Internship Project

Refinement of Deep Learning approaches for cells clustering

By:

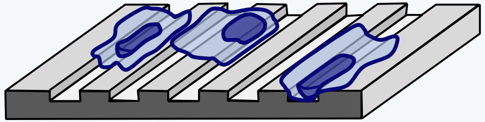
Souheib BEN MABROUK

Supervisors:

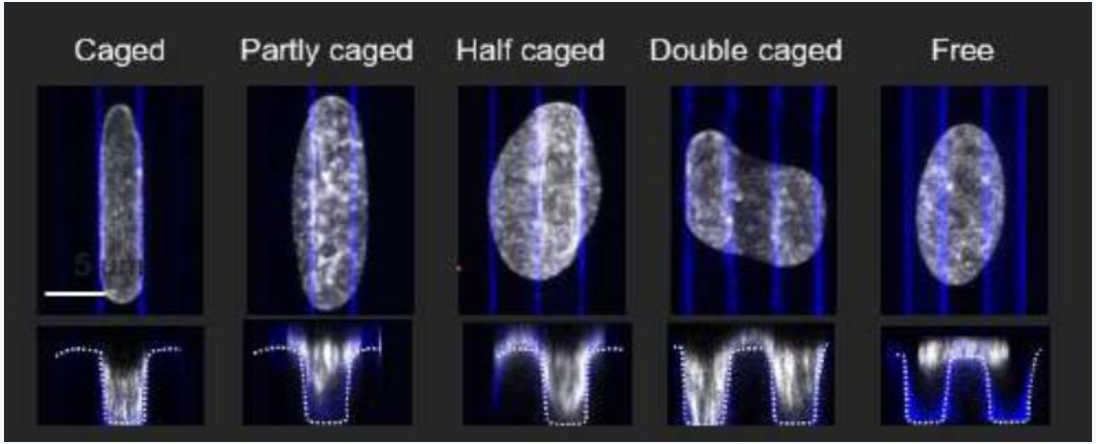
Prof : Elsa ANGELINI
Post doc : Bettina ROELLINGER
Prof : Abdul BARAKAT

Project Context (1/2)

The need for innovative tools for early cancer detection



Response of cancerous cells vs non tumorous cells on micropattern surfaces: **deformations on trench lines**



Project Context (2/2)

A continuation of the PRIM project of François Thenier

Specific interest

- Deformations induced by **mutations** compared to control cells
- **Different kinds** of mutations

New challenge

- Another dataset with **new processing challenges**
- Need for a more **sophisticated architecture**

Agenda

I. Scope Statement

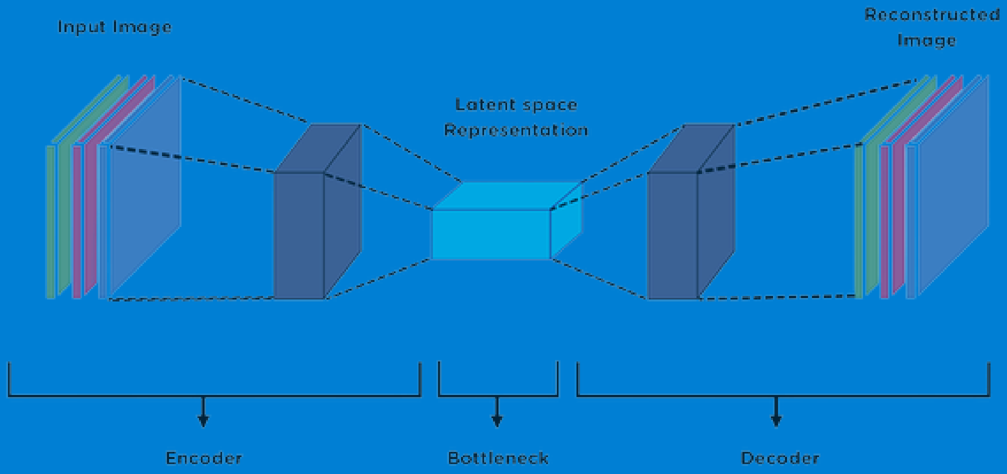
II. Project Design

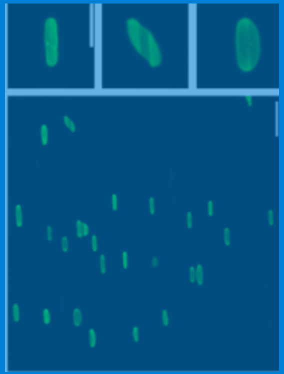
III. Methods

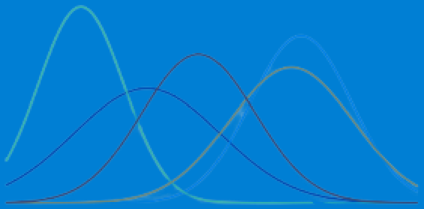
IV. Results & Discussions

V. Refinement of the architecture

VI. Conclusion & Perspective







I. Scope Statement

I. Scope Statement (1/2)

Project's Goal:

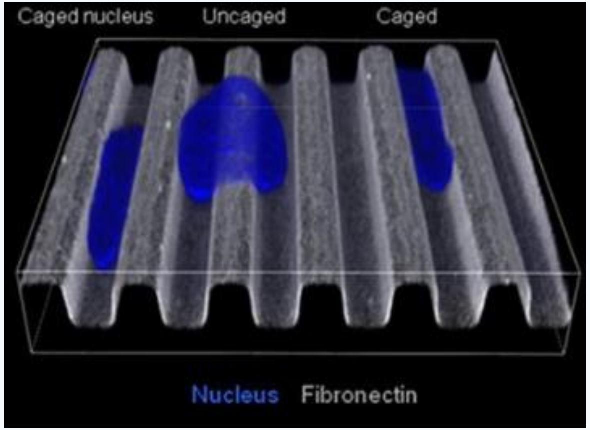
Develop a deep learning pipeline for weakly supervised cells segmentation and clustering

- Adapt image pre-processing to the density of cell images
- Refine a CNN solution for cell segmentation : Cellopse
- Generate the segmentation masks and label the cells
- Develop VAE / cVAE architectures for cell clustering
- Evaluate the models and infer the results

I. Scope Statement (2/2)

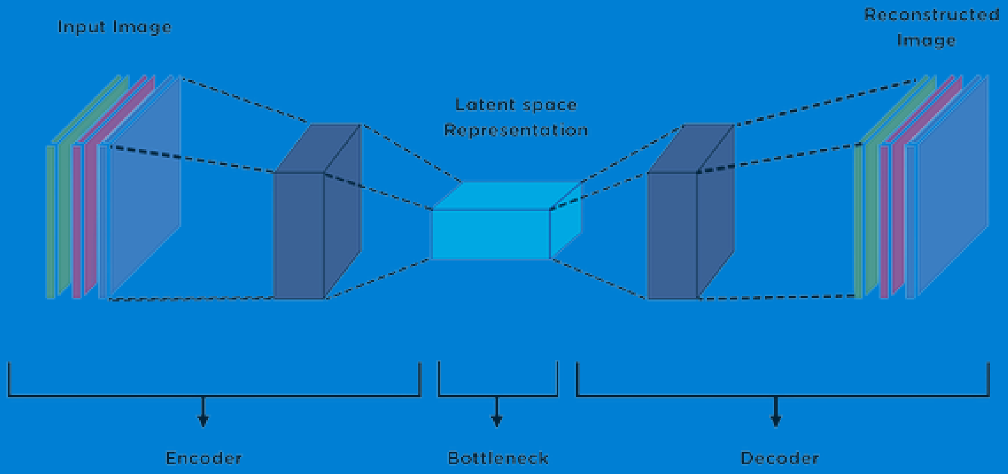
Dataset

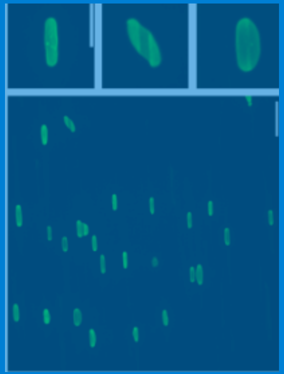
- Images of MCF10A cells
- Cultured on grooves of 5 µm width spaced by 5 µm

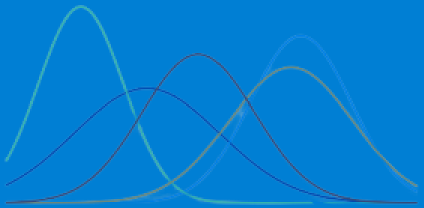


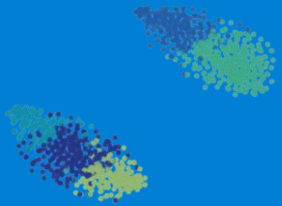
Types

- Control line
- BRAFV600E
- DCIS.com
- PIK3CA
- Rac1









II. Project Design

II. Project Design (1/2)

Preprocessing

- Rotation
- CLAHE
- Normalize
- Clip intensity
- Denoising

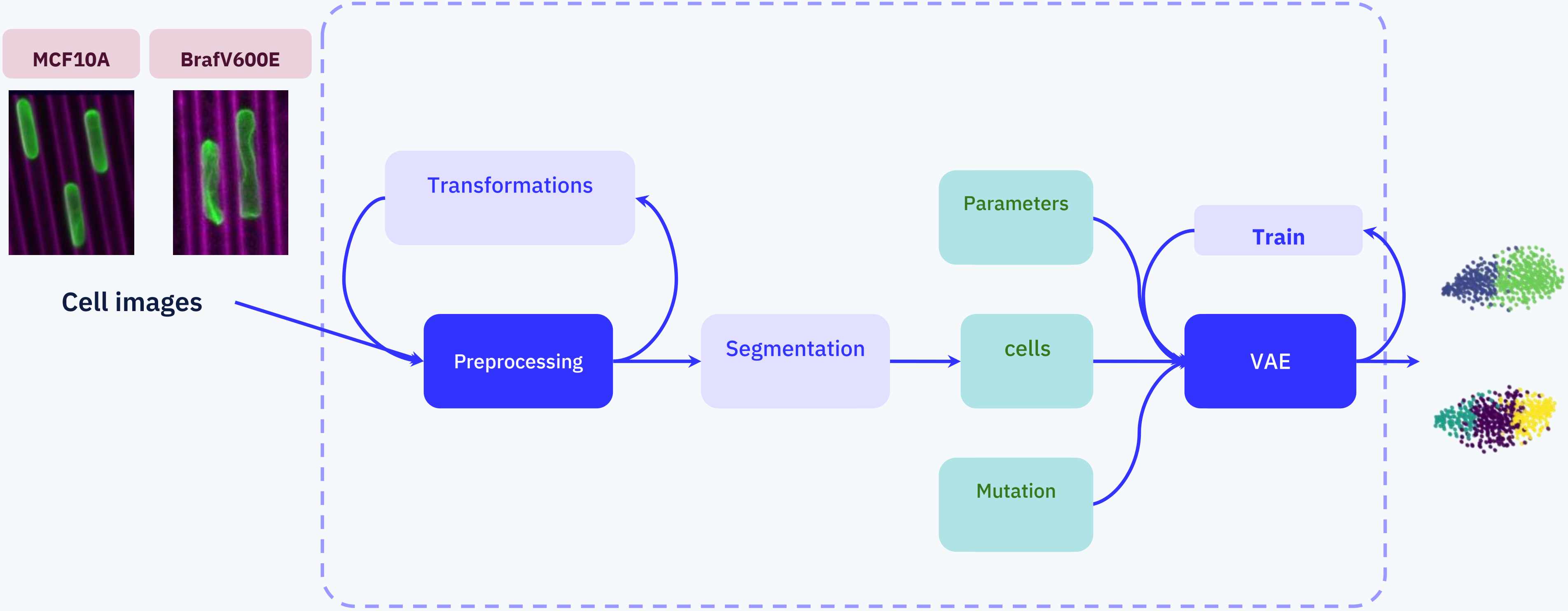
Segmentation

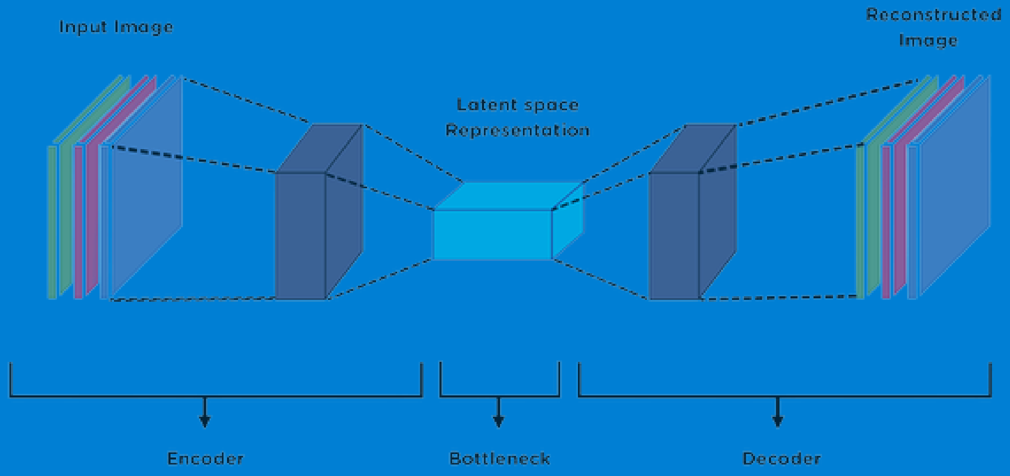
- Fine tune a Cellpose model
- Segmentation Masks
- Crop Cell patches

Cells clustering

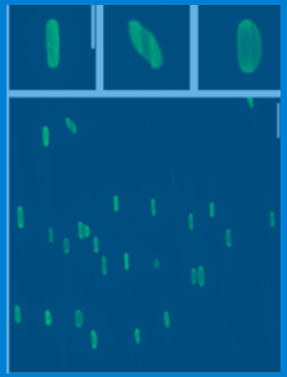
- VAE
- cVAE

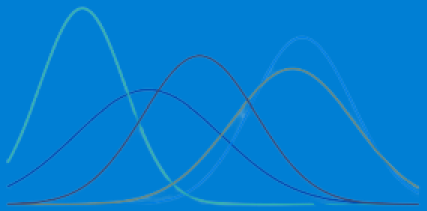
II. Project Design (2/2)

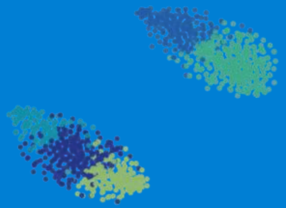




Input Image
Reconstructed Image
Latent space Representation
Encoder
Bottleneck
Decoder







III. Methods

III. Methods : Preprocessing (1/5)

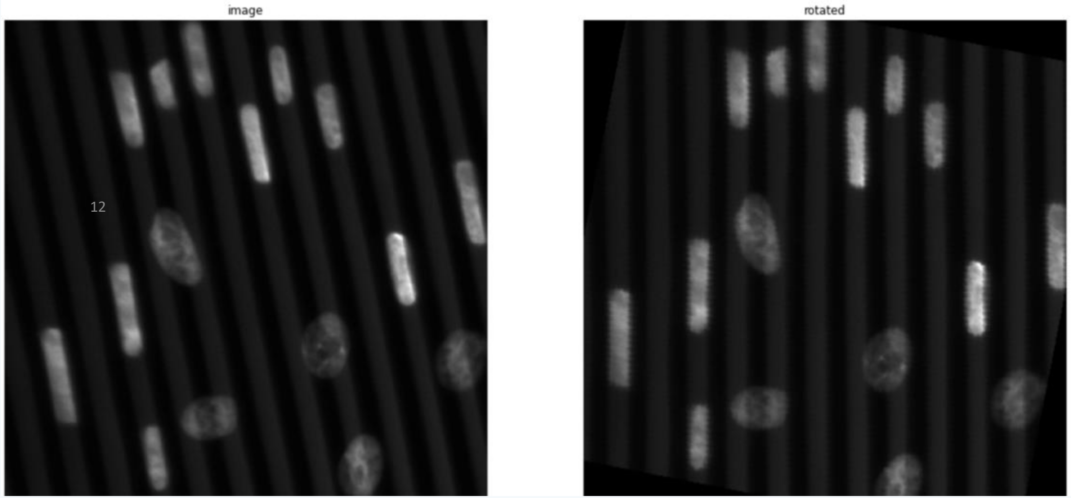
Preprocessing

Rotation | CLAHE | Normalize | Clip intensity | Denoising

- **Angle Detection**: based on Hough line transform with Canny edge detection

- **Image Rotation :** using a rotation matrix of the detected angle

Eliminates **orientation-related noise** in the training data



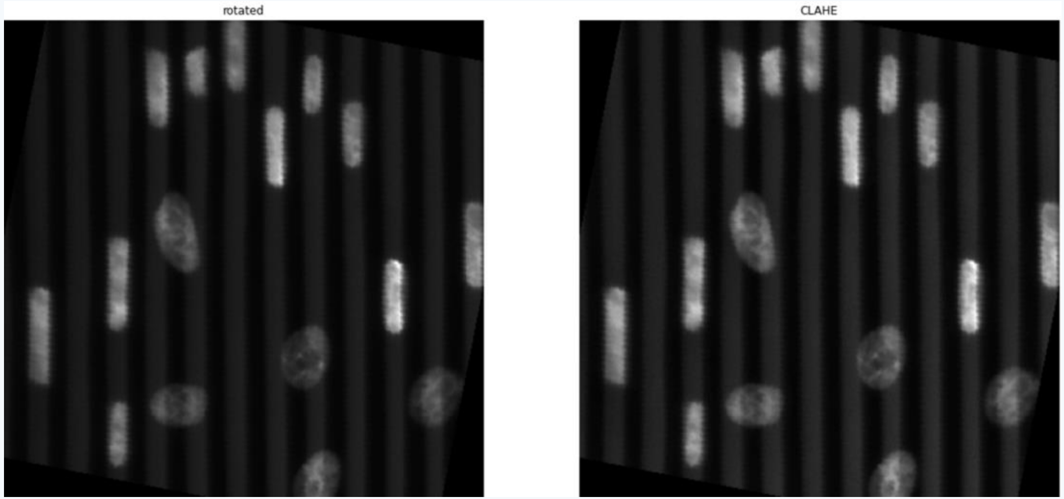
III. Methods : Preprocessing (2/5)

Preprocessing

Rotation | **CLAHE** | Normalize | Clip intensity | Denoising

- **Poor** contrast images, challenging identification of cell features
- Contrast Limited Adaptive **Histogram Equalization**

➡ **Improves** image **contrast**



III. Methods : Preprocessing (3/5)

Preprocessing

Rotation | CLAHE | **Normalize** | Clip intensity | Denoising

- **Scale** the pixel values linearly to a common range, typically [0, 1].
- **Prepare** images for intensity clipping

14

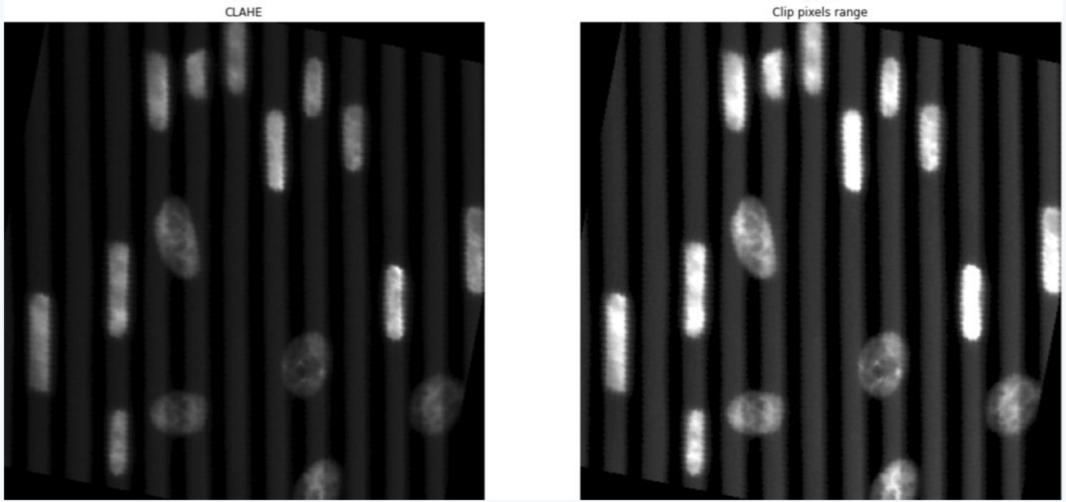
➡ **Mitigates issues** related to varying image intensities

III. Methods : Preprocessing (4/5)

Preprocessing: Rotation → CLAHE → Normalize → **Clip intensity** → Denoising

- **Outliers:** extreme pixel values
- Use of the **percentiles** of the image intensity distribution.

➡ **Enhances** image **quality** by reducing extreme pixel values



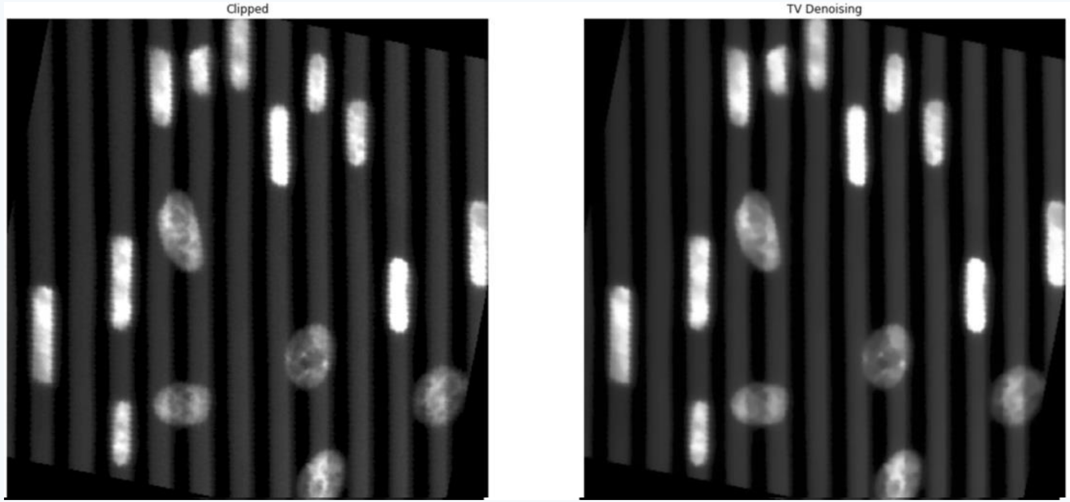
III. Methods : Preprocessing (5/5)

Preprocessing

Rotation | CLAHE | Normalize | Clip intensity | **Denoising**

- Total Variation **(TV) denoising**

➡ **Enhances** image **clarity** by **removing** unwanted **noise and** keeping cell **textures**



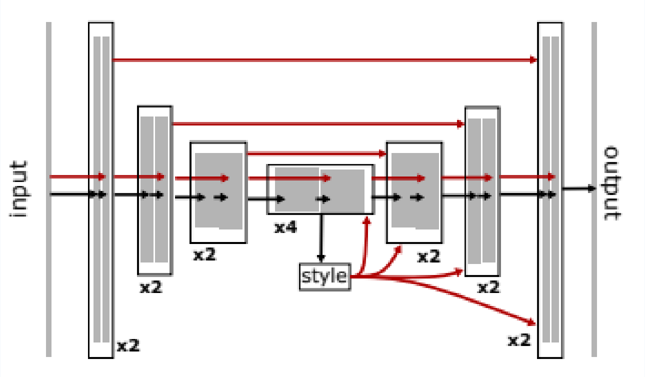
III. Methods : Segmentation (1/5)

Segmentation

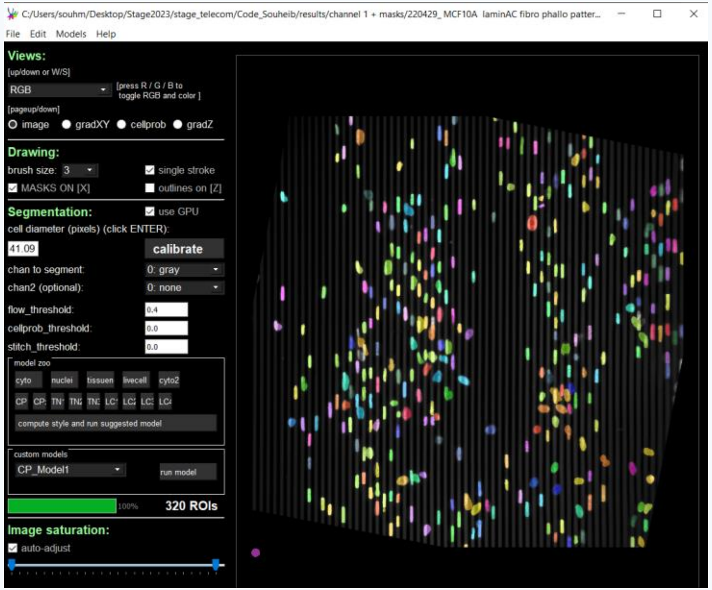
Fine tune a Cellpose model

Segmentation Masks

Crop Cell patches



Cellpose architecture : **UNET**



Cellpose GUI: "**Human in the loop**" training

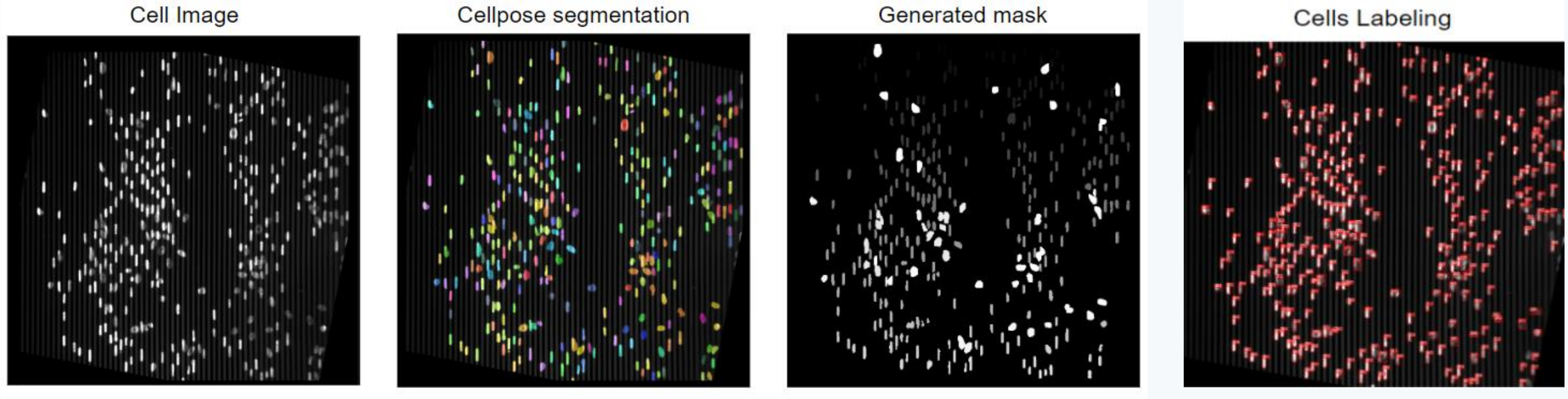
III. Methods : Segmentation (2/5)

Segmentation

Fine tune a Cellpose model

Segmentation Masks

Crop Cell patches



III. Methods : Segmentation (3/5)

Segmentation

Fine tune a Cellpose model → Segmentation Masks → Crop Cell patches

Method 1 : Crop + Resize

- Resize the cell image while preserving its aspect ratio.

\+ Flexibility of the resized shape

\- Information loss : Real size of the cell

Method 2 : Crop + Padd

- Crop or pad the cell image while preserving its original size.

\+ Take into account the real size of the cell as a feature

\- Limited choice of the target shape

→ **A study of cell properties is required**

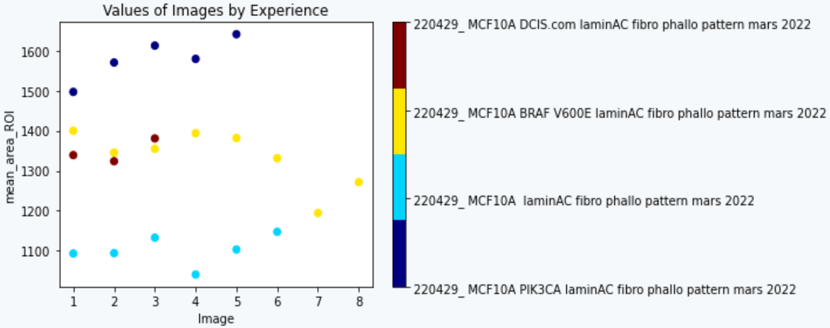
III. Methods : Segmentation (4/5)

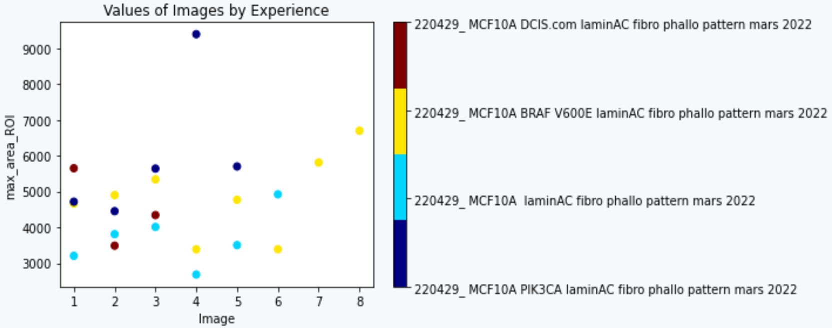
Segmentation

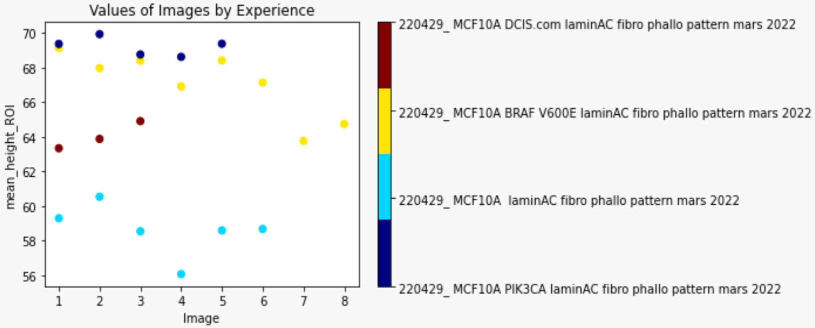
Fine tune a Cellpose model

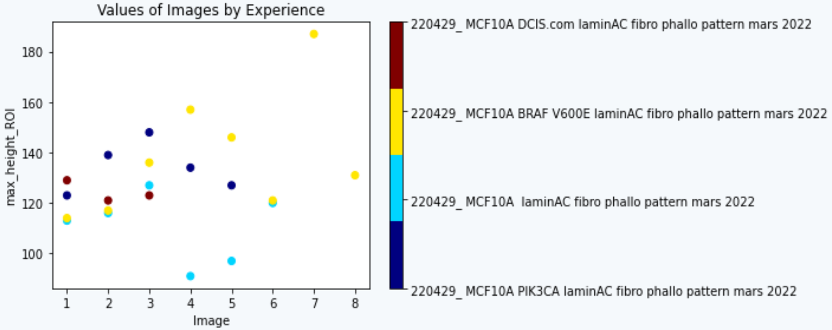
Segmentation Masks

Crop Cell patches









III. Methods : Segmentation (5/5)

Segmentation | Fine tune a Cellpose model | Segmentation Masks | Crop Cell patches

Control cell | PIK3CA mutant

Method 2 : Crop + Padd

- Crop or pad the cell image while preserving its original size.

\+ Take into account the real size of the cell as a feature

\- Limited choice of the target shape

\+ **Eliminate background**

III. Methods : Cells clustering (1/3)

Cells clustering

VAE

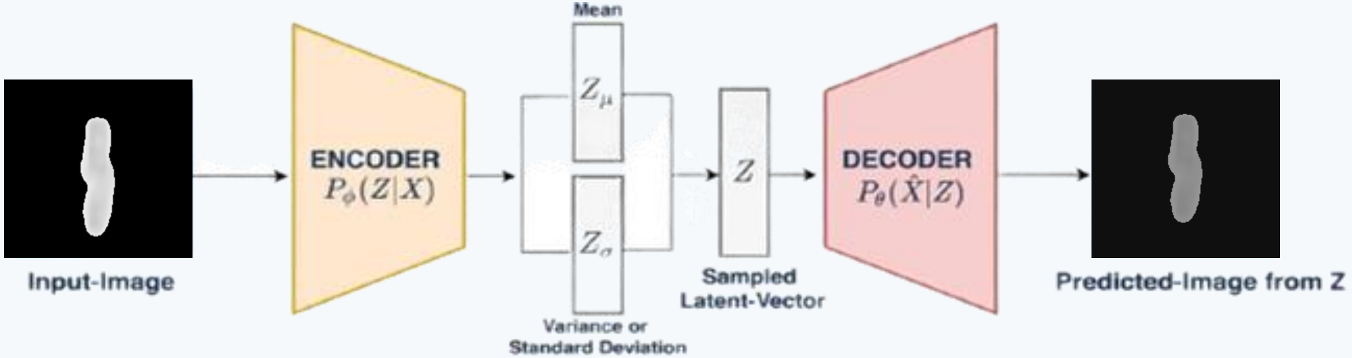
cVAE

III. Methods : Cells clustering (1/3)

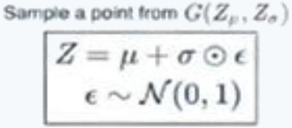
Cells clustering

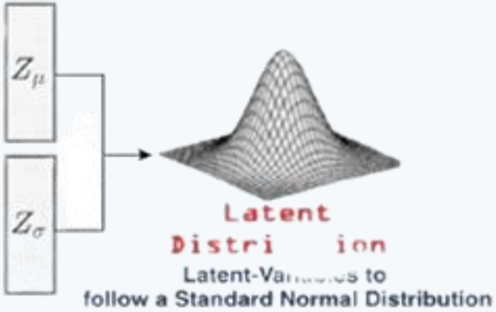
VAE

cVAE



$$Loss = L(x,\hat{x}) + \beta KL(q(z|x)||p(z))$$



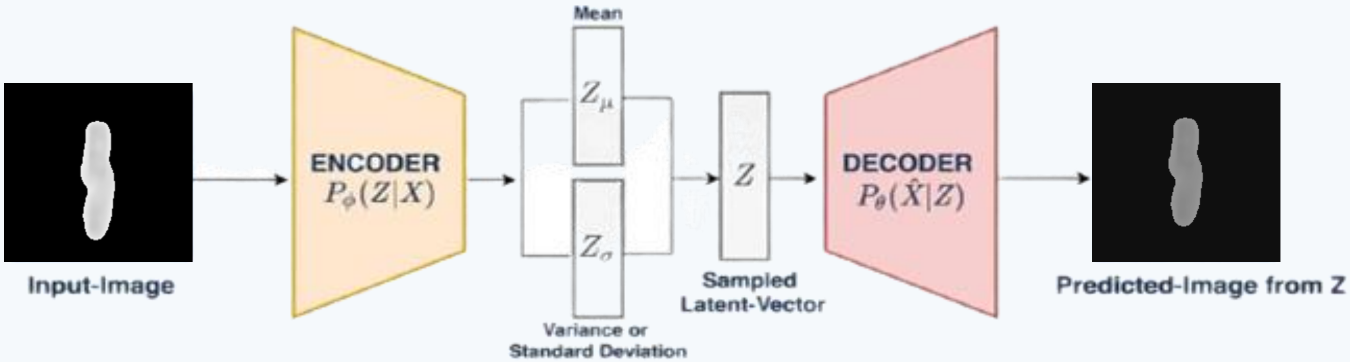


III. Methods : Cells clustering (2/3)

Cells clustering

VAE

cVAE



$$Loss = L(x,\hat{x}) + \beta KL(q(z|x)||p(z))$$

$L(x,\hat{x})$: Reconstruction loss (MSE , BCE ...)

$KL(q(z|x)||p(z))$: Kullback–Leibler divergence

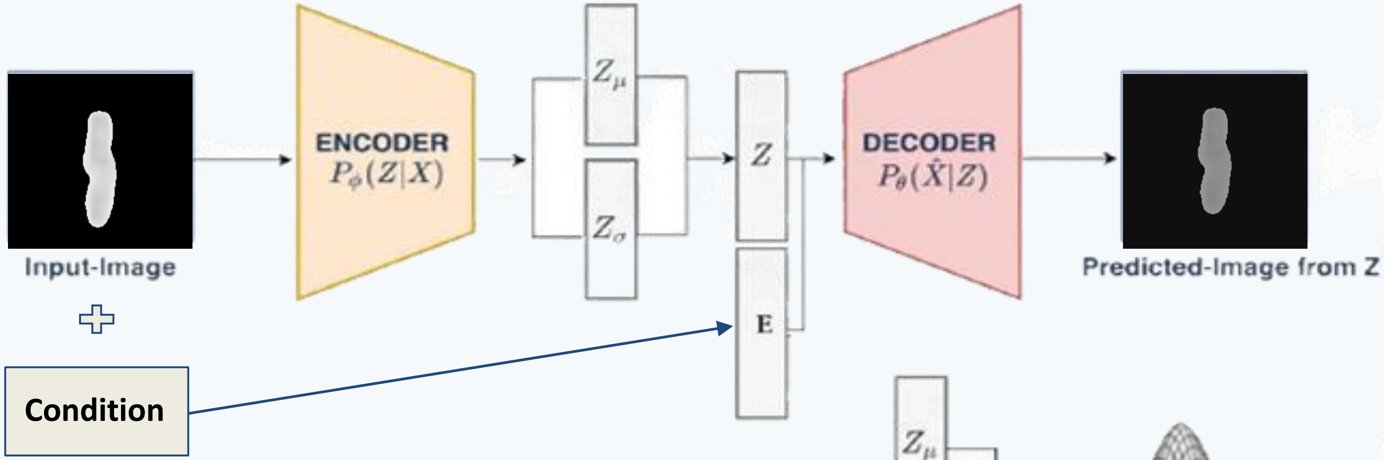
$p(z) \sim \mathcal{N}(0,1)$

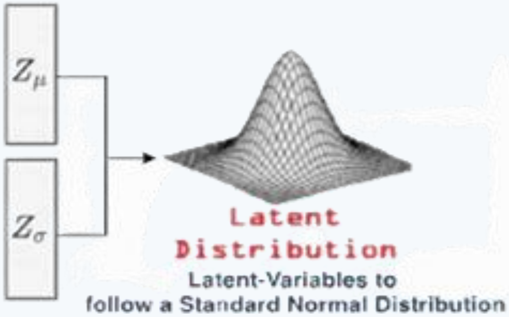
III. Methods : Cells clustering (3/3)

Cells clustering

VAE

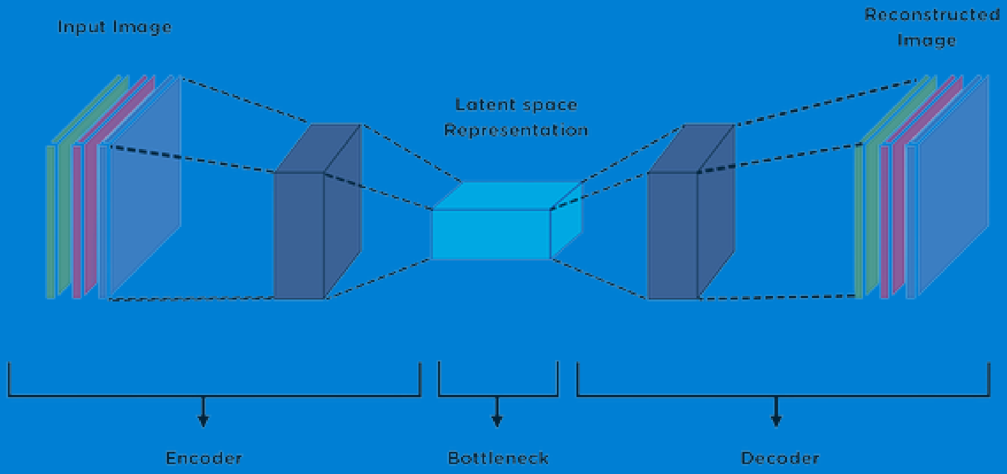
cVAE

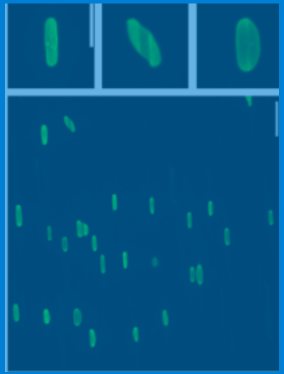


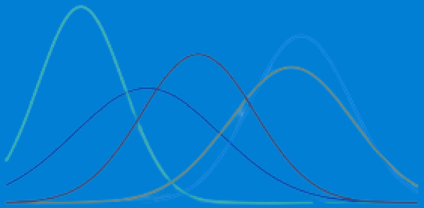


$$E = Embedding(condition)$$

$$\boxed{Z' = Z \oplus E}$$





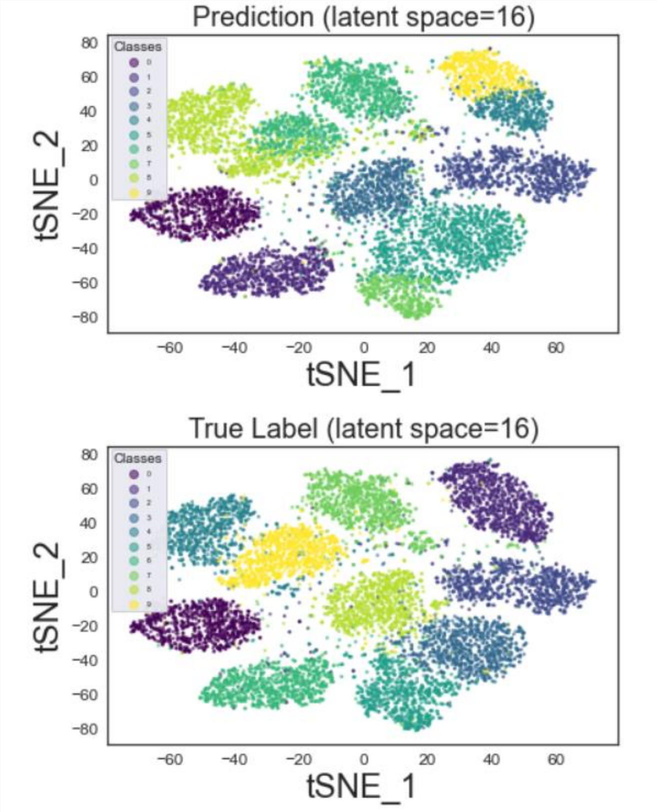


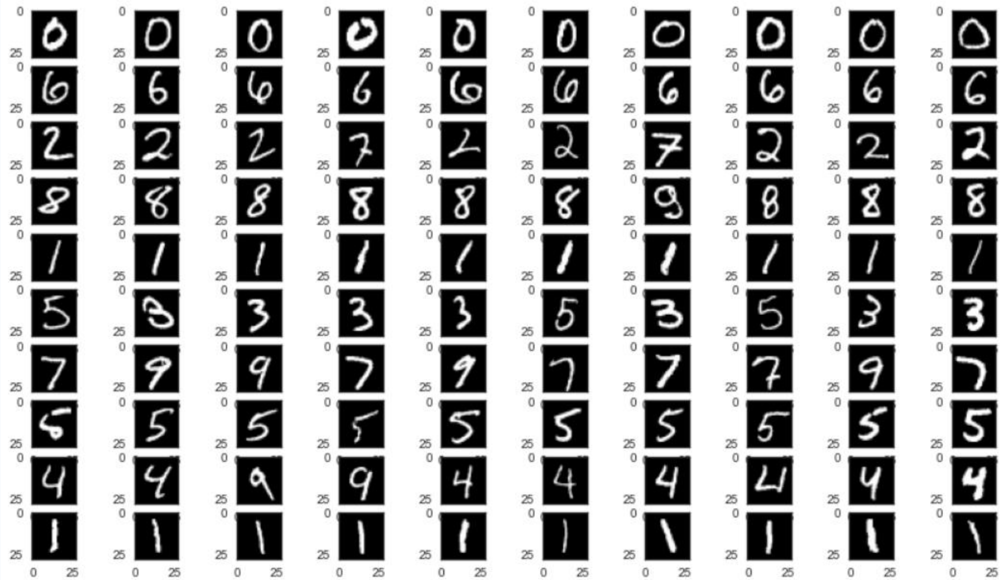
IV. Results & Discussions

IV. Results & Discussions:

- Comparison of VAE and cVAE performances on MNIST and FashionMNIST clustering
- The role of beta parameter
- Evaluation of cells clustering and limitation of the conventional models

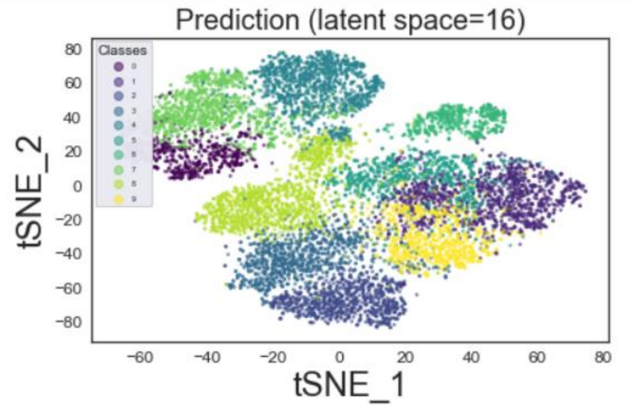
IV. Results & Discussions: VAE on MNIST and Fashion MNIST (1/2)

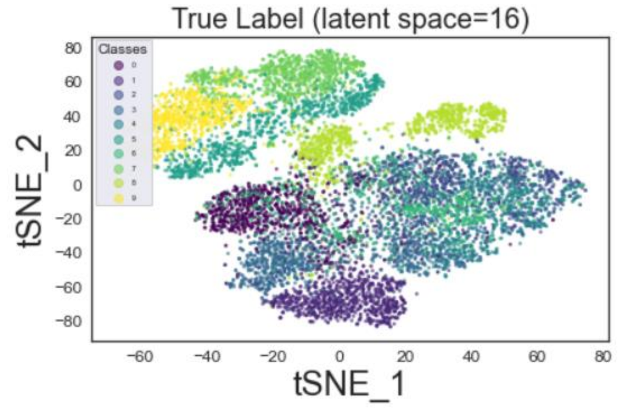


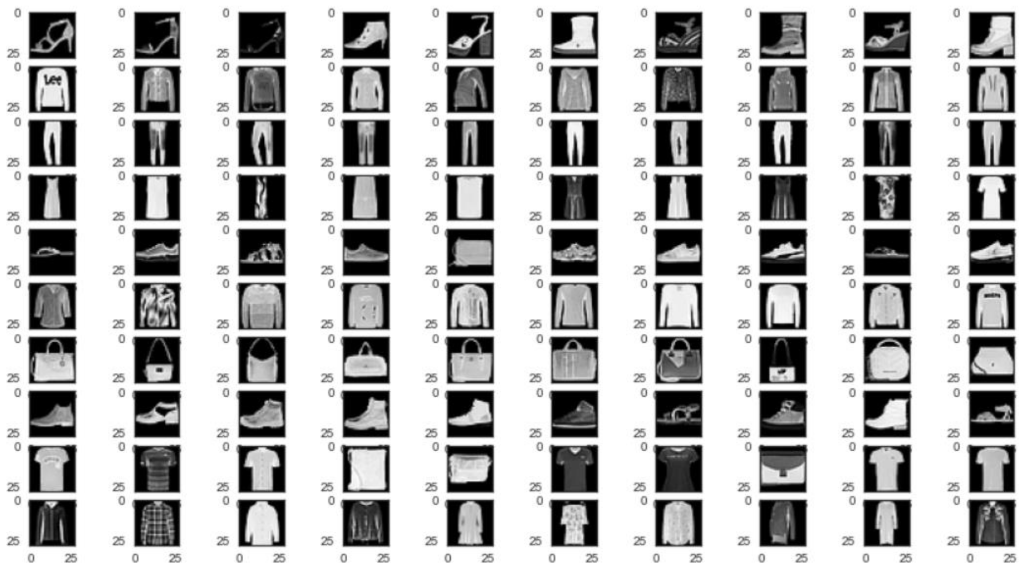


Dataset_Z_latentDim	Accuracy	Silhouette	DB	CH
Mnist_Z_16	0.811	0.0773455	2.44891	429.722
Mnist_Z_32	0.8167	0.0368477	3.84555	174.857

IV. Results & Discussions: VAE on MNIST and Fashion MNIST (2/2)

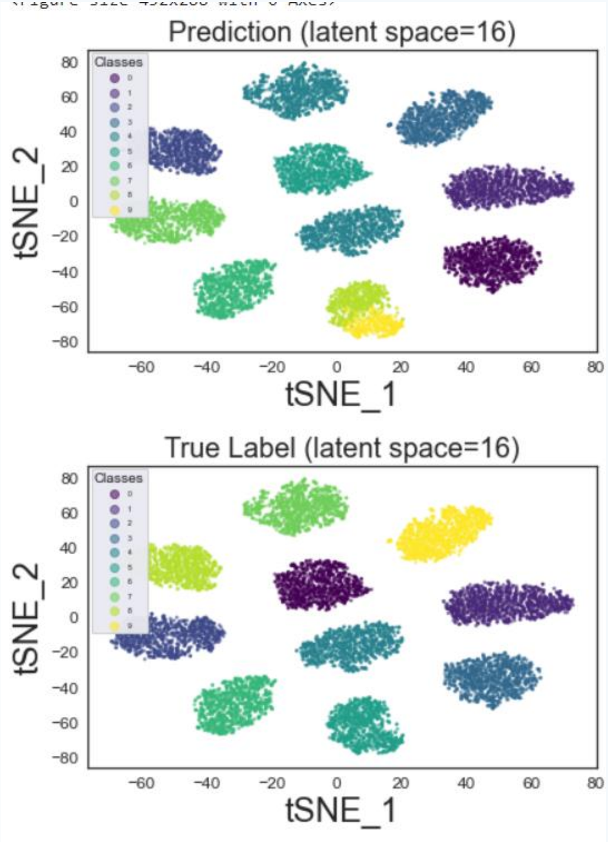


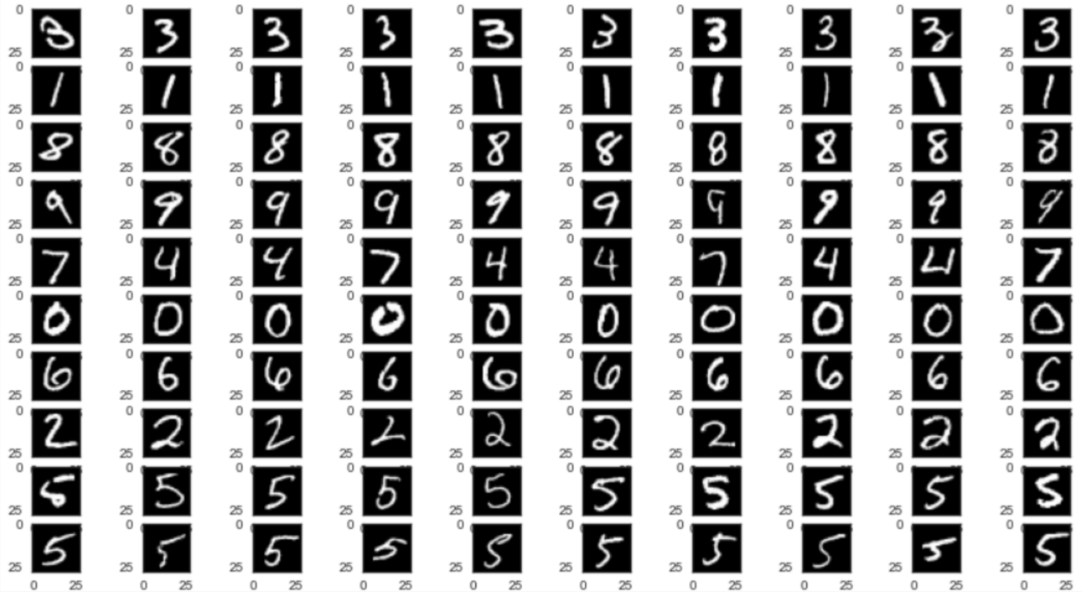




Dataset_Z_latentDim	Accuracy	Silhouette	DB	CH
FashionMnist_Z_16	0.5848	0.0628801	2.91779	399.27
FashionMnist_Z_32	0.5934	0.0277953	4.66206	172.193

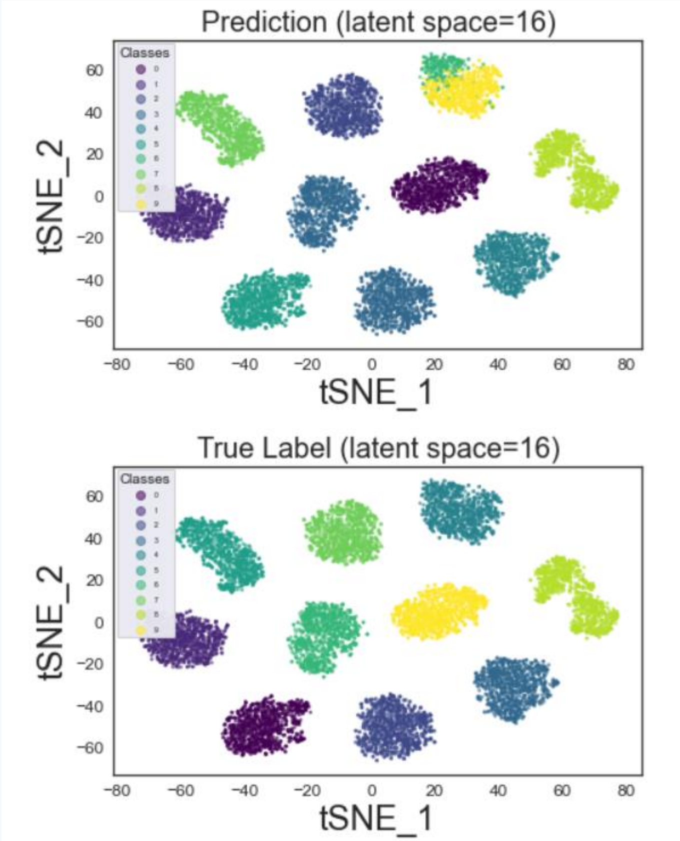
IV. Results & Discussions: cVAE on MNIST and Fashion MNIST (1/2)

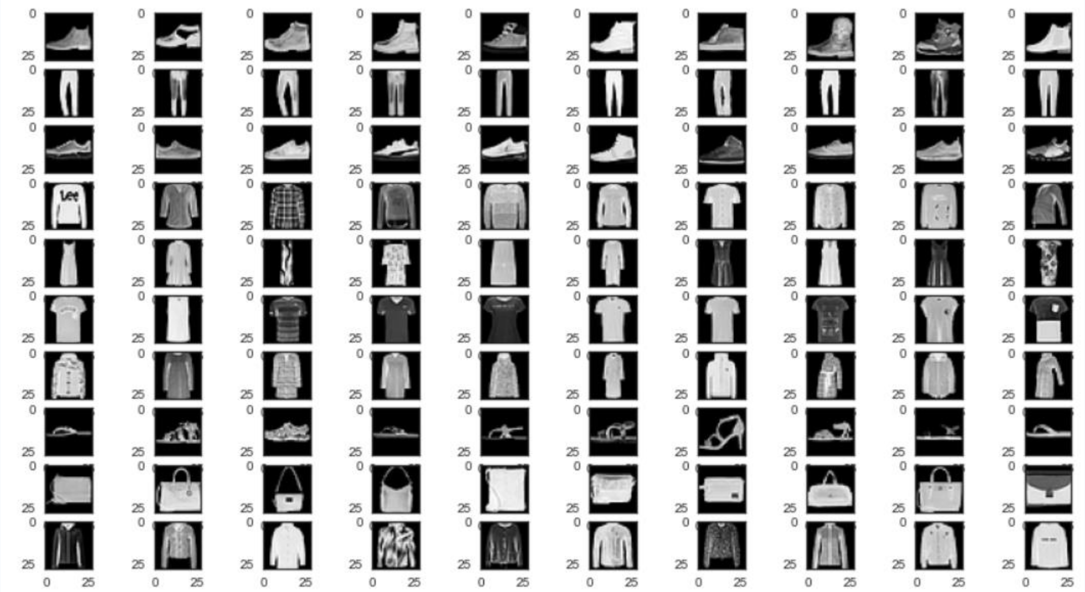




Dataset_Z_latentDim	Accuracy	Silhouette	DB	CH
Mnist_Z_16	0.9018	0.251985	1.8037	1107.44
Mnist_Z_32	0.9042	0.184489	2.4797	712.183

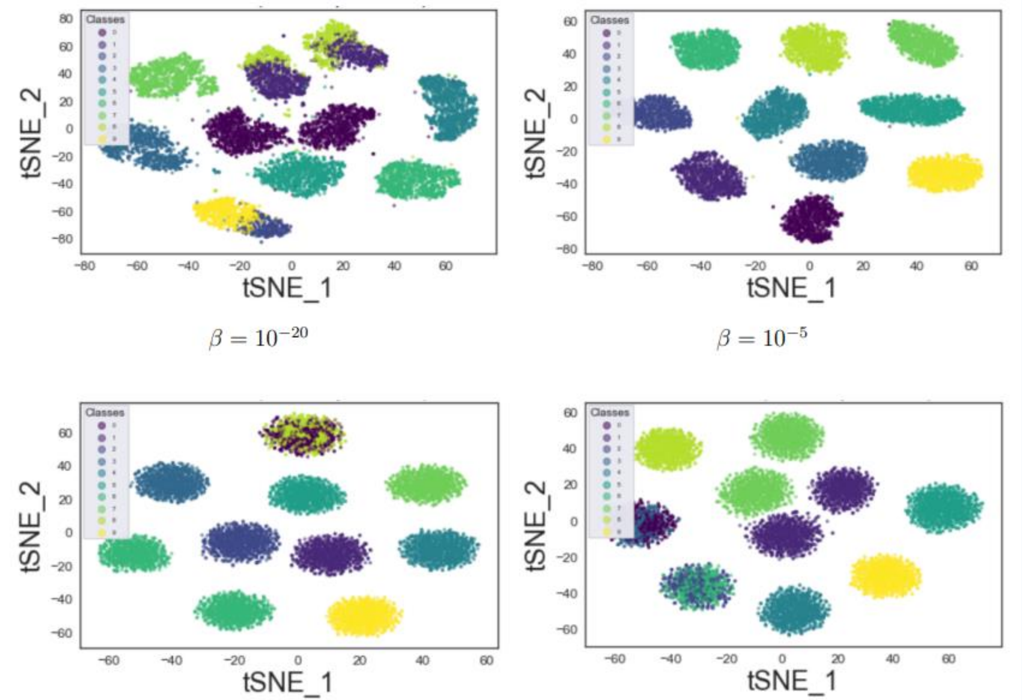
IV. Results & Discussions: VAE on MNIST and Fashion MNIST (2/2)





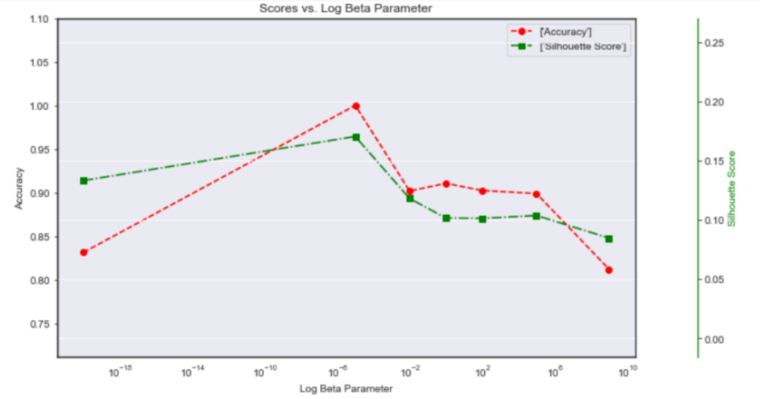
Dataset_Z_latentDim	Accuracy	Silhouette	DB	CH
FashionMnist_Z_16	0.9	0.209225	2.02346	944.545
FashionMnist_Z_32	0.975	0.217652	1.75011	807.943

IV. Results & Discussions: Beta parameter



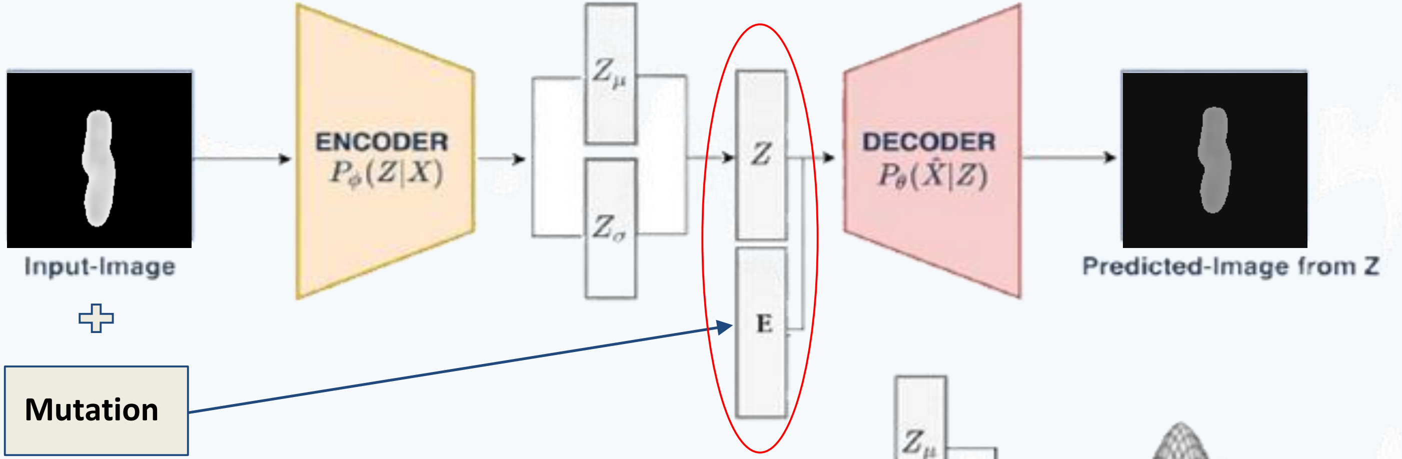
$\beta = 10^{-20}$ $\beta = 10^{-5}$

$\beta = 1$ $\beta = 10^{9}$



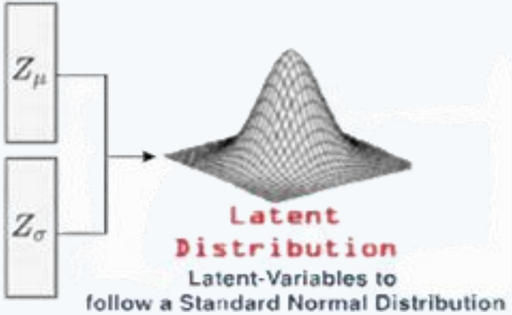
Accuracy and Silhouettes scores vs β for Mnist

IV. Results & Discussions: cVAE for cells clustering (1/3)

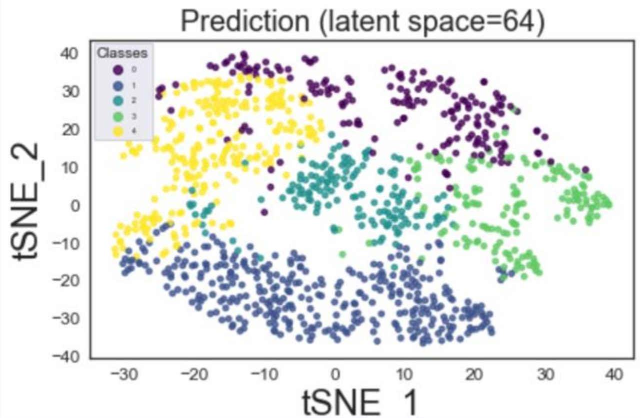


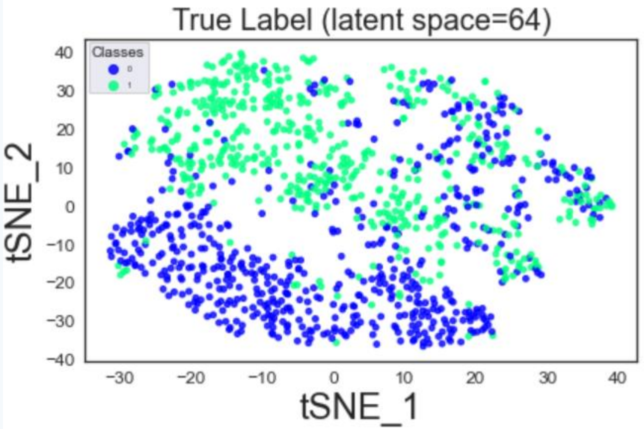
$$E = Embedding(condition)$$

$$\boxed{Z' = Z \oplus E}$$

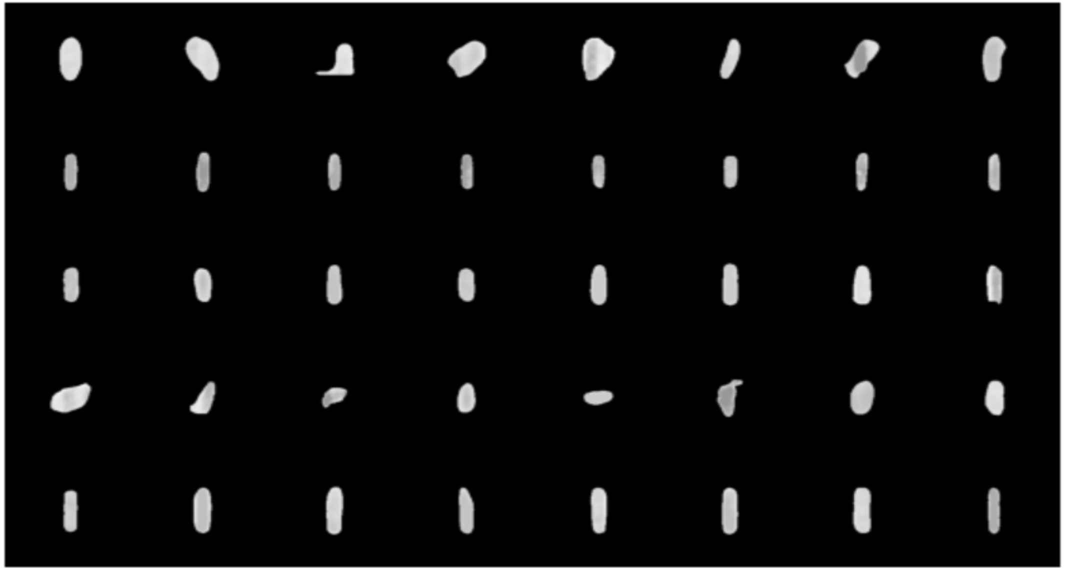


V. Refinement of the architecture





each line correponds to a cluster



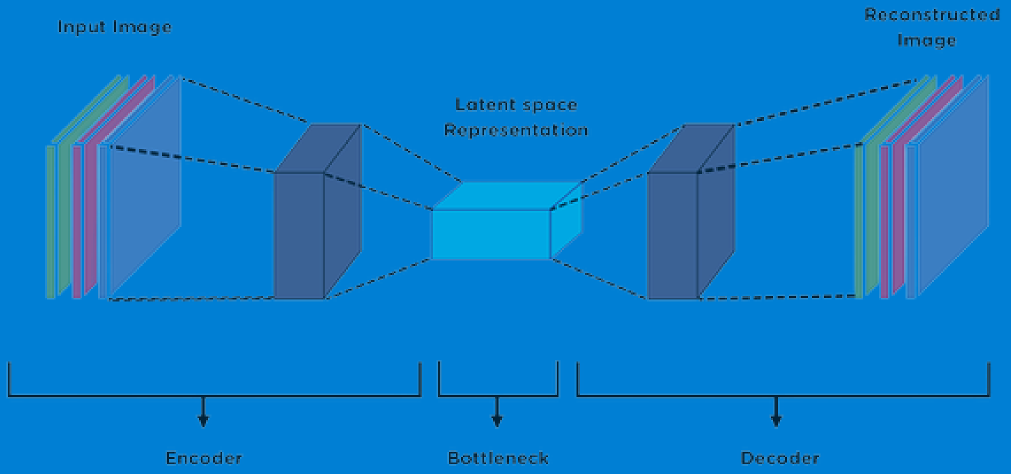
Label	Number of Cells	% Occurrences of Wild Cells	% Occurrences of Mutation Cells
0	177	36.72%	**63.28%**
1	347	**90.20%**	9.80%
2	152	28.29%	**71.71%**
3	159	50.31%	49.69%
4	235	25.11%	**74.89%**

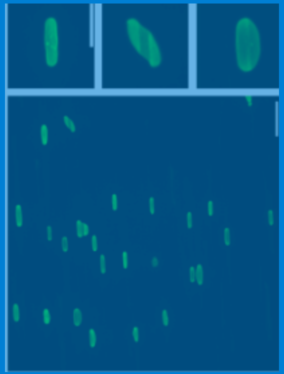
Number of clusters	Silhouette score	Davies-Bouldin Index	Calinski-Harabasz Index
4	**0.1922**	**1.8014**	**232.8991**
5	0.1588	1.8083	212.1467

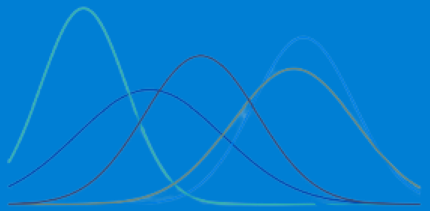
IV. Results & Discussions: cVAE for cells clustering (3/3)

- Divergence of the Kullback-Leibler loss in the training
- Overlap of control and mutation cells at some clusters that are very close in terms of deformation / shape

+ A statistical inference of the current results shows potential features for the cancerous cells developed in the case of mutations





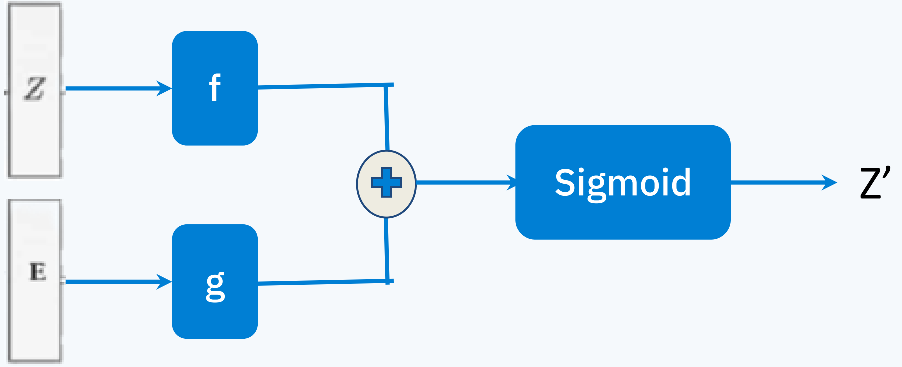


V. Refinement of the architecture

V. Refinement of the architecture (1/4)

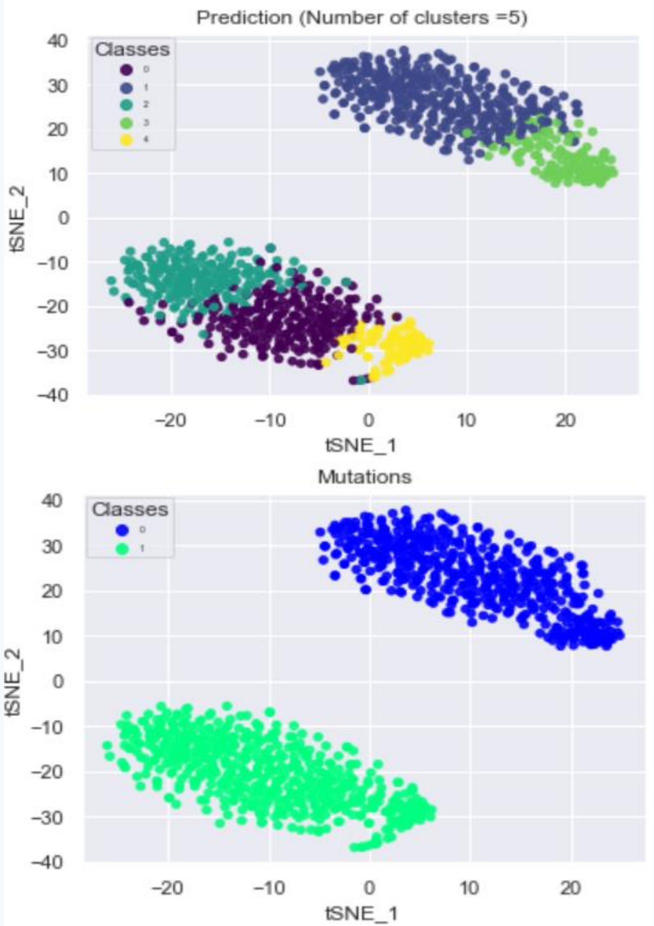
- Divergence of the Kullback-Leibler loss in the training : $\boxed{Z' = Z \oplus E}$

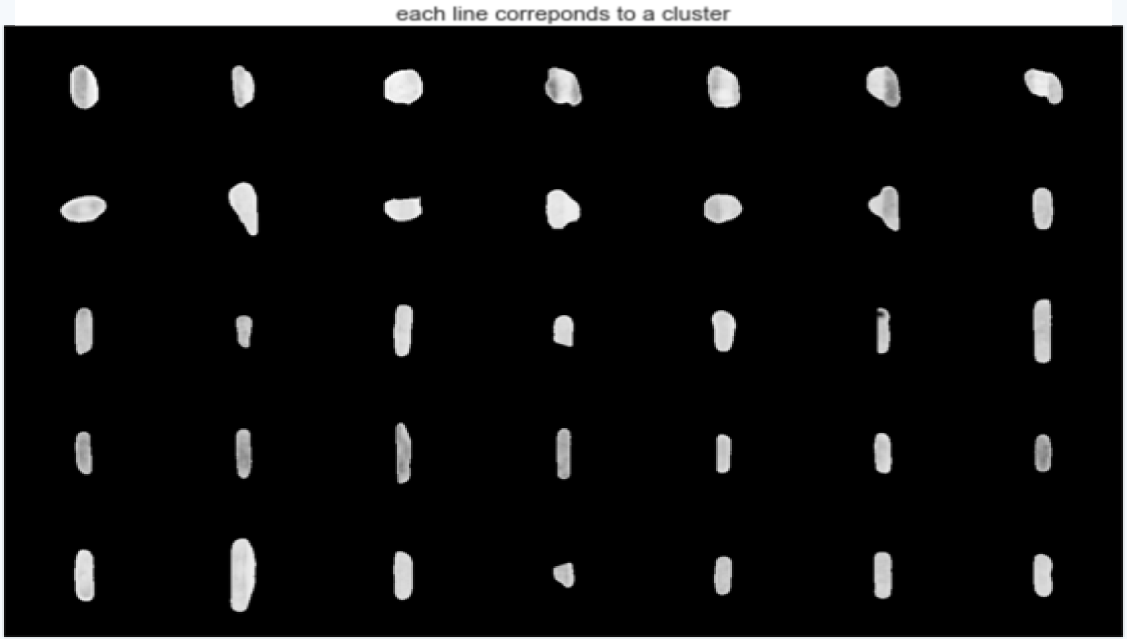
-> Refine the architecture so that the latent space becomes more "flexible" in order to resemble a Gaussian distribution.



$$\boxed{Z' = Sig(f(Z) \oplus g(E))}$$

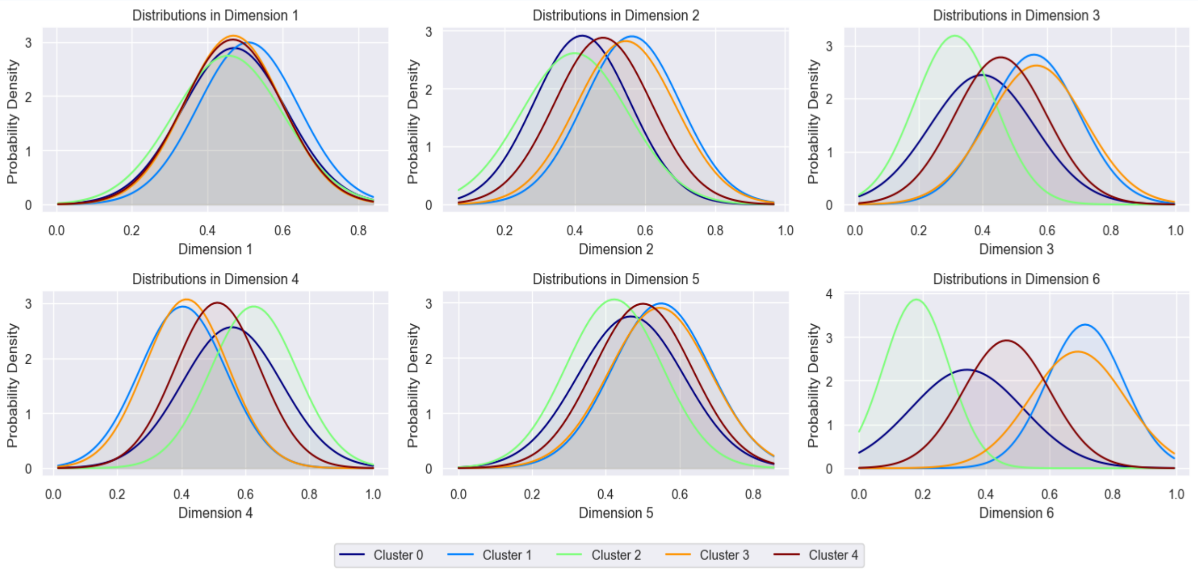
V. Refinement of the architecture (2/4)



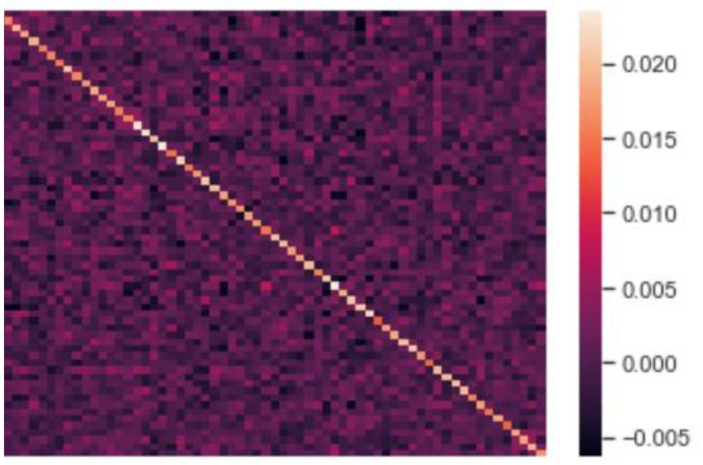


7 Samples of each cluster

V. Refinement of the architecture (3/4)



Distribution of the clusters at 6 dimensions of the latent space



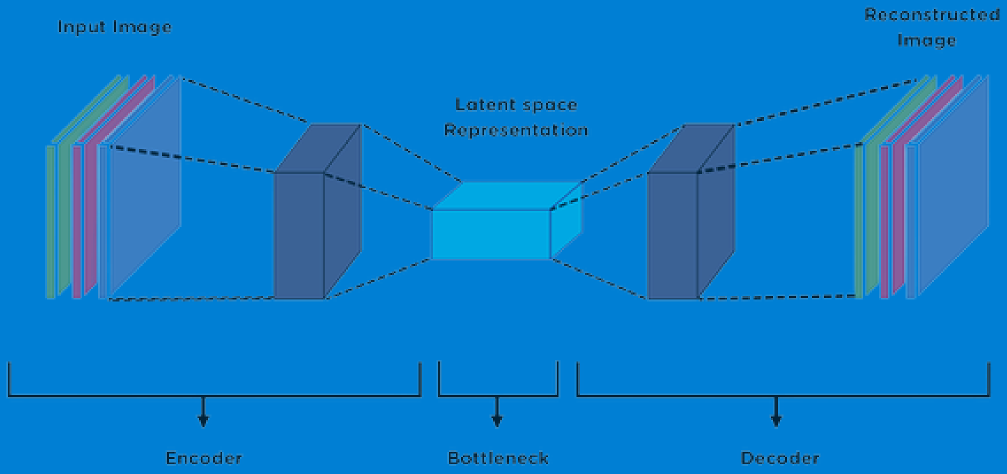
Correlation matrix of the cluster1 with the others at each of Latent space dimensions

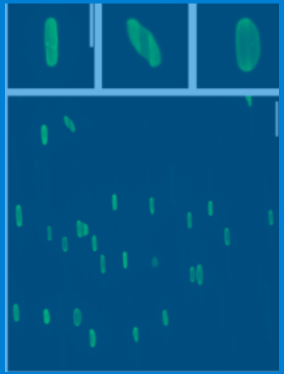
V. Refinement of the architecture (4/4)

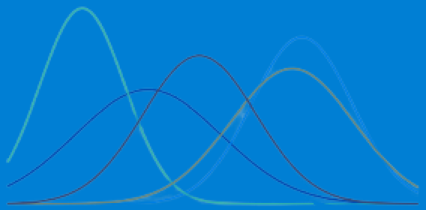
- Alpha is an architectural Hyperparameter.
- It controls the contribution of Z and E to the final latent space Z'

Alpha	nbclusters	Silhouette	DB	CH
16	5	0.394324	0.820397	4109
16	7	0.301135	0.961042	3552.71
8	5	0.424161	0.785045	3869.34
8	7	0.274534	1.10468	3167.72

The dimension of g(E) is $\frac{1}{\alpha}$.latent_dim The dimension of f(Z) is $\frac{\alpha-1}{\alpha}$. latent_dim







Conclusion & Perspective

Conclusion & Perspective (1/2)

Summary

- Enhanced and preprocessed data
- Fine-tuned models of Cellpose for MCF10A cells
- Build different Variational Autoencoder architectures for cell clustering
- Deployed models and documented code on github repository

souheib1

Conclusion & Perspective (2/2)

Challenges

- Defining the preprocessing methods for a rich data
- Using cellpose GUI (GPU needed)
- Plenty hyperparameters to fine-tune
- Optimizing the code for future uses

Perspective

- End-to-end pipeline
- Study the correlations of the dynamic aspects of each cluster

Thank you for your attention

souheib1