



$\beta$ -VAE : Learning Basic Visual Concepts with a Constrained Variational Autoencoders

# DEEP LEARNING

---

Armen SULEJMANI & Gauthier EVRAERD  
20192557 20190839

# PRÉSENTATION

- Modèle: Intégration de  $\beta$  dans un Variational Autoencoder (VAE)
- Base de données: MNIST avec rotations aléatoires
- Objectif : Capturer les facteurs de variation

# BASE DE DONNÉES



**Dataset MNIST** : 60 000 images (entraînement) + 10 000 images (test)



**Redimensionnement** : 28x28 à 64x64 pour aligner avec l'architecture



**Dataset Rotated MNIST** : Images avec rotation aléatoire (-90° à 90°)



**Entraînement** : Combinaison Rotated MNIST + MNIST



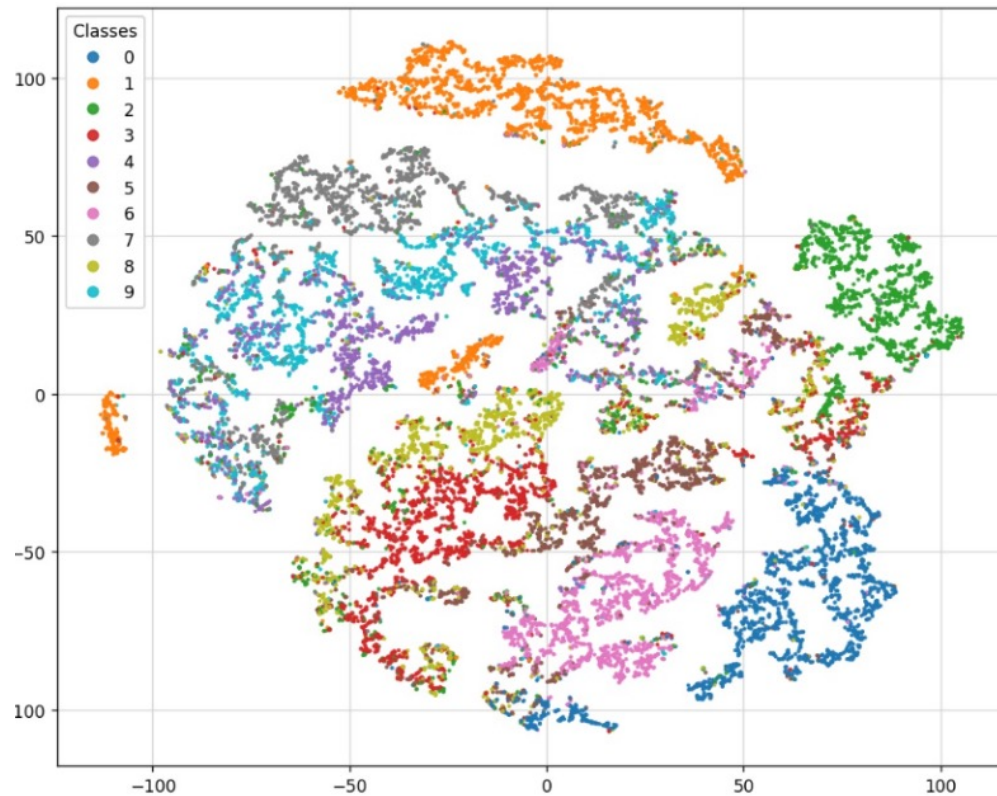
**Objectif** : Améliorer la capacité du VAE à générer chiffres avec ou sans rotation

# $\beta$ -VAE

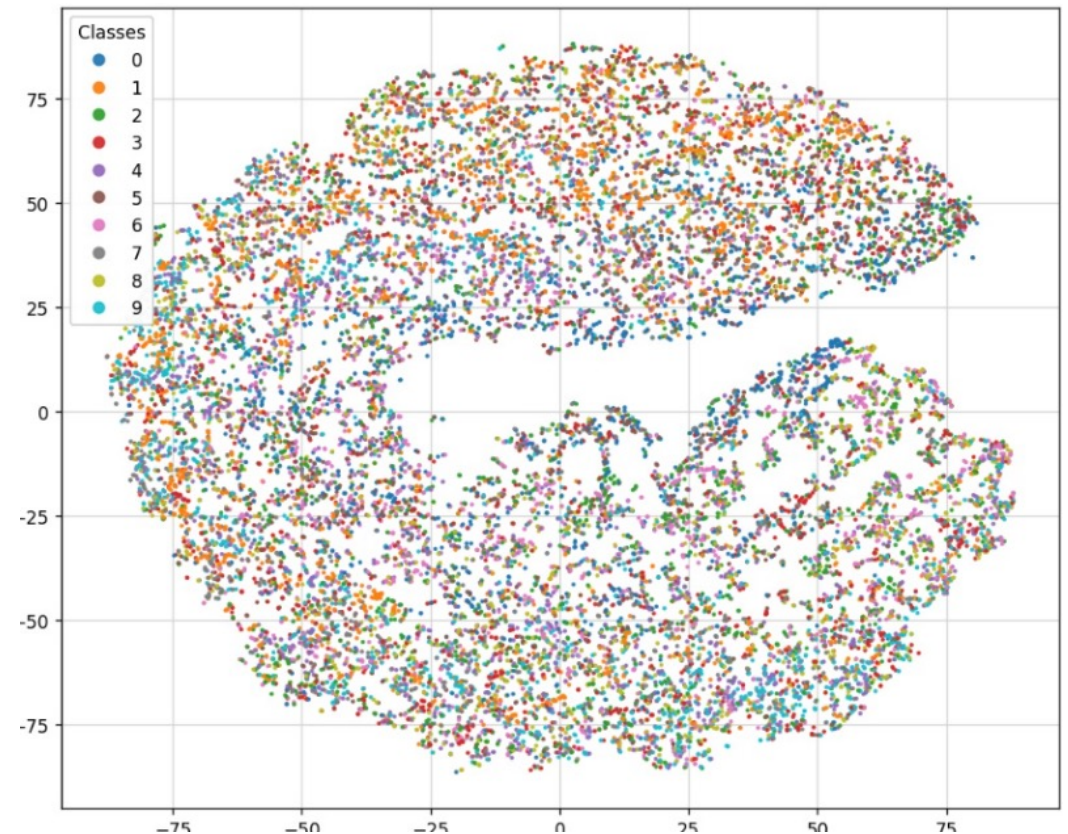
- Réguler la représentation des données dans l'espace latent du VAE
- Paramètre  $\beta$  : Contrôle l'équilibre fidélité/organisation
- Influence de  $\beta$  : Impact sur séparation des facteurs de variation
- Compromis : Précision de reconstruction vs. Séparation (ajustement de  $\beta$ )
- Valeur  $\beta$  :  $[1 ; +\infty]$
- Principe: Améliorer la séparation des facteurs de variation à l'aide de  $\beta$

# ÉTUDE : EXPLORER IMPACT DE BETA SUR NOTRE ENSEMBLE DE DONNÉES

$\beta = 4$



$\beta = 128$

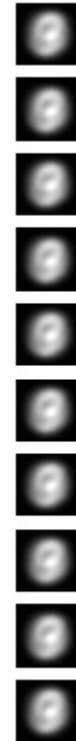


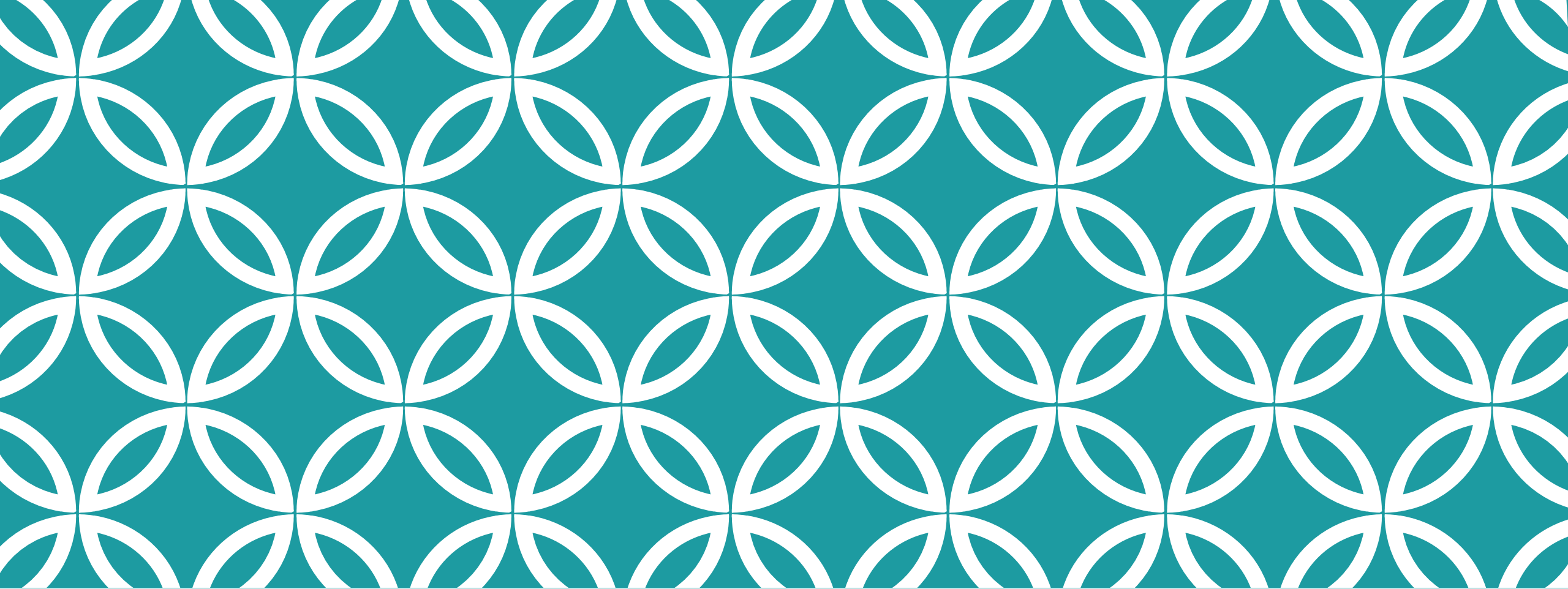
# ÉTUDE : EXPLORER IMPACT DE BETA SUR NOTRE ENSEMBLE DE DONNÉES

$\beta = 4$



$\beta = 128$



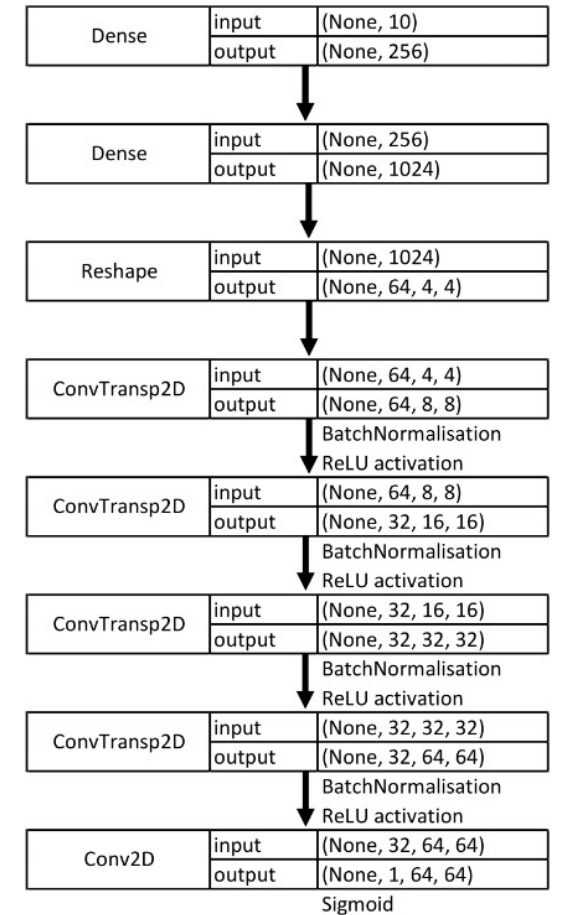
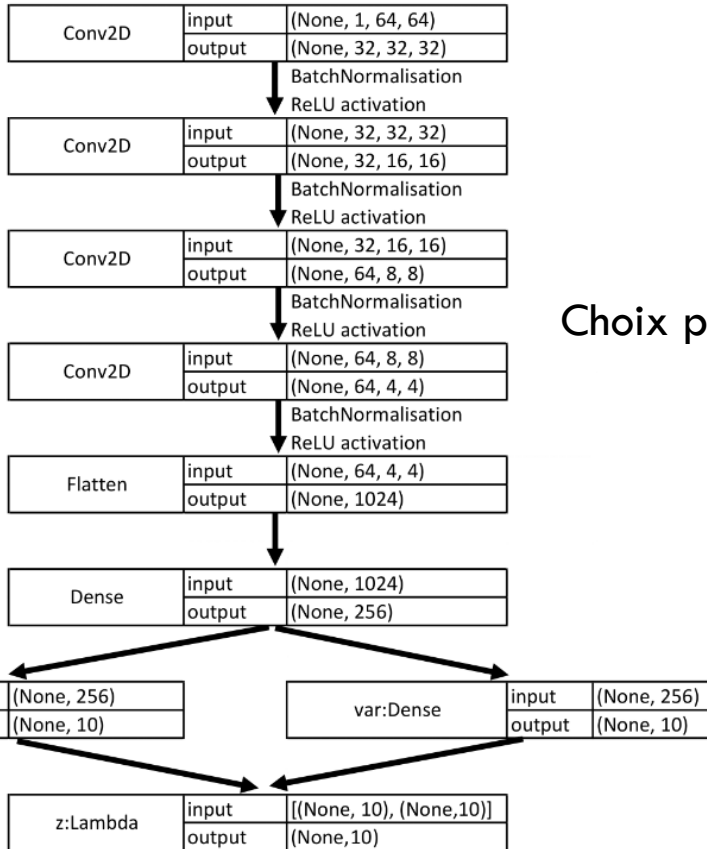


# ARCHITECTURE DU VAE



# ARCHITECTURE DE L'ARTICLE

Deux architectures VAE : MLP et CNN  
Choix pour images : Utilisation de l'architecture CNN  
4 couches (32, 32, 64, 64 channels)

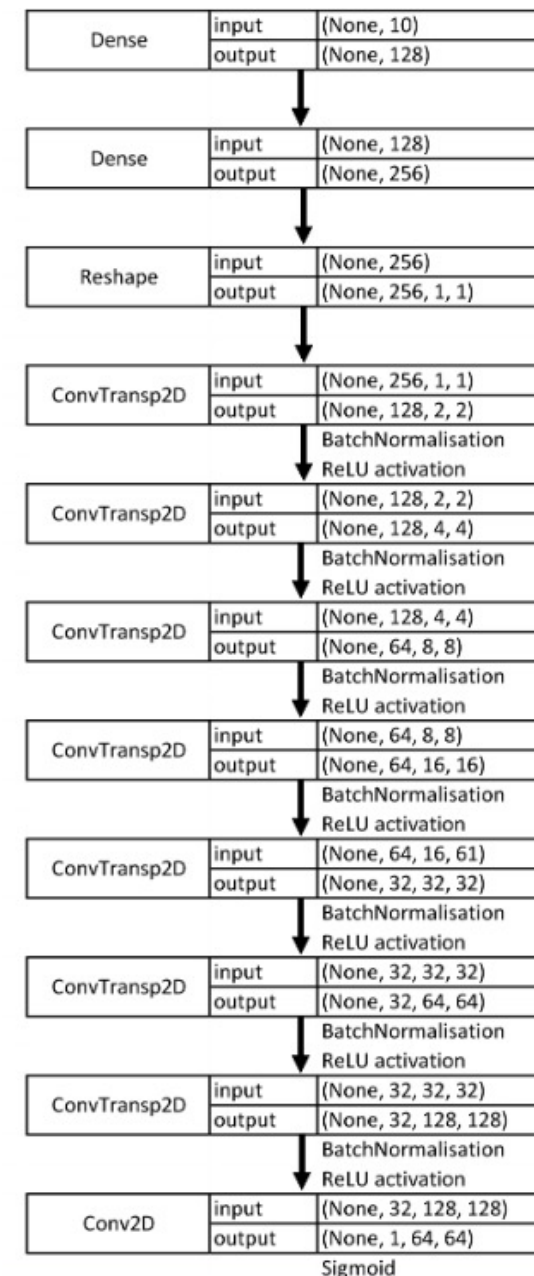
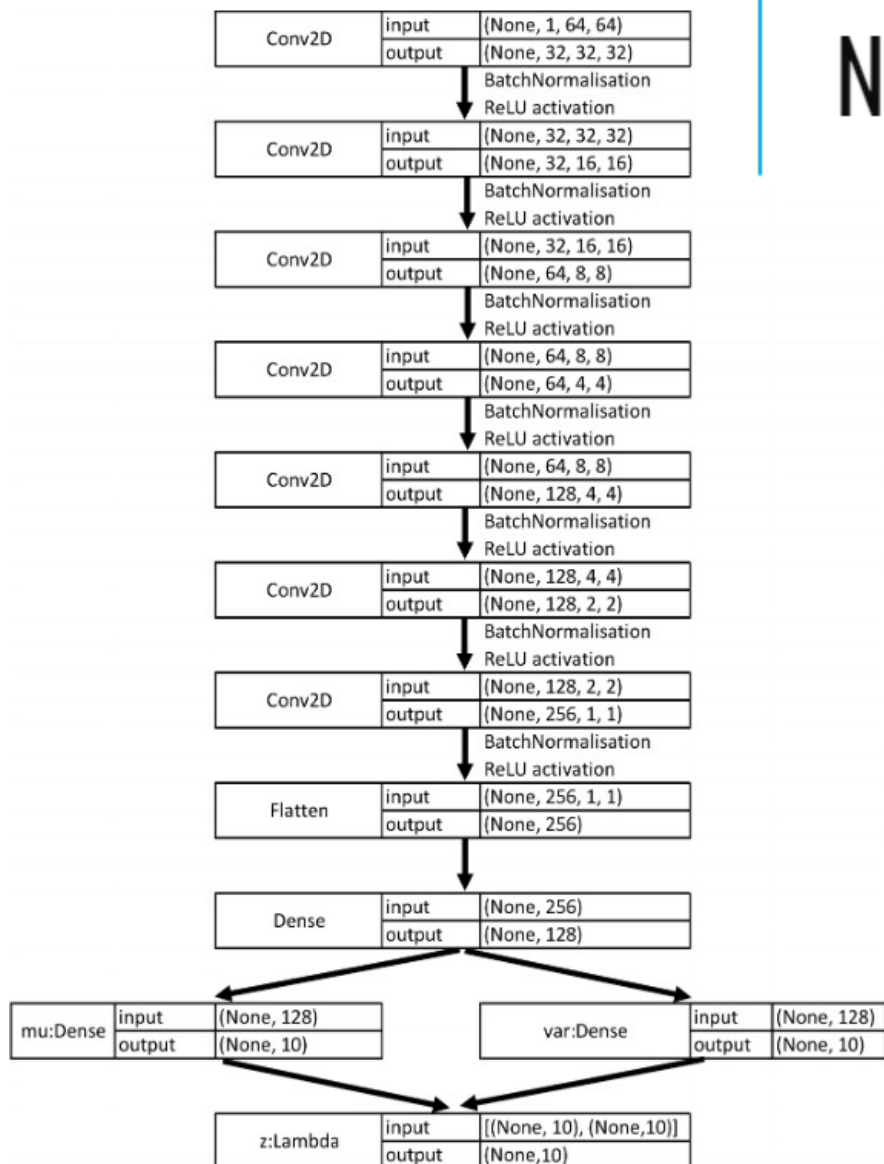


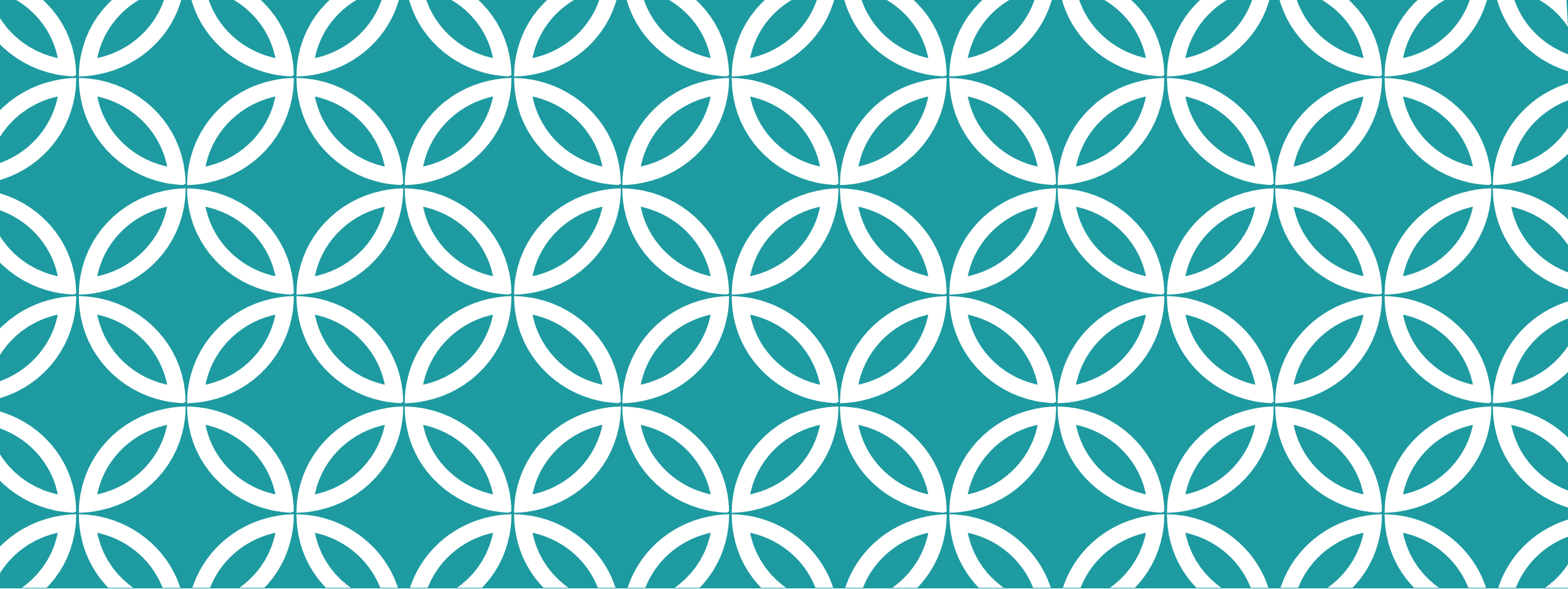


# NOTRE ARCHITECTURE

Inspirée de l'article (VAE-CNN)  
 Complexification : Ajout de couches de convolution  
 Objectif : Capturer de meilleurs facteurs de variation

- Entre chaque couche : Batch Normalization, Activation linéaire
- Encodeur : 32, 32, 64, 64, 128, 128, 256 (channels)
- Décodeur : Inverse de l'encodeur, Sigmoid à la dernière couche



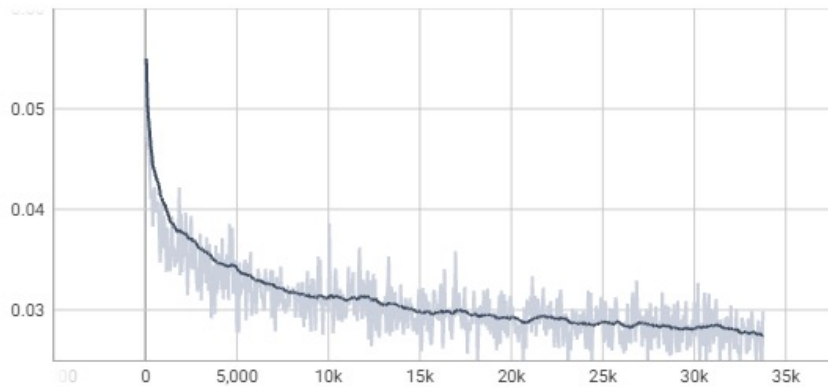


# RÉSULTAT

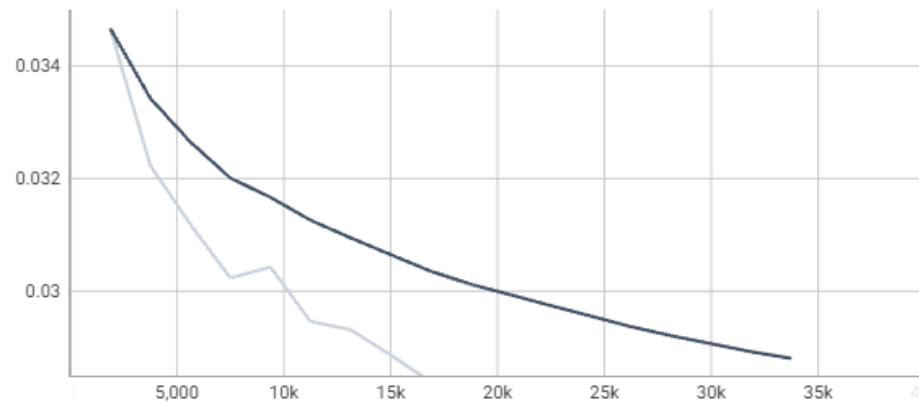
# RÉSULTATS OBTENUS AVEC NOTRE ARCHITECTURE PROPOSÉE

Pour l'entrainement:

- 15 epochs
- $\beta = 4$
- Dimension d'espace latent = 10



Training Loss

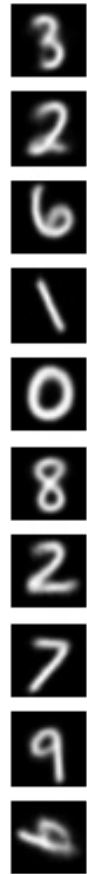


Validation Loss

# RÉSULTATS OBTENUS AVEC NOTRE ARCHITECTURE PROPOSÉE

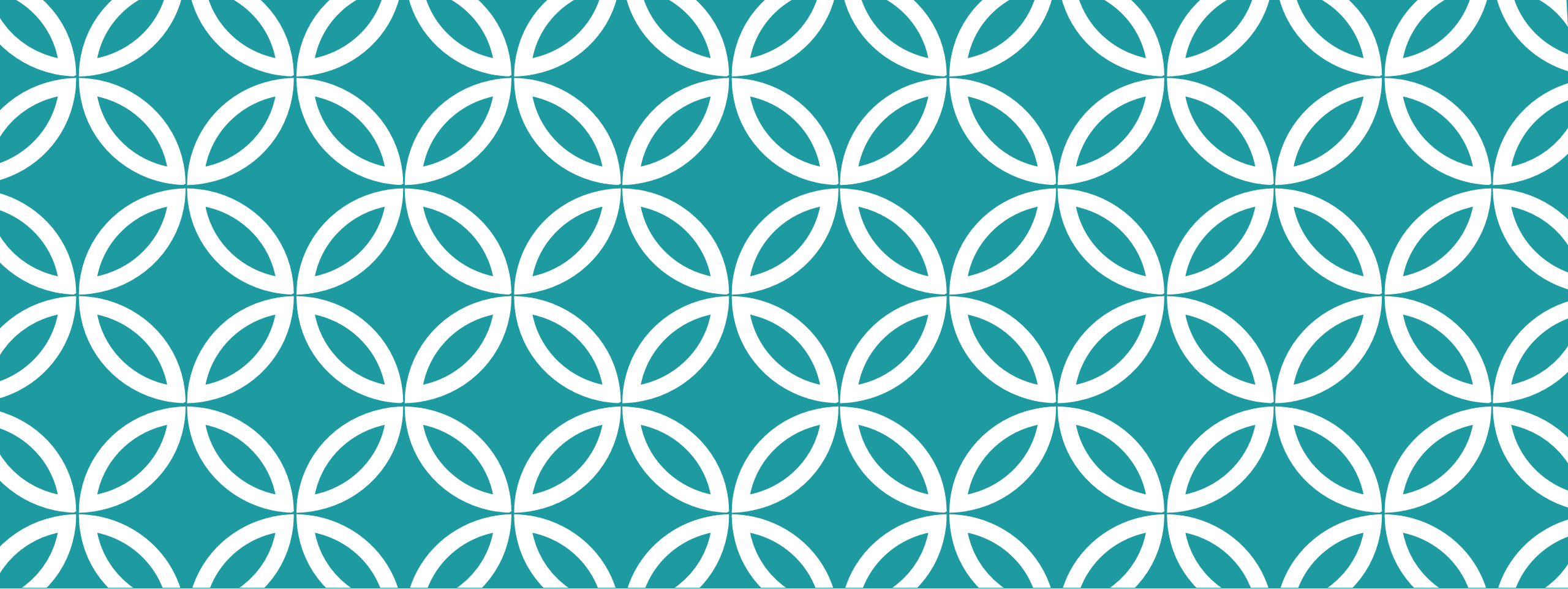


Reconstruction



Génération





# VISUALISATION DE L'ESPACE LATENT

# VISUALISATION DE L'ESPACE LATENT

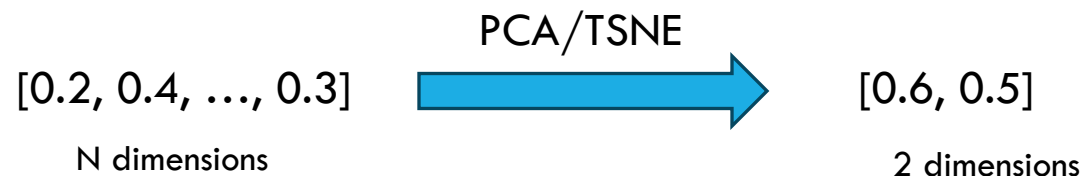
L'espace latent possède plusieurs dimensions



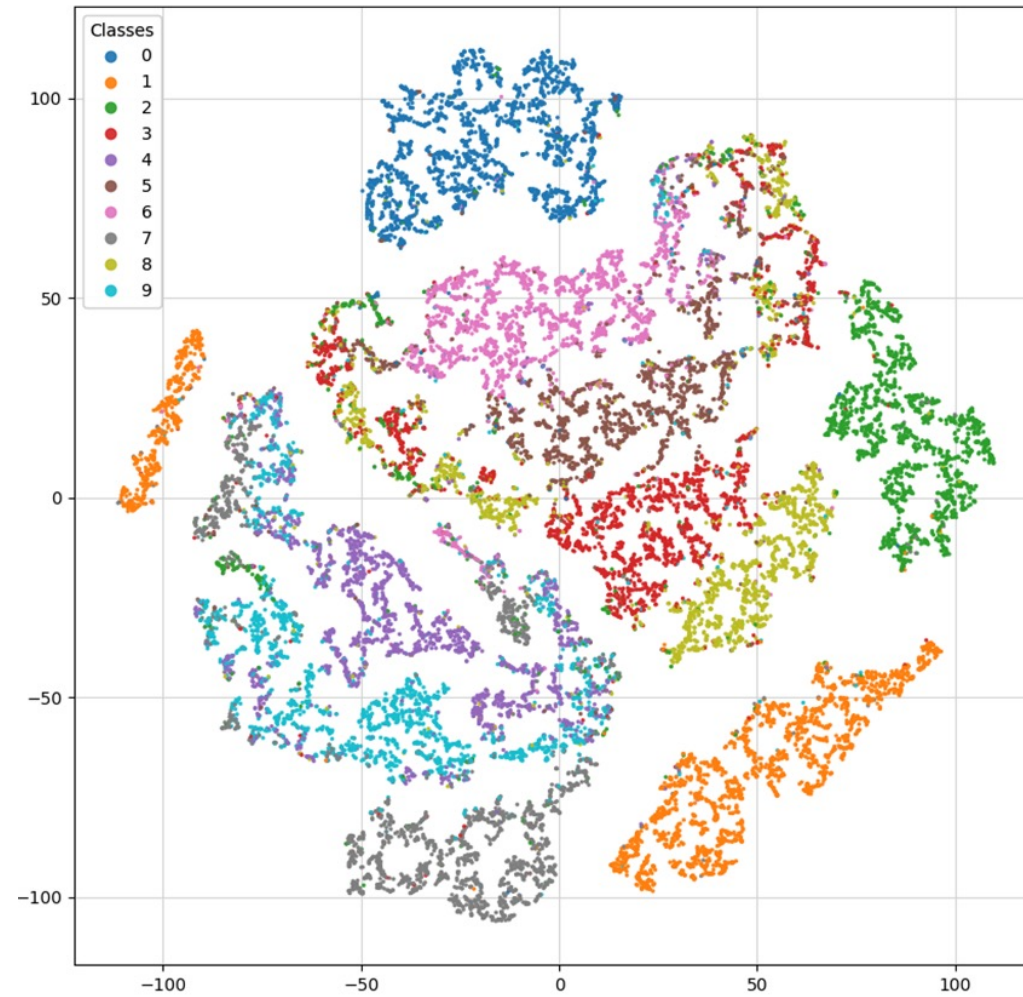
Impossible à visualiser dans un graphe



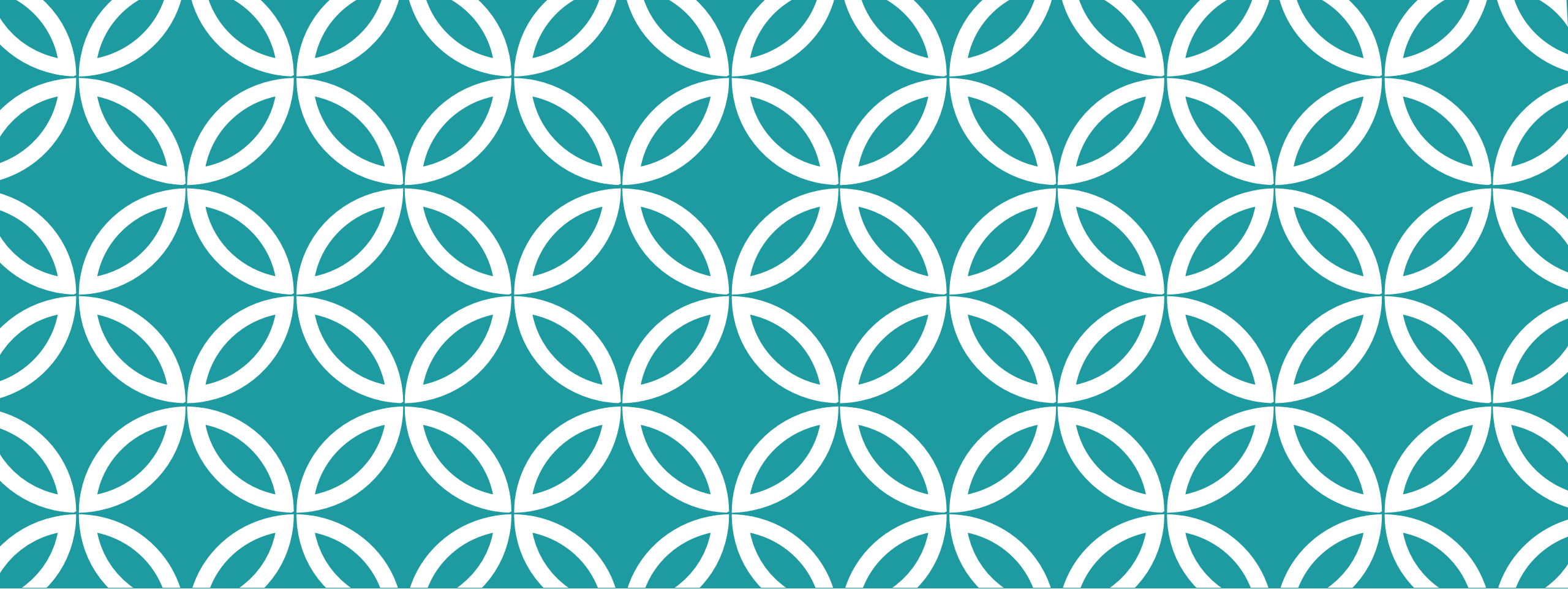
- Utilisation de la PCA ou TSNE pour réduire les dimensions à 2
- Permet de garder au maximum l'information originale en réduisant la dimension



# ESPACE LATENT



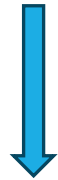




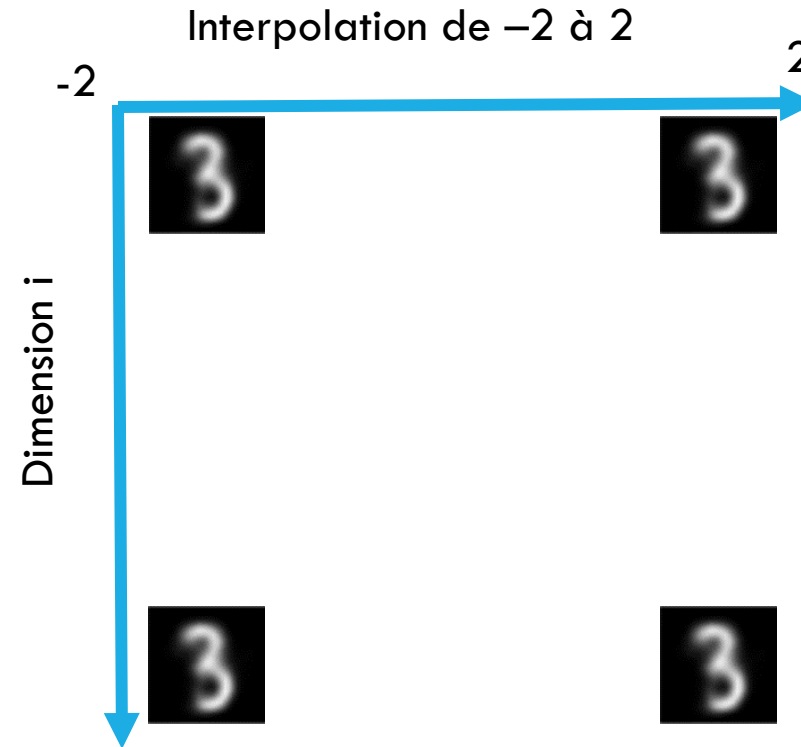
# CAPTURE DES FACTEURS DE VARIATION

# CAPTURE À PARTIR D'UN POINT ALÉATOIRE

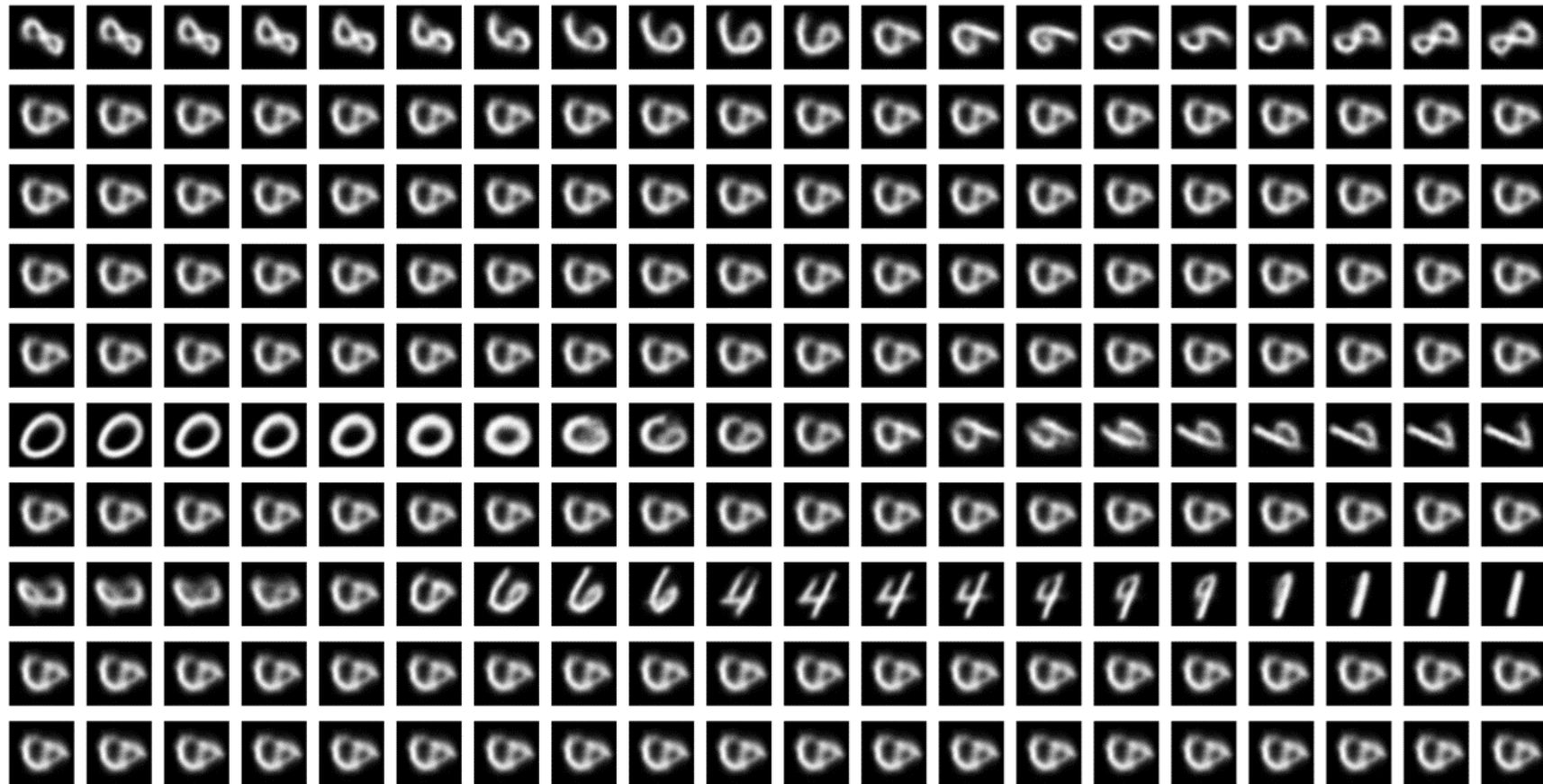
Tentative: Déterminer les dimensions de l'espace latent responsable des variations



Observer à partir d'un point aléatoire de l'espace latent l'effet de chaque dimension à intervalles réguliers

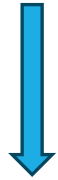


# VARIATION À PARTIR D'UN POINT ALÉATOIRE

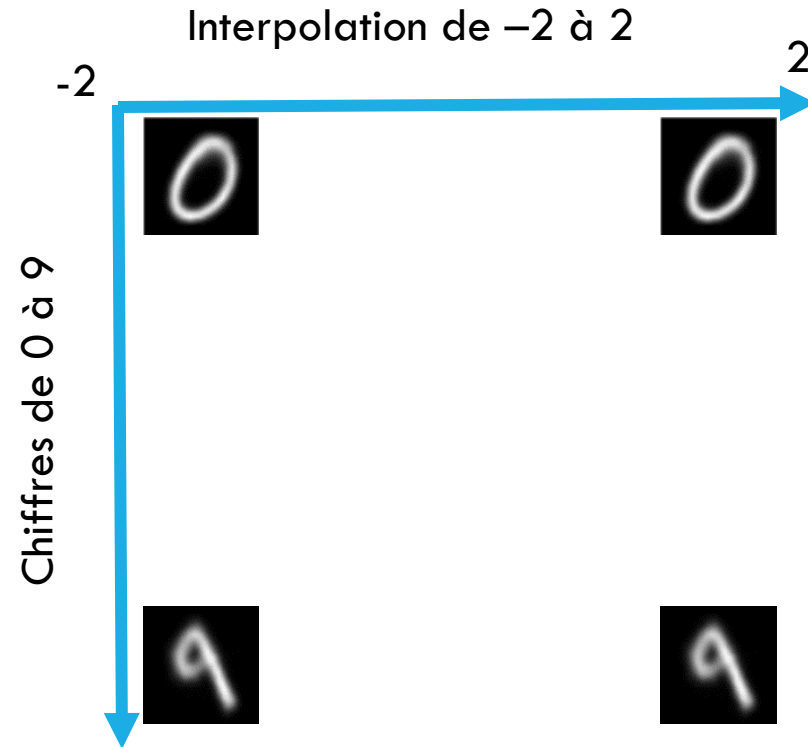


# CAPTURE À PARTIR DE L'ENCODAGE D'IMAGES

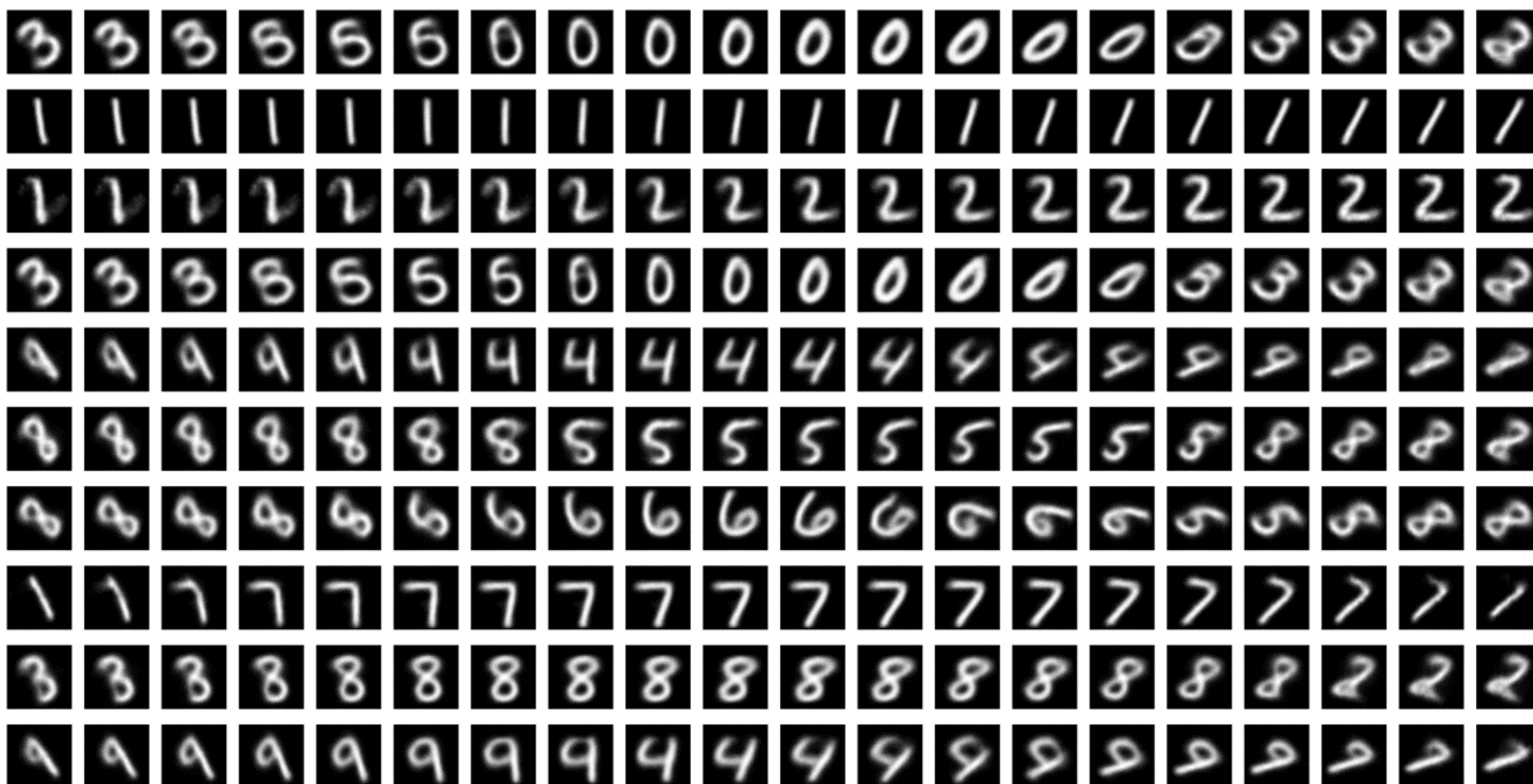
Un point aléatoire dans l'espace latent ne donne pas toujours un résultat exploitable



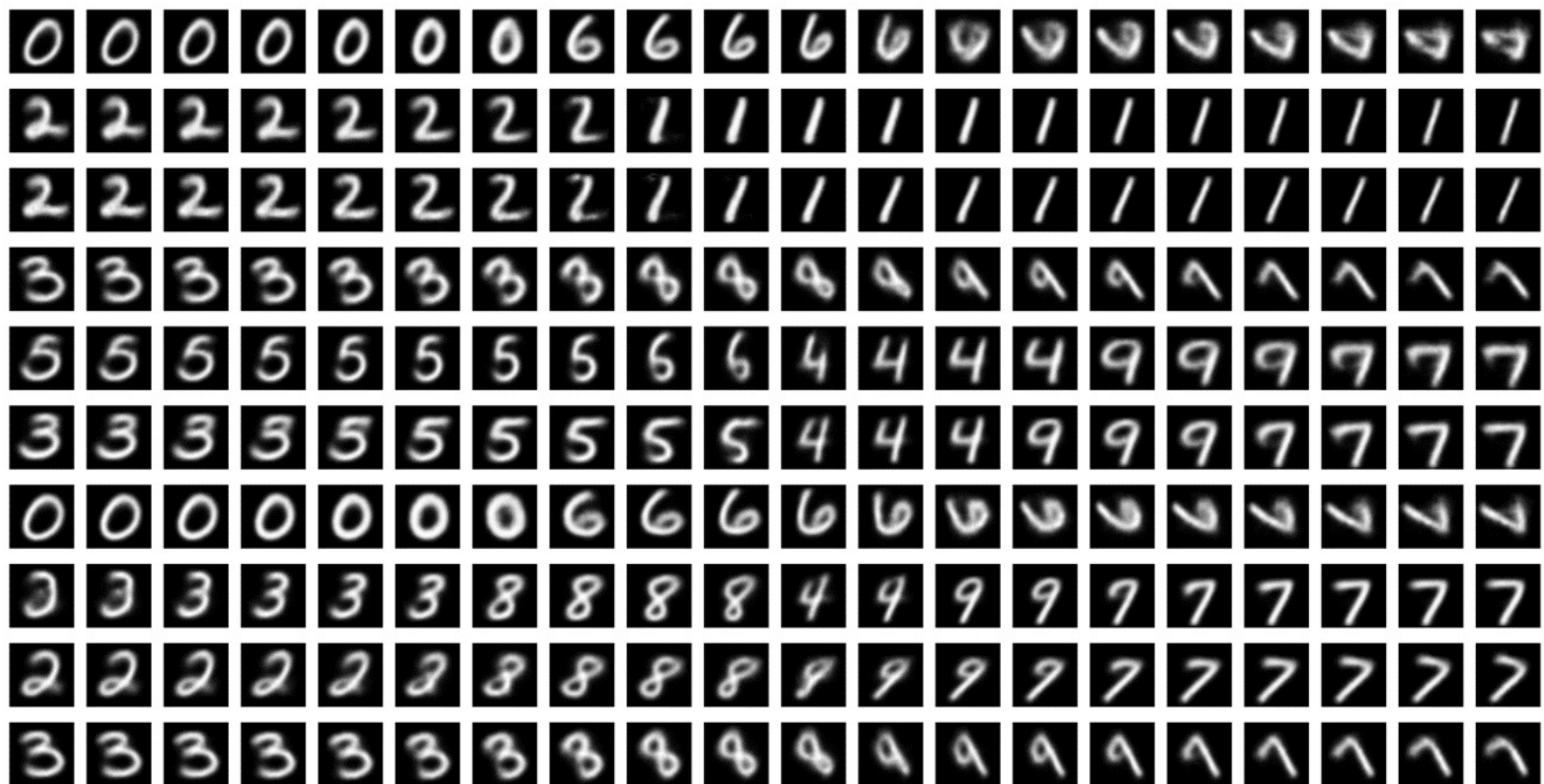
Observer à partir de l'encodage d'une image de chaque chiffre



# DIMENSION 1



# DIMENSION 6

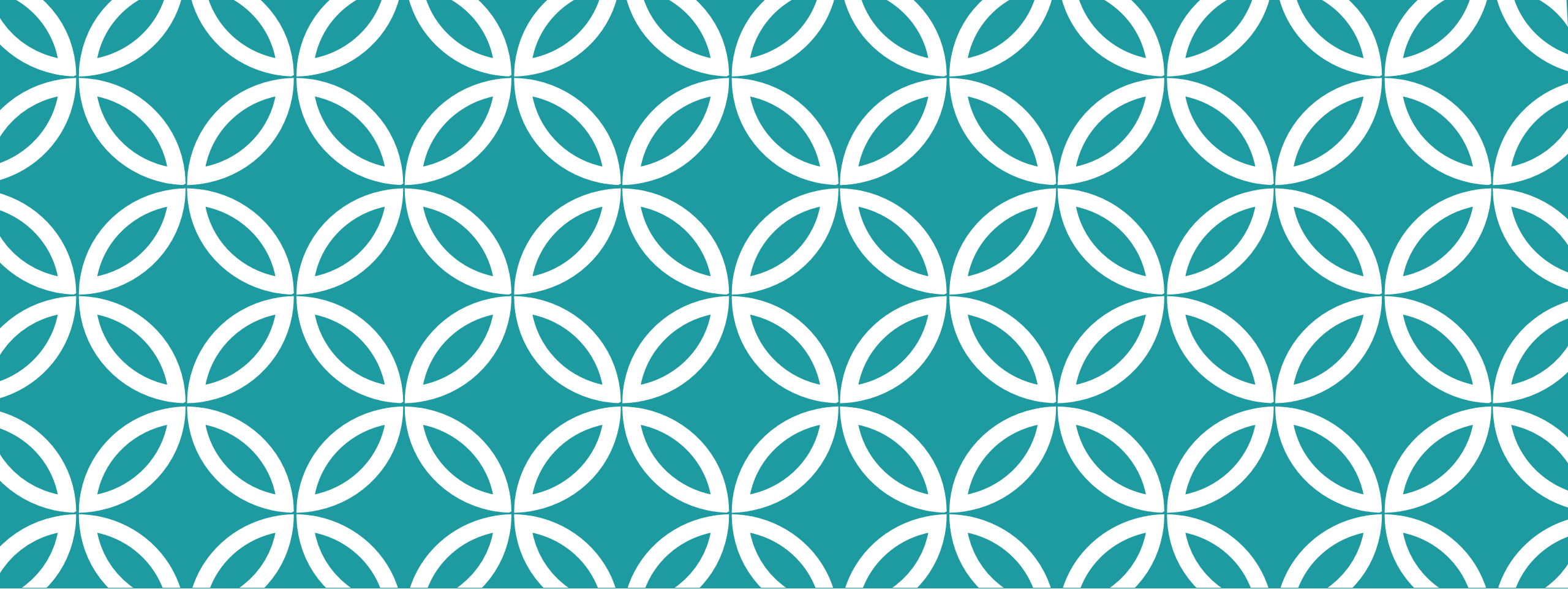


# POUR RÉSUMER

Notre architecture ne donne pas des résultats concluants très concluant

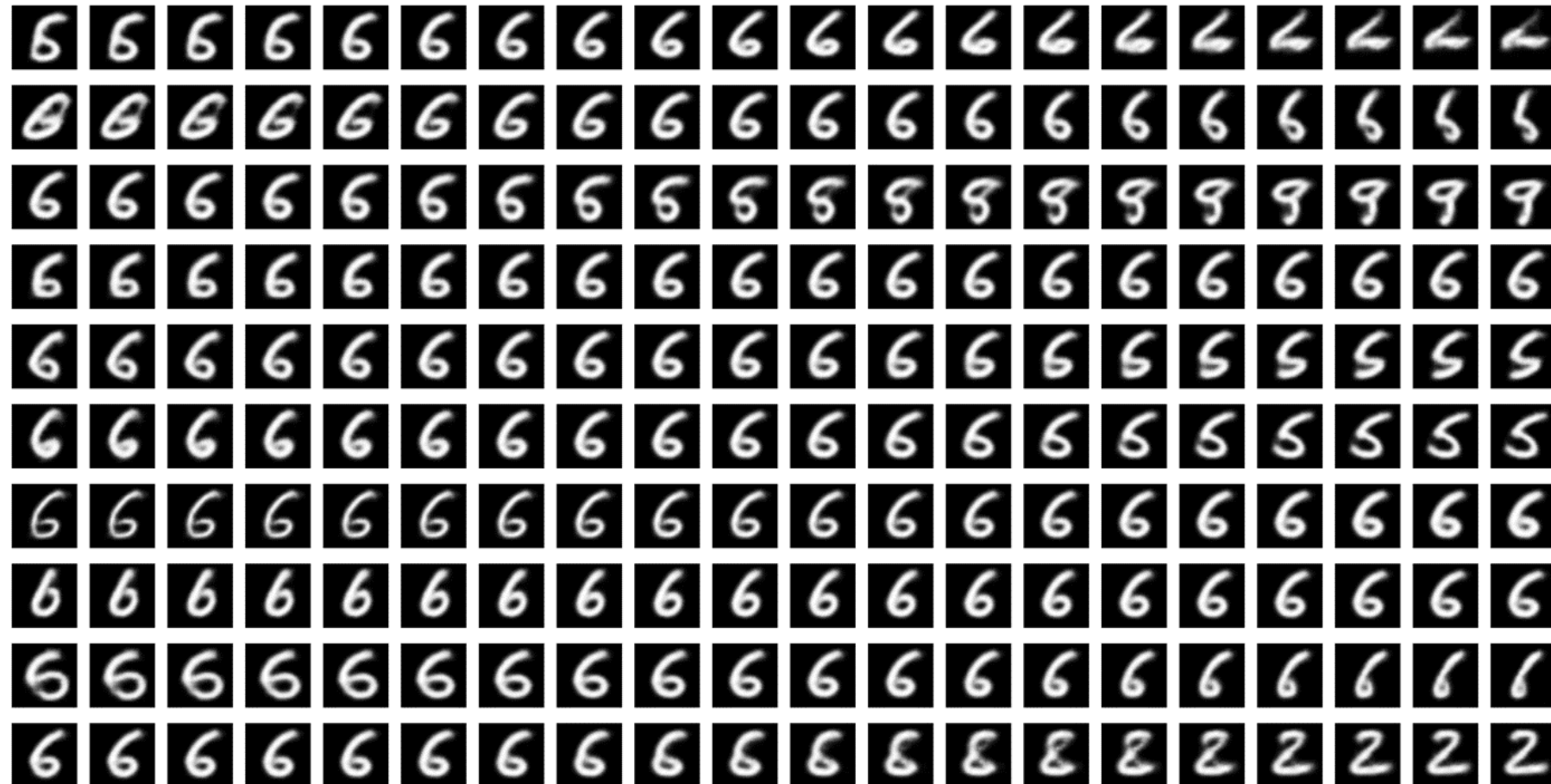
- Peu de dimensions sont responsables d'une variation (seulement la 1, 6 et 8)
- Les facteurs de variation sont liés entre eux
- Le paramètre  $\beta$  n'a pas permis de mieux les séparer





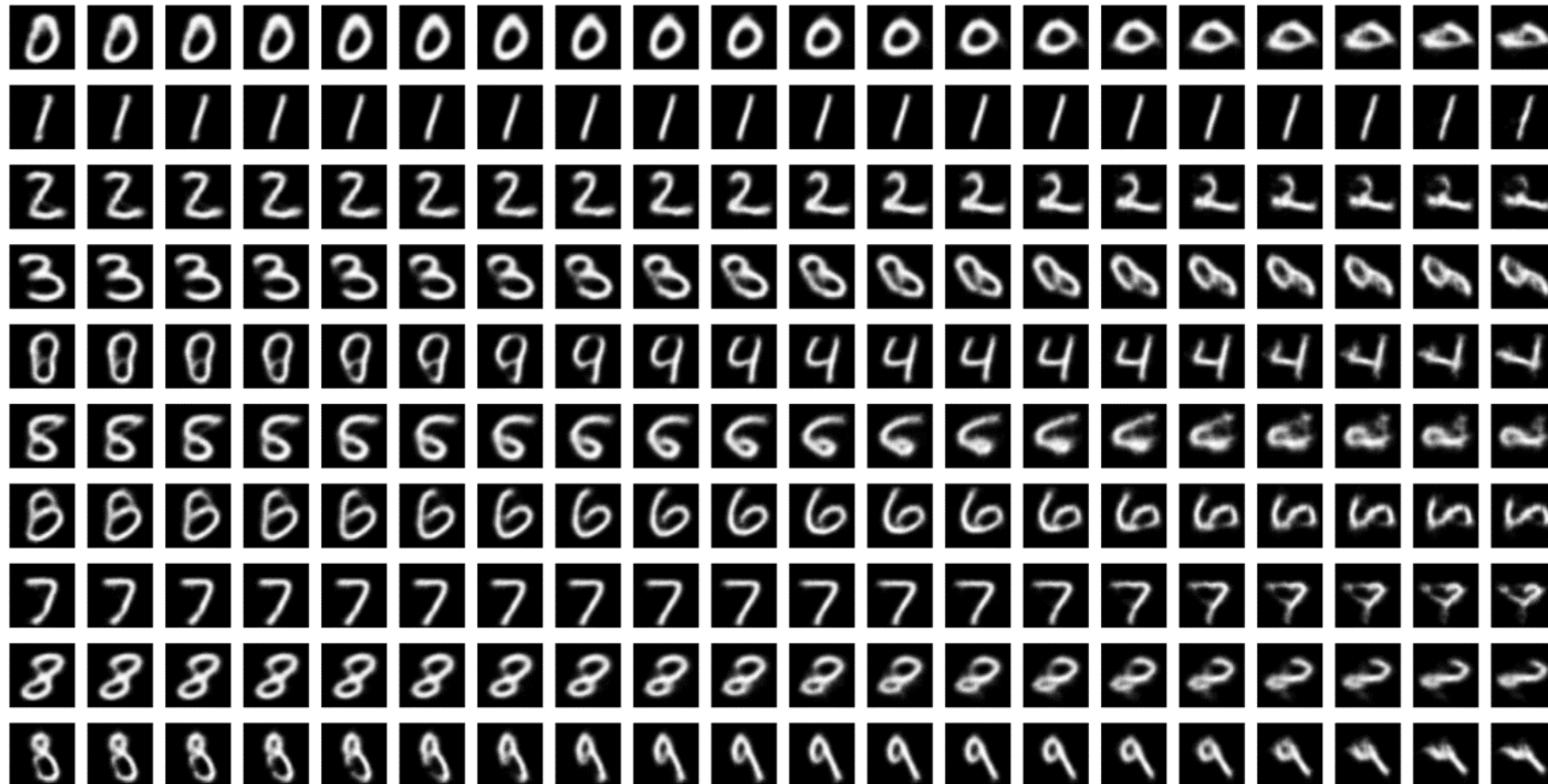
# COMPARAISON AVEC L'ARCHITECTURE DE L'ARTICLE

# VARIATION À PARTIR D'UN POINT ALÉATOIRE

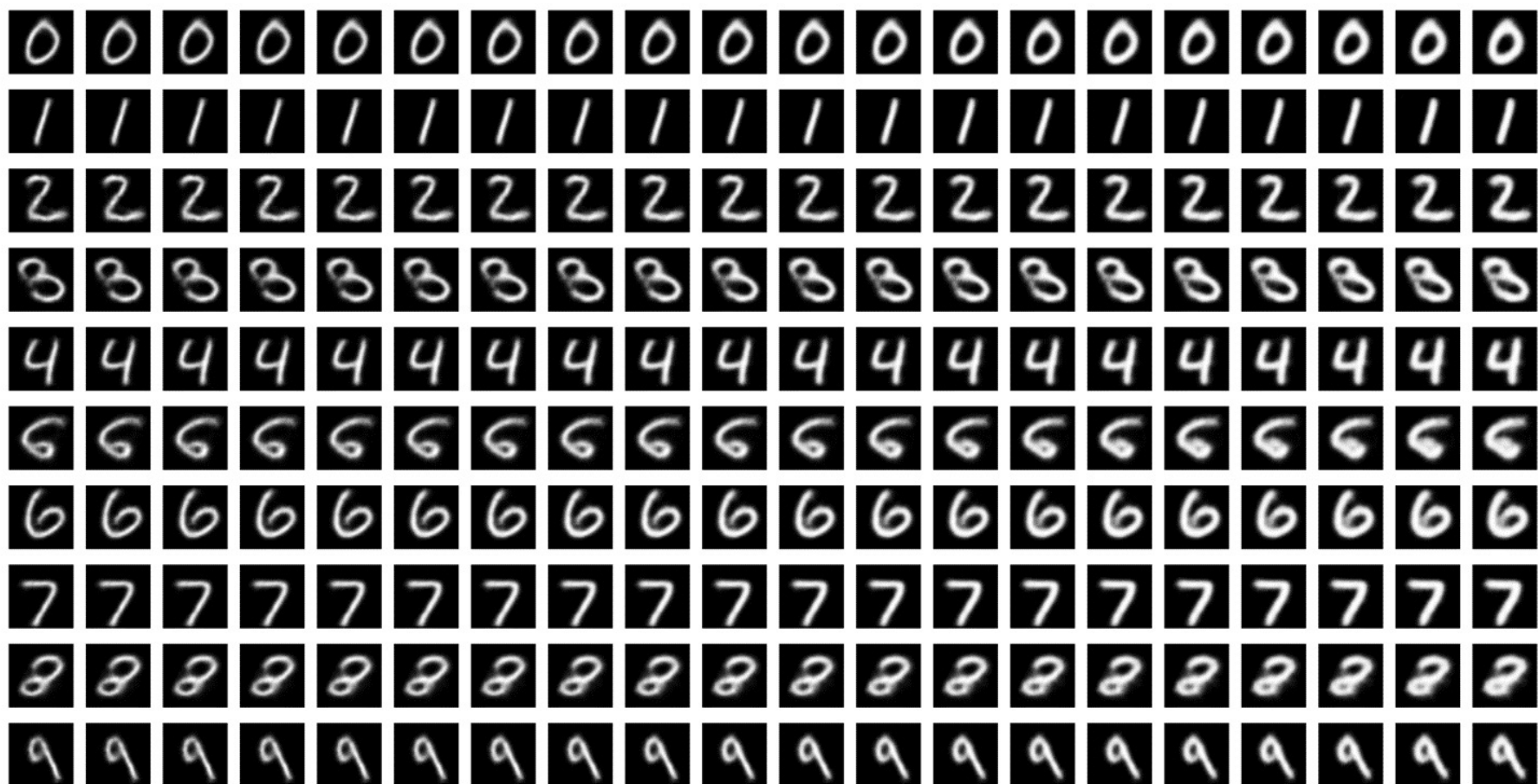


# VARIATION À PARTIR DE L'ENCODAGE D'IMAGES

## DIMENSION 1

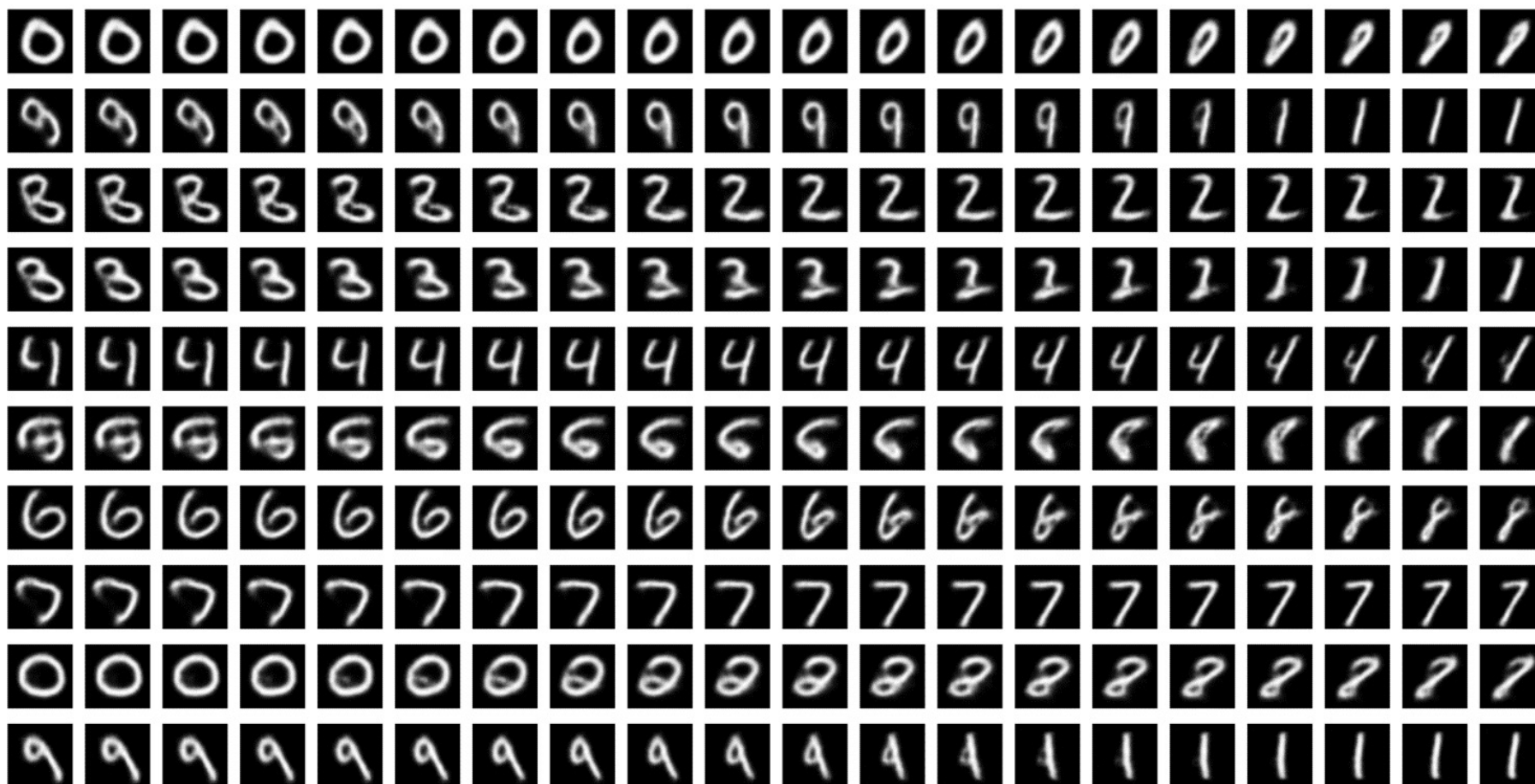


# DIMENSION 7





# DIMENSION 9

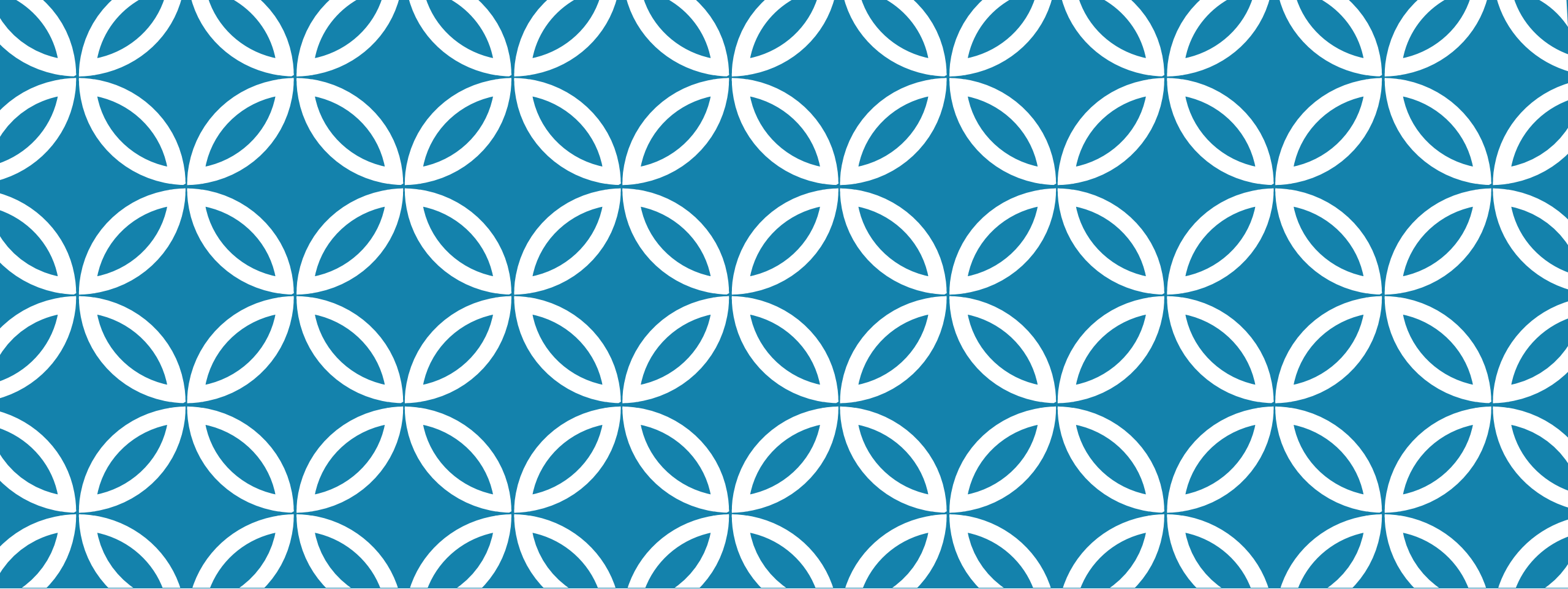


# POUR RÉSUMER

L'architecture de l'article donne de bien meilleurs résultats

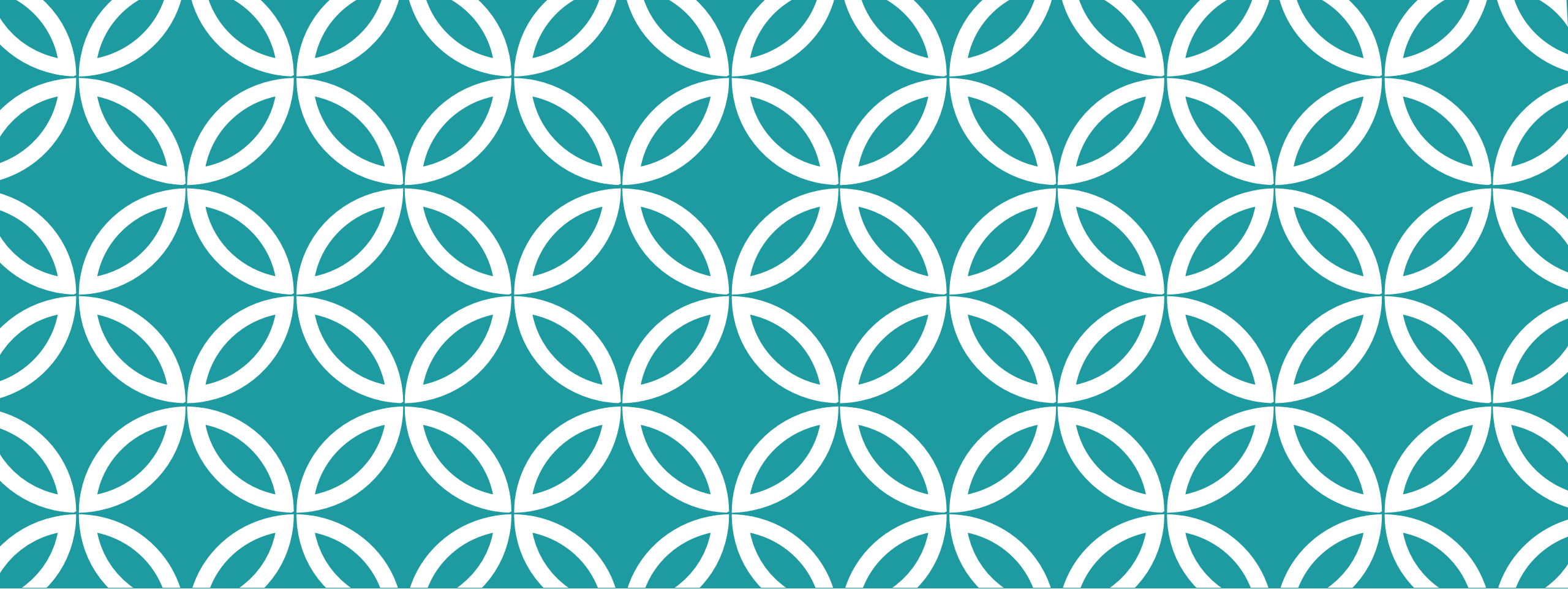
- Toutes les dimensions sont responsables d'une variation
- Les facteurs de variation sont indépendants (il y a une seule variation par dimension)
- Le paramètre  $\beta$  a bien permis de mieux séparer les facteurs de variation

Point négatif: La rotation n'a pas été capturée



**POUR CONCLURE** |





**POSSIBILITÉ D'AJOUT**

**DEMO**