The Astrophysical Journal, 893:15 (9pp), 2020 April 10 https://doi.org/10.3847/1538-4357/ab7925

Deep Learning the Morphology of Dark Matter Substructure

Stephon Alexander[1], Sergei Gleyzer[2], Evan McDonough[1], Michael W. Toomey[3], and Emanuele Usai[3]
[1] Brown Theoretical Physics Center and Department of Physics, Brown University, Providence, RI, USA
[2] Department of Physics and Astronomy, University of Alabama, Tuscaloosa, AL, USA
[3] Department of Physics, Brown University, Providence, RI, USA; michael_toomey@brown.edu
Received 2019 November 17; revised 2020 February 18; accepted 2020 February 21; published 2020 April 8

Abstract

Strong gravitational lensing is a promising probe of the substructure of dark matter halos. Deep-learning methods have the potential to accurately identify images containing substructure, and differentiate weakly interacting massive particle dark matter from other well motivated models, including vortex substructure of dark matter condensates and superfluids. This is crucial in future efforts to identify the true nature of dark matter. We implement, for the first time, a classification approach to identifying dark matter based on simulated strong lensing images with different substructure. Utilizing convolutional neural networks trained on sets of simulated images, we demonstrate the feasibility of deep neural networks to reliably distinguish among different types of dark matter substructure. With thousands of strong lensing images anticipated with the coming launch of Vera C. Rubin Observatory, we expect that supervised and unsupervised deep-learning models will play a crucial role in determining the nature of dark matter.

Unified Astronomy Thesaurus concepts: Dark matter (353); Strong gravitational lensing (1643); Convolutional neural networks (1938)

1. Introduction

The canonical candidate for dark matter is a weakly interacting massive particle (WIMP). Indeed, extensions of the standard model (SM) generally include WIMPs of mass 100 GeV that accurately reproduce the observed dark matter density; realizing what is known as the WIMP miracle. However, WIMPS have thus far evaded detection, both by direct detection (Goodman & Witten 1985; Drukier et al. 1986; Akerib et al. 2017; Cui et al. 2017; Aprile et al. 2018) and colliders (e.g., Aaboud et al. 2019). There are also hints at cracks in the WIMP paradigm, for example, the core versus cusp problem: observations of halos have consistently shown that actual dark matter halos lack cusps (Burkert 1996) like that of the Navarro–Frenk–White (NFW) profile found from simulation (Navarro et al. 1996). This motivates the consideration of alternatives to the WIMP paradigm.

An interesting possibility is condensate models of dark matter, both Bose–Einstein (BEC; Sin 1994; Hu et al. 2000; Silverman & Mallett 2002; Sikivie & Yang 2009; Berezhiani & Khoury 2015; Hui et al. 2017; Ferreira et al. 2019) and Bardeen–Cooper–Schreifer (BCS; Alexander & Cormack 2017; Alexander et al. 2018). These build on the decades-long study of axion dark matter (Abbott & Sikivie 1983; Dine & Fischler 1983; Preskill et al. 1983) and the realization that axions, arising as the Goldstone boson of a spontaneously broken global U(1) symmetry, are the field theory definition of superfluidity (Schmitt 2015). In these models, dark matter is a quasi-particle excitation of the fundamental degrees of freedom that comprise the condensate. For a specific choice of the effective field theory of the superfluid, this reproduces the baryonic Tully–Fisher relation (Berezhiani & Khoury 2015, 2016).

These condensate models have the interesting property that they can form vortices (Rindler-Daller & Shapiro 2012), line-like defects that are a nonrelativistic analog to cosmic strings (Brandenberger 1994, 2014). If they exist, vortices constitute a substructure component for dark matter halos. The detection of vortices would be a smoking gun for superfluid dark matter (SFDM). We are thus lead to discriminate between different models of dark matter by probing substructure in halos.

A promising method to detect substructure is from strong gravitational lensing images. Observations of strongly lensed quasars have been used to infer the presence of substructure (Mao & Schneider 1998; Dalal & Kochanek 2002; Hsueh et al. 2017). Additionally, high resolution images with the Atacama Large Millimeter/submillimeter Array have inferred the presence of subgalactic structure (Hezaveh et al. 2016). Extended lensing images, in particular, can serve as a very sensitive probe (Koopmans 2005; Vegetti & Koopmans 2009a, 2009b). Given strong lensing has already proven to be a powerful probe of dark matter substructure, it is logical to extend this to distinguishing between the different types that may be characteristic of specific dark matter models.

Bayesian likelihood analyses can be implemented to determine if a given dark matter model is consistent with a set of lensing images. Indeed, such analyses have been conducted searching for particle dark matter substructure (Vegetti et al. 2010; Daylan et al. 2018). However, Bayesian analysis of strong lensing images is notoriously computationally expensive, and machine-learning methods have been demonstrated to achieve a speed-up of several orders of magnitude (Hezaveh et al. 2017). In this work we take the latter approach, and with condensate models of dark matter in mind, implement a deep-learning algorithm to identify specific types of dark matter in simulated lensing images; that is, we consider the search for substructure as a classification problem.

Applications of machine- and deep-learning methods are abound in cosmology (Ntampaka et al. 2019) and the physical sciences more broadly (Carleo et al. 2019). In particular, this approach has been applied to strong gravitational lensing (Hezaveh et al. 2017; Perreault Levasseur et al. 2017; Morningstar et al. 2018, 2019), and most recently, to the study of particle dark matter subhalos (Brehmer et al. 2019).

The treatment of substructure searches as a classification problem compliments the existing approaches of statistical detection (Cyr-Racine et al. 2016, 2019; Brennan et al. 2019; Diaz Rivero et al. 2018; Díaz Rivero et al. 2018) and identification of individual substructures (e.g., Hezaveh et al. 2016). This work can be interpreted as an intermediate step before the latter: we train and implement a convolutional neural network (CNN) to distinguish different classes of substructure in lensing images that can then be further processed to find the position, mass, and other properties, of individual substructures.

The overarching goal of this work is to undertake a theory-agnostic approach to dark matter searches. As a first step, we first present the results of an implementation of a supervised neural network to distinguish between two different types of dark matter. Given the vast number of models and considerable theoretical uncertainty on the nature of dark matter, it would then be advantageous to implement an unsupervised machine-learning algorithm to identify various potential dark matter signals in the strong lensing images.

The structure of this paper is as follows: in Section 2 we review dark matter substructure, and in Section 3 we consider as a prototypical example the substructure of SFDM. We construct simulated lensing images in Section 4 and in Section 5 a neural network to analyze them. We present our results in Section 6, and discuss the implications for detection in Section 7. We close with a discussion of future work in Section 8.

2. Dark Matter Substructure and Strong Gravitational Lensing

The ΛCDM paradigm predicts that density fluctuations present in the the early universe evolve to become the large-scale structure of the universe via hierarchical structure formation. This model envisions small halos merge together forming larger and larger structures leading to the DM halos that we see today (Kauffmann et al. 1993). It is also expected from simulation that subhalos can avoid significant tidal disruption and remain largely intact. On large scales ΛCDM is consistent with the cosmic microwave background, galaxy clustering, and weak lensing (Heymans et al. 2012; Anderson et al. 2014; Planck Collaboration 2016). However, on smaller scales the verdict is less clear. A classic example is the missing satellites problem (Bullock & Boylan-Kolch 2017; though see Kim et al. 2018 for a differing take). Indeed, different types of particle dark matter can have vastly different substructure on subgalactic scales. For example, the greater streaming length of WDM (Bode et al. 2001; Abazajian 2006) and emergent properties of self-interacting dark matter (Spergel & Steinhardt 2000) can prevent the formation of small-scale substructure. Thus, while large-scale structure for different types of DM can appear identical, careful attention to structure on subgalactic scales can be a powerful tool to distinguish DM models.

A powerful probe of the gravitationally bound structures of dark matter is strong gravitational lensing. Given a matter over or under density, the deflection angle along the line of sight is given by an integral over the induced gravitational potential (Narayan & Bartelmann 1997),

$$\boldsymbol{\alpha} = \frac{2}{c^2}\boldsymbol{\nabla}_\theta \int d\chi \frac{\chi_s - \chi}{\chi\chi_s}\Psi(\boldsymbol{r}), \tag{1}$$

where χ is the distance along the line of sight, χ_s is the distance to the source, and $\Psi(\boldsymbol{r})$ is the gravitational potential. In the thin lens approximation (Narayan & Bartelmann 1997), valid in the limit that the thickness of the lensing galaxy is small compared to the distance to the lens, this takes a simplified form,

$$\boldsymbol{\alpha} = \frac{2}{c^2}\frac{D_{\rm LS}}{D_S D_L}\boldsymbol{\nabla}\int dz \Psi(\boldsymbol{r}), \tag{2}$$

where $D_{\rm LS}$, D_L, and D_S are the angular diameter distances from the lens to the source, from the observer to the lens, and from the observer to the source, respectively, and z is the distance along the line of sight. From this expression one can straightforwardly compute the lensing due to any gravitational potential Ψ.

The gravitational potential is in turn determined by matter density via the Poisson equation, $\nabla^2\Psi \propto \rho$. The linearity of this equation implies that the total lensing due to the separate contributions, e.g., of a halo and halo substructure, is simply the sum of the individual contributions. That is,

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\rm LSS} + \boldsymbol{\alpha}_{\rm halo} + \boldsymbol{\alpha}_{\rm halo-sub}, \tag{3}$$

where $\boldsymbol{\alpha}_{\rm LSS}$ is the external shear due to large-scale structure, and $\boldsymbol{\alpha}_{\rm halo,halo-sub}$ are the lensing due to the halo and halo substructure respectively.

The well studied case is the lensing due to the spherical substructures expected from hierarchical structure formation in the context of noninteracting particle dark matter. However, as mentioned in the 1, other types of substructure can exist in models of dark matter outside the WIMP paradigm. As a prototypical example, we will consider dark matter condensates, namely superfluids, which exhibit substructure in the form of vortices. We now proceed to develop this in detail.

3. Case Study: Dark Matter Superfluidity

The canonical example of a condensate dark matter model is axion dark matter. Axions were introduced as a solution to the strong-CP problem of the SM (Peccei & Quinn 1977; Weinberg 1978; Wilczek 1978), and soon thereafter proposed as a dark matter candidate (Abbott & Sikivie 1983; Dine & Fischler 1983; Preskill et al. 1983). It was later argued that axions could form a Bose–Einstein condensate and exhibit superfluidity (Sin 1994; Hu et al. 2000; Silverman & Mallett 2002; Hui et al. 2017; Sikivie & Yang 2009; Berezhiani & Khoury 2015; Ferreira et al. 2019).

To emphasize the superfluid nature of axions, we can rewrite the field equations in terms of fluid equations. Being comprised of extremely light particles at an incredibly high number density, axion dark matter is well described by a coherent scalar field. Moreover, because dark matter as we observe it is cold, the system is well described by a nonrelativistic limit. The Euler and continuity equations of classical fluid mechanics emerge in this nonrelativistic limit, defined via the decomposition

$$\varphi(x, t) = \sqrt{\frac{\hbar^3 c}{2m}}\left(\phi(x, t)e^{-imc^2 t/\hbar} + c.c.\right) \tag{4}$$

and the limit $|\ddot{\phi}| \ll mc^2|\dot{\phi}|/\hbar$ (Hui et al. 2017). If we now define the fluid density ρ and velocity $\boldsymbol{v}$ by

$$\phi \equiv \sqrt{\frac{\rho}{m}}e^{i\theta}, \quad v \equiv \frac{\hbar}{R\,m}\nabla\theta = \frac{\hbar}{2miR}\left(\frac{1}{\phi}\nabla\phi - \frac{1}{\phi^*}\nabla\phi^*\right) \tag{5}$$

then the nonrelativistic limit of the Klein–Gordon equation becomes

$$\begin{aligned}&\dot{\rho} + 3H\rho + \frac{1}{R}\nabla\cdot(\rho\boldsymbol{v}) = 0,\\ &\dot{\boldsymbol{v}} + H\boldsymbol{v} + \frac{1}{R}(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\frac{1}{R}\nabla\Phi + \frac{\hbar^2}{2R^3m^2}\nabla\left(\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}}\right).\end{aligned} \tag{6}$$

These are the Madelung equations in an expanding universe, which are the continuity and Euler equations of fluid mechanics, with the addition of the second term on the right of the lower equation, referred to as the quantum pressure.

Parallel to the development of axions has been the study of condensate phases of non-Abelian gauge theories such as the SM's Quantum Chromo-Dynamics (QCD). Starting from the realization that neutrons can undergo a BCS transition to a superfluid in the interior of neutron stars (Baym et al. 1969), it was found that at high enough densities, the quarks themselves could form Cooper pairs and undergo a BCS transition to a superfluid or superconducting state (Alford et al. 1998; for a review see, e.g., Alford et al. 2008). These developments have spurred on the study of neutron star physics (for reviews, see Lombardo & Schulze 2001; Dean & Hjorth-Jensen 2003; Page et al. 2013; Haskell & Sedrakian 2018), which with the observation of gravitational waves from a neutron star binary merger (Abbott et al. 2017), may be on the cusp of a breakthrough.

Bringing together these disparate developments, it was shown in Alexander et al. (2018) that a QCD-like theory could lead to superfluidity on cosmological scales, and constitute a scenario for SFDM, providing a BCS analog to the axion's BEC. The natural embedding of this scenario in inflationary cosmology, with the fundamental degrees of freedom produced in huge numbers as a side effect of baryogenesis, leads to a scenario of dark matter with observables from all stages of the evolution of the universe.

What these variant scenarios of SFDM have in common is the existence of vortices. Let us now see explicitly how vortex structures come out of superfluid halos. As per Equation (4), we can model SFDM with a macroscopic complex scalar function $\phi(r)$ that is described by the Lagrangian,

$$L = \int d^3x\left(\frac{\hbar^2}{2m}|\nabla\phi|^2 + \frac{1}{4}\lambda|\phi|^4 + \frac{1}{2}m\Phi|\phi|^2\right). \tag{7}$$

The first term on the r.h.s of Equation (7) is the kinetic term, the second an effective interaction potential with coupling strength $\lambda = 4\pi\hbar^2 a/m$ where a is the s-wave scattering length with a cross section $\sigma = 8\pi a^2$, and the last term is the coupling to the Newtonian gravitational potential Φ. This system is completely described by the time-dependent Gross–Pitaevski equation and the Poisson equation,

$$i\hbar\dot{\phi} = \left(-\frac{\hbar^2}{2m}\nabla^2 + m\Phi - \frac{\lambda}{m^2}|\phi|^2\right)\phi, \tag{8}$$

$$\nabla^2\Phi = 4\pi Gm|\phi|^2, \tag{9}$$

where ∇^2 is the spatial Laplacian and we take $c = 1$. The wave function ϕ can be written in terms of the modulus and phase,

$$\phi(r, t) = |\phi|(r, t)e^{iS(r,t)}, \tag{10}$$

in terms of which the equation of motion becomes two equations,

$$\begin{aligned}&-\frac{2m}{\hbar}|\phi|\frac{\partial S}{\partial t} + \Delta|\phi| - |\phi|(\nabla S)^2 = \frac{2m}{\hbar^2}(m\Phi + g|\phi|^2)|\phi|,\\ &\frac{\partial|\phi|^2}{\partial t} + \nabla\cdot\left[|\phi|^2\frac{\hbar}{m}\nabla S\right] = 0.\end{aligned} \tag{11}$$

These equations form a system of quantum–mechanical hydrodynamic equations with a bulk velocity $v = \hbar/m\nabla S$.

Note that the curl of the bulk velocity is null. This seems to imply that an SFDM halo would not have angular momentum. However, this would not be the case in reality. Thus the bulk velocity must contain a singularity; this is the vortex. Note that an integral over a closed contour around such a vortex would be nonzero

$$\oint_C \nabla S\cdot dl = 2\pi d\frac{\hbar}{m}. \tag{12}$$

Here d is an integer called the winding number. This implies that angular momentum of the vortex is quantized.

Solutions describing vortices in dark matter halos were found in Rindler-Daller & Shapiro (2012). The vortex solution is characterized by a density profile that can be parameterized as Rindler-Daller & Shapiro (2012),

$$\rho_v(r, z) = \begin{cases}0, & r > r_v\\ \rho_{v0}\left[\left(\frac{r}{r_v}\right)^{\alpha_v} - 1\right], & r \leqslant r_v\end{cases},$$

where r is the radial coordinate in cylindrical coordinates, r_v is the core-radius of the vortex solution, and α_v is a scaling exponent. This effectively models the vortex as a tube. On distance scales much larger than r_v, the vortex can be approximated as a line of density ρ_{v0}.

The values of these parameters, most importantly the density and total mass of the vortices, as well as the expected number density in realistic dark matter halos, varies widely across the literature. For example, the total amount of vortices in halos ranges from 340 vortices in the M31 halo with assumed constituent particle mass $m = 10^{-23}$ eV (Silverman & Mallett 2002) to $N = 10^{23}$ vortices in a typical DM halo for $m = 1$ eV (Berezhiani & Khoury 2015). In Banik & Sikivie (2013) it is shown that vortices can have mutual attraction and, over time, coalesce into a single, more massive vortex.

We also note that other substructure exists in superfluid scenarios, such as the recently found thin-disk solutions (Alexander et al. 2019). These have a lensing signal which interpolates between that of a vortex and a spherical halo, depending on the orientation of the disk.

Finally, we note the relation of vortices to cosmic strings. The latter is often explained as the relativistic analog to vortices

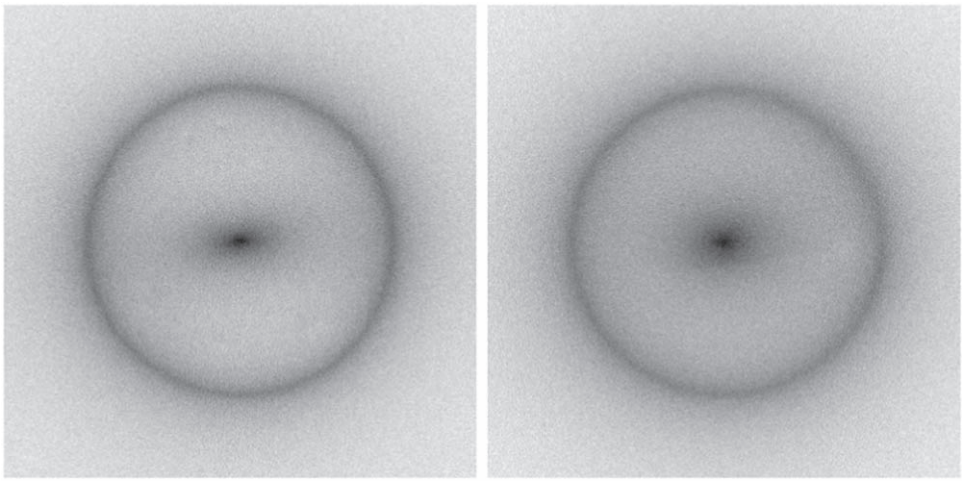

Figure 1. Lens images for particle substructure (left) and superfluid substructure (right). Simulated with *PyAutoLens*.

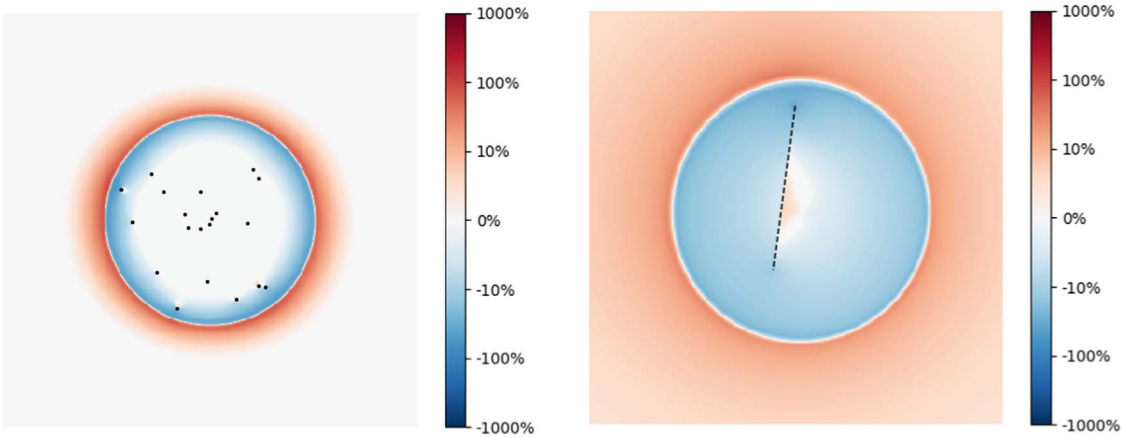

Figure 2. Residuals image for particle substructure (left) and superfluid-like substructure (right). Both are 1% of the halo mass. The black dots represent positions of subhalos and the dashed line represents the position of a vortex.

(see, e.g., Brandenberger 1994, 2014), formed during a phase transition in a relativistic quantum field theory. As such, cosmic stings have a transverse velocity that is close to the speed of light. In spite of this, much of the work on strong lensing by cosmic strings (Sazhin et al. 2007; Gasparini et al. 2008; Morganson et al. 2010) has approximated them as stationary or nonrelativistic, and hence effectively behaving as vortices. While we do not use these results directly, the lensing images generated in the following section agree with the results obtained in the cosmic string literature.

4. Strong Lensing Images

At this moment strong lensing data is limited to a handful of images. However, the upcoming completion of the Vera C. Rubin Observatory (VRO), formally the Large Synoptic Survey Telescope , will lead to thousands of strong lensing images that can be analyzed (Verma et al. 2019). In this work we have chosen to simulate our lensing images using the package *PyAutoLens* (Nightingale & Dye 2015; Nightingale et al. 2018). Written in Python, it can produce a variety of simulated strong lensing images where the user can adjust, among many possibilities, the mass of the halo, include substructure, light profiles, and mass profiles.

In addition to the simulation of the lensing itself, we also consider the addition of noise and the modifications induced by a point-spread function (PSF) on our observation. Thus, we can vary the level of noise in our images and include a PSF that is in line with real world instruments like *Hubble* or the future VRO, in this case subarcsecond resolution. Following Daylan et al. (2018), we approximate the PSF as an Airy disk whose first zero-crossing occurs at a radius of $\sigma_{\rm psf} \lesssim 1''$. This approximation is valid when noise is dominated by diffraction, which we assume to be the case.

The lensing image due to a single vortex embedded in a halo, with the vortex mass 1% that of the halo, and that of spherical substructure, as studied in Daylan et al. (2018), are shown in Figure 1. To quantify the effect of the substructure we subtract from each image the lensing image due the halo alone, and show the result (the "residuals") in Figure 2 for vortex and spherical substructure. From these images one can appreciate that the difference in lensing is primarily in the the morphology of the signal, making this an ideal task for classification with a CNN.

Lensing images were generated with parameters and their distributions given in Table 1. We have included the light from the lensing galaxy and nonnegligible backgrounds and noise. We have also accounted for other instrumental effects like the PSF, which we have modeled after the expected resolution of VRO, as well as shear effects. Example images can be found for each class in Figure 3.

Table 1
Parameters with Distributions and Priors Used in the Simulation of Strong Lensing Images

Lensing Galaxy—Sérsic Light Profile			
Parameter	Distribution	Priors	Details
θ_x	Fixed	0	x position
θ_y	Fixed	0	y position
z	Fixed |uniform	0.5 |[0.4, 0.6]	Redshift
e	Uniform	[0.5, 1.0]	Axis ratio
ϕ	Uniform	$[0, 2\pi]$	Orientation relative to y axis
I	Fixed	1.2	Intensity of emission (arbitrary units)
n	Fixed	2.5	Sérsic index
R	Fixed |uniform	0.5 |[0.5, 2]	Effective radius
Dark Matter Halo—Spherical Isothermal			
Parameter	Distribution	Priors	Details
θ_x	Fixed	0	x position
θ_y	Fixed	0	y position
θ_E	Fixed	1.2	Einstein radius
External Shear			
Parameter	Distribution	Priors	Details
γ_{ext}	Uniform	[0.0, 0.3]	Magnitude
ϕ_{ext}	Uniform	$[0, 2\pi]$	Angle
Lensed Galaxy—Sérsic Profile			
Parameter	Distribution	Priors	Details
r	Uniform	[0, 1.2]	Radial distance from center
ϕ_{bk}	Uniform	$[0, 2\pi]$	Angular position of galaxy from y-axis
z	Fixed |uniform	1.0 |[0.8, 1.2]	Redshift
e	Uniform	[0.7, 1.0]	Axis ratio
ϕ	Uniform	$[0, 2\pi]$	Orientation relative to y-axis
I	Uniform	[0.7, 0.9]	Intensity of emission (arbitrary units)
n	Fixed	1.5	Sérsic index
R	Fixed	0.5	Effective radius
Vortex			
Parameter	Distribution	Priors	Details
θ_x	Fixed |normal	0 |[0.0, 0.5]	x position
θ_y	Fixed |normal	0 |[0.0, 0.5]	y position
l	Fixed |uniform	1.0 |[0.5, 2.0]	Length of vortex
ϕ_v	Uniform	$[0, 2\pi]$	Orientation from y-axis
m_{vort}	Fixed	0.01 M_{Halo}	Total mass of vortex
Spherical			
Parameter	Distribution	Priors	Details
r	Uniform	[0, 1.0]	Radial distance from center
ϕ_{sph}	Uniform	$[0, 2\pi]$	Angular position of substructure from y-axis
N	Fixed |poisson	25 |μ=25	Number of substructures
m_{sub}	Fixed	0.01 M_{Halo}	Total mass of subhalos

Note. Where two values are given, the first corresponds to Section 6.1. and the second corresponds to Section 6.2. Note that only a single type of substructure was used per image.

5. Network and Training

In this work we take a supervised approach to establish a set of performance benchmarks for identifying different types of dark matter substructure. We, therefore, simulate the expected lensing effects from a variety of substructures and train a CNN to distinguish among them.

As the total mass constrained in the substructure is likely a small fraction of the total lensing mass, it may prove challenging to distinguish dark matter with traditional methods.

The addition of noise and other astrophysical backgrounds makes this an even more challenging task. Deep-learning methods are more amenable to identifying subtle morphologies in images. This is the approach we take in this work.

CNN are the natural choice for working with images. There are several pretrained networks openly available, e.g., *ResNet* (or *residual neural network;* He et al. 2015), *AlexNet* (Krizhevsky et al. 2017), *DenseNet* (Huang et al. 2016), and *VGG* (Simonyan & Zisserman 2014). For *ResNet*, the defining feature is that residual networks can skip layers all together in training. This, in practice, helps speed up the learning rate of the network by allowing the network to train fewer layers in the initial stages of learning. For this reason we will focus on *ResNet*, and return to a detailed algorithm comparison in Section 6.

During training we make use of data augmentation (see, e.g., Krizhevsky et al. 2017) via translation and rotations up to 90°. These all constitute invariant transformations with respect to the underlying substructure that allow the network to learn the actual structure in images. Thus data augmentation aims to increase our learning efficiency by seeing the same image in new ways several times.

We utilize 150,000 training and 15,000 validation images. The binary cross–entropy loss was minimized with the Adam optimizer in batches of 200 over a total of at most 20 epochs. The learning rate starts with a value of 1×10^{-4} and is reduced by a factor of 10 when the validation loss is not improved for three consecutive epochs. The networks were implemented using the PyTorch package and run on a single NVIDIA Titan K80 GPU. We use the well-established area under the ROC curve (AUC) as a metric for classifier performance

6. Results

As discussed in previous sections, we are interested in identifying and classifying substructure in strong gravitational lensing images. We do this with a supervised CNN, which requires that the classes be identified from the outset. In addition to the vortex and spherical subhalo substructure classes discussed in Section 4, there remains the possibility that an image may not have any detectable substructure at all, e.g., if the Einstein radius of the substructure is predominantly smaller than the PSF of the detector. Given this, we introduce an additional class: no substructure present.

In what follows we will train a multiclass classifier to predict the three classes: vortex, spherical, and no-substructure. We train three additional binary classifiers to distinguish between the two most probable classes predicted by the multiclass classifier. Thus, we are training an initial, higher level classifier, our multiclass classifier, which specializes in determining the best two models for a given image. We can then pass the image to a binary classifier that specializes in distinguishing between

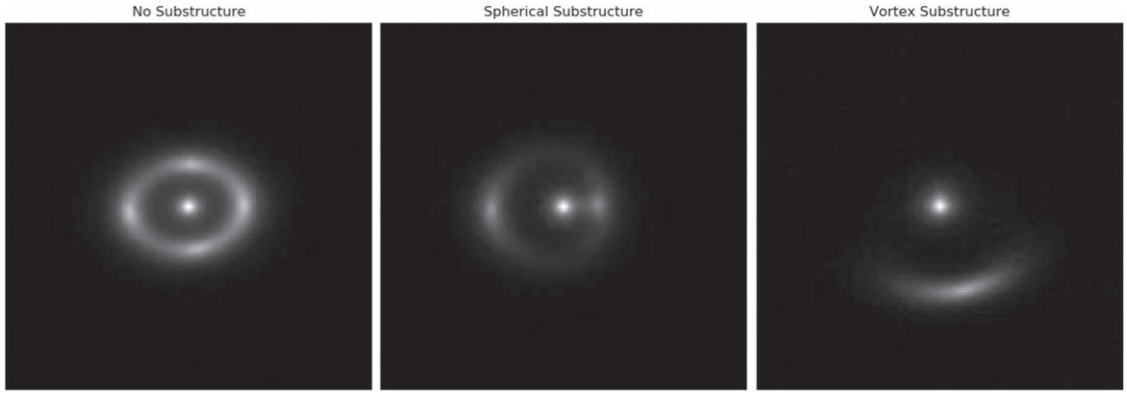

Figure 3. Example simulated images for all three classes.

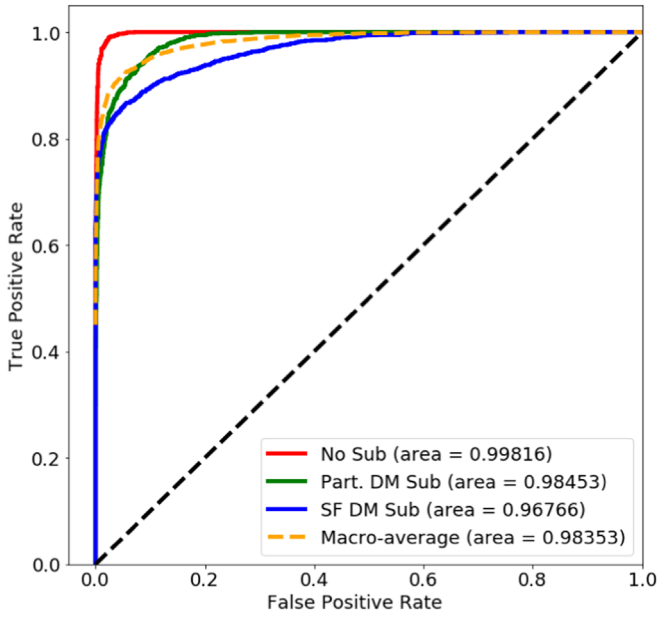

Figure 4. ROC curve for multiclass substructure classification with *ResNet*, as discussed in Section 6.1.

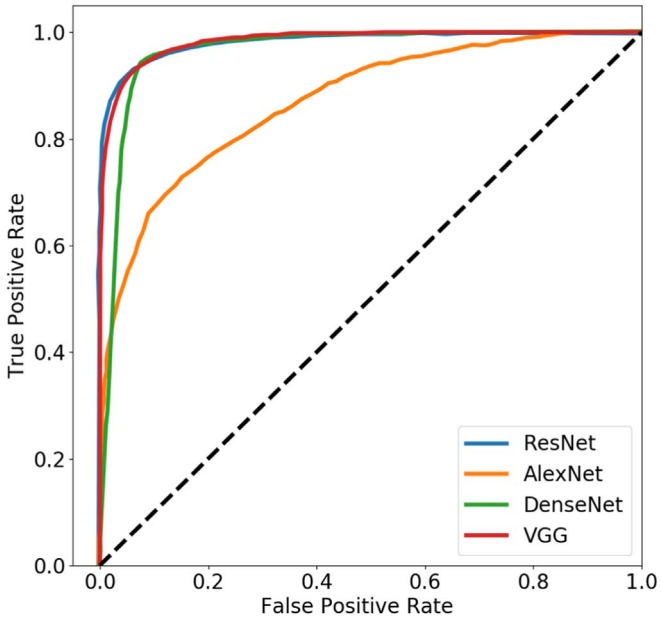

Figure 5. Comparison of architechures. Shown is the ROC curve averaged across substructure types for *Resnet*, *Alexnet*, *VGG*, and *DenseNet*.

the two most probable classes as chosen by the multiclass classifier. We do this using realistic mock lensing images as described in Section 4, with parameters and their distributions given in Table 1.

We start by considering an idealized population of physical systems, with, e.g., the distance to the lensed and lensing galaxy the same in each image. This could plausibly be the case if a very large data set was first divided into subsets exhibiting roughly constant properties. This is useful for comparison to Daylan et al. (2018), which held fixed the number of subhalos, and as a playground for performance tests of differing network architectures. We then consider a less-idealized population of images, with several additional physical properties of the halo and substructure allowed to vary. The parameters for each case are given in Table 1.

6.1. A Multiclass Classifier for Substructure Morphology

To begin, we consider similar parameters as images simulated in Daylan et al. (2018), which used probabilistic cataloging to identify spherical substructure. We differ with Daylan et al. (2018) by modeling the spherical substructure as point masses as opposed to truncated–NFW profiles, and of course by the inclusion of vortex-like substructure. Similar to Daylan et al. (2018), we have chosen a fixed value of 25 subhalos.

We use a pretrained network as described in Section 5. The ROC-AUC curves for substructure classification by *ResNet* are shown in Figure 4. As can be appreciated from the AUC scores of 0.998, 0.985, and 0.968, for images with no substructure, spherical subhalos, and vortices, respectively, our algorithm achieves excellent classification.

For the sake of comparison, we repeat this with differing choices of architecture, namely *VGG*, *DenseNet*, and *AlexNet*. The resulting ROC curves (averaged over substructure types) are shown in Figure 5. While *ResNet*, *VGG*, and *DenseNet* show comparable performance, *AlexNet* performs considerably worse. The other three architectures were otherwise indistinguishable, with the exception that, as expected, *ResNet* is more computationally efficient than both *VGG* and *DenseNet*.

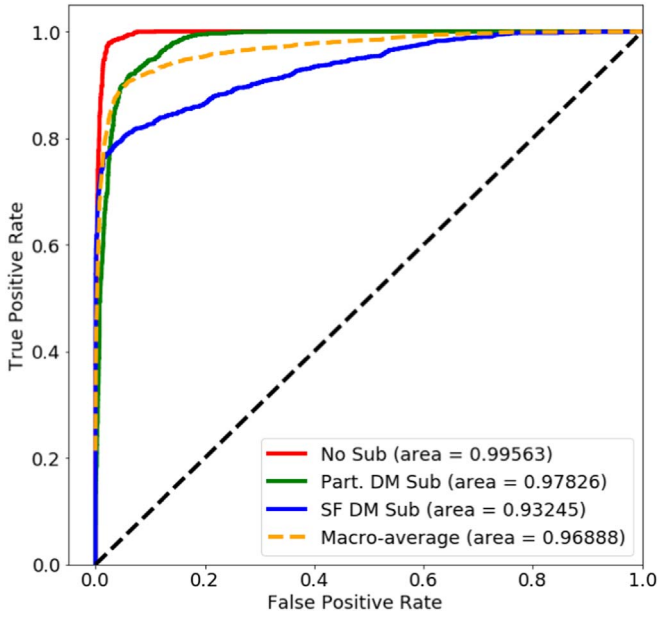

Figure 6. ROC curve for multiclass substructure classification with *ResNet*, including additional variations across the population of images, as discussed in Section 6.2.

6.2. Toward a Representative Population of Images

We now allow the training data to be a more diverse set of physical systems, as may be the case with actual data. We allow for variation in the distance to the lensing and lensed galaxy, the galaxy size, and importantly, the intensity of the background and noise, allowing the background to become nonnegligible. With regards to substructure, we vary the position of the vortex, and for spherical substructure consider the number of halos to be taken from a Poisson draw with mean 25. The details of all parameters and the specific distributions from which we draw values for simulation are all included in Table 1.

Utilizing the same *ResNet* architecture and running for 20 epochs while updating the learning rates, training our multi-class classifier obtains good results with a macro-averaged AUC of 0.969. The ROC curve is shown in Figure 6. Training of binary classifiers for 20 epochs results in near perfect classification with AUC scores near unity for distinguishing between images with and without substructure. The network also excels at distinguishing between different types of substructure achieving an AUC value 0.955. ROC curves for the binary classifiers are shown in Figure 7.

7. Toward the Detection of Substructure

To complete the analysis of this work, we establish the detection threshold for our network. To do so, we change the total mass of the substructure while holding all other parameters constant. Of course, in practice, it would be possible to train for more epochs, to use a deeper network, to add more training images, etc., and to push our threshold further. We implement this by simulating sets of 50,000 training and 5000 validation images at different total fractions of the halo mass for each class. We train each set on the same architecture, here *ResNet*, for 10 epochs. The metric we use to parameterize the ability of the network to learn is the AUC.

The AUC score for varying fraction of the halo mass contained in both types of substructure is shown in Figure 8. From this one can appreciate that the AUC rapidly deteriorates for a substructure mass below $10^{-2.5} \approx 0.3\%$ of the halo mass. From this we conclude that a CNN, given the fixed computing resources stated above, can reliably identify lensing images containing a substructure provided that its collective mass constitutes at least a fraction of a percent of the dark matter in the halo.

8. Discussion and Conclusion

It is well established that substructure can constrain dark matter models. In this work we have proposed it may even identify the nature of dark matter. Motivated by the significant theoretical uncertainty as to the nature of dark matter, in this work we have considered the study of substructure as a classification problem, and investigated the feasibility of using a machine-learning architecture to distinguish different types of substructure in strong lensing images.

Utilizing a simple supervised CNN, trained on simulated images, we have demonstrated that it is indeed feasible for a network to reliably distinguish among different types of dark matter substructure. This compliments existing approaches to substructure, namely the statistical detection (Cyr-Racine et al. 2016, 2019; Diaz Rivero et al. 2018; Díaz Rivero et al. 2018; Brennan et al. 2019) and the pinpointing of individual substructures (e.g., Hezaveh et al. 2016), and could be used as a part of a data analysis pipeline in the latter task.

A major caveat of this analysis has been the reliance on simulated lensing images. There are theoretical uncertainties in both the properties of individual substructures, and the population level statistics of host halos and their substructure. The latter leads to an uncertainty as to whether the training sample is an accurate representation of the underlying population. To overcome this difficulty, one may implement domain adaption, wherein the network is trained with the express goal of performing well on a target data set with different properties from the training set.

It is also interesting to consider future implementations of unsupervised models for identifying the morphology of the substructure. This would allow the analysis to be fully agnostic as to the true nature of dark matter, in recognition of the possibility that dark matter could be outside the scope of current theoretical expectations. One path in this direction is to implement an auto-encoder to learn the underlying substructure of real images. With thousands of galaxy–galaxy strong lensing images expected in the next few years (Verma et al. 2019), there should be ample data for training.

Finally, we note that deep learning may be amenable to searching for dark matter vortices in other observational windows, analogous to searches for cosmic strings in the cosmic microwave background (Ciuca et al. 2019; Ciuca & Hernández 2019) and 21 cm (Brandenberger et al. 2010; McDonough & Brandenberger 2013). We leave this, and the development of an unsupervised approach, to future work.

The authors thank Cora Dvorkin, Javad Hashemi, Shirley Ho, and David Spergel, for useful discussions. One of the authors thanks Robert Brandenberger for encouragement to work on this topic more than 20 years ago.

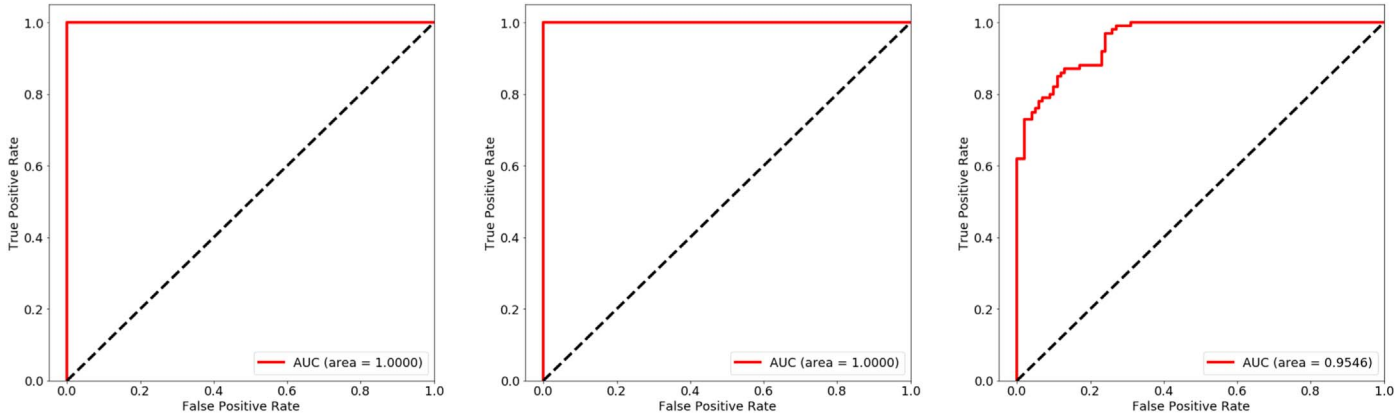

Figure 7. ROC-AUC curves for the binary classifiers (left to right: no substructure vs. particle, no substructure vs. vortex, and particle vs. vortex) which were trained in addition to our multiclass classifiers. The binary classifiers play the role of distinguishing between the two most probable classes of our multiclass classifier.

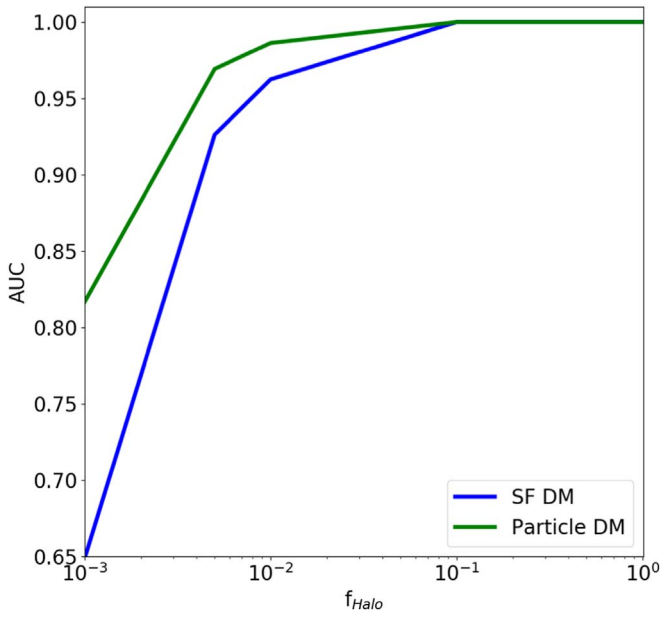

Figure 8. AUC as a function of the ratio of substructure to host halo mass.

ORCID iDs

Sergei Gleyzer https://orcid.org/0000-0002-6222-8102
Evan McDonough https://orcid.org/0000-0002-2130-3903
Michael W. Toomey https://orcid.org/0000-0003-1205-4033
Emanuele Usai https://orcid.org/0000-0001-9323-2107

References

Aaboud, M., Aad, G., Abbott, B., et al. 2019, JHEP, 2019, 142
Abazajian, K. 2006, PhRvD, 73, 063513
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, PhRvL, 119, 161101
Abbott, L. F., & Sikivie, P. 1983, PhLB, 120, 133
Akerib, D. S., Aslum, S., Araújo, H. M., et al. 2017, PhRvL, 118, 021303
Alexander, S., Bramburger, J. J., & McDonough, E. 2019, PhLB, 797, 134871
Alexander, S., & Cormack, S. 2017, JCAP, 1704, 005
Alexander, S., McDonough, E., & Spergel, D. N. 2018, JCAP, 1805, 003
Alford, M. G., Rajagopal, K., & Wilczek, F. 1998, PhLB, 422, 247
Alford, M. G., Schmitt, A., Rajagopal, K., & Schäfer, T. 2008, RvMP, 80, 1455
Anderson, L., Aubourg, E., et al. 2014, MNRAS, 441, 24
Aprile, E., Aalbers, J., Agostini, F., et al. 2018, PhRvL, 121, 111302
Banik, N., & Sikivie, P. 2013, PhRvD, 88, 123517
Baym, G., Pethick, C., & Pines, D. 1969, Natur, 224, 673
Berezhiani, L., & Khoury, J. 2015, PhRvD, 92, 103510
Berezhiani, L., & Khoury, J. 2016, PhLB, 753, 639
Bode, P., Ostriker, J. P., & Turok, N. 2001, ApJ, 556, 93
Brandenberger, R. H. 1994, IJMPA, A9, 2117
Brandenberger, R. H. 2014, NuPhS, 246, 45
Brandenberger, R. H., Danos, R. J., Hernandez, O. F., & Holder, G. P. 2010, JCAP, 1012, 028
Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G., & Cranmer, K. 2019, ApJ, 886, 49
Brennan, S., Benson, A. J., Cyr-Racine, F.-Y., et al. 2019, MNRAS, 488, 5085
Bullock, J. S., & Boylan-Kolch, M. 2017, ARA&A, 55, 343
Burkert, A. 1996, ApJL, 447, L25
Carleo, G., Cirac, I., Cranmer, K., et al. 2019, RvMP, 92, 045002
Ciuca, R., & Hernández, O. F. 2019, MNRAS, 483, 5179
Ciuca, R., Hernández, O. F., & Wolman, M. 2019, MNRAS, 485, 1377
Cui, X., Abdukerim, A., Chen, W., et al. 2017, PhRvL, 119, 181302
Cyr-Racine, F.-Y., Keeton, C. R., & Moustakas, L. A. 2019, PhRvD, 100, 023013
Cyr-Racine, F.-Y., Moustakas, L. A., Keeton, C. R., Sigurdson, K., & Gilman, D. A. 2016, PhRvD, 94, 043505
Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25
Daylan, T., Cyr-Racine, F.-Y., Diaz Rivero, A., et al. 2018, ApJ, 854, 141
Dean, D. J., & Hjorth-Jensen, M. 2003, RvMP, 75, 607
Diaz Rivero, A., Cyr-Racine, F.-Y., & Dvorkin, C. 2018, PhRvD, 97, 023001
Díaz Rivero, A., Dvorkin, C., Cyr-Racine, F.-Y., Zavala, J., & Vogelsberger, M. 2018, PhRvD, 98, 103517
Dine, M., & Fischler, W. 1983, PhLB, 120, 137
Drukier, A. K., Freese, K., & Spergel, D. N. 1986, PhRvD, 33, 3495
Ferreira, E. G. M., Franzmann, G., Khoury, J., & Brandenberger, R. 2019, JCAP, 2019, 027
Gasparini, M. A., Marshall, P., Treu, T., Morganson, E., & Dubath, F. 2008, MNRAS, 385, 1959
Goodman, M. W., & Witten, E. 1985, PhRvD, 31, 3059
Haskell, B., & Sedrakian, A. 2018, Astrophysics and Space Science Library, Vol. 457 (Cham: Springer), 401
He, K., Zhang, X., Ren, S., & Sun, J. 2015, arXiv:1512.03385
Heymans, C., van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146
Hezaveh, Y. D., Perreault Levasseur, L., & Marshall, P. J. 2017, Natur, 548, 555
Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, ApJ, 823, 37
Hsueh, J. W., Oldham, L., Spingola, C., et al. 2017, MNRAS, 469, 3713
Hu, W., Barkana, R., & Gruzinov, A. 2000, PhRvL, 85, 1158
Huang, G., Liu, Z., & Weinberger, K. Q. 2016, arXiv:1608.06993
Hui, L., Ostriker, J. P., Tremaine, S., & Witten, E. 2017, PhRvD, 95, 043541
Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201
Kim, S. Y., Peter, A. H. G., & Hargis, J. R. 2018, PhRvL, 121, 211302
Koopmans, L. V. E. 2005, MNRAS, 363, 1136
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2017, Commun. ACM, 60, 84
Lombardo, U., & Schulze, H. J. 2001, LNP, 578, 30
Mao, S., & Schneider, P. 1998, MNRAS, 295, 587
McDonough, E., & Brandenberger, R. H. 2013, JCAP, 1302, 045
Morganson, E., Marshall, P., Treu, T., Schrabback, T., & Blandford, R. D. 2010, MNRAS, 406, 2452

Morningstar, W. R., Hezaveh, Y. D., Perreault Levasseur, L., et al. 2018, arXiv:1808.00011
Morningstar, W. R., Perreault Levasseur, L., Hezaveh, Y. D., et al. 2019, ApJ, 883, 14
Narayan, R., & Bartelmann, M. 1997, arXiv:9606001
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Nightingale, J. W., & Dye, S. 2015, MNRAS, 452, 2940
Nightingale, J. W., Dye, S., & Massey, R. J. 2018, MNRAS, 478, 4738
Ntampaka, M., Avestruz, C., Boada, S., et al. 2019, BAAS, 51, 14
Page, D., Lattimer, J. M., Prakash, M., & Steiner, A. W. 2013, arXiv:1302.6626
Peccei, R. D., & Quinn, H. R. 1977, PhRvL, 38, 1440
Perreault Levasseur, L., Hezaveh, Y. D., & Wechsler, R. H. 2017, ApJL, 850, L7
Planck Collaboration 2016, A&A, 594, A63
Preskill, J., Wise, M. B., & Wilczek, F. 1983, PhLB, 120, 127
Rindler-Daller, T., & Shapiro, P. R. 2012, MNRAS, 422, 135
Sazhin, M. V., Khovanskaya, O. S., Capaccioli, M., et al. 2007, MNRAS, 376, 1731
Schmitt, A. 2015, LNP, 888, 1
Sikivie, P., & Yang, Q. 2009, PhRvL, 103, 111301
Silverman, M. P., & Mallett, R. L. 2002, GReGr, 34, 633
Simonyan, K., & Zisserman, A. 2014, arXiv:1409.1556
Sin, S.-J. 1994, PhRvD, 50, 3650
Spergel, D. N., & Steinhardt, P. J. 2000, PhRvL, 84, 3760
Vegetti, S., & Koopmans, L. V. E. 2009a, MNRAS, 392, 945
Vegetti, S., & Koopmans, L. V. E. 2009b, MNRAS, 400, 1583
Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, MNRAS, 408, 1969
Verma, A., Collett, T., Smith, G. P., et al. 2019, arXiv:1902.05141
Weinberg, S. 1978, PhRvL, 40, 223
Wilczek, F. 1978, PhRvL, 40, 279