Inferring dark matter substructure with astrometric lensing beyond the power spectrum

Siddharth Mishra-Sharma
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
Massachusetts Institute of Technology
Harvard University
New York University
smsharma@mit.edu

Abstract

Astrometry—the precise measurement of positions and motions of celestial objects—has emerged as a promising avenue for characterizing the dark matter population in our Galaxy. By leveraging recent advances in simulation-based inference and neural network architectures, we introduce a novel method to search for global dark matter-induced gravitational lensing signatures in astrometric datasets. Our method based on neural likelihood-ratio estimation shows significantly enhanced sensitivity to a cold dark matter population and more favorable scaling with measurement noise compared to existing approaches based on two-point correlation statistics, establishing machine learning as a powerful tool for characterizing dark matter using astrometric data.

1 Introduction and background

Although there exists plenty of evidence for dark matter (DM) on galactic scales and above (see Ref. [1] for a recent overview), the distribution of DM clumps—subhalos—on sub-galactic scales is less well-understood and remains an active area of cosmological study. This distribution additionally correlates with and may provide clues about the underlying particle physics nature of dark matter (see *e.g.*, Refs. [2–4]), highlighting its relevance across multiple domains.

While more massive dark matter subhalos can be detected and studied through their association with luminous tracers such as bound stellar populations, subhalos with smaller masses $\lesssim 10^9 \, \mathrm{M_\odot}$ are not generally associated with luminous matter [5, 6], rendering their characterization challenging. Gravitational effects provide one of the few avenues to probe the distribution of these otherwise-invisible subhalos [7]. Gravitational lensing *i.e.*, the bending of light from a background source due to a foreground mass, is one such effect and has been proposed in various incarnations as a probe of dark subhalos. Strong gravitational lensing, for example, has been used to infer the presence of dark matter substructure in galaxies outside of our own [8–11]. Astrometric lensing, on the other hand, has recently emerged as a promising way to characterize the dark matter subhalo population within the Milky Way.

Astrometry refers to the precise measurement of the positions and motions of luminous celestial objects like stars and galaxies. Gravitational lensing of these background objects by a moving foreground mass, such as a dark matter subhalo, can imprint a characteristic, correlated signal on their measured kinematics (angular velocities and/or accelerations). Ref. [12] introduced several methods for extracting this signature, including computing convolutions of the expected lensing signal on astrometric datasets and detecting local kinematic outliers. Ref. [13] applied the former method to data from the *Gaia* satellite, obtaining constraints on the abundance of dark compact objects in the

Milky Way and showcasing the applicability of astrometric dark matter searches in a practical setting. Finally, Ref. [14] proposed using the angular power spectrum of the astrometric field as an observable to infer the population properties of subhalos in our Galaxy, leveraging the collective, correlated signal of a large subhalo sample.

Astrometric datasets are inherently high-dimensional, consisting of positions and kinematics of potentially millions of objects. Especially when the expected signal consists of the collective imprint of a large number of lenses, characterizing their population properties involves marginalizing over all possible configurations of subhalos, rendering the likelihood intractable and usually necessitating the use of simplified data representations like the power spectrum. While effective, such simplification can result in loss of information compared to that contained in the original dataset when the expected signal is non-Gaussian in nature. The existence of systematic effects that are degenerate with a putative signal in the low-dimensional summary domain can further inhibit sensitivity.

The dawn of the era of precision astrometry, with the *Gaia* satellite [15] having recently delivered the most precise astrometric dataset to-date [16–18] and surveys including the Square Kilometer Array (SKA) [19, 20] and Roman Space Telescope [21] set to achieve further leaps in sensitivity over the next decade, calls for methods that can extract more information from these datasets than is possible using existing techniques. In this paper, we introduce such a method that uses spherical convolutional neural networks in conjunction with parameterized classifiers [22, 23] in order to estimate likelihood ratios associated with the abundance of a cold dark matter population directly from a binned map of the astrometric velocity field. We show that our method outperforms established proposals based on the two-point correlation statistics of the astrometric field, both in absolute sensitivity as well as its scaling with measurement noise.
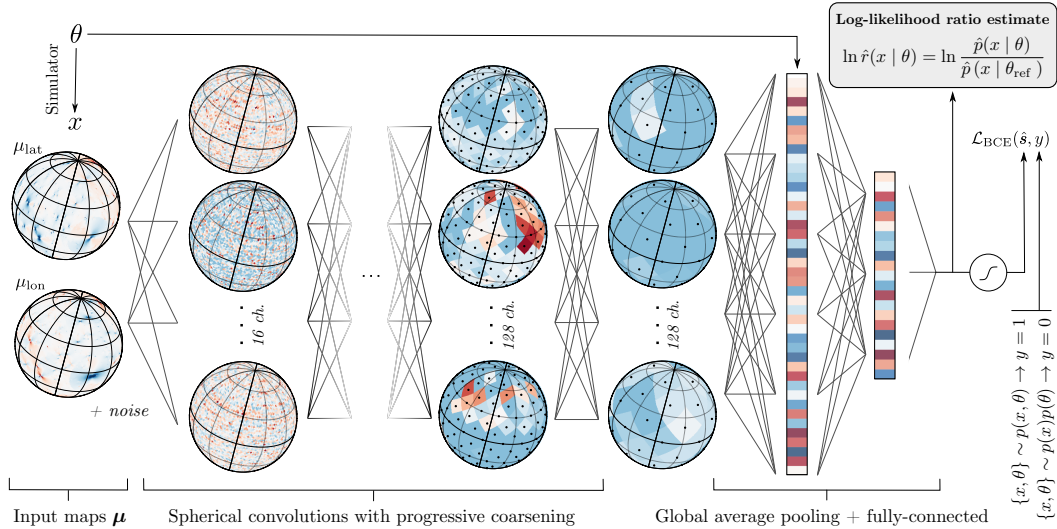
2   Model and inference

Figure 1: A schematic illustration of the method and neural network architecture used in this work.

The forward model   We consider a population of Navarro-Frenk-White (NFW) [24] subhalos following a power-law mass function, $\mathrm{d}n/\mathrm{d}m \propto m^{\alpha}$, with slope $\alpha = -1.9$ as expected if the population is sourced from nearly scale-invariant primordial fluctuations in the canonical Λ Cold Dark Matter (ΛCDM) scenario. The concentration-mass relation from Ref. [25] is used to model the concentrations associated with density profiles of individual subhalos.

Subhalos between 10^7–$10^{10}\,\mathrm{M}_\odot$ are simulated, assuming the influence of lighter subhalos to be too small to be discernable [14]. The subhalo fraction f_{sub}, quantifying the expected fraction of the mass of the Milky Way contributed by subhalos in the range 10^{-6}–$10^{10}\,\mathrm{M}_\odot$, is taken to be the parameter of interest. The spatial distribution of subhalos in the Galactocentric frame is modeled using results from the Aquarius simulation following Refs. [26, 27]. Since this spatial distribution accounts for

the depletion of subhalos towards the Galactic Center due to gravitational tidal effects, the angular number density of subhalos looking out from the Sun frame can be considered to be effectively isotropic.

The asymptotic velocities of subhalos in the Galactocentric frame are taken to follow a truncated Maxwell-Boltzmann distribution [28, 29] $f_{\mathrm{Gal}}(\mathbf{v}) \propto e^{-\mathbf{v}^2/v_0^2} \cdot H(v_{\mathrm{esc}} - |\mathbf{v}|)$, where $v_{\mathrm{esc}} = 550\,\mathrm{km\,s^{-1}}$ is the Galactic escape velocity [30], $v_0 = 220\,\mathrm{km\,s^{-1}}$ [31], and H is the Heaviside step function. Once instantiated, the positions and velocities of subhalos are transformed into the Galactic frame, assuming $R_\odot = 8.2\,\mathrm{kpc}$ to be the distance of the Sun from the Galactic Center [32, 33] and $\mathbf{v}_\odot = (11, 232, 7)\,\mathrm{km\,s^{-1}}$ its Galactocentric velocity [34]. Note that the asymmetry in the direction of motion of the Sun in the Milky Way introduces a preferred direction for the Sun-frame velocities of subhalos, breaking strict rotation invariance in the forward model. Although not explicitly pursued here, this asymmetry can be used as an additional distinguishing handle for the lensing signal, as was done in Ref. [14].

Our datasets consist of the 2-dimensional angular velocity map of background sources on the celestial sphere. Given a spherically-symmetric subhalo lens moving with transverse velocity $\mathbf{v}_l$, the expected lens-induced velocity for a quasar at impact parameter $\mathbf{b}$ is given by [12]

$$\boldsymbol{\mu}(\mathbf{b}) = 4G_{\mathrm{N}} \left\{ \frac{M(b)}{b^2} \left[2\hat{\mathbf{b}}\left(\hat{\mathbf{b}} \cdot \mathbf{v}_l\right) - \mathbf{v}_l \right] - \frac{M'(b)}{b}\hat{\mathbf{b}}\left(\hat{\mathbf{b}} \cdot \mathbf{v}_l\right) \right\} \qquad (1)$$

where $M(b)$ and $M'(b)$ are the projected mass of the subhalo at a given impact parameter distance $b = |\mathbf{b}|$ and its gradient. An example of the induced velocity signal on part of the celestial sphere, projected along the Galactic latitudinal and longitudinal directions and exhibiting dipole-like structures, is shown in the leftmost column of Fig. 1.

We take our source population to consist of remote, point-like galaxies known as quasars which, due to their large distances from the Earth, are not expected to have significant intrinsic angular velocities. We assume the sources to be isotropically-distributed, although this assumption can be easily relaxed for a realistic source sample. The velocity maps are assumed to be spatially binned, and we use a HEALPix binning [35] with resolution parameter nside=64, corresponding to $N_{\mathrm{pix}} = 49{,}152$ pixels over the full sky with pixel area $\sim 0.8\,\mathrm{deg}^2$. The values within each pixel then quantify the average latitudinal and longitudinal velocity components of quasars within that pixel, with the impact parameter $\mathbf{b}$ representing the vector from the center of a subhalo to the center of the pixel.

In order to enable a comparison with traditional approaches—which are generally not expected to be sensitive to a cold dark matter subhalo population with next-generation astrometric surveys [12, 14]—we benchmark using an optimistic observational configuration corresponding to measuring the proper motions of $N_q = 10^8$ quasars with noise $\sigma_\mu = 0.1\,\mu\mathrm{as\,yr^{-1}}$.

The power spectrum approach   Ref. [14] introduced an approach for extracting the astrometric signal due to a dark matter subhalo population by decomposing the observed map into its angular (vector) power spectrum. The power spectrum is a summary statistic ubiquitous in astrophysics and cosmology and quantifies the amount of correlation contained at different spatial scales. In the case of data on a sphere, the basis of spherical harmonics is often used, and the power spectrum then encodes the correlation structure on different multipoles ℓ. The power spectrum effectively captures the linear component of the signal and, when the underlying signal is a Gaussian random field, captures *all* of the relevant information contained in the map(s) [36]. The expected signal in the power spectrum domain can be evaluated semi-analytically using the formalism described in Ref. [14] and, assuming a Gaussian likelihood, the expected sensitivity can be computed using a Fisher forecasting approach. We use this prescription as a comparison point to the method introduced here.

While effective, reduction of the full astrometric map to its power spectrum results in loss of information; this can be seen from the fact that the signal in the leftmost column of Fig. 1 is far from Gaussian. Furthermore, the existence of correlations on large angular scales due to *e.g.*, biases in calibration of celestial reference frames [37] or systematic variations in measurements taken over different regions of the sky introduces degeneracies with a putative signal and precludes their usage in the present context. For this reason multipoles $\ell < 10$ were discarded in Ref. [14], degrading the projected sensitivity.

Likelihood-ratio estimation using parameterized classifiers   Recent advances in machine learning have enabled methods that can be used to efficiently perform inference on models defined through complex simulations; see Ref. [38] for a recent review. Here, we make use of neural likelihood-ratio estimation [22, 23, 39–42], previously applied to the problem of inferring dark matter substructure using observations of strong gravitational lenses [43] and cold stellar streams [44].

Given a classifier that can distinguish between samples $\{x\} \sim p(x \mid \theta)$ drawn from parameter points θ and those from a fixed reference hypothesis $\{x\} \sim p(x \mid \theta_{\mathrm{ref}})$, the decision function output by the optimal classifier $s(x, \theta) = p(x \mid \theta)/(p(x \mid \theta) + p(x \mid \theta_{\mathrm{ref}}))$ is one-to-one with the likelihood ratio, $r(x \mid \theta) \equiv p(x \mid \theta)/p(x \mid \theta_{\mathrm{ref}}) = s(x, \theta)/(1 - s(x, \theta))$, a fact appreciated as the likelihood-ratio trick [22, 45]. The classifier $s(x, \theta)$ in this case is a neural network that can work directly on the high-dimensional data x, and is parameterized by θ by having it included as an input feature. In order to improve numerical stability and reduce dependence on the fixed reference hypothesis θ_{ref}, we follow Ref. [42] and train a classifier to distinguish between data-sample pairs from the joint distribution $\{x, \theta\} \sim p(x, \theta)$ and those from a product of marginal distributions $\{x, \theta\} \sim p(x)p(\theta)$ (defining the reference hypothesis and in practice obtained by shuffling samples within a batch) using the binary cross-entropy (BCE) loss as the optimization objective.

Extracting information from high-dimensional astrometric maps   Since our dataset consists of a velocity field sampled on a sphere, we use a spherical convolutional neural network in order to directly learn useful representations from these maps that are efficiently suited for the downstream classification task. Specifically, we make use of `DeepSphere` [46, 47], a graph-based convolutional neural network tailored to data sampled on a sphere. For this purpose, the `HEALPix` grid can be cast as a weighted undirected graph with N_{pix} vertices and edges connecting each pixel vertex to its set of 8 neighboring pixels. The weighted adjacency matrix over neighboring pixels (i, j) is given by $A_{ij} = \exp\left(-\Delta r_{ij}^2/\rho^2\right)$ where Δr_{ij} specifies the 3-dimensional Euclidean distance between the pixel centers and the widths ρ are obtained from Ref. [46]. `DeepSphere` then efficiently performs convolutions in the spectral domain using a basis of Chebychev polynomials as convolutional kernels [48]; here, we set $K = 4$ as the maximum polynomial order.

All inputs are normalized to zero mean and unit standard deviation across the training sample. Starting with 2 scalar input channels representing the two orthogonal (Galactic latitude and longitude) components of the velocity vector map,[1] we perform a graph convolution operation, increasing the channel dimension to 16 followed by a batch normalization, ReLU nonlinearity, and downsampling the representation by a factor of 4 with max pooling into the next coarser `HEALPix` resolution. Pooling leverages scale separation, preserving important characteristics of the signal across different resolutions. Four more such layers are employed, increasing the channel dimension by a factor of 2 at each step until a maximum of 128, with maps after the last convolutional layer having resolution `nside=2` corresponding to 48 pixels. At this stage, we average over the spatial dimension (known as global average pooling [49]) in order to encourage approximate rotation invariance, outputting 128 features onto which the parameter of interest f_{sub} is appended. These features are passed through a fully-connected network with (1024, 128) hidden units and ReLU activations outputting the classifier decision $\hat{s}$ by applying a sigmoidal projection.

Model training and evaluation 10^5 maps from the forward model were produced, with 15% of these held out for validation. The estimator was trained using a batch size of 64 for up to 50 epochs with early stopping if the validation loss had not improved after 10 epochs. The `Adam` optimizer [50] was used with initial learning rate 10^{-3} decayed through cosine annealing. A coarse grid search was used to inform the architecture and hyperparameter choices.

For a given test map, the log-likelihood ratio profile can be obtained by evaluating the trained estimator for different values of f_{sub} while keeping the input map fixed. The network output prior to the final sigmoidal projection directly gives the required log-likelihood ratio estimate: $\ln \hat{r} = S^{-1}(\hat{s})$, where S is the sigmoid function [42, 44]. Figure 1 presents an illustrative summary of the neural network architecture and method used in this work.

---

[1]We note that by representing the input angular velocity vector field in terms of two input scalar channels, we break the desired rotation equivariance of spherical convolutions due to differences in how scalar and vector representations transform under rotations. Although this will have a downstream effect on rotation invariance, a detailed study of how this influences the performance of our model is beyond the scope of this paper.
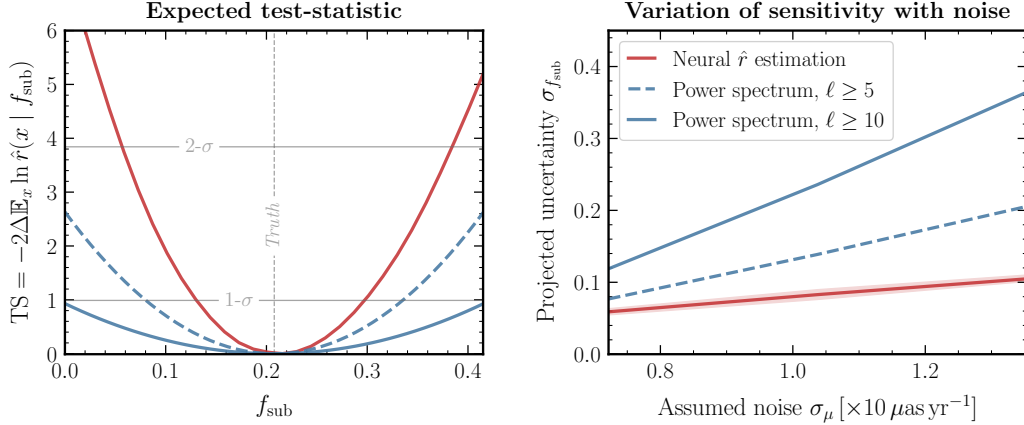
Figure 2: *(Left)* The expected log-likelihood ratio test-statistic (TS) profile for a cold dark matter population as a function of substructure fraction $f_{\rm sub}$ obtained using the neural likelihood-ratio estimation method introduced in this work (red line) compared with the corresponding profiles for existing approaches using power spectrum summaries with different multipole thresholds $\ell \gtrsim 5$ (dashed blue line) and $\ell \gtrsim 10$ (solid blue line). The vertical dotted line indicates the true benchmark value of the parameter $f_{\rm sub}$ in the test dataset. Our method shows enhanced sensitivity to a cold dark matter population compared to traditional approaches. *(Right)* Scaling of the expected sensitivities, quantified by the respective 1-σ uncertainties, with per-object instrumental noise. For the machine learning-based approach, the band quantifies the middle-95% containment of the inferred 1-σ uncertainty. Our method shows a more favorable scaling with assumed measurement noise.

3 Experiments on simulated data

We evaluate our trained likelihood-ratio estimator on maps drawn from a benchmark configuration motivated by Refs. [26, 27], containing 150 subhalos in expectation between 10^8–$10^{10} \, {\rm M}_\odot$ and corresponding to $f_{\rm sub} \simeq 0.2$. The left panel of Fig. 2 shows the expected log-likelihood ratio test-statistic (TS) as a function of substructure fraction $f_{\rm sub}$ for this nominal configuration. This is obtained by evaluating the trained estimator on 100 test maps over a uniform grid in $f_{\rm sub}$ and taking the point-wise mean. Corresponding curves using the power spectrum approach are shown in blue, using minimum multipoles of $\ell \geq 5$ (dashed) and $\ell \geq 10$ (solid). Thresholds corresponding to 1- and 2-σ significance assuming a χ^2-distributed TS are shown as the horizontal grey lines. We see that sensitivity gains of over a factor of ~ 2 can be expected for this particular benchmark when using the machine learning approach compared to the traditional power spectrum approach. No significant bias on the central value of the inferred DM abundance relative to the overall uncertainty scale is observed.

The right panel of Fig. 2 shows the scaling of expected 1-σ uncertainty on substructure fraction $f_{\rm sub}$ with assumed noise per quasar, keeping the number of quasars fixed (red, with the line showing the median and shaded band corresponding to the middle-95% containment of the uncertainty inferred over 50 test datasets) compared to the power spectrum approach (blue lines). A far more favorable scaling of the machine learning approach is seen compared to the power spectrum approach, suggesting that it is especially advantageous in low signal-to-noise regimes that are generally most relevant for dark matter searches.

Since the existence of measurement noise correlated on large spatial scales is a potential source of systematic uncertainty when working with astrometric maps, we test the susceptibility of our method to such effects by creating simulated data containing large-scale noise not previously seen by the trained estimator. Instead of assuming a scale-invariant noise power spectrum $C_\ell^{\rm noise} = 4\pi\sigma_\mu^2/N_q$ [14], in this case we model noise with an order of magnitude excess in power on scales $\ell \lesssim 10$, parameterized as $C_\ell^{\rm noise} = 4\pi\sigma_\mu^2/N_q \cdot (10 - 9S(\ell - 10))$ where S denotes the sigmoid function. The left panel of Fig. 3 illustrates this noise model (thicker green line) as well as the power spectrum of one simulated realization from this model (thinner green line, obtained using the `HEALPix` module `anafast`) contrasted with the standard scale-invariant noise case (red lines). The

right panel of Fig. 3 shows the expected log-likelihood ratio test-statistic profile for the two cases. Although a bias in the maximum-likelihood estimate of f_{sub} is seen when the test data has unmodeled noise (green line), the true test parameter value (dashed vertical line) is seen to lie well within the inferred 1-σ confidence interval. This suggests that the method is only marginally susceptible to substantive amounts of correlated noise on large spatial scales.
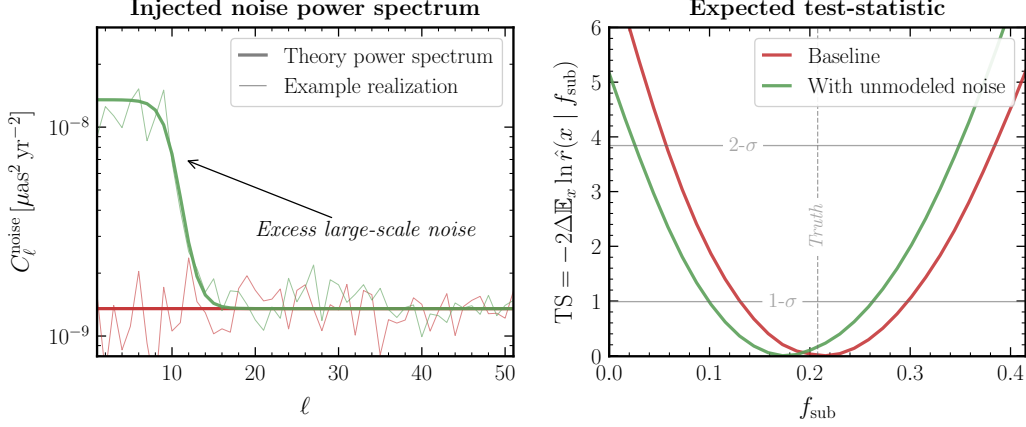

Figure 3: *(Left)* The power spectrum of the noise model (thicker green line) used to study the impact of correlated noise on large spatial scales, not modeled during training, on the performance of the likelihood-ratio estimator. The thinner green line shows the power spectrum of an example noise realization instantiated from this noise model. The red lines show corresponding power spectra for a scale-invariant noise model. *(Right)* The expected test-statistic profile for a model evaluated on maps containing excess large-scale noise (green line) compared to the model evaluated on maps with scale-invariant noise (red line). A bias in the maximum-likelihood estimate returned by the model is seen when substantial unaccounted-for noise is presented in the test maps.

4 Conclusions and outlook

We have introduced a method to analyze astrometric datasets over large regions of the sky using techniques based on machine learning with the aim of inferring the lensing signature of a dark matter substructure. We have shown our method to be significantly more sensitive to a cold dark matter subhalo population compared to established methods based on global summary statistics, with more favorable scaling as a function of measurement noise. Since the collection and reduction of astrometric data is an expensive endeavor, the use of methods that can take advantage of more of the available information can be equated to long periods of data-taking, underscoring their importance. Additionally, unlike the power spectrum approach, the current method does not require the construction of a numerically-expensive estimator to account for non-uniform exposure, selection effects, and instrumental noise in realistic datasets. These, as well as any other modeled observational effects, can be incorporated directly at the level of the forward model.

We have focused in this work on assessing sensitivity to a cold dark matter-like subhalo population with quasar velocity astrometry, which is within the scope of upcoming radio surveys like the SKA [19, 20]. Our method can also be applied in a straightforward manner to look for the *acceleration* lensing signal imprinted on Milky Way stars, in particular sourced by a population of more compact subhalos than those expected in the cold dark matter scenario. These features are expected to imprint a larger degree of non-Gaussianity compared to the signal explored here (as can be seen, *e.g.*, from Fig. 1 of Ref. [14]), and machine learning methods may provide larger relative sensitivity gains when deployed in that context. Such analyses are within purview of the upcoming Roman exoplanet microlensing survey [51] as well as future *Gaia* data releases.

Several improvements and extensions to the method presented in this paper are possible. The use of architectures that can equivariantly handle vector inputs [52] can aid in learning more efficient representations of the astrometric map. Using convolutions based on fixed rather than learned filters can additionally reduce model complexity and produce more interpretable representations [53–57].

The use of methods for likelihood-ratio estimation that can leverage additional latent information in the forward model can significantly enhance the sample efficiency of the analysis [39–41, 58]. We leave the study of these extensions as well as application of our method to other dark matter population scenarios to future work.

Astrometric lensing has been established as a promising way to characterize the Galactic dark matter population, with theoretical progress in recent years going in step with advances on the observational front. While this work is a first attempt at bringing principled machine learning techniques to this field, with the availability of increasingly complex datasets we expect machine learning to be an important general-purpose tool for future astrometric dark matter searches.

Code used for reproducing the results presented in this paper is available at https://github.com/smsharma/neural-global-astrometry.

Acknowledgments and disclosure of funding

References

[1] A. M. Green, *"Dark Matter in Astrophysics/Cosmology,"* in *Les Houches summer school on Dark Matter* (2021) arXiv:2109.05854 [hep-ph].

[2] K. Schutz, *"Subhalo mass function and ultralight bosonic dark matter,"* Phys. Rev. D **101**, 123026 (2020), arXiv:2001.05503 [astro-ph.CO].

[3] P. Bode, J. P. Ostriker, and N. Turok, *"Halo formation in warm dark matter models,"* Astrophys. J. **556**, 93 (2001), arXiv:astro-ph/0010389.

[4] J. J. Dalcanton and C. J. Hogan, *"Halo cores and phase space densities: Observational constraints on dark matter physics and structure formation,"* Astrophys. J. **561**, 35 (2001), arXiv:astro-ph/0004381.

[5] A. Fitts *et al.*, *"FIRE in the Field: Simulating the Threshold of Galaxy Formation,"* Mon. Not. Roy. Astron. Soc. **471**, 3547 (2017), arXiv:1611.02281 [astro-ph.GA].

[6] J. I. Read, G. Iorio, O. Agertz, and F. Fraternali, *"The stellar mass-halo mass relation of isolated field dwarfs: a critical test of ΛCDM at the edge of galaxy formation,"* Mon. Not. Roy. Astron. Soc. **467**, 2019 (2017), arXiv:1607.03127 [astro-ph.GA].

[7] M. R. Buckley and A. H. G. Peter, *"Gravitational probes of dark matter physics,"* Phys. Rept. **761**, 1 (2018), arXiv:1712.06615 [astro-ph.CO].

[8] Y. D. Hezaveh *et al.*, *"Detection of lensing substructure using ALMA observations of the dusty galaxy SDP.81,"* Astrophys. J. **823**, 37 (2016), arXiv:1601.01388 [astro-ph.CO].

---

[2]https://github.com/deepsphere/deepsphere-pytorch

[9] S. Vegetti, L. V. E. Koopmans, A. Bolton, T. Treu, and R. Gavazzi, *"Detection of a Dark Substructure through Gravitational Imaging,"* Mon. Not. Roy. Astron. Soc. **408**, 1969 (2010), arXiv:0910.0760 [astro-ph.CO].

[10] D. Gilman, S. Birrer, A. Nierenberg, T. Treu, X. Du, and A. Benson, *"Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses,"* Mon. Not. Roy. Astron. Soc. **491**, 6077 (2020), arXiv:1908.06983 [astro-ph.CO].

[11] S. Vegetti, D. J. Lagattuta, J. P. McKean, M. W. Auger, C. D. Fassnacht, and L. V. E. Koopmans, *"Gravitational detection of a low-mass dark satellite at cosmological distance,"* Nature **481**, 341 (2012), arXiv:1201.3643 [astro-ph.CO].

[12] K. Van Tilburg, A.-M. Taki, and N. Weiner, *"Halometry from Astrometry,"* JCAP **07**, 041 (2018), arXiv:1804.01991 [astro-ph.CO].

[13] C. Mondino, A.-M. Taki, K. Van Tilburg, and N. Weiner, *"First Results on Dark Matter Substructure from Astrometric Weak Lensing,"* Phys. Rev. Lett. **125**, 111101 (2020), arXiv:2002.01938 [astro-ph.CO].

[14] S. Mishra-Sharma, K. Van Tilburg, and N. Weiner, *"Power of halometry,"* Phys. Rev. D **102**, 023026 (2020), arXiv:2003.02264 [astro-ph.CO].

[15] Gaia Collaboration, *"The Gaia mission,"* Astron. Astrophys. **595**, A1 (2016), arXiv:1609.04153 [astro-ph.IM].

[16] Gaia Collaboration, *"Gaia Data Release 2. Summary of the contents and survey properties,"* Astron. Astrophys. **616**, A1 (2018), arXiv:1804.09365 [astro-ph.GA].

[17] L. Lindegren *et al.*, *"Gaia Data Release 2. The astrometric solution,"* Astron. Astrophys. **616**, A2 (2018), arXiv:1804.09366 [astro-ph.IM].

[18] Gaia Collaboration, *"Gaia Early Data Release 3. Summary of the contents and survey properties,"* Astron. Astrophys. **649**, A1 (2021), arXiv:2012.01533 [astro-ph.GA].

[19] E. B. Fomalont and M. Reid, *"Microarcsecond astrometry using the SKA,"* New Astron. Rev. **48**, 1473 (2004), arXiv:astro-ph/0409611.

[20] M. J. Jarvis, D. Bacon, C. Blake, M. L. Brown, S. N. Lindsay, A. Raccanelli, M. Santos, and D. Schwarz, *"Cosmology with SKA Radio Continuum Surveys,"* (2015), arXiv:1501.03825 [astro-ph.CO].

[21] WFIRST Astrometry Working Group, *"Astrometry with the Wide-Field Infrared Space Telescope,"* Journal of Astronomical Telescopes, Instruments, and Systems **5**, 044005 (2019), arXiv:1712.05420 [astro-ph.IM].

[22] K. Cranmer, J. Pavez, and G. Louppe, *"Approximating Likelihood Ratios with Calibrated Discriminative Classifiers,"* (2015), arXiv:1506.02169 [stat.AP].

[23] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, *"Parameterized neural networks for high-energy physics,"* Eur. Phys. J. C **76**, 235 (2016), arXiv:1601.07913 [hep-ex].

[24] J. F. Navarro, C. S. Frenk, and S. D. M. White, *"The Structure of cold dark matter halos,"* Astrophys. J. **462**, 563 (1996), arXiv:astro-ph/9508025.

[25] M. A. Sánchez-Conde and F. Prada, *"The flattening of the concentration–mass relation towards low halo masses and its implications for the annihilation signal boost,"* Mon. Not. Roy. Astron. Soc. **442**, 2271 (2014), arXiv:1312.1729 [astro-ph.CO].

[26] M. Hütten, C. Combet, G. Maier, and D. Maurin, *"Dark matter substructure modelling and sensitivity of the Cherenkov Telescope Array to Galactic dark halos,"* JCAP **09**, 047 (2016), arXiv:1606.04898 [astro-ph.HE].

[27] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White, *"The Aquarius Project: the subhalos of galactic halos,"* Mon. Not. Roy. Astron. Soc. **391**, 1685 (2008), arXiv:0809.0898 [astro-ph].

[28] S. Chandrasekhar, *An introduction to the study of stellar structure* (1939).

[29] M. Lisanti, *"Lectures on Dark Matter Physics,"* in *Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings* (2017) pp. 399–446, arXiv:1603.03797 [hep-ph].

[30] T. Piffl *et al.*, *"The RAVE survey: the Galactic escape speed and the mass of the Milky Way,"* Astron. Astrophys. **562**, A91 (2014), arXiv:1309.4293 [astro-ph.GA].

[31] F. J. Kerr and D. Lynden-Bell, *"Review of galactic constants,"* Mon. Not. Roy. Astron. Soc. **221**, 1023 (1986).

[32] Gravity Collaboration, *"A geometric distance measurement to the Galactic center black hole with 0.3% uncertainty,"* Astron. Astrophys. **625**, L10 (2019), arXiv:1904.05721 [astro-ph.GA].

[33] J. Bovy, *"A purely acceleration-based measurement of the fundamental Galactic parameters,"* , arXiv:2012.02169 (2020), arXiv:2012.02169 [astro-ph.GA].

[34] R. Schönrich, J. Binney, and W. Dehnen, *"Local kinematics and the local standard of rest,"* Mon. Not. Roy. Astron. Soc. **403**, 1829 (2010), arXiv:0912.3693 [astro-ph.GA].

[35] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman, *"HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere,"* Astrophys. J. **622**, 759 (2005), arXiv:astro-ph/0409513 [astro-ph].

[36] M. Tegmark, *"How to measure CMB power spectra without losing information,"* Phys. Rev. D **55**, 5895 (1997), arXiv:astro-ph/9611174.

[37] Gaia Collaboration, *"Gaia Data Release 2. The celestial reference frame (Gaia-CRF2),"* Astron. Astrophys. **616**, A14 (2018), arXiv:1804.09377 [astro-ph.GA].

[38] K. Cranmer, J. Brehmer, and G. Louppe, *"The frontier of simulation-based inference,"* Proc. Nat. Acad. Sci. **117**, 30055 (2020), arXiv:1911.01429 [stat.ML].

[39] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *"A Guide to Constraining Effective Field Theories with Machine Learning,"* Phys. Rev. D **98**, 052004 (2018), arXiv:1805.00020 [hep-ph].

[40] J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, *"Mining gold from implicit models to improve likelihood-free inference,"* Proc. Nat. Acad. Sci. **117**, 5242 (2020), arXiv:1805.12244 [stat.ML].

[41] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *"Constraining Effective Field Theories with Machine Learning,"* Phys. Rev. Lett. **121**, 111801 (2018), arXiv:1805.00013 [hep-ph].

[42] J. Hermans, V. Begy, and G. Louppe, *"Likelihood-free MCMC with Amortized Approximate Ratio Estimators,"* (2019), arXiv:1903.04057 [stat.ML].

[43] J. Brehmer, S. Mishra-Sharma, J. Hermans, G. Louppe, and K. Cranmer, *"Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning,"* Astrophys. J. **886**, 49 (2019), arXiv:1909.02005 [astro-ph.CO].

[44] J. Hermans, N. Banik, C. Weniger, G. Bertone, and G. Louppe, *"Towards constraining warm dark matter with stellar streams through neural simulation-based inference,"* (2020), arXiv:2011.14923 [astro-ph.GA].

[45] S. Mohamed and B. Lakshminarayanan, "Learning in implicit generative models," (2017), arXiv:1610.03483 [stat.ML].

[46] M. Defferrard, M. Milani, F. Gusset, and N. Perraudin, *"DeepSphere: a graph-based spherical CNN,"* (2020), arXiv:2012.15000 [cs.LG].

[47] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, *"DeepSphere: Efficient spherical convolutional neural network with HEALPix sampling for cosmological applications,"* Astronomy and Computing **27**, 130 (2019), arXiv:1810.12186.

[48] M. Defferrard, X. Bresson, and P. Vandergheynst, *"Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering,"* (2016), arXiv:1606.09375 [cs.LG].

[49] M. Lin, Q. Chen, and S. Yan, "Network in network," (2014), arXiv:1312.4400 [cs.NE].

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," (2017), arXiv:1412.6980 [cs.LG].

[51] K. Pardo and O. Doré, *"Detecting dark matter subhalos with the Nancy Grace Roman Space Telescope,"* (2021), arXiv:2108.10886 [astro-ph.CO].

[52] C. Esteves, A. Makadia, and K. Daniilidis, "Spin-Weighted Spherical CNNs," (2020), arXiv:2006.10731 [cs.CV].

[53] S. Cheng, Y.-S. Ting, B. Ménard, and J. Bruna, *"A new approach to observational cosmology using the scattering transform,"* Mon. Not. Roy. Astron. Soc. **499**, 5902 (2020), arXiv:2006.08561 [astro-ph.CO].

[54] W. Ha, C. Singh, F. Lanusse, S. Upadhyayula, and B. Yu, *"Adaptive wavelet distillation from neural networks through interpretations,"* (2021), arXiv:2107.09145 [stat.ML].

[55] A. K. Saydjari and D. P. Finkbeiner, *"Equivariant Wavelets: Fast Rotation and Translation Invariant Wavelet Scattering Transforms,"* (2021), arXiv:2104.11244 [cs.CV].

[56] J. D. McEwen, C. G. R. Wallis, and A. N. Mavor-Parker, *"Scattering Networks on the Sphere for Scalable and Rotationally Equivariant Spherical CNNs,"* (2021), arXiv:2102.02828 [cs.CV].

[57] G. Valogiannis and C. Dvorkin, *"Towards an Optimal Estimation of Cosmological Parameters with the Wavelet Scattering Transform,"* (2021), arXiv:2108.07821 [astro-ph.CO].

[58] M. Stoye, J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, *"Likelihood-free inference with an improved cross-entropy estimator,"* (2018), arXiv:1808.00973 [stat.ML].

[59] T. P. Robitaille *et al.* (Astropy), *"Astropy: A Community Python Package for Astronomy,"* Astron. Astrophys. **558**, A33 (2013), arXiv:1307.6212 [astro-ph.IM].

[60] A. Price-Whelan *et al.*, *"The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package,"* Astron. J. **156**, 123 (2018), arXiv:1801.02634.

[61] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski, *"healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python,"* Journal of Open Source Software **4**, 1298 (2019).

[62] F. Pérez and B. E. Granger, *"IPython: a system for interactive scientific computing,"* Computing in Science and Engineering **9**, 21 (2007).

[63] T. Kluyver *et al.*, *"Jupyter notebooks - a publishing format for reproducible computational workflows,"* in *ELPUB* (2016).

[64] J. D. Hunter, *"Matplotlib: A 2D graphics environment,"* Computing In Science & Engineering **9**, 90 (2007).

[65] A. Chen *et al.*, *"Developments in mlflow: A system to accelerate the machine learning lifecycle,"* in *Proceedings of the fourth international workshop on data management for end-to-end machine learning* (2020) pp. 1–4.

[66] C. R. Harris *et al.*, *"Array programming with NumPy,"* Nature **585**, 357 (2020).

[67] M. Defferrard, L. Martin, R. Pena, and N. Perraudin, *"PyGSP: Graph Signal Processing in Python,"* (2017).

[68] A. Paszke *et al.*, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019) pp. 8024–8035.

[69] M. Fey and J. E. Lenssen, *"Fast graph representation learning with PyTorch Geometric,"* in *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019).

[70] W. Falcon *et al.*, *"Pytorchlightning/pytorch-lightning: 0.7.6 release,"* (2020).

[71] A. Tejero-Cantero *et al.*, *"sbi: A toolkit for simulation-based inference,"* Journal of Open Source Software **5**, 2505 (2020).

[72] P. Virtanen *et al.*, *"SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,"* Nature Methods **17**, 261 (2020).

[73] M. Waskom *et al.*, *"mwaskom/seaborn: v0.8.1 (september 2017),"* (2017).