Conditional Diffusion Models for Generating Images of SDSS-Like Galaxies

Mikaeel Yunus
Department of Physics and Astronomy
Johns Hopkins University
Baltimore, MD 21218
myunus1@jh.edu

John F. Wu
Space Telescope Science Institute
Baltimore, MD 21218
jowu@stsci.edu

Timothy M. Heckman
Department of Physics and Astronomy
Johns Hopkins University
Baltimore, MD 21218
theckma1@jhu.edu

Benne W. Holwerda
Department of Physics and Astronomy
University of Louisville
Louisville, KY 40208
benne.holwerda@louisville.edu

Abstract

We present a novel application of conditional diffusion models for creating synthetic images of galaxies based on their physical properties. While previous work has focused on conditioning on a single parameter, we demonstrate that diffusion models can generate galaxies conditioned on multiple physical properties simultaneously, allowing us to explore how combinations of these properties correlate with galaxy appearance. Our model, trained on data from the Sloan Digital Sky Survey (SDSS), generates galaxy images conditioned on redshift, stellar mass, star formation rate, and gas-phase metallicity. Notably, the model captures expected astrophysical trends, such as the relationship between metallicity and galaxy color or morphology. However, the generated images disagree with SDSS validation images as measured by Gini coefficients, M_{20} coefficients, and Concentration-Asymmetry-Smoothness statistics, which is consistent with systematic underprediction of diffuse flux. While modern generative models are capable of producing realistic images, applying these models to astrophysics may still prove challenging.

1 Introduction

Artificial intelligence has revolutionized astrophysics, but the subfield of galaxy formation and evolution faces a unique challenge: while data is plentiful, the connection between a galaxy's physical properties and its observable characteristics remains poorly understood. Due to the complex and diverse range of galaxy appearances, we are unable to generate realistic-looking galaxies based on their actual properties, despite our knowledge of galaxy physics. Even simulations can struggle to match the realistic morphologies of all galaxy populations. This limitation hinders our ability to connect the underlying physics of galaxies with their detailed morphologies, impeding a vital component of our understanding of galaxy evolution.

Recent advances in generative machine learning offer promising solutions for understanding galaxy properties and their astrophysical traits. Methods including CNNs [17, 15], VAEs [14], and GANs [6] have been applied to generate galaxy images [8, 16, 30, 3, 12]. More recently, researchers have turned to diffusion models [24, 18], which have been shown to demonstrate superior image sample quality [5]. In particular, [18] showed that conditional diffusion models can generate galaxy images at a specified redshift.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

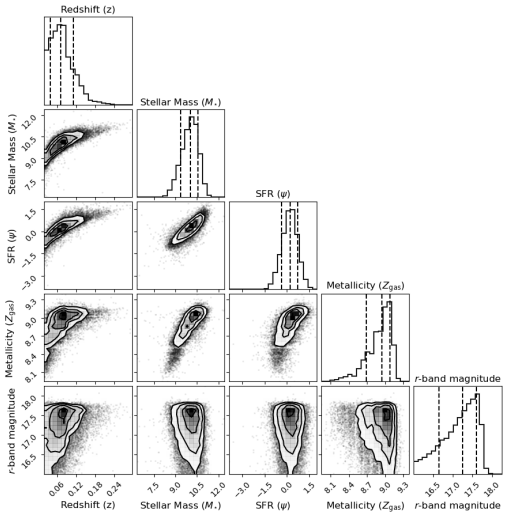

Figure 1: Univariate and bivariate distributions of SDSS galaxy properties and r-band magnitudes. The 16th, 50th, and 84th percentiles of the univariate distributions are labeled with dashed lines.

We propose a generative model that can produce realistic galaxy images conditioned on physical parameters by taking a standard diffusion model and giving it embedded knowledge of observed galaxy properties. The specific physical parameters we use are fundamentally linked to galaxy evolution: stellar mass indicates the total number of stars, star formation rate measures ongoing stellar birth, and metallicity traces chemical enrichment history. Successfully generating images conditioned on these properties would demonstrate that our model captures the complex relationships between galaxy physics and appearance.

In the following sections, we describe our data (§2), detail our methodology (§3), present results evaluating the model's effectiveness (§4), and discuss the limitations and future directions of our work (§5).

2 Data

We generate artificial galaxies that resemble those from the Sloan Digital Sky Survey [29, 1] Main Galaxy Sample ($r < 17.78$; [27]). Starting with the SDSS `galSpecExtra` catalog of spectroscopically confirmed galaxies [13, 2, 28, 23], we remove objects with artificially low gas metallicities ($Z_{\rm gas} < 0$), stellar masses ($M_\star < M_\odot$), and star formation rates ($\psi < 10^{-10} M_\odot$ yr^{-1}). We also discard objects with model magnitudes brighter than $r < 16$ (see Figure 1). We also remove galaxies classified as LINERs and AGN, since they may have biased properties. To save compute resources, we draw a random 20% of the remaining sample (21,364 galaxies). Due to GPU memory constraints during diffusion model inference, we reserve 512 samples for validation and train on the rest. In Figure 1, we show distributions of the redshifts z and physical properties ($M_\star$, ψ, $Z_{\rm gas}$) of the galaxies in our total sample.

We download galaxy image cutouts in JPG format from the SDSS SkyServer. Each 128×128 gri-band image cutout is centered on the galaxy and delivered at the native SDSS pixel scale ($0.396''$ pixel^{-1}). We remove sky background flux from the images by creating a segmentation map using the `statmorph` Python package [20] and then subtracting the median value from each color channel. This sky-subtraction process ensures that there is no extra background flux, which makes training easier for our generative machine learning pipeline. Sky background subtraction is also a standard practice in astronomical image processing. While this preprocessing could potentially affect the generation of low surface brightness features, it ensures our model learns intrinsic galaxy properties rather than observational artifacts. The systematic underprediction of diffuse flux noted in Section 5 could potentially be attributed, at least in part, to this preprocessing step. This highlights an important trade-off between using clean training data and capturing extended low-surface-brightness features.

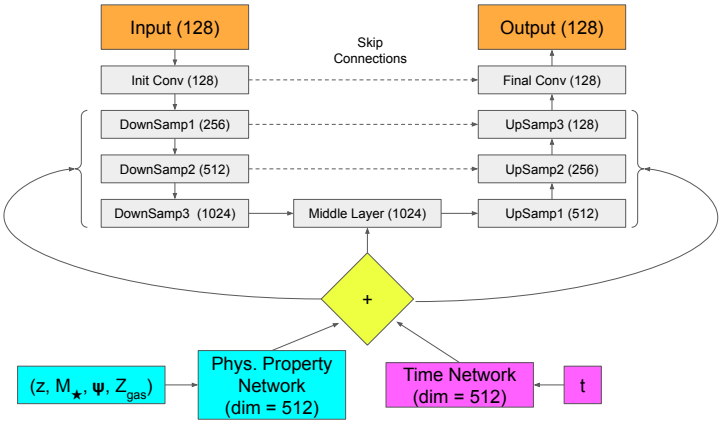

Figure 2: Architecture of the U-Net in our conditional diffusion model. "DownSamp" and "UpSamp" blocks represent downsampling and upsampling layers, respectively. Numbers indicate output channels per layer. The central yellow "+" combines the 512-dimensional outputs from the physical property and time neural networks, feeding the result to all U-Net layers.

3 Methodology

Conditional Diffusion Models. Diffusion models are able to generate highly realistic images based on a training data set [25, 9]. Beginning with a image of noise, diffusion models iteratively subtract Gaussian-distributed noise until an image emerges after T iterations. A U-Net [21] learns to predict how much noise remains at every iteration $t \in \{0, \ldots, T\}$. This trajectory can also be expressed as the solution to a reverse-iteration stochastic differential equation that depends *only* on the score of the data distribution [26].

The U-Net can adjust its predictions using information other than t [10]. We optimize the U-Net to predict noise as a function of galaxy properties $(z, M_\star, \psi, Z_{\rm gas})$ and iteration step t. Because the model is now conditioned on physical properties of galaxies, we refer to our approach as a conditional diffusion model. We also normalize galaxy properties to zero mean and unit variance. Figure 2 shows the U-Net architecture in our conditional diffusion model, where the "Time Network" incorporates the denoising iteration t, and the "Phys. Property Network" conditions on galaxy properties. The "Time Network" processes the denoising iteration t through two linear layers of width 512. Similarly, the "Physical Property Network" processes the normalized galaxy properties $(z, M_\star, \psi, Z_{\rm gas})$ through one linear layer of width 512. The outputs of both networks are added together and fed to all U-Net layers.

Our conditioning approach uses separate fully-connected neural networks to process both diffusion model timesteps and physical properties, combining the outputs of these networks through simple addition. Although more complex conditioning strategies exist, our results demonstrate that this simple approach successfully captures key astrophysical trends in the generated images.

Optimization. We train our conditional diffusion model against the L_1 loss using the Adam optimizer with no weight decay.[1] Gaussian noise is added according to a linear variance schedule [9]. We choose $T = 1000$ iterations for the denoising procedure. We train for 100 epochs with a batch size of 128; training takes 5.5 hours when parallelized across four NVIDIA V100 GPUs.

4 Results

Conditional Generation of Galaxy Images. We are able to generate realistic images of galaxies conditioned on their physical properties. From the SDSS validation set, we sample galaxies with different gas-phase metallicities (e.g., left panel of Figure 3). Then, we condition our trained diffusion

[1]We find that L_1 loss produces more realistic galaxy images than the mean squared error loss. Other works have explored this topic in more detail [22]. We also find that non-zero weight decay produces poor results.

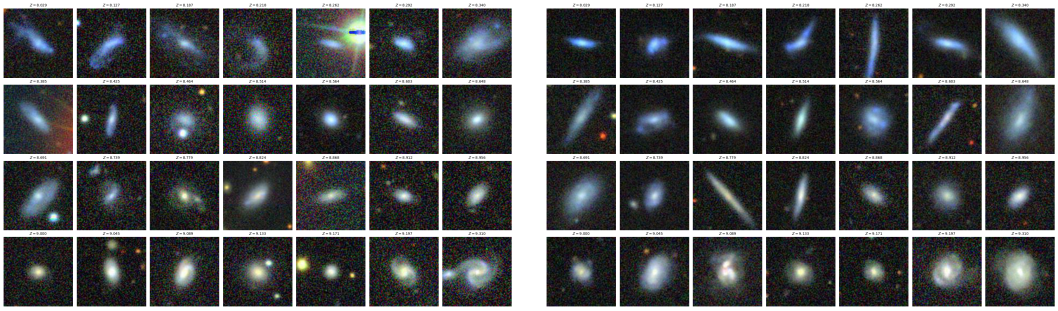

Figure 3: SDSS validation images (*left*) and generated images (*right*). Both sets of images are selected to have the same gas-phase metallicities, and are shown in increasing order of metallicity.

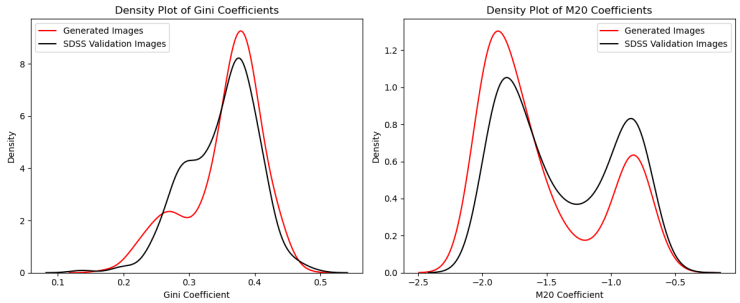

Figure 4: Gini and M_{20} coefficients of generated (red) and validation (black) images.

model to generate galaxy images corresponding to the same metallicites (shown in right panel of Figure 3). The generated images appear realistic: the galaxies exhibit spiral arms, star-forming regions bulges, and bars, and sometimes the generated image cutouts also feature other objects that resemble foreground stars or companion galaxies. However, we note the generated images contain less noise than the validation images. We find similar results when conditioning on other galaxy properties like the specific star formation rate (sSFR $\equiv \psi/M_\star$).

Morphology Metrics. We quantify the morphological similarity between generated and validation images. Using `statmorph` [20], we compute the Gini and M_{20} coefficients [19], as well as concentration, asymmetry, and smoothness (known as the CAS statistics; [4]). These morphology metrics are commonly used in astronomy to describe galaxy shapes [11].

In Figure 4, we present a comparison of the Gini and M_{20} coefficients of the generated and validation images. The Gini coefficient measures how equitably light is distributed throughout a galaxy, such that a high Gini coefficient implies that most of the light resides in a small number of pixels. The M_{20} coefficient measures the spatial spread of the brightest 20% of a galaxy's pixels. We conduct a Kolmogorov-Smirnov (KS) test to determine whether generated images are drawn from the same sample as validation images; we find $p = 5 \times 10^{-3}$ for Gini coefficients, and $p \ll 10^{-5}$ for M_{20} coefficients. This indicates that the generated and validation images are from distinct samples.

In Figure 5, we show CAS statistics for the generated and validation images. Concentration quantifies how concentrated the light of a galaxy is towards its center, asymmetry measures the degree to which a galaxy image is not rotationally symmetric, and smoothness indicates the presence of small-scale structures relative to the underlying large-scale structure (unintuitively, a high smoothness value implies a galaxy that is not smooth). Like we did for our Gini and M_{20} coefficients above, we conduct a KS test and find $p = 2 \times 10^{-4}$ for concentration, $p = 10^{-3}$ for asymmetry, and $p \ll 10^{-5}$ for smoothness. Again, this shows that the generated and validation images are drawn from distinct samples.

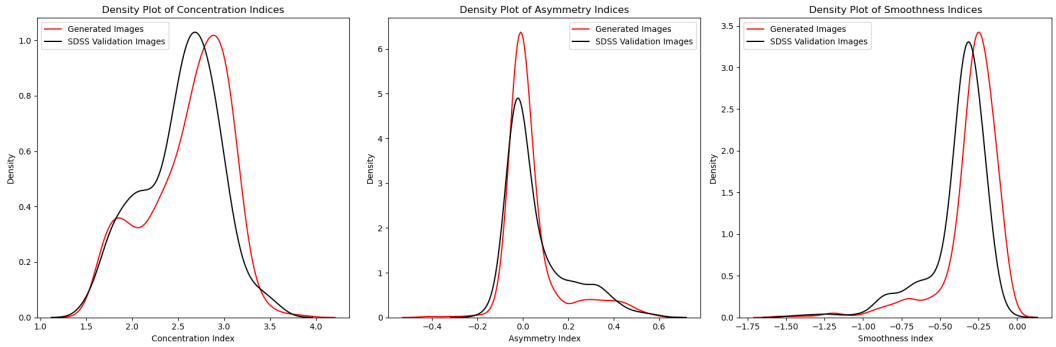

Figure 5: CAS statistics of generated (red) and validation (black) images.

5 Discussion

We have developed a conditional diffusion model capable of generating SDSS-like galaxy images conditioned on redshift, stellar mass, star formation rate, and gas-phase metallicity. Our generated images align with astrophysical intuition and exhibit the expected trends (e.g., lower-metallicity galaxies are bluer in Figure 3). However, when measured against Gini-M_{20} coefficients and CAS statistics, our conditional diffusion model results disagree with SDSS validation images.

In Section 4, we report KS statistics with p-values that are order 10^{-3} at best and are several orders of magnitude smaller at worst. This reflects a poor alignment between generated and validation images for the morphological metrics we select. We theorize that the generated galaxies may be missing diffuse light that can be seen in the outskirts of some real SDSS galaxy images (see, e.g., Figure 3). If low surface-brightness emission is systematically absent from generated images, then we expect M_{20} to be biased low and concentration and smoothness to be biased high (see Figures 4 and 5). In the future, we can address this issue by training and validating on larger galaxy samples, tweaking our model architecture and optimization hyperparameters, and benchmarking on a larger set of morphology metrics used throughout astronomy and computer vision [7].

References

[1] Kevork N. Abazajian, Jennifer K. Adelman-McCarthy, Marcel A. Agüeros, Sahar S. Allam, Carlos Allende Prieto, Deokkeun An, Kurt S. J. Anderson, Scott F. Anderson, James Annis, Neta A. Bahcall, C. A. L. Bailer-Jones, J. C. Barentine, Bruce A. Bassett, Andrew C. Becker, Timothy C. Beers, Eric F. Bell, Vasily Belokurov, Andreas A. Berlind, Eileen F. Berman, Mariangela Bernardi, Steven J. Bickerton, Dmitry Bizyaev, John P. Blakeslee, Michael R. Blanton, John J. Bochanski, William N. Boroski, Howard J. Brewington, Jarle Brinchmann, J. Brinkmann, Robert J. Brunner, Tamás Budavári, Larry N. Carey, Samuel Carliles, Michael A. Carr, Francisco J. Castander, David Cinabro, A. J. Connolly, István Csabai, Carlos E. Cunha, Paul C. Czarapata, James R. A. Davenport, Ernst de Haas, Ben Dilday, Mamoru Doi, Daniel J. Eisenstein, Michael L. Evans, N. W. Evans, Xiaohui Fan, Scott D. Friedman, Joshua A. Frieman, Masataka Fukugita, Boris T. Gänsicke, Evalyn Gates, Bruce Gillespie, G. Gilmore, Belinda Gonzalez, Carlos F. Gonzalez, Eva K. Grebel, James E. Gunn, Zsuzsanna György, Patrick B. Hall, Paul Harding, Frederick H. Harris, Michael Harvanek, Suzanne L. Hawley, Jeffrey J. E. Hayes, Timothy M. Heckman, John S. Hendry, Gregory S. Hennessy, Robert B. Hindsley, J. Hoblitt, Craig J. Hogan, David W. Hogg, Jon A. Holtzman, Joseph B. Hyde, Shin-ichi Ichikawa, Takashi Ichikawa, Myungshin Im, Željko Ivezić, Sebastian Jester, Linhua Jiang, Jennifer A. Johnson, Anders M. Jorgensen, Mario Jurić, Stephen M. Kent, R. Kessler, S. J. Kleinman, G. R. Knapp, Kohki Konishi, Richard G. Kron, Jurek Krzesinski, Nikolay Kuropatkin, Hubert Lampeitl, Svetlana Lebedeva, Myung Gyoon Lee, Young Sun Lee, R. French Leger, Sébastien Lépine, Nolan Li, Marcos Lima, Huan Lin, Daniel C. Long, Craig P. Loomis, Jon Loveday, Robert H. Lupton, Eugene Magnier, Olena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Bruce Margon, John P. Marriner, David Martínez-Delgado, Takahiko Matsubara, Peregrine M. McGehee, Timothy A. McKay, Avery Meiksin, Heather L. Morrison, Fergal Mullally, Jeffrey A. Munn, Tara Murphy, Thomas Nash, Ada Nebot, Jr. Neilsen, Eric H., Heidi Jo Newberg, Peter R. Newman, Robert C. Nichol, Tom Nicinski, Maria Nieto-Santisteban, Atsuko Nitta, Sadanori Okamura, Daniel J. Oravetz, Jeremiah P. Ostriker, Russell Owen, Nikhil Padmanabhan, Kaike Pan, Changbom Park, George Pauls, Jr. Peoples, John, Will J. Percival, Jeffrey R. Pier, Adrian C. Pope, Dimitri Pourbaix, Paul A. Price, Norbert Purger, Thomas Quinn, M. Jordan Raddick, Paola Re Fiorentin, Gordon T. Richards, Michael W. Richmond, Adam G. Riess, Hans-Walter Rix, Constance M. Rockosi, Masao Sako, David J. Schlegel, Donald P. Schneider, Ralf-Dieter Scholz, Matthias R. Schreiber, Axel D. Schwope, Uroš Seljak, Branimir Sesar, Erin Sheldon, Kazu Shimasaku, Valena C. Sibley, A. E. Simmons, Thirupathi Sivarani, J. Allyn Smith, Martin C. Smith, Vernesa Smolčić, Stephanie A. Snedden, Albert Stebbins, Matthias Steinmetz, Chris Stoughton, Michael A. Strauss, Mark SubbaRao, Yasushi Suto, Alexander S. Szalay, István Szapudi, Paula Szkody, Masayuki Tanaka, Max Tegmark, Luis F. A. Teodoro, Aniruddha R. Thakar, Christy A. Tremonti, Douglas L. Tucker, Alan Uomoto, Daniel E. Vanden Berk, Jan Vandenberg, S. Vidrih, Michael S. Vogeley, Wolfgang Voges, Nicole P. Vogt, Yogesh Wadadekar, Shannon Watters, David H. Weinberg, Andrew A. West, Simon D. M. White, Brian C. Wilhite, Alainna C. Wonders, Brian Yanny, D. R. Yocum, Donald G. York, Idit Zehavi, Stefano Zibetti, and Daniel B. Zucker. The Seventh Data Release of the Sloan Digital Sky Survey. *The Astrophysical Journal Supplement Series*, 182(2):543–558, June 2009.

[2] J. Brinchmann, S. Charlot, S. D. M. White, C. Tremonti, G. Kauffmann, T. Heckman, and J. Brinkmann. The physical properties of star-forming galaxies in the low-redshift Universe. *Monthly Notices of the Royal Astronomical Society*, 351(4):1151–1179, July 2004.

[3] Brandon Buncher, Awshesh Nath Sharma, and Matias Carrasco Kind. Survey2Survey: a deep learning generative model approach for cross-survey image mapping. *Monthly Notices of the Royal Astronomical Society*, 503(1):777–796, May 2021.

[4] Christopher J. Conselice. The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories. *The Astrophysical Journal Supplement Series*, 147(1):1–28, July 2003.

[5] Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis, 2021.

[6] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[7] S. Hackstein, V. Kinakh, C. Bailer, and M. Melchior. Evaluation metrics for galaxy image generators. *Astronomy and Computing*, 42:100685, January 2023.

[8] Ryan Hausen and Brant E. Robertson. Morpheus: A Deep Learning Framework for the Pixel-level Analysis of Astronomical Image Data. *The Astrophysical Journal Supplement Series*, 248(1):20, May 2020.

[9] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.

[10] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022.

[11] B W Holwerda. *Galaxy Morphology*. 2514-3433. IOP Publishing, 2021.

[12] Benjamin J Holzschuh, Conor M O'Riordan, Simona Vegetti, Vicente Rodriguez-Gomez, and Nils Thuerey. Realistic galaxy images and improved robustness in machine learning tasks from generative modelling. *Monthly Notices of the Royal Astronomical Society*, 515(1):652–677, May 2022.

[13] Guinevere Kauffmann, Timothy M. Heckman, Christy Tremonti, Jarle Brinchmann, Stéphane Charlot, Simon D. M. White, Susan E. Ridgway, Jon Brinkmann, Masataka Fukugita, Patrick B. Hall, Željko Ivezić, Gordon T. Richards, and Donald P. Schneider. The host galaxies of active galactic nuclei. *Monthly Notices of the Royal Astronomical Society*, 346(4):1055–1077, December 2003.

[14] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022.

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[16] 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, July 2021.

[17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[18] 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.

[19] Jennifer M. Lotz, Joel Primack, and Piero Madau. A new nonparametric approach to galaxy morphological classification. *The Astronomical Journal*, 128(1):163–182, July 2004.

[20] Vicente Rodriguez-Gomez, Gregory F. Snyder, Jennifer M. Lotz, Dylan Nelson, Annalisa Pillepich, Volker Springel, Shy Genel, Rainer Weinberger, Sandro Tacchella, Rüdiger Pakmor, Paul Torrey, Federico Marinacci, Mark Vogelsberger, Lars Hernquist, and David A. Thilker. The optical morphologies of galaxies in the IllustrisTNG simulation: a comparison to Pan-STARRS observations. *Monthly Notices of the Royal Astronomical Society*, 483(3):4140–4159, March 2019.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[22] Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models, 2022.

[23] Samir Salim, Stéphane Charlot, R. Michael Rich, Guinevere Kauffmann, Timothy M. Heckman, Tom A. Barlow, Luciana Bianchi, Yong-Ik Byun, Jose Donas, Karl Forster, Peter G. Friedman, Patrick N. Jelinsky, Young-Wook Lee, Barry F. Madore, Roger F. Malina, D. Christopher Martin, Bruno Milliard, Patrick Morrissey, Susan G. Neff, David Schiminovich, Mark Seibert, Oswald H. W. Siegmund, Todd Small, Alex S. Szalay, Barry Y. Welsh, and Ted K. Wyder. New Constraints on the Star Formation Histories and Dust Attenuation of Galaxies in the Local Universe from GALEX. *The Astrophysical Journal Letters*, 619(1):L39–L42, January 2005.

[24] 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, January 2022.

[25] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics, 2015.

[26] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations, 2021.

[27] Michael A. Strauss, David H. Weinberg, Robert H. Lupton, Vijay K. Narayanan, James Annis, Mariangela Bernardi, Michael Blanton, Scott Burles, A. J. Connolly, Julianne Dalcanton, Mamoru Doi, Daniel Eisenstein, Joshua A. Frieman, Masataka Fukugita, James E. Gunn, Željko Ivezić, Stephen Kent, Rita S. J. Kim, G. R. Knapp, Richard G. Kron, Jeffrey A. Munn, Heidi Jo Newberg, R. C. Nichol, Sadanori Okamura, Thomas R. Quinn, Michael W. Richmond, David J. Schlegel, Kazuhiro Shimasaku, Mark SubbaRao, Alexander S. Szalay, Dan Vanden Berk, Michael S. Vogeley, Brian Yanny, Naoki Yasuda, Donald G. York, and Idit Zehavi. Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. *The Astronomical Journal*, 124(3):1810–1824, September 2002.

[28] Christy A. Tremonti, Timothy M. Heckman, Guinevere Kauffmann, Jarle Brinchmann, Stéphane Charlot, Simon D. M. White, Mark Seibert, Eric W. Peng, David J. Schlegel, Alan Uomoto, Masataka Fukugita, and Jon Brinkmann. The Origin of the Mass-Metallicity Relation: Insights from 53,000 Star-forming Galaxies in the Sloan Digital Sky Survey. *The Astrophysical Journal*, 613(2):898–913, October 2004.

[29] Donald G. York, J. Adelman, Jr. Anderson, John E., Scott F. Anderson, James Annis, Neta A. Bahcall, J. A. Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, William N. Boroski, Steve Bracker, Charlie Briegel, John W. Briggs, J. Brinkmann, Robert Brunner, Scott Burles, Larry Carey, Michael A. Carr, Francisco J. Castander, Bing Chen, Patrick L. Colestock, A. J. Connolly, J. H. Crocker, István Csabai, Paul C. Czarapata, John Eric Davis, Mamoru Doi, Tom Dombeck, Daniel Eisenstein, Nancy Ellman, Brian R. Elms, Michael L. Evans, Xiaohui Fan, Glenn R. Federwitz, Larry Fiscelli, Scott Friedman, Joshua A. Frieman, Masataka Fukugita, Bruce Gillespie, James E. Gunn, Vijay K. Gurbani, Ernst de Haas, Merle Haldeman, Frederick H. Harris, J. Hayes, Timothy M. Heckman, G. S. Hennessy, Robert B. Hindsley, Scott Holm, Donald J. Holmgren, Chi-hao Huang, Charles Hull, Don Husby, Shin-Ichi Ichikawa, Takashi Ichikawa, Željko Ivezić, Stephen Kent, Rita S. J. Kim, E. Kinney, Mark Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, John Korienek, Richard G. Kron, Peter Z. Kunszt, D. Q. Lamb, B. Lee, R. French Leger, Siriluk Limmongkol, Carl Lindenmeyer, Daniel C. Long, Craig Loomis, Jon Loveday, Rich Lucinio, Robert H. Lupton, Bryan MacKinnon, Edward J. Mannery, P. M. Mantsch, Bruce Margon, Peregrine McGehee, Timothy A. McKay, Avery Meiksin, Aronne Merelli, David G. Monet, Jeffrey A. Munn, Vijay K. Narayanan, Thomas Nash, Eric Neilsen, Rich Neswold, Heidi Jo Newberg, R. C. Nichol, Tom Nicinski, Mario Nonino, Norio Okada, Sadanori Okamura, Jeremiah P. Ostriker, Russell Owen, A. George Pauls, John Peoples, R. L. Peterson, Donald Petravick, Jeffrey R. Pier, Adrian Pope, Ruth Pordes, Angela Prosapio, Ron Rechenmacher, Thomas R. Quinn, Gordon T. Richards, Michael W. Richmond, Claudio H. Rivetta, Constance M. Rockosi, Kurt Ruthmansdorfer, Dale Sandford, David J. Schlegel, Donald P. Schneider, Maki Sekiguchi, Gary Sergey, Kazuhiro Shimasaku, Walter A. Siegmund, Stephen Smee, J. Allyn Smith, S. Snedden, R. Stone, Chris Stoughton, Michael A. Strauss, Christopher Stubbs, Mark SubbaRao, Alexander S. Szalay, Istvan Szapudi, Gyula P. Szokoly, Anirudda R. Thakar, Christy Tremonti, Douglas L. Tucker, Alan Uomoto, Dan Vanden Berk, Michael S. Vogeley, Patrick Waddell, Shu-i. Wang, Masaru Watanabe, David H. Weinberg, Brian Yanny, Naoki Yasuda, and SDSS Collaboration. The Sloan Digital Sky Survey: Technical Summary. *The Astronomical Journal*, 120(3):1579–1587, September 2000.

[30] Lorenzo Zanisi, Marc Huertas-Company, François Lanusse, Connor Bottrell, Annalisa Pillepich, Dylan Nelson, Vicente Rodriguez-Gomez, Francesco Shankar, Lars Hernquist, Avishai Dekel, Berta Margalef-Bentabol, Mark Vogelsberger, and Joel Primack. A deep learning approach to test the small-scale galaxy morphology and its relationship with star formation activity in hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 501(3):4359–4382, March 2021.