Realistic galaxy image generation in color using Variational AutoEncoders

Thomas Sainrat, Marc Huertas-Company, Hubert Bretonnière
Alexandre Boucaud, Cécile Roucelle, Eric Aubourg, Bastien Arcelin

Motivations

Weak lensing (measuring the deformation of galaxies due to gravitational lensing by foreground objects) is a crucial probe for cosmology : by doing a statistical treatment, one can recover maps of the Universe and estimate cosmological parameters ; this is one of the goals of the upcoming **Euclid mission,** which was the main target for this work.

However, this requires very accurate shape measurement and handling various sources of bias ; hence the need for realistic simulations. Simulating **color** is of particular interest for tackling the problem of **blending** and for dealing with redshifts.

Proposed solution

We expand on the work of Lanusse et al. (see Related work) to generate color images. We use the **CANDELS dataset** as a source of multiband images.

This required dealing with different observing conditions in the different filters, and being able to achieve competitive performances while using a smaller dataset than the original paper. We explored **data augmentation** techniques, as well as **domain transfer**.

Dataset Description

The CANDELS dataset is a multiband survey realized with the Hubble Space Telescope, with 4 to 7 different filters on 5 different fields. We use the catalog introduced by Dinauro et al. (2018), which contains approximately 17600 galaxies. We chose to use the F160W (infrared) and F606W (visible) filters to represent the two instruments from the Euclid mission.

We also performed some cleaning on the catalog, including filtering out blended images and clearing parasite galaxies in the images.

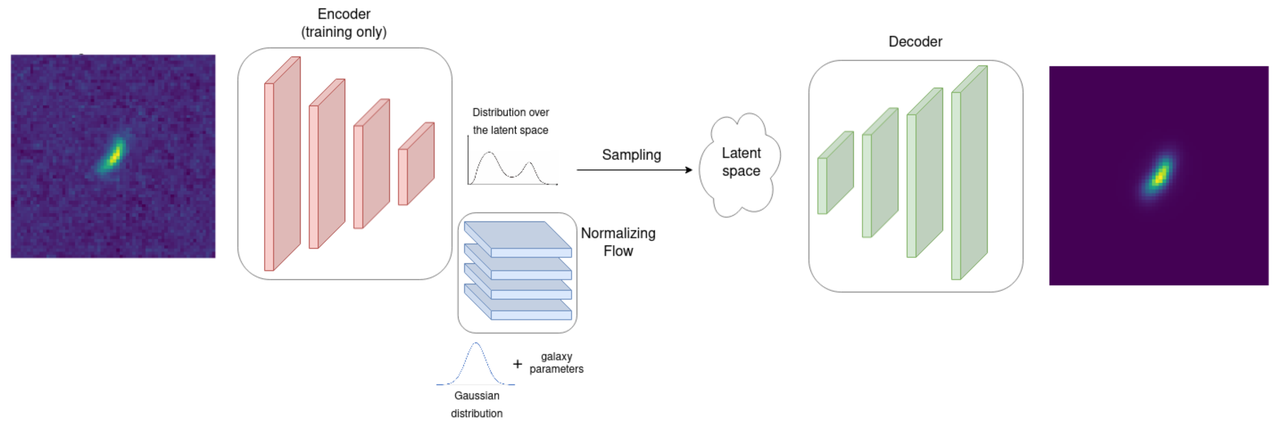


Figure 1 : Architecture of the Flow-VAE model

Related Work

A recent approach was proposed in Lanusse et al. (2021), making use of a **Variational AutoEncoder** (VAE) and a **Normalizing Flow**. The VAE is first trained to reproduce galaxies by encoding them as a distribution on the latent space, sampling from the distribution and decoding the result. The Flow is then trained to map a Gaussian distribution to a distribution of the latent space corresponding to a galaxy with specific input parameters (ellipticity, size, magnitude…). Putting the two together allows to generate new galaxies with the specified parameters.

Results

We successfully adapted the architecture of Lanusse et al. to generate multiband images. We achieve satisfying results on the different parameters, especially color which is well reproduced. The performance on ellipticity and size is not as good as the original article, although this was to be expected with a smaller dataset.

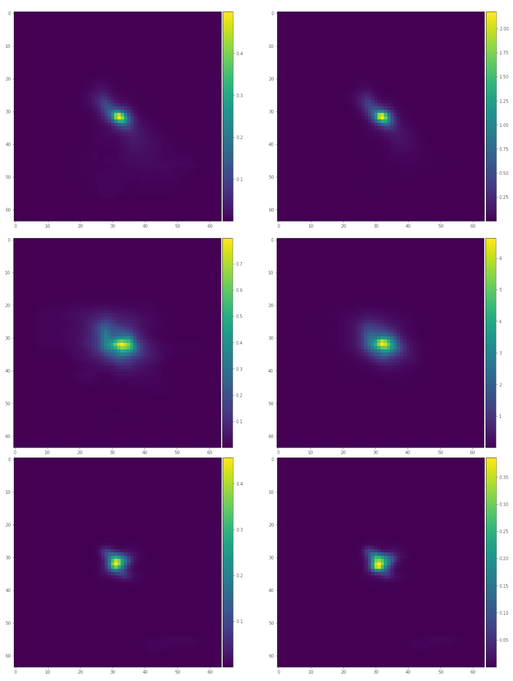


Figure 2 : Examples of generated galaxies (left : visible, right : infrared)

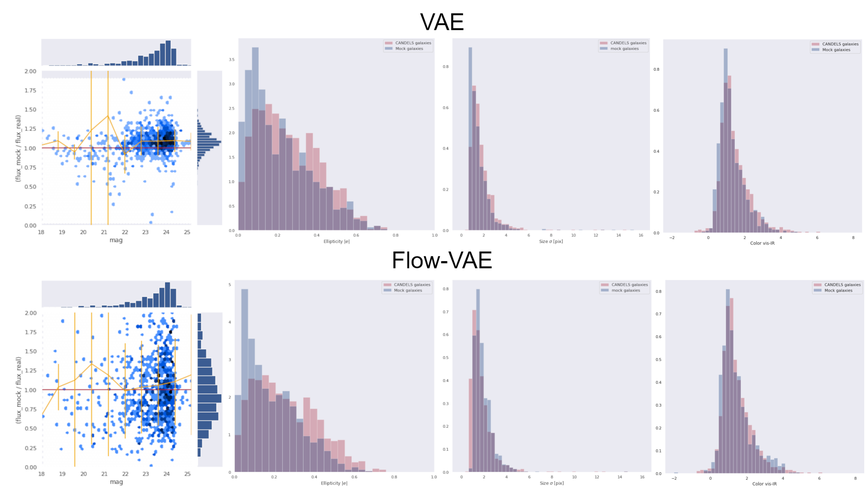


Figure 3 : Distributions of the target parameters : flux, ellipticity, size, color (red : true distribution, blue : distribution in the generated galaxies)

References

- Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., Li, C.L., Freeman, P., & Póczos, B. (2021). Deep generative models for galaxy image simulations. *Monthly Notices of the Royal Astronomical Society, 504(4), 5543-5555.*
- Dimauro, P., Huertas-Company, M., Daddi, E., Pérez-González, P., Bernardi et al. (2018). A catalog of polychromatic bulge-disc decompositions of ~17.600 galaxies in CANDELS. *Monthly Notices of the Royal Astronomical Society, 478(4), 5410-5426.*

This work is supported by diiP, IdEx Université de Paris, ANR-18-IDEX-0001