Mapping Galaxy Images across Ultraviolet, Visible, and Infrared Bands Using Generative Deep Learning

Youssef Zaazou[1] , Alex Bihlo[1] , and Terrence S. Tricco[2]
[1] Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada; yazaazou@mun.ca, abihlo@mun.ca
[2] Department of Computer Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada; tstricco@mun.ca

Abstract

We demonstrate that generative deep learning can translate galaxy observations across ultraviolet, visible, and infrared photometric bands. Leveraging mock observations from the Illustris simulations, we develop and validate a supervised image-to-image model capable of performing both band interpolation and extrapolation. The resulting trained models exhibit high fidelity in generating outputs, as verified by both general image comparison metrics (mean absolute error, SSIM, PSNR) and specialized astronomical metrics (Gini coefficient, M20). Moreover, we show that our models can be used to predict real-world observations, using data from the DECaLS survey as a case study. These findings highlight the potential of generative learning to augment astronomical data sets, enabling efficient exploration of multiband information in regions where observations are incomplete. This work opens new pathways for optimizing mission planning, guiding high-resolution follow-ups, and enhancing our understanding of galaxy morphology and evolution.

1. Introduction

Galaxies can appear quite different when observed across various photometric bands due to the wavelength-dependent nature of the light they emit (K. Kouroumpatzakis et al. 2023). Each photometric band captures a specific range of wavelengths, highlighting different physical properties of galaxies (S. Van den Bergh 1998). In shorter wavelength bands, like the ultraviolet (U) band, galaxies reveal the presence of hot, young stars and regions of active star formation. Meanwhile, in optical bands (e.g., G, R), the light primarily comes from older, cooler stars. At longer wavelengths, such as in infrared bands, the emission is dominated by dust and the cooler, more evolved stellar populations.

Studying galaxies across multiple photometric bands is essential because each band reveals different aspects of a galaxy's physical properties, offering a more complete picture of its structure, composition, and evolution. By combining observations from various bands, it is possible to trace a galaxy's star formation history, stellar populations, dust content, and even interactions with its environment. Observations made in multiple wavelengths require fewer inferences and lead to stronger conclusions being drawn about the observed galaxies. This makes multiband astronomy critical for developing a detailed understanding of a galaxy's life cycle and the processes driving its evolution. For further details, see S. Van den Bergh (1998) and K. Kouroumpatzakis et al. (2023).

1.1. Image-to-image Translation

Image-to-image translation refers to a class of computer vision modeling techniques that aims to transform an image from one domain into another while preserving essential visual content (Y. Pang et al. 2021). This process has gained popularity due to its applications in various fields, such as medical imaging (J. Long et al. 2015), artistic style transfer (L. A. Gatys et al. 2016), and even astronomical observations (V. Kinakh et al. 2024). The main challenge lies in learning the complex relationships between the source and target domains while maintaining consistency in features such as structure, texture, and color. Common approaches involve deep-learning techniques, particularly convolutional neural networks.

One of the most prominent methods in image-to-image translation is generative adversarial networks (GANs) introduced in I. Goodfellow et al. (2020). The pix2pix framework was later developed by P. Isola et al. (2017) as a conditional GAN-based approach and quickly became a foundation for many advancements in this area. It leverages paired data sets of input and ground truth images to learn the mapping between the two domains. However, in scenarios where paired data are not available, unpaired approaches like CycleGAN (J.-Y. Zhu et al. 2017) are employed. CycleGAN allows translation between two domains without paired images by introducing cycle consistency loss, which ensures that translating an image from one domain to another and then back again yields the original image.

More recently, however, J. Ho et al. (2020) introduced denoising diffusion probabilistic models (DDPMs), which have succeeded GAN models as the state of the art in image-to-image translation (e.g., P. Esser et al. 2024). C. Saharia et al. (2022) developed a DDPM-based image-to-image model that consistently outperforms GAN-based implementations.

1.2. Deep-learning Applications in Astronomy

Traditional data analysis methods in astronomy, particularly for studying galaxies, have been vastly outpaced by the enormous volume of data generated by modern telescopes and surveys, as outlined by Y. Zhang & Y. Zhao (2015). These traditional approaches often involved painstaking manual inspection and interpretation, making them time consuming and less efficient. An example of this is the Galaxy Zoo project

(see K. W. Willett et al. 2013) where crowdsourcing was used to assign morphological labels to images of galaxies. However, deep-learning algorithms have revolutionized this process by automating many tasks once performed manually. For example, deep learning has proven instrumental in classifying galaxies by identifying their various types and structures, as in J. De La Calleja & O. Fuentes (2004). This automation has significantly accelerated the analysis process and increased reliability, allowing astronomers to handle larger data sets and uncover new insights more effectively. For a comprehensive dive into machine learning and deep-learning applications in astronomy, see C. J. Fluke & C. Jacobs (2019).

Initially, discriminative learning was favored over generative learning because it was more straightforward and computationally feasible given the technology and resources available. Discriminative models, which focus on classifying data and predicting labels by learning the boundaries between different classes, were simpler to implement and required less computational power relative to generative models. Recently, however, generative learning has become popular in astronomy due to significant advancements in computational power and the development of sophisticated algorithms. Examples of generative learning in astronomy can be seen in A. Spindler et al. (2021) and M. J. Smith et al. (2022). However, these approaches are for unconditioned generative modeling where the output is not conditioned on any input. Image-to-image models are a subset of generative models that differ from their unconditioned counterparts by conditioning the output image on a particular input image.

Image-to-image modeling has seen some applications in astrophysics in recent years. Initially, image-to-image modeling was aimed at reconstructing noise signatures and denoising input images. Q. Lin et al. (2021) were the first attempt to apply image-to-image techniques to reconstruct noise signatures and perform translation between sky surveys. A. Vojtekova et al. (2020) proposed a U-Net-based approach for image denoising and enhancement. Similarly, T. Liu et al. (2025) presented a denoising method that differs from A. Vojtekova et al. (2020) by requiring no labeled data, instead relying on self-supervision (Y. Quan et al. 2020) to train their models. More recently, V. Kinakh et al. (2024) trained and compared several image-to-image models, including both GAN- and DDPM-based models, to translate Hubble Space Telescope data into James Webb Space Telescope imagery. However, their approach is for indiscriminate patches of sky and does not isolate any one class of astronomical phenomena (e.g., galaxies, stars, nebulae).

1.3. Our Research

In this paper, we explore the problem of translating across different photometric bands (M. S. Bessell 2005) for galaxy observations using a custom generative image-to-image machine learning model. The challenge of translating images of galaxies across photometric bands using image-to-image modeling remains relatively unexplored. This offers a promising area for further research and exploration. Despite the great successes of image-to-image machine learning models in computer vision—for example, CycleGAN (J.-Y. Zhu et al. 2017), pix2pix (P. Isola et al. 2017), and Palette (C. Saharia et al. 2022)—the adoption of such methods in the astronomy and astrophysics community is significantly lagging. This is despite an abundance of data that could be used to train and validate image-to-image machine learning models to perform various tasks.

A successful image-to-image machine learning model could be utilized for a variety of applications. One option is to optimize observational mission planning. Existing data could be leveraged to simulate or predict what unexplored wavelengths or regions of the sky might look like, allowing prioritization of areas of interest that are likely to yield important discoveries. This could assist efficient allocation of telescope time, but could also help in the planning stages for the next generation of telescopes. Another potential application is to help guide high-resolution follow-up observations. The idea here is that translated images could predict structures or regions of interest that are unresolved in low-resolution data. These regions could be prioritized for further high-resolution studies, saving observational time and resources.

We present a fully supervised image-to-image machine learning model that is capable of predicting output bands using a variety of input bands. We also present two types of band translation: the first is band interpolation, where an output band is predicted based on two delineating input bands, and the second is band extrapolation, where a sequence of bands is extended. We demonstrate that the same model architecture can be used to perform either type of translation, requiring no modifications when switching between approaches or when changing the target output band. To clarify, each set of inputs and corresponding target band outputs requires retraining the model, but it is the model architecture and training procedure that is identical for all input and output band combinations.

The majority of our results are obtained using mock observations from Illustris simulations (M. Vogelsberger et al. 2014; P. Torrey et al. 2015). Illustris offers the advantage of having a wide range of bands to map between without having to account for instrumental differences, as would be the case using inputs from different surveys or instruments. To confirm the applicability of our model on real-world data, we additionally present a mapping using real observations from the DESI Legacy Imaging Surveys (DECaLS; see, e.g., A. Dey et al. 2019).

All the models' outputs were investigated on two grounds: qualitative and quantitative. Through qualitative visual inspection, we confirm that the generated images are nearly identical to their respective ground truth images in structure. Several quantitative metrics are used to confirm our qualitative findings. General image comparison metrics are used, such as the structural similarity index (SSIM), as seen in D. Brunet et al. (2011), and the peak signal-to-noise ratio (PSNR), both of which function as measures of proximity between the generated and ground truth images. Additionally, we use galaxy morphology indicators, specifically Gini and M20 measurements, as presented in J. M. Lotz et al. (2004), to confirm that the generated images do indeed match the ground truth images from a galaxy morphology perspective.

To ensure reproducibility and foster further research, the code developed for this study is publicly available in (Y. Zaazou et al. 2025).

2. Proposed Approach

2.1. Data Set

Two distinct data sets are used throughout this work. The first data set is a collection of mock observations obtained from Illustris simulations. Most of our testing and prototyping was conducted using Illustris mock observations. The second data set is a collection of real observations of galaxies taken

Table 1
Summary of Photometric Bands Used

Band	FUV	NUV	U	G	R	Z	K
Corresponding telescope	GALEX	GALEX	SDSS	SDSS	SDSS	SDSS	2MASS
Effective wavelength (nm)	151.3	230	357.3	472.4	620.1	891.7	2162.0

from the Galaxy10 DECaLS data set courtesy of the Python package astroNN by H. W. Leung & J. Bovy (2018). This data set provides a proof of concept to demonstrate that our models can be trained on and make inferences from real data.

2.1.1. Illustris Data Set

The Illustris mock observation catalog (P. Torrey et al. 2015) is a synthetic data set generated from Illustris cosmological simulations (M. Vogelsberger et al. 2014), which model the formation and evolution of galaxies over cosmic time. The Illustris project simulates a wide range of physical processes that govern galaxy formation, including dark matter dynamics, gas cooling, star formation, supernova feedback, and black hole accretion. These simulations capture the complex interplay between baryonic matter and dark matter, allowing for the detailed modeling of galaxy evolution across cosmic time. Additionally, the simulations incorporate radiative transfer, enabling a more accurate representation of how light interacts with gas, dust, and stars within galaxies.

There are many unique advantages and disadvantages of using mock observations as opposed to real observations. First, the images are ideal in the sense that they contain no instrumental noise or artifacts. Additionally, Illustris mock observations are not limited by factors such as atmospheric inference and survey depth. The most significant contribution that Illustris offers is its wide range of bands, spanning all the way from the far ultraviolet to the mid-infrared. This is significantly wider than any single survey can offer and allows us to test various band combinations without having to switch between different surveys and account for cross-survey differences such as differing noise signatures, resolutions, or misalignment issues.

Our data set comprises paired galaxy images from the Illustris mock observation catalog where pairs are grouped by individual galaxies. Images are scaled using asinh scaling (R. H. Lupton et al. 1999), followed by normalization to a range of $[0, 1]$, as is standard in machine learning (see P. J. M. Ali & R. H. Faraj 2014). We used the far-ultraviolet (FUV), near-ultraviolet (NUV), U, G, R, Z, and K bands in various combinations. These bands correspond to bands from the Galaxy Evolution Explorer survey (GALEX; D. C. Martin et al. 2005), the Sloan Digital Sky Survey (SDSS; D. G. York et al. 2000), and the Two Micron All Sky Survey (2MASS; M. Skrutskie et al. 2006). See Table 1 for a breakdown of the photometric bands used in this work.

2.1.2. Galaxy10 DECaLS Data Set

The Galaxy10 DECaLS data set, as obtained through the astroNN Python package from H. W. Leung & J. Bovy (2018), contains over 17,000 colored galaxy images available in G, R, and Z bands. The observations are collected from DECaLS (A. Dey et al. 2019). The images are normalized to scale the image pixel values from $[0, 255]$ to $[0, 1]$, but

otherwise the images are not altered or preprocessed in any other way.

2.2. Metrics

Various metrics are used to assess the quality of the output images compared to the target images. Some of these metrics are general image comparison metrics in the sense that they can be applied to any set of input images. Others are specialized metrics specifically designed for analyzing galaxy morphology. We use the following set of metrics:

1. *Mean absolute error (MAE)*. The MAE between output images and target images. This averages the absolute per pixel differences.
2. *SSIM*. The SSIM (D. Brunet et al. 2011) measures the distance between a pair of input images. Unlike the pointwise MAE, the SSIM can account for local structural differences. The SSIM has a range of $[-1.0, 1.0]$; values closer to 1.0 indicate high agreement between the images, 0.0 indicates no relation, and values closer to -1.0 indicate disagreement between the inputs.
3. *PSNR*. The PSNR is a widely used metric for assessing the similarity between two images, particularly in the context of image compression and reconstruction. It measures the ratio between the maximum possible pixel value (signal) and the power of the noise (difference) between the original and the reconstructed image. PSNR is expressed in decibels, with higher values indicating greater similarity, as they imply lower error between the images. A PSNR value around 30–50 dB is generally considered to imply that the reconstructed image is of high fidelity, with the reconstructed image almost indistinguishable from the original to the human eye for PSNR around 40 dB.
4. *Gini and M20*. Gini and M20 measurements, first presented in J. M. Lotz et al. (2004), are used to characterize the morphological features of galaxies. The Gini coefficient is a statistical measure originally used to quantify income inequality, but in galaxy morphology, it quantifies the distribution of light among a galaxy's pixels. The M20, on the other hand, is a measure of the relative contribution of the brightest 20% of a galaxy's light to its overall light distribution. Specifically, it quantifies the spatial distribution of the brightest regions within a galaxy, helping to identify whether the galaxy has a single dominant core or multiple bright regions, which could indicate complex internal structures or recent mergers.

   When used together, Gini and M20 provide a powerful tool for differentiating between various morphological types, especially in galaxies with irregular or disturbed structures. We will be comparing Gini and M20 measurements of output and target images to confirm that the outputs are free of spurious effects and reliably resemble natural images. Larger agreement between the distributions for ground truth label values and the

generated image values indicates higher model performance. To measure the distance between these distributions, we will use the first Wasserstein distance ($\mathcal{W}_1$), also known as the earth mover's distance (Y. Rubner et al. 2000). To ensure the interpretability of the $\mathcal{W}_1$ measurements, we scaled the Gini coefficient and M20 measurements to a range of [0.0, 1.0]. However, the maximum and minimum values used for rescaling were selected globally to preserve the relative scale of the distributions. Due to this normalization, the $\mathcal{W}_1$ distance for our Gini and M20 distributions has values of 0.0 for identical distributions and 1.0 for maximally disjointed distributions.

2.3. Inference Techniques

Most sky surveys rarely capture information in one band. For example, in conducting translations from 2MASS's K-band observations (M. Skrutskie et al. 2006), information from 2MASS's H-band observations could be leveraged with minimal additional extraction and preprocessing overheads. To fully exploit all the information available from any given survey, we introduce band extrapolation. With extrapolation, a sequence of bands is extended to infer bands further along the sequence. For instance, we could select bands G and R as the input and train a model to extend this sequence in either direction by inferring lower-frequency bands (such as Z and K), or higher-frequency bands (such as U or NUV).

We also introduce band interpolation, where instead of extending a sequence of observations, intermediate observations are generated at specific bands delineated by both input bands. For example, for inputs such as U and K, we could generate G, R, and Z observations. To clarify, each output band requires a new instance of the model. This leads to the training of multiple models to any combinations of input bands to different output bands. This applies to both the interpolation and extrapolation approaches.

There are inherent pros and cons for each approach. The extrapolation models are at a disadvantage compared to the interpolation models, as they must extend in wavelength away from the input bands. Interpolation is able to use information from both high- and low-wavelength bands, by comparison. However, band interpolation will often require that the data are gathered from multiple surveys, making it logistically more challenging.

2.4. Generator Training

The inputs of the model are denoted as X_{ph}, where ph gives the photometric band such that X_{U} and X_{K} denote inputs from the U and K photometric bands, respectively. The output of the model will be denoted by $\hat{Y}$, where Y is to be substituted with the target output band. In this manner, a model accepting inputs X_{NUV} and X_{K} with outputs $\hat{R}$ is a model trained to interpolate the R band from corresponding NUV and K inputs. In the previous example, the ground truth labels would be denoted as R, and this notation will carry forward in subsequent sections.

We elected to employ a fully supervised training process to train our generator models. Our models can be considered a complex nonlinear transformation with the model weights controlling the transformation applied to the input(s). Training our models consists of finding the model weights that minimize the distance between the output and the ground truth label. This distance is referred to as the loss function.

Our loss function is the combination of the $\mathcal{L}_1$ (mean absolute) error and the inverse SSIM loss, given by

$$\mathcal{L}_{\mathrm{SSIM}} = 1 - \mathrm{SSIM}(Y, \hat{Y}), \qquad (1)$$

where $\mathrm{SSIM}(Y, \hat{Y})$ corresponds to the SSIM between the ground truth and reconstructed images. This loss function was determined through a series of trial and error. The full objective to be minimized during training is

$$\mathcal{L} = \mathcal{L}_1 + \lambda \mathcal{L}_{\mathrm{SSIM}}, \qquad (2)$$

where λ is a regularization parameter greater than 0. We experimented with setting $\lambda = 0$, but found that setting $\lambda = 1$ yielded higher-quality results. In setting $\lambda = 1$, we found an improvement across all mentioned image quality metrics (see Section 2.2), compared to $\lambda = 0$. We were also able to visually detect a subtle but consistent increase in the correspondence between the output and target images using $\lambda = 1$. We experimented with discarding the $\mathcal{L}_1$ loss component altogether; however, our experimentation shows that the resulting images were noisier compared to a loss that incorporated the $\mathcal{L}_1$ loss. Future work may involve investigating the effects of λ values greater than 1.

2.5. Generator Architecture

We employ a ResNet-like architecture (K. He et al. 2016), primarily composed of residual blocks. This architecture emphasizes residual learning, making it easier for the model to learn identity mappings if needed. The generator itself follows the architecture of the generator networks in the CycleGAN implementation (J.-Y. Zhu et al. 2017). However, after performing hyperparameter tuning, we found that the default hyperparameters in CycleGAN did not perform best compared to other configurations. In summary, we decreased the number of downsampling and upsampling blocks to increase the resolution of the latent residual layers along with increasing the number of residual blocks from six to nine. For more details on model selection, hyperparameter tuning, and training, see the Appendix.

Future work may see the use of a diffusion-based architecture, as in C. Saharia et al. (2022). The probabilistic nature of a diffusion model's image generation process, which may yield several outputs for a single input, may be used to quantify the model's uncertainty. However, note that the uncertainty of a diffusion model comes from the stochasticity inherent to the model and does not correspond to the uncertainty of the physical system.

3. Results

3.1. Interpolation

We now investigate the performance of three models tasked with band interpolation based on the same input. The input bands are the NUV and K bands, which correspond, in wavelength, to GALEX's NUV and 2MASS's K bands, respectively. We note significant differences between both input bands since we have selected the bands to be quite far apart to provide the models with the opportunity to showcase their ability to handle nontrivial transformations.

We trained three models to map to G, R, and Z bands corresponding in wavelength to SDSS bands. The reason for selecting these bands is to investigate the theoretical

augmentation of SDSS's library using GALEX and 2MASS observations. A paired set of observations taken from both GALEX and 2MASS could be used to augment the SDSS database significantly. The task of obtaining and preprocessing such a paired data set is not one we have investigated, but it is theoretically possible and would potentially produce high-quality images to augment SDSS observations with minimal inference cost.

The inputs and the interpolated results can be seen in Figure 1. The generated outputs are nearly identical to the ground truth data. This is due to the model being able to take advantage of the information present in both bands. The model can detect areas of high star formation from the NUV band while also having access to the distribution of the majority of the stellar matter through the K band. Leveraging information from both bands allows it to make predictions that are more precise than is possible with only one band's input.

Table 2 shows performance metrics for our three interpolation models. All the shown metrics indicate that the models consistently produce outputs that are in line with the ground truth data. The MAE is less than 1%. SSIM scores of ∼0.99 indicate that the generated and ground truth images are virtually identical, indicating excellent similarity in image quality assessment. For perspective, the SSIM between the NUV input band and the ground truth labels of bands G, $R,$ and Z are 0.53, 0.45, and 0.43, respectively, while the SSIM between the K input band and ground truth bands G, R, and Z are 0.86, 0.82, and 0.85, respectively.

The mean PSNR measurements are ∼35 dB, which generally implies that the generated images are of high quality. Departures from the ground truth data are small and only noticeable upon close inspection.

Figure 2 shows the distributions of the Gini and M20 values for the input and generated data. The distributions for the input bands are represented, along with the distributions of the ground truth and generated images in the target band. The shape of the distributions between the generated and ground truth images for the target band are in excellent agreement. This highlights that the model is capturing the breadth of diversity within galaxies for the target band, avoiding issues such as mode collapse. Table 2 quantifies the Gini and M20 measurements by comparing the Wasserstein distance between the two distributions. Relatively low $\mathcal{W}_1$ values are observed, indicating a significant agreement between the ground truth data and generated images, as far as galaxy morphology is concerned.

3.2. Extrapolation

This section investigates the performance of three models tasked with band extrapolation based on similar inputs. The input bands are the G and R bands, which correspond, in wavelength, to SDSS's G and R bands. The two input bands are proximal in wavelength but still provide the model with sufficient information as to the progression of band observations. Further experimentation is needed to investigate the effect of the input bands' proximity. A different band combination (such as G and Z) may provide the model with better information.

We trained three models to map to U, NUV, and FUV bands corresponding in wavelength to the SDSS U band and the GALEX NUV and FUV bands, respectively. All three output bands are lower in wavelength than the input bands, which

presents the models with more challenging transformations. Lower-wavelength inferences are more demanding than their higher-wavelength counterparts because of the inherent complexity in regions containing high star formation rates. Compared to band interpolation, this task is, practically, more straightforward since it is highly probable that all of the input data can be obtained from a single survey.

The inputs and the extrapolated results can be seen in Figure 3. As with the band interpolation, the generated images are nearly identical to the ground truth data. The model is able to recreate the structure present in the ground truth data with a very high degree of accuracy. Of particular note is the models' ability to preserve their accuracy across the sequence even as the output band becomes further removed from the input bands, such as in the FUV band inferences, though it can be noticed that the magnitude of residuals increases as the extrapolation distance increases.

The performance metrics in Table 3 indicate that the models consistently produce output that is in line with the ground truth data across multiple extrapolated bands. The MAE scores are ∼1% for the three bands, with the error increasing the further the extrapolated band is from the source bands.

The U and NUV models have SSIM scores of 0.99 and 0.95, respectively. In general, SSIM scores >0.95 indicate that two images are virtually identical, meaning that the images generated by our models are exceptionally representative of the ground truth images. While the FUV model's SSIM score of 0.91 indicates slightly lower agreement, this is still a perfectly acceptable result, particularly given how distant the FUV band is from the input. For perspective, the SSIM between the G input band and the ground truth labels of bands U, NUV, and FUV are 0.84, 0.52, and 0.49, respectively, while the SSIM between the R input band and ground truth bands U, NUV, and FUV are 0.73, 0.44, and 0.41, respectively. The generated images are indeed high-quality transformations of the input bands.

A similar pattern is noticeable along the PSNR measurements of the three models. The U model's PSNR of above 40 dB indicates nearly perfect reconstruction of the ground truth data. The FUV and NUV models average around 30 dB, which indicates visible, but minor, differences between the generated and ground truth images.

Figure 4 shows the distributions of the Gini and M20 values for our data, along with the input bands' distributions. The models do an excellent job of reproducing the distributions for the target band. Of note are the M20 values for the FUV band, as the input bands are bimodal and highly overlapped, whereas the FUV target band is unimodal, but non-Gaussian, with little overlap to the input band distributions. The Gini and M20 distributions for the generated images match the ground truth distributions almost perfectly across all three extrapolated bands. Table 3 shows the $\mathcal{W}_1$ distances between the generated and ground truth distributions, indicating high agreement, reflected by the relatively low $\mathcal{W}_1$ values. This provides strong evidence that the models produce representative galaxy morphologies across the data set.

3.3. DECaLS Observations

We now train our model architecture on real observational data from DECaLS (see the data description in Section 2.1.2). A single model is trained to interpolate the R band based on input G and Z bands, respectively. Training on DECaLS data is
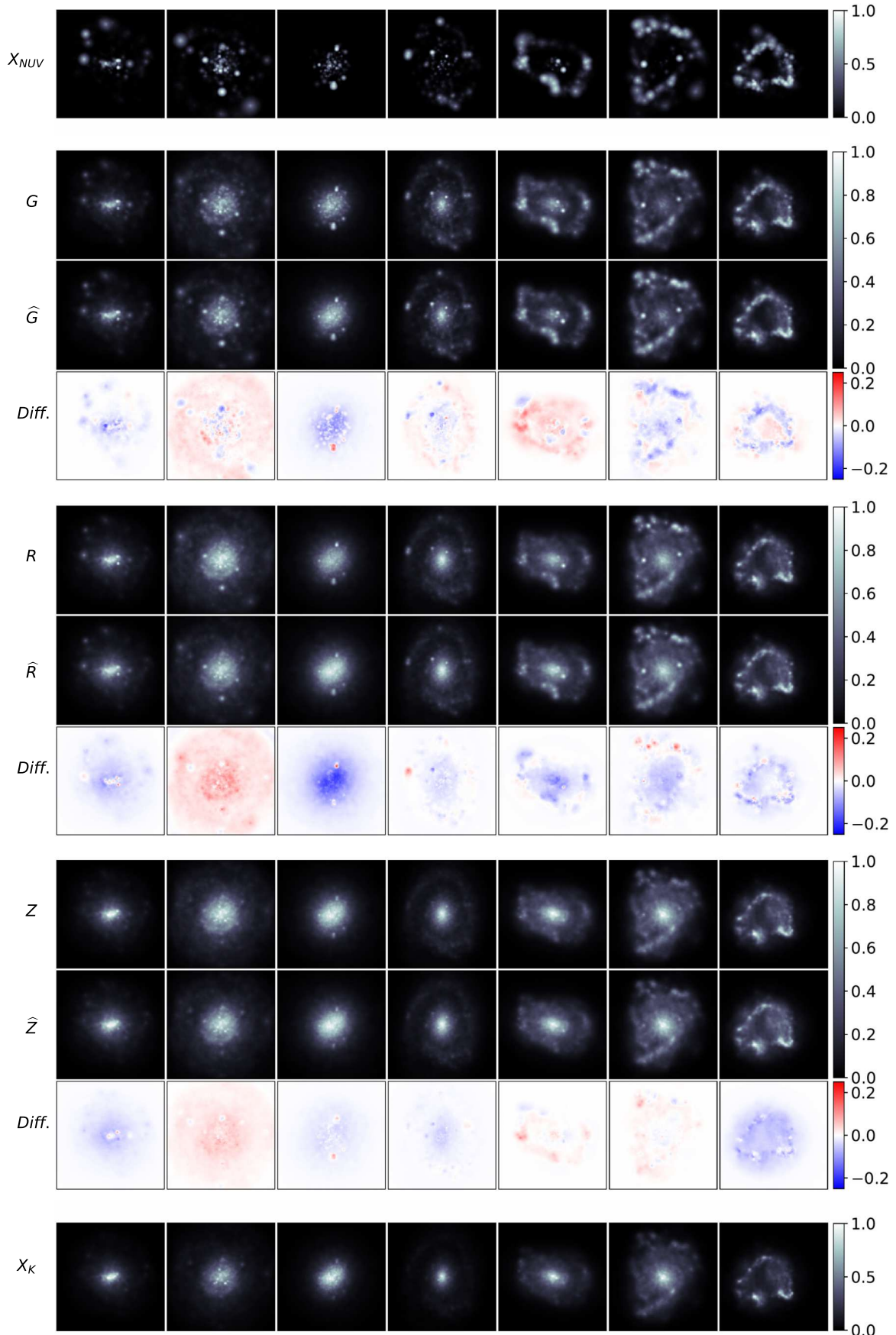
Figure 1. The results for our three trained interpolation models; each column is reserved for a single galaxy. The NUV inputs and the K inputs (X_{NUV} and X_K) are delineated by the G, R, and Z groupings (one for each model). The first row of each grouping shows the ground truth labels, the second row shows the respective generated images, and the last row is the residual taken to be the ground truth label minus the model output.
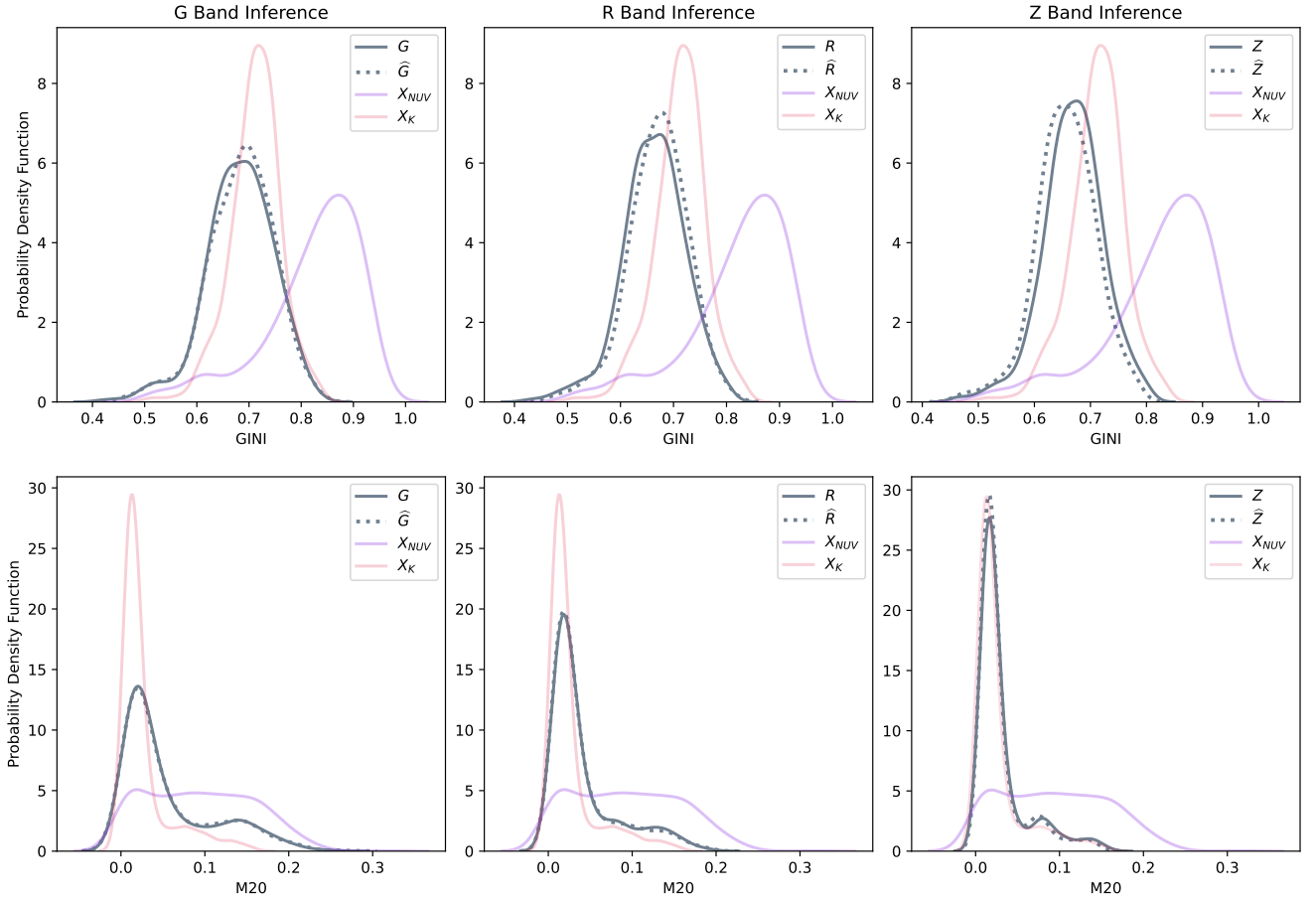
Figure 2. Distributions of Gini and M20 values for each of the three interpolation models alongside the input data's distributions for reference. We note significant agreement in the ground truth and generated distributions, particularly for the M20 measurement. While the Gini coefficient distributions are not perfectly aligned, they still indicate a successful transformation, especially when taken in the context of the initial distributions of the inputs.

Table 2
Performance Metrics for the Three Interpolation Models

Inferred Band (λ_{eff})	MAE ↓	SSIM ↑	PSNR ↑	Gini : $\mathcal{W}_1$↓	M20 : $\mathcal{W}_1$↓
G (472 nm)	0.0078	0.984	35.3	0.0061	0.0064
R (620 nm)	0.0086	0.986	34.3	0.0206	0.0060
Z (891 nm)	0.0066	0.990	37.4	0.038	0.0098

Note. Arrows indicate the desirable direction for each metric: ↑ indicates higher is better; ↓ indicates lower is better.

highly significant, as it represents our model's capacity to learn and generate from real data, rather than mock observational data from the Illustris simulations. While the wavelength difference between the input bands and output band is smaller than the results of the models presented in Section 3.1, this concession is necessary to gauge how noise and various background artifacts in real observational data can complicate the training process.

Figure 5 shows the input images (G and Z bands), the ground truth and generated R-band images, and the residuals between the generated and ground truth images. A significant level of agreement between the ground truth data and the generated images is evident. Of note is that the signatures of the central bulge are much more similar to the ground truth images compared to either of the inputs. Similarly, the outline of the spiral arms is in line with the ground truth images. The

residuals contain mostly noise and are free from any significant signs of structure. The model is recreating the galaxy structure present in the ground truth data.

Table 4 shows our performance metrics. The MAE is ∼2%, which is higher than the Illustris-trained models. This is evident from the residuals in Figure 5, which show a general level of background noise that is not present to such a degree in the clean Illustris data. The MAE metric will average this background noise into a nonzero value, even though this noise is largely random.

The SSIM score of 0.86 is not as strong an agreement compared to the results of Section 3.1. However, the majority of the discrepancy is between the differing noise signatures and backgrounds of the generated and ground truth images. For perspective, the SSIM between the R labels and the G and Z inputs is 0.53 and 0.38, respectively. An increase to SSIM values of 0.86 from either input certainly indicates that the model has learned a correct transformation to the target band.

The PSNR score of ∼30 is similar to the results found when interpolating Illustris mock observational data, as given in Table 2. This implies that the generated images are representative of the ground truth data.
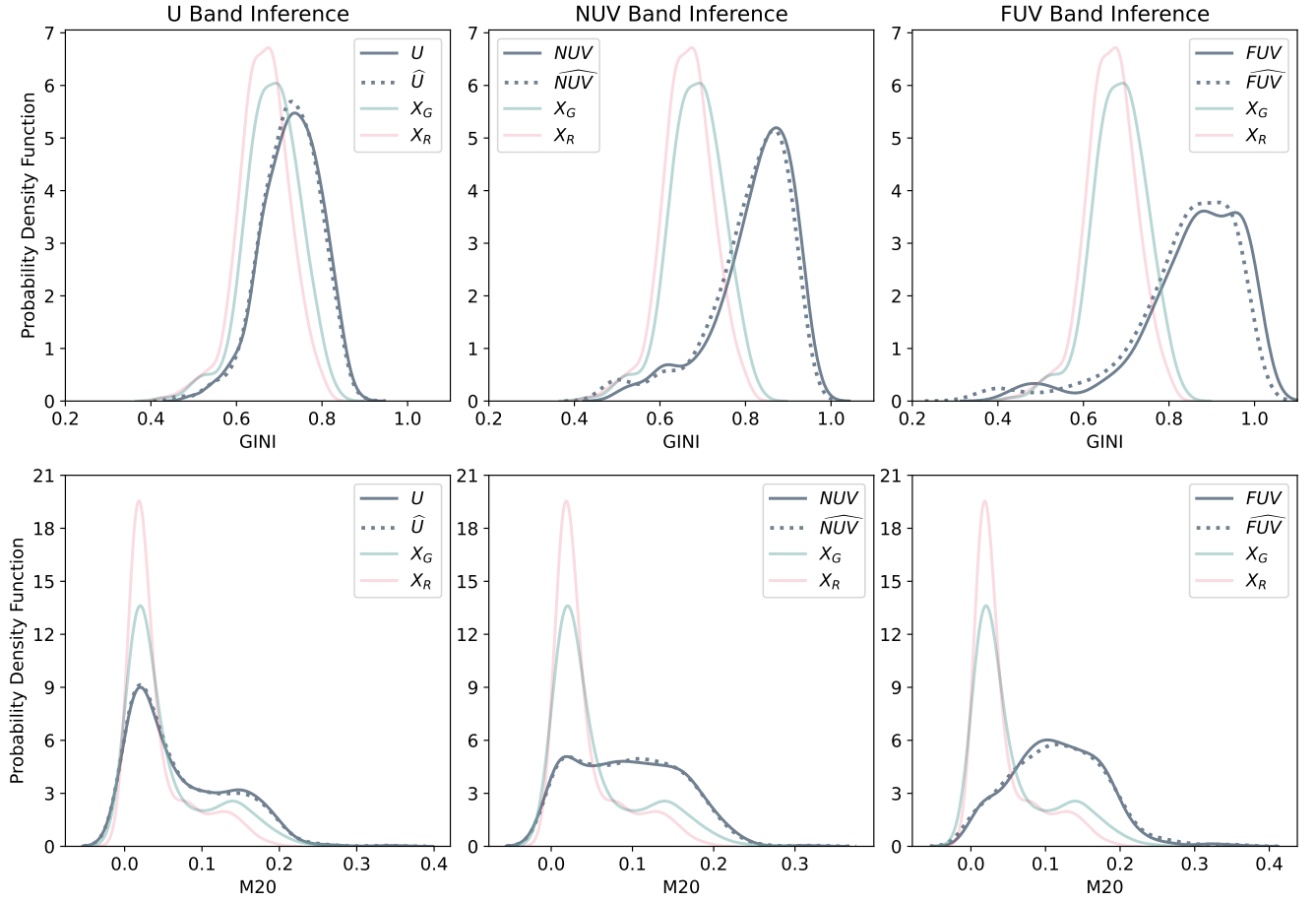
Figure 6 shows the distributions of the Gini and M20 values for the input bands (G and Z), and the generated and ground truth bands (R). The model is able to reproduce the target distribution, though with some minor differences in the Gini distribution. As given in Table 2, the $\mathcal{W}_1$ distances between the distributions are found to be quite small, with similar
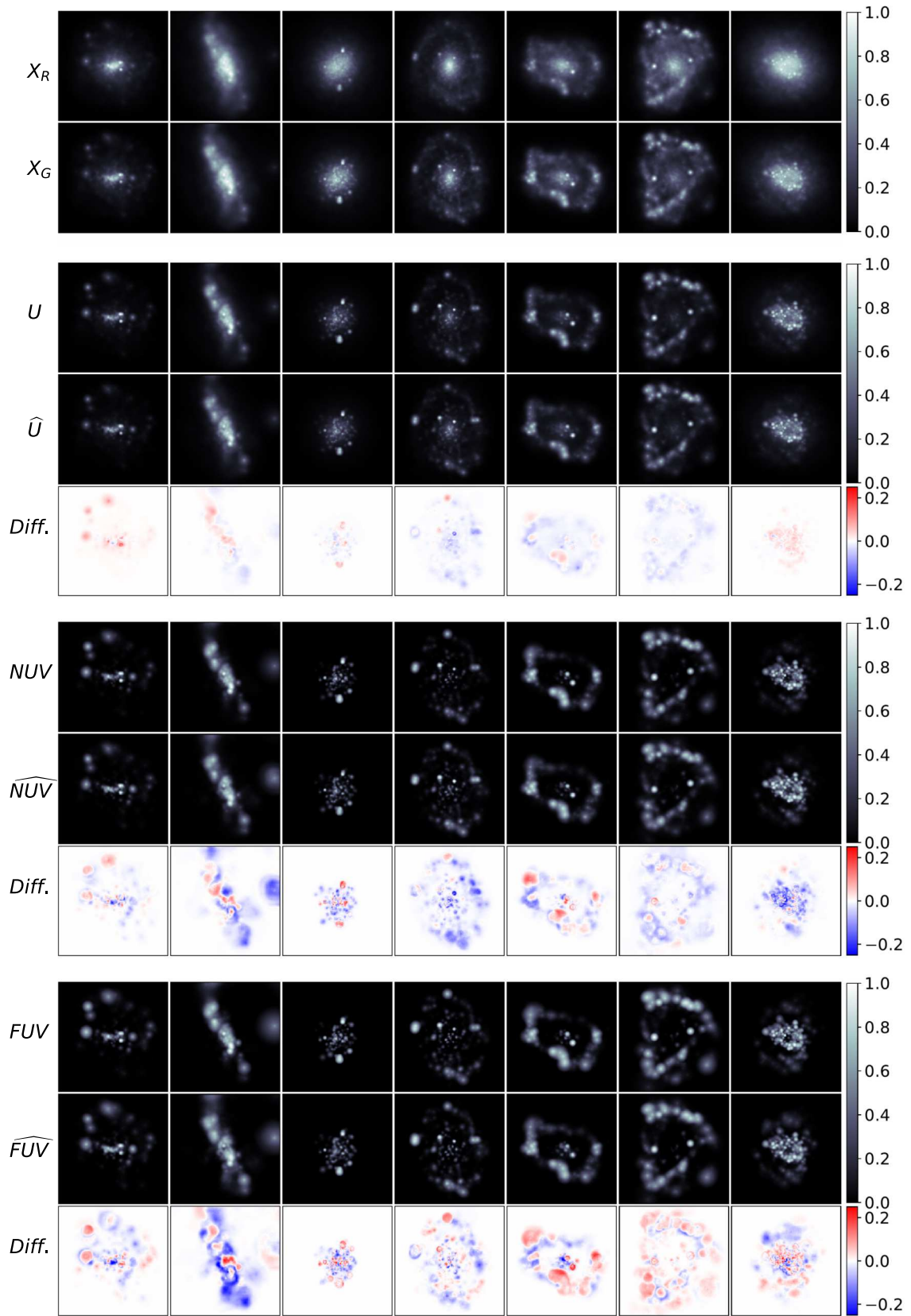
Figure 3. The results for our three trained extrapolation models; each column is reserved for a single galaxy. The R inputs and the G inputs (X_R and X_G) precede the U, NUV, and FUV groupings (one for each model). The first row of each grouping shows the ground truth labels, the second row shows the respective generated images, and the last row is the residual taken to be the ground truth label minus the model output.

Figure 4. Distributions of the Gini and M20 values for each of the three extrapolation models alongside the input data's distributions for reference. We note significant agreement in the ground truth and generated distributions, particularly for the M20 measurement. While the Gini coefficient distributions are not perfectly aligned, they still indicate a successful transformation, especially when taken in the context of the initial distributions of the inputs.

Table 3
Performance Metrics for the Three Extrapolation Models

Inferred Band ($\lambda_{\rm eff}$)	MAE ↓	SSIM ↑	PSNR ↑	Gini : $\mathcal{W}_1$↓	M20 : $\mathcal{W}_1$↓
U (357 nm)	0.0034	0.99	40.3	0.0086	0.0079
NUV (230 nm)	0.0089	0.96	32.3	0.035	0.0081
FUV (151 nm)	0.0127	0.91	28.4	0.049	0.014

Note. Arrows indicate the desirable direction for each metric: ↑ indicates higher is better; ↓ indicates lower is better.

magnitude to the $\mathcal{W}_1$ distances for the band interpolation models trained on Illustris mock observational data. As with the other metrics explored, this indicates high agreement between ground truth data and generated images in terms of galaxy morphology.

A further exploration of the residuals between the model outputs and ground truth labels is shown in Figure 7. The histogram, computed over the entire test data set, reveals a highly symmetric distribution centered around zero, indicating that the model does not exhibit a systematic bias in over- or underpredicting pixel values. Moreover, the histogram reveals that the vast majority of residual pixel values lie within the range $[-0.1, 0.1]$. To quantify this observation, we computed the absolute values of the residuals and found the 99th percentile to be 0.091, indicating that over 99% of all residuals have a magnitude less than 0.1.

3.4. Data Set Augmentation

To test the robustness of our model to variations in the input data, we trained a version of the model on an augmented data set. We selected the NUV extrapolation model as a reference and trained an identical architecture to the one described in Section 3.2, with the only difference being the use of data augmentation. The augmentations included random horizontal and vertical flips, small-angle rotations, and the addition of Gaussian noise. These transformations were applied on the fly using data augmentation layers placed at the top of the model architecture, ensuring that each input image was perturbed differently at every iteration. Figure 8 shows an example galaxy image alongside several randomly augmented variants.

The goal of this experiment was to evaluate whether the model's performance would be sensitive to such low-level perturbations. As can be seen in Table 5, the resulting model exhibited nearly identical behavior to the version trained on the original data set. This suggests that the model's architecture and training paradigm are inherently robust to these common image-space transformations. It also supports the notion that the features learned by the network are more likely tied to global morphology and structural cues rather than superficial pixel-level patterns.

3.5. Uncertainty Quantification

To estimate the variability and uncertainty inherent in our approach, we trained the same model eight separate times using
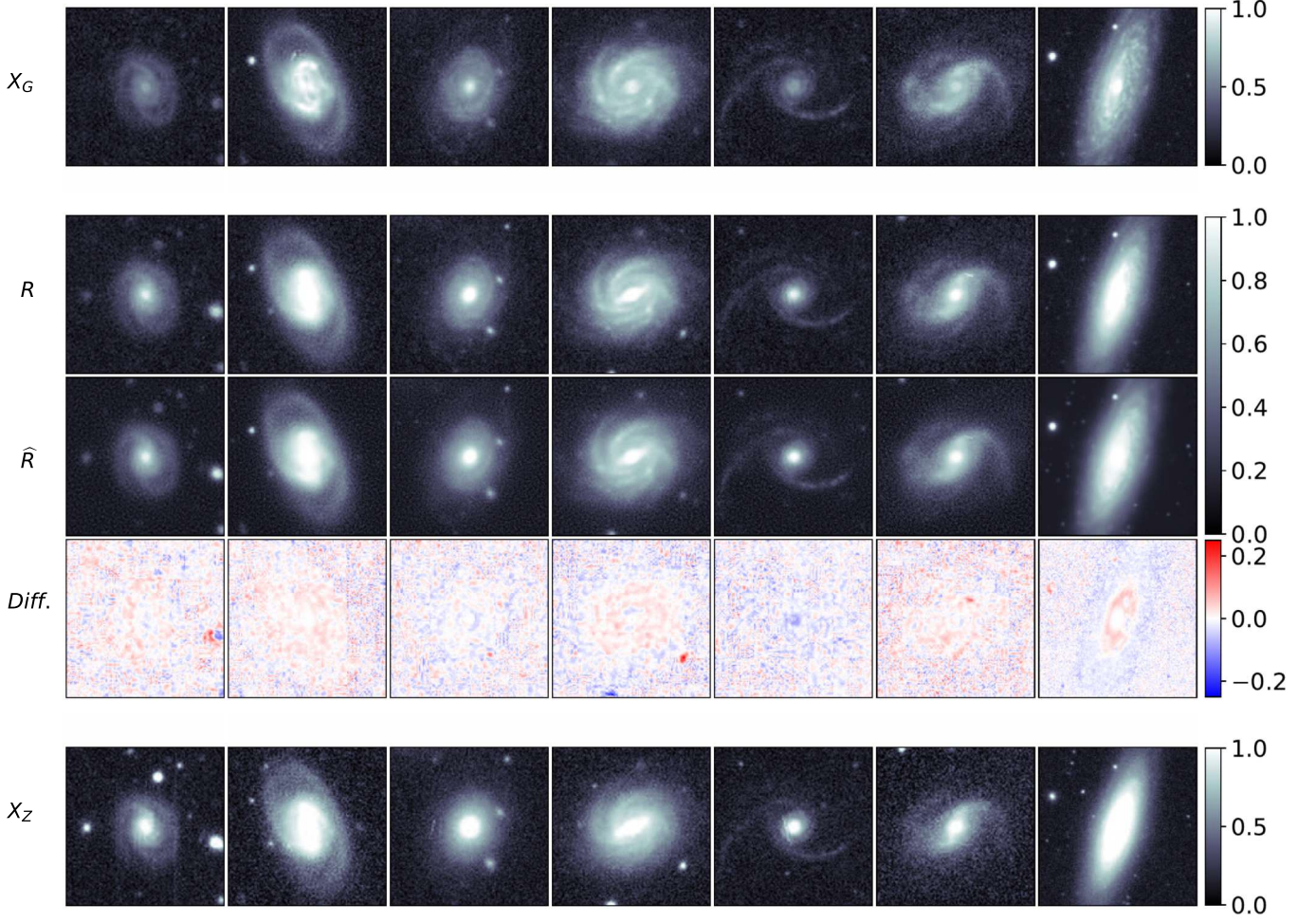
Figure 5. The results for our DECaLS model; each column is reserved for a single galaxy. The G and Z inputs (X_G and X_Z) are delineated by the R grouping, the first row of which shows the ground truth labels, the second shows the respective generated images, and the last row is the residual taken to be the ground truth label minus the model output.

Table 4
Performance Metrics for the DECaLS-trained Interpolation Model

| Inferred Band (λ_{eff})	MAE ↓	SSIM ↑	PSNR ↑	Gini : $\mathcal{W}_1$↓	M20 : $\mathcal{W}_1$↓
R (620 nm)	0.022	0.86	30.5	0.034	0.0012

Note. Arrows indicate the desirable direction for each metric: ↑ indicates higher is better; ↓ indicates lower is better.

identical data sets and hyperparameters—changing only the random seed used for weight initialization. As in Section 3.4, we selected the NUV extrapolation model as an example for the uncertainty quantification used in this section. Essentially, the NUV extrapolation model from Section 3.2 was run eight different times with different random initializations.

Figure 9 shows the standard deviation of the resulting predictions made by our models alongside ground truth labels. These pixel-wise standard deviation calculations show us the areas in the output with the most model uncertainty. We observe the most model uncertainty around areas of most active star formation. Note the donut-like structure in the uncertainty; this indicates that the location of the structure is well resolved but the size and luminosity vary somewhat between different models.

In addition, Table 6 presents the same quantitative metrics reported in Sections 3.1 through 3.3, along with aggregate statistics across the eight-model ensemble. Specifically, each value in the table represents the mean of a given metric computed across the eight models, with the accompanying standard deviation reflecting the variability across those models. For each model, the metric itself is calculated as the mean over the 500 samples in the test data set. Notably, the relatively low variability observed across all quantitative metrics suggests that the model's performance is stable and consistent across different training runs. This consistency not only reinforces the robustness of the model but also validates the reliability of the chosen metrics as meaningful indicators of performance.

While quantifying model uncertainty is a useful diagnostic tool, it is important to note that the variability captured here reflects epistemic uncertainty—that is, uncertainty stemming from the model itself, such as sensitivity to initialization or stochastic training dynamics. This does not necessarily correspond to true physical uncertainty in the underlying astrophysical processes. However, we observe that the model uncertainty is concentrated in regions where there is a strong signal and that this uncertainty is roughly proportional to the strength of that signal—patterns that are physically intuitive.
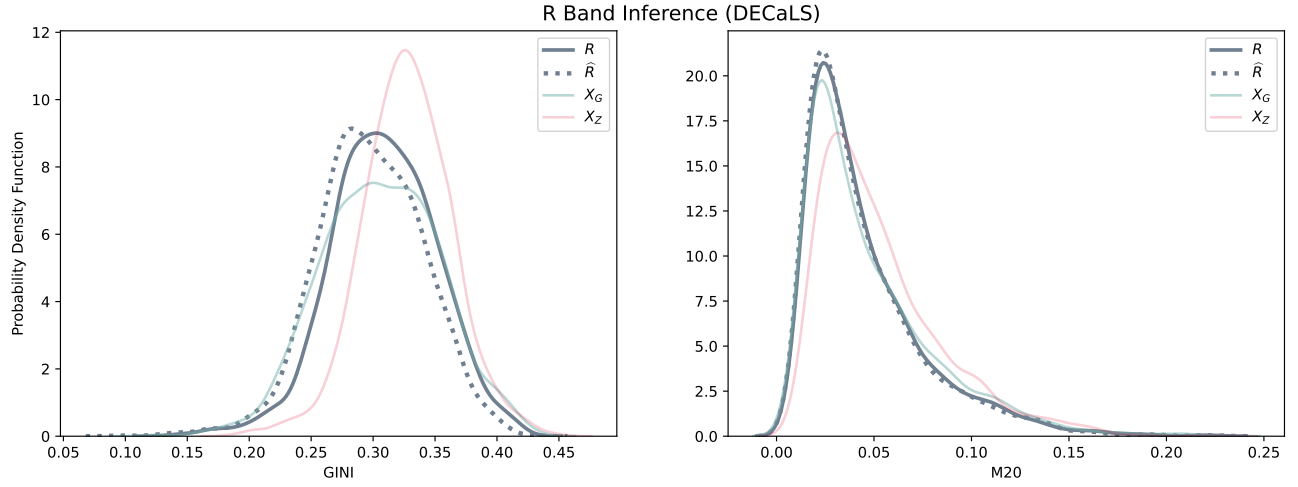
Figure 6. Distributions of the Gini and M20 values for the DECaLS interpolation model, with the input distributions for reference.
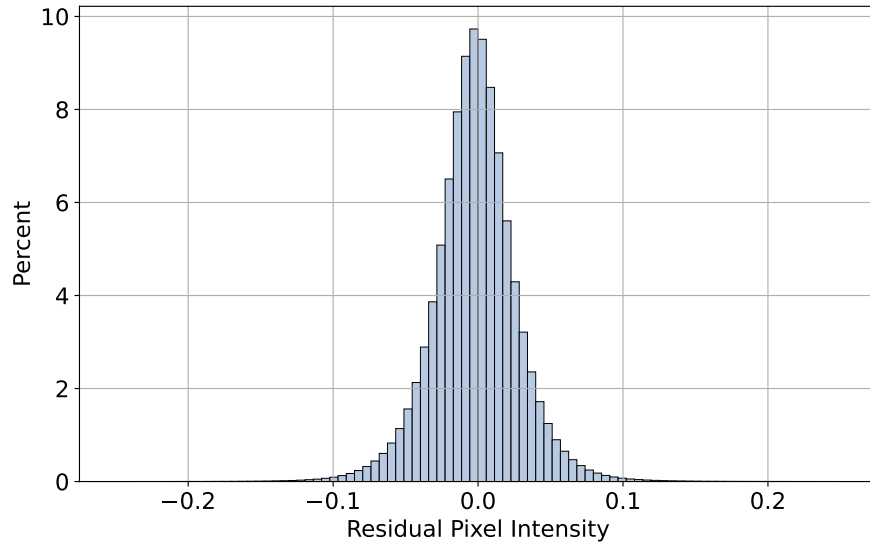
Figure 7. Histogram of residual (pixel) values aggregated over the entire test set. The horizontal axis delineates bins of residual values, while the vertical axis indicates the percentage of pixels falling into each bin. The distribution is approximately symmetric about zero, with most residuals lying within the range $[-0.1, 0.1]$.
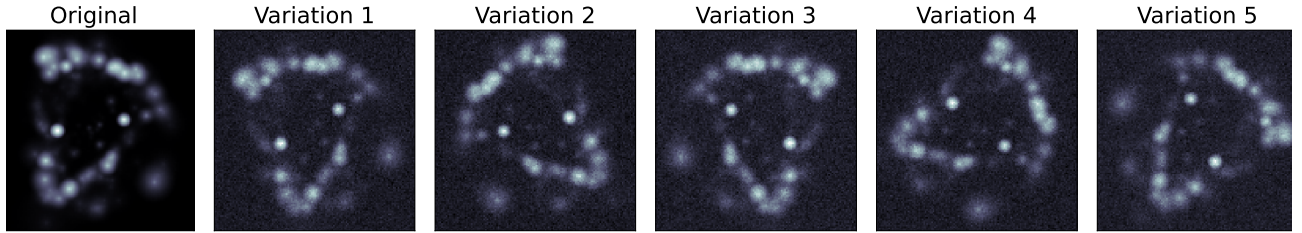

Figure 8. The effect of our data augmentation layers. The original image can be seen on the far left along with five variations alongside it.

Table 5
Performance Metrics for the NUV Extrapolation Model Comparing Original and Augmented Data Sets

Training Data Set	MAE ↓	SSIM ↑	PSNR ↑	Gini $: \mathcal{W}_1 \downarrow$	M20 $: \mathcal{W}_1 \downarrow$
Original	0.0089	0.96	32.3	0.035	0.0081
Augmented	0.0088	0.948	33.5	0.041	0.0078

Note. Arrows indicate the desirable direction for each metric: ↑ indicates higher is better; ↓ indicates lower is better.
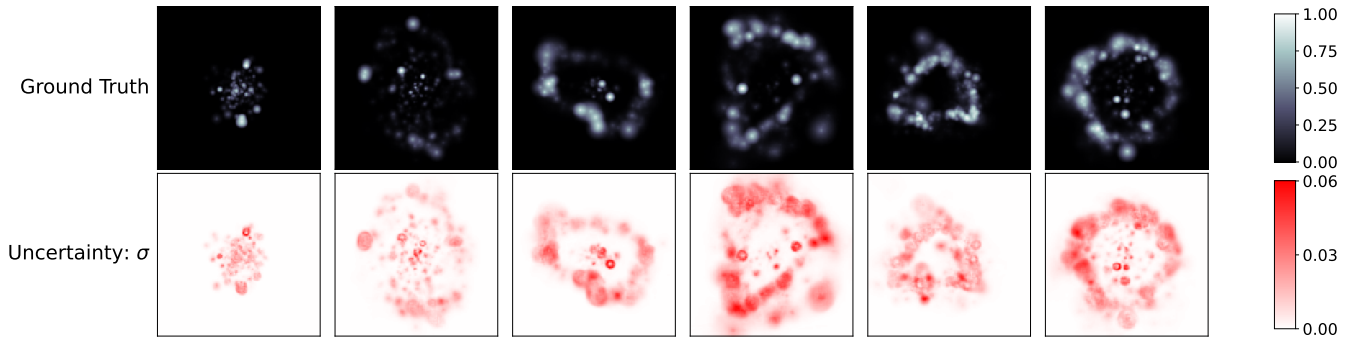
Figure 9. The standard deviations of the resulting predictions made by our eight models. The top row contains ground truth labels, and the bottom row contains the corresponding standard deviation of the outputs of the eight models.

Table 6
Summary Statistics (Mean and Standard Deviation) across the Eight NUV Models Representing the Uncertainty in Our Performance Metrics across the Models

Inferred Band (λ_{eff})	MAE ↓	SSIM ↑	PSNR ↑	Gini : $\mathcal{W}_1$↓	M20 : $\mathcal{W}_1$↓
NUV	0.0089 ± 0.00041	0.96 ± 0.0049	33.2 ± 0.42	0.034 ± 0.0021	0.0081 ± 0.00032

Note. Arrows indicate the desirable direction for each metric: ↑ indicates higher is better; ↓ indicates lower is better.

This alignment suggests that, while the source of uncertainty is algorithmic, it may still carry meaningful information about the underlying structures being modeled.

4. Conclusion

In this study, we have demonstrated the feasibility and effectiveness of generative deep-learning models for translating galaxy observations across photometric bands. Our model is a convolutional neural network with a ResNet-like architecture that uses a combination of MAE loss and SSIM score as a custom loss function. The model is a supervised image-to-image model that can perform interpolation or extrapolation of photometric bands from input band sources. Our experiments use mock galaxy observations from Illustris simulation data and real-world observational data from DECaLS.

The main contributions of this paper include: (i) outlining a straightforward image-to-image model that relies on supervised training using a custom training loss; (ii) demonstrating that our image-to-image model successfully performs a variety of mappings between different photometric bands; and (iii) demonstrating that the Illustris mock observation catalog can be used to develop and prototype image-to-image models (and that successes with the Illustris data set carry over into observed data sets). Our main contribution is a proof of concept that image-to-image models can be used to translate observations of galaxies across different photometric bands.

Experiments with Illustris-trained models performed band interpolation to *G*, *R*, and *Z* bands using *K* and NUV bands as inputs, and band extrapolation using *G* and *R* as inputs to predict *U*, NUV, and FUV. Exceptional results across all metrics were achieved for both band interpolation and extrapolation. SSIM values ranged between 0.98 and 0.99 for interpolation and above 0.91 for extrapolation, indicating that the generated images were nearly identical to the ground truth. Similarly, mean PSNR values ranged from 30 to 40 dB, demonstrating minimal perceptible differences. Furthermore, specialized galaxy morphology metrics, such as the Gini coefficient and M20, displayed strong alignment with the ground truth, evidenced by low $\mathcal{W}_1$ values between ground

truth and generated distributions. This high performance reflects the advantages of working with noise-free, artifact-free data in the Illustris data set, enabling the models to effectively capture the intricate relationships between photometric bands.

The model trained on real observational data from DECaLS performed band interpolation, taking *G*- and *Z*-band data as inputs and interpolating the *R* band. The performance of the DECaLS model was slightly reduced compared to the Illustris data-trained model, which can be attributed to the complexities of real-world observations, such as noise and artifacts in the DECaLS data. Despite this, the mean SSIM score of 0.86 and mean PSNR values of approximately 30 dB reflect a high degree of quality in the generated images. The low $\mathcal{W}_1$ values between Gini and M20 distributions are comparable to values obtained with the Illustris-trained models.

By utilizing mock observations from the Illustris simulations, we have shown that these models can perform both band interpolation and extrapolation with high fidelity, producing outputs that align well with ground truth images across multiple evaluation metrics. The ability of our models to generate real-world data, as evidenced by the DECaLS observations, highlights the robustness and practical applicability of our approach.

On a practical note, our supervised training method becomes feasible when a sufficient amount of labeled observations have already been collected for a subset of the survey area. In such cases, a model can be trained on this partial data set to learn the mapping between available and target photometric bands. Once trained, the model can then be deployed to infer observations in regions where direct observations are unavailable, effectively extending the reach of the survey. This approach is particularly valuable when acquiring additional observations is costly or time constrained. Rather than replacing observational science, our method serves as a complement, providing scientifically plausible predictions in data-sparse regions and helping prioritize future observations. While a supervised model does require some initial training data, it does not necessitate complete area coverage; even a modest set of labeled observations can bootstrap the learning

process and offer practical utility in planning and decision-making.

Another noteworthy practical consideration is that of image scaling. For all our data sets, we scaled each image independently to the range $[0, 1]$ rather than using a global min–max normalization across the entire data set. This design choice is motivated by the fact that the flux ranges in the original, unprocessed images can vary by several orders of magnitude. A global scaling would result in many images having their dynamic range severely compressed, making it difficult for the model to learn meaningful representations. The per-image normalization circumvents this compression issue but prevents us from using a global inverse function to recover physical flux values directly from model outputs. However, this limitation does not preclude the possibility of recovering fluxes in practice. Recent work in the field has shown that machine learning models can effectively predict missing flux values in multiband surveys using available photometric measurements (e.g., N. Chartab et al. 2023; L. Doorenbos et al. 2024). In principle, one could train a secondary model that, given the known physical fluxes of the input bands and the normalized model output, recovers the physical flux of the predicted band. We consider this an exciting direction for future work, especially in contexts where quantitative flux recovery is critical.

Our findings open new opportunities for augmenting astronomical data sets in regions where multiband observations are sparse or unavailable. This can significantly aid in mission planning and enhance high-resolution follow-ups, ultimately contributing to a deeper understanding of galaxy morphology and evolution. Future work will focus on extending the model's capabilities to address limitations such as noise variability and calibration differences across surveys, as well as exploring unsupervised or semisupervised approaches to further enhance the model's generalization capabilities. We would also like to experiment with different model architectures, such as diffusion models.

Acknowledgments

Appendix
Training Details

We trained each of our models for 128 epochs on a V100 GPU, which took approximately 2 hr for each training instance. The model uses around 12 MB of memory and, once trained, can be used to make around 1000 inferences in about 5 minutes (using the same V100 GPU), making it very practical to deploy and run.

A.1. Data Sets

Following the preprocessing discussed in Section 2.1, the images in both the Illustris and DECaLS data sets were resized to a resolution of 128×128. The Illustris training data set comprises 2000 samples (a sample is a set of inputs along with an associated ground truth label), while the testing data set contains 500 samples. The DECaLS training data set contains 8000 samples, while the testing data set contains 2000 samples. We conducted hyperparameter tuning using the training data set only. For a detailed outline of our hyperparameter tuning, see Appendix A.2. All of the results shown in this paper, including the metrics, were obtained using the testing data sets.

The reason for the DECaLS data sets being larger than the Illustris data sets is that the DECaLS samples contain more noise and background artifacts, which makes learning from the DECaLS data set more challenging compared to the Illustris data set. To compensate for this, we selected a larger number of samples for the DECaLS data set compared to the Illustris data set.

We sorted the Illustris synthetic image catalog by galaxy mass. From the original set of 6000 images, we selected the top 2500 highest-mass galaxies. During the subsequent train–test split, we applied stratified sampling based on star formation rate to ensure that our subsets were representative of the overall diversity in galaxy morphology (we found star formation rate to have a very high correspondence to galaxy morphology).

For the DECaLS data set, we benefited from prior morphological classifications provided by Galaxy Zoo 2 (K. W. Willett et al. 2013) and available through the astroNN package, which contains approximately 17,000 galaxy images categorized into 10 imbalanced classes. From this collection, we selected images from classes 5, 6, and 7, which correspond to spiral galaxy morphologies (see H. W. Leung & J. Bovy 2018 for astroNN's class details). Our motivation for focusing on spiral galaxies is that they tend to exhibit the highest variation across photometric bands, making them well suited for evaluating band-to-band transformations.

A.2. Generator Hyperparameters

We based our model architecture on the model presented in J.-Y. Zhu et al. (2017), which proposed a generator architecture with a ResNet backbone, as in K. He et al. (2016). The architecture comprises downsampling blocks followed by residual blocks, first presented in K. He et al. (2016), and finally upsampling blocks. The downsampling blocks reduce the spatial dimensions of the input image while capturing high-level features. This is achieved using strided convolutions, which compress the image's resolution, allowing the network to focus on broader patterns and context. The residual block preserves the input's spatial information while transforming features. It consists of convolutional layers with skip connections that add the input directly to the output, facilitating gradient flow during training. The upsampling block restores the compressed spatial dimensions back to the original size using transposed convolutions. This process reconstructs finer details in the output image. For more information on different types of convolutions used in deep learning, see V. Dumoulin & F. Visin (2016).

Note that hyperparameters are the configuration parameters of the model and are not learned during training. We conducted extensive hyperparameter tuning. Initially, various hyperparameters were investigated individually to conclude if they had any tangible effect on the resulting images. The hyperparameters investigated include: the optimizer learning rate, the number of filters in the convolutional layers, the activation functions, and different normalization layers, such as batch normalization, instance normalization, and group

normalization (D. Ulyanov et al. 2016; N. Bjorck et al. 2018; Y. Wu & K. He 2018). We also experimented with inserting attention layers (A. Vaswani et al. 2017) between the model's residual blocks. None of the previously mentioned adjustments had a noticeable impact on model performance from a qualitative or quantitative standpoint. Furthermore, in the case of the model with attention layers, the model took significantly longer to train. The only factors that had a reliably noticeable effect on model performance were the number of down-sampling and upsampling blocks as well as the number of residual blocks in the latent space, and so further testing was done by varying these two hyperparameters while keeping all other hyperparameters unchanged.

We performed a grid search through the hyperparameter space and tested models with one, two, and three down-sampling and upsampling blocks and with three, six, and nine residual blocks as the latent space for a total of nine different hyperparameter configurations. J.-Y. Zhu et al. (2017) proposed that, for images of size 128×128, the generator architecture should comprise two downsampling blocks followed by a latent space containing six residual blocks and two upsampling blocks. We found that one downsampling and upsampling block and nine residual blocks works best. Our reasoning for this configuration working better is that it contains a wider bottleneck (in other words, less down-sampling or compression of the input) and relies more on residual learning. A small bottleneck does not transmit high-frequency features, such as spirals and bars, but only transmits low-frequency features such as overall shape and orientation. Given that the most significant differences between photo-metric bands are mainly high-frequency features, it follows that the more aggressive the bottleneck, the less adept our network is at resolving these high-frequency features.

ORCID iDs

Youssef Zaazou ⓘ https://orcid.org/0009-0007-1095-9935
Alex Bihlo ⓘ https://orcid.org/0000-0002-4024-3352
Terrence S. Tricco ⓘ https://orcid.org/0000-0002-6238-9096

References

Ali, P. J. M., & Faraj, R. H. 2014, Data Normalization and Standardization: A Technical Report, Machine Learning Technical Reports, Koya University
Bessell, M. S. 2005, ARA&A, 43, 293
Bjorck, N., Gomes, C. P., Selman, B., & Weinberger, K. Q. 2018, Advances in Neural Information Processing Systems 31, ed. S. Bengio, (NeurIPS) https://proceedings.neurips.cc/paper_files/paper/2018/hash/36072923bfc3cf47745d704feb489480-Abstract.html
Brunet, D., Vrscay, E. R., & Wang, Z. 2011, ITIP, 21, 1488
Chartab, N., Mobasher, B., Cooray, A. R., et al. 2023, ApJ, 942, 91
De La Calleja, J., & Fuentes, O. 2004, MNRAS, 349, 87
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Doorenbos, L., Sextl, E., Heng, K., et al. 2024, ApJ, 977, 131
Dumoulin, V., & Visin, F. 2016, arXiv:1603.07285
Esser, P., Kulal, S., Blattmann, A., et al. 2024, arXiv:2403.03206
Fluke, C. J., & Jacobs, C. 2019, WIREs Data Min. Knowl. Discov., 10, e1349
Gatys, L. A., Ecker, A. S., & Bethge, M. 2016, in 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 2414
Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2020, Commun. ACM, 63, 139
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 770
Ho, J., Jain, A., & Abbeel, P. 2020, Advances in Neural Information Processing Systems 33, ed. H. Larochelle et al. (NeurIPS) https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html
Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. 2017, in 2017 IEEE Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 5967
Kinakh, V., Belousov, Y., Quétant, G., et al. 2024, Senso, 24, 1151
Kouroumpatzakis, K., Zezas, A., Kyritsis, E., Salim, S., & Svoboda, J. 2023, A&A, 673, A16
Leung, H. W., & Bovy, J. 2018, MNRAS, 483, 3255
Lin, Q., Fouchez, D., & Pasquet, J. 2021, in 2020 25th Int. Conf. on Pattern Recognition (ICPR) (Piscataway, NJ: IEEE), 5634
Liu, T., Quan, Y., Su, Y., et al. 2025, NatAs, 9, 608
Long, J., Shelhamer, E., & Darrell, T. 2015, in 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 3431
Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163
Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, AJ, 118, 1406
Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJ, 619, L1
Pang, Y., Lin, J., Qin, T., & Chen, Z. 2021, IEEE Trans. Multimed., 24, 3859
Quan, Y., Chen, M., Pang, T., & Ji, H. 2020, in 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 1887
Rubner, Y., Tomasi, C., & Guibas, L. J. 2000, Int. J. Comput. Vis., 40, 99
Saharia, C., Chan, W., Chang, H., et al. 2022, in ACM SIGGRAPH 2022 Conf. Proc., ed. M. Nandigjav, N. J. Mitra, & A. Hertzmann (New York: Association for Computing Machinery), 15
Skrutskie, M., Cutri, R., Stiening, R., et al. 2006, AJ, 131, 1163
Smith, M. J., Geach, J. E., Jackson, R. A., et al. 2022, MNRAS, 511, 1808
Spindler, A., Geach, J. E., & Smith, M. J. 2021, MNRAS, 502, 985
Torrey, P., Snyder, G. F., Vogelsberger, M., et al. 2015, MNRAS, 447, 2753
Ulyanov, D., Vedaldi, A., & Lempitsky, V. 2016, arXiv:1607.08022
Van den Bergh, S. 1998, Galaxy Morphology and Classification (Cambridge: Cambridge Univ. Press)
Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, Advances in Neural Information Processing Systems 30, ed. I. Guyon, (NeurIPS) https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, Natur, 509, 177
Vojtekova, A., Lieu, M., Valtchanov, I., et al. 2020, MNRAS, 503, 3204
Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835
Wu, Y., & He, K. 2018, in Computer Vision – ECCV 2018, ed. V. Ferrari et al. (Cham: Springer), 3
York, D. G., Adelman, J., Anderson, J. E., Jr, et al. 2000, AJ, 120, 1579
Zaazou, Y., Bihlo, A., & Tricco, T. 2025, Code for Mapping Galaxy Images Across Ultraviolet, Visible and Infrared Bands Using Generative Deep Learning, v1.0, Zenodo, doi:10.5281/zenodo.15316079
Zhang, Y., & Zhao, Y. 2015, DatSJ, 14, 11
Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. 2017, in 2017 IEEE Int. Conf. on Computer Vision (Piscataway, NJ: IEEE), 2242