
From Simulation to Survey: Benchmarking Super-Resolution for LSST-like Lensing Data

Aleksandr Duplinskii

Department of Physics
University of Oxford
Oxford, OX1 3PU, UK

al.duplinsky@gmail.com

Pranath Reddy

University of Florida
Gainesville, FL 32611, USA

Michael W Toomey

Center for Theoretical Physics – a Leinweber Institute,
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Sergei Gleyzer

Department of Physics
Astronomy, University of Alabama
Tuscaloosa, AL 35401, USA

Abstract

High-resolution imaging of strong gravitational lenses have the potential to provide unique insights into dark matter substructure and galaxy evolution, but these data are currently limited to a handful of examples from expensive space-based observations. Ground-based surveys like LSST will deliver orders of magnitude more lens systems, but at substantially lower resolution and higher noise levels. To bridge this gap, we present a synthetic dataset of paired high- and low-resolution gravitational lens images, designed to mimic the differences between space-based and ground-based instruments. Our dataset, generated using the Mejiro simulation framework, integrates physically motivated lens and source models with realistic instrumental effects, yielding 20,000 paired samples suitable for training and benchmarking super-resolution methods. We evaluate four representative approaches - RCAN, SwinIR, a conditional diffusion model, and SatGAN - spanning CNN, Transformer, diffusion, and adversarial paradigms. Our benchmarks on the Mejiro data set establish a controlled testbed for conditional super-resolution in astronomy, with potential to enhance the scientific return of upcoming wide-field surveys.

1 Introduction

Galaxy-galaxy strong lensing, the distortion of distant galaxies by an intervening dark matter halo, has emerged as a powerful probe of dark matter, cosmology, and galaxies more generally Bartelmann [2010]. For the problem of identifying dark matter, understanding and constraining perturbations in gravitational lenses from substructure will be of critical importance Hezaveh et al. [2016]. However, this is only possible with high-resolution images from expensive space-based instruments like Hubble Space Telescope and the forthcoming Roman Space Telescope along with detailed lens modeling. However, space-based imaging is observationally expensive and covers limited sky area. In contrast, ground-based surveys like the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST) will observe millions of galaxies and discover thousands of new lenses, but with much lower resolution due to atmospheric seeing and detector noise. This resolution gap poses a critical challenge: the majority of future strong lensing data will lack the sharpness required for precise scientific analyses.

Image super-resolution (SR) offers a promising path forward. By learning to map noisy, low-resolution images to their high-resolution counterparts, SR models could effectively restore information lost

in ground-based observations, enabling analyses previously restricted to space-based data. While SR has been extensively studied in computer vision, astronomy presents unique challenges: realistic astrophysical structures, lensing distortions, and instrumental effects must all be modeled accurately to ensure scientific validity. Progress in this direction requires benchmark datasets that couple physical realism with well-controlled degradations, as well as systematic evaluations across diverse SR approaches.

In this work, we introduce a synthetic dataset of paired high- and low-resolution gravitational lensing images, generated with the Mejiro framework Wedig et al. [2025] by combining lens mass models, source galaxy morphologies, and realistic instrument simulations. The dataset consists of 20,000 pairs reflecting the resolution gap between Roman-like space imaging and LSST-like ground-based imaging. We use this dataset to benchmark four distinct SR paradigms: a CNN baseline (RCAN), a Transformer (SwinIR), a probabilistic diffusion model, and an adversarially trained GAN. Our results provide the first controlled comparison of modern SR methods for gravitational lensing, offering a foundation for future work at the intersection of computer vision and astrophysics.

2 Dataset Generation

We construct a synthetic dataset of strong gravitational lensing systems with paired high- and low-resolution images, designed to reflect the gap between space-based and ground-based observations. The dataset is generated using the Mejiro simulation framework, which internally combines physically motivated mass models from `pyHalo` Gilman et al. [2020] with gravitational lensing solvers from `lenstronomy` Birrer and Amara [2018] and image rendering via `GalSim` Rowe et al. [2015]. Each system consists of a background source galaxy lensed by a foreground galaxy-scale halo, with additional perturbations to mimic the diversity observed in real data.

Lens population. The main deflector is modeled as an elliptical power-law mass distribution with external shear. We introduce diversity by randomly perturbing the Einstein radius ($\theta_E \times U[0.5, 1.5]$), ellipticity components ($e_1, e_2 \sim U[-0.15, 0.15]$ with clipping to $e < 0.6$), and external shear strength/orientation. For a subset of samples, we enrich the lens with line-of-sight halos and bound subhalos using `pyHalo` Gilman et al. [2020], drawing from a cold dark matter mass function with $M_{\text{sub}} \in [10^6, 10^{10}] M_\odot$, truncated by host halo properties.

Source population. The background galaxy is modeled as a Sérsic profile, with parameters randomly varied to increase morphological diversity: centroid offsets ($\Delta x, \Delta y \sim U[-0.6'', 0.6'']$), effective radius ($R_{\text{eff}} \times U[0.5, 2.0]$), ellipticity ($e_1, e_2 \sim U[-0.15, 0.15]$ with clipping), and Sérsic index ($n \sim U[0.5, 4.0]$). This produces a mixture of compact and extended sources with diverse brightness distributions.

High-resolution (HR) images. HR targets are rendered using the Roman Space Telescope instrument model Spergel et al. [2015] with a pixel scale of $0.11''$ and a $5''$ field of view. To provide clean ground-truth targets, we omit PSF convolution and detector noise.

Low-resolution (LR) images. LR counterparts are generated by degrading HR images to LSST-like conditions Ivezic et al. [2009]. We convolve each image with a Moffat PSF ($\text{FWHM} \sim 0.7''$, $\beta = 3.5$, ellipticity $e \simeq 0.03$, orientation $\theta \simeq 20^\circ$). Detector effects are then simulated: a pair of 15 s exposures are co-added, and Poisson-distributed sky background and dark current are added together with Gaussian read noise ($\sigma_{\text{read}} = 5 e^-$). This reproduces the noise properties of a single LSST visit.

Dataset format. We generate 20,000 paired HR/LR images, each stored as a single-channel 48×48 .npy array. Padding is applied to standardize dimensions, ensuring direct compatibility with CNN, Transformer, diffusion, and GAN-based architectures. Normalization strategies are applied at the preprocessing stage and differ depending on the model (e.g., min–max scaling for diffusion models, global percentile scaling for CNN/Transformer-based baselines).

Overall, the dataset combines physical realism (via halo substructure and Sérsic-profile sources), instrumental realism (via PSF and noise modeling), and statistical diversity (via randomized parameters), making it a challenging and controlled benchmark for conditional super-resolution in astronomy.

3 Super-Resolution Models

We benchmark four distinct super-resolution (SR) approaches, chosen to span classical CNN baselines, Transformer-based methods, probabilistic diffusion, and GAN-based adversarial training. Each model is trained on the same HR/LR dataset described above, with consistent data splits and evaluation metrics.

RCAN (Residual Channel Attention Network). RCAN Zhang et al. [2018] is a widely used CNN architecture for SR, relying on very deep residual-in-residual blocks with channel attention. It is optimized with pixel-wise L1 loss, providing a strong deterministic baseline for structure-preserving super-resolution.

SwinIR (Swin Transformer for Image Restoration). SwinIR Liang et al. [2021] replaces CNN feature extraction with shifted-window self-attention. This model has shown strong performance across natural image restoration tasks. We adapt the official implementation to single-channel astronomical inputs and train with a global percentile normalization.

Conditional Diffusion Model. We design a custom U-Net-based conditional diffusion model, where the low-resolution image acts as the conditioning input. Our approach is inspired by the *DiffLense* framework Reddy et al. [2024], which has demonstrated the potential of diffusion models for super-resolving strong gravitational lensing data. We adopt the standard DDPM formulation with 1000 denoising steps and a cosine noise schedule for time embeddings. Conditioning is provided both by concatenating the low-resolution image with the noisy input and by injecting a cosine-embedded representation of the LR image into the U-Net.

SatGAN. We implement a lightweight UNet-like generator trained adversarially against a patch discriminator with spectral normalization. The generator is supervised by a combination of pixel-wise losses (L1, flux-conservation), structural similarity (MS-SSIM), and a data-consistency term that enforces agreement with the observed LR image after PSF degradation. Adversarial training encourages sharper reconstructions compared to purely regression-based approaches.

Evaluation metrics. We report four standard quantitative metrics, averaged over the held-out test set by comparing the outputs of the SR model to the high-resolution labels: mean absolute error (MAE), mean squared error (MSE), peak signal-to-noise ratio (PSNR, in dB), and structural similarity index (SSIM). Higher PSNR and SSIM values correspond to better perceptual quality, while lower MAE/MSE indicate pixel-level fidelity.

Table 1: Comparison of SR methods on the lensing dataset (test set). Lower is better for MAE/MSE; higher is better for PSNR/SSIM.

Model	MAE \downarrow	MSE \downarrow	PSNR (dB) \uparrow	SSIM \uparrow
RCAN	0.000126	0.000001007	64.86	0.9927
SwinIR	0.000166	0.000000822	62.10	0.9993
Cond. Diffusion	0.001059	0.000005567	53.62	0.9946
SatGAN	0.000089	0.000000590	66.83	0.9994

We also conducted a downstream task of Einstein radius estimation based on the generated images. Since conditional diffusion showed significantly worse performance, we excluded it from further analysis. The dataset was divided into two equal parts, each containing 10,000 image pairs. For each pair the Einstein radius parameter is saved during the simulation stage and later used as a label for the regression task. Each SR model was trained on the first part (training set) and then evaluated on the second part (test set) to produce enhanced images. Subsequently, a ResNet18 model was trained on the test sets produced by each SR model, as well as on the original low-resolution and high-resolution test datasets. The results are shown in Table 2.

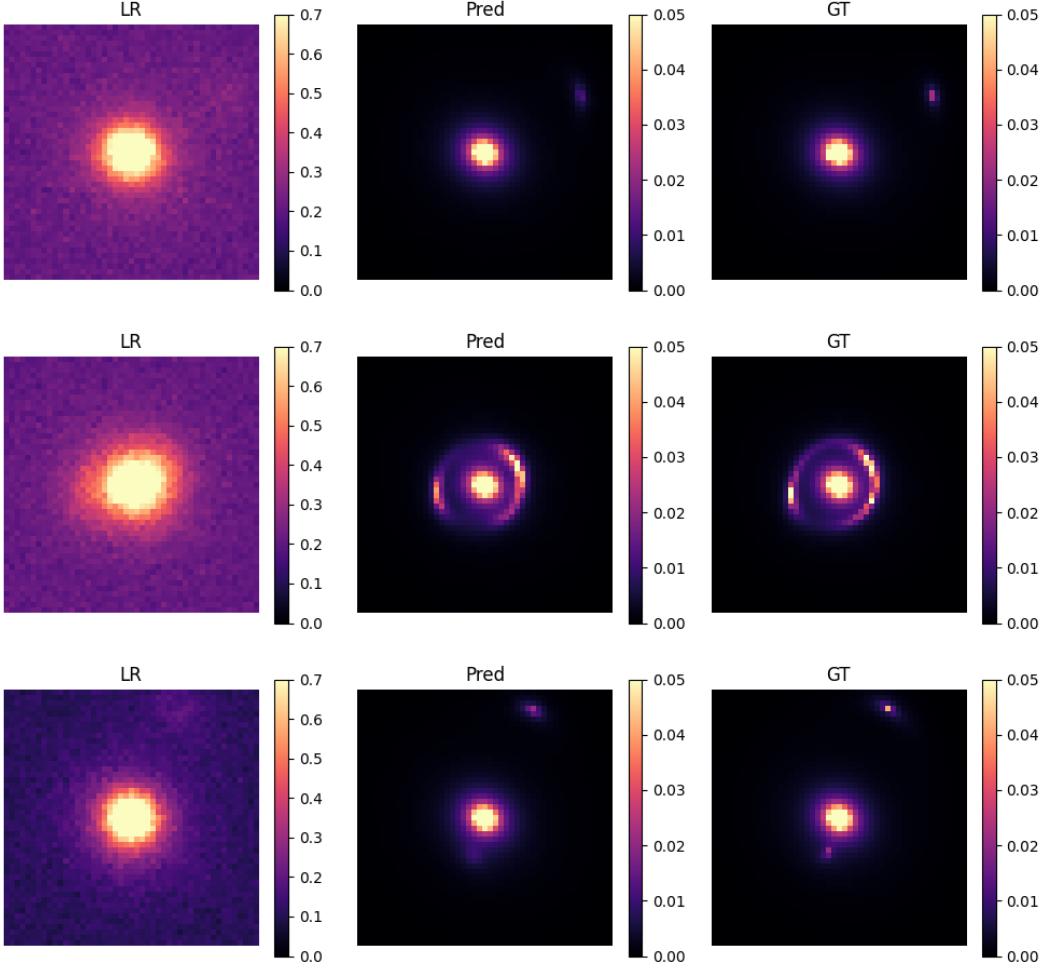


Figure 1: Example of the SwinIR performance on the generated dataset. Left to right: Low resolution image - input to the model, simulating the performance of the LSST telescope; Prediction of high resolution image by the model; Ground truth high resolution image.

Table 2: Einstein radius estimation performance

Metric	High-Res	Low-Res	RCAN	SwinIR	SatGAN
MAE	0.0425	0.0898	0.0834	0.0907	0.0782
MSE	0.0027	0.0168	0.0134	0.0173	0.0223

4 Conclusion

We presented a new synthetic dataset of paired high- and low-resolution strong lensing images, designed to mimic the gap between Roman-like space imaging and LSST-like ground-based observations. Using this benchmark, we compared four super-resolution paradigms: CNNs, Transformers, conditional diffusion, and GAN-based approaches.

Among the models, SatGAN achieves the strongest overall results, with SwinIR performing very closely and RCAN only slightly behind. However, unlike SatGAN, SwinIR does not use point spread function as an input, which might be a benefit in cases where it is not explicitly known. The conditional diffusion model, which in DiffLense benefited from the harder task setup to deliver improved perceptual quality, falls short in our setting due to known diffusion artifacts such as residual noise. Taken together, these results suggest that super-resolution of LSST-like images is

highly promising: under controlled conditions, modern deep learning approaches can already recover space-like detail with near-perfect fidelity, paving the way for robust application to real survey data.

More broadly, the advancement of super-resolution techniques, such as those explored here, has great potential to impact some of the most pressing questions in physics and cosmology. In particular, high-fidelity reconstructions of lensed images will be critical for extracting information about substructure signals, whose morphology is dictated by the underlying dark matter properties. Our study suggests that the RCAN manages to slightly improve the Einstein radius estimation compared to the low resolution baseline, while for SwinIR and SatGan models the results are controversial. The fact that all of the results are pretty far from the original high-resolution ones reflects the reality that some information is completely erased from the images and may not be accessed through digital processing algorithms. By benchmarking super-resolution applications for lensing, our work helps lay the groundwork for ensuring that the wealth of data from LSST and other high-volume, low-resolution surveys can be fully leveraged for scientific discovery.

References

- Matthias Bartelmann. Gravitational lensing. *Classical and Quantum Gravity*, 27(23):233001, November 2010. ISSN 1361-6382. doi: 10.1088/0264-9381/27/23/233001. URL <http://dx.doi.org/10.1088/0264-9381/27/23/233001>.
- Simon Birrer and Adam Amara. lenstronomy: multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, 22:189–201, 2018.
- Daniel Gilman, Jo Bovy, Tommaso Treu, Charles R. Keeton, and Anna M. Nierenberg. Pyhalo: an open-source python package for substructure lensing with line-of-sight halos. *Monthly Notices of the Royal Astronomical Society*, 491(4):6077–6090, 2020.
- 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, et al. Detection of lensing substructure using alma observations of the dusty galaxy sdp. 81. *The Astrophysical Journal*, 823(1):37, 2016.
- Zeljko Ivezic et al. Lsst science book, version 2.0. *arXiv preprint arXiv:0912.0201*, 2009.
- Jingyun Liang, Jiezhang Cao, Kai Zhang Fan, Yulun Zhang, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1833–1844, 2021.
- Pranath Reddy, Michael W. Toomey, Hanna Parul, and Sergei Gleyzer. DiffLense: A conditional diffusion model for super-resolution of gravitational lensing data, 2024. *arXiv preprint*.
- Barnaby T. P. Rowe, M. Jarvis, R. Mandelbaum, J. Bosch, M. Simet, J. E. Meyers, T. Kacprzak, R. Nakajima, J. Zuntz, and et al. Galsim: The modular galaxy image simulation toolkit. *Astronomy and Computing*, 10:121–150, 2015.
- David Spergel, Neil Gehrels, Charles Baltay, David Bennett, James Breckinridge, Megan Donahue, Alan Dressler, B Scott Gaudi, Tom Greene, Olivier Guyon, and et al. Wide-field infrared survey telescope-astrophysics focused telescope assets wfIRST-afta final report. *arXiv preprint arXiv:1503.03757*, 2015.
- Bryce Wedig et al. The Roman View of Strong Gravitational Lenses. *Astrophys. J.*, 986(1):42, 2025. doi: 10.3847/1538-4357/adc24f.
- Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision (ECCV)*, pages 286–301, 2018.