Strong-Lensing Source Reconstruction with Denoising Diffusion Restoration Models

Konstantin Karchev
Theoretical and Scientific Data Science
SISSA, Trieste, Italy
kkarchev@sissa.it

Noemi Anau Montel
GRAPPA Institute
University of Amsterdam, The Netherlands
n.anaumontel@uva.nl

Adam Coogan
Ciela – Computation and Astrophysical Data Analysis Institute, Montréal, Quebec, Canada
Département de Physique, Université de Montréal, Canada
Mila – Quebec AI Institute, Montreal, Canada
adam.coogan@umontreal.ca

Christoph Weniger
GRAPPA Institute
University of Amsterdam, The Netherlands
c.weniger@uva.nl

Abstract

Analysis of galaxy–galaxy strong lensing systems is strongly dependent on any prior assumptions made about the appearance of the source. Here we present a method of imposing a data-driven prior / regularisation for source galaxies based on denoising diffusion probabilistic models (DDPMs). We use a pre-trained model for galaxy images, AstroDDPM, and a chain of conditional reconstruction steps called denoising diffusion restoration model (DDRM) to obtain samples consistent both with the noisy observation and with the distribution of training data for AstroDDPM. We show that these samples have the qualitative properties associated with the posterior for the source model: in a low-to-medium noise scenario they closely resemble the observation, while reconstructions from uncertain data show greater variability, consistent with the distribution encoded in the generative model used as prior.

1 Introduction

Gravitational lensing, the phenomenon of light bending trajectory under the influence of gravitating mass, has enabled progress in diverse areas of physics: from discovering some of the furthest observed galaxies in the Universe [1, 2] and analysing them [e.g. 3] to inferring the dark matter content of clusters and its distribution on galactic and sub-galactic scales [4–9], including detections of individual light dark matter halos without luminous counterparts [10–12], and measuring the Hubble constant [13, 14]. Critical in most endeavours is the ability to model the complex morphology of lensed sources, either as a goal in and of itself, or in order to disentangle their surface brightness inhomogeneity from perturbations in the lens.

Existing strong-lensing source models can be roughly classified in four categories with increasing complexity: analytic parametrisations like the Sérsic profile [15, 16]; regularised pixellation of the source plane [17–21] (where the regularisation can be implicit in the use of e.g. a Gaussian process prior [22] or continuous neural fields [23]); basis function regression onto e.g. wavelets

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

[24] or shapelets [25, 26]; and deep learning approaches with e.g. recurrent inference machines [27–29] or variational autoencoders [30]. While the former three categories are based on specific model assumptions, the deep-learning approaches are data-driven: it aims to learn from observations what typical galaxies look like and steer reconstructions appropriately. And while the current set of galaxy–galaxy strong lensing observations number on the order of hundreds, mainly coming from dedicated campaign like SLACS [31–33] and BELLS [34, 35], future general-purpose cosmological surveys are expected to deliver *hundreds of thousands* more [36], which underlines the need for fast and robust inference methodologies.

In this work we demonstrate galaxy–galaxy strong-lensing source reconstruction using denoising diffusion, the state-of-the-art deep-learning generative technique at the core of recent striking text-to-image models like DALL-E 2 [37]. The aim of any generative model (see e.g. Bond-Taylor et al. [38] for a "recent" review) is to learn from (usually very high-dimensional) data an approximation to the underlying distribution from which it has been drawn and enable easy sampling of new high-fidelity examples. Denoising diffusion probabilistic models (DDPMs), introduced by Sohl-Dickstein et al. [39] and elaborated by Ho et al. [40], achieve this by learning to reverse the gradual degradation of an input with random noise. By carefully designing both the noising and denoising processes, one can arrive at a particularly simple structure of the overall model, where a neural network (NN) is trained to predict the mean of a Gaussian used in denoising.

We use a DDPM pre-trained on galaxy images called AstroDDPM [41] and a modified sampling procedure called denoising diffusion restoration model (DDRM) [42] to condition the generation on a particular strong-lensing observation. We verify that this results in samples that exhibit desirable properties of the Bayesian posterior: when the noise in the observation is low, reconstructions follow it closely, while when noise is significant, samples are dictated by the data-driven prior encoded in AstroDDPM and show significant variation while still being consistent with the data. We expect the denoising diffusion approach to source reconstruction to prove instrumental in generating constrained training examples for simulation-based inference of dark matter substructure properties.

2 Background

2.1 Galaxy–galaxy strong gravitational lensing

Galaxy–galaxy strong gravitational lenses are usually modelled in the *thin-lens approximation* whereby all observed light is assumed to have originated from a specified *source plane* and been deflected by mass concentrated in a *lens* (or *image*) *plane* located between the source and the observer. Thin-lensing is entirely defined by the field of *deflection angles*, which is calculated from, and thus encodes, the mass distribution in the lens plane: see Meneghetti [43] for full details.

Importantly, gravitational lensing preserves surface brightness since it does not create or destroy photons, and so the observed flux in the image plane is simply the flux of the source at the origin of the ray. This means that lensing is a linear process, and source reconstruction can be phrased as a linear inversion problem, if the source is modelled on a (possibly irregular) grid, as recognised by [17, 18]. Since the grid can be made as fine as possible, while the observations have a fixed (usually coarse) resolution, and due to the almost complete degeneracy between lens and source, the *regularisation* and/or *Bayesian prior* on the source model has a crucial role both for the quality of the reconstruction, and for subsequent analysis performed with it (e.g. lens substructure inference).

Usually, the lensing configuration is a priori unknown (or weakly constrained by observations of the light of the lens galaxy) and often itself a target for inference. In this work, however, we focus only on source reconstruction, so we assume we know the details of the lens perfectly: i.e. we know both the mass configuration and the light of the lensing galaxy, which we can perfectly subtract from the observation. Our model, then, can be stated as

$$\mathbf{y}_{\text{obs}} = \mathbf{H}_{\text{lens}}\mathbf{x}_{\text{src}} + \mathbf{z}, \tag{1}$$

where $\mathbf{y}_{\text{obs}} \in \mathbb{R}^n$ is the observed image (flattened to a vector), $\mathbf{x}_{\text{src}} \in \mathbb{R}^m$ is the gridded source model, and $\mathbf{z}$ is observational noise, assumed i.i.d. Gaussian in each pixel: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma_y^2\mathbb{1})$. The n-by-m matrix $\mathbf{H}_{\text{lens}}$ encodes the lensing distortions, instrumental effects (such as a point-spread function), and interpolation of $\mathbf{x}_{\text{src}}$ across the grid on which it is defined. We use a ray-tracing code built with `PyTorch` in order to calculate $\mathbf{H}_{\text{lens}}$ with automatic differentiation of a forward simulation (since eq. (1) is linear, the particular values used for $\mathbf{x}_{\text{src}}$ in the forward pass are immaterial).

2.2 Denoising diffusion probabilistic models (DDPM)

DDPMs are a class of unsupervised density estimation techniques that aim to learn the underlying distribution $\mathrm{q}(\mathbf{x})$ of data $\{\mathbf{X}_i\}_{i=1}^N$ in a way that is then easy to sample from. They achieve this by introducing T *latent spaces* $\mathbf{x}_t$ for $t = 1 : T$, which are modelled in two ways: via a forward (diffusion) and a reverse (generative) processes. The forward process is a Markov chain that slowly adds Gaussian noise with increasing variance[1] β_t to the initial data point: $\mathrm{q}^{(\mathrm{t})}(\mathbf{x}_t \,|\, \mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbb{1})$, ending up with essentially pure noise. The model then learns the inverse (iteration-dependent) *denoising* operation $\mathrm{p}_\Theta^{(\mathrm{t})}(\mathbf{x}_t \,|\, \mathbf{x}_{t+1})$, which is usually again modelled as a Gaussian with pre-determined variance, whose mean is provided by a neural network $f_\Theta(\mathbf{x}_{t+1})$. Optimisation is performed with gradient ascent on the evidence lower bound (ELBO) of this model: a measure of the similarity over the training data between the forward and reverse distributions of the latent spaces:

$$\mathrm{q}(\mathbf{x}_{0:T}) = \mathrm{q}(\mathbf{x}_0) \prod_{t=1}^{T} \mathrm{q}^{(\mathrm{t})}(\mathbf{x}_t \,|\, \mathbf{x}_{t-1}) \qquad \leftrightarrow \qquad \mathrm{p}_\Theta(\mathbf{x}_{0:T}) = \mathrm{p}_\Theta^{(\mathrm{T})}(\mathbf{x}_T) \prod_{t=0}^{T-1} \mathrm{p}_\Theta^{(\mathrm{t})}(\mathbf{x}_t \,|\, \mathbf{x}_{t+1}). \quad (2)$$

While $\mathrm{q}(\mathbf{x}_0)$ is approximated with the training data, $\mathrm{p}_\Theta^{(\mathrm{T})}(\mathbf{x}_T)$ is set to a Gaussian with unit variance, so that one can draw pure noise and iteratively denoise it to obtain a new sample for $\mathbf{x}_0$.

2.3 DDPM as a prior: denoising diffusion restoration model (DDRM)

We would like to use the learnt approximation to $\mathrm{q}(\mathbf{x})$ as a prior for the linear inversion problem stated in eq. (1), i.e. sample from the posterior $\mathrm{p}(\mathbf{x} \,|\, \mathbf{y}) \propto \mathrm{q}(\mathbf{x})\, \mathrm{p}(\mathbf{y} \,|\, \mathbf{x})$ with a Gaussian likelihood $\mathrm{p}(\mathbf{y} \,|\, \mathbf{x}) = \mathcal{N}(\mathbf{y} \,|\, \mathbf{H}\mathbf{x}, \sigma_y^2 \mathbb{1})$. This can be achieved by conditioning $\mathrm{q}^{(\mathrm{t})}$ and $\mathrm{p}_\Theta^{(\mathrm{t})}$ on $\mathbf{y}$ and training bespoke density estimators (i.e. denoisers) for each observation. However, this is obviously very computationally expensive and does not scale to analyses of multiple systems that differ only by the lensing matrix $\mathbf{H}_{\text{lens}}$ but not by the assumed prior on $\mathbf{x}$.

Kawar et al. [42, 44] showed that, under certain conditions, a *pre-trained* DDPM model can be indeed used as a prior in a linear inversion model in order to sample from a *constrained* generative process $\mathrm{p}_\Theta(\mathbf{x}_{0:T} \,|\, \mathbf{y}) = \mathrm{p}_\Theta^{(\mathrm{T})}(\mathbf{x}_T \,|\, \mathbf{y}) \prod_{t=0}^{T-1} \mathrm{p}_\Theta^{(\mathrm{t})}(\mathbf{x}_t \,|\, \mathbf{x}_{t+1}, \mathbf{y})$. Their solution is expressed in the singular space of $\mathbf{H}$ and therefore starts with computing its singular value decomposition (SVD):

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^\top, \quad (3)$$

and applying the transformations $\bar{\mathbf{y}} \equiv \mathbf{S}^+\mathbf{U}^\top\mathbf{y}$ (where $\mathbf{S}^+$ is a Moore–Penrose pseudo-inverse) and $\bar{\mathbf{x}} \equiv \mathbf{V}^\top\mathbf{x}$. The sampling procedure then considers separately components which are constrained by the data—i.e. those that have a positive singular value $s_i > 0$—from those that are not ($s_i = 0$). Crucially, in the initial iterations, in which the noise level in the Markov chain is larger than the observational uncertainty, the denoising procedure is steered towards the observation with a weight $\eta_b = 2\sigma_t^2/(\sigma_t^2 + \sigma_y^2/s_i^2)$, where σ_t^2 is the accumulated noise variance at step t. At each step, the pre-trained denoiser f_Θ is only used to calculate the mean for the following step: $\mathbf{x}_{\Theta,t} = f_\Theta(\mathbf{x}_{t+1})$, which is then rotated into $\bar{\mathbf{x}}_{\Theta,t} \equiv \mathbf{V}^\top\mathbf{x}_{\Theta,t}$. DDRM has one hyperparameter, η, which relates to the specific way the denoising network has been trained and also influences the amount by which denoising is steered towards the observation. The specific form of the DDRM updates is given in eqs. (4) and (5) in appendix A.

3 Demonstration on mock data and discussion

In this section, we apply DDRM to realistic mock observations of galaxy–galaxy strong lensing. We use the AstroDDPM[2] network [41], pre-trained on the PROBES dataset [45, 46], which contains 1962 images of late-type galaxies that exhibit fine structure and details. We note that the PROBES dataset may not be representative of high-redshift source galaxies appearing in strong lenses, and so future analyses should check for possible biases due to the choice of training data. Since AstroDDPM is a multi-channel model (with channels corresponding to the g, r and z photometric bands), in order

[1] There are two conventions for scheduling the noise, termed "variance exploding" and "variance preserving". Here, we present the latter, which is used to train AstroDDPM, even though the DDRM implementation we use is variance exploding: for the conversion between the two, see Appendix B of Kawar et al. [42].

[2] `https://github.com/Smith42/astroddpm`, released under the AGPL-3.0 open-source license

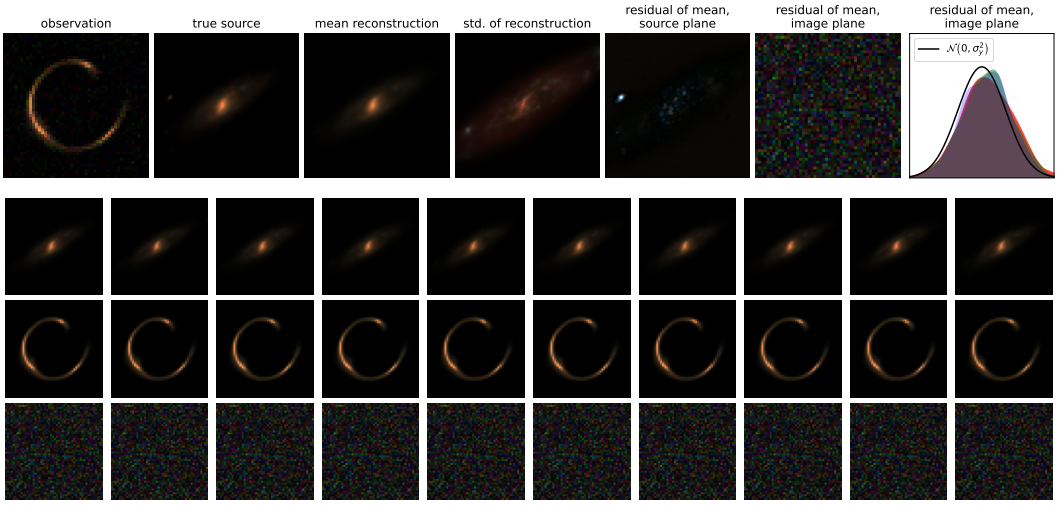

Figure 1: Top: from left to right, the mock observation, $\mathbf{y}$ (with a medium noise level), the true source, $\mathbf{x}$ (an unconstrained sample from AstroDDPM), the mean and standard deviation of 100 posterior samples from DDRM, $\mathbf{x}_{0,i} \sim \mathrm{p}_\Theta(\mathbf{x}_0 \,|\, \mathbf{y})$, and the residual of the mean with respect to the true source and with respect to the observation in the image plane; finally, a histogram of the latter compared to a Gaussian. Bottom: each column is a random posterior sample (top row), which is then lensed to produce the respective noiseless image $\mathbf{H}\mathbf{x}_{0,i}$ (middle row). Shown (bottom row) are also the residuals between $\mathbf{H}\mathbf{x}_{0,i}$ and the observation. In residual plots, negative values in one channel are shown as positive values in the other two (red $\leftrightarrow$ cyan, green $\leftrightarrow$ magenta, blue $\leftrightarrow$ yellow), considering complementary colors as "negative".

to avoid biases, we simulate and analyse multi-channel images as well, considering an independent likelihood as in eq. (1) for each channel. We note, still, that it is possible to include any linear operation on the channels (e.g. "selecting a channel" or "averaging channels") inside $\mathbf{H}_{\text{lens}}$.

We set up a uniform source-plane grid spanning $1'' \times 1''$ with $m = 256 \times 256$ pixels, to match AstroDDPM. We choose a resolution of $0.05''$ and $n = 50 \times 50$ pixels in the image plane, as appropriate for Hubble observations and set up a lens configuration so that multiple images are formed. In the end, only $\sim$1000 image pixels trace back to the source grid. As a source image we use an unconditioned sample from AstroDDPM.

To account for integrating the flux within an pixel, which is especially important in highly-magnified parts of the image, when calculating $\mathbf{H}_{\text{lens}}$, we simulate at a 10-times higher image resolution and then downsample with local averaging. As a cross-check, we have verified that multiplying $\mathbf{H}_{\text{lens}}\mathbf{x}_{\text{src}}$ matches the full simulation output to numerical precision. Finally, we add independent pixel noise. We test two settings: a "medium-noise" regime where σ_y is set to $1/30$ of the brightest image pixel, consistent with typical SLACS lenses; and a "high-noise" regime with peak signal-to-noise ratio of only 6 in order to verify that the DDRM produces enough variation when the observation is not very constraining.

Our main results are displayed in fig. 1. We set $\eta = 1$ in accordance with the theorem of Kawar et al. [42] and sample 100 realisations, which takes $\sim$10 min on an NVIDIA A-100 GPU. We verify that in this medium-noise setting, the true source is reconstructed with high fidelity even from only $\sim$1000 pixels, owing to the multiple observed projections and properly taking integration within a pixel into account. Standardised residuals between the lensed mean reconstruction and the observation follow a unit normal distribution and show no structure or signs of bias. Individual samples vary to a degree appropriate for the observational noise.

If the noise level is increased by a factor of 5, the reconstructions show accordingly higher variability (see fig. 2 in the appendix). Conditioned samples now follow more closely the prior and display a larger variety of morphologies, sizes and brightnesses, while still being consistent with the observation (the residuals of the mean are still approximately normal), although reconstructions seem to be slightly dimmer in the red (brightest) channel.

4 Conclusion

We have shown that one can use a pre-trained denoising diffusion model and the procedure in DDRM to reconstruct source galaxies from noisy strong gravitational lensing data with high fidelity. The reconstructions exhibit a qualitative variability necessary for them to be interpreted as samples from a posterior for the source's appearance, and we intend to perform quantitative tests, e.g. using the classification 2-sample test [47, 48], over a large number of mock observations to verify this. Such tests will also aid in setting the hyperparameter η of DDRM. In future work we will also unite DDRM source reconstruction with a scheme for inferring the mass distribution of the lens galaxy, which defines the distortion matrix $\mathbf{H}_{\text{lens}}$. Finally, our intended application for the method presented here is for generating training data for simulation-based inference of dark matter substructure, which will require an extension of the methodology to handle the correlated and spatially varying noise present in real lensing observations. We are confident that even in its present form, strong-lensing source reconstruction with DDRMs can be a useful tool for astrophysics and cosmology.

Broader Impact This work is focused on the precision analysis of strong gravitational lensing data via diffusion models, a class of generative models. Unfortunately, there are numerous well-known malicious uses of generative models (e.g. sample generation techniques can be employed to produce fake images and videos that can impact people's lives). On the other hand, through our analysis of source reconstruction in strong lensing, we have proven diffusion models to be useful for solving high-dimensional Bayesian inference problems thanks to their ability to capture the statistics of natural datasets. Although we do not anticipate potential for misuse of the presented application, the usual caution has to be exercised when drawing scientific conclusions based on a complex analysis machinery.

Acknowledgments and Disclosure of Funding

This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 864035 – Undark). A. C. acknowledges funding from the Schmidt Futures Foundation. K. K. acknowledges the hospitality of the Galileo Galilei Institute, Florence.

References

[1] Adi Zitrin, Ivo Labbé, Sirio Belli, Rychard Bouwens, Richard S. Ellis, Guido Roberts-Borsani, Daniel P. Stark, Pascal A. Oesch, and Renske Smit. Lymanα Emission from a Luminous z = 8.68 Galaxy: Implications for Galaxies as Tracers of Cosmic Reionization. *ApJL*, 810(1):L12, September 2015. doi: 10.1088/2041-8205/810/1/L12.

[2] Rohan P. Naidu, Pascal A. Oesch, Pieter van Dokkum, Erica J. Nelson, Katherine A. Suess, Katherine E. Whitaker, Natalie Allen, Rachel Bezanson, Rychard Bouwens, Gabriel Brammer, Charlie Conroy, Garth Illingworth, Ivo Labbe, Joel Leja, Ecaterina Leonova, Jorryt Matthee, Sedona H. Price, David J. Setton, Victoria Strait, Mauro Stefanon, Sandro Tacchella, Sune Toft, John R. Weaver, and Andrea Weibel. Two Remarkably Luminous Galaxy Candidates at $z \approx 11 - 13$ Revealed by JWST. *arXiv e-prints*, art. arXiv:2207.09434, July 2022.

[3] T. Treu, K. B. Schmidt, G. B. Brammer, B. Vulcani, X. Wang, M. Bradač, M. Dijkstra, A. Dressler, A. Fontana, R. Gavazzi, A. L. Henry, A. Hoag, K. H. Huang, T. A. Jones, P. L. Kelly, M. A. Malkan, C. Mason, L. Pentericci, B. Poggianti, M. Stiavelli, M. Trenti, and A. von der Linden. The Grism Lens-Amplified Survey from Space (GLASS). I. Survey Overview and First Data Release. *ApJ*, 812(2):114, October 2015. doi: 10.1088/0004-637X/812/2/114.

[4] N. Dalal and C. S. Kochanek. Direct Detection of Cold Dark Matter Substructure. *ApJ*, 572(1):25–33, June 2002. doi: 10.1086/340303.

[5] S. Vegetti, L. V. E. Koopmans, M. W. Auger, T. Treu, and A. S. Bolton. Inference of the cold dark matter substructure mass function at z = 0.2 using strong gravitational lenses. *MNRAS*, 442(3):2017–2035, August 2014. doi: 10.1093/mnras/stu943.

[6] Daniel Gilman, Simon Birrer, Anna Nierenberg, Tommaso Treu, Xiaolong Du, and Andrew 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. *MNRAS*, 491(4):6077–6101, February 2020. doi: 10.1093/mnras/stz3480.

[7] J. W. Hsueh, W. Enzi, S. Vegetti, M. W. Auger, C. D. Fassnacht, G. Despali, L. V. E. Koopmans, and J. P. McKean. SHARP - VII. New constraints on the dark matter free-streaming properties and substructure abundance from gravitationally lensed quasars. *MNRAS*, 492(2):3047–3059, February 2020. doi: 10.1093/mnras/stz3177.

[8] Massimo Meneghetti, Guido Davoli, Pietro Bergamini, Piero Rosati, Priyamvada Natarajan, Carlo Giocoli, Gabriel B. Caminha, R. Benton Metcalf, Elena Rasia, Stefano Borgani, Francesco Calura, Claudio Grillo, Amata Mercurio, and Eros Vanzella. An excess of small-scale gravitational lenses observed in galaxy clusters. *Science*, 369(6509):1347–1351, September 2020. doi: 10.1126/science.aax5164.

[9] James W. Nightingale, Qiuhan He, Xiaoyue Cao, Aristeidis Amvrosiadis, Amy Etherington, Carlos S. Frenk, Richard G. Hayes, Andrew Robertson, Shaun Cole, Samuel Lange, Ran Li, and Richard Massey. Scanning For Dark Matter Subhalos in Hubble Space Telescope Imaging of 54 Strong Lenses. *arXiv e-prints*, art. arXiv:2209.10566, September 2022.

[10] S. Vegetti, L. V. E. Koopmans, A. Bolton, T. Treu, and R. Gavazzi. Detection of a dark substructure through gravitational imaging. *Monthly Notices of the Royal Astronomical Society*, 408(4):1969–1981, Oct 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.16865.x. URL `http://dx.doi.org/10.1111/j.1365-2966.2010.16865.x`.

[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 galaxy at cosmological distance. *Nature*, 481(7381): 341–343, Jan 2012. ISSN 1476-4687. doi: 10.1038/nature10669. URL `http://dx.doi.org/10.1038/nature10669`.

[12] Yashar D. Hezaveh, Neal Dalal, Daniel P. Marrone, Yao-Yuan Mao, Warren Morningstar, Di Wen, Roger D. Bland ford, John E. Carlstrom, Christopher D. Fassnacht, Gilbert P. Holder, Athol Kemball, Philip J. Marshall, Norman Murray, Laurence Perreault Levasseur, Joaquin D. Vieira, and Risa H. Wechsler. Detection of Lensing Substructure Using ALMA Observations of the Dusty Galaxy SDP.81. *ApJ*, 823(1): 37, May 2016. doi: 10.3847/0004-637X/823/1/37.

[13] SH Suyu, S Huber, R Cañameras, M Kromer, S Schuldt, S Taubenberger, A Yıldırım, V Bonvin, JHH Chan, F Courbin, et al. Holismokes-i. highly optimised lensing investigations of supernovae, microlensing objects, and kinematics of ellipticals and spirals. *Astronomy & Astrophysics*, 644:A162, 2020.

[14] S Birrer, AJ Shajib, A Galan, M Millon, T Treu, A Agnello, M Auger, GC-F Chen, L Christensen, T Collett, et al. Tdcosmo-iv. hierarchical time-delay cosmography–joint inference of the hubble constant and galaxy density profiles. *Astronomy & Astrophysics*, 643:A165, 2020.

[15] J. L. Sérsic. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina*, 6:41–43, Feb 1963.

[16] Joel R Brownstein, Adam S Bolton, David J Schlegel, Daniel J Eisenstein, Christopher S Kochanek, Natalia Connolly, Claudia Maraston, Parul Pandey, Stella Seitz, David A Wake, et al. The boss emission-line lens survey (bells). i. a large spectroscopically selected sample of lens galaxies at redshift 0.5. *The Astrophysical Journal*, 744(1):41, 2011.

[17] S. J. Warren and S. Dye. Semilinear gravitational lens inversion. *The Astrophysical Journal*, 590:673–682, Jun 2003. ISSN 0004-637X. doi: 10/bws44f. URL `https://ui.adsabs.harvard.edu/abs/2003ApJ...590..673W`. ADS Bibcode: 2003ApJ...590..673W.

[18] S. H. Suyu, P. J. Marshall, M. P. Hobson, and R. D. Blandford. A bayesian analysis of regularized source inversions in gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 371:983–998, Sep 2006. ISSN 0035-8711. doi: 10/cgt4n6. URL `https://ui.adsabs.harvard.edu/abs/2006MNRAS.371..983S`. ADS Bibcode: 2006MNRAS.371..983S.

[19] S. Vegetti and L. V. E. Koopmans. Bayesian strong gravitational-lens modelling on adaptive grids: objective detection of mass substructure in galaxies. *Monthly Notices of the Royal Astronomical Society*, 392(3): 945–963, Jan 2009. doi: 10.1111/j.1365-2966.2008.14005.x.

[20] J. W. Nightingale and S. Dye. Adaptive semi-linear inversion of strong gravitational lens imaging. *Monthly Notices of the Royal Astronomical Society*, 452:2940–2959, Sep 2015. ISSN 0035-8711. doi: 10/f7q7xh. URL `https://ui.adsabs.harvard.edu/abs/2015MNRAS.452.2940N`. ADS Bibcode: 2015MNRAS.452.2940N.

[21] J. W. Nightingale, S. Dye, and Richard J. Massey. Autolens: automated modeling of a strong lens's light, mass, and source. *Monthly Notices of the Royal Astronomical Society*, 478:4738–4784, Aug 2018. ISSN 0035-8711. doi: 10/gd4sks. URL `https://ui.adsabs.harvard.edu/abs/2018MNRAS.478.4738N`. ADS Bibcode: 2018MNRAS.478.4738N.

[22] Konstantin Karchev, Adam Coogan, and Christoph Weniger. Strong-lensing source reconstruction with variationally optimized gaussian processes. *Monthly Notices of the Royal Astronomical Society*, 512: 661–685, May 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac311. URL `https://ui.adsabs.harvard.edu/abs/2022MNRAS.512..661K`. ADS Bibcode: 2022MNRAS.512..661K.

[23] Siddharth Mishra-Sharma and Ge Yang. Strong Lensing Source Reconstruction Using Continuous Neural Fields. *arXiv e-prints*, art. arXiv:2206.14820, June 2022.

[24] A. Galan, A. Peel, R. Joseph, F. Courbin, and J. L. Starck. SLITRONOMY: Towards a fully wavelet-based strong lensing inversion technique. *A&A*, 647:A176, March 2021. doi: 10.1051/0004-6361/202039363.

[25] Simon Birrer, Adam Amara, and Alexandre Refregier. Gravitational Lens Modeling with Basis Sets. *ApJ*, 813(2):102, November 2015. doi: 10.1088/0004-637X/813/2/102.

[26] Simon Birrer and Adam Amara. lenstronomy: Multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, 22:189–201, December 2018. doi: 10.1016/j.dark.2018.11.002.

[27] Warren R. Morningstar, Yashar D. Hezaveh, Laurence Perreault Levasseur, Roger D. Blandford, Philip J. Marshall, Patrick Putzky, and Risa H. Wechsler. Analyzing interferometric observations of strong gravitational lenses with recurrent and convolutional neural networks. *arXiv e-prints*, art. arXiv:1808.00011, July 2018.

[28] Warren R. Morningstar, Laurence Perreault Levasseur, Yashar D. Hezaveh, Roger Blandford, Phil Marshall, Patrick Putzky, Thomas D. Rueter, Risa Wechsler, and Max Welling. Data-driven Reconstruction of Gravitationally Lensed Galaxies Using Recurrent Inference Machines. *ApJ*, 883(1):14, September 2019. doi: 10.3847/1538-4357/ab35d7.

[29] Alexandre Adam, Laurence Perreault-Levasseur, and Yashar Hezaveh. Pixelated Reconstruction of Gravitational Lenses using Recurrent Inference Machines. *arXiv e-prints*, art. arXiv:2207.01073, July 2022.

[30] Marco Chianese, Adam Coogan, Paul Hofma, Sydney Otten, and Christoph Weniger. Differentiable strong lensing: uniting gravity and neural nets through differentiable probabilistic programming. *MNRAS*, 496(1): 381–393, May 2020. doi: 10.1093/mnras/staa1477.

[31] Adam S. Bolton, Scott Burles, Léon V. E. Koopmans, Tommaso Treu, and Leonidas A. Moustakas. The Sloan Lens ACS Survey. I. A Large Spectroscopically Selected Sample of Massive Early-Type Lens Galaxies. *ApJ*, 638(2):703–724, February 2006. doi: 10.1086/498884.

[32] Adam S. Bolton, Scott Burles, Léon V. E. Koopmans, Tommaso Treu, Raphaël Gavazzi, Leonidas A. Moustakas, Randall Wayth, and David J. Schlegel. The Sloan Lens ACS Survey. V. The Full ACS Strong-Lens Sample. *ApJ*, 682(2):964–984, August 2008. doi: 10.1086/589327.

[33] Yiping Shu, Joel R. Brownstein, Adam S. Bolton, Léon V. E. Koopmans, Tommaso Treu, Antonio D. Montero-Dorta, Matthew W. Auger, Oliver Czoske, Raphaël Gavazzi, Philip J. Marshall, and Leonidas A. Moustakas. The Sloan Lens ACS Survey. XIII. Discovery of 40 New Galaxy-scale Strong Lenses. *ApJ*, 851(1):48, December 2017. doi: 10.3847/1538-4357/aa9794.

[34] Joel R. Brownstein, Adam S. Bolton, David J. Schlegel, Daniel J. Eisenstein, Christopher S. Kochanek, Natalia Connolly, Claudia Maraston, Parul Pandey, Stella Seitz, David A. Wake, W. Michael Wood-Vasey, Jon Brinkmann, Donald P. Schneider, and Benjamin A. Weaver. The BOSS Emission-Line Lens Survey (BELLS). I. A Large Spectroscopically Selected Sample of Lens Galaxies at Redshift $\sim$0.5. *ApJ*, 744(1): 41, January 2012. doi: 10.1088/0004-637X/744/1/41.

[35] Yiping Shu, Adam S. Bolton, Christopher S. Kochanek, Masamune Oguri, Ismael Pérez-Fournon, Zheng Zheng, Shude Mao, Antonio D. Montero-Dorta, Joel R. Brownstein, Rui Marques-Chaves, and Brice Ménard. The BOSS Emission-line Lens Survey. III. Strong Lensing of Lyα Emitters by Individual Galaxies. *ApJ*, 824(2):86, June 2016. doi: 10.3847/0004-637X/824/2/86.

[36] Thomas E. Collett. The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *ApJ*, 811(1):20, September 2015. doi: 10.1088/0004-637X/811/1/20.

[37] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv e-prints*, art. arXiv:2204.06125, April 2022.

[38] Sam Bond-Taylor, Adam Leach, Yang Long, and Chris G. Willcocks. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. *arXiv e-prints*, art. arXiv:2103.04922, March 2021.

[39] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics, 2015. URL `https://arxiv.org/abs/1503.03585`.

[40] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. URL `https://arxiv.org/abs/2006.11239`.

[41] Michael J Smith, James E Geach, Ryan A Jackson, Nikhil Arora, Connor Stone, and Stéphane Courteau. Realistic galaxy image simulation via score-based generative models. *Monthly Notices of the Royal Astronomical Society*, 511(2):1808–1818, jan 2022. doi: 10.1093/mnras/stac130. URL `https://doi.org/10.1093%2Fmnras%2Fstac130`.

[42] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models, 2022. URL https://arxiv.org/abs/2201.11793.

[43] Massimo Meneghetti. *Introduction to Gravitational Lensing*, volume 956 of *Lecture Notes in Physics*. Springer Nature, 2021. ISBN 978-3-030-73581-4. URL https://link.springer.com/book/10.1007/978-3-030-73582-1.

[44] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. In *Advances in Neural Information Processing Systems*, Oct 2021. URL https://openreview.net/forum?id=pBKOx_dxYAN.

[45] Connor Stone and Sté phane Courteau. The intrinsic scatter of the radial acceleration relation. *The Astrophysical Journal*, 882(1):6, aug 2019. doi: 10.3847/1538-4357/ab3126. URL https://doi.org/10.3847%2F1538-4357%2Fab3126.

[46] Connor Stone, Sté phane Courteau, and Nikhil Arora. The intrinsic scatter of galaxy scaling relations. *The Astrophysical Journal*, 912(1):41, may 2021. doi: 10.3847/1538-4357/abebe4. URL https://doi.org/10.3847%2F1538-4357%2Fabebe4.

[47] Jerome H. Friedman. On multivariate goodness of fit and two sample testing. *eConf*, C030908:THPD002, 2003.

[48] David Lopez-Paz and Maxime Oquab. Revisiting classifier two-sample tests. In *International Conference on Learning Representations*, 2017.

A Appendix: denoising diffusion restoration model (DDRM)

Starting with the SVD of $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ and the transformed observation $\bar{\mathbf{y}} \equiv \mathbf{S}^+\mathbf{U}^\top\mathbf{y}$ (with $^+$ a Moore-Penrose pseudo inverse), DDRM consists of applying the following updates:

$$\mathrm{p}^{(\mathrm{T})}_\Theta(\bar{\mathbf{x}}^{(i)}_T \,|\, \mathbf{y}) \qquad = \begin{cases} \mathcal{N}\big(0, \sigma_T^2\big) & \text{if } s_i = 0, \\ \mathcal{N}\Big(\bar{\mathbf{y}}^{(i)}, \sigma_T^2 - \frac{\sigma_y^2}{s_i^2}\Big) & \text{if } s_i > 0; \end{cases} \tag{4}$$

$$\mathrm{p}^{(\mathrm{t})}_\Theta(\bar{\mathbf{x}}^{(i)}_t \,|\, \mathbf{x}_{t+1}, \mathbf{y}) = \begin{cases} \mathcal{N}\Big(\bar{\mathbf{x}}^{(i)}_{\Theta,t} + \sqrt{1-\eta^2}\frac{\sigma_t}{\sigma_{t+1}}\Big(\bar{\mathbf{x}}^{(i)}_{t+1} - \bar{\mathbf{x}}^{(i)}_{\Theta,t}\Big), \eta^2\sigma_t^2\Big) & \text{if } s_i = 0, \\ \mathcal{N}\Big(\bar{\mathbf{x}}^{(i)}_{\Theta,t} + \sqrt{1-\eta^2}\frac{\sigma_t}{\sigma_y/s_i}\Big(\bar{\mathbf{y}}^{(i)} - \bar{\mathbf{x}}^{(i)}_{\Theta,t}\Big), \eta^2\sigma_t^2\Big) & \text{if } \sigma_t < \sigma_y/s_i, \\ \mathcal{N}\Big((1-\eta_b)\bar{\mathbf{x}}^{(i)}_{\Theta,t} + \eta_b\bar{\mathbf{y}}^{(i)}, \sigma_t^2 - \eta_b^2\frac{\sigma_y^2}{s_i^2}\Big) & \text{if } \sigma_t \geq \sigma_y/s_i. \end{cases} \tag{5}$$

Here $\cdot^{(i)}$ labels the i^{th} component of a vector. At the beginning of each iteration, the current transformed prediction $\bar{\mathbf{x}}_{t+1}$ is de-rotated into $\mathbf{x}_{t+1} = \mathbf{V}\bar{\mathbf{x}}_{t+1}$, which is then denoised: $\mathbf{x}_{\Theta,t} = f_\Theta(\mathbf{x}_{t+1})$, and rotated back into $\bar{\mathbf{x}}_{\Theta,t} \equiv \mathbf{V}^\top\mathbf{x}_{\Theta,t}$.

Equation (5) allows for controlling the relative information content carried by noise versus that encoded in the network: when $\eta = 1$, unconstrained pixels (first case) are sampled independently at each denoising step, whereas setting $\eta = 0$ connects them deterministically to the initial noise realisation. Furthermore, in high-noise scenarios, which correspond to the second case of eq. (5), η controls how strongly denoising is steered towards the particular observation, with low values leading to stronger conditioning.

Kawar et al. [42] prove the equivalence of the DDRM and DDPM ELBO objectives, which allows one to use a pre-trained unconditioned DDPM model as a denoiser in DDRM, under the condition $\eta = 1$. They show that for other choices of η (and even of η_b, which may also be considered a hyperparameter), the objectives remain similar, so *approximate* DDRM can still be performed.

In figs. 3 and 4 we briefly explore the effect setting a low $\eta = 0.03$ has on reconstructions. In the medium-noise scenario, fig. 3, results are similar to using $\eta = 1$, but now the generative process produces artefacts like spots, which are common in unconditioned AstroDDPM samples (see fig. 2 of Smith et al. [41]) but unwarranted by data. In the high-noise setting, fig. 4, residuals are much improved from the case of $\eta = 1$ due to the stronger conditioning on the observation. These qualitative tests show the importance of tuning η so as to match the regime in which AstroDDPM has been trained (in fact, Smith et al. [41] use the equivalence between DDPM and score-matching described in Ho et al. [40] to train their model, so they do not need to explicitly set a parameter like η). We plan to optimise η in the future by quantitatively measuring the quality of posterior samples with the classifier 2-sample test [47, 48].

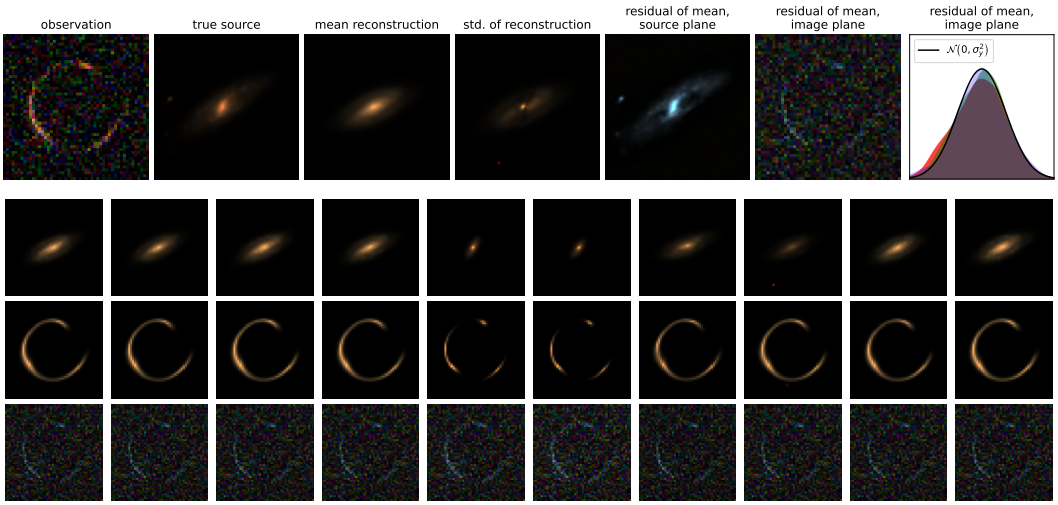

Figure 2: Same as fig. 1, but with the high noise setting (peak signal-to-noise ratio 6).

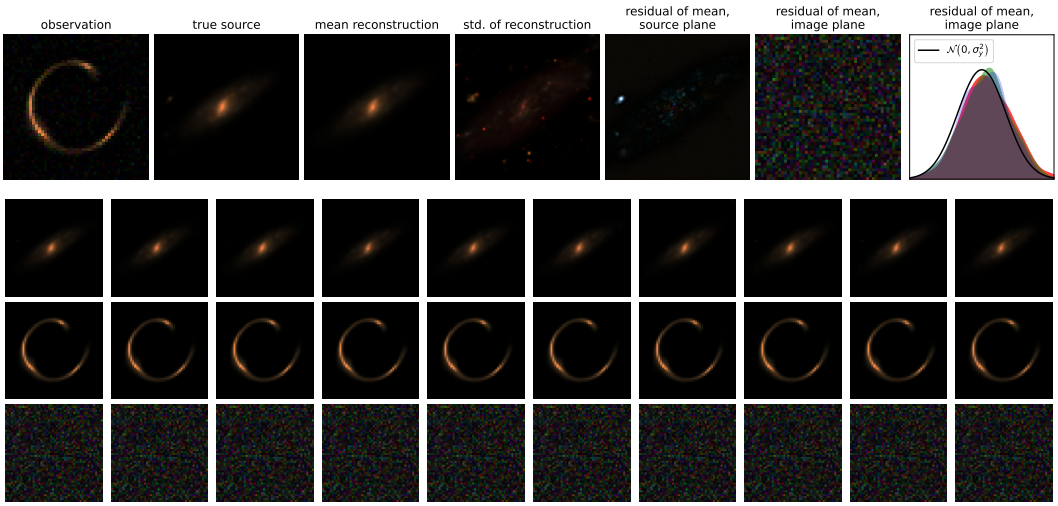

Figure 3: Same as fig. 1 (medium-noise setting), but inference has been performed with $\eta = 0.03$.

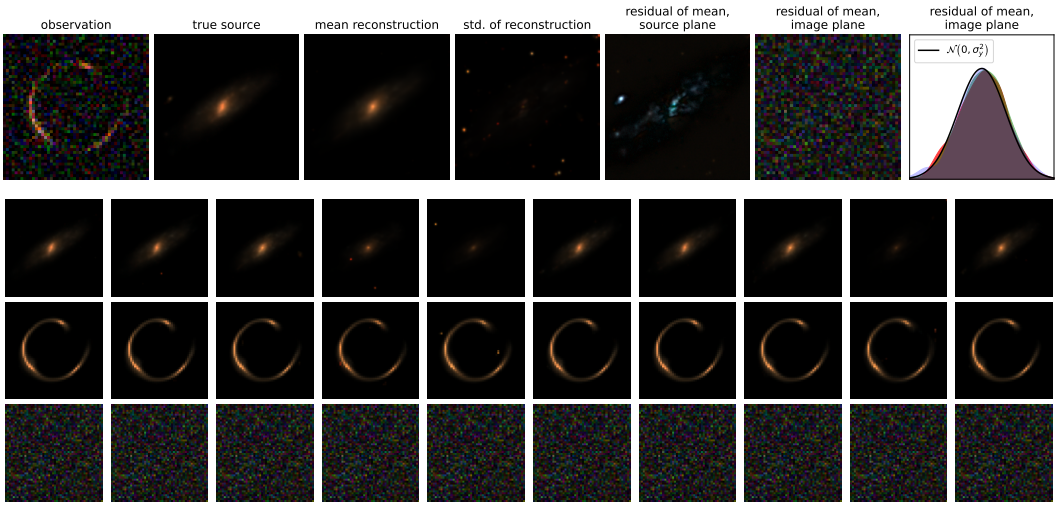

Figure 4: Same as fig. 2 (high-noise setting), but inference has been performed with $\eta = 0.03$.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] We have explored different regimes (medium and high-noise scenarios) and the effect of the only hyperparameter, η. We have presented work-in-progress results with qualitative descriptions and described how we plan to go about quantifying performance in the future.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] We explain in the impact statement that we show an application to astronomy but realise that high-fidelity generative models can have serious impact on individuals and societies through fake content.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] Our experiments were run using publicly available models and code, while the lensing framework we have used is not yet public, although we are working on releasing it.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] We investigate in some detail variance in the samples we obtain in the figures and in section 3
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] in section 3
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [No] We have not produced any new (useful and hard to obtain) assets
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] We only use explicitly public models and open-source code.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] We only use synthetic data coming from a generative model of astronomical observations of galaxies.
5. If you used crowdsourcing or conducted research with human subjects... [No] We did not.
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]