First steps in simulations and morphology classification of JWST/MIRI deep field galaxies

Martin Topinka*†

Dublin Institute for Advanced Studies
E-mail: `martin.topinka@gmail.com`

The JWST/MIRI mid-infrared deep field imaging is expected to reveal many thousands of high redshifted galaxies. To be ready for the morphological studies of these galaxies a simulation of the deep field is being prepared using the MIRISim simulator. Since running the full astrophysical simulation of the sources would be very computer demanding, a new machine learning approach based on the use of a Variational AutoEncoder (VAE) is tested to generate examples of the early galaxies of different morphologies. The AE is also used to detect morphological anomalies/irregularities in the images of galactic core that may be related to an existence of a black hole binary system and a post merger product. Here I describe the method and show the preliminary results on dummy data.

---

*Speaker.

†Thanks to the JWST/MIRI high-z team.

1. Introduction

The James Webb Space Telescope (JWST) [1, 2], often presented as the next generation Hubble Space Telescope, is being constructed in a collaboration between NASA, ESA and CSA. The main components of the telescope are the primary, 6.6 m in diameter mirror, the spacecraft bus with the instruments and a huge sun shield of a size of 26 m × 14 m. The overall build is shown in the Figure 1. JWST will become the largest space telescope mission ever launched. The mission life time, mostly determined by the fuel for manoeuvring and pointing, is expected to be around 10 years, with the minimum of of 5 years. To this date, the launch of the rocket Ariane 5 with the JWST folded in the cargo top of the rocket has been scheduled for March 2021.
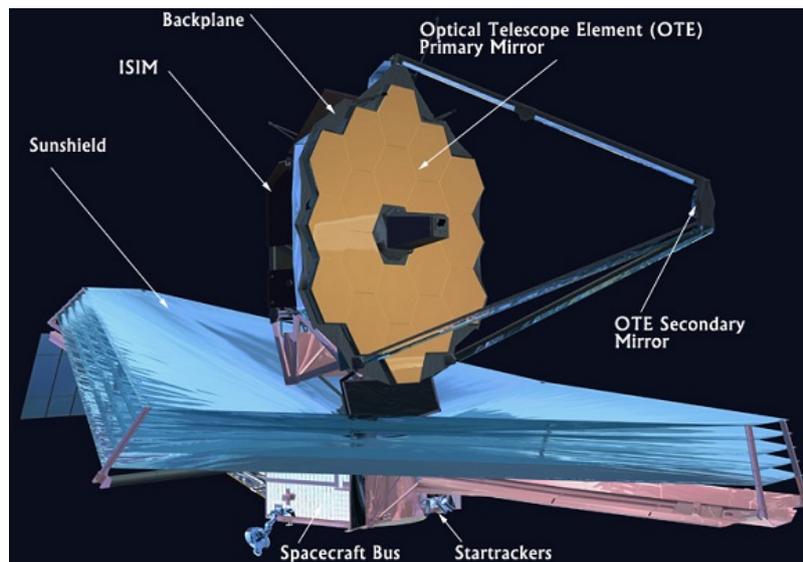

Figure 1: A concept of JWST with the primary mirror, the secondary mirror, the sun shield and the spacecraft bus and star-trackers. Credit: NASA.

2. JWST/MIRI deep field

The JWST instruments cover the mid-infrared wavelength range, e.g. NIRCam 0.6–5 μm, MIRI 5–28 μm [3, 4, 5, 6, 7, 8, 9, 10, 11]. The main scientific goals that the JWST mission is going to explore [12, 13] the origin of the stars and protoplanetary systems; assembly and evolution of galaxies; the end of the dark ages, the first light and the re-ionisation phase of the early Universe; Exoplanets and planetary systems (including our solar system) and the origins of life Being more than one order of magnitude sensitive in some bands than any other observatory, a lot of unpredicted discoveries (the unknown unknown) are expected.

In this work I focus to the key project from the JWST/MIRI high-z group core programme related of the JWST/MIRI deep field. This over 60 hours long deep exposure would shed light on the formation and evolution of galaxies in the high-z universe, possibly answering the key questions about the nature of the first galaxies, how the galaxies are formed and what is the role of the central black holes. The rate and impact of the galaxy mergers would be studied.

3. MIRISim

To plan the MIRI observations and to analyse the incoming deep field data, an advanced end-to-end simulator, the MIRISim software package has been developed [14, 15]. MIRISim allows users to see a simulated datacube as the result from observing the virtual celestial sources predefined by a user. A user can customise the virtual sky view defined by a variety of parameters such as the customised input sky scene, dither patterns, number of integrations, etc. An example of the workflow using MIRISim is shown in the Figure 2. The ability of MIRISim to simulate a field densely covered by galaxies is presented in the Figure 3.
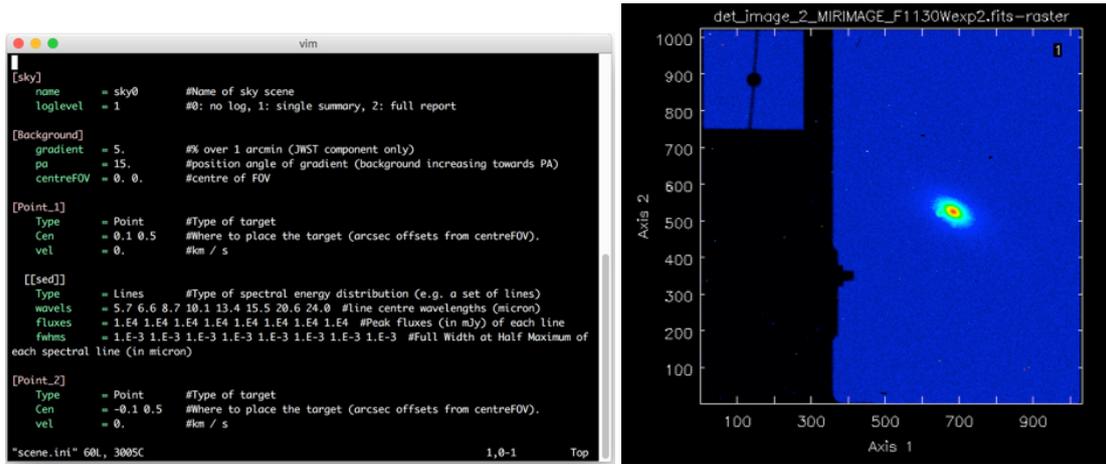

Figure 2: An example of MIRISim in action, starting from a scene description in a text file (left) to a simulated output image on the detector (right). Note, that a small number of the detector pixels are blocked due to mechanical setup of the detector. The left "black" column" is dedicated for a projection of the on-board coronagraphs. Credit: Pamela Klaassen on behalf of the MIRISim team, private communication.

4. Galaxy morphology classification

One part of the extensive study of the early type galaxy formation process and the galaxy evolution is based on extracting information about the high-z galaxy morphology from the deep field image.

JWST is expected to reveal high redshifted galaxies up to the redshift 20 that have never been observed before. The HST record holder of the furthest observed galaxy is GN-z11 with redshift $z \sim 11$ [16]. Therefore, a preparation of a realistic set of high redshifted galaxies is not straightforward. In principle, one can simulate the morphologies of the high-z galaxies using a state-of-the art galaxy simulator, such as Illustris [17], or extrapolate the existing galaxy images to higher redshifts, or a combination of these two. The first is pure astrophysical but very CPU consuming and limited to our models, the second is faster to develop but may create non-astrophysical artefacts.

A general workflow of preparing a realistic simulation of the JWST/MIRI deep field consists of the following steps:

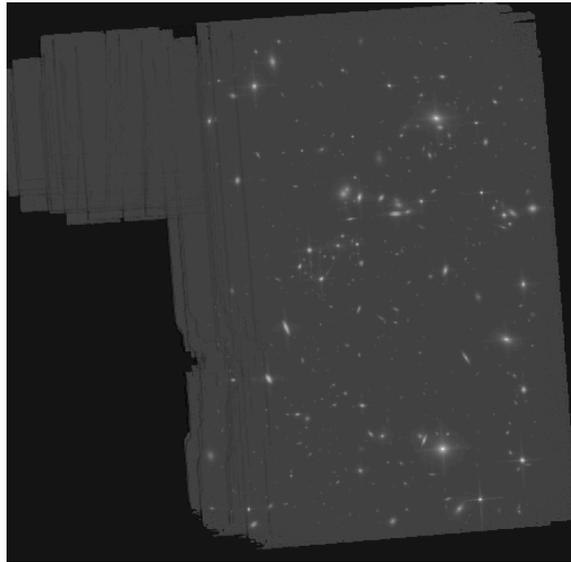- to generate a representative sample of the high-z galaxies

Figure 3: A result of a MIRISim simulation of a deep field exposure that contains 1000 galaxies. Credits: Tuomo Tikkanen.

- to use these galaxies as input for the MIRISim

- to run the script to obtain the simulated JWST/MIRI deep field image

- galaxies identified in the output simulated mage undergo the morphology classification and the morphology statistics is built

Having the full control over the input scene one can determine how sensitive the classification algorithm is to individual galaxy classes and the accuracy of identifying the merging features.

5. Variational Auto-Encoders as a galaxy generator

As a first step in the generation of the artificial galaxy catalog, an auto-encoder (AE) neural network architecture is introduced. The scheme is based i on two sets of layers typical for a convolutional neural network stuck together in a shape of an hour-glass. The input image and the output reconstructed image are set to be identical in the ideal case and the parameters of the network, the values at each node, are calculated to minimise the reconstruction error. This is common for all auto-encoders. The layer in the middle is called the latent space and it is a reduced representation of the input image. Applying the decoder on a vector in the latent space representation of the input image one can obtain an image close to the input image.

The modification that makes the auto-encoder variational happens in the nodes of the bottleneck in which each feature is not a single value but it is composed of a pair of values: the mean value of the feature and its variance. This approach is very useful in generating images that the algorithm has never seen during the training phase. The new set of images is obtained by random sampling from the Gaussian distribution around the mean value μ with the variance σ in each of the features of the latent space. The decoder is applied on the sampled values from the latent space.
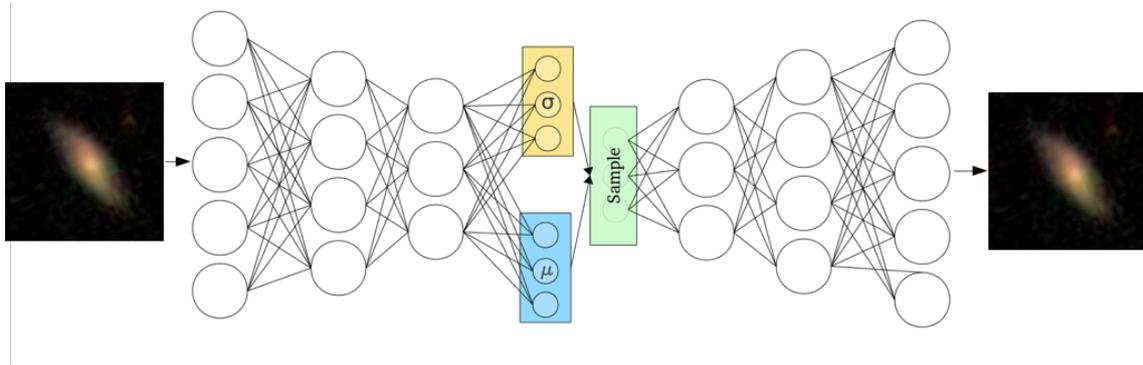
Figure 4: A schematic representation of a neural network architecture based on a variational auto-encoder. The galaxy image is encoded to a bottle-neck, the latent space of a low dimension and decoded back to the original image. The goal is to keep the reconstruction error minimal.

An example of artificial galaxies generated from the existing set of images obtained from the SDSS archive is shown in the Figure 7.
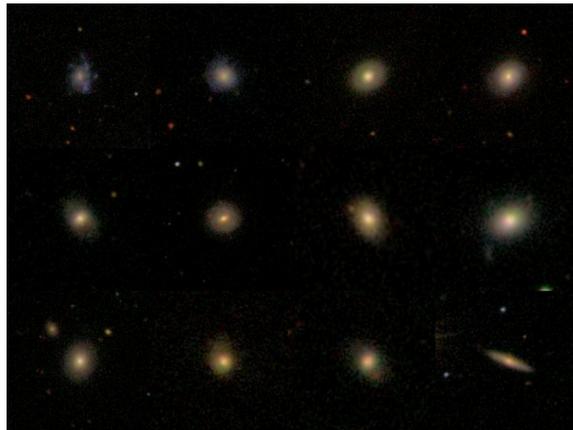

Figure 5: An example of artificial galaxies sampled by VAE from a training dataset based on the SDSS images of galaxies.

An example of the small, 2-dim latent space (while the input images were 32×32 pixels) is shown in the Figure 6 together with colouring the galaxies as a function of true redshift known for those SDSS galaxies. Sampling from the high-redshift end of the latent space ensures that only high redshift galaxies are generated.

Surprisingly, the latent space allows certain arithmetic operations with the latent space vectors, such as adding and subtraction, that makes no sense in the full space of the input images. For example a vector from the latent space trained to represent spiral galaxies added to a vector from the latent space trained to represent galaxies with a bar would lead to a representation of a spiral galaxy with a bar, while if we added the images of the two in a simple way as input images it would create mess.
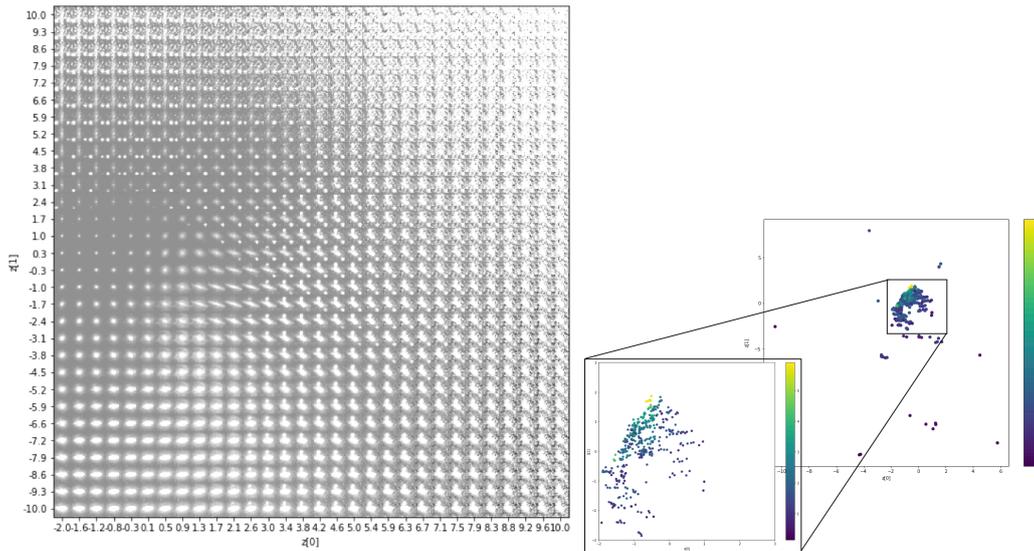
Figure 6: An example of galaxies of different shapes as they were sampled from different regions of the full latent space (left). Representation of the galaxies from the train sample in the latent space including the redshift information. It shows the yellow region of the latent space where the high-redshift galaxies are concentrated (right).

6. Outlier detection

AEs are very powerful in finding non-gaussian outliers and anomaly detection. The network is trained to code and decode images of a certain type of galaxies, e.g. single core galaxies. If a different type of a galaxy, e.g. a merger with two cores is used as an input, the reconstruction error of the AE increases because the AE has been trained to minimise the reconstruction error on single galaxies. The preliminary test of the method is shown in the Figure **??**
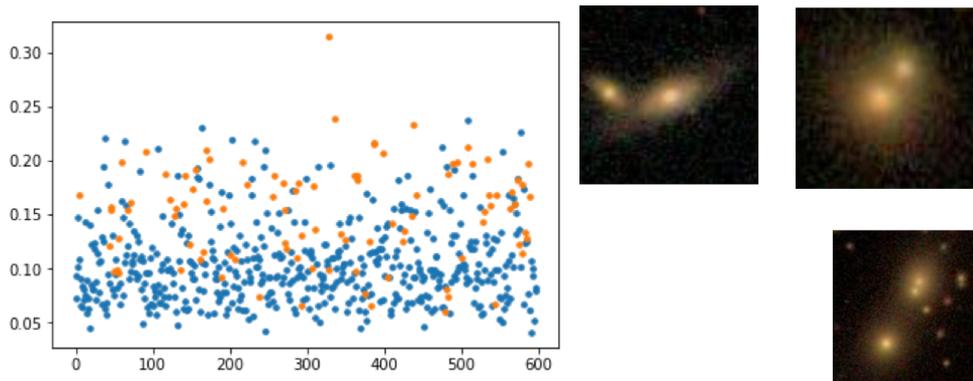

Figure 7: An example of an application of the AE anomaly detection algorithm (left). The AE has been trained using the SDSS images of single (non-merger) galaxies. Then the AE is fed with a galaxy sample that contains a mixture of single and merging galaxies. The reconstruction error for each galaxy is plot: single (blue dots) and merging (orange dots). The mergers tend to have a higher reconstruction error. Three examples of the galaxies with the highest reconstruction error are shown (right).

7. Conclusion

It was demonstrated that AE and VAE are promising tools to generate a wide variety of artificial galaxies not equal to the train images but keeping the main properties of the train sample.

In the next step, galaxies simulated from an analytical transcription in MIRISim, combined with the galaxies generated from existing astrophysical images (e.g. from Spitzer IR telescope) and the combination of both strategies may generate reliable test galaxy catalogs. More advanced models of galaxies, beyond the analytical descriptive capabilities of MIRISim, capturing detailed astrophysical aspects of the early galaxies would require to use products from a more advanced simulations of galaxies in Illustris as a training set for VAE.

References

[1] Clampin, M. 2008, Advances in Space Research, Vol. 41, 1983

[2] https://www.jwst.nasa.gov

[3] https://jwst.stsci.edu/instrumentation

[4] Rieke, G. H., Wright, G. S., Böker, T., et al. 2015, PASP, Vol. 127, 584

[5] Wright, G. S., Wright, D., Goodson, G. B., et al. 2015, PASP, Vol. 127, 595

[6] Bouchet, P., García-Marín, M., Lagage, P.-O., et al. 2015, PASP, Vol. 127, 612

[7] Kendrew, S., Scheithauer, S., Bouchet, P., et al. 2015, PASP, Vol. 127, 623

[8] Boccaletti, A., Lagage, P.-O., Baudoz, P., et al. 2015, PASP, Vol. 127, 633

[9] Wells, M., Pel, J.-W., Glasse, A., et al. 2015, PASP, Vol. 127, 646

[10] Rieke, G. H., Ressler, M. E., Morrison, J. E., et al. 2015, PASP, Vol. 127, 665

[11] Ressler, M. E., Sukhatme, K. G., Franklin, B. R., et al. 2015, PASP, Vol. 127, 675

[12] Gardner et al., 2006, Space Science Reviews, Vol. 123, 485

[13] Kalirai, J. 2018, Contemporary Physics, 59, 251

[14] https://miricle.org

[15] Gordon, K. D., Chen, C. H., Anderson, R. E., et al. 2015, PASP, 127, 696

[16] Oesch, P.; Brammer, G.; van Dokkum, P. et al., 2016 The Astrophysical Journal, Vol .819, 129-140

[17] https://www.illustris-project.org