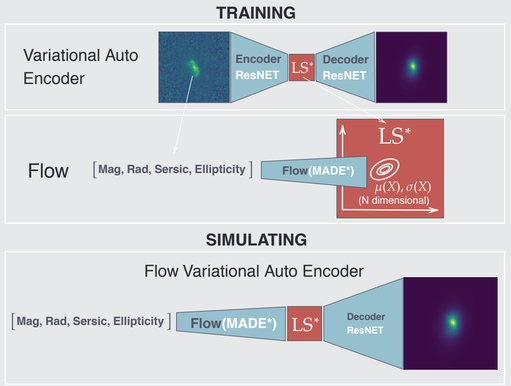
Simulation and Segmentation of Galaxies With Deep Learning for Euclid

Hubert Bretonniere*
Hervé Dole, Marc Huertas-Company, Alexandre Boucaud

Forecast for Galaxy morphology forecasts

Model

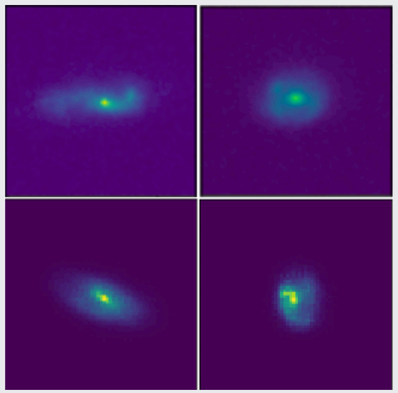
We used a Flow Variational Auto Encoders (Lanusse et al. [1]) to produce complex galaxies. The Variational AutoEncoder learns to produce images of galaxies from real images (here from the Hubble Telescope). The Flow allows to control the parameters of the simulated galaxies.



*LS = Latent Space ; *MADE = Masked AutoRegressive Flow

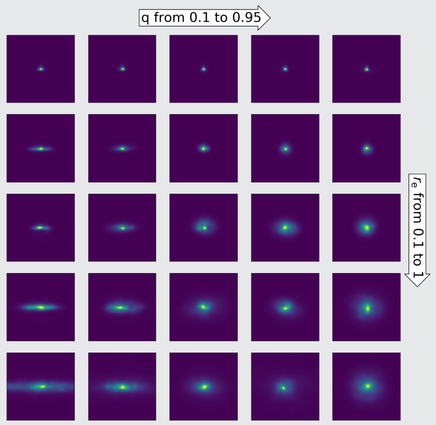
Simulation of Complex Galaxies

Learning from real images, our network can model complex galaxy structures such as spiral arms, bars, rings, bulges and disks.

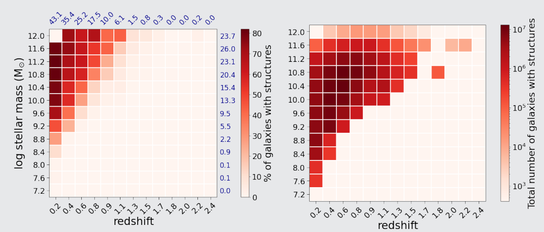


Control of the shape parameters

Galaxies simulated with a catalogue of galaxies with decreasing ellipticity from left to right, and increasing radius from top to bottom. The fact that the galaxies are bigger from top to bottom and rounder from left to right shows that our model reproduces the input parameters, while producing complex galaxies.



Forecast of Euclid's galaxy morphology capacities

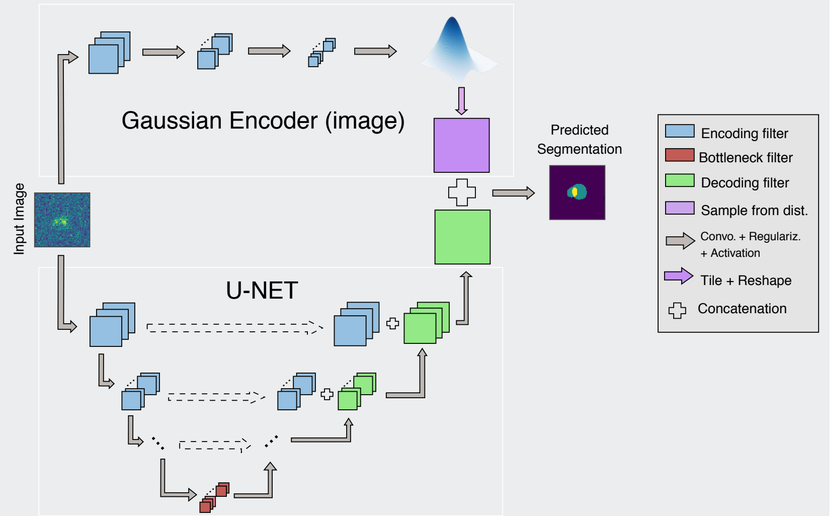


Thanks to our noiseless and 'Euclidised' simulation, we were able to forecast the number of galaxies with interesting morphological features Euclid (will be able to resolve. This is crucial to know which impact Euclid will have in 'Legacy' science, here galaxy morphology or evolution. We found that after the six years of survey, around 250 millions of galaxies will be interesting for morphology science. In this figure, we show the percentage (left) and the total number (right) of those galaxies in the Euclid Wide Survey regarding their redshift and stellar mass.

Segmentation of Blended Galaxies

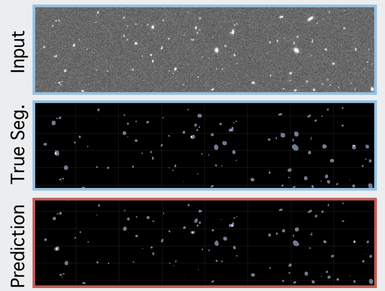
Model

We adapted a so called Probabilistic U-Net to detect galaxies. The output of the model is a segmentation map: the pixels belonging only to the background are set to 0, the ones with only one galaxy are set to 1, and the pixels where there are two (or more) galaxies overlapping to 2. The U-Net (bottom part of the figure) is a state of the art model to produce this kind of segmentation, by encoding and decoding the necessary informations using convolution layers. The Gaussian Encoder (top part of the figure) gives a probabilistic behaviour to the U-Net, capturing the uncertainty of the prediction in the dispersion of a Gaussian. Detecting galaxies which are blended is crucial to the missions that probe weak lensing (by computing the shape of the galaxies), as it is for Euclid.

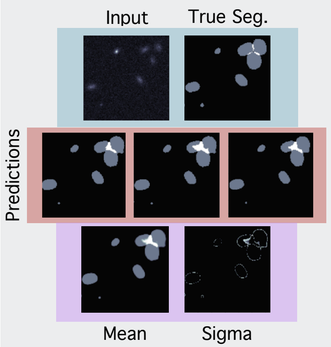


Reconstruction of a segmentation map

Our model is trained on stamps. But it can reproduce a large field by cutting the field in stamp of the training size. Because the training set I is also made of a big field gutted into stamps, it can recognise galaxies that are cut in the edges of the stamps. Here is an example of a big field segmented with this method.



Probabilistic Behaviour



Our model is probabilistic. By sampling many times the gaussian encoder, we will produce different realisations of the segmentation. It encodes the uncertainty of the prediction. You can see here that the difficult part of the task (the area blended by three galaxies) is more uncertain (bigger sigma) than the isolated galaxies.

[1] Deep Generative Models for Galaxy Image Simulations, F.Lanusse et al. 2020 ; [2] Euclid Definition Study Report, Laureijs et al. 2011, [3] A Probabilistic U-Net for Segmentation of Ambiguous, S.Kohl et al., 2019