Learning Galaxy Evolution via Diffusion Models

Anonymous Author(s)
Affiliation
Address
email

Abstract

In astrophysics, understanding the evolution of galaxies in large part through imaging data is fundamental to comprehending the formation of the Universe. This paper introduces a new approach to conditioning Denoising Diffusion Probabilistic Models (DDPM) on redshifts for generating galaxy images. We explore whether this advanced generative model can 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 step in using generative models to enhance our scientific insight into cosmic phenomena.

1 Introduction

Understanding galaxy formation and evolution is central to astrophysics, but observational limitations restrict our ability to capture galaxies across cosmic time. Redshift-conditioned generative models help fill these gaps by simulating galaxies in underexplored regions, 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 applied diffusion models in astronomy by discretizing continuous redshift values to adapt to the discrete-time framework of these models. This discretization process inherently leads to information loss, which in turn limits the model's ability to accurately learn the continuous distribution $p(X^z \mid z)$ 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.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

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, 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 Zendo (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 64×64 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 DDPM

Utilizing DDPMs [1], we introduce a novel approach to learn the conditional distribution $p(X^z \mid z)$ 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 during 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 64×64 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 1×10^{-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 set to a learning rate of 2×10^{-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, 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 and implementations are to be released in a publicly available open sourced github.*

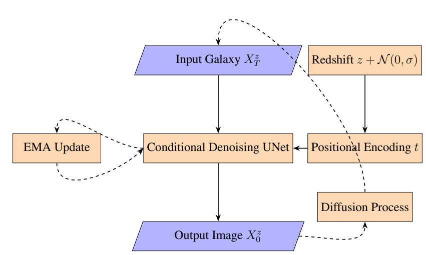

Figure 1: Model Architecture

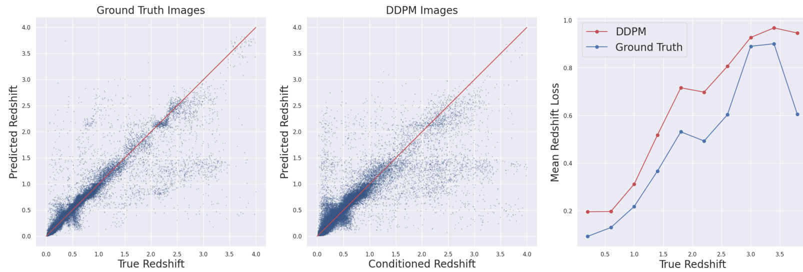

Figure 2: From left to right, the figure displays: 1) a scatter plot comparing predicted redshifts to true redshifts for ground truth images, 2) a similar scatter plot for DDPM-generated images, 3) a plot of true redshift versus mean redshift loss, highlighting the performance accuracy across the redshift range.

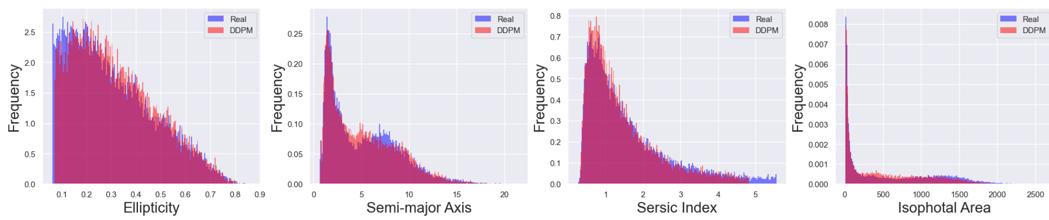

Figure 3: From left to right, the figure displays histograms comparing the frequency distribution of DDPM-generated and real galaxies in terms of 1) ellipticity, 2) semi-major axis, 3) Sersic index, and 4 isophotal area).

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) [13] and Inception Score (IS) [14] 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 properites that astronomers typically use to characterize galaxies, 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 [15]. 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 challenge because these redshifts represent less than 2% of the training dataset.

6.2 Galaxy Morphology

We calculate standard metrics on both the test data and the DDPM-generated images conditioned on the test data's redshifts. Our findings confirm that the DDPM successfully learns the physical

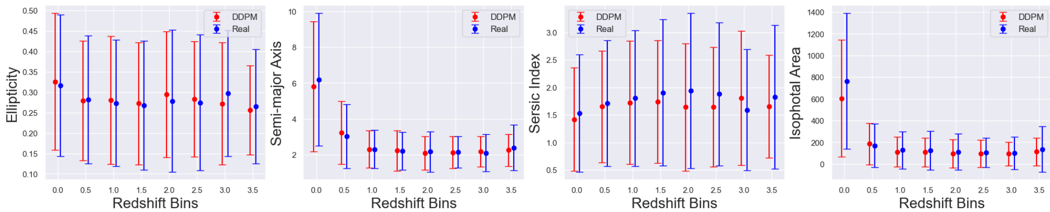

Figure 4: From left to right, the figure displays 95% CIs comparing DDPM-generated and real galaxies across redshift bins: 1) ellipticity, 2) semi-major axis, 3) Sersic index, and 4 isophotal area)

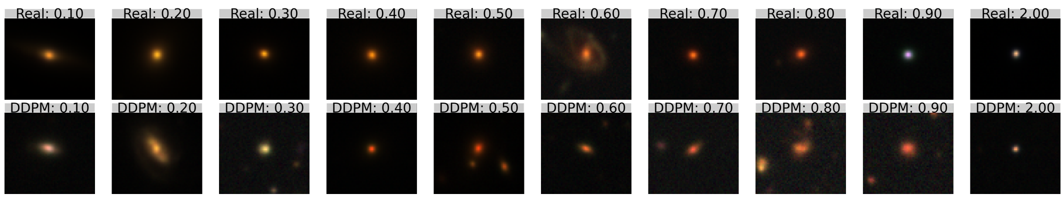

Figure 5: (Top) Real galaxies and corresponding redshifts and (Bottom) DDPM generated galaxies. Both rows correspond to respective redshifts 0.10 to 0.90 and the final image at redshift 2.00.

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 that the DDPM model is able to associate redshifts with morphological characteristics of galaxies observed at that redshift.

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.

6.3 Limitations

While our model successfully captures key physical properties of galaxies, it is limited by the training dataset's bias toward lower redshifts, which affects its performance at higher redshift values (See Fig. 2). Additionally, the generated images may exhibit increased variability (Fig. 5), particularly in underrepresented redshift ranges, potentially blending characteristics from neighboring redshifts.

7 Conclusion

In this work, we introduced a novel approach to generating galaxy images using Denoising Diffusion Probabilistic Models (DDPM), conditioned on continuous redshift values. Our empirical analysis demonstrates that conditioning the model solely on redshift enables it to implicitly learn key morphological characteristics of galaxies without requiring explicit morphological information. This finding suggests that redshift, a measure of both age and distance, can serve as a robust predictor of galaxy structure.

Our results show that the DDPM captures essential physical attributes, such as semi-major axis, isophotal area, ellipticity, and Sersic index, with high fidelity to the true data distribution. The model's ability to generalize these attributes, conditioned solely on redshift and image data, supports the hypothesis that redshift is intricately linked to galaxy morphology. This finding not only enhances our understanding of galaxy formation but also establishes DDPMs as a valuable tool for simulating realistic galaxy populations across cosmic timescales.

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 Zooey 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] 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.

[14] 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.

[15] 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.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

1. **Claims**

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

   Answer: [Yes]

   Justification: Claims are made in the abstract and in Sec. 1 are discussed through out the paper. See Sec. 4, Sec. 5, Sec. 6.

   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: See Sec. 6.3.

   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: The paper is an empirical analysis without theoretical results.

   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: Methods, data and experimental setup are provided in detail in Sec. 5, Sec. 4 and Sec. 6 respectively.

   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: [Yes]

   Justification: Details of the model architecture and training are fully described in Sec. 5 and the model weights and training scripts are planned to be open sourced via github. Data uses an opensource dataset as described in 4 and is readily available at: Zendo (GalaxiesML: `https://zenodo.org/records/11117528` CC-BY 4.0)

   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: See Sec. 5 and Sec. 6.

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: Reported in Fig. 4 and discussed further in Sec. 6.

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: Discussed in Sec. 5, the model is trained on a single NVIDIA A6000 GPU.

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: None.

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: [NA]

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: [NA]

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: Liscenses for data are found at Zendo (GalaxiesML: `https://zenodo.org/records/11117528` CC-BY 4.0) and is cited in Sec. 4.

    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: [NA]

    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: [NA]

    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: [NA]

    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.