universe

MDPI

Article

From Non-Parametric Statistics to Images: Generating Galaxy Morphologies

Quanfeng Xu [1,2], Mi Chen [3], Renhao Ye [1,2], Siyuan Chen [1,2], Nan Li [4] and Shiyin Shen [2,5,*]

[1] School of Astronomy and Space Science, University of Chinese Academy of Sciences, 1 East Yanqi Lake Rd., Beijing 100049, China; xuquanfeng@shao.ac.cn (Q.X.); renhaoye@shao.ac.cn (R.Y.); chensiyuan@shao.ac.cn (S.C.)
[2] Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Rd., Shanghai 200030, China
[3] Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands; chenmi@astro.rug.nl
[4] National Astronomical Observatory, Chinese Academy of Sciences, Beijing 100101, China; nan.li@nao.cas.cn
[5] Key Lab for Astrophysics, Shanghai 200234, China
* Correspondence: ssy@shao.ac.cn

Abstract

Non-parametric statistics characterize complex galaxy morphologies but do not specify how to construct images that reproduce those statistics. We present Flow-VAE, which combines a variational autoencoder with a conditional normalizing flow to generate images conditioned on concentration, asymmetry, smoothness (CAS), Gini coefficient (G), and M_{20}. A Gaussian mixture model completes unspecified descriptors and measures the training-distribution support of requested combinations. Using a dataset of 240,000 DESI Legacy Imaging Surveys r-band images, we evaluate fidelity by remeasuring morphological statistics from model outputs and comparing them with the inputs. With a tolerance of $\pm 10\%$ of each descriptor's adopted range, descriptor-specific recovery rates are 96.62–99.97% for reconstruction and 87.81–99.83% for single-parameter generation among valid measurements. Generated descriptors show regression toward typical training values, while less-supported joint conditions produce broader residuals and systematic offsets. These results quantify the extent to which Flow-VAE can generate galaxy images consistent with prescribed non-parametric morphological statistics.

Keywords: galaxy morphology; variational autoencoders; normalizing flows; non-parametric statistics

1. Introduction

Galaxies are intrinsically complex three-dimensional structures that are projected into two-dimensional morphological images in observational astronomy. Analyzing these morphologies provides essential insights into galaxy formation and evolution [1]. Traditionally, visual classification schemes such as the Hubble sequence [2] were used to categorize these structures; however, such subjective approaches are often too simplistic to capture their full structural diversity and evolutionary intricacies. By mapping the distribution of surface brightness [3], features such as spiral arms, bulges, and disks reveal underlying physical processes, including mergers, star formation, and internal evolution. This morphological information helps determine the mass distribution of a galaxy, its interactions with the environment, and the evolution of large-scale cosmic structures, serving as a vital observational test for cosmological models.

Check for updates

Academic Editor: Marc S. Seigar

Received: 12 October 2026
Revised: 7 November 2026
Accepted: 13 November 2026
Published: 18 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

A widely used approach to quantitative morphology is to represent the galaxy surface-brightness distribution with an analytical profile. For example, the Sérsic profile describes the radial light distribution using the Sérsic index (n), with lower and higher values generally corresponding to disk-dominated and centrally concentrated systems, respectively. Many profile-fitting tools have been developed to efficiently fit galaxy surface-brightness distributions, including `GALFIT` [4], `ProFit` [5], and `GalMOSS` [6]. Although these models provide compact and interpretable descriptions of the global galaxy structure, their results depend on the adopted profile, initialization, and fitting procedure. Moreover, a small number of smooth analytical components cannot always reproduce irregular substructures, asymmetries, or clumpy emission, particularly for high-redshift and interacting galaxies, potentially biasing the inferred morphological properties [7].

Non-parametric statistics provide a complementary description of galaxy morphology by measuring the pixel-level flux distribution without assuming an explicit analytical surface-brightness profile. One of the most widely used systems is Concentration, Asymmetry, and Smoothness (CAS) [8], where concentration measures the central concentration of light, asymmetry quantifies deviations from rotational symmetry, and smoothness characterizes small-scale structures. Other commonly used descriptors include the Gini coefficient, M_{20}, and the multimode, intensity, and deviation statistics (MID, [9]), which capture complementary aspects of flux inequality, the spatial distribution of bright regions, and multi-component structure. These measurements can be computed efficiently using tools such as `statmorph` [10] and its CSST-oriented extension `statmorph_csst` (https://github.com/pentyum/statmorph_csst.git, accessed on 6 November 2026). Non-parametric statistics have consequently been widely applied to studies of galaxy mergers [8], star formation [11], and morphological evolution [12], as well as to the study of the morphological evolution of the galaxy population as a whole [13].

Two properties make these indices particularly well suited to the present problem. First, they are model-independent: they are computed from the pixel distribution of the galaxy and require no assumed surface-brightness profile, so they remain well-defined for clumpy, asymmetric, and interacting systems for which a Sérsic fit is ill-conditioned. Second, because they are defined by a measurement recipe rather than by a fitted model, the same estimator can be applied without modification to observed images, to mock images drawn from cosmological simulations [10], and to synthetic images produced by a generative model. This last property is what makes the closed-loop validation performed in this work possible.

Because these five descriptors are strongly correlated in the observed galaxy population, specifying one of them does not specify the others, and sampling the remaining four independently would produce combinations that are rare or absent in the data. We therefore model the joint distribution of (C, A, S, G, M_{20}) with a Gaussian mixture model (GMM), a weighted sum of K multivariate Gaussians whose weights, means, and covariances are fitted to the training sample. A GMM provides a flexible density estimate in a low-dimensional space, and we can evaluate both its conditional distributions and its density in closed form. This is what allows a partially specified condition vector to be completed in a way that respects the observed correlations (Section 3.4), and it supplies an explicit and interpretable measure of how well a requested combination is represented in the training data (Section 4.2.2).

However, the absence of an explicit generative profile makes the inverse problem fundamentally underdetermined. Non-parametric statistics compress a two-dimensional surface-brightness distribution into a small set of scalar quantities, and morphologically distinct galaxies can therefore share similar values of C, A, S, G, and M_{20}. Although these statistics can be measured directly from an observed image, no unique analytical mapping

exists from a prescribed set of descriptors to a realistic galaxy image. This limitation becomes particularly important when the goal is not only to characterize observed galaxies but also to generate simulated galaxies with controlled morphological properties.

Traditional simulation tools such as `GalSim` [14] efficiently generate galaxy images from analytical surface-brightness profiles. However, these profiles do not naturally reproduce or directly control the asymmetric, clumpy, and multi-component structures captured by non-parametric statistics. Deep generative models offer a potential solution by learning the distribution of galaxy images directly from observational data, thereby generating realistic morphologies beyond simple analytical profiles. GalSim-Hub [15], for example, developed a Flow-VAE conditioned on galaxy size, magnitude, and photometric redshift. For Euclid image simulations, Euclid Collaboration et al. [16] conditioned a generative model on the half-light radius, axis ratio, and Sérsic index. Compared with conventional parametric simulations, these data-driven approaches can reproduce a richer range of galaxy morphologies and more realistic substructures. Table 1 summarizes how their conditioning variables compare with those adopted here.

Table 1. Comparison of the conditioning variables used by data-driven galaxy image simulation and generation methods.

Study	Conditioning Variables	Type of Variable
`GalSim` [14]	Profile parameters	Parametric
GalSim-Hub [15]	Magnitude, size, z_{phot}	Global photometric
Euclid [16]	Size, magnitude, z, axis ratio, Sérsic n	Mixed parametric
This work	C, A, S, G, M_{20}	Non-parametric, measured

Nevertheless, the conditioning variables employed by these two generative models predominantly encode global properties of galaxy structure, and thus provide limited control over higher-order morphological features, such as asymmetry, smoothness, and clumpiness. In addition, parameters such as the half-light radius and the Sérsic index are sensitive to the specific measurement protocol or profile-fitting methodology adopted, complicating the assessment of whether a stochastically generated image faithfully satisfies the prescribed constraints. Therefore, Conditioning generation on directly measurable, non-parametric summary statistics would not only be beneficial for simulation studies that require detailed morphological features of galaxies, such as the modeling of the morphological characteristics of sources in strong gravitational lensing [17], but also allow quantitative checks that generated galaxies match requested properties.

In fact, for stochastic generative models, visual realism and agreement with the population distribution do not ensure that individual outputs satisfy their prescribed conditions. Remeasuring the conditioning statistics from generated images and comparing them directly with the inputs therefore provides an essential object-level validation. In this study, we develop a Flow-VAE model conditioned on five widely used non-parametric morphological statistics: concentration (C), asymmetry (A), smoothness (S), the Gini coefficient (G), and M_{20}. Our objective is to quantify conditional fidelity by comparing the requested morphological statistics with those remeasured from generated images. This closed-loop evaluation is a key methodological contribution of our framework, as it quantifies conditional accuracy and reveals systematic biases across the sampled parameter space.

The remainder of this paper is organized as follows. Section 2 describes the galaxy sample, image preprocessing, and measurement of the non-parametric statistics. Section 3 introduces the Flow-VAE architecture and the GMM used to model the joint descriptor distribution. Section 4 evaluates reconstruction fidelity, single-parameter control, and generation under low-density conditions, and validates the generated images against

structural diagnostics that are not used as conditioning variables. Section 5 discusses the model's capabilities and limitations, and Section 6 summarizes the main conclusions.

2. Datasets

2.1. Galaxy Images and Preprocessing

We construct a galaxy sample from the SDSS main galaxy sample (MGS; $r_{\rm petro} < 17.77$) [18], using imaging from the DESI Legacy Imaging Surveys [19]. Table 2 summarizes the selection procedure and the number of objects remaining at each step. For model training, we use 64×64 r-band cutouts with a pixel scale of $0''.262$. This cutout size was chosen to keep the computational cost manageable.

Following the catalogue matching and source-detection requirements listed in Table 2, we apply quality and range cuts to the 42 parameters measured by SExtractor and `statmorph`. The final selection retains galaxies with $r \gtrsim 15$ and $z \leq 0.3$, excludes the most extended and most compact sources, and requires sufficient space within each cutout for both the galaxy measurement aperture and a background annulus. We draw a working sample of 240,000 galaxies from the surviving objects. These requirements favour compact galaxies of intermediate surface brightness at fixed magnitude, so the model describes this selected population rather than the full MGS.

Because non-parametric morphological statistics are derived from the image flux distribution, contamination from foreground stars and neighbouring galaxies can bias their measurements. We therefore perform source detection on each cutout using `SExtractor` [20], adopting a detection threshold of 1.5σ above the local background and a minimum detection area of five connected pixels. Each cutout is centred on the catalogued position of an MGS galaxy, and the detection containing the central pixel is identified as the target. All other detections are treated as potential contaminants. From the resulting segmentation map, we construct a binary mask that retains pixels assigned to the target galaxy and the background, while excluding pixels assigned to other detected sources. Figure 1 illustrates this procedure for a representative cutout. The resulting masks are used in the morphological measurements and the reconstruction loss to reduce contamination from neighbouring sources.

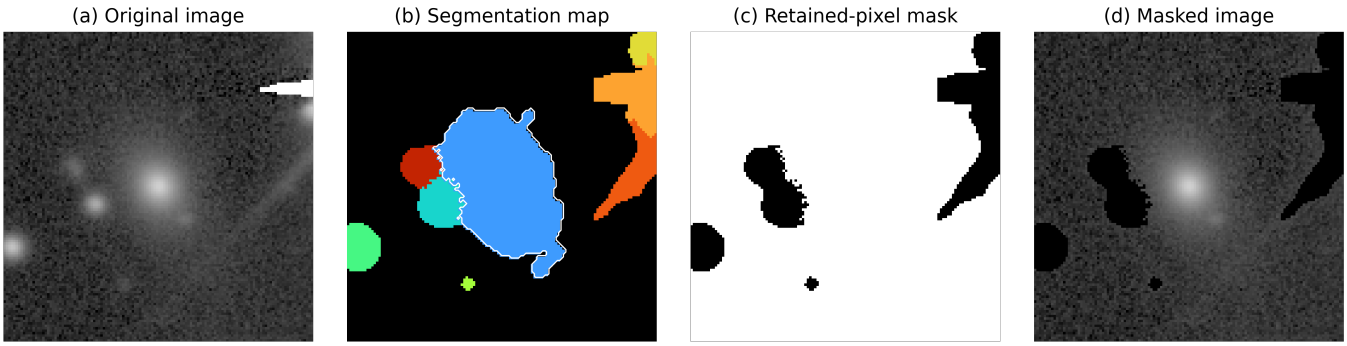


Figure 1. Example of contaminant masking. (**a**) Original image. (**b**) SExtractor segmentation map. (**c**) Binary mask retaining the target galaxy and background. (**d**) Masked image. The blue contour in (**b**) outlines the target galaxy; black regions in (**c**,**d**) indicate excluded sources.

Table 2. Sample selection and the number of objects remaining after each step. The final row combines the quality and range cuts with the selection of the working sample.

Step	Selection Criterion	*N*
0	SDSS MGS: (VAGC_SELECT & 4) $\neq 0$, $0 < z < 0.5$, and $r_{\rm petro} < 17.77$	633,568
1	DECaLS southern imaging footprint: NGC with $\delta < 32.375^{\circ}$, and the full SGC	410,184

Table 2. *Cont.*

Step	Selection Criterion	N
2	Match to a DR9 BGS_BRIGHT target within 0.″79	394,465
3	Availability of a 64 × 64 cutout	374,203
4	SExtractor detection within the central 8 × 8-pixel region in the g, r, and z bands	371,490
5	Quality and range cuts, followed by selection of the working sample	240,000

2.2. Non-Parametric Statistics

Before performing non-parametric statistical analyses on each galaxy sample, we use the previously generated `SExtractor` masks to ensure that only pixels belonging to the target galaxy contribute to the calculations. We then utilize `Statmorph` [10] on the masked images to extract the galaxies' non-parametric morphological statistics, which comprise *concentration* (C), *asymmetry* (A), *smoothness* (S) (CAS, [8]), the Gini coefficient (G, [21]), and the second-order moment of the brightest 20% of the galaxy's flux (M_{20}, [21]). The definitions of these non-parametric morphological statistics are listed below.

- **CAS**
  - *Concentration* (C) quantifies how concentrated the light is within a galaxy by comparing the radii that contain 20% and 80% of the total light. It is calculated using the formula:
$$C = 5 \log_{10}\left(\frac{r_{80}}{r_{20}}\right) \tag{1}$$
  - *Asymmetry* (A) measures the rotational symmetry of a galaxy by rotating its image 180 degrees and comparing it to the original. The calculation is given by:
$$A = \frac{\sum |I - I_{180^\circ}|}{\sum |I|} - B_A \tag{2}$$
where I is the galaxy image and B_A is the dimensionless background-asymmetry correction.
  - *Smoothness* (S) quantifies the presence of small-scale structures, such as star clusters, by comparing a smoothed version of the galaxy image to the original. It is calculated as:
$$S = \frac{\sum |I - I_S|}{\sum |I|} - B_S \tag{3}$$
where I_S is the galaxy's image smoothed by a boxcar of width 0.25 times the petrosian radius r_p. The term B_S is a background correction that accounts for small-scale fluctuations introduced by sky noise.
- **G**: The *Gini coefficient* (G) measures the distribution of light across the pixels of a galaxy, with values ranging from 0 (uniform light distribution) to 1 (highly concentrated light). It is defined as:
$$G = \frac{1}{2\bar{I}n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |I_i - I_j| \tag{4}$$
where $\bar{I}$ represents the mean brightness of all pixels. A higher Gini coefficient reflects light concentrated in a few regions.
- M_{20}: The M_{20} metric represents the second-order moment of the brightest 20% of the pixels, capturing the spread of light and highlighting bright off-centre regions. It is calculated as:
$$M_{\text{tot}} = \sum_{i}^{n} M_i = \sum_{i}^{n} f_i \cdot ((x_i - x_{\text{cen}})^2 + (y_i - y_{\text{cen}})^2) \tag{5}$$

$$M_{20} = \log_{10}\left(\frac{\sum_i M_i}{M_{\rm tot}}\right), \sum_i f_i < 0.2 f_{\rm tot} \quad (6)$$

where $(x_{\rm cen}, y_{\rm cen})$ denotes the galaxy's centroid, and M_i is the product of each bright pixel's flux f_i and the square of its distance from the galaxy center.

We measured the non-parametric morphological statistics of 240,000 r-band galaxy images using the `Statmorph_csst` package (version 1.0). Although several of these statistics are formally defined over much broader domains, such as $M_{20} \in (-\infty, 0]$ and $G \in [0, 1]$, the values measured for real galaxies occupy substantially narrower regions of parameter space. To exclude rare extreme measurements and restrict the generative model to morphologies representative of the DESI MGS sample, we adopt ranges that contain approximately 99% of the observed values: $C \in [0.5, 4.5]$, $A \in [-0.7, 1.0]$, $S \in [-0.25, 0.2]$, $G \in [0.4, 0.95]$, and $M_{20} \in [-2.5, -0.5]$.

Among the galaxies analysed 23.14% have $A < 0$, and 34.43% have $S < 0$. Negative values can arise when the background correction exceeds the normalized image residual, but do not by themselves establish that the corresponding morphological structure is undetectable. We retain these values as conditioning targets because the model aims to reproduce measured, background-corrected descriptors.

Figure 2 illustrates how each statistic relates to galaxy structure within these adopted ranges. For each row, one target parameter is uniformly sampled across its allowed interval, while the remaining four parameters are constrained near their median values in the dataset. This controlled sampling isolates the morphological variation associated with each statistic and provides a direct visual interpretation of the parameter space used for model training and generation.

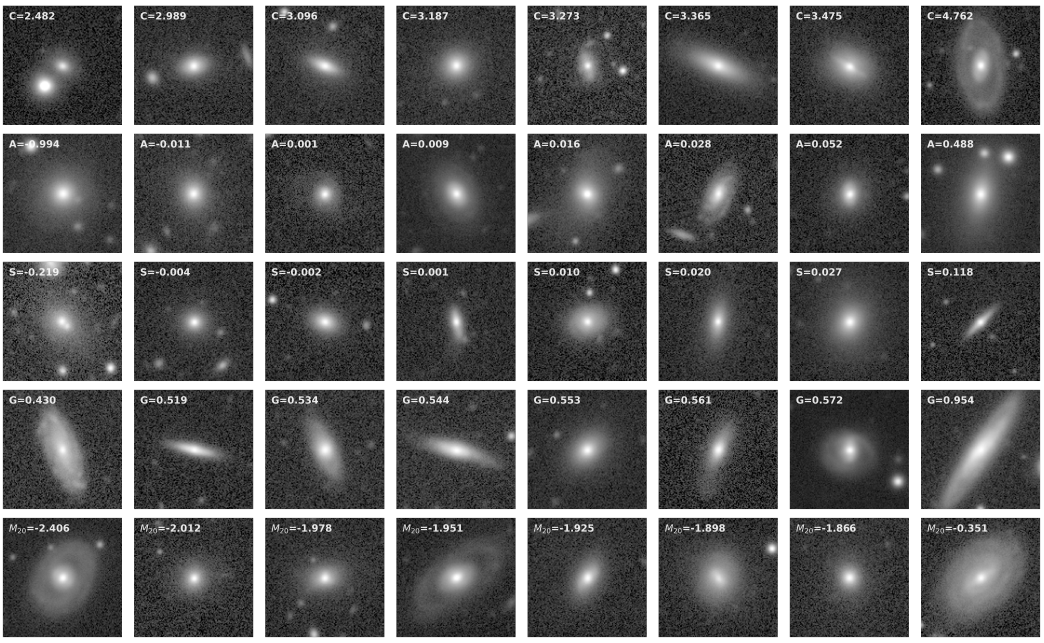


Figure 2. Visualizing the physical interpretation of non-parametric morphological statistics through controlled sampling of real galaxy images. Each row isolates the effect of a single target parameter, which is uniformly sampled across its numerical range from left to right, while the other four parameters are constrained near their overall median values. The visual sequences explicitly demonstrate the structural transitions captured by each statistic: C (Diffuse $\rightarrow$ Concentrated), A (Symmetric $\rightarrow$ Asymmetric), S (Smooth $\rightarrow$ Clumpy/Structured), G (Uniform $\rightarrow$ Concentrated Flux), and M_{20} (Core-dominated $\rightarrow$ Extended/Multi-core). The exact value of the varying parameter is annotated in the top-left corner of each panel.

3. Methods

In this section, we first briefly introduce the principles of the VAE (Section 3.1) models and Flow (Section 3.2). We then use Flow-VAE (Section 3.3) to map the distributions of non-parametric statistics, latent variables, and galaxy images to each other.

3.1. Variational Auto-Encoders

Variational Autoencoders (VAE, [22]) learn probabilistic latent representations of complex data distributions [23]. Unlike conventional autoencoders, which map each input to a fixed latent representation, a VAE represents each input as a probability distribution in latent space. A VAE consists of three main components: an encoder, a variational inference mechanism, and a decoder. The encoder maps a galaxy image $\mathbf{x}$ to a distribution in an N-dimensional latent space, typically parameterized as a Gaussian distribution with mean $\mu(\mathbf{x})$ and variance $\sigma(\mathbf{x})^2$:

$$\mathbf{z} \sim q_\phi(\mathbf{z} \mid \mathbf{x}) = \mathcal{N}\Big(\mu(\mathbf{x}), \mathrm{diag}\Big[\sigma(\mathbf{x})^2\Big]\Big), \qquad \mathbf{z} \in \mathbb{R}^N. \tag{7}$$

Here, $q_\phi(\mathbf{z} \mid \mathbf{x})$ denotes the approximate posterior distribution of the latent variables inferred from the input galaxy image. During training, a latent vector $\mathbf{z}$ is sampled from this posterior and passed to the decoder, which maps it back to image space to produce the reconstructed image $\hat{\mathbf{x}}$. To regularize the latent space, the VAE minimizes the *Kullback–Leibler* divergence (KL, [24]) between the approximate posterior and a predefined prior distribution. The overall loss is therefore

$$\mathcal{L}_{\mathrm{VAE}} = \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{KL}} D_{\mathrm{KL}}\big(q_\phi(\mathbf{z} \mid \mathbf{x}) || p_{\mathrm{prior}}(\mathbf{z})\big), \tag{8}$$

where $\mathcal{L}_{\mathrm{rec}}$ measures the reconstruction error between $\mathbf{x}$ and $\hat{\mathbf{x}}$, $p_{\mathrm{prior}}(\mathbf{z})$ is the adopted latent prior, and λ_{KL} controls the balance between reconstruction fidelity and posterior regularization. This formulation encourages a continuous and regularized latent space while preserving sufficient information for image reconstruction and generation.

3.2. Normalizing Flow

Normalizing flows model complex probability densities by transforming a simple base distribution through a sequence of invertible and differentiable mappings, $f_\theta = f_{\theta_N} \circ \cdots \circ f_{\theta_1}$ [25,26]. In the context of our VAE framework, normalizing flows are utilized to learn the conditional latent density $p_\theta(\mathbf{z} \mid \mathbf{y})$ [27]. Here, $\mathbf{z}$ represents the VAE latent vector, and $\mathbf{y} = (C, A, S, G, M_{20})$ denotes the vector of non-parametric morphological statistics.

For a fixed context $\mathbf{y}$, the forward transformation maps the VAE latent vectors $\mathbf{z}$ to a simple base distribution variable $\mathbf{u}$:

$$\mathbf{u} = f_\theta(\mathbf{z}; \mathbf{y}), \qquad \mathbf{z} = f_\theta^{-1}(\mathbf{u}; \mathbf{y}), \qquad \mathbf{z}, \mathbf{u} \in \mathbb{R}^N. \tag{9}$$

Crucially, the morphological descriptors $\mathbf{y}$ condition the transformation; they are not its transformed output. Key properties of normalizing flows include differentiability, which ensures the Jacobian determinant can be computed, and invertibility enables the probability density to be evaluated exactly through the change-of-variables formula.

We employ a Masked Autoregressive Flow (MAF, [27]) for its expressiveness and computational efficiency. The parameters $\mu_b(\mathbf{y})$ and $\sigma_b(\mathbf{y})$ are predicted from the conditioning vector by a learned neural mapping. The conditional base distribution is parameterized as a diagonal Gaussian:

$$p_{\mathrm{b}}(\mathbf{u} \mid \mathbf{y}) = \mathcal{N}\Big(\boldsymbol{\mu}_{\mathrm{b}}(\mathbf{y}), \mathrm{diag}\Big[\sigma_{\mathrm{b}}^2(\mathbf{y})\Big]\Big). \tag{10}$$

The same context is supplied to every autoregressive transform. Thus, the morphological descriptors control both the Gaussian base density and the transformation to the latent distribution.

The training of normalizing flows involves optimizing the parameters θ by maximizing the log-likelihood of the observed data. Following the change-of-variables formula, the conditional log-likelihood is computed as:

$$\log p_\theta(\mathbf{z} \mid \mathbf{y}) = \log p_{\mathrm{b}}(f_\theta(\mathbf{z}; \mathbf{y}) \mid \mathbf{y}) + \log \left| \det \frac{\partial f_\theta(\mathbf{z}; \mathbf{y})}{\partial \mathbf{z}} \right|. \tag{11}$$

For image-derived latent vectors and their paired morphological contexts, the flow loss is expressed as the negative log-probability:

$$\mathcal{L}_{\mathrm{flow}} = -\frac{1}{B} \sum_{i=1}^{B} \log p_\theta(\mathbf{z}_i \mid \mathbf{y}_i). \tag{12}$$

During the generation phase, the model draws $\mathbf{u} \sim p_{\mathrm{b}}(\mathbf{u} \mid \mathbf{y})$, applies the inverse mapping $\mathbf{z} = f_\theta^{-1}(\mathbf{u}; \mathbf{y})$, and generates the galaxy image $\hat{\mathbf{x}}$ via the VAE decoder parameterizes $p_\psi(\mathbf{x} \mid \mathbf{z})$. Sampling the conditional base distribution allows different image realizations to be generated from the same prescribed morphological descriptors.

3.3. Our Framework

Figure 3 summarizes the proposed Flow-VAE framework for generating galaxy images conditioned on non-parametric morphological statistics. Each r-band galaxy image $\mathbf{x}$ is paired with its morphological descriptor vector $\mathbf{y} = (C, A, S, G, M_{20})$ and a source mask $\mathcal{M}$ derived from `SExtractor`. The VAE encoder maps the input image to a latent representation $\mathbf{z}$, while the conditional normalizing flow models the distribution $p_\theta(\mathbf{z} \mid \mathbf{y})$. In this formulation, the morphological descriptors act as conditioning variables that constrain the latent distribution rather than as direct outputs of the flow.

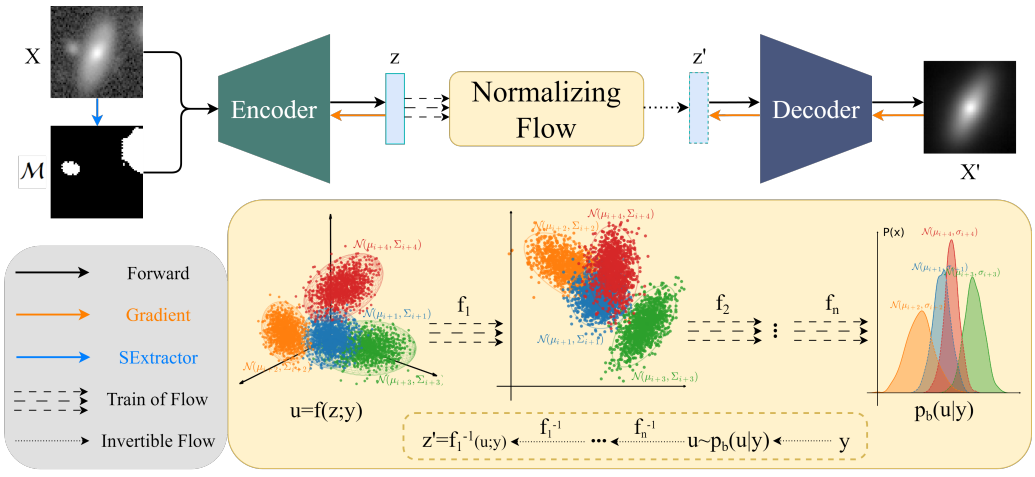


Figure 3. Overview of the Flow-VAE framework. Galaxy images are encoded into VAE latent representations, while the non-parametric morphological descriptors condition the normalizing flow. Source masks are applied to the VAE reconstruction loss to reduce contamination from neighbouring sources.

To reduce the influence of neighbouring sources, the reconstruction loss is evaluated over the unmasked pixels, which include the target galaxy and surrounding background. The masked reconstruction loss is defined as

$$\mathcal{L}_{\mathrm{rec}_{\mathcal{M}}} = \mathcal{M} \odot \|\mathbf{x} - \hat{\mathbf{x}}\| \tag{13}$$

where $\hat{\mathbf{x}}$ is the reconstructed image and $\mathcal{M}$ is the binary source mask described in Section 2.1. This masking helps reduce the influence of contaminating sources when optimizing the reconstruction of the target galaxy.

The VAE and conditional flow are optimized to preserve both image fidelity and the statistical structure of the latent space. The overall objective is written as

$$\mathcal{L}_{\text{Flow-VAE}} = \mathcal{L}_{\text{rec}_{\mathcal{M}}} + \lambda_{\text{flow}} \cdot \mathcal{L}_{\text{flow}} + \lambda_{\text{KL}} \cdot \mathcal{L}_{\text{KL}}. \tag{14}$$

where $\mathcal{L}_{\text{KL}}$ regularizes the VAE posterior toward its prior and $\mathcal{L}_{\text{flow}}$ maximizes the conditional likelihood of the image-derived latent vectors given the morphological descriptors. The coefficient λ_{flow} controls the contribution of the conditional flow objective and is set to 0.1 in this work.

During generation, a target descriptor vector $\mathbf{y}$ is first specified and a base-space sample $\mathbf{u}$ is drawn from the corresponding conditional base distribution. The inverse flow maps this sample to a VAE latent vector, $\mathbf{z} = f_\theta^{-1}(\mathbf{u}; \mathbf{y})$, which is then passed to the decoder to generate a galaxy image $\hat{\mathbf{x}}$. Because the sampling process is stochastic, multiple distinct galaxy images can be generated from the same prescribed morphological statistics, allowing the model to represent the intrinsic morphological diversity that is not captured by the five scalar descriptors alone. A practical implementation of generation from a complete five-dimensional descriptor vector is provided in Appendix A.

We train the model using the Adam optimizer [28] with a learning rate of 0.001, a batch size of 256, and 100 training epochs. The 240,000 preprocessed *r*-band galaxy images are randomly divided into 216,000 training images and 24,000 test images. The model is implemented in PyTorch (2.15.0) and trained on a single NVIDIA A100 GPU with 80 GB of memory.

3.4. Gaussian Mixture Model for Parameter

Flow-VAE generates images from specified morphological descriptors, but its conditional accuracy may depend on how well the requested parameter combination is represented in the training data. We model their joint distribution using a Gaussian mixture model (GMM, [29]), which represents the density of $\mathbf{y} = (C, A, S, G, M_{20})$ as

$$p_{\text{GMM}}(\mathbf{y}) = \sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \qquad w_k \geq 0, \quad \sum_{k=1}^{K} w_k = 1, \tag{15}$$

where w_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$ are the weight, mean, and covariance of component k. We adopt a GMM because it captures correlations and possible multimodality in the joint distribution while allowing analytic conditioning and density evaluation. These properties enable us to complete partially specified condition vectors using $p_{\text{GMM}}(\mathbf{y}_{-j} \mid y_j)$ and to assess training-set support for requested parameter combinations.

This joint model supplies the missing conditions when only one descriptor is specified. Writing the fixed descriptor as y_j and the remaining four as $\mathbf{y}_{-j}$, their conditional distribution is

$$p_{\text{GMM}}(\mathbf{y}_{-j} \mid y_j) = \frac{p_{\text{GMM}}(\mathbf{y})}{p_{\text{GMM}}(y_j)}. \tag{16}$$

For a GMM, conditioning gives another Gaussian mixture, with component weights, means, and covariances determined by the specified value. We sample $\mathbf{y}_{-j}$ from this distribution and combine it with y_j to form a complete condition for Flow-VAE. Repeating the procedure keeps the selected descriptor fixed while allowing the others to vary according to the fitted correlations.

To assess the empirical support of these completed vectors and subsequently perturbed inputs, we evaluate their joint log-density:

$$s(\mathbf{y}) = \log p_{\mathrm{GMM}}(\mathbf{y}) = \log \left[\sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right]. \tag{17}$$

The score accounts for the full parameter combination rather than its individual ranges. Section 4.2.2 uses its empirical quantiles to examine generation errors in low-density regions. This density reflects the selected sample and its measurement procedure; low density indicates limited support under the fitted model, not physical impossibility.

4. Results

We apply the Flow-VAE to extract latent variables from the r-band DECaLS galaxy images and map them to physical parameters. Following Xu et al. [23], we adopt a 40-dimensional latent space, which optimally balances reconstruction fidelity, computational efficiency, and the training stability of the normalizing flow.

4.1. Model Reconstruction Fidelity

We evaluate reconstruction on the 24,000-image test set to assess how well the VAE preserves image structure and morphological statistics. Each image is encoded, passed through the VAE latent transformation, and decoded. The reconstructed images achieve a mean structural similarity index (SSIM, [30]) of 0.892, with values closer to 1 indicating greater similarity between the original and reconstructed images. To assess morphological fidelity, we also remeasure C, A, S, G, and M_{20} using the same procedure applied to the original images. Figure 4 compares these measurements with their original values, with the identity line indicating exact agreement. For each bin of the original descriptor values, the figure also shows the median reconstructed value and the 16th–84th percentile interval, which corresponds to approximately $\pm 1\sigma$ for a Gaussian distribution. These binned summaries reveal both systematic offsets from the identity relation and the scatter in the reconstructed measurements. No additional sky background or noise is added to the generated images before the morphological descriptors are measured.

We apply the same measurement procedure to original, reconstructed, and generated images, using identical SExtractor configurations and statmorph parameter settings. Because synthesized images may differ from the originals in both morphology and contaminating sources, we construct a new segmentation mask for each image rather than reuse the original mask. The detection containing the central pixel is treated as the target, while all other detections are masked before the morphological descriptors are measured.

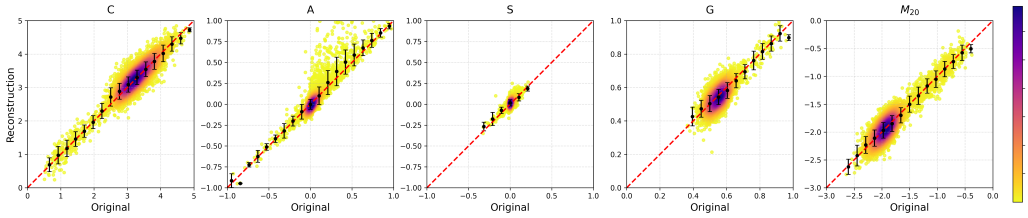


Figure 4. Comparison of non-parametric morphological statistics measured from original and VAE-reconstructed galaxy images.

To compare morphological fidelity across reconstruction and the subsequent generation experiments, we define the signed normalized residual for image i and descriptor j as

$$\delta_{ij} = \frac{\widehat{y}_{ij} - y_{ij}^{\mathrm{ref}}}{R_j}, \tag{18}$$

where $\widehat{y}_{ij}$ is the descriptor measured from the output image and R_j is the fixed range width (see Section 2.2 in detail) adopted for descriptor j throughout the analysis. Normalization by R_j enables comparison across descriptors without dividing by individual values that may be close to zero. Here, y_{ij}^{ref} is the descriptor value provided as input to the model, and it is measured from the original image, including the modified value in the perturbation experiment. SSIM and the descriptor recovery rates are calculated separately: SSIM is computed from the original and reconstructed image pairs, whereas the recovery rates quantify agreement between morphological descriptor values.

We define successful recovery as $|\delta_{ij}| \leq 0.10$, corresponding to an absolute deviation within 10% of the adopted descriptor range. Calculated over all valid reconstructed images, the recovery rates are 96.62%, 97.14%, 99.97%, 97.22%, and 96.96% for C, A, S, G, and M_{20}, respectively. Thus, each descriptor is recovered within this tolerance in more than 96% of valid reconstructions. Table 3 also reports recovery rates at the stricter 5% tolerance.

Table 3. Morphological descriptor recovery rates for reconstruction, single-parameter generation, and a constant-median baseline. R_j denotes the adopted descriptor range width, and δ_{ij} is defined in Equation (18). For each descriptor, the baseline always predicts its median in the test set, independently of the input value.

Descriptor	R_j	**Reconstruction**		**Generation**		**Null Model (Median)**	
		$\lvert\delta\rvert \leq 0.05$	$\lvert\delta\rvert \leq 0.10$	$\lvert\delta\rvert \leq 0.05$	$\lvert\delta\rvert \leq 0.10$	$\lvert\delta\rvert \leq 0.05$	$\lvert\delta\rvert \leq 0.10$
C	4.00	75.2%	96.62%	61.6%	90.95%	39.37%	69.09%
A	1.70	91.3%	97.14%	72.1%	94.92%	65.04%	88.74%
S	0.45	98.5%	99.97%	88.7%	99.83%	77.39%	87.49%
G	0.55	72.7%	97.22%	56.3%	87.81%	47.11%	78.58%
M_{20}	2.00	74.2%	96.96%	56.2%	88.79%	48.43%	75.74%

4.2. Model Generation Capability

As illustrated in Figure 3, Flow-VAE generates galaxy images from a complete condition vector $\mathbf{y} = (C, A, S, G, M_{20})$ using the procedure described in Appendix A. Figure 5 shows each observed galaxy alongside five images generated from its measured morphology statistics. The five descriptors do not uniquely determine the light distribution, as illustrated by the different major-axis orientations in the bottom row. We therefore assess conditional fidelity by comparing the requested descriptor values with those measured from the generated images.

We evaluate conditional generation fidelity from two complementary perspectives. First, Section 4.2.1 quantifies descriptor-specific control by fixing one of the five statistics and sampling the remaining four from their GMM conditional distribution. This experiment measures how accurately the model recovers each prescribed descriptor while retaining statistically supported combinations of the other morphological properties. Section 4.2.2 evaluates complete condition vectors and compares their generation errors across different GMM likelihoods. The GMM likelihood measures training-distribution support, allowing us to test whether joint conditional fidelity decreases for less probable parameter combinations.



Figure 5. Observed test galaxies and Flow-VAE images generated using the five morphological descriptors (C, A, S, G, and M_{20}).

4.2.1. Single-Parameter Conditional Generation

We first perform a single-parameter conditional-generation experiment. For a target descriptor y_j spanning its test-set distribution, the specified value is held fixed while the remaining four descriptors, denoted by $\mathbf{y}_{-j}$, are sampled from the GMM conditional distribution in Equation (16). The completed five-dimensional vectors are then supplied to Flow-VAE to generate an ensemble of galaxy images for each target value. Single-parameter generation is a special case of the more general partial-parameter procedure, which supports any specified subset of the five descriptors and is described in Appendix B.

To assess single-parameter generation fidelity, we remeasure the specified descriptor from each synthesized image and compare it with its requested value (Figure 6). Table 3 reports recovery rates at both the 5% and 10% tolerances, together with the constant-median baseline. Reconstruction and generation exceed this baseline for all five descriptors at both tolerances. For S, however, the baseline alone achieves recovery rates of 77.39% and 87.49% at the 5% and 10% tolerances, respectively, compared with 88.7% and 99.83% for conditional generation. These baseline rates show that the concentration of descriptor values can produce substantial recovery without input-dependent predictions. The high recovery rate of S alone should therefore not be interpreted as evidence of stronger conditional control than for the other descriptors.

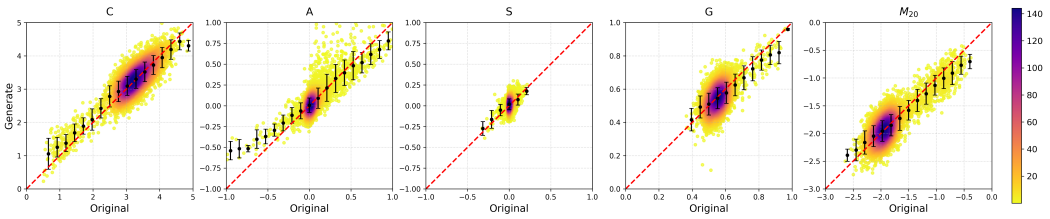


Figure 6. Comparison of non-parametric morphological Statistics between observed and single-parameter generated galaxies.

The binned input–output relations show regression toward the center of the training distribution: low requested values tend to be overestimated, whereas high values tend to be underestimated, particularly for C, A, and G. This compression qualitatively resembles the attenuation of extreme predictions discussed by Ting [31]; however, the present experiment

does not establish whether measurement-error-induced attenuation is responsible for the trend.

4.2.2. Joint-Parameter Conditional Generation

Here, we focus on complete condition vectors, $\mathbf{y} = (C, A, S, G, M_{20})$, and examine whether their generation fidelity depends on their representation in the joint training distribution. Because the five descriptors are correlated, a condition vector can lie within the adopted range of every individual descriptor while remaining rare as a five-dimensional combination. We quantify this training-set support using the GMM log-density score ($s = \log p(\mathbf{y})$).

Because the distribution of s is not generally Gaussian, we convert its empirical percentile rank into a Gaussian-equivalent scale:

$$n_\sigma(s) = -\Phi^{-1}\left[\widehat{F}_s(s)\right], \tag{19}$$

where $\widehat{F}_s(s)$ is the empirical cumulative distribution function of the training-set scores and Φ^{-1} is the inverse standard normal cumulative distribution function. Larger n_σ values therefore indicate condition vectors with lower GMM density and weaker representation in the training sample. In particular, $n_\sigma > 3$ corresponds to the lowest approximately 0.135% of the training-score distribution. We use this threshold as a descriptive reference for the low-density tail.

To probe the model under progressively less supported conditions, we construct a bounded perturbation sample from the test set. For each condition vector, we randomly select one descriptor and perturb it by up to 10% of its adopted range, while keeping the other four descriptors fixed. The perturbed value is clipped to the corresponding training-set range, ensuring that every descriptor remains marginally valid. Nevertheless, changing one descriptor independently can weaken its consistency with the other four and move the complete condition vector into a lower-density region of the joint distribution.

Figure 7 shows that the residual distributions generally broaden as the condition vectors move toward larger n_σ. Several descriptors also develop systematic offsets in the low-density regime, indicating that the generated measurements are increasingly drawn toward combinations better represented in the training sample. Conditional fidelity therefore depends on the joint support of the complete input vector, rather than only on whether each descriptor lies within its individual training range.

These results show that the model can accept bounded combinations of all five descriptors, but its conditional fidelity is highest for combinations well represented in the training data and decreases toward lower-density regions. The GMM density score therefore provides a useful measure of training-set support for a requested condition, but it does not represent the probability of successful generation or determine whether the corresponding morphology is physically feasible. Even in the low-density regime ($n_\sigma > 3$), 22.5% of valid generated images recover all five requested descriptors within 10% of their adopted ranges. This indicates some capacity to generalize to sparsely represented combinations within the tested perturbation setting, despite the decline in conditional fidelity.

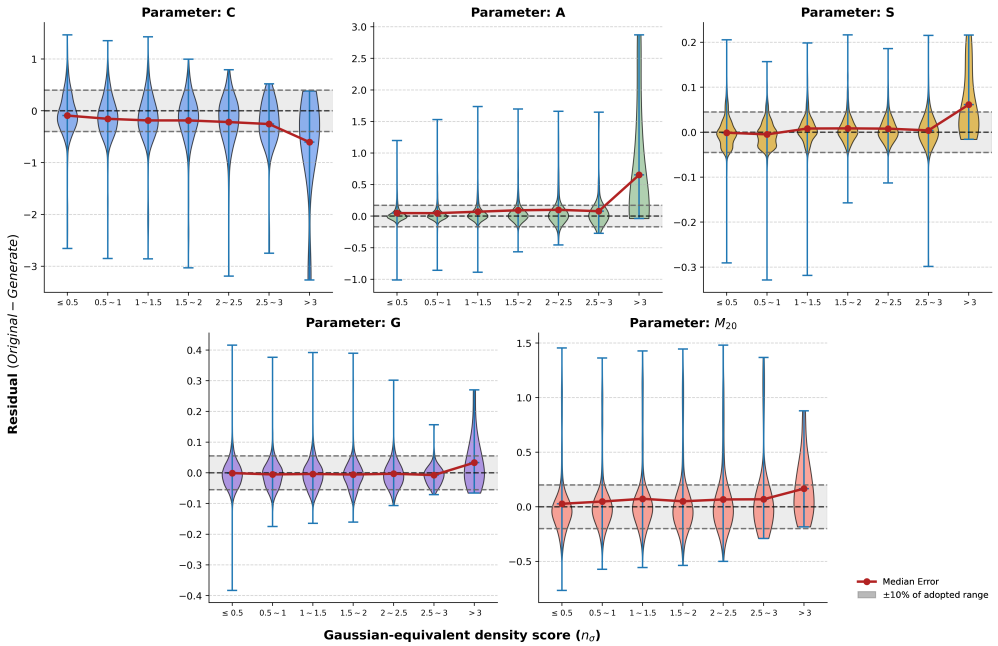


Figure 7. Residual distributions of the generated morphological statistics as a function of GMM input-density score. The violin shapes represent error densities within discrete 0.5σ bins, while the connected line traces the median residual in each bin.

5. Discussion

Flow-VAE extends previous galaxy-generation methods conditioned mainly on magnitude, size, and Sérsic index by conditioning image generation on non-parametric morphological descriptors [15,16]. Under the same recovery criterion, single-parameter generation yields descriptor recovery rates 0.14–9.41 percentage points below those of reconstruction. This difference shows that preserving morphology during reconstruction does not guarantee equally accurate conditional generation. Smoothness has the highest generation recovery rate and the smallest reduction from reconstruction, while G and M_{20} have the lowest recovery rates.

The lower fidelity of G and M_{20} may reflect their sensitivity to pixel-level flux variations and, for M_{20}, the spatial distribution of bright regions, which may not be fully captured by the adopted 40-dimensional latent space. Related work shows that high reconstruction fidelity does not necessarily imply readily accessible physical information in the learned representation [32]. Increasing the latent dimensionality could improve morphological fidelity but may also complicate conditional density modeling with normalizing flows. A systematic investigation of this trade-off is beyond the scope of this work.

To complement the visual comparison in Section 4.2, we examined how closely the conditionally generated sample reproduces the observed distributions of three structural properties beyond the five conditioning descriptors (Figure 8). We applied the same measurement procedure to both samples without PSF correction. We defined R_{50} as the circular radius enclosing half the flux measured within 1.5 times the circular Petrosian radius. A single Sérsic fit provided the effective index n, while a fit with a Sérsic bulge and an exponential disc provided B/T, the fraction of total model light assigned to the bulge. The marginal distributions of R_{50}, n, and B/T show close agreement, with Kolmogorov–Smirnov distances of 0.023, 0.022, and 0.038, respectively. Both samples peak in apparent galaxy size at $R_{50} \simeq 6$–7 pixels, with fitted Sérsic indices just above unity, indicating similar characteristic sizes and predominantly near-exponential light profiles at the observed resolution. The largest of the three distances occurs for B/T, whose distribution shows a slightly sharper peak near 0.27 in the observed sample.

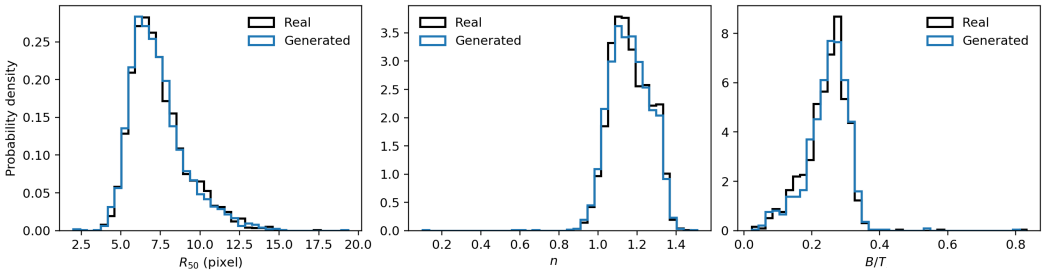


Figure 8. Distributions of the circular half-light radius R_{50} (**left**), effective single-Sérsic index n (**middle**), and model-based bulge-to-total light ratio B/T (**right**) for real (black) and generated (blue) images.

Flow-VAE is trained on observed r-band images from the DESI Legacy Imaging Surveys, with the five conditioning descriptors measured from those same images. The descriptor recovery rates therefore quantify agreement with morphology at the resolution of the observations. Effective image resolution also affects the interpretation of the additional structural measurements, particularly how well the central and outer light components can be distinguished when estimating B/T. The fixed 64×64 field of view also limits the applicability to extended nearby galaxies, for which the cutout may truncate the outer light distribution or the Petrosian measurement aperture and leave insufficient source-free area for the background corrections to A and S. For small, weakly resolved galaxies, the measured A, S, and M_{20} remain sensitive to PSF smoothing, whose physical scale varies with redshift.

The masked MAE reconstruction loss measures absolute pixel differences over unmasked pixels, without the inverse-variance weighting used in a χ^2 objective [33]. It does not explicitly account for spatial variations in noise uncertainty, so the reconstruction scores alone do not establish how accurately the model recovers the underlying noise-free signal. The training sample is dominated by low-redshift ($z < 0.1$) galaxies; applications to other surveys, bands, or redshift ranges therefore need validation with representative data, particularly where spatial resolution and noise properties differ from those of the training sample. Nevertheless, the framework is expected to be transferable to these other surveys, subject to appropriate retraining and validation on representative data.

6. Conclusions

We present Flow-VAE, a framework for generating galaxy images conditioned on five non-parametric morphological statistics: C, A, S, G, and M_{20}. The model combines a variational autoencoder with a conditional normalizing flow, while a GMM represents the joint distribution of the conditioning variables. This design allows any subset of the five descriptors to be prescribed, with the remaining descriptors sampled conditionally to avoid unsupported independent combinations.

The main results are as follows:

- **Model Reconstruction Fidelity:** The VAE achieves a mean SSIM of 0.892 on the 24,000-image test set. Using $|\delta_{ij}| \leq 0.10$ as the recovery criterion, the reconstruction success rates are 96.62% for C , 97.14% for A, 99.97% for S, 97.22% for G, and 96.96% for M_{20}.
- **Single-parameter morphological control:** Conditional GMM sampling enables generation with one prescribed descriptor while accounting for the empirical relationships among the remaining four. The generated measurements broadly follow the requested values, although systematic departures remain toward the ends of the sampled ranges.
- **Training-distribution support:** The joint GMM density provides an empirical measure of generation reliability across the five-dimensional morphological parameter space.

The Gaussian-equivalent 3σ boundary marks the input regions where conditional performance has been evaluated.

The principal limitations are systematic regression toward common descriptor values and reduced conditional accuracy for sparsely represented parameter combinations. These effects should be considered when generating samples with rare or extreme morphological statistics. Within well-supported regions of the training distribution, Flow-VAE provides a practical method for generating galaxy samples with controlled morphological descriptors without imposing analytical surface-brightness profiles. However, agreement in these descriptors does not by itself establish visual realism or the recovery of specific morphological structures. Subject to validation for the intended application, the generated samples could support survey-pipeline tests, selection-effect studies, and assessments of morphology-dependent biases in surveys such as Euclid and CSST.

The current conditioning scheme does not directly specify morphological classes or individual structures such as bars, spiral arms, rings, or tidal features. Agreement in the five descriptors alone does not establish visual realism, and the conditional distributions of size, axis ratio, and Sérsic index remain unvalidated.

Author Contributions: Conceptualization, Q.X., N.L. and S.S.; methodology, Q.X.; software, Q.X.; validation, Q.X.; formal R.Y., M.C., S.C., Q.X. and S.S.; investigation, Q.X.; data curation, R.Y.; writing—original draft preparation, Q.X. and S.S.; writing—review and editing, Q.X. and S.S.; visualization, Q.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (Nos.2025YFF0510603), the National Natural Science Foundation of China (Nos.12141302), and the China Manned Space Project (CMS-CSST-2025-A07).

Data Availability Statement: This work uses the galaxy image (DECaLS) data underlying this article are available at https://www.legacysurvey.org/ (accessed on 6 November 2026). Our code is publicly available: https://github.com/xuquanfeng/Galaxy_Flow_VAE (accessed on 6 November 2026).

Acknowledgments: The DESI Legacy Imaging Surveys consist of three complementary projects: the Dark Energy Camera Legacy Survey (DECaLS), Beijing–Arizona Sky Survey (BASS), and Mayall z-band Legacy Survey (MzLS). DECaLS, BASS, and MzLS include data obtained, respectively, with the Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO), operated by NSF's National Optical-Infrared Astronomy Research Laboratory (NOIRLab); the Bok telescope at Steward Observatory; and the Mayall telescope at Kitt Peak National Observatory (KPNO), operated by NOIRLab. Pipeline processing was supported by NOIRLab and Lawrence Berkeley National Laboratory (LBNL). This project used data obtained with the Dark Energy Camera (DECam), constructed by the Dark Energy Survey (DES) collaboration; funding for DES was provided by the U.S. Department of Energy (DOE), the National Science Foundation (NSF), and numerous international agencies (full list at https://www.darkenergysurvey.org). BASS is a key project of the Telescope Access Program (TAP), funded by the National Astronomical Observatories, Chinese Academy of Sciences (NAOC), and the Chinese Academy of Sciences. The Legacy Surveys team thanks the Tohono O'odham Nation for access to Iolkam Du'ag (Kitt Peak). NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with NSF; LBNL is managed by the Regents of the University of California under contract to the U.S. Department of Energy. The Legacy Surveys imaging of the DESI footprint is supported by DOE Contract No. DE-AC02-05CH1123 and NSF Contract No. AST-0950945.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Generation from Complete Morphological Conditions

This appendix describes conditional galaxy generation when all five non-parametric morphological statistics are specified. The user supplies a complete condition vector

$\mathbf{y} = (C, A, S, G, M_{20})$. Before generation, the pre-fitted GMM evaluates the joint density of the requested vector, providing an empirical indication of how well the condition is represented in the training sample. This density score characterizes statistical support and should not be interpreted as a probability of successful generation or as a criterion of physical feasibility.

Given the complete condition vector, a base-space latent variable $\mathbf{u}$ is sampled from the conditional flow distribution and transformed through the inverse flow as

$$\mathbf{z} = f_\theta^{-1}(\mathbf{u}; \mathbf{y}). \tag{A1}$$

The resulting latent vector $\mathbf{z}$ is passed to the VAE decoder to generate a galaxy image. Repeated sampling at the same $\mathbf{y}$ produces different realizations associated with the same requested morphological conditions. Because the five statistics do not uniquely specify an image, these realizations are not expected to reproduce a particular observed galaxy at the pixel level. Conditional fidelity should instead be evaluated by remeasuring the five statistics from the generated images and comparing them with their requested values.

```
import joblib
import numpy as np
import GMMPercentileScorer
import torch
import Flow_VAE
state = joblib.load('gmm_state.pkl')
scorer = GMMPercentileScorer(n_components=500)
scorer.load_state(state)
y = np.array([3.05, 0.02, 0.01, 0.52, -1.65])
res = scorer.score_point(y)
print("percentile %:", res['percentile_pct'])
# percentile %: 81.2210939945057
model = Flow_VAE()
model.load_state_dict(torch.load('Flow_VAE.pth'))
u = model.sample_base(condition=y)
z = model.flow_inverse(u,condition=y)
imgs = model.decode(z)
```

Appendix B. Generation from Partially Specified Morphological Conditions

This appendix describes conditional generation when only a subset of the five morphological statistics is specified. Let $\mathbf{y}_\mathcal{K}$ denote the specified descriptors and $\mathbf{y}_\mathcal{U}$ the unspecified descriptors. The pre-fitted GMM is used to derive the conditional distribution

$$p_{\mathrm{GMM}}(\mathbf{y}_\mathcal{U} \mid \mathbf{y}_\mathcal{K}), \tag{A2}$$

from which the missing values are sampled. Combining the specified and sampled values produces a complete five-dimensional condition vector that preserves the correlations represented in the training population.

The completed vector is subsequently passed to Flow-VAE using the complete-condition generation procedure described in Appendix A. Specifying one descriptor and sampling the remaining four corresponds to the single-parameter experiment in Section 4.2.1; the same procedure also supports any other subset of the five descriptors.

```
import cal_gmm
import sample_gmm
# Input
know_idx = [0, 2] # Assuming known x0, x2
know = [3.22, -0.06]
num_missing = 3

gmm = state['gmm']
miss_idx = [i for i in range(5) if i not in known_idx]
w, means, covs = cal_gmm(gmm, know, know_idx, miss_idx)
samples = sample_gmm(w, means, covs, n_samples=1)
final = np.zeros(5)
final[know_idx] = x_known
final[missing_idx] = samples
print(', '.join([f'{final[i]}' for i in range(5)]))
# 3.22, -0.28, -0.06, 0.53, -1.72
```

References

1. Robertson, B.E.; Banerji, M.; Brough, S.; Davies, R.L.; Ferguson, H.C.; Hausen, R.; Kaviraj, S.; Newman, J.A.; Schmidt, S.J.; Tyson, J.A.; et al. Galaxy formation and evolution science in the era of the Large Synoptic Survey Telescope. *Nat. Rev. Phys.* **2019**, *1*, 450–462. [CrossRef]
2. Hubble, E.P. Extragalactic nebulae. *Astrophys. J.* **1926**, *64*, 321–369. [CrossRef] [PubMed]
3. Abraham, R.G.; van den Bergh, S. The Morphological Evolution of Galaxies. *Science* **2001**, *293*, 1273–1278. [CrossRef] [PubMed]
4. Peng, C.Y.; Ho, L.C.; Impey, C.D.; Rix, H.W. Detailed Decomposition of Galaxy Images. II. Beyond Axisymmetric Models. *Astron. J.* **2010**, *139*, 2097. [CrossRef]
5. Robotham, A.S.G.; Taranu, D.S.; Tobar, R.; Moffett, A.; Driver, S.P. ProFit: Bayesian profile fitting of galaxy images. *Mon. Not. R. Astron. Soc.* **2017**, *466*, 1513–1541. [CrossRef]
6. Chen, M.; S. de Souza, R.; Xu, Q.; Shen, S.; Chies-Santos, A.L.; Ye, R.; Canossa-Gosteinski, M.A.; Cong, Y. Galmoss: A package for GPU-accelerated galaxy profile fitting. *Astron. Comput.* **2024**, *47*, 100825. [CrossRef]
7. Andrae, R.; Jahnke, K.; Melchior, P. Parametrizing arbitrary galaxy morphologies: Potentials and pitfalls. *Mon. Not. R. Astron. Soc.* **2011**, *411*, 385–401. [CrossRef]
8. Conselice, C.J. The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories. *Astrophys. J. Suppl. Ser.* **2003**, *147*, 1. [CrossRef] [PubMed]
9. Freeman, P.E.; Izbicki, R.; Lee, A.B.; Newman, J.A.; Conselice, C.J.; Koekemoer, A.M.; Lotz, J.M.; Mozena, M. New image statistics for detecting disturbed galaxy morphologies at high redshift. *Mon. Not. R. Astron. Soc.* **2013**, *434*, 282–295. [CrossRef]
10. Rodriguez-Gomez, V.; Snyder, G.F.; Lotz, J.M.; Nelson, D.; Pillepich, A.; Springel, V.; Genel, S.; Weinberger, R.; Tacchella, S.; Pakmor, R.; et al. The optical morphologies of galaxies in the IllustrisTNG simulation: A comparison to Pan-STARRS observations. *Mon. Not. R. Astron. Soc.* **2018**, *483*, 4140–4159. [CrossRef]
11. Wuyts, S.; Förster Schreiber, N.M.; van der Wel, A.; Magnelli, B.; Guo, Y.; Genzel, R.; Lutz, D.; Aussel, H.; Barro, G.; Berta, S.; et al. Galaxy Structure And Mode Of Star Formation In The Sfr–Mass Plane From z 2.5 To z 0.1. *Astrophys. J.* **2011**, *742*, 96. [CrossRef]
12. Joshi, H.; Kumar, B.; Singh, K.P.; Sagar, R.; Sharma, S.; Pandey, J.C. Multiwavelength study of a young open cluster NGC 7419. *Mon. Not. R. Astron. Soc.* **2008**, *391*, 1279–1300. [CrossRef]
13. Conselice, C.J. The evolution of galaxy structure over cosmic time. *Annu. Rev. Astron. Astrophys.* **2014**, *52*, 291–337. [CrossRef]
14. Rowe, B.T.P.; Jarvis, M.; Mandelbaum, R.; Bernstein, G.M.; Bosch, J.; Simet, M.; Meyers, J.E.; Kacprzak, T.; Nakajima, R.; Zuntz, J.; et al. GALSIM: The modular galaxy image simulation toolkit. *Astron. Comput.* **2015**, *10*, 121–150. [CrossRef]
15. Lanusse, F.; Mandelbaum, R.; Ravanbakhsh, S.; Li, C.L.; Freeman, P.; Póczos, B. Deep generative models for galaxy image simulations. *Mon. Not. R. Astron. Soc.* **2021**, *504*, 5543–5555. [CrossRef]
16. Euclid Collaboration; Bretonnière, H.; Huertas-Company, M.; Boucaud, A.; Lanusse, F.; Jullo, E.; Merlin, E.; Tuccillo, D.; Castellano, M.; Brinchmann, J.; et al. Euclid preparation—XIII. Forecasts for galaxy morphology with the Euclid Survey using deep generative models. *A&A* **2022**, *657*, A90. [CrossRef]
17. He, Q.; Nightingale, J.W.; Amvrosiadis, A.; Robertson, A.; Cole, S.; Frenk, C.S.; Massey, R.; Li, R.; Cao, X.; Lange, S.C.; et al. Unveiling lens light complexity with a novel multi-Gaussian expansion approach for strong gravitational lensing. *Mon. Not. R. Astron. Soc.* **2024**, *532*, 2441–2462. [CrossRef]

18. Ahn, C.P.; Alexandroff, R.; Allende Prieto, C.; Anderson, S.F.; Anderton, T.; Andrews, B.H.; Aubourg, É.; Bailey, S.; Balbinot, E.; Barnes, R.; et al. The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *Astrophys. J. Suppl. Ser.* **2012**, *203*, 21. [CrossRef]
19. Dey, A.; Schlegel, D.J.; Lang, D.; Blum, R.; Burleigh, K. Overview of the DESI Legacy Imaging Surveys. *Astron. J.* **2019**, *157*, 168. [CrossRef]
20. Bertin, E.; Arnouts, S. SExtractor: Software for source extraction. *Astron. Astrophys. Suppl. Ser.* **1996**, *117*, 393–404. [CrossRef]
21. Lotz, J.M.; Primack, J.; Madau, P. A New Nonparametric Approach to Galaxy Morphological Classification. *Astron. J.* **2004**, *128*, 163. [CrossRef] [PubMed]
22. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
23. Xu, Q.; Shen, S.; de Souza, R.S.; Chen, M.; Ye, R.; She, Y.; Chen, Z.; Ishida, E.E.O.; Krone-Martins, A.; Durgesh, R. From images to features: Unbiased morphology classification via variational auto-encoders and domain adaptation. *Mon. Not. R. Astron. Soc.* **2023**, *526*, 6391–6400. [CrossRef]
24. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
25. Dinh, L.; Krueger, D.; Bengio, Y. NICE: Non-linear Independent Components Estimation. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
26. Durkan, C.; Bekasov, A.; Murray, I.; Papamakarios, G. Neural Spline Flows. In *Proceedings of the NeurIPS*; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2019; pp. 7509–7520.
27. Papamakarios, G.; Pavlakou, T.; Murray, I. Masked Autoregressive Flow for Density Estimation. *arXiv* **2018**, arXiv:1705.07057v4. [CrossRef]
28. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
29. Reynolds, D. *Gaussian Mixture Models*; Springer: Boston, MA, USA, 2009; Volume 741, p. 3.
30. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
31. Ting, Y.S. Why Machine Learning Models Systematically Underestimate Extreme Values. *Open J. Astrophys.* **2025**, *8*, 95. [CrossRef]
32. Sanjaripour, S.; Smith, M.J.; Pérez-Carrasco, M.; Martínez-Galarza, J.R.; Mobasher, B.; Canalizo, G. The Galaxy's Guide to the Tokenizer: A Benchmark for Scientific Foundation Models. *arXiv* **2026**, arXiv:2606.25610v1. [CrossRef]
33. Ye, R.; Shen, S.; Xu, Q. From Observation to Ground Truth: Reconstruction Losses of Galaxy Imaging Under Heteroscedastic Noise. *Universe* **2026**, *12*, 285. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.