Birrer, A. Amara, and A. Refregier, *ApJ* **813**, 102 (2015), arXiv:1504.07629 [astro-ph.CO].
- [31] B. Ostdiek, A. D. Rivero, and C. Dvorkin, to appear (2020).
- [32] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White, *Mon. Not. R. Astron. Soc.* **391**, 1685 (2008), arXiv:0809.0898 [astro-ph].
- [33] P. Ade *et al.* (Planck), *Astron. Astrophys.* **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].
- [34] R. Kormann, P. Schneider, and M. Bartelmann, *A&A* **284**, 285 (1994).
- [35] J. F. Navarro, C. S. Frenk, and S. D. M. White, *ApJ* **462**, 563 (1996), arXiv:astro-ph/9508025 [astro-ph].
- [36] D. P. Kingma and J. Ba, arXiv e-prints, arXiv:1412.6980 (2014), arXiv:1412.6980 [cs.LG].