

Deep learning the astrometric signature of dark matter substructure

Kyriakos Vattis^{*,†}, Michael W. Toomey[†], and Savvas M. Koushiappas[‡]

*Department of Physics, Brown University, Providence, Rhode Island 02912-1843, USA
and Brown Theoretical Physics Center, Brown University, Providence, Rhode Island 02912-1843, USA*



(Received 28 August 2020; accepted 18 November 2021; published 20 December 2021)

We study the application of machine learning techniques for the detection of the astrometric signature of dark matter substructure. In this proof of principle, a population of dark matter subhalos in the Milky Way will act as lenses for sources of extragalactic origin such as quasars. We train RESNET-18, a state-of-the-art convolutional neural network to classify angular velocity maps of a population of quasars into lensed and nonlensed classes. We show that an SKA-like survey with extended operational baseline can be used to probe the substructure content of the Milky Way and demonstrate how axiomatic attribution can be used to localize substructures in lensing maps.

DOI: [10.1103/PhysRevD.104.123541](https://doi.org/10.1103/PhysRevD.104.123541)

I. INTRODUCTION

The Standard Cosmological Model with 70% dark energy in the form of a cosmological constant, 25% of cold dark matter, and 5% of baryonic matter, has been quite successfully tested in recent years from galactic scales to cosmological scales [1–5]. However, one of its main components, dark matter, has only been observed through its gravitational effects despite all the efforts through direct and indirect detection [6–19] and colliders like the Large Hadron Collider [20,21] to observe weakly interacting massive particles or other popular candidates like axions [22–25]. While the efforts continue, the null results highlight the necessity of finding alternative gravitational signatures that would shed light on the elusive nature of dark matter.

A very promising way to probe various dark matter models gravitationally is to study their effects on the distribution of dark matter on subgalactic scales, an idea that has been studied thoroughly in the literature. Several methods have been proposed, including the utilization of tidal streams [26–31], astrometric observations [32–35], pulsar timing array observations [36], stellar wakes [37], and images of strongly lensed background galaxies or quasars [38–52] looking for dark subhalos. For most of the aforementioned methods, the expected signal is small and difficult to observe.

With the increase of computational power and the improvement of machine learning algorithms, there have been multiple attempts to involve neural network-based

approaches for such analyses. In the last few years, machine learning has been used in an array of problems in cosmology [53] as well as in other problems in the physical sciences [54]. Examples include applications in large-scale structure [55], the cosmic microwave background [56–58], the cosmological 21 cm signal [59], and lensing studies [60–68]. In the search for dark matter, supervised machine learning algorithms have been trained to identify substructure properties with simulated galaxy-galaxy strong lensing images. More recently, convolutional neural networks (CNNs) have been used for classification of different subhalo mass cutoffs [69], classifying dark matter (DM) halos with and without substructure [70,71], to distinguish dark matter models with disparate substructure morphology [70], and also inference of population level properties for DM substructure [72].

In this work, we investigate the possibility of applying machine learning techniques and in particular a CNN, to detect the presence of lensing effects from dark matter substructure on a population of quasars in astrometric data from future surveys based on simulated data as a proof of principle. We also demonstrate how axiomatic attribution can be used to localize substructures in lensing maps. In Sec. II, we outline the formalism followed in order to simulate the lensing signal followed by a description of creating simulated data maps. We discuss the implementation of the CNN in Sec. III and our results in Sec. IV, and our conclusions are shown in Sec. V.

II. ASTROMETRIC LENSING

A. Formalism

We begin by considering the lensing effect of a dark matter halo on the astrometric parameters of a background source (position, velocity, and acceleration), similar in

*kyriakos_vattis@brown.edu

†michael_toomey@brown.edu

‡koushiappas@brown.edu

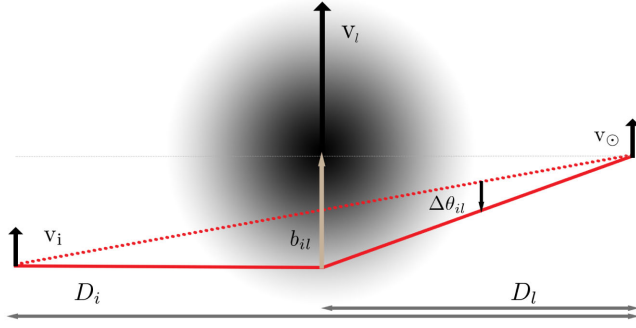


FIG. 1. Astrometric lensing of a source i by a lens l moving with velocities $\mathbf{v}_i$ and $\mathbf{v}_l$, respectively, as observed by an observer moving with velocity $\mathbf{v}_\odot$. The impact parameter vector $\mathbf{b}_{il}$ is shown in gold, and $\Delta\theta_{il}$ is the induced angular deflection. All vectors are two dimensional, and the distances D_l and D_i are considered constant.

spirit to Refs. [33,34]. We utilize a modified version of the code provided in Ref. [34] to extract the astrometric signal in a suite of models that are used to create the simulated data set used in deep learning model training.

The system we consider is shown in Fig. 1; a source i moving with velocity $\mathbf{v}_i$ is lensed by a lens l with velocity $\mathbf{v}_l$ as seen by an observer with velocity $\mathbf{v}_\odot$. The distance from observer to a lens is D_l and from observer to a source is D_i . Both are assumed to be constant, and thus all vectors involved in the calculation are two dimensional lying on the celestial sphere.

For such a system, the deflection angle $\Delta\theta_{il}$ between the real position and the position of the image of the source is given by

$$\Delta\theta_{il} = -\left(1 - \frac{D_l}{D_i}\right) \frac{4G_N M(b_{il})}{c^2 b_{il}} \hat{\mathbf{b}}_{il}, \quad (1)$$

where G_N is Newton's Gravitational constant, c is the speed of light, $\mathbf{b}_{il}$ is the impact parameter vector pointing toward the center of the lens, and $M(b_{il})$ is the enclosed lens mass, which, given a spherically symmetric density profile $\rho(r)$ for the lens, can be calculated as

$$M(b_{il}) = 2\pi \int_{-\infty}^{+\infty} dx \int_0^{b_{il}} db' b' \rho(\sqrt{x^2 + b'^2}). \quad (2)$$

Because all the components of the system are nonstationary, the impact parameter vector $\mathbf{b}_{il}$ is itself time dependent, and as a consequence, the deflection vector $\Delta\theta_{il}$ has a time dependence as well. To calculate the lensing induced velocity, we take the time derivative of Eq. (1), resulting in

$$\begin{aligned} \mathbf{u} &= \dot{\Delta\theta}_{il} \\ &= -\left(1 - \frac{D_l}{D_i}\right) \frac{4G_N}{c^2} \left\{ \frac{M'(b_{il})|b_{il}|}{b_{il}} \hat{\mathbf{b}}_{il} \right. \\ &\quad \left. + \frac{M(b_{il})}{b_{il}^2} [\mathbf{v}_{il} - 2|b_{il}| \hat{\mathbf{b}}_{il}] \right\} \end{aligned} \quad (3)$$

with M' the derivative of the enclosed mass with respect to the impact factor. The quantity $\mathbf{v}_{il}$ is the effective velocity of the lens defined as

$$\mathbf{v}_{il} = \mathbf{v}_l - \left(1 - \frac{D_l}{D_i}\right) \mathbf{v}_\odot - \frac{D_l}{D_i} \mathbf{v}_i, \quad (4)$$

where $\mathbf{v}_l$ is the velocity vector of the lens, $\mathbf{v}_i$ is the velocity vector of the source, and $\mathbf{v}_\odot$ is the velocity vector of the observer. The time derivative of the magnitude of the impact factor is given by

$$|\dot{b}_{il}| = \hat{\mathbf{b}}_{il} \cdot \mathbf{v}_{il}. \quad (5)$$

Similarly, we take the time derivative of Eq. (3) to calculate the lensing induced acceleration as

$$\begin{aligned} \mathbf{a} &= \ddot{\Delta\theta}_{il} \\ &= -\left(1 - \frac{D_l}{D_i}\right) \frac{4G_N}{c^2} \left\{ \frac{M''(b_{il})|b_{il}|^2}{b_{il}} \hat{\mathbf{b}}_{il} \right. \\ &\quad + \frac{M'(b_{il})}{b_{il}^2} [2|b_{il}|\mathbf{v}_{il} + |b_{il}|\ddot{\mathbf{b}}_{il} - 4|\dot{b}_{il}|^2 \hat{\mathbf{b}}_{il}] \\ &\quad \left. - \frac{2M(b_{il})}{b_{il}^3} [2|b_{il}|\mathbf{v}_{il} + |b_{il}|\ddot{\mathbf{b}}_{il} + 3|\dot{b}_{il}|^2 \hat{\mathbf{b}}_{il}] \right\} \end{aligned} \quad (6)$$

with M'' the second derivative of the enclosed mass with respect to the impact factor and

$$|\ddot{b}_{il}| = \frac{\mathbf{v}_{il} \cdot (b_{il}\mathbf{v}_{il} - |\dot{b}_{il}|\mathbf{b}_{il})}{b_{il}^2}. \quad (7)$$

In order to compute the deflection angle, velocity, and acceleration [Eqs. (1), (3), and (6)], we need to specify the distribution of matter in the lens, i.e., lens density profile $\rho(r)$. We assume that the density profile of the halos is described by the Navarro-Frenk-White (NFW) profile [73]

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

where r_s is the characteristic scale and ρ_s is the characteristic density of the profile.

In order to demonstrate the aforementioned effects, Fig. 2 shows an example of induced deflections, velocities, and accelerations in arbitrary units (see also Ref. [33]). The lens is located at the origin moving toward the $+\hat{\mathbf{y}}$ direction

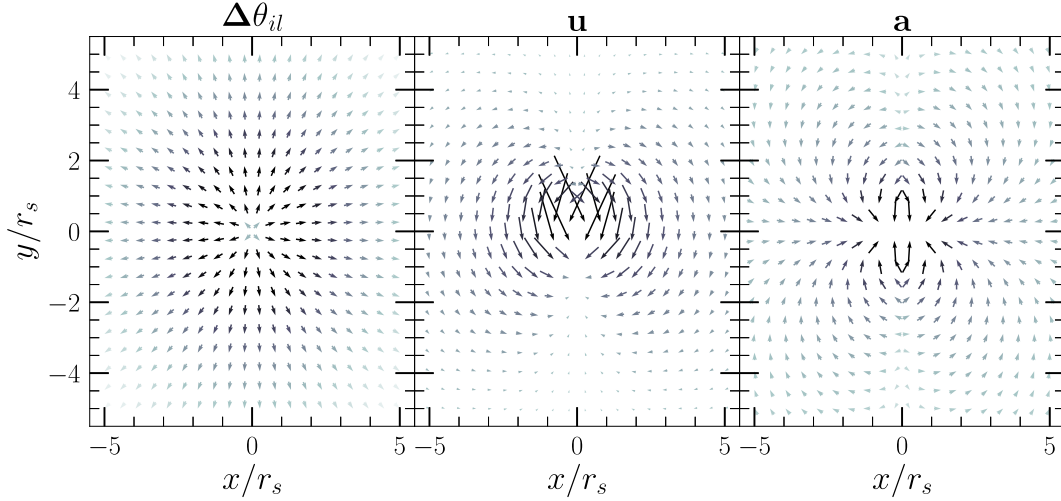


FIG. 2. The angular deflection $\Delta\theta_{il}$ and its first two time derivatives in arbitrary units induced by a NFW halo centered on the origin and moving in the $+\hat{\mathbf{y}}$ direction. Darker colors correspond to the magnitude of the vector.

in all three plots. The left panel depicts the angular deflection. The maximum angular deflection occurs for impact parameters approximately equal to the characteristic scale, as for $b_{il} < r_s$ the enclosed mass $M(< b_{il})$ decreases rapidly. On the other hand, if $b_{il} \gg r_s$, the enclosed mass approaches a constant, and therefore the deflection decreases as $\sim 1/b_{il}$. The middle panel in Fig. 2 shows the corresponding induced velocity; the largest amplitude in the velocity vector occurs near the center and in a direction opposite to the direction of motion of the lens with a dipolelike behavior in the far-field limit. Finally, the right panel of Fig. 2 shows the acceleration pattern which exhibits a quadrupolelike behavior in the far field.¹ All these characteristics are in agreement to what has been demonstrated in Ref. [33].

B. Population of lenses

We now proceed to simulate the lensing effect on a population of sources by a population of lenses. Following Ref. [34], we choose to simulate the lens population as NFW subhalos in the Milky Way that follow a mass function of the form

$$\frac{dN}{dM} = A_0 M^{-1.9} \quad (9)$$

normalized so that there are 150 subhalos in the mass range $10^8 - 10^{10} M_\odot$ [74], consistent with results from recent hydrodynamical simulations [75,76].

We then define the spatial distribution of the subhalos, taking into account the tidal disruption due to the gradient

¹Note that the amplitude of velocity and acceleration decreases faster than the displacement field as each one of these picks an additional factor of $1/b_{il}$.

of the Galactic potential toward the Galactic Center that depletes the fraction of mass bound in substructures in this region, by assuming an Einasto profile fitted to the results of the Aquarius simulation [77]

$$\rho_g(r) = \exp \left\{ -\frac{2}{\gamma_e} \left[\left(\frac{r}{r_e} \right)^{\gamma_e} - 1 \right] \right\}. \quad (10)$$

with $r_e = 199$ kpc and $\gamma_e = 0.678$. We use this distribution to sample the radii at which the subhalos lie.

The next ingredient for defining a lens population is the dark matter velocity distribution. We assume that in the Galactic frame and far away from the Sun's gravitational potential the dark matter velocity distribution is described by the Standard Halo Model as

$$f_\infty(\mathbf{v}) = N \left(\frac{1}{\pi v_0^2} \right)^{3/2} e^{-v^2/v_0^2} \quad (11)$$

for $|\mathbf{v}| < v_{\text{esc}}$ and 0 otherwise (however, see also Ref. [78]). N is a normalization factor to account for the truncation at v_{esc} , $v_0 = 220$ km/s [79], and the escape velocity is $v_{\text{esc}} = 550$ km/s [80]. We transform this velocity distribution at the position of Earth (as we are interested in measurements at the Earth's frame) by applying a Galilean transformation from the Galactic frame to the Earth frame,

$$f_\oplus(\mathbf{v}) \approx f_\infty(\mathbf{v} + \mathbf{v}_\odot), \quad (12)$$

where $\mathbf{v}_\odot = \{11, 232, 7\}$ km/s is the velocity of the Sun in $\{U_\odot, V_\odot, W_\odot\}$ coordinates [81]. We ignore the annual modulation due to the motion of the Earth around the Sun as it is a factor of 10 smaller compared to the velocity components.

We normalize the density profile of each subhalo in the following way. The NFW profile is defined by two

parameters, ρ_s and r_s , or alternatively, the mass of the halo, M_{200} , and its concentration c_{200} . The concentration is a measure of the compactness of the halos, defined as $c_{200} = R_{200}/r_s$, where R_{200} is the radius that encloses M_{200} whose density is 200 times the critical density of the Universe $\rho_c = 3H_0^2/8\pi G_N$. With this definition, the characteristic density ρ_s and scale radius r_s are obtained from

$$M_{200} = \int_0^{r_{200}} dr' 4\pi r'^2 \rho(r') = 4\pi \rho_s r_s^3 f(c_{200}), \quad (13)$$

where

$$f(x) = \ln(1+x) - \frac{x}{1+x}$$

and [82]

$$c_{200}(M_{200}, z=0) = \sum_{i=0}^5 c_i \times \left[\ln \left(\frac{M_{200}}{h^{-1} M_\odot} \right) \right]^i \quad (14)$$

with the factors c_i having values $c_i = [37.5153, -1.5093, 1.636 \times 10^{-2}, 3.66 \times 10^{-4}, -2.89237 \times 10^{-5}, 5.32 \times 10^{-7}]$.

It is clear that this choice of the description of subhalo profile parameters is unlikely as subhalos are evolved systems in the potential well of the Milky Way, while the prescription we use refers to field halos. Nevertheless, we make this choice not because it is realistic but because it can serve as an example to the proof of principle concept we explore here and also because it leads to a maximal signal. We plan to explore a thorough study of multiple other prescriptions of the mass distribution characteristics of substructure in future work.

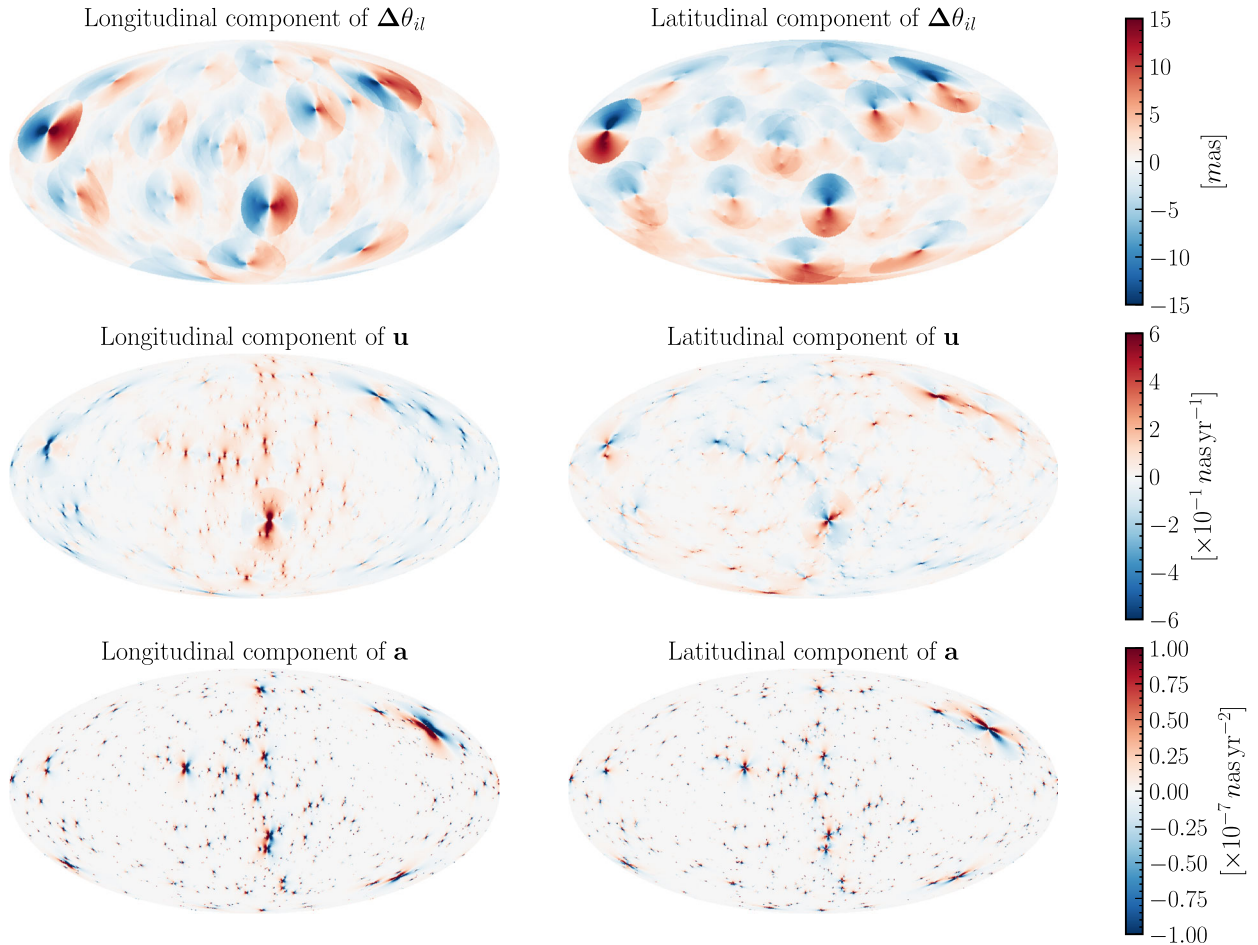


FIG. 3. The induced astrometric lensing effect on a population of quasar sources by a population of NFW subhalos distributed in the Milky Way, under the assumption of pure signal and no instrumental noise. We truncate the effect within 20 deg of the center of each subhalo. The calculation was done on a HEALPIX map with $n_{\text{side}} = 128$, and we simulate ~ 1200 subhalos with masses between 10^7 and $10^{10} M_\odot$. The first row shows the angular displacement $\Delta\theta_{il}$, the second row shows the induced velocity $\mathbf{u}$, and finally the third row shows the induced acceleration $\mathbf{a}$.

C. Population of sources

We generate a population of sources with $D_i = \infty$ and $v_i = 0$ at the center of each pixel of a HEALPIX map with $n_{\text{side}} = 128$, and we simulate ~ 1200 subhalos between 10^7 and $10^{10} M_\odot$. This approach is similar to future observations of quasars with the Square Kilometer Array (SKA), or other instruments with similar capabilities. In the absence of instrumental noise, the resulting signal within 20 deg from the center of each halo is shown in shown in Fig. 3.

In practice, the sensitivity and precision of the instruments play a significant role. For this proof of principle example, we assume an SKA-like sensitivity [83]. Given the current forecasts [83] after 10 years of operation, SKA will be able to observe up to $N_q \sim 10^8$ quasars with a peculiar velocity uncertainty of $\sigma_u \approx 1 \mu\text{as/yr}$, which we label scenario A. Additionally, we consider four more sets of parameters (scenarios B, C, D, and E) as shown in Table I. As we will see, the weakness of the signal makes the significance of detection highly dependent on these two parameters N_q and σ_u .

III. ASTROMETRIC LENSING SIGNAL

A. Characteristics of the signal in simulated data sets

In Fig. 3, we show the expected astrometric signal for the ideal case with no instrumental noise. We observe qualitative characteristics similar to Fig. 2, such as the gradient of the magnitude of the displacement field as a function of distance is small compared to the gradient in the velocity and acceleration fields (due to the $1/b_{il}$ factors;—see Footnote 1). In addition, the color bar depicts the magnitude of the expected effect. For the displacement field, the median amplitude is 1 mas, while for the velocity and acceleration fields the median values are $10^{-2} \text{ nas yr}^{-1}$ and $5 \times 10^{-10} \text{ nas yr}^{-2}$, respectively.

For the purposes of this work, the velocity vector $\mathbf{u}$ is an adequate probe of substructure lensing as quasars are stationary in the galactic frame. However, note that the acceleration can also be used for some halo mass and concentration functions as was shown in Ref. [33].

In order to apply machine learning methods in our efforts to extract dark matter substructure from the astrometric data, we first need to create a library of fake data sets which we will use in order to test the method outlined above. While Fig. 3 shows the pure signal effect, we must take into account instrumental noise. To create simulated data sets, we distribute the N_q quasars on a HEALPIX grid with $n_{\text{side}} = 32$ corresponding to 12,288 equal-area pixels. We assume quasars follow a Poisson distribution among pixels with a mean value given by the total number of quasars divided by the total number of pixels.

At each pixel, we sample a longitudinal and latitudinal velocity component from a normal distribution centered at the expected value in the presence of lensing and with a

TABLE I. Scenarios of astrometric parameters N_q and σ_u used in this work.

Scenario	N_q	$\sigma_u (\mu\text{as/yr})$
A	10^8	1
B	10^8	0.1
C	10^9	1
D	10^9	0.1
E	3×10^9	1

standard deviation σ_u . With all quasars being assigned a measured velocity under the assumption that all quasars in a pixel will experience the same lensing effect, we average all quasars and assign a single velocity to each pixel. For the case of only instrumental noise, i.e., no signal, we follow the same procedure with the normal distribution always centered at 0. We repeat the process for all five scenarios of Table I.

Figure 4 shows an example of the simulated data after taking into account the sensitivity of the instrument for parameter set C. The upper two projections contain the lensing signal, while the bottom two projections show the no-signal, only instrumental noise case.

Next, we will use this method to generate big data sets for machine learning training and validation data sets. The hope is that by training a model to distinguish between a Gaussian noise from the non-Gaussian signal maps it will be possible to extract information about the underlying population of lenses in future surveys.

B. Machine learning implementation

The structure of the data leads naturally to an implementation of a CNN. One can appreciate this from Fig. 4, in which it is clear that correlations among pixels will likely serve a key role in helping to distinguish between the signal and no-signal cases. Given this intuition, we trained state-of-the-art architecture RESNET-18 [84] as a classifier between these two classes. We selected RESNET-18 to train across all our sets after testing other algorithms (VGG [85] and ALEXNET [86]), as it consistently achieved noticeably better performance—see Fig. 5.

RESNET-18 differs from a normal CNN in that it has residual blocks which are characterized by their use of skip connections, helping to alleviate the problem of vanishing gradients. One problem, however, is that our simulations are based on data in HEALPIX format. While there are architectures that are designed to be trained on this data structure, notably graph-based CNN DEEPSPHERE [87], we instead opt to project our maps to Cartesian coordinates, allowing us to train on a greater diversity of established architectures like RESNET.

For each scenario of Table 1, our training set consists of 25 000 training and 2500 validation images per class (signal and no-signal cases). With a large number of

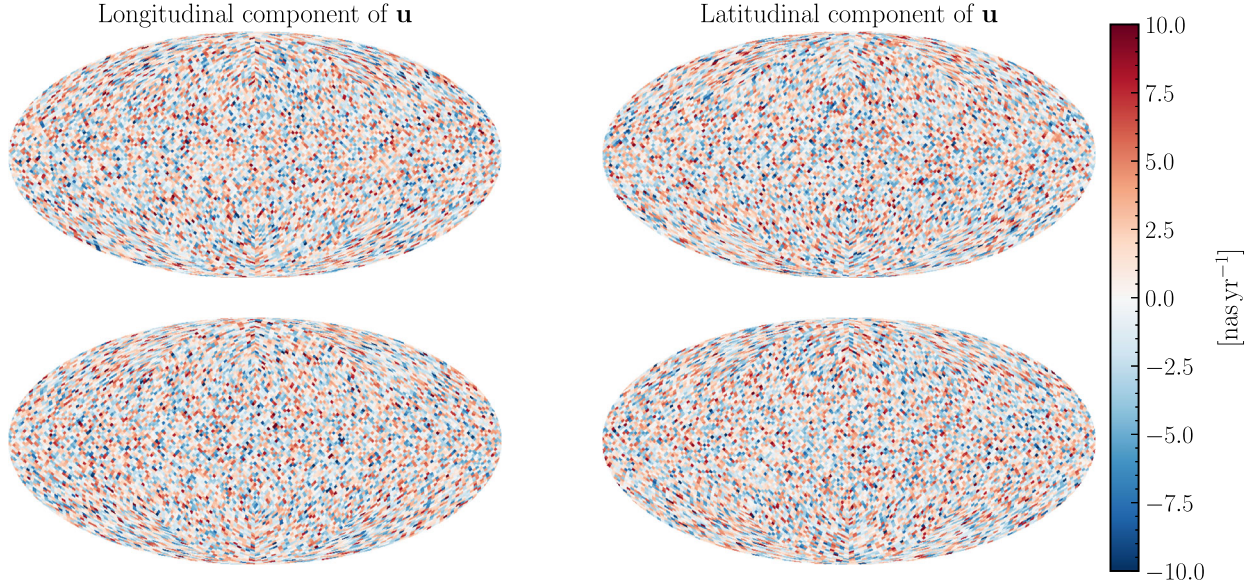


FIG. 4. Top: the induced astrometric lensing velocity on a population of quasar sources by a population of subhalos distributed in the Milky Way superimposed with Gaussian noise based on the astrometric parameters $N_q = 10^9$ and $\sigma_u = 1 \mu\text{as/yr}$ (scenario C in Table 1). Bottom: similar to top but with no signal (only noise).

simulated data, cross-validation is unnecessary, as is the case for a lot of deep learning applications. The cross-entropy loss was minimized over at most 50 epochs in batches of 32 with the Adam optimizer where the learning rate was initialized with a value of 1×10^{-1} and decayed by a factor of 10 if the validation loss was not improved after five epochs.

We implement RESNET-18 with the PYTORCH package and run on four NVIDIA Tesla V100 GPUs at the Brown Center

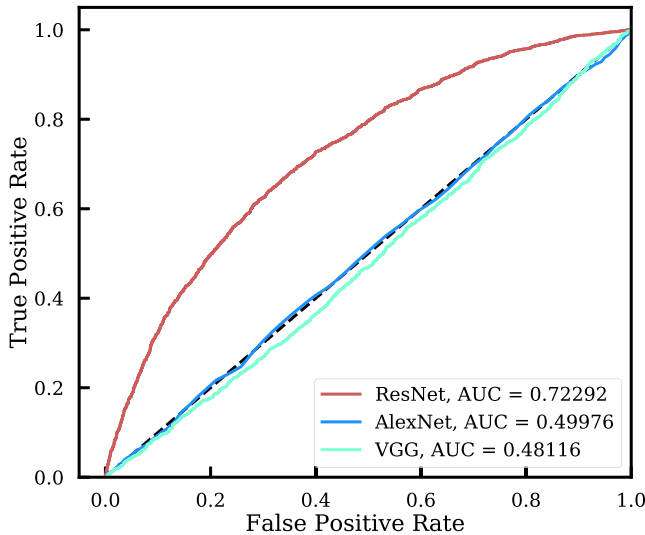


FIG. 5. ROC curves of architecture performance between RESNET-18, ALEXNET, and VGG-11 for set C. The black dashed line indicates equal true positive and false positive rate corresponding to the case that a network is guessing in its classifications.

for Computation and Visualization. As a metric for classifier performance, we utilize the well-established Area Under the receiver operating characteristic (ROC) Curve (AUC).

C. Integrated gradients

Neural networks are often referred to as *black boxes*—effectively a magician pulling a rabbit out of their hat. A true black box, however, would be of little interest. It is of critical importance if one is to use neural networks, say, on real data, to garner some understanding of how the machine comes to making its decision.

One such method that has been developed recently is the method of attributing a networks prediction to its inputs. This is most frequently realized by analyzing gradients of the network output with respect to its input. A popular method is *integrated gradients* [88], which is realized as a path integral of gradients from a baseline to the desired input,

$$IG_i(x, x') ::= (x_i - x'_i) \int_{\alpha=0}^1 \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha. \quad (15)$$

Here, F corresponds to a trained model (perhaps a classifier between two types of dark matter), and x corresponds to an input (perhaps an astrometric lensing image velocity map). For a two-dimensional image input, one can construct an assignment map where each pixel is assigned an attribution score. A positive (negative) attribution score contributes favorably (negatively) to the final network prediction, whereas pixels with no attribution do not contribute. Inputs with the largest magnitude attribution score are the most influential in the networks final decision.

IV. RESULTS

In Fig. 6, we present the ROC curves and their corresponding AUC values for each set of parameters of Tab. I. This is the main result of this work. The red lines correspond to $\sigma_u = 1 \mu\text{as/yr}$, while the blue lines are for $\sigma_u = 0.1 \mu\text{as/yr}$, and the total number of quasars N_q is designated by the line's width. As expected, larger N_q and

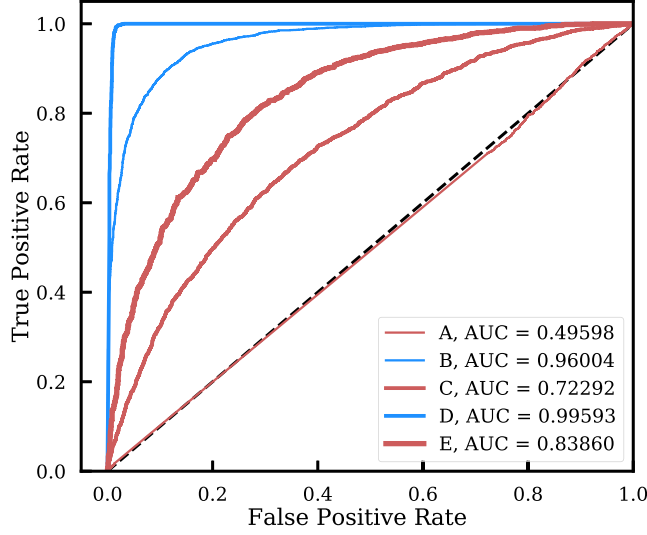


FIG. 6. ROC curves and their corresponding AUC values for each set of values of N_q and σ_u as described in Tab. I. The red lines correspond to $\sigma_u = 1 \mu\text{as/yr}$ while the blue lines are for $\sigma_u = 0.1 \mu\text{as/yr}$ and the total number of quasars N_q is designated by the line's width. The black dashed line indicates equal True Positive and False Positive Rate corresponding to the case that a network is guessing in its classifications.

smaller σ_u lead to higher AUC values and higher detection significance since the noise level of each pixel scales as $\sim \sigma_u / \sqrt{n_q}$ where n_q is the number of quasars per pixel. Additionally, we can see that a decrease in the uncertainty σ_u is more impactful than increasing the number of quasars N_q and thus a more precise survey will be better suited to detect the signal than a more sensitive one.

Our model trained on an SKA-like survey, realized here as parameter scenario A, is not sensitive to the signal induced by substructure after achieving an AUC of ~ 0.5 ; thus; in this configuration, the data are likely too noise dominated even for a CNN. However, a more optimistic survey with ten times more quasars does manage to achieve limited power at discerning between signal and noise as evidenced by a marginal AUC score of 0.723 for scenario C, but this is still not very well equipped to detect dark matter substructure based on a 10 year operational baseline.

In order to achieve a more robust performance, corresponding to AUC values $\gtrsim 0.8$, we need either the number of quasars to increase even more by a factor of at least 3 compared to scenario C and/or σ_u to decrease by approximately a factor of 10, or some combination of the above. Encouragingly, it is very promising that the detection significance improves drastically with parameter scenario D where $N_q = 10^9$ and $\sigma_u = 0.1 \mu\text{as/yr}$, allowing for almost perfect classification using the RESNET-18 architecture.

In other words, if SKA runs for longer than a decade, it is possible that the proposed technique can be successfully applied in the future to detect Milky Way substructure. For example, doubling its operational time reduces the uncertainty to $\sigma_u = 0.5 \mu\text{as/yr}$, and with the currently planned $N_q \sim 10^8$ quasars, it is possible to obtain an AUC of $\gtrsim 0.8$

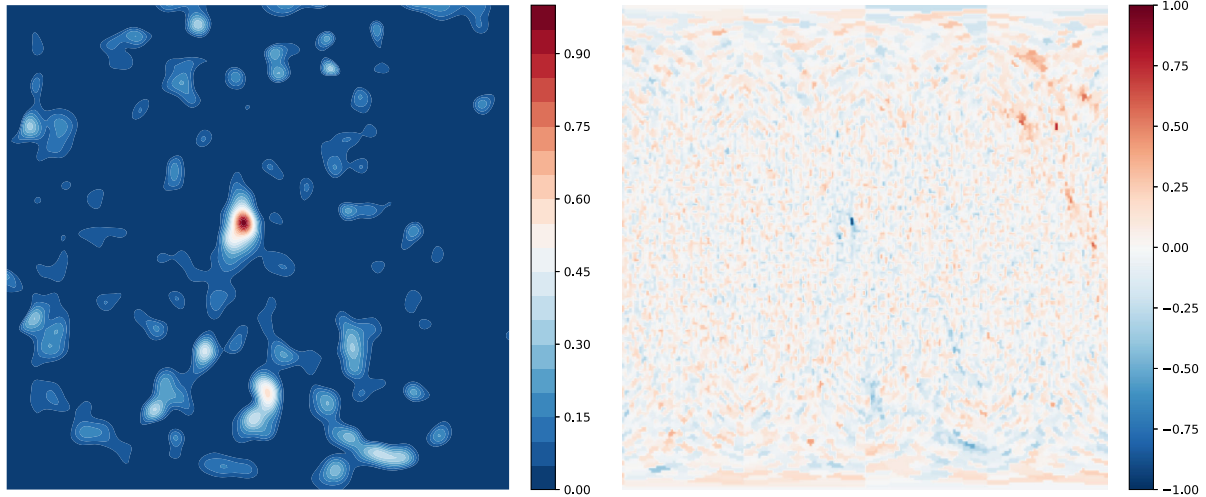


FIG. 7. Example assignment map for *positive* attributions (left), which has been convolved with a Gaussian filter with $\sigma = 3$ pixels, calculated from a simulated lensing map with substructure from scenario D (right). The attribution score and astrometric lensing velocities have been normalized for convenience. The power of axiomatic attribution is now manifest in this example where it demonstrates that high, positive attribution scores can be useful in localization of substructure.

and thus reliably probe the substructure content of the Milky Way.

With several classifiers in hand, it is possible to peer into the black box using integrated gradients and extract information beyond our classification label. In Fig. 7, we show the assignment map for positive attributions calculated from an example lensing map with substructure from scenario D. Since positive attributions correspond to pixels that voted for the substructure class, in principle they can be used to localize substructure in a lensing map.

Figure 7 demonstrates the application of integrated gradients by showing that high attributions are correlated with substructure locations. Not surprisingly, it is prohibitively difficult to discern such features for data sets with lower AUC values as the usefulness of an attribution map is correlated with architecture performance.

We would like to point out that this method for substructure localization supplements other techniques that could be applied like image segmentation—for example, in the context of strong lensing, see Ref. [89]—but should prove to be computationally more tractable since training is restricted to classification. As a final note, beyond the information contained in positive attributions, features in the *negative* attribution map should prove critical in understanding various noise and backgrounds in the data set in addition to improvements in the quality of data that may come by longer observation times.

V. CONCLUSION

We investigate the possibility of using supervised deep learning techniques to detect the astrometric lensing effect on a population of quasars from dark matter subhalos in the Milky Way as a proof of principle using simulated data. In particular, we apply state-of-the-art convolutional neural network RESNET-18 to our data set, and we show how a CNN could be used to detect the astrometric lensing signal from future astrometric surveys provided an architecture trained on simulations were to be applied to a real data set.

We observe that for the current best velocity sensitivity estimates for a 10 year run of an SKA-like survey $\sigma_u \sim 1 \mu\text{as/yr}$ a total number of $N_q \sim 3 \times 10^9$ quasars would be required for the classifier to detect the signal with an AUC value > 0.8 . Much better performance with an AUC value of ~ 0.96 is achieved for $N_q \sim 10^8$ but with a much smaller velocity uncertainty of $\sigma_u \sim 0.1 \mu\text{as/yr}$. We see that a smaller uncertainty has much higher impact on the ability of the classifier to detect the dark matter substructure and it

can be achieved by expanding the duration of the survey to more than a decade. Our results are in qualitative agreement with the results of Ref. [34], though we cannot quantitatively compare since the AUC values used here do not directly map to the discovery significance shown in Fig. 6 of that work.

We have further shown a method for localizing subhalos using integrated gradients, a method of axiomatic attribution. Concretely, we find that the map for positive attributions, pixels which voted for the substructure class, are found to be consistent with the location of known strong lenses in our simulations. While we do not investigate in detail here, the negative attributions are likely to encode crucial information about various noise and backgrounds present in the data set.

While it may be possible that there is an architecture that could be better suited to tackle the complexity of the data set at the level of expectations for SKA, it should be pointed out that the results presented here are based upon an assumed dark matter model defined by the mass and distribution of its substructure. It may be the case that dark matter is something other than the CDM picture presented here, and thus prospects for detection may differ. Similar to the spirit of Ref. [90], in which anomaly detection, a form of *unsupervised* machine learning, was used to detect the presence of dark matter substructure in strong lensing images, it seems natural to extend this work to the unsupervised scenario where an autoencoder is trained on sets of images corresponding to SKA measurements with no substructure and qualifying the constraining power of identifying any DM substructure in SKA data. In this way, one is taking a *theory agnostic* approach to the detection of dark matter—i.e., our classifier performance will not depend on an arbitrary choice of model and its various parameters. Additionally, the inclusion of acceleration data on top of the velocity maps can further strengthen the detection of the signal. We leave this for future work.

ACKNOWLEDGMENTS

We acknowledge useful conversations with Stephon Alexander, Jatan Buch, and John Leung. This research was conducted using computational resources and services at the Center for Computation and Visualization, at Brown University. This work was partially supported by NSF Grant No. PHY-2014052. We gratefully acknowledge the support of Brown University.

- [1] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A6 (2020).
- [2] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, S. Allam, F. Andrade-Oliveira, J. Annis, S. Avila, M. Banerji, N. Banik, K. Bechtol *et al.*, *Mon. Not. R. Astron. Soc.* **483**, 4866 (2019).
- [3] H. Gil-Marín, J. Guy, P. Zarrouk, E. Burtin, C.-H. Chuang, W. J. Percival, A. J. Ross, R. Ruggeri, R. Tojerio, G.-B. Zhao *et al.*, *Mon. Not. R. Astron. Soc.* **477**, 1604 (2018).
- [4] E. Corbelli and P. Salucci, *Mon. Not. R. Astron. Soc.* **311**, 441 (2000).
- [5] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradač, *Astrophys. J.* **679**, 1173 (2008).
- [6] A. K. Drukier, K. Freese, and D. N. Spergel, *Phys. Rev. D* **33**, 3495 (1986).
- [7] M. W. Goodman and E. Witten, *Phys. Rev. D* **31**, 3059 (1985).
- [8] D. S. Akerib *et al.* (LUX Collaboration), *Phys. Rev. Lett.* **118**, 021303 (2017).
- [9] X. Cui *et al.* (PandaX-II Collaboration), *Phys. Rev. Lett.* **119**, 181302 (2017).
- [10] E. Aprile *et al.* (XENON Collaboration), *Phys. Rev. Lett.* **121**, 111302 (2018).
- [11] F. Froberg and A. R. Duffy, *J. Phys. G* **47**, 094002 (2020).
- [12] Fermi LAT Collaboration, *J. Cosmol. Astropart. Phys.* **09** (2015) 008.
- [13] A. Geringer-Sameth, S. M. Koushiappas, and M. G. Walker, *Phys. Rev. D* **91**, 083535 (2015).
- [14] A. Albert, R. Alfaro, C. Alvarez, J. D. Álvarez, R. Arceo, J. C. Arteaga-Velázquez, D. Avila Rojas, H. A. Ayala Solares, N. Bautista-Elivar, A. Becerril *et al.*, *Astrophys. J.* **853**, 154 (2018).
- [15] S. Archambault, A. Archer, W. Benbow, R. Bird, E. Bourbeau, T. Brantseg, M. Buchovecky, J. H. Buckley, V. Bugaev, K. Byrum *et al.*, *Phys. Rev. D* **95**, 082001 (2017).
- [16] J. Rico, *Galaxies* **8**, 25 (2020).
- [17] MAGIC Collaboration, M. L. Ahnen, S. Ansoldi, L. A. Antonelli, P. Antoranz, A. Babic, B. Banerjee, P. Bangale, U. Barres de Almeida, J. A. Barrio *et al.*, *J. Cosmol. Astropart. Phys.* **02** (2016) 039.
- [18] IceCube Collaboration, M. G. Aartsen, M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers, M. Ahrens, I. A. Samarai, D. Altmann, K. Andeen *et al.*, *Eur. Phys. J. C* **77**, 627 (2017).
- [19] K. Choi, K. Abe, Y. Haga, Y. Hayato, K. Iyogi, J. Kameda, Y. Kishimoto, M. Miura *et al.* (The Super-Kamiokande Collaboration), *Phys. Rev. Lett.* **114**, 141301 (2015).
- [20] M. Aaboud *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **05** (2019) 142.
- [21] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **10** (2017) 073.
- [22] N. Du, N. Force, R. Khatiwada, E. Lentz, R. Ottens, L. J. Rosenberg, G. Rybka, G. Carosi, N. Woollett, D. Bowring *et al.*, *Phys. Rev. Lett.* **120**, 151301 (2018).
- [23] P. W. Graham, I. G. Irastorza, S. K. Lamoreaux, A. Lindner, and K. A. van Bibber, *Annu. Rev. Nucl. Part. Sci.* **65**, 485 (2015).
- [24] K. Kannike, M. Raidal, H. Veermäe, A. Strumia, and D. Teresi, *Phys. Rev. D* **102**, 095002 (2020).
- [25] J. Buch, M. A. Buen-Abad, J. Fan, and J. S. Chau Leung, *J. Cosmol. Astropart. Phys.* **10** (2020) 051.
- [26] W. H. W. Ngan and R. G. Carlberg, *Astrophys. J.* **788**, 181 (2014).
- [27] R. G. Carlberg, *Astrophys. J.* **820**, 45 (2016).
- [28] J. Bovy, *Phys. Rev. Lett.* **116**, 121301 (2016).
- [29] D. Erkal, V. Belokurov, J. Bovy, and J. L. Sanders, *Mon. Not. R. Astron. Soc.* **463**, 102 (2016).
- [30] R. E. Sanderson, C. Vera-Ciro, A. Helmi, and J. Heit, *arXiv:1608.05624*.
- [31] A. Bonaca, D. W. Hogg, A. M. Price-Whelan, and C. Conroy, *Astrophys. J.* **880**, 38 (2019).
- [32] R. Feldmann and D. Spolyar, *Mon. Not. R. Astron. Soc.* **446**, 1000 (2015).
- [33] K. Van Tilburg, A.-M. Taki, and N. Weiner, *J. Cosmol. Astropart. Phys.* **07** (2018) 041.
- [34] S. Mishra-Sharma, K. Van Tilburg, and N. Weiner, *Phys. Rev. D* **102**, 023026 (2020).
- [35] C. Mondino, A.-M. Taki, K. Van Tilburg, and N. Weiner, *Phys. Rev. Lett.* **125**, 111101 (2020).
- [36] J. A. Dror, H. Ramani, T. Trickle, and K. M. Zurek, *Phys. Rev. D* **100**, 023003 (2019).
- [37] M. Buschmann, J. Kopp, B. R. Safdi, and C.-L. Wu, *Phys. Rev. Lett.* **120**, 211101 (2018).
- [38] J. Chen and S. M. Koushiappas, *Astrophys. J.* **724**, 400 (2010).
- [39] J. Chen, S. M. Koushiappas, and A. R. Zentner, *Astrophys. J.* **741**, 117 (2011).
- [40] A. Diaz Rivero, F.-Y. Cyr-Racine, and C. Dvorkin, *Phys. Rev. D* **97**, 023001 (2018).
- [41] A. Díaz Rivero, C. Dvorkin, F.-Y. Cyr-Racine, J. Zavala, and M. Vogelsberger, *Phys. Rev. D* **98**, 103517 (2018).
- [42] S. Mao and P. Schneider, *Mon. Not. R. Astron. Soc.* **295**, 587 (1998).
- [43] J. W. Hsueh, L. Oldham, C. Spingola, S. Vegetti, C. D. Fassnacht, M. W. Auger, L. V. E. Koopmans, J. P. McKean, and D. J. Lagattuta, *Mon. Not. R. Astron. Soc.* **469**, 3713 (2017).
- [44] N. Dalal and C. S. Kochanek, *Astrophys. J.* **572**, 25 (2002).
- [45] Y. D. Hezaveh *et al.*, *Astrophys. J.* **823**, 37 (2016).
- [46] S. Vegetti and L. V. E. Koopmans, *Mon. Not. R. Astron. Soc.* **392**, 945 (2009).
- [47] L. V. E. Koopmans, *Mon. Not. R. Astron. Soc.* **363**, 1136 (2005).
- [48] S. Vegetti and L. V. E. Koopmans, *Mon. Not. R. Astron. Soc.* **400**, 1583 (2009).
- [49] S. Vegetti, L. V. E. Koopmans, A. Bolton, T. Treu, and R. Gavazzi, *Mon. Not. R. Astron. Soc.* **408**, 1969 (2010).
- [50] S. Vegetti, D. J. Lagattuta, J. P. McKean, M. W. Auger, C. D. Fassnacht, and L. V. E. Koopmans, *Nature (London)* **481**, 341 (2012).
- [51] S. Vegetti, L. V. E. Koopmans, M. W. Auger, T. Treu, and A. S. Bolton, *Mon. Not. R. Astron. Soc.* **442**, 2017 (2014).
- [52] E. Ritondale, S. Vegetti, G. Despali, M. Auger, L. Koopmans, and J. McKean, *Mon. Not. R. Astron. Soc.* **485**, 2179 (2019).
- [53] M. Ntampaka *et al.*, *arXiv:1902.10159*.

- [54] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [55] S. Pan, M. Liu, J. Forero-Romero, C. G. Sabiu, Z. Li, H. Miao, and X.-D. Li, *Sci. China Phys. Mech. Astron.* **63**, 110412 (2020).
- [56] A. Mishra, P. Reddy, and R. Nigam, *Mach. Learn.* **1**, 045012 (2020).
- [57] F. Farsian, N. Krachmalnicoff, and C. Baccigalupi, *J. Cosmol. Astropart. Phys.* **07** (2020) 017.
- [58] J. Caldeira, W. L. K. Wu, B. Nord, C. Avestruz, S. Trivedi, and K. T. Story, *Astron. Comput.* **28**, 100307 (2019).
- [59] S. Hassan, S. Andrianomena, and C. Doughty, *Mon. Not. R. Astron. Soc.* **494**, 5761 (2020).
- [60] D. Gruen, S. Seitz, J. Koppenhoefer, and A. Riffeser, *Astrophys. J.* **720**, 639 (2010).
- [61] G. Nurbæva, M. Tewes, F. Courbin, and G. Meylan, *Astron. Astrophys.* **577**, A104 (2015).
- [62] J. Schmelzle, A. Lucchi, T. Kacprzak, A. Amara, R. Sgier, A. Réfrégier, and T. Hofmann, [arXiv:1707.05167](https://arxiv.org/abs/1707.05167).
- [63] J. Fluri, T. Kacprzak, A. Lucchi, A. Refregier, A. Amara, T. Hofmann, and A. Schneider, *Phys. Rev. D* **100**, 063514 (2019).
- [64] Y. D. Hezaveh, L. Perreault Levasseur, and P. J. Marshall, *Nature (London)* **548**, 555 (2017).
- [65] L. Perreault Levasseur, Y. D. Hezaveh, and R. H. Wechsler, *Astrophys. J.* **850**, L7 (2017).
- [66] W. R. Morningstar, Y. D. Hezaveh, L. Perreault Levasseur, R. D. Blandford, P. J. Marshall, P. Putzky, and R. H. Wechsler, [arXiv:1808.00011](https://arxiv.org/abs/1808.00011).
- [67] W. R. Morningstar, L. Perreault Levasseur, Y. D. Hezaveh, R. Blandford, P. Marshall, P. Putzky, T. D. Rueter, R. Wechsler, and M. Welling, *Astrophys. J.* **883**, 14 (2019).
- [68] R. Canameras, S. Schuldt, S. H. Suyu, S. Taubenberger, T. Meinhardt, L. Leal-Taixe, C. Lemon, K. Rojas, and E. Savary, *Astron. Astrophys.* **644**, A163 (2020).
- [69] S. Varma, M. Fairbairn, and J. Figueroa, [arXiv:2005.05353](https://arxiv.org/abs/2005.05353).
- [70] S. Alexander, S. Gleyzer, E. McDonough, M. W. Toomey, and E. Usai, *Astrophys. J.* **893**, 15 (2020).
- [71] A. Diaz Rivero and C. Dvorkin, *Phys. Rev. D* **101**, 023515 (2020).
- [72] J. Brehmer, S. Mishra-Sharma, J. Hermans, G. Louppe, and K. Cranmer, *Astrophys. J.* **886**, 49 (2019).
- [73] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **462**, 563 (1996).
- [74] M. Hütten, C. Combet, G. Maier, and D. Maurin, *J. Cosmol. Astropart. Phys.* **09** (2016) 047.
- [75] P. Mollitor, E. Nezri, and R. Teyssier, *Mon. Not. R. Astron. Soc.* **447**, 1353 (2015).
- [76] T. Sawala, C. S. Frenk, A. Fattahi, J. F. Navarro, R. G. Bower, R. A. Crain, C. Dalla Vecchia, M. Furlong, J. C. Helly, A. Jenkins *et al.*, *Mon. Not. R. Astron. Soc.* **457**, 1931 (2016).
- [77] M. Roos, *J. Mod. Phys.* **03**, 1152 (2012).
- [78] K. Vattis and S. M. Koushiappas, *Phys. Rev. D* **97**, 103003 (2018).
- [79] F. J. Kerr and D. Lynden-Bell, *Mon. Not. R. Astron. Soc.* **221**, 1023 (1986).
- [80] T. Piffl, C. Scannapieco, J. Binney, M. Steinmetz, R. D. Scholz, M. E. K. Williams, R. S. de Jong, G. Kordopatis, G. Matijević, O. Bienaymé *et al.*, *Astron. Astrophys.* **562**, A91 (2014).
- [81] R. Schönrich, J. Binney, and W. Dehnen, *Mon. Not. R. Astron. Soc.* **403**, 1829 (2010).
- [82] M. A. Sánchez-Conde and F. Prada, *Mon. Not. R. Astron. Soc.* **442**, 2271 (2014).
- [83] E. Fomalont and M. Reid, *New Astron. Rev.* **48**, 1473 (2004).
- [84] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2016).
- [85] K. Simonyan and A. Zisserman, [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [86] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Commun. ACM* **60**, 84 (2017).
- [87] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, *Astron. Comput.* **27**, 130 (2019).
- [88] M. Sundararajan, A. Taly, and Q. Yan, [arXiv:1703.01365](https://arxiv.org/abs/1703.01365).
- [89] B. Ostdiek, A. Diaz Rivero, and C. Dvorkin, [arXiv:2009.06663](https://arxiv.org/abs/2009.06663).
- [90] S. Alexander, S. Gleyzer, H. Parul, P. Reddy, M. W. Toomey, E. Usai, and R. Von Klar, [arXiv:2008.12731](https://arxiv.org/abs/2008.12731).