

---

# Learning the Evolution of Physical Structure of Galaxies via Diffusion Models

---

Andrew Lizarraga<sup>1</sup>, Eric Hanchen Jiang<sup>1</sup>, Jacob Nowack<sup>2</sup>,  
Yun Qi Li<sup>3</sup>, Ying Nian Wu<sup>1</sup>, Bernie Boscoe<sup>2</sup>, Tuan Do<sup>4</sup>

<sup>1</sup>Department of Statistics and Data Science, UCLA

<sup>2</sup>Department of Computer Science, Southern Oregon University

<sup>3</sup>Department of Physics and Astronomy, University of Washington

<sup>4</sup>Department of Physics and Astronomy, UCLA

{andrewlizarraga, ericjiang0318, yunqil}@g.ucla.edu,  
{boscoe, nowackj}@sou.edu, ywu@stat.ucla.edu, tdo@astro.ucla.edu  
[https://github.com/astrodatalab/lizarraga\\_2024](https://github.com/astrodatalab/lizarraga_2024)

## Abstract

In astrophysics, understanding the evolution of galaxies is primarily through imaging data is fundamental to comprehending the formation of the Universe. This paper introduces a novel approach to conditioning Denoising Diffusion Probabilistic Models (DDPM) on redshifts for generating galaxy images. We explore whether this advanced generative model can accurately capture the physical characteristics of galaxies based solely on their images and redshift measurements. Our findings demonstrate that this model not only produces visually realistic galaxy images but also encodes the underlying changes in physical properties with redshift that are the result of galaxy evolution. This approach marks a significant advancement in using generative models to enhance our scientific insight into cosmic phenomena.

## 1 Introduction

Understanding galaxy formation and evolution is central to astrophysics, yet observational limitations restrict our ability to capture galaxies across cosmic timescales. Redshift-conditioned generative models help by simulating galaxies in underexplored regions, thus offering new insights into galaxy evolution and cosmic structure. Recently, Denoising Diffusion Probabilistic Models (DDPM) models [1] have emerged as a promising generative model class, achieving state-of-the-art results in generating high-fidelity images [1, 2, 3].

DDPMs operate by gradually adding noise to data through a forward diffusion process and then learning to reverse this process to generate new samples. Their ability to model complex distributions makes them suitable candidates for generating galaxy images conditioned on specific properties, such as redshift, which corresponds approximately to the distance of a galaxy.

## 2 Related Work

Recent efforts [4, 5] have employed diffusion models in astronomy by discretizing continuous redshift values to adapt to the discrete-time framework of these models. However, this discretization process inherently leads to information loss, which in turn limits the model's ability to accurately learn the continuous distribution  $p(X^z | z)$  thereby impacting the precision of the generated galaxy images conditioned on redshift. Similar approaches, such as those by Xue et al. [6], have explored the use of

DDPMs for Point Spread Function (PSF) deconvolution, but their method, distinct from ours, does not address the limitations of discrete stepwise conditioning. Lanusse et al. [7] and Margalef et al. [8] utilized Generative Adversarial Networks (GANs) with redshift as a conditional input to generate synthetic galaxy images, simulating the visual characteristics of galaxies across different distances and observational scenarios. However these GANs struggle with mode collapse and benchmarks were compared with perceptual scores as opposed to true galaxy morphology.

### 3 Contributions

To overcome these limitations, we propose a novel adaptation of DDPMs, specifically tailored for generating galaxy images across a continuous range of redshifts without the need for discretization or the introduction of a secondary redshift encoding model. Our main contributions are as follows:

- We develop a new approach that directly conditions the DDPM on continuous redshift values, significantly enhancing the model’s accuracy and fidelity.
- Our findings demonstrate that our model can implicitly learn the morphological characteristics of galaxies without explicit input regarding these attributes, thereby suggesting that redshift alone is predictive of galaxy morphology.

### 4 Data

For our analysis, we employ a subset of the *Hyper Suprime-Cam Galaxy Dataset* curated by Do et al. [9], which is publicly accessible at Zenodo (GalaxiesML: <https://zenodo.org/records/11117528> CC-BY 4.0). This dataset is based on the data released by the Hyper Suprime-Cam survey, as detailed by Aihara et al. [10]. It comprises 286,401 galaxies, spanning redshifts from 0 to 4. Each galaxy is represented by images taken in five visible wavelength bands—( $g, r, i, z, y$ ) filters. We use the  $64 \times 64$  pixel images from GalaxiesML. The dataset includes accurate spectroscopic measurements of each galaxy’s true redshift (or distance from Earth). Due to the selection process, the dataset exhibits a bias toward lower redshifts, with approximately 92.8% of the galaxies having redshifts less than 1.5. We adhere to the training and testing split proposed by Li et al. [4], resulting in a training set comprising 204,513 images and a testing set containing 40,914 images.

### 5 Methods

#### 5.1 Continuous Conditioning of DDPMs

Utilizing DDPMs [1], we introduce a novel approach to learn the conditional distribution  $p(X^z | z)$  by integrating redshift values into the U-Net architecture’s time steps [4, 5]. To prevent model overfitting and ensure learning is concentrated within a Gaussian neighborhood around specific redshifts  $z$ , Gaussian noise  $\mathcal{N}(0, \sigma)$  is added to the redshifts during training, enhancing the model’s ability to interpolate between nearby redshifts. Our Conditional Denoising U-Net starts with a noisy initial galaxy image  $X_T^z$  and, through iterative denoising informed by both time step and the adjusted redshifts, aims to produce a clean galaxy image  $X_0^z$ . To additionally stabilize the training, we implement an Exponential Moving Average (EMA) [11] and adhere to a standard variance schedule [1, 12] to balance noise addition and preserve data structure.

The model’s diffusion process starts with  $64 \times 64$  pixel galaxies images with 5 channels, which are passed to a noising schedule across 1000 time steps, linearly interpolating noise levels from a Beta Start of  $1 \times 10^{-4}$  to a Beta End of 0.02. Training utilizes Huber Loss for its robustness to outliers, gradient clipping with a max norm of 1.0, and an AdamW optimizer [13] set to a learning rate of  $2 \times 10^{-5}$ . Redshifts are perturbed with Gaussian noise (std dev 0.01) to prevent overfitting and improve generalization. Our UNet model, equipped with self-attention layers, varies channels by resolution stage and includes 4 attention heads with layer normalization and GELU activation [14], applied before and after attention. Temporal and conditional redshift information is encoded using sinusoidal positional encoding of the time step  $t$ , transformed into a 256-dimensional vector. This vector is further modified by adding Gaussian noise to the redshift value  $z + \mathcal{N}(0, 0.01)$ , prior to being fed into the U-Net (refer to 5.1). The model was trained on a single NVIDIA A6000 GPU. *Exact architecture details are made publicly available: [https://github.com/astrodatalab/lizarraga\\_2024](https://github.com/astrodatalab/lizarraga_2024)*

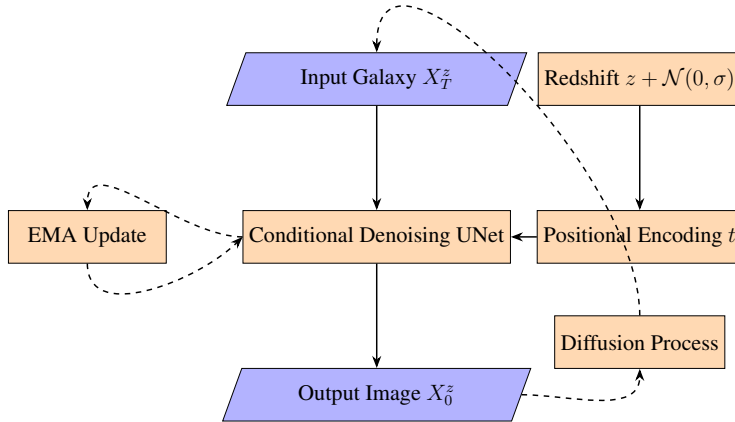


Figure 1: Model Architecture

## 5.2 Evaluation

Our evaluation focuses on the measured physical attributes of galaxies to gauge the physical consistency of our generated images, which involve five color filters ( $g, r, i, z, y$ ). While perceptual quality metrics like Fréchet Inception Distance (FID) [15] and Inception Score (IS) [16] indicate general similarity to true images, they fail to assess critical morphological properties of galaxies and their evolution over time. Our evaluation involves generating synthetic images conditioned on redshifts from the test dataset and comparing to physical properties that astronomers typically use to characterize galaxies [e.g. 17], such as the shape (ellipticity, semi-major axis), size (isophotal area), and brightness distribution (Sersic index). Furthermore, using the CNNRedshift predictor established by Li et al. [4], we assess the redshift accuracy against the ground truth, utilizing the redshift loss from [18]. This redshift predictor was trained on real galaxy images using spectroscopic ground truth and produces good predictions on real data (Fig. 2). These comparisons help verify the physical plausibility of the diffusion model’s output.

## 6 Results

### 6.1 Redshift Prediction

We find that the generated images have redshift predictions that are in good agreement with the redshift that they were generated with as evaluated by the CNNRedshift predictor (Fig. 2). The DDPM produces images with redshift predictions that have slightly larger scatter than with real images, but follows the 1:1 line between conditioned redshift and predicted redshift well up to a redshift about 2. Redshifts beyond 2 are challenging because these redshifts represent less than 2% of the training dataset.

### 6.2 Modeling the Physical Characteristics of Galaxies

We calculate standard metrics for both the test data and the DDPM-generated images, which are conditioned on the test data’s redshifts. Our findings confirm that the DDPM successfully learns the physical characteristics of galaxies—such as the ellipticity, semi-major axis, Sersic index, and isophotal area even though these attributes were never explicitly provided to the model. When comparing the frequencies of each metric between the DDPM and the true distribution, we see in Fig. 3 that the overall shape of the distributions is very close.

Moreso, Fig. 4 illustrates that for each redshift bin, the mean values (represented by red dots) of each metric for DDPM-generated galaxies closely match the means of the true test distribution (blue dots). The ranges of these metrics generally fall within the true distribution’s ranges. This suggests

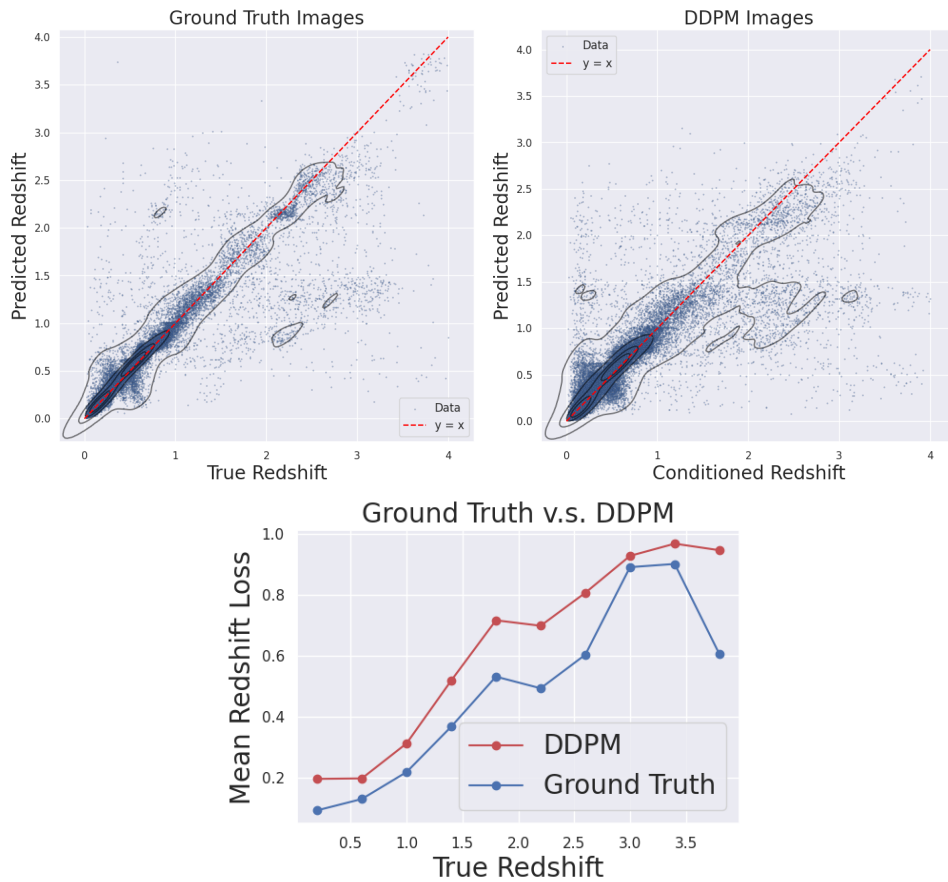


Figure 2: (TOP-LEFT) a scatter plot comparing predicted redshifts to true redshifts for ground truth images, (TOP-RIGHT) a similar scatter plot for DDPM-generated images, (BOTTOM) a plot of true redshift versus mean redshift loss, highlighting the performance accuracy across the redshift range.

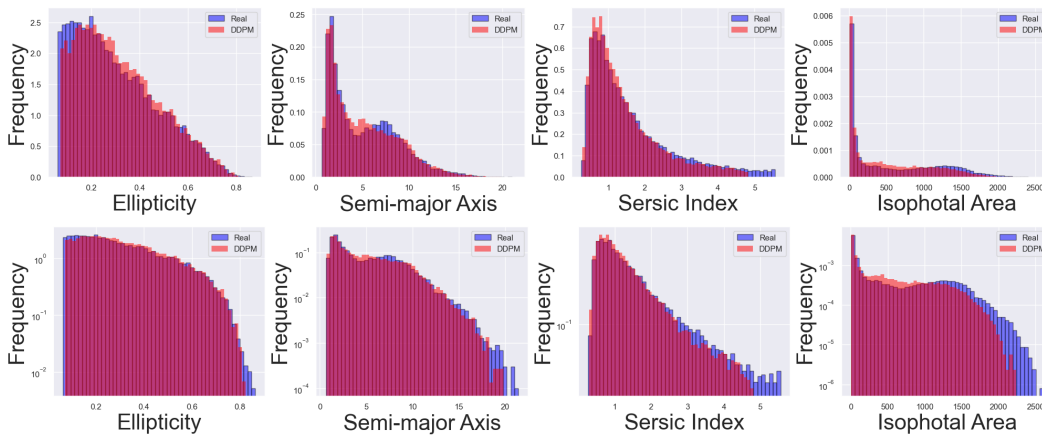


Figure 3: (TOP) From left to right, the figure displays histograms comparing the frequency distribution of DDPM-generated and real galaxies in terms of 1) ellipticity, 2) semi-major axis, 3) Sersic index, and 4) isophotal area. (BOTTOM) Log-Scale of (TOP)

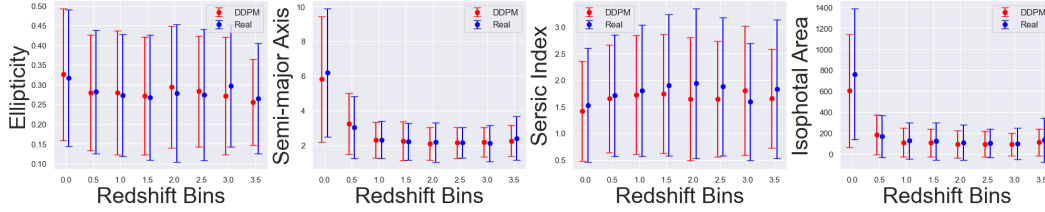


Figure 4: From left to right, the figure displays 95% CIs comparing DDPM-generated and real galaxies across redshift bins: 1) ellipticity, 2) semi-major axis, 3) Sersic index, and 4) isophotal area).

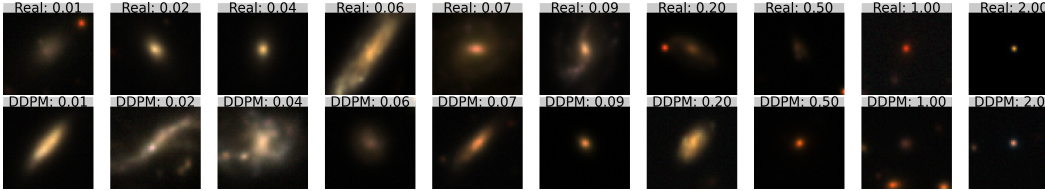


Figure 5: (Top) Real galaxies and corresponding redshifts and (Bottom) DDPM generated galaxies. Both rows correspond to respective redshifts.

that the DDPM model is able to associate redshifts with morphological characteristics of galaxies observed at that redshift. For example, the galaxies tend to be more compact at higher redshifts but the distribution of ellipticity does not change much with redshift, consistent with the testing dataset.

Recall that Fig. 2 indicates a greater variance in detected redshifts. We anticipate the model to produce a broader range of generated images, potentially blending characteristics from neighboring redshift values. This effect is evident in Fig. 5, where the model generates images that display increased diversity and variability.

## 7 Conclusion

In this work, we introduced a novel approach to generating galaxy images using Denoising Diffusion Probabilistic Model (DDPM), conditioned on continuous redshift values. Our results show that the DDPM effectively captures essential physical attributes of galaxies, such as semi-major axis, isophotal area, ellipticity, and Sersic index, with high fidelity to the true data distribution. This suggests that redshift, a measure of both age and distance, serves as a strong predictor of galaxy structure, even without direct morphological inputs.

Future work should focus on extending this approach toward models that can learn the physical evolution of galaxies more directly. Reproducing the morphological characteristics [e.g., 17] is the first step to embed the physics of galaxy evolution into a neural network. To demonstrate more direct connection to physics, one should also apply more stringent tests. For example, it's uncertain if models produce galaxies that have the same star formation rate density evolution [e.g., 19] or physical changes through galaxy mergers [20].

Moreover, considering DDPM's ability to interpolate between modes of the learned probability distribution, we propose raises the question if DDPM's can be utilized for dynamic visualizations of galaxy evolution as a function of redshift. Such a framework could serve as a powerful tool for studying galaxy formation and evolution across cosmic timescales.

## Acknowledgments and Disclosure of Funding

This work was partially supported by CSREx - SOU, NSF DMS-2015577, NSF DMS-2415226, and a gift fund from Amazon.

## References

- [1] 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.
- [2] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021.
- [3] Prafulla Dhariwal and Alexander Nichol. Diffusion Models Beat GANs on Image Synthesis. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794. Curran Associates, Inc., 2021.
- [4] Yun Qi Li, Tuan Do, Evan Jones, Bernie Boscoe, Kevin Alfaro, and Zooley Nguyen. Using Galaxy Evolution as Source of Physics-Based Ground Truth for Generative Models, 2024.
- [5] 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, 01 2022.
- [6] Zhiwei Xue, Yuhang Li, Yash J. Patel, and Jeffrey Regier. Diffusion Models for Probabilistic Deconvolution of Galaxy Images. *ArXiv*, abs/2307.11122, 2023.
- [7] François Lanusse, Rachel Mandelbaum, Siamak Ravanbakhsh, Chun-Liang Li, Peter Freeman, and Barnabás Póczos. Deep generative models for galaxy image simulations. *Monthly Notices of the Royal Astronomical Society*, 504(4):5543–5555, 05 2021.
- [8] Berta Margalef-Bentabol, Marc Huertas-Company, Tom Charnock, Carla Margalef-Bentabol, Mariangela Bernardi, Yohan Dubois, Kate Storey-Fisher, and Lorenzo Zanisi. Detecting outliers in astronomical images with deep generative networks. *Monthly Notices of the Royal Astronomical Society*, 496(2):2346–2361, 06 2020.
- [9] Tuan Do, Evan Jones, Bernie Boscoe, Yunqi (Billy) Li, and Kevin Alfaro. GalaxiesML: an imaging and photometric dataset of galaxies for machine learning, June 2024.
- [10] Makoto Ando Hiroaki Aihara, Yusra AlSayyad and et al. Second data release of the Hyper Suprime-Cam Subaru Strategic Program. *Publications of the Astronomical Society of Japan*, 71(6):114, 10 2019.
- [11] Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, and Samuli Laine. Analyzing and Improving the Training Dynamics of Diffusion Models. In *Proc. CVPR*, 2024.
- [12] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. *ArXiv*, abs/2010.02502, 2020.
- [13] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR)*, 2020.
- [14] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.
- [15] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pages 6626–6637, 2017.
- [16] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved Techniques for Training GANs. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, pages 2234–2242, 2016.
- [17] Christopher J. Conselice. The Evolution of Galaxy Structure Over Cosmic Time. *Annual Review of Astronomy and Astrophysics*, 52(1):291–337, 2014.

- [18] Atsushi J. Nishizawa, Bau-Ching Hsieh, Masayuki Tanaka, and Tadafumi Takata. Photometric Redshifts for the Hyper Suprime-Cam Subaru Strategic Program Data Release 2, 2020.
- [19] Piero Madau and Mark Dickinson. Cosmic Star-Formation History. *Annual Review of Astronomy and Astrophysics*, 52:415–486, August 2014.
- [20] Jennifer M. Lotz, Patrik Jonsson, T. J. Cox, and Joel R. Primack. Galaxy merger morphologies and time-scales from simulations of equal-mass gas-rich disc mergers. *Monthly Notices of the Royal Astronomical Society*, 391:1137–1162, December 2008.