High-Fidelity Reconstructions of Strong Lenses in the Data-Driven Generative Modeling Era

Anonymous Author(s)
Affiliation
Address
email

Abstract

We achieve state-of-the-art reconstructions of strong gravitational lensing systems from the Sloan Lens ACS (SLACS) survey by leveraging score-based diffusion models as high-dimensional priors over major components of the lensing system: the background source, foreground lens light, and point-spread function (PSF). Our approach produces high-resolution models that substantially reduce residuals compared to previous lens modeling attempts. To our knowledge, this is the first application of data-driven generative priors to real strong lensing observations, establishing a new benchmark for precision lens modeling in preparation for upcoming large-scale imaging surveys.

1 Introduction

Strong gravitational lensing, in which the light from a distant galaxy is deflected and magnified by a foreground mass, is a cornerstone tool for precision cosmology [Wong et al., 2019, Shajib et al., 2020, Qi et al., 2021, Colaço et al., 2025], dark-matter studies [Mao and Schneider, 1998, Vegetti et al., 2010, Hezaveh et al., 2016, Gilman et al., 2019, Powell et al., 2023, Ballard et al., 2024, Lange et al., 2025], and probing galaxy structure at high redshift [Coe et al., 2012, Fan et al., 2019, Shajib, Anowar J. et al., 2023]. Unlocking this potential requires lens modeling—inferring the physical components of a lensing system, including the mass distribution of the lens galaxy and the light profiles of both the lens and background source.

Traditional lens modeling has long been constrained by computational cost and simplifying assumptions. Parametric models impose rigid functional forms, while handcrafted pixel-based priors often fail to capture the true morphological complexity of galaxies [Bolton et al., 2008, Savary, E. et al., 2022, Knabel et al., 2023, Huang et al., 2025, Cao et al., 2025]. These limitations can prevent accurate modeling of the data and bias the inferred lens parameters [Xu et al., 2015, Sonnenfeld, 2018, Galan, A. et al., 2024, Ballard et al., 2024].

Recent advances in generative modeling provide a powerful alternative. Data-driven approaches, such as diffusion models [Song and Ermon, 2019, Ho et al., 2020, Song et al., 2020, 2021, Yang et al., 2022], can learn rich, high-dimensional priors directly from data, enabling reconstructions that incorporate realistic astrophysical structure [Dia et al., 2023, Adam et al., 2023, Legin et al., 2023b, Bourdin et al., 2024, Cuesta-Lazaro and Mishra-Sharma, 2024, Adam et al., 2025, Barco et al., 2025, Boruah et al., 2025].

In this work, we revisit high-quality strong lensing observations from the SLACS survey and achieve state-of-the-art lens models by incorporating high-dimensional, data-driven priors into the lens modeling process. We focus on galaxy-galaxy strong lensing systems, jointly inferring the mass distribution of the foreground lens, the light profiles of both the lens and background source galaxies, and instrumental components such as the PSF of the telescope.

2 Methods

Lens Data

We model 30 strong gravitational lensing systems from the SLACS survey [Bolton et al., 2008], observed with the Hubble Space Telescope (HST) in the F814W filter (program ID 10886, 10174, 10587) available from the Mikulski Archive for Space Telescopes (MAST)[1]. For each system, we use four dithered FLT exposures and extract 128×128 pixel cutouts centered on the lens galaxy as our data for lens modeling, denoted $\mathbf{D} = \{\mathbf{d}_i\}_{i=1}^4$.

Lens Modeling

For each lensing system, we jointly infer the lens light, lens mass, source, per-exposure PSFs, and alignment shifts. We denote the full set of parameters as

$$\mathbf{\Theta} = \{\mathbf{L}, \mathbf{M}, \mathbf{S}\}, \quad \mathbf{P} = \{\mathbf{P}_i\}_{i=1}^4, \quad \boldsymbol{\delta} = \{\boldsymbol{\delta}_i\}_{i=1}^4,$$

where $\mathbf{L}$, $\mathbf{M}$, and $\mathbf{S}$ describe the lens light, lens mass, and source, while $\mathbf{P}_i$ and $\boldsymbol{\delta}_i$ denote the PSF and alignment shift for each exposure i.

The goal is to sample from the posterior distribution

$$p(\boldsymbol{\delta}, \mathbf{P}, \mathbf{\Theta} \mid \mathbf{D}) \propto \prod_{i=1}^{4} p(\mathbf{d}_i \mid \boldsymbol{\delta}_i, \mathbf{P}_i, \mathbf{\Theta})\, p(\boldsymbol{\delta}_i)\, p(\mathbf{P}_i)\, p(\mathbf{L}) p(\mathbf{M}) p(\mathbf{S}), \tag{1}$$

where the per-exposure likelihood $p(\mathbf{d}_i \mid \boldsymbol{\delta}_i, \mathbf{P}_i, \mathbf{\Theta})$ is assumed Gaussian with pixel-wise variance obtained from the FITS file `ERR` arrays. Cosmic rays and other artifacts are masked using data-quality `DQ` flags and manual inspection.

The lens mass $\mathbf{M}$ is represented by a parametric model consisting of an Elliptical Power-Law (EPL) profile (Einstein radius R_{E}, axis ratio q, orientation ϕ_{EPL}, and slope t), an external shear (amplitude γ_{ext} and orientation ϕ_{ext}), and multipoles of order $m = 3, 4$, with amplitudes a_3, a_4 and orientations ϕ_3, ϕ_4 [Tessore, Nicolas and Benton Metcalf, R., 2015, Xu et al., 2015, Meneghetti, 2022]. We adopt broad uniform priors for all parameters, except $t \sim \mathcal{N}(1.0, 0.2)$ and $a_3, a_4 \sim \mathcal{N}(0, 0.01)$. Alignment shifts $\boldsymbol{\delta}$ follow uniform priors over the full field of view of the data cutouts.

We model the lens light $\mathbf{L}$, source $\mathbf{S}$, and PSFs $\mathbf{P}$ as pixelated images (256^2, 256^2, and 128^2 pixels, respectively). Bilinear interpolation is used to evaluate $\mathbf{S}$ and $\mathbf{L}$ at arbitrary coordinates when performing lens raytracing simulations.

Data-Driven Priors

We train score-based diffusion models [Song et al., 2020] with the `scoremodels` package [Adam, 2025], adopting NCSN++ U-Net architectures to learn high-dimensional priors for the source $\mathbf{S}$, lens light $\mathbf{L}$, and PSF $\mathbf{P}$. The source prior was trained on ∼2,000 galaxies from the PROBES dataset [Stone and Courteau, 2019, Stone et al., 2021] for 1000 epochs with a batch size of 4, while the lens light prior was trained for 24 hours with a batch size of 16 on on-the-fly simulations containing 1 to 5 Sérsic components [Sérsic, 1963]. The PSF prior was trained for 50 epochs with a batch size of 32 on ∼10,000 upsampled empirical PSFs derived from HST/ACS WFC star cutouts [Anderson and King, 2006]. For all models, we adopt a Variance Exploding SDE with a geometric noise schedule, using $(\sigma_{\min}, \sigma_{\max}) = (10^{-5}, 500)$ for $\mathbf{S}$, $(10^{-5}, 200)$ for $\mathbf{L}$, and $(10^{-3}, 100)$ for $\mathbf{P}$. All models were trained on NVIDIA A100 40 GB GPUs.

Joint Posterior Sampling

We draw samples from the posterior distribution in Equation 1 using a Gibbs sampling scheme, cycling through the components one at a time while holding the others fixed:

$$\mathbf{S} \;\to\; (\mathbf{M}, \boldsymbol{\delta}) \;\to\; \mathbf{L} \;\to\; \mathbf{P}.$$

The lens mass $\mathbf{M}$ and alignment shifts $\boldsymbol{\delta}$ are sampled jointly using the Metropolis-Adjusted Langevin Algorithm (MALA) [Roberts and Tweedie, 1996]. The source $\mathbf{S}$, lens light $\mathbf{L}$, and PSFs $\mathbf{P}$ are sampled

[1] https://mast.stsci.edu

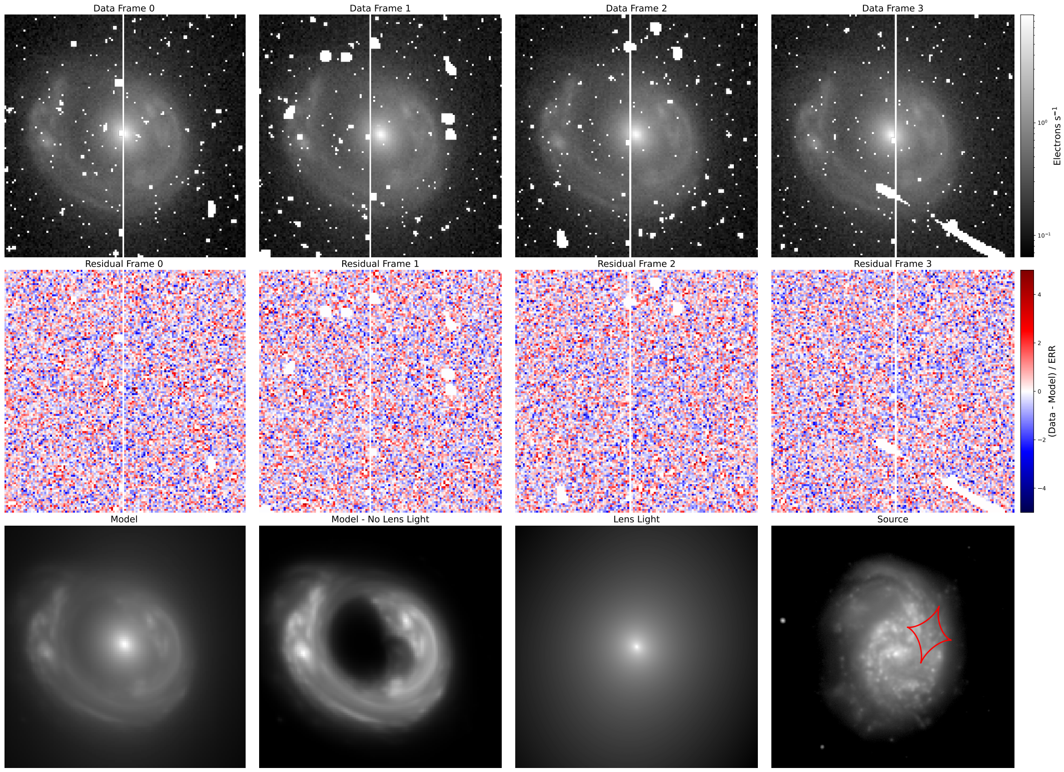

Figure 1: Lens model of SDSSJ1430+4105. From top to bottom: the four data exposures (log-scaled), normalized data-model residuals per exposure with colormap centered and clipped at ± 5 (5σ); and a summary of the lens model. The bottom row shows, from left to right, the lens model, the lens model without lens light, the lens light model, and the reconstructed background source (all in log-scale). The red curve indicates the lensing caustic overlaid on the source model: source features inside the caustic are lensed into four observable images, while those outside produce only two.

by solving the reverse-time posterior SDE under the convolved Gaussian likelihood approximation following Adam et al. [2022]

The PSF models $\mathbf{P}$ are treated separately from the physical lens model: for each exposure, $\mathbf{P}_i$ is sampled from cutouts of stars within the same data frame, excluding the region containing the main lensing system. Therefore, our PSF posterior sampling is conditionally independent of the physical lens model parameters $\mathbf{\Theta} = \{\mathbf{L}, \mathbf{M}, \mathbf{S}\}$.

Our Gibbs procedure proceeds in three stages: **(1) Initialization:** a short chain to move the sampler toward a region of high posterior probability; **(2) Annealed sampling:** 100 Gibbs steps (20 burn-in) with inflated likelihood variance to encourage exploration; **(3) Fine-tuning:** starting from each annealed sample, we run additional Gibbs updates with the true likelihood to obtain posterior samples without annealing. The full Gibbs procedure for one lensing system requires approximately 160 GPU hours on an NVIDIA A100 40 GB GPU. For the first stage, we initialize the foreground mass model parameters $\mathbf{M}$ using the results of Bolton et al. [2008].

3 Results and Discussion

For each strong lensing system, we present data–model residuals from posterior samples (Figure 1, Appendix 3). A gallery of sampled sources across the SLACS lensing systems is shown in Figure 2. We also present full joint posterior statistics of SDSSJ1430+4105 in Figure 3 and for a subset of lenses, foreground mass-model parameter uncertainties in Table 1 marginalized over $\mathbf{S}$, $\mathbf{L}$, $\mathbf{P}$, and $\boldsymbol{\delta}$.

The results demonstrate that the vast majority of signal in Hubble-resolution data can be effectively modeled by combining parametric lens mass models with high-dimensional, data-driven

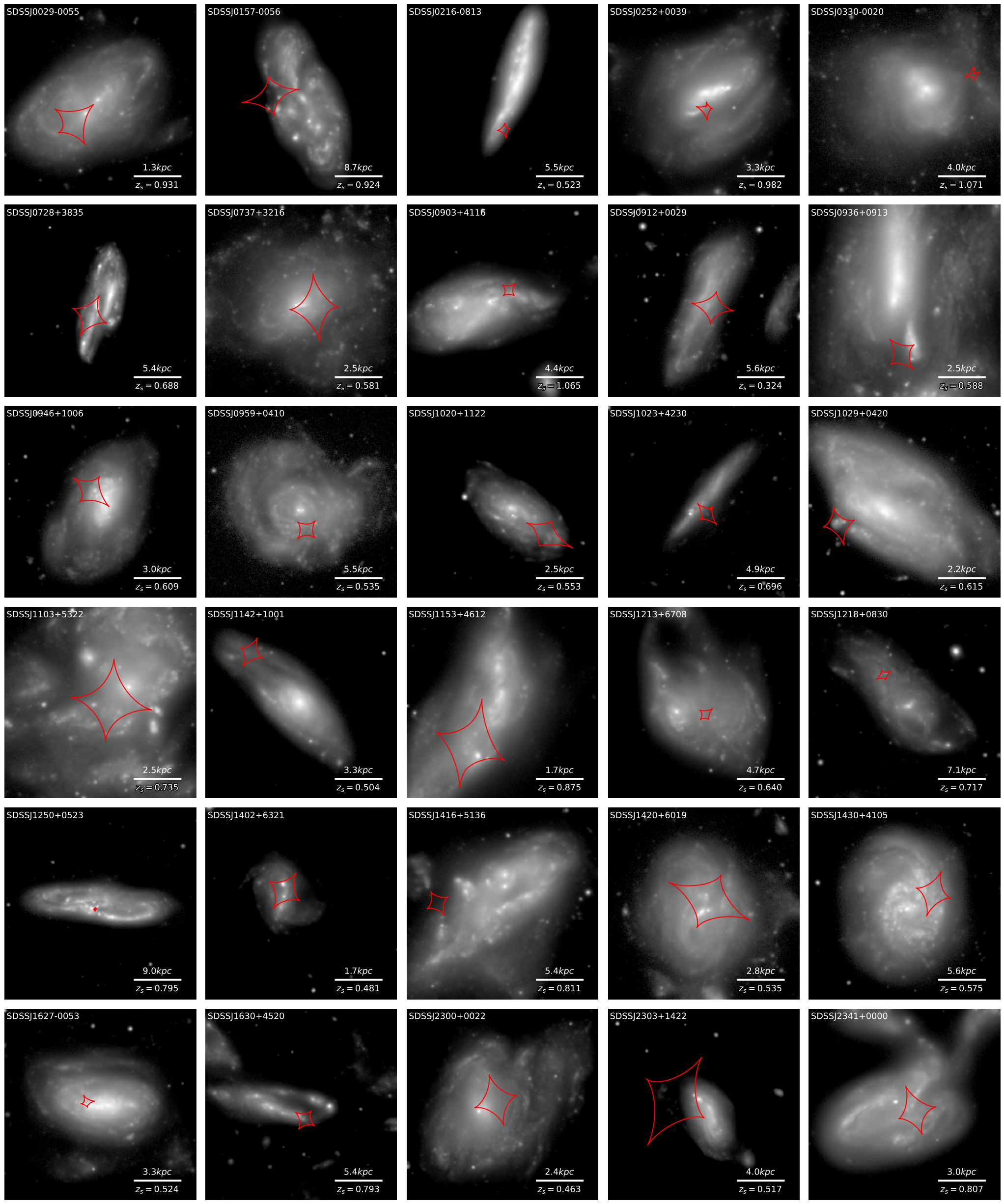

Figure 2: Example posterior sample of the background source for representative lensing systems. Each panel shows the reconstructed source with the lensing caustic overlaid, along with the name of the lensing system, source redshift and a physical reference scale in kiloparsecs (kpc). The physical scale is computed assuming a flat ΛCDM cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

prior models for the source, lens light, and PSF. Nonetheless, there are a few current limitations. First, a few systems, in particular SDSSJ2341+0000 (Figure 32), SDSSJ1103+5322 (Figure 19) and SDSSJ0959+0410 (Figure 15), show residuals traceable to the lens light prior trained largely on parametric Sérsic profiles, indicating the need for training on real or simulated galaxies with richer morphologies. Second, our Gaussian likelihood ignores known non-Gaussian HST noise, so integrating learned noise statistics (e.g., score-based, Legin et al. [2023a]) into the likelihood model would yield improved residuals.

References

Alexandre Adam. score_models: Score-based generative models (python package). https://github.com/AlexandreAdam/score_models, 2025. Accessed: 2025-07-21.

Alexandre Adam, Adam Coogan, Nikolay Malkin, Ronan Legin, Laurence Perreault-Levasseur, Yashar Hezaveh, and Yoshua Bengio. Posterior samples of source galaxies in strong gravitational lenses with score-based priors, 2022. URL https://arxiv.org/abs/2211.03812.

Alexandre Adam, Laurence Perreault-Levasseur, Yashar Hezaveh, and Max Welling. Pixelated reconstruction of foreground density and background surface brightness in gravitational lensing systems using recurrent inference machines. *The Astrophysical Journal*, 951(1):6, jun 2023. doi: 10.3847/1538-4357/accf84. URL https://dx.doi.org/10.3847/1538-4357/accf84.

Alexandre Adam, Connor Stone, Connor Bottrell, Ronan Legin, Yashar Hezaveh, and Laurence Perreaul-Levasseur. Echoes in the noise: Posterior samples of faint galaxy surface brightness profiles with score-based likelihoods and priors. *The Astronomical Journal*, 169(5):254, apr 2025. doi: 10.3847/1538-3881/adb039. URL https://dx.doi.org/10.3847/1538-3881/adb039.

Jay Anderson and Ivan R. King. PSFs, Photometry, and Astronomy for the ACS/WFC. Instrument Science Report ACS 2006-01, 34 pages, February 2006.

Daniel J. Ballard, Wolfgang J. R. Enzi, Thomas E. Collett, Hannah C. Turner, and Russell J. Smith. Gravitational imaging through a triple source plane lens: revisiting the ΛCDM-defying dark subhalo in SDSSJ0946+1006. , 528(4):7564–7586, March 2024. doi: 10.1093/mnras/stae514.

Gabriel Missael Barco, Alexandre Adam, Connor Stone, Yashar Hezaveh, and Laurence Perreault-Levasseur. Tackling the problem of distributional shifts: Correcting misspecified, high-dimensional data-driven priors for inverse problems. *The Astrophysical Journal*, 980(1):108, feb 2025. doi: 10.3847/1538-4357/ad9b92. URL https://dx.doi.org/10.3847/1538-4357/ad9b92.

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*. *The Astrophysical Journal*, 682(2):964, aug 2008. doi: 10.1086/589327. URL https://dx.doi.org/10.1086/589327.

Supranta S. Boruah, Michael Jacob, and Bhuvnesh Jain. Diffusion-based mass map reconstruction from weak lensing data, 2025. URL https://arxiv.org/abs/2502.04158.

Antoine Bourdin, Ronan Legin, Matthew Ho, Alexandre Adam, Yashar Hezaveh, and Laurence Perreault-Levasseur. Inpainting galaxy counts onto n-body simulations over multiple cosmologies and astrophysics, 2024. URL https://arxiv.org/abs/2408.00839.

Xiaoyue Cao, Ran Li, Nan Li, Yun Chen, Rui Li, Huanyuan Shan, and Tian Li. Csst strong lensing preparation: fast modelling of galaxy–galaxy strong lenses in the big data era. *Monthly Notices of the Royal Astronomical Society*, 540(4):3121–3134, 06 2025. ISSN 0035-8711. doi: 10.1093/mnras/staf891. URL https://doi.org/10.1093/mnras/staf891.

Dan Coe, Adi Zitrin, Mauricio Carrasco, Xinwen Shu, Wei Zheng, Marc Postman, Larry Bradley, Anton Koekemoer, Rychard Bouwens, Tom Broadhurst, Anna Monna, Ole Host, Leonidas A. Moustakas, Holland Ford, John Moustakas, Arjen van der Wel, Megan Donahue, Steven A. Rodney, Narciso Benítez, Stephanie Jouvel, Stella Seitz, Daniel D. Kelson, and Piero Rosati. Clash: Three strongly lensed images of a candidate z 11 galaxy. *The Astrophysical Journal*, 762(1):32, dec 2012. doi: 10.1088/0004-637X/762/1/32. URL https://dx.doi.org/10.1088/0004-637X/762/1/32.

L. R. Colaço, R. F. L. Holanda, Z. C. Santana, and R. Silva. A joint analysis of strong lensing and type ia supernovae to determine the hubble constant. *The European Physical Journal C*, 85(5), May 2025. ISSN 1434-6052. doi: 10.1140/epjc/s10052-025-14315-6. URL http://dx.doi.org/10.1140/epjc/s10052-025-14315-6.

Carolina Cuesta-Lazaro and Siddharth Mishra-Sharma. A point cloud approach to generative modeling for galaxy surveys at the field level, 2024. URL https://arxiv.org/abs/2311.17141.

Noe Dia, M. J. Yantovski-Barth, Alexandre Adam, Micah Bowles, Pablo Lemos, Anna M. M. Scaife, Yashar Hezaveh, and Laurence Perreault-Levasseur. Bayesian imaging for radio interferometry with score-based priors, 2023. URL https://arxiv.org/abs/2311.18012.

Xiaohui Fan, Feige Wang, Jinyi Yang, Charles R. Keeton, Minghao Yue, Ann Zabludoff, Fuyan Bian, Marco Bonaglia, Iskren Y. Georgiev, Joseph F. Hennawi, Jiangtao Li, Ian D. McGreer, Rohan Naidu, Fabio Pacucci, Sebastian Rabien, David Thompson, Bram Venemans, Fabian Walter, Ran Wang, and Xue-Bing Wu. The discovery of a gravitationally lensed quasar at z = 6.51. *The Astrophysical Journal Letters*, 870(2):L11, jan 2019. doi: 10.3847/2041-8213/aaeffe. URL `https://dx.doi.org/10.3847/2041-8213/aaeffe`.

Galan, A., Vernardos, G., Minor, Q., Sluse, D., Van de Vyvere, L., and Gomer, M. Exploiting the diversity of modeling methods to probe systematic biases in strong lensing analyses. *AA*, 692:A87, 2024. doi: 10.1051/0004-6361/202451095. URL `https://doi.org/10.1051/0004-6361/202451095`.

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. *Monthly Notices of the Royal Astronomical Society*, 491(4): 6077–6101, 12 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz3480. URL `https://doi.org/10.1093/mnras/stz3480`.

Yashar D. Hezaveh, Neal Dalal, Daniel P. Marrone, Yao-Yuan Mao, Warren Morningstar, Di Wen, Roger D. Blandford, 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. *The Astrophysical Journal*, 823(1):37, may 2016. doi: 10.3847/0004-637X/823/1/37. URL `https://dx.doi.org/10.3847/0004-637X/823/1/37`.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf`.

X. Huang, S. Baltasar, N. Ratier-Werbin, C. Storfer, W. Sheu, S. Agarwal, M. Tamargo-Arizmendi, D. J. Schlegel, J. Aguilar, S. Ahlen, G. Aldering, S. Banka, S. BenZvi, D. Bianchi, A. Bolton, D. Brooks, A. Cikota, T. Claybaugh, A. de la Macorra, A. Dey, P. Doel, J. Edelstein, A. Filipp, J. E. Forero-Romero, E. Gaztanaga, S. Gontcho A Gontcho, A. Gu, G. Gutierrez, K. Honscheid, E. Jullo, S. Juneau, R. Kehoe, D. Kirkby, T. Kisner, A. Kremin, K. J. Kwon, A. Lambert, M. Landriau, D. Lang, L. Le Guillou, J. Liu, A. Meisner, R. Miquel, J. Moustakas, A. D. Myers, S. Perlmutter, I. Perez-Rafols, F. Prada, G. Rossi, D. Rubin, E. Sanchez, M. Schubnell, Y. Shu, E. Silver, D. Sprayberry, N. Suzuki, G. Tarle, B. A. Weaver, and H. Zou. Desi strong lens foundry i: Hst observations and modeling with giga-lens, 2025. URL `https://arxiv.org/abs/2502.03455`.

Shawn Knabel, B W Holwerda, J Nightingale, T Treu, M Bilicki, S Brough, S Driver, L Finnerty, L Haberzettl, S Hegde, A M Hopkins, K Kuijken, J Liske, K A Pimbblet, R C Steele, and A H Wright. Modelling strong lenses from wide-field ground-based observations in kids and gama. *Monthly Notices of the Royal Astronomical Society*, 520(1):804–827, 01 2023. ISSN 0035-8711. doi: 10.1093/mnras/stad133. URL `https://doi.org/10.1093/mnras/stad133`.

Samuel C Lange, Aristeidis Amvrosiadis, James W Nightingale, Qiuhan He, Carlos S Frenk, Andrew Robertson, Shaun Cole, Richard Massey, Xiaoyue Cao, Ran Li, and Kaihao Wang. Galaxy mass modelling from multiwavelength jwst strong lens analysis: dark matter substructure, angular mass complexity, or both? *Monthly Notices of the Royal Astronomical Society*, 539(2):704–726, 03 2025. ISSN 0035-8711. doi: 10.1093/mnras/staf491. URL `https://doi.org/10.1093/mnras/staf491`.

Ronan Legin, Alexandre Adam, Yashar Hezaveh, and Laurence Perreault-Levasseur. Beyond gaussian noise: A generalized approach to likelihood analysis with non-gaussian noise. *The Astrophysical Journal Letters*, 949 (2):L41, jun 2023a. doi: 10.3847/2041-8213/acd645. URL `https://dx.doi.org/10.3847/2041-8213/acd645`.

Ronan Legin, Matthew Ho, Pablo Lemos, Laurence Perreault-Levasseur, Shirley Ho, Yashar Hezaveh, and Benjamin Wandelt. Posterior sampling of the initial conditions of the universe from non-linear large scale structures using score-based generative models. *Monthly Notices of the Royal Astronomical Society: Letters*, 527(1):L173–L178, 10 2023b. ISSN 1745-3925. doi: 10.1093/mnrasl/slad152. URL `https://doi.org/10.1093/mnrasl/slad152`.

Shude Mao and Peter Schneider. Evidence for substructure in lens galaxies? *Monthly Notices of the Royal Astronomical Society*, 295(3):587–594, 04 1998. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1998.01319.x. URL `https://doi.org/10.1046/j.1365-8711.1998.01319.x`.

Massimo Meneghetti. *Introduction to Gravitational Lensing: With Python Examples*. 2022.

Devon M Powell, Simona Vegetti, J P McKean, Simon D M White, Elisa G M Ferreira, Simon May, and Cristiana Spingola. A lensed radio jet at milli-arcsecond resolution – ii. constraints on fuzzy dark matter from an extended gravitational arc. *Monthly Notices of the Royal Astronomical Society: Letters*, 524(1): L84–L88, 06 2023. ISSN 1745-3925. doi: 10.1093/mnrasl/slad074. URL `https://doi.org/10.1093/mnrasl/slad074`.

Jing-Zhao Qi, Jia-Wei Zhao, Shuo Cao, Marek Biesiada, and Yuting Liu. Measurements of the hubble constant and cosmic curvature with quasars: ultracompact radio structure and strong gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 503(2):2179–2186, 03 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab638. URL `https://doi.org/10.1093/mnras/stab638`.

Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. ISSN 13507265. URL `http://www.jstor.org/stable/3318418`.

Savary, E., Rojas, K., Maus, M., Clément, B., Courbin, F., Gavazzi, R., Chan, J. H. H., Lemon, C., Vernardos, G., Cañameras, R., Schuldt, S., Suyu, S. H., Cuillandre, J.-C., Fabbro, S., Gwyn, S., Hudson, M. J., Kilbinger, M., Scott, D., and Stone, C. Strong lensing in unions: Toward a pipeline from discovery to modeling. *AA*, 666:A1, 2022. doi: 10.1051/0004-6361/202142505. URL `https://doi.org/10.1051/0004-6361/202142505`.

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, February 1963.

A J Shajib, S Birrer, T Treu, A Agnello, E J Buckley-Geer, J H H Chan, L Christensen, C Lemon, H Lin, M Millon, J Poh, C E Rusu, D Sluse, C Spiniello, G C-F Chen, T Collett, F Courbin, C D Fassnacht, J Frieman, A Galan, D Gilman, A More, T Anguita, M W Auger, V Bonvin, R McMahon, G Meylan, K C Wong, T M C Abbott, J Annis, S Avila, K Bechtol, D Brooks, D Brout, D L Burke, A Carnero Rosell, M Carrasco Kind, J Carretero, F J Castander, M Costanzi, L N da Costa, J De Vicente, S Desai, J P Dietrich, P Doel, A Drlica-Wagner, A E Evrard, D A Finley, B Flaugher, P Fosalba, J García-Bellido, D W Gerdes, D Gruen, R A Gruendl, J Gschwend, G Gutierrez, D L Hollowood, K Honscheid, D Huterer, D J James, T Jeltema, E Krause, N Kuropatkin, T S Li, M Lima, N MacCrann, M A G Maia, J L Marshall, P Melchior, R Miquel, R L C Ogando, A Palmese, F Paz-Chinchón, A A Plazas, A K Romer, A Roodman, M Sako, E Sanchez, B Santiago, V Scarpine, M Schubnell, D Scolnic, S Serrano, I Sevilla-Noarbe, M Smith, M Soares-Santos, E Suchyta, G Tarle, D Thomas, A R Walker, and Y Zhang. Strides: a 3.9 per cent measurement of the hubble constant from the strong lens system des j04085354. *Monthly Notices of the Royal Astronomical Society*, 494(4):6072–6102, 03 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa828. URL `https://doi.org/10.1093/mnras/staa828`.

Shajib, Anowar J., Mozumdar, Pritom, Chen, Geoff C.-F., Treu, Tommaso, Cappellari, Michele, Knabel, Shawn, Suyu, Sherry H., Bennert, Vardha N., Frieman, Joshua A., Sluse, Dominique, Birrer, Simon, Courbin, Frederic, Fassnacht, Christopher D., Villafaña, Lizvette, and Williams, Peter R. Tdcosmo - xii. improved hubble constant measurement from lensing time delays using spatially resolved stellar kinematics of the lens galaxy. *AA*, 673:A9, 2023. doi: 10.1051/0004-6361/202345878. URL `https://doi.org/10.1051/0004-6361/202345878`.

Yang Song and Stefano Ermon. Generative Modeling by Estimating Gradients of the Data Distribution. *arXiv e-prints*, art. arXiv:1907.05600, July 2019. doi: 10.48550/arXiv.1907.05600.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv e-prints*, art. arXiv:2011.13456, November 2020. doi: 10.48550/arXiv.2011.13456.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving Inverse Problems in Medical Imaging with Score-Based Generative Models. *arXiv e-prints*, art. arXiv:2111.08005, November 2021. doi: 10.48550/arXiv.2111.08005.

Alessandro Sonnenfeld. On the choice of lens density profile in time delay cosmography. , 474(4):4648–4659, March 2018. doi: 10.1093/mnras/stx3105.

Connor Stone and Stéphane 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://dx.doi.org/10.3847/1538-4357/ab3126`.

Connor Stone, Stéphane 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://dx.doi.org/10.3847/1538-4357/abebe4`.

Tessore, Nicolas and Benton Metcalf, R. The elliptical power law profile lens. *AA*, 580:A79, 2015. doi: 10.1051/0004-6361/201526773. URL https://doi.org/10.1051/0004-6361/201526773.

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, 10 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.16865.x. URL https://doi.org/10.1111/j.1365-2966.2010.16865.x.

Kenneth C Wong, Sherry H Suyu, Geoff C-F Chen, Cristian E Rusu, Martin Millon, Dominique Sluse, Vivien Bonvin, Christopher D Fassnacht, Stefan Taubenberger, Matthew W Auger, Simon Birrer, James H H Chan, Frederic Courbin, Stefan Hilbert, Olga Tihhonova, Tommaso Treu, Adriano Agnello, Xuheng Ding, Inh Jee, Eiichiro Komatsu, Anowar J Shajib, Alessandro Sonnenfeld, Roger D Blandford, Léon V E Koopmans, Philip J Marshall, and Georges Meylan. H0licow – xiii. a 2.4 per cent measurement of h0 from lensed quasars: 5.3 tension between early- and late-universe probes. *Monthly Notices of the Royal Astronomical Society*, 498(1):1420–1439, 09 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz3094. URL https://doi.org/10.1093/mnras/stz3094.

Dandan Xu, Dominique Sluse, Liang Gao, Jie Wang, Carlos Frenk, Shude Mao, Peter Schneider, and Volker Springel. How well can cold dark matter substructures account for the observed radio flux-ratio anomalies. *Monthly Notices of the Royal Astronomical Society*, 447(4):3189–3206, 01 2015. ISSN 0035-8711. doi: 10.1093/mnras/stu2673. URL https://doi.org/10.1093/mnras/stu2673.

Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion Models: A Comprehensive Survey of Methods and Applications. *arXiv e-prints*, art. arXiv:2209.00796, September 2022. doi: 10.48550/arXiv.2209.00796.

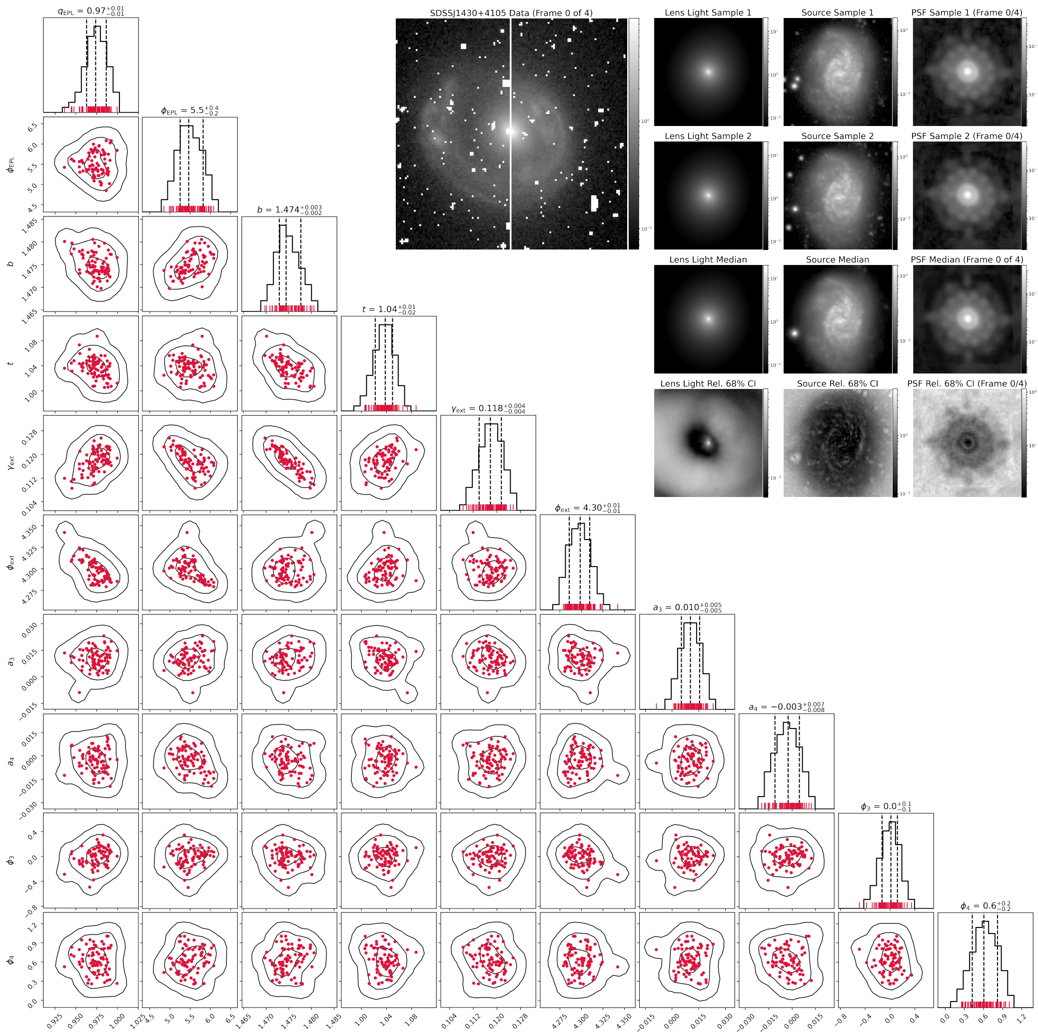

Figure 3: Joint posterior distribution for SDSSJ1430+4105. The corner plot shows posterior samples (red) of the foreground mass parameters $\mathbf{M}$, with contours (black) marking the 1σ, 2σ and 3σ regions. The 1-D corner plot marginals display the posterior median of the foreground mass parameters $\mathbf{M}$ with 16th–84th percentile uncertainties, indicated by dashed vertical lines and reported above each panel. Posterior samples and uncertainties for the source $\mathbf{S}$, lens light $\mathbf{L}$, and PSF model $\mathbf{P}_0$ (frame 0 of 4) are also shown. The upper-right panels display, from top to bottom, two posterior samples, the median model, and the relative 68% credible interval (interval width divided by the median). All results are based on 80 posterior samples and are marginalized over alignment parameters $\boldsymbol{\delta}$.

System Name	q	$\phi_{\rm EPL}$	b	t	γ	$\phi_{\rm ext}$	a_3	a_4
SDSSJ0029−0055	$0.475^{+0.067}_{-0.045}$	$1.641^{+0.080}_{-0.100}$	$0.954^{+0.003}_{-0.004}$	$1.802^{+0.056}_{-0.053}$	$0.053^{+0.010}_{-0.006}$	$2.361^{+0.129}_{-0.085}$	$0.008^{+0.010}_{-0.006}$	$0.010^{+0.008}_{-0.007}$
SDSSJ0157−0056	$0.445^{+0.078}_{-0.063}$	$5.324^{+0.111}_{-0.144}$	$0.761^{+0.170}_{-0.032}$	$1.207^{+0.089}_{-0.233}$	$0.277^{+0.024}_{-0.060}$	$5.204^{+0.157}_{-0.125}$	$0.009^{+0.011}_{-0.005}$	$0.009^{+0.013}_{-0.006}$
SDSSJ0216−0813	$0.766^{+0.100}_{-0.098}$	$5.605^{+0.168}_{-0.074}$	$1.120^{+0.011}_{-0.016}$	$1.349^{+0.075}_{-0.106}$	$0.165^{+0.020}_{-0.016}$	$2.295^{+0.017}_{-0.016}$	$0.034^{+0.011}_{-0.011}$	$0.030^{+0.010}_{-0.010}$
SDSSJ0252+0039	$0.911^{+0.015}_{-0.015}$	$5.791^{+0.042}_{-0.028}$	$1.030^{+0.002}_{-0.002}$	$0.301^{+0.112}_{-0.073}$	$0.035^{+0.006}_{-0.005}$	$5.742^{+0.041}_{-0.075}$	$0.003^{+0.002}_{-0.002}$	$0.007^{+0.003}_{-0.002}$
SDSSJ0330−0020	$0.898^{+0.024}_{-0.022}$	$6.090^{+0.070}_{-0.045}$	$1.134^{+0.004}_{-0.004}$	$1.224^{+0.056}_{-0.042}$	$0.003^{+0.007}_{-0.003}$	$6.243^{+0.173}_{-0.189}$	$0.007^{+0.007}_{-0.005}$	$0.006^{+0.008}_{-0.005}$
SDSSJ0728+3835	$0.765^{+0.047}_{-0.043}$	$0.210^{+0.032}_{-0.065}$	$1.196^{+0.012}_{-0.014}$	$0.891^{+0.093}_{-0.126}$	$0.015^{+0.019}_{-0.013}$	$5.067^{+0.128}_{-0.510}$	$0.006^{+0.007}_{-0.003}$	$0.009^{+0.008}_{-0.006}$
SDSSJ0737+3216	$0.849^{+0.020}_{-0.022}$	$6.124^{+0.078}_{-0.085}$	$0.966^{+0.004}_{-0.002}$	$1.341^{+0.066}_{-0.065}$	$0.139^{+0.011}_{-0.010}$	$4.550^{+0.016}_{-0.015}$	$0.013^{+0.005}_{-0.005}$	$0.031^{+0.005}_{-0.004}$
SDSSJ0903+4116	$0.792^{+0.033}_{-0.028}$	$4.758^{+0.130}_{-0.084}$	$1.296^{+0.005}_{-0.004}$	$1.266^{+0.056}_{-0.064}$	$0.036^{+0.010}_{-0.017}$	$4.427^{+0.189}_{-0.117}$	$0.017^{+0.007}_{-0.008}$	$0.008^{+0.006}_{-0.006}$
SDSSJ0912+0029	$0.889^{+0.027}_{-0.171}$	$0.812^{+0.372}_{-0.356}$	$1.276^{+0.133}_{-0.195}$	$0.115^{+0.019}_{-0.011}$	$0.050^{+0.103}_{-0.025}$	$1.002^{+0.226}_{-0.125}$	$0.025^{+0.009}_{-0.010}$	$0.014^{+0.005}_{-0.011}$
SDSSJ0936+0913	$0.340^{+0.121}_{-0.090}$	$0.291^{+0.115}_{-0.072}$	$0.548^{+0.335}_{-0.071}$	$0.669^{+0.128}_{-0.126}$	$0.069^{+0.050}_{-0.030}$	$2.091^{+0.447}_{-0.219}$	$0.013^{+0.007}_{-0.009}$	$0.009^{+0.009}_{-0.006}$
SDSSJ0946+1006	$0.853^{+0.013}_{-0.013}$	$4.339^{+0.033}_{-0.036}$	$1.413^{+0.002}_{-0.002}$	$1.351^{+0.086}_{-0.086}$	$0.083^{+0.005}_{-0.006}$	$3.742^{+0.022}_{-0.021}$	$0.043^{+0.008}_{-0.007}$	$0.021^{+0.008}_{-0.009}$
SDSSJ1020+1122	$0.675^{+0.222}_{-0.278}$	$6.080^{+0.807}_{-2.299}$	$1.049^{+0.040}_{-0.031}$	$1.688^{+0.122}_{-0.173}$	$0.172^{+0.063}_{-0.055}$	$3.042^{+0.148}_{-0.239}$	$0.015^{+0.010}_{-0.011}$	$0.015^{+0.011}_{-0.010}$
SDSSJ1023+4230	$0.769^{+0.039}_{-0.034}$	$4.697^{+0.074}_{-0.076}$	$1.417^{+0.005}_{-0.004}$	$0.785^{+0.205}_{-0.259}$	$0.065^{+0.014}_{-0.011}$	$4.589^{+0.051}_{-0.114}$	$0.007^{+0.007}_{-0.005}$	$0.046^{+0.013}_{-0.013}$
SDSSJ1029+0420	$0.350^{+0.034}_{-0.107}$	$5.629^{+0.139}_{-0.320}$	$0.952^{+0.070}_{-0.099}$	$0.790^{+0.266}_{-0.096}$	$0.188^{+0.043}_{-0.107}$	$6.040^{+0.242}_{-0.242}$	$0.007^{+0.009}_{-0.005}$	$0.005^{+0.009}_{-0.002}$
SDSSJ1213+6708	$0.459^{+0.049}_{-0.095}$	$1.076^{+0.058}_{-0.044}$	$0.986^{+0.136}_{-0.237}$	$1.410^{+0.143}_{-0.194}$	$0.163^{+0.011}_{-0.010}$	$4.176^{+0.070}_{-0.028}$	$0.008^{+0.010}_{-0.005}$	$0.035^{+0.019}_{-0.026}$
SDSSJ1218+0830	$0.651^{+0.054}_{-0.096}$	$0.999^{+0.105}_{-0.066}$	$1.694^{+0.016}_{-0.013}$	$0.856^{+0.048}_{-0.048}$	$0.157^{+0.034}_{-0.019}$	$0.972^{+0.079}_{-0.058}$	$0.016^{+0.017}_{-0.012}$	$0.076^{+0.013}_{-0.019}$
SDSSJ1250+0523	$0.971^{+0.022}_{-0.049}$	$3.406^{+2.514}_{-1.025}$	$1.140^{+0.002}_{-0.003}$	$1.411^{+0.091}_{-0.177}$	$0.012^{+0.005}_{-0.005}$	$4.338^{+0.131}_{-0.271}$	$0.005^{+0.007}_{-0.004}$	$0.042^{+0.010}_{-0.011}$
SDSSJ1402+6321	$0.731^{+0.035}_{-0.035}$	$0.464^{+0.043}_{-0.041}$	$1.368^{+0.003}_{-0.004}$	$1.087^{+0.099}_{-0.138}$	$0.011^{+0.008}_{-0.009}$	$0.776^{+0.797}_{-0.429}$	$0.007^{+0.008}_{-0.004}$	$0.010^{+0.007}_{-0.007}$
SDSSJ1416+5136	$0.831^{+0.055}_{-0.049}$	$1.467^{+0.200}_{-0.154}$	$1.316^{+0.007}_{-0.018}$	$0.983^{+0.056}_{-0.034}$	$0.044^{+0.013}_{-0.014}$	$5.573^{+0.270}_{-0.183}$	$0.010^{+0.007}_{-0.006}$	$0.012^{+0.010}_{-0.008}$
SDSSJ1420+6019	$0.456^{+0.010}_{-0.014}$	$5.921^{+0.009}_{-0.014}$	$1.144^{+0.006}_{-0.009}$	$0.448^{+0.065}_{-0.026}$	$0.254^{+0.009}_{-0.025}$	$5.924^{+0.012}_{-0.018}$	$0.009^{+0.004}_{-0.003}$	$0.015^{+0.006}_{-0.004}$
SDSSJ1430+4105	$0.974^{+0.013}_{-0.011}$	$5.466^{+0.360}_{-0.212}$	$1.474^{+0.003}_{-0.002}$	$1.038^{+0.012}_{-0.016}$	$0.118^{+0.004}_{-0.004}$	$4.299^{+0.011}_{-0.013}$	$0.010^{+0.005}_{-0.005}$	$0.006^{+0.006}_{-0.004}$
SDSSJ1627−0053	$0.761^{+0.026}_{-0.107}$	$1.372^{+0.187}_{-0.033}$	$1.233^{+0.007}_{-0.003}$	$1.578^{+0.133}_{-0.558}$	$0.001^{+0.055}_{-0.001}$	$0.556^{+1.084}_{-0.544}$	$0.012^{+0.011}_{-0.008}$	$0.011^{+0.006}_{-0.007}$
SDSSJ1630+4520	$0.756^{+0.010}_{-0.014}$	$0.198^{+0.033}_{-0.022}$	$1.828^{+0.003}_{-0.005}$	$0.868^{+0.079}_{-0.060}$	$0.062^{+0.005}_{-0.006}$	$0.148^{+0.056}_{-0.031}$	$0.023^{+0.012}_{-0.016}$	$0.036^{+0.006}_{-0.011}$
SDSSJ2300+0022	$0.669^{+0.032}_{-0.055}$	$0.006^{+0.102}_{-0.061}$	$1.296^{+0.009}_{-0.005}$	$1.312^{+0.113}_{-0.088}$	$0.019^{+0.014}_{-0.016}$	$2.711^{+1.228}_{-2.560}$	$0.010^{+0.008}_{-0.008}$	$0.008^{+0.007}_{-0.005}$
SDSSJ2303+1422	$0.661^{+0.023}_{-0.026}$	$1.208^{+0.039}_{-0.038}$	$1.644^{+0.009}_{-0.007}$	$0.839^{+0.090}_{-0.069}$	$0.069^{+0.013}_{-0.010}$	$1.737^{+0.098}_{-0.074}$	$0.037^{+0.008}_{-0.006}$	$0.016^{+0.008}_{-0.012}$
SDSSJ2341+0000	$0.878^{+0.080}_{-0.080}$	$4.815^{+1.516}_{-0.194}$	$1.284^{+0.005}_{-0.007}$	$1.342^{+0.098}_{-0.143}$	$0.100^{+0.025}_{-0.022}$	$4.963^{+0.121}_{-0.031}$	$0.058^{+0.013}_{-0.008}$	$0.041^{+0.011}_{-0.022}$

Table 1: Summary of foreground mass lens parameters $\mathbf{M}$ marginalized over the source $\mathbf{S}$, lens light $\mathbf{L}$, the set of per-frame PSF models $\mathbf{P}$, and relative frame shifts $\boldsymbol{\delta}$. Reported values are posterior medians with 68% credible intervals.

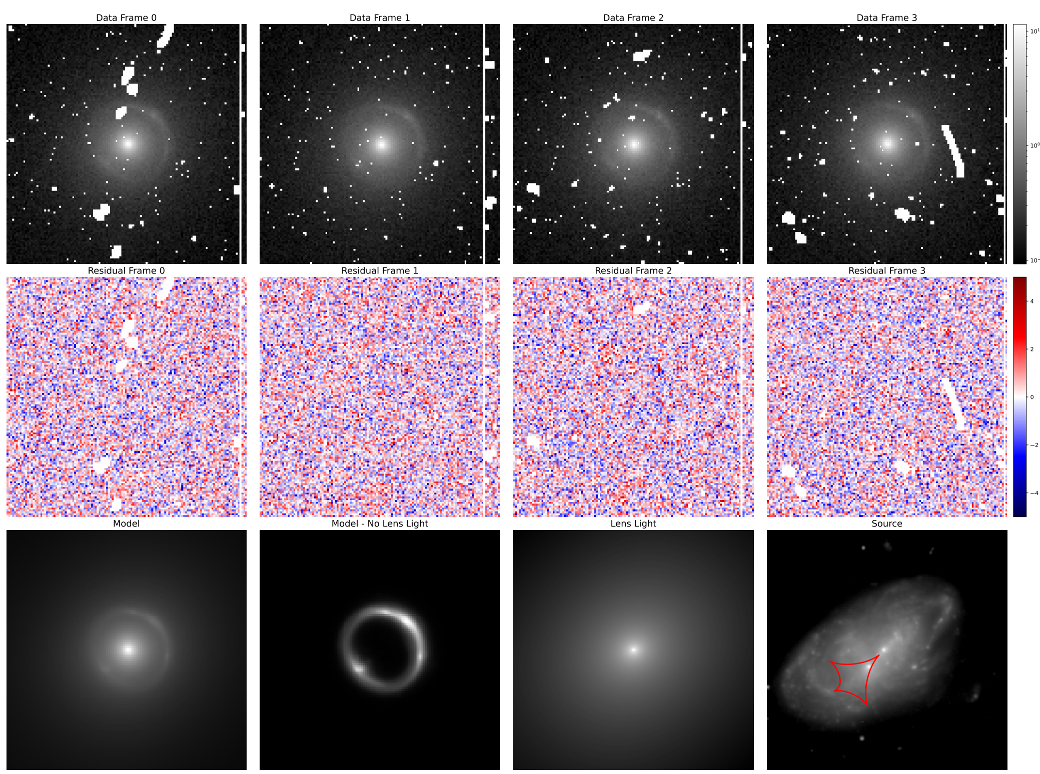

Figure 4: Same as Fig. 1, but for SDSSJ0029-0055.

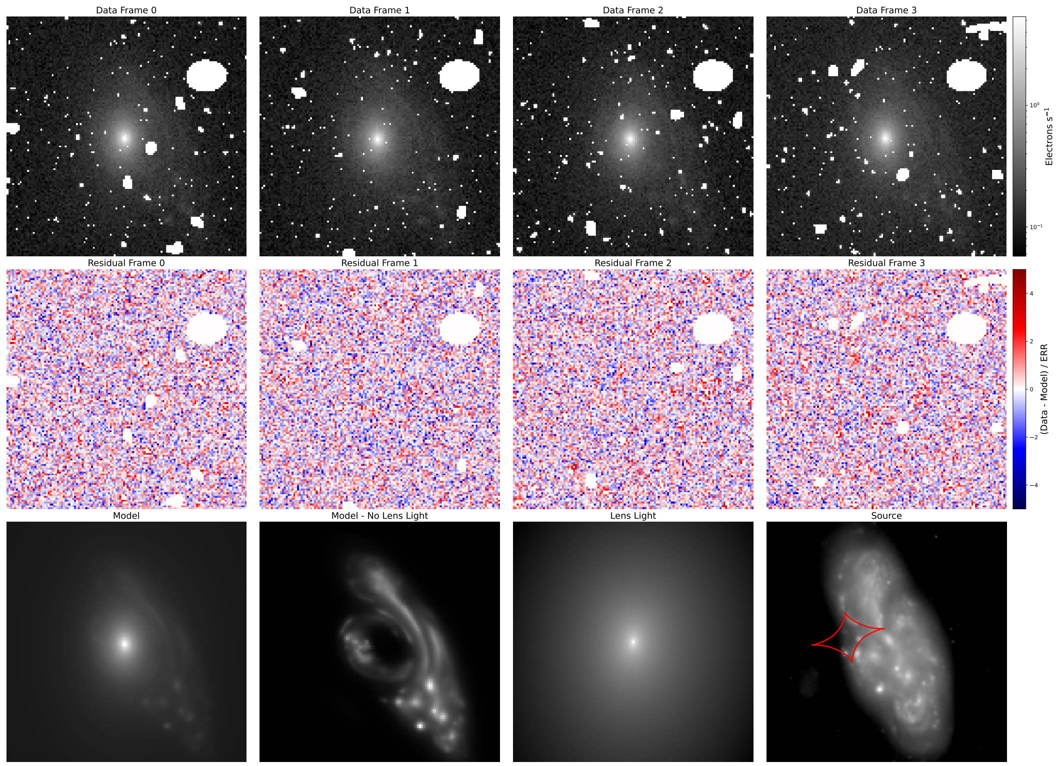

Figure 5: Same as Fig. 1, but for SDSSJ0157-0056.

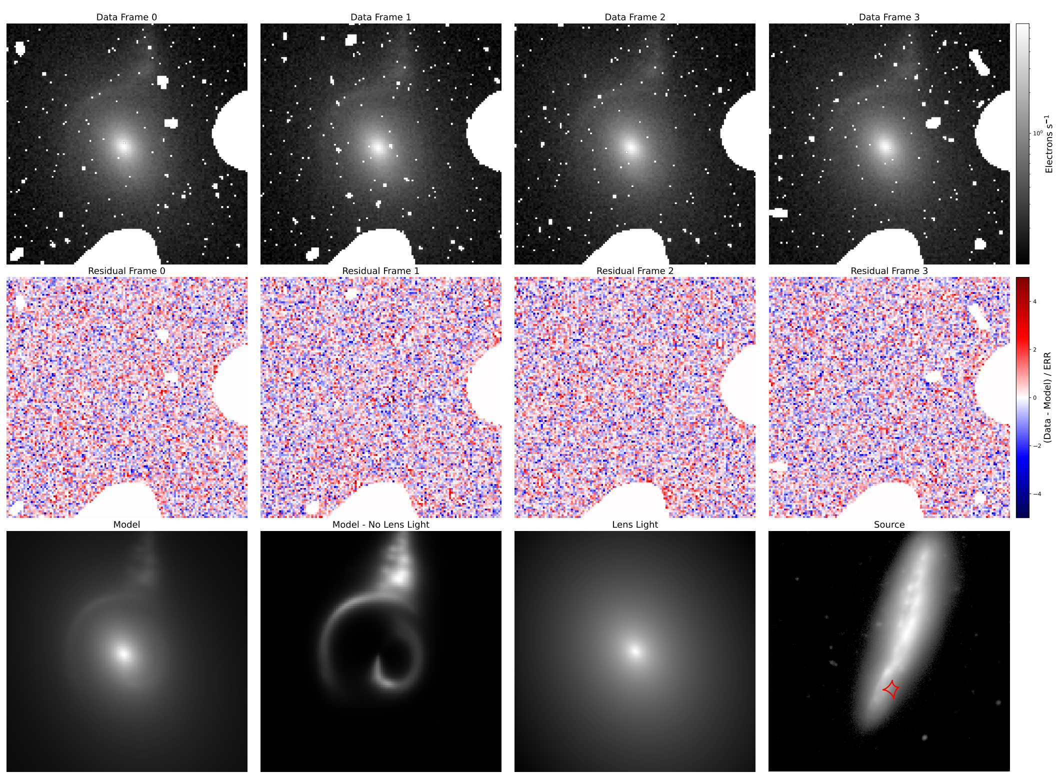

Figure 6: Same as Fig. 1, but for SDSSJ0216-0813.

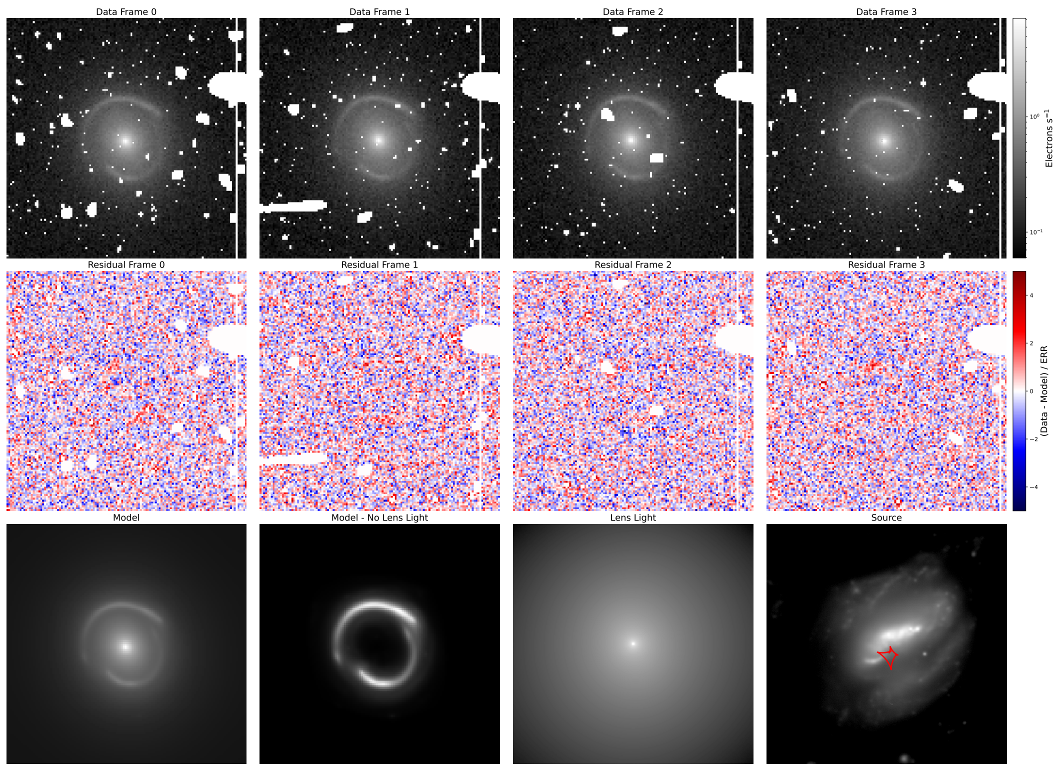

Figure 7: Same as Fig. 1, but for SDSSJ0252+0039.

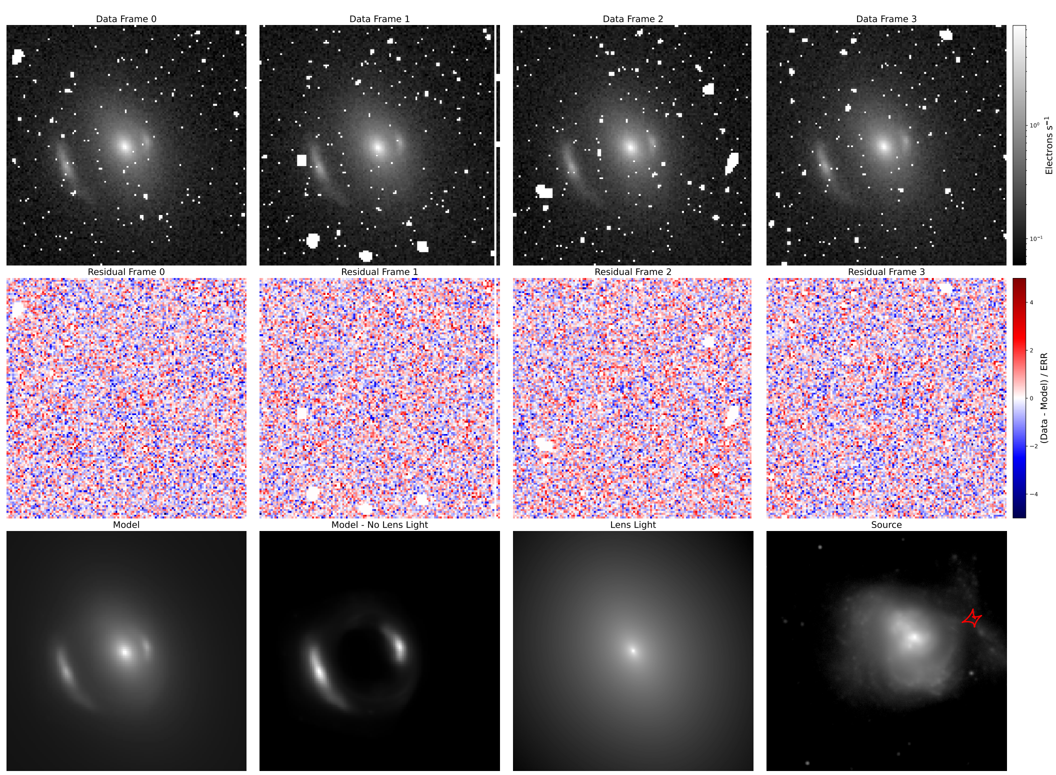

Figure 8: Same as Fig. 1, but for SDSSJ0330-0020.

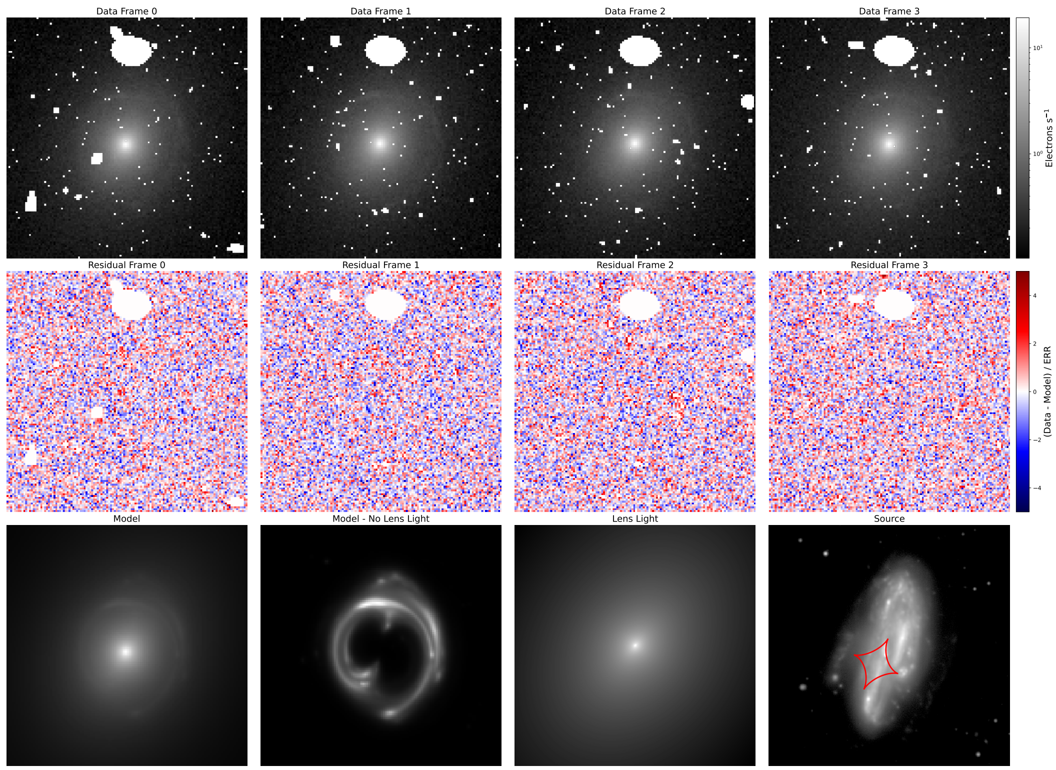

Figure 9: Same as Fig. 1, but for SDSSJ0728+3835.

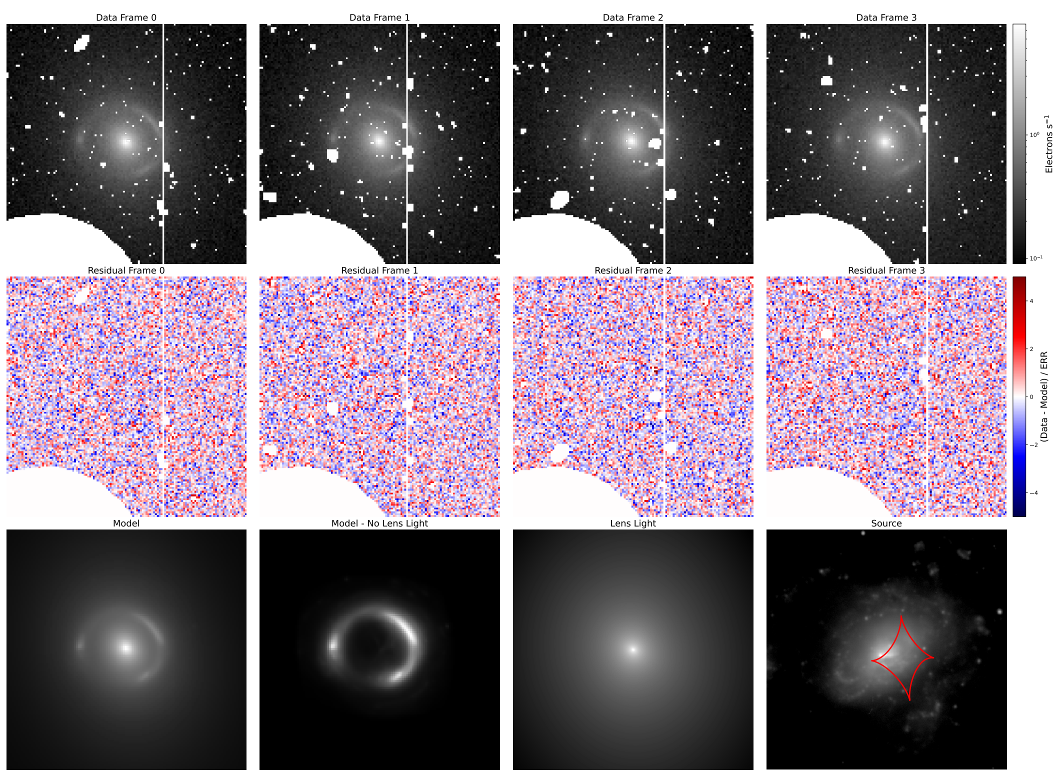

Figure 10: Same as Fig. 1, but for SDSSJ0737+3216.

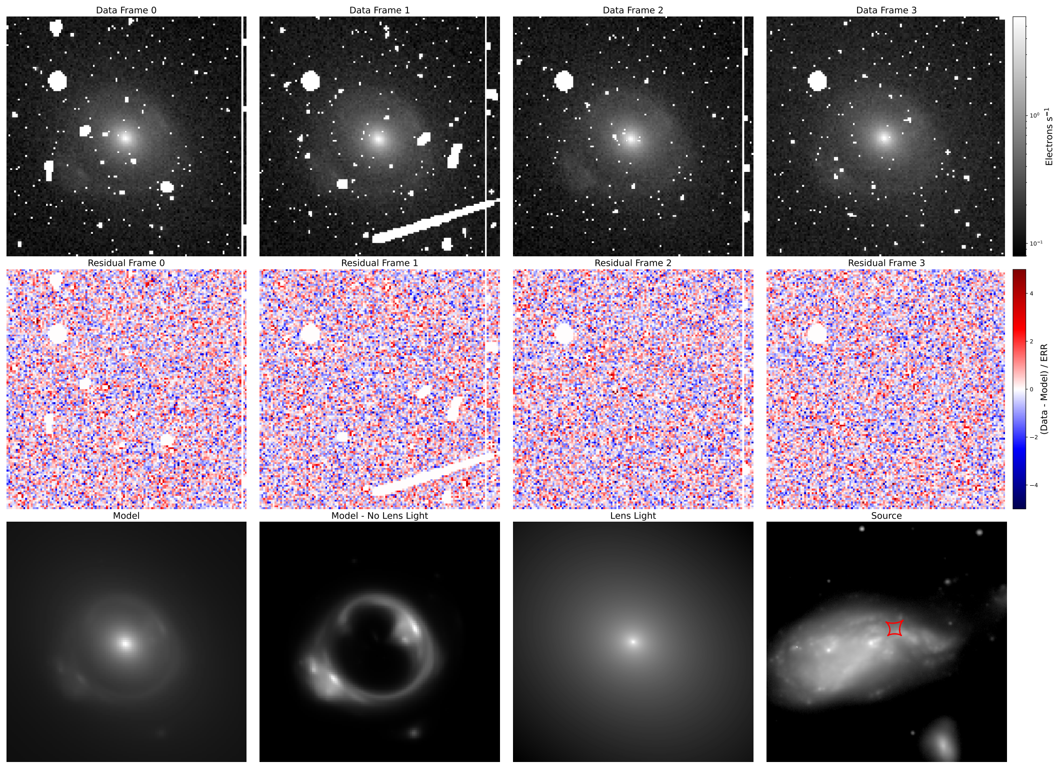

Figure 11: Same as Fig. 1, but for SDSSJ0903+4116.

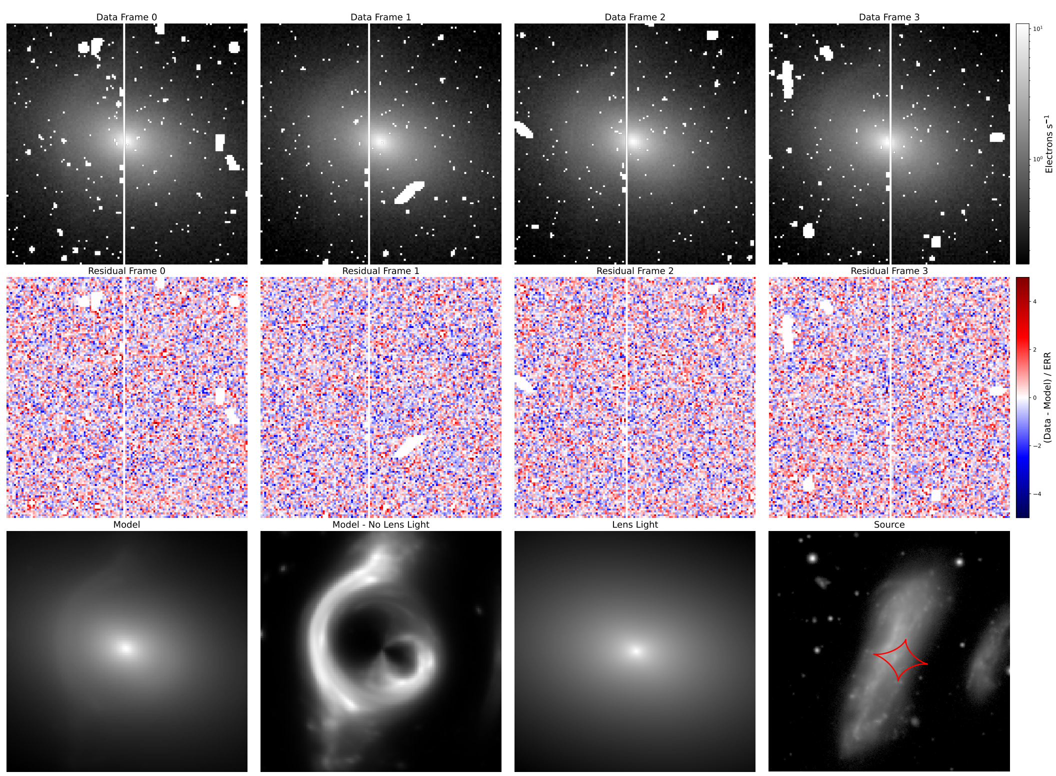

Figure 12: Same as Fig. 1, but for SDSSJ0912+0029.

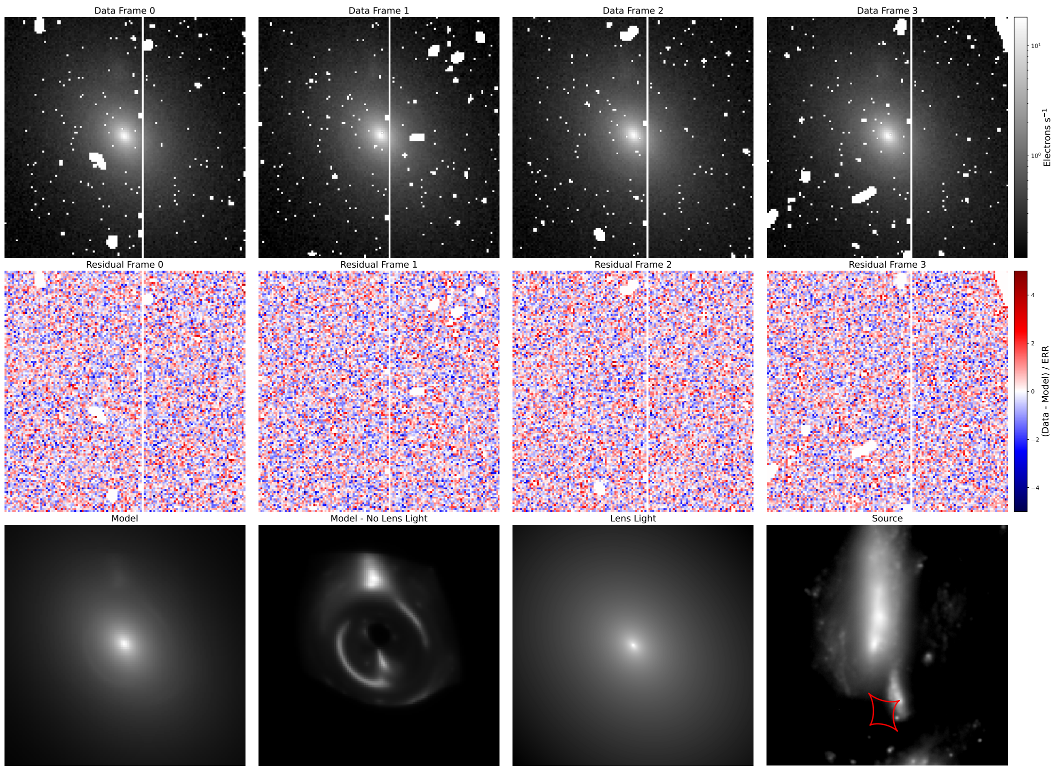

Figure 13: Same as Fig. 1, but for SDSSJ0936+0913.

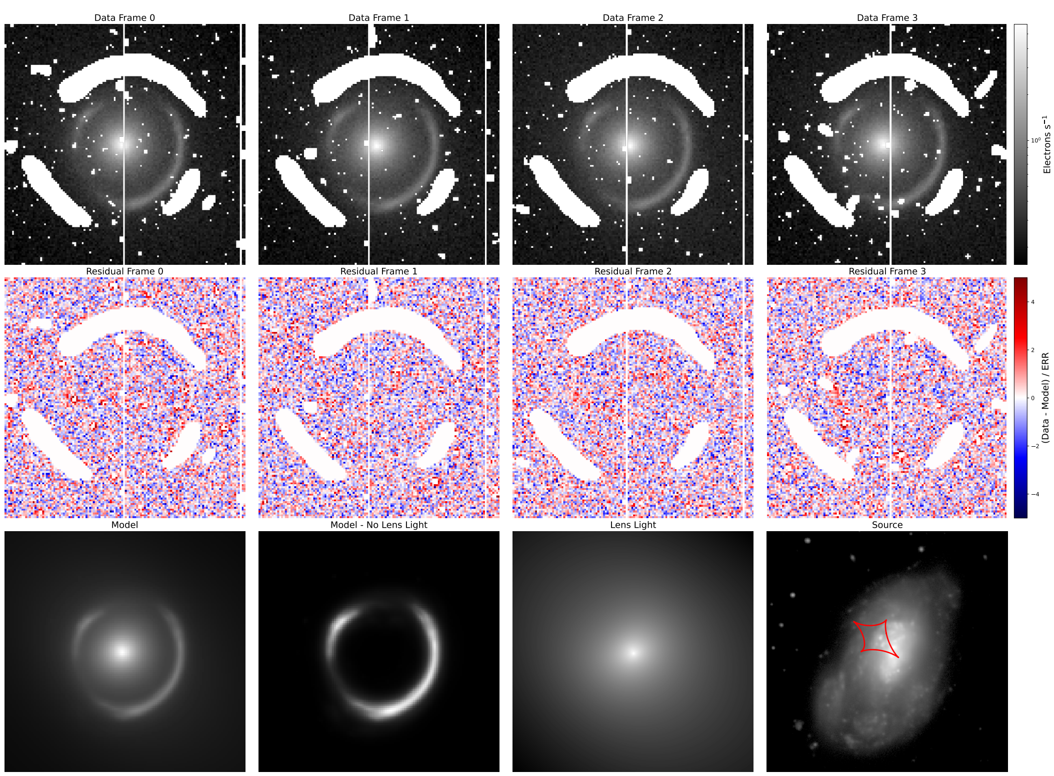

Figure 14: Same as Fig. 1, but for SDSSJ0946+1006.

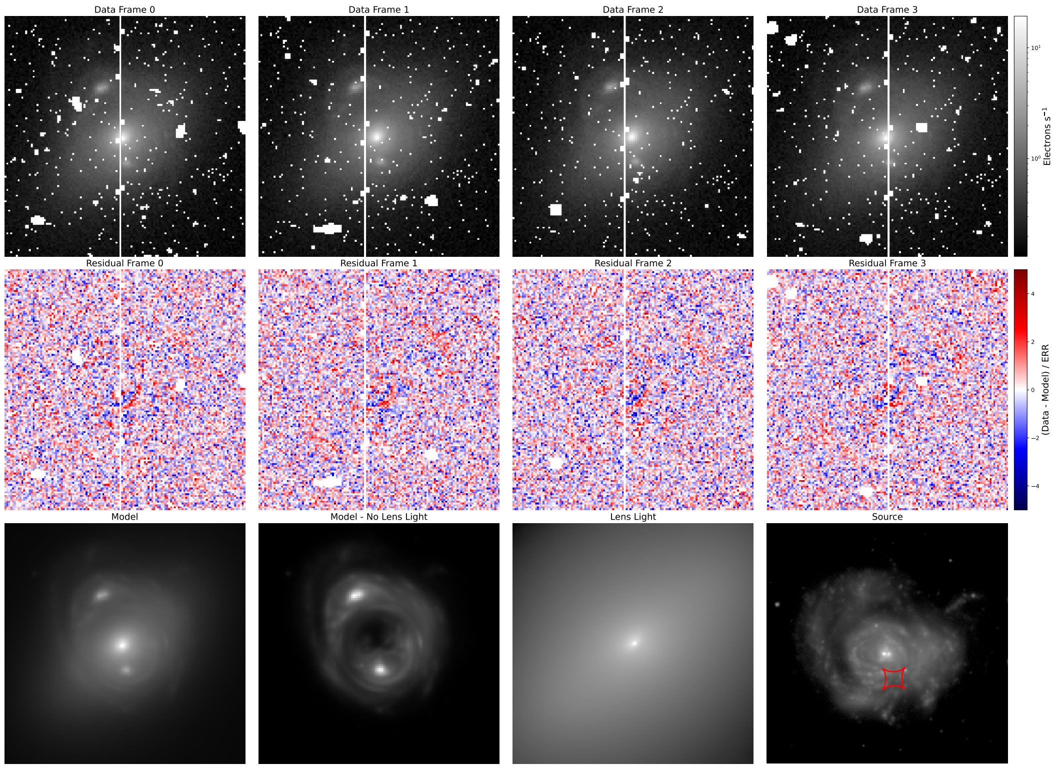

Figure 15: Same as Fig. 1, but for SDSSJ0959+0410.

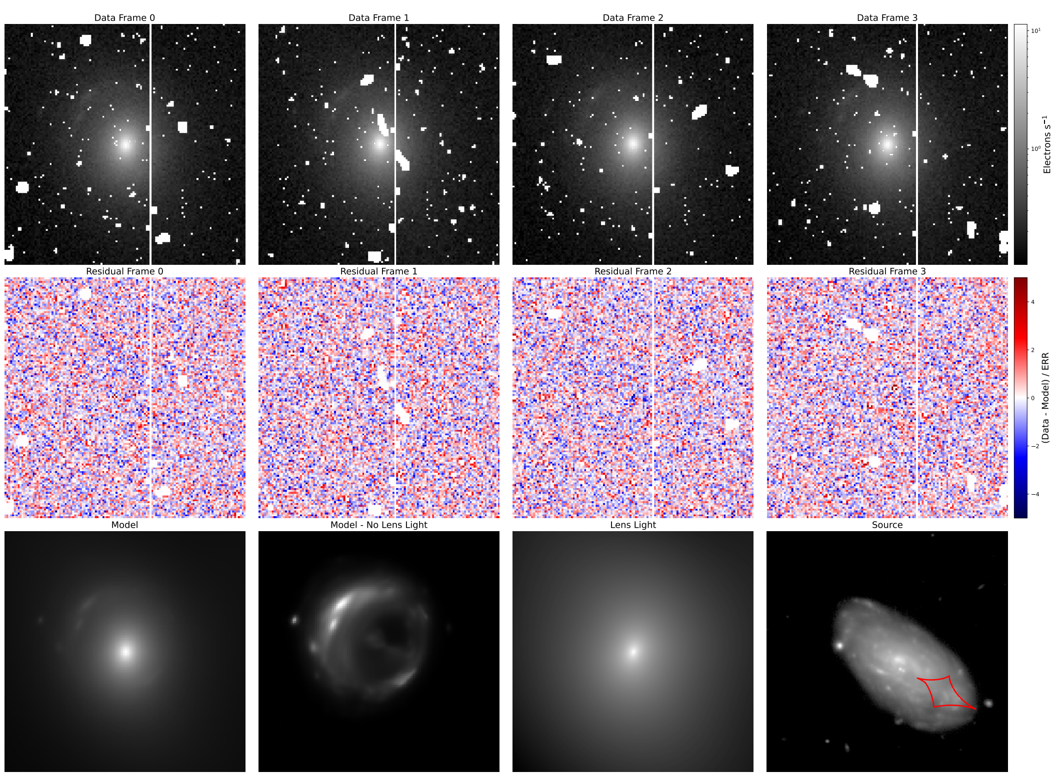

Figure 16: Same as Fig. 1, but for SDSSJ1020+1122.

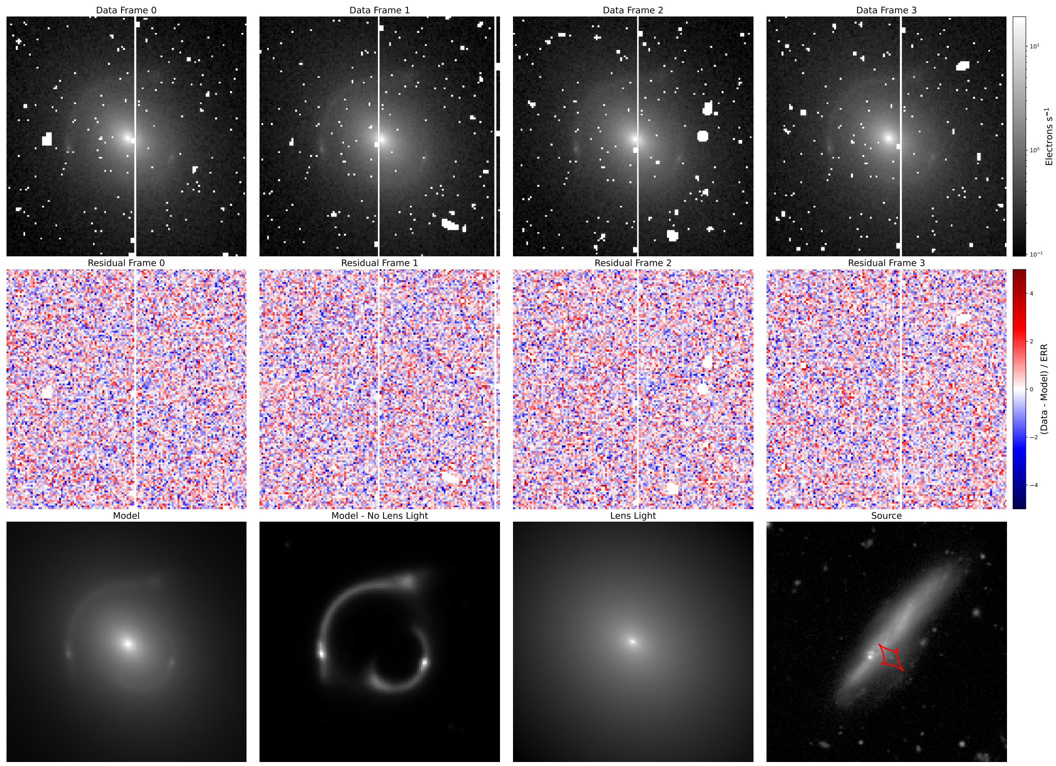

Figure 17: Same as Fig. 1, but for SDSSJ1023+4230.

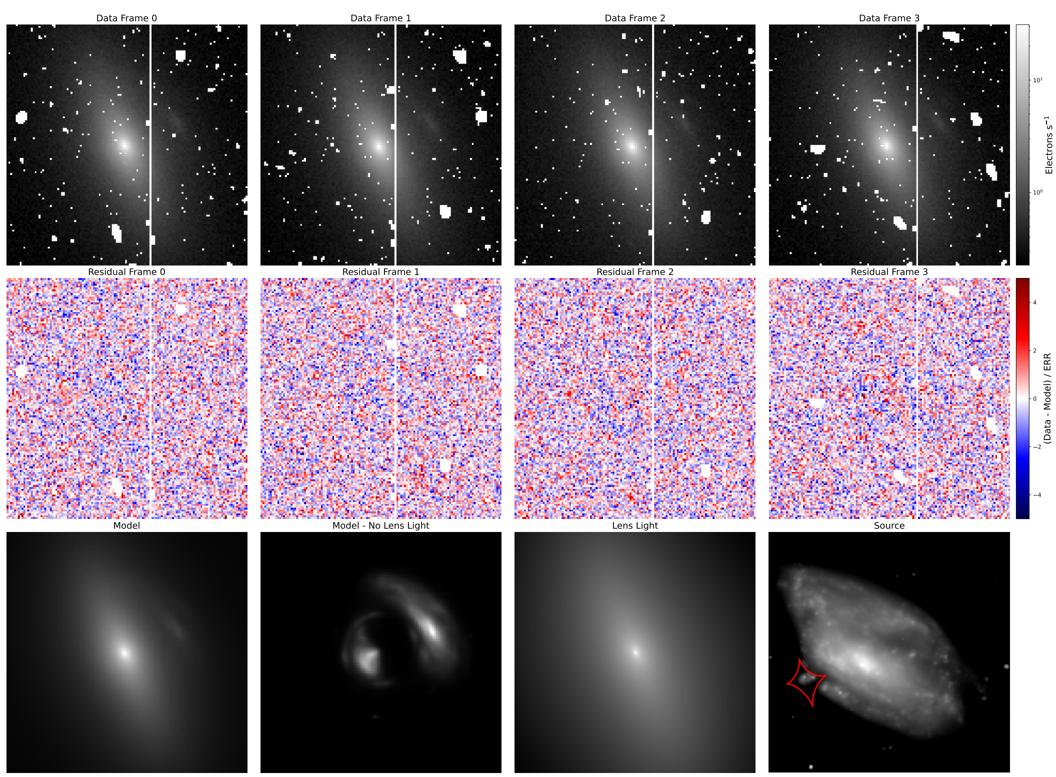

Figure 18: Same as Fig. 1, but for SDSSJ1029+0420.

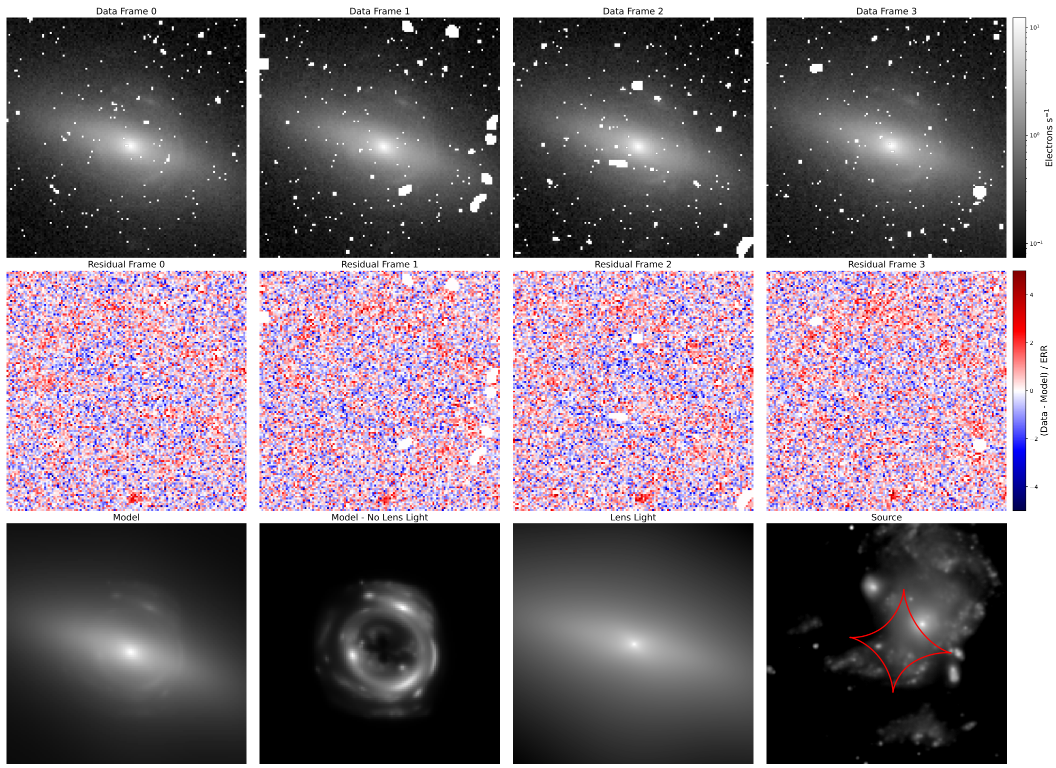

Figure 19: Same as Fig. 1, but for SDSSJ1103+5322.

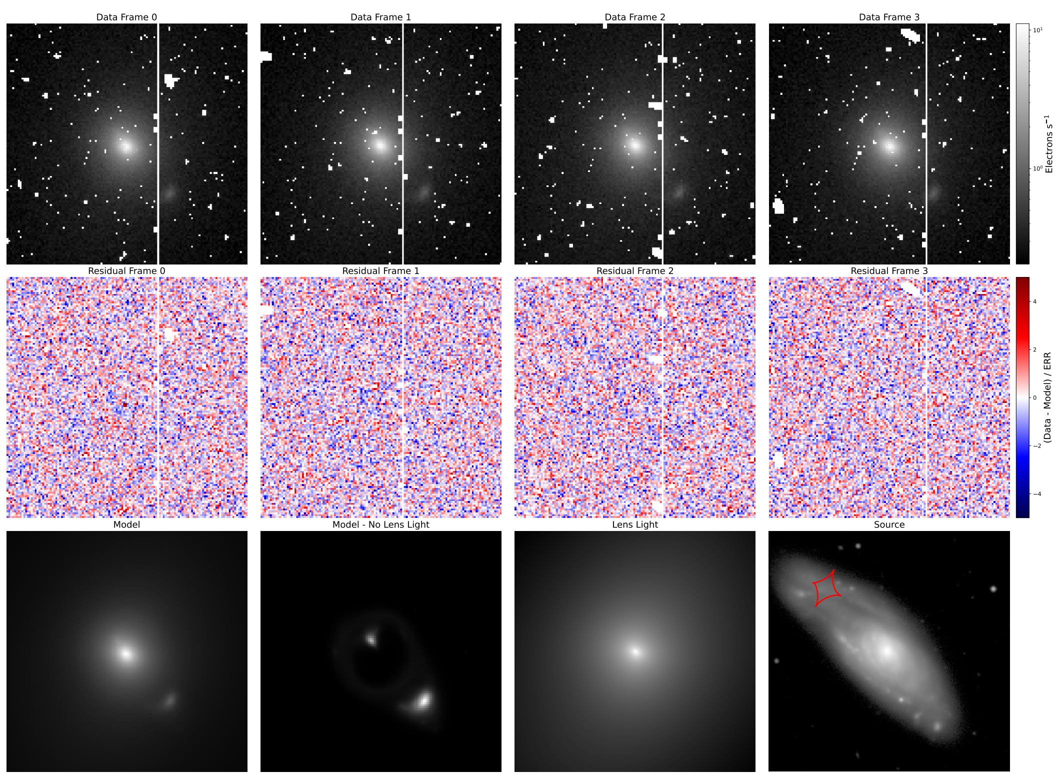

Figure 20: Same as Fig. 1, but for SDSSJ1142+1001.

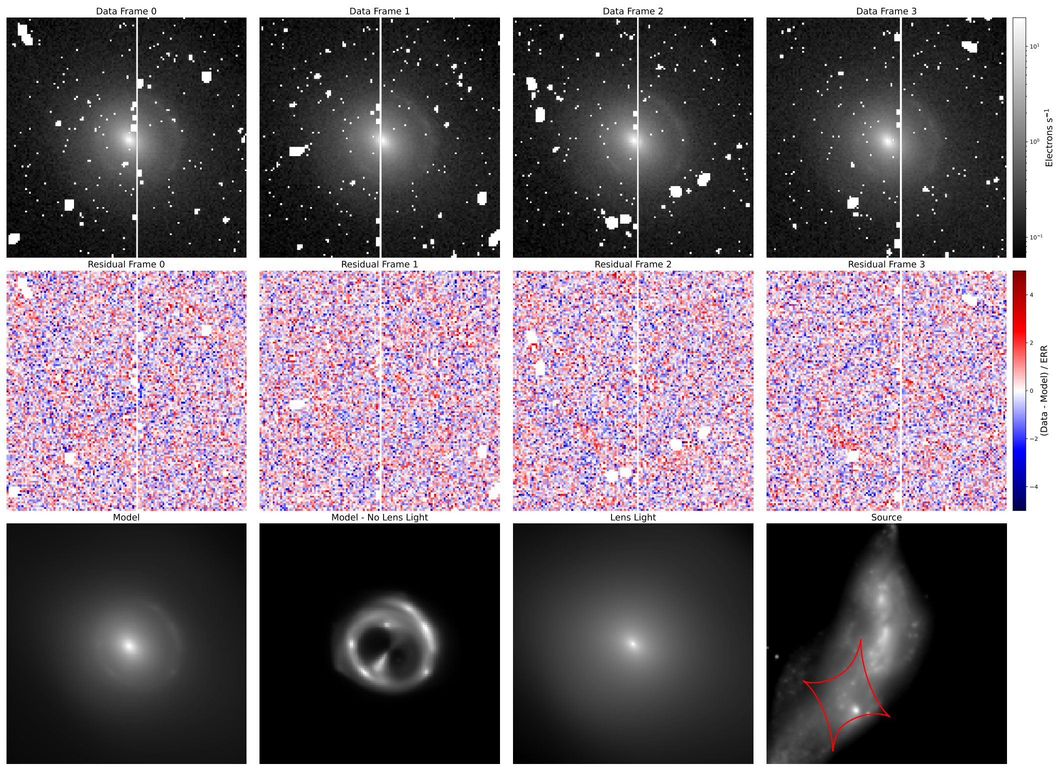

Figure 21: Same as Fig. 1, but for SDSSJ1153+4612.

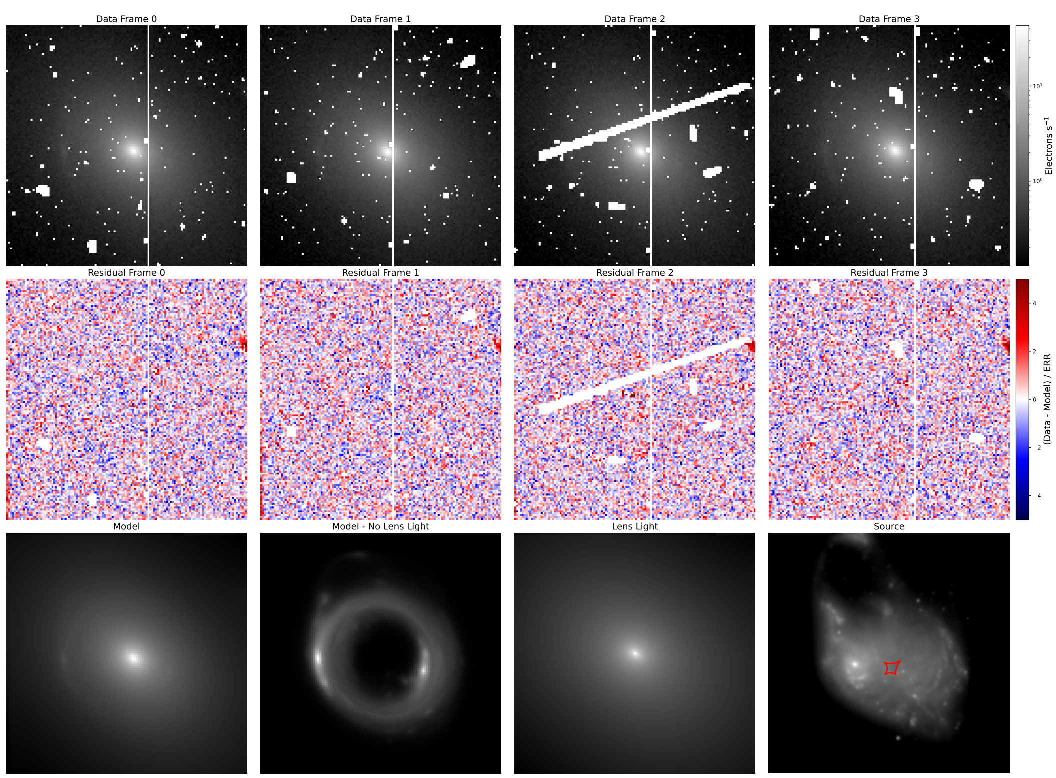

Figure 22: Same as Fig. 1, but for SDSSJ1213+6708.

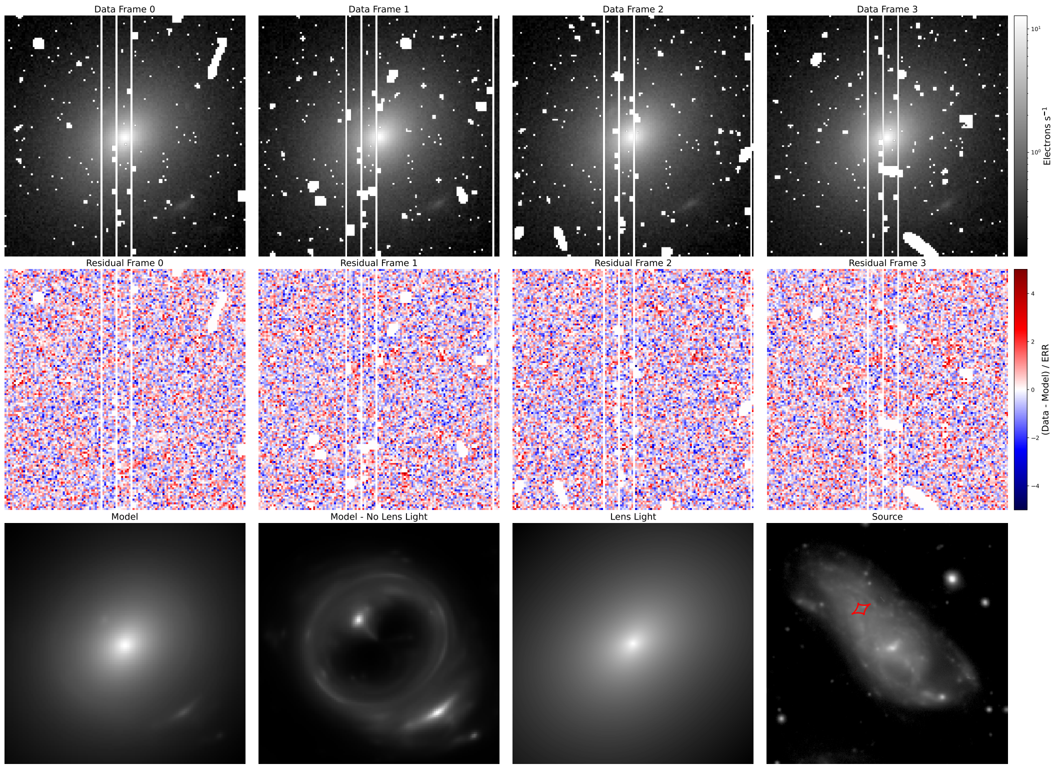

Figure 23: Same as Fig. 1, but for SDSSJ1218+0830.

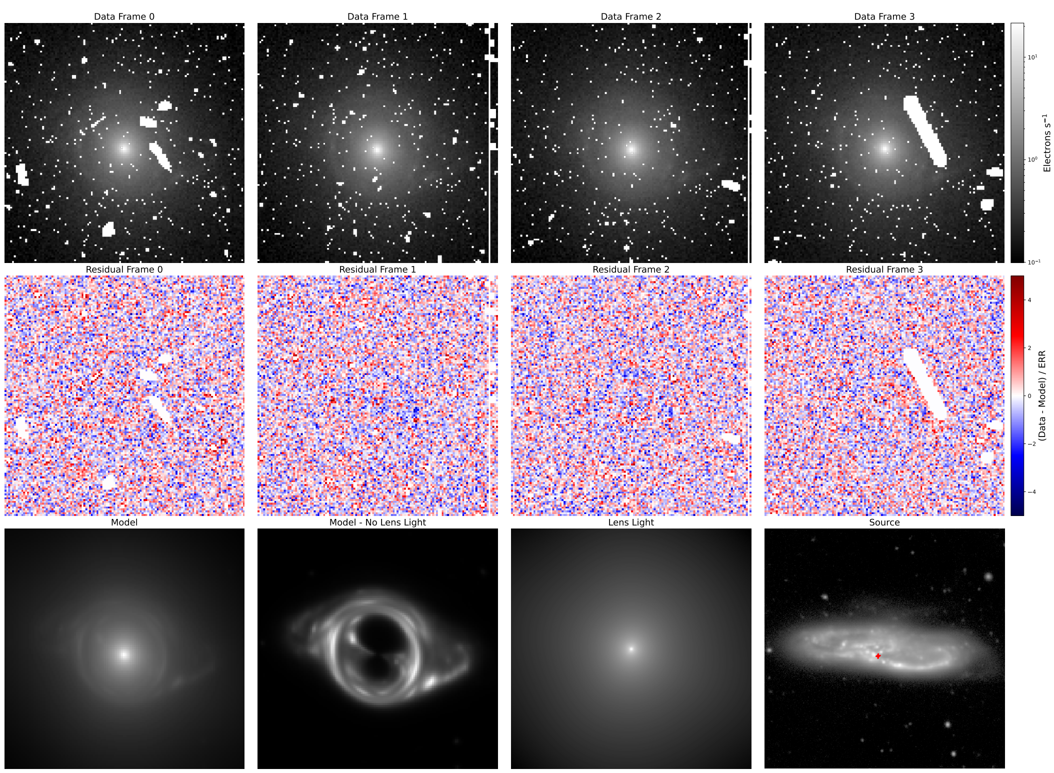

Figure 24: Same as Fig. 1, but for SDSSJ1250+0523.

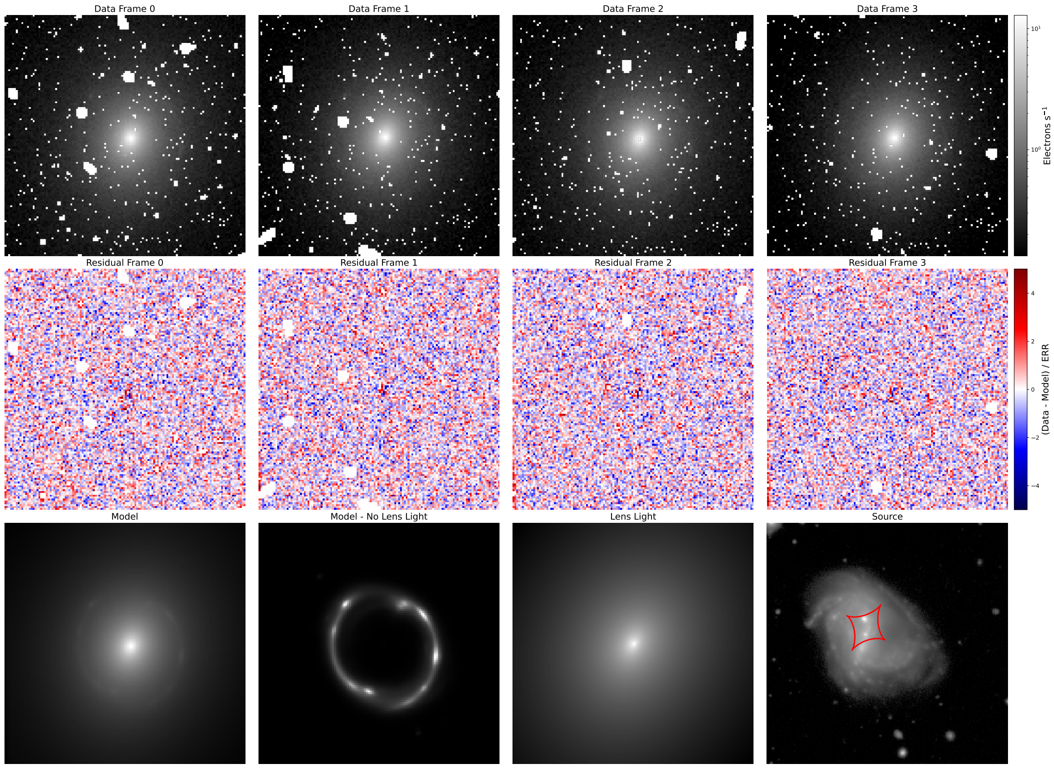

Figure 25: Same as Fig. 1, but for SDSSJ1402+6321.

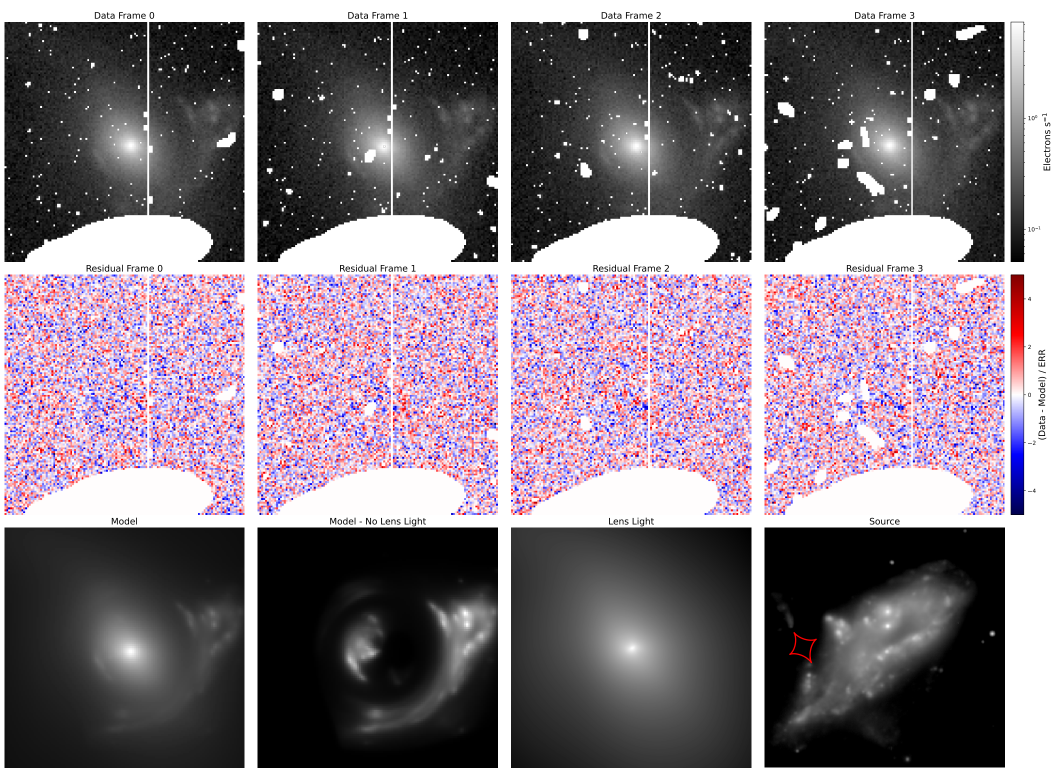

Figure 26: Same as Fig. 1, but for SDSSJ1416+5136.

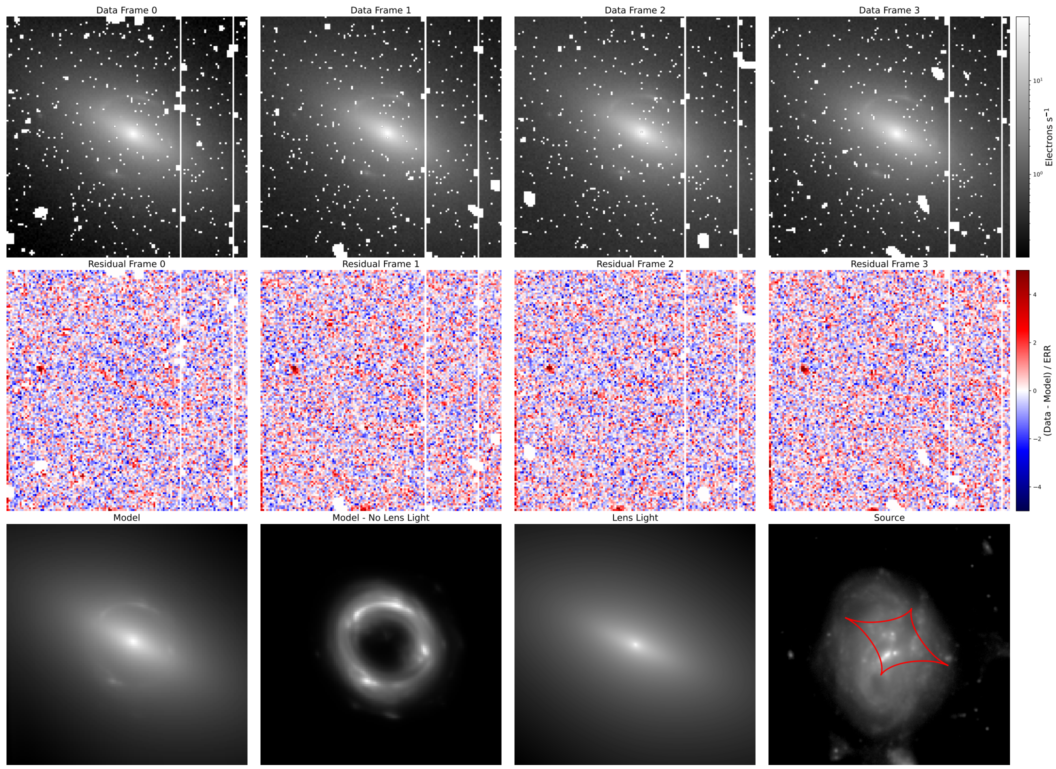

Figure 27: Same as Fig. 1, but for SDSSJ1420+6019.

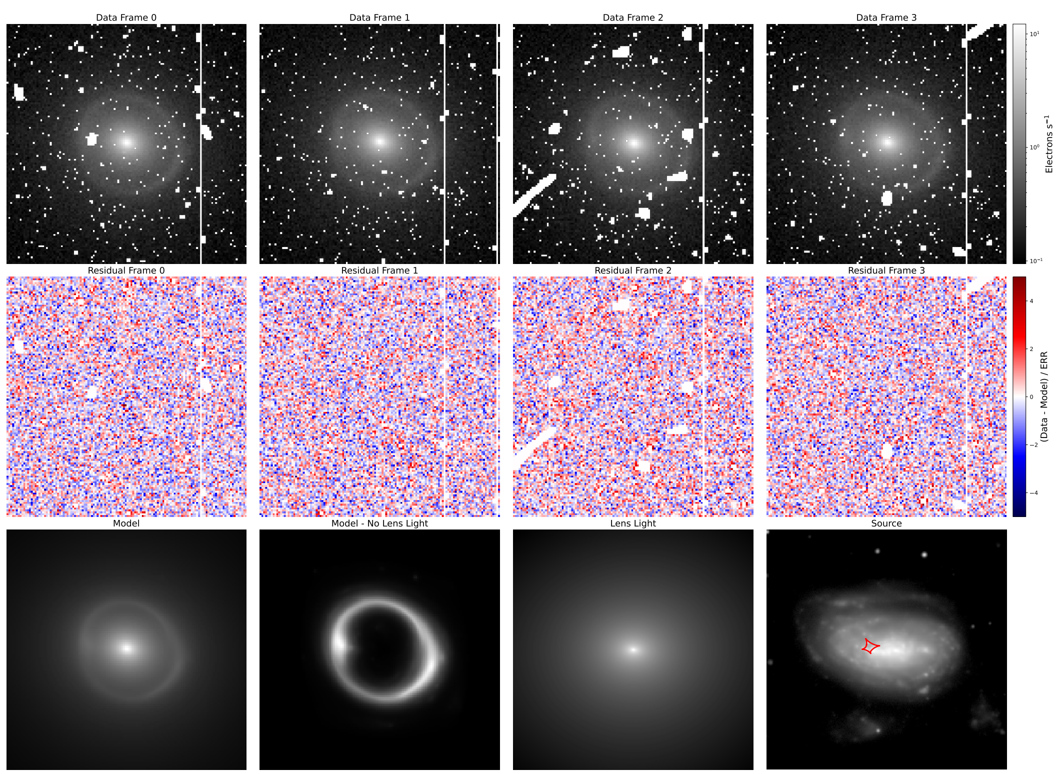

Figure 28: Same as Fig. 1, but for SDSSJ1627-0053.

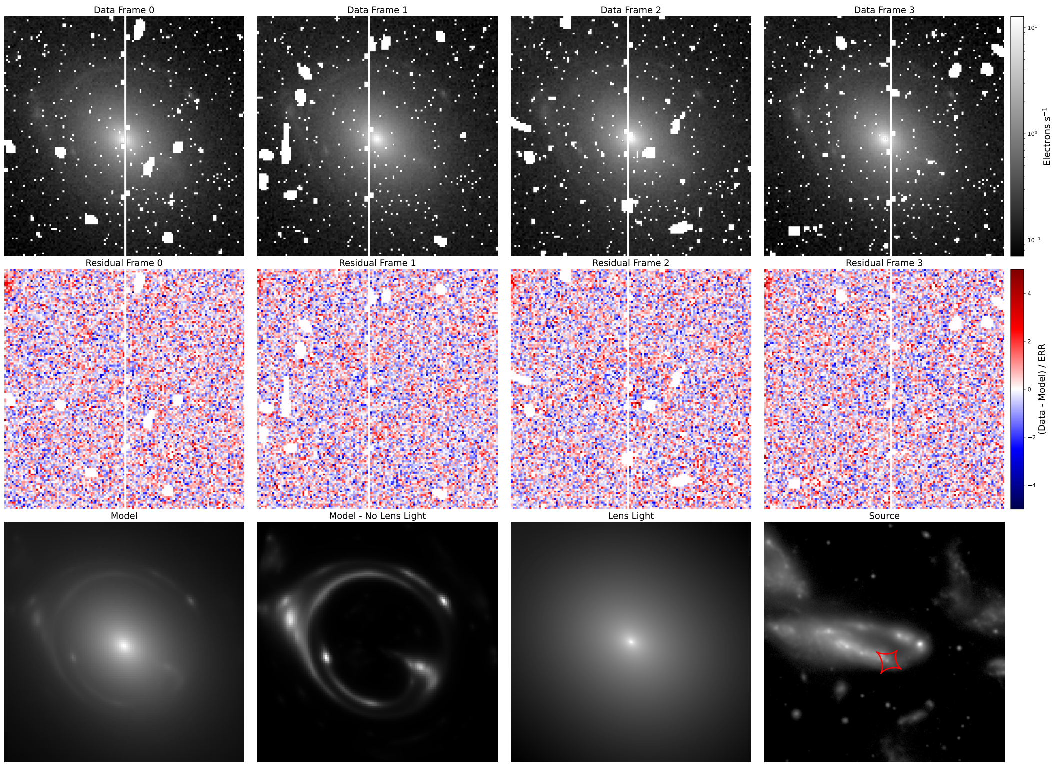

Figure 29: Same as Fig. 1, but for SDSSJ1630+4520.

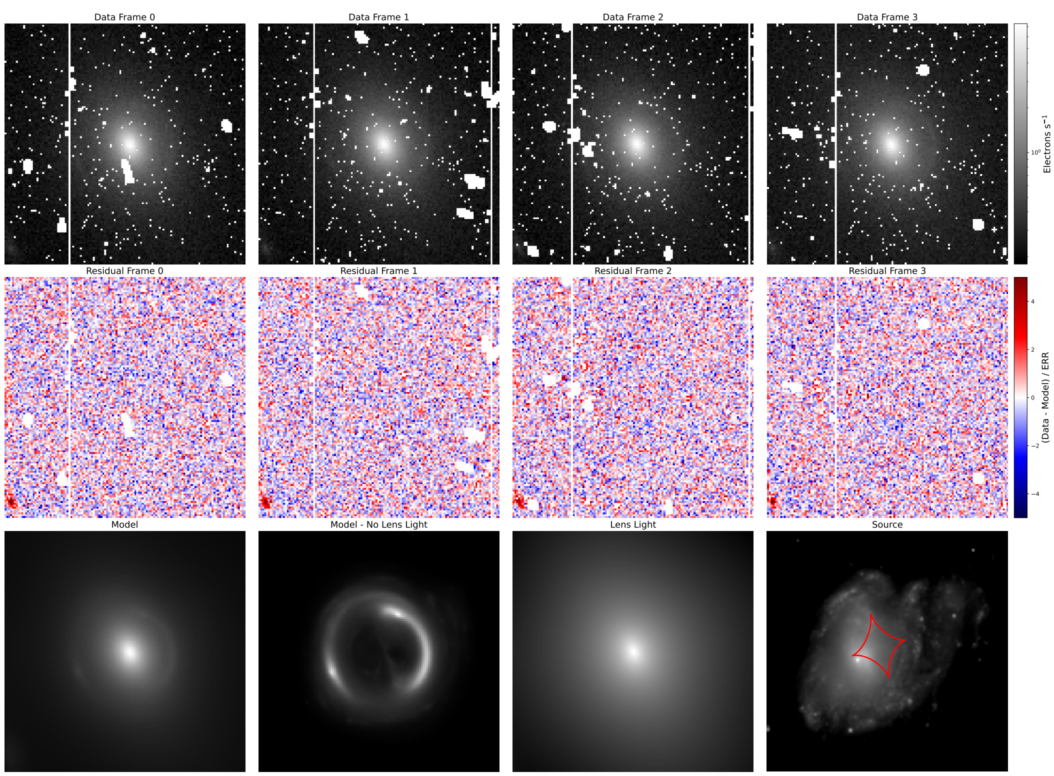

Figure 30: Same as Fig. 1, but for SDSSJ2300+0022.

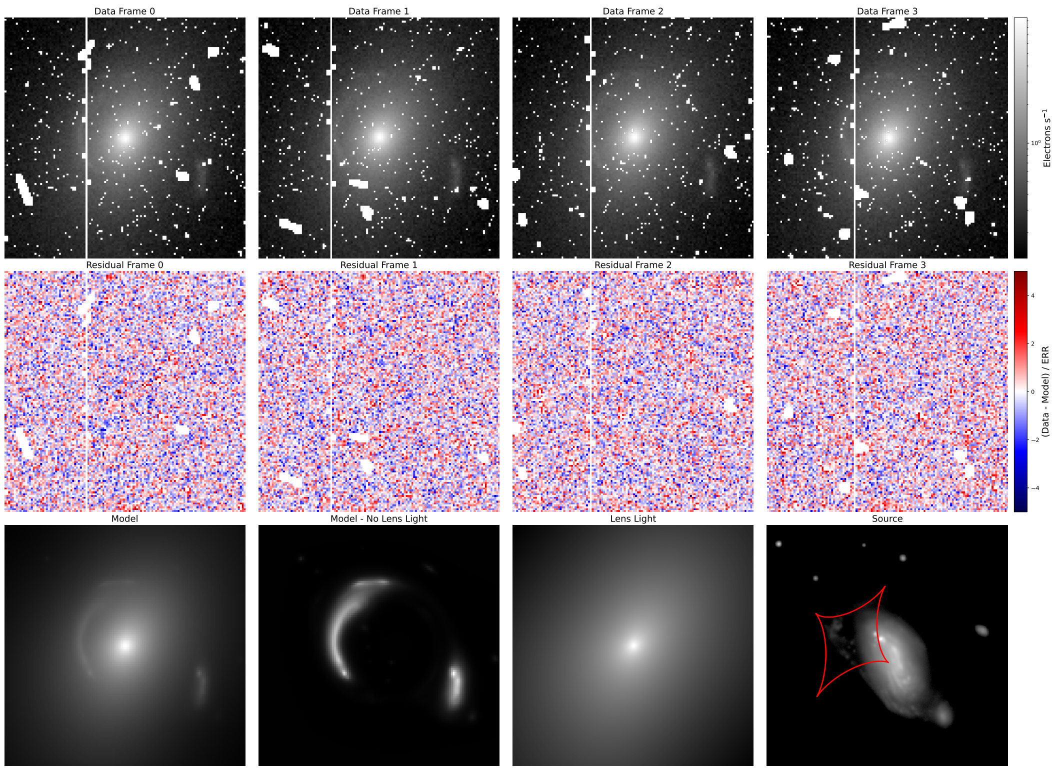

Figure 31: Same as Fig. 1, but for SDSSJ2303+1422.

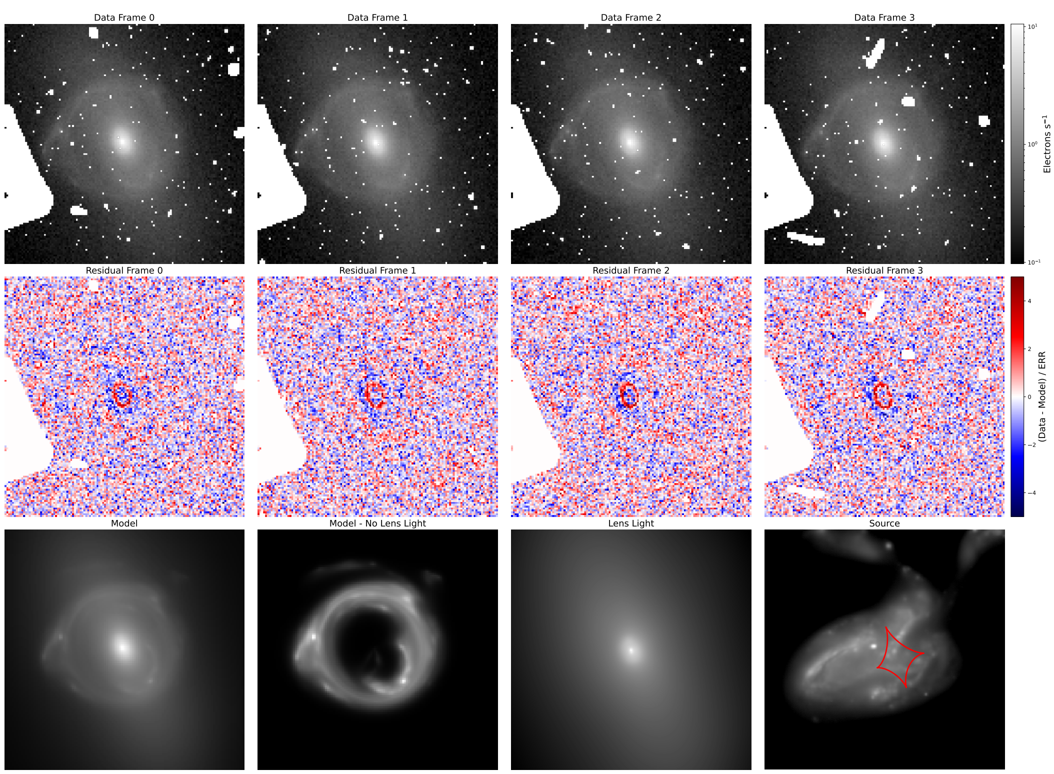

Figure 32: Same as Fig. 1, but for SDSSJ2341+0000.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction claim state-of-the-art modeling of real strong lenses, which is supported by the data–model residuals we present.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss two main limitations in the Results and Discussion: training the lens-light score model only on parametric profiles, and adopting a Gaussian likelihood despite the non-Gaussian nature of HST noise.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: There are no theoretical results presented in the paper, so proofs are not applicable. We do make assumptions detailed in the text on the chosen models we use to reconstruct the lensing systems (e.g., foreground mass distribution)

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide details on the dataset, modeling choices, and the specific neural network architectures and training data used, ensuring the main results can be reproduced.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We provide detailed descriptions of the dataset and methods, but do not release our code. Nonetheless, the information in the manuscript is sufficient to faithfully reproduce the main results.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We specify the number of epochs, batch sizes, training datasets, and chosen architecture (U-Net, NCSN++), along with other key training details necessary to understand the results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We report error bars for the parametric foreground mass models and provide the full joint posterior statistics for one representative system (SDSSJ1430+4105), ensuring the statistical significance of the results is clearly conveyed.

Guidelines:

- The answer NA means that the paper does not include experiments.

- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We specify the GPU resources used for both training and sampling (NVIDIA A100, 40 GB), providing sufficient information on the compute requirements to reproduce the experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics, and our research conforms fully to its principles.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

Justification: Our methods advance astrophysical inference by enabling more accurate reconstructions of strong gravitational lenses, with positive impact for cosmology and dark matter studies. A potential negative impact is that the underlying generative modeling techniques could, in principle, be repurposed for harmful image manipulation, though our models are domain-specific and not designed for such uses.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: We do not consider our models, trained on astrophysical data and simulations, to pose a high risk of misuse. Moreover, all data analyzed in this work is publicly available.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We properly credit the creators of all external assets, explicitly referencing the publicly available datasets used. All code was developed by us, and we respect the licenses and terms of use of the cited resources.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: We do not introduce or release any new assets in this work, so no additional documentation is required.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: Our paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: We do not use LLMs in our work, so no declaration is required.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.