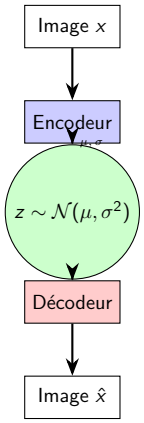
Implémentation et Analyse de VAE sur MNIST

avec Introduction aux DDPM

BELHASSAN Adam KEDDARI Bilel LEGENDRE Antoine

Master 2 - Data Science
Encadrant : Pr. RICHARD Frédéric

15 novembre 2025



Caractéristiques :

- Code latent **probabiliste**
- Deux objectifs : reconstruire **ET** générer
- Différence avec AE classique : régularisation de l'espace latent

Applications :

- Compression d'images
- Génération de données
- Détection d'anomalies

Partie exploratrice

Nous verrons également les **DDPM** (modèles de diffusion)

Problématique

Quel impact du paramètre β sur le compromis reconstruction/régularisation ?

Trois axes d'étude :

1. Influence de β : 10^{-3}, 1, 20
2. Comparaison **Conv vs Dense**
3. Effet de la dimension latente : 2 vs 4

Dataset : MNIST

- 60,000 images 28×28
- 10 classes (chiffres 0-9)
- Split : 80% train / 20% test

Plan de présentation :

- VAE (cœur de l'étude)
- DDPM

Formulation

$$\mathcal{L} = \text{BCE} + \beta \cdot \text{KLD}$$

Binary Cross-Entropy (BCE) :

$$\text{BCE} = -\sum_i [x_i \log(\hat{x}_i) + (1 - x_i) \log(1 - \hat{x}_i)]$$

- Mesure la qualité de reconstruction
- Plus elle est faible, meilleure est la reconstruction

Kullback-Leibler Divergence (KLD) :

$$\text{KLD} = -\frac{1}{2} \sum_j \left[1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2\right]$$

- Régularise l'espace latent vers $\mathcal{N}(0, 1)$
- Permet la génération et structure l'espace

Paramètre β

Contrôle l'équilibre entre reconstruction et régularisation

Reparameterization trick : $z = \mu + \sigma \odot \epsilon$ avec $\epsilon \sim \mathcal{N}(0, I)$

Présentation :

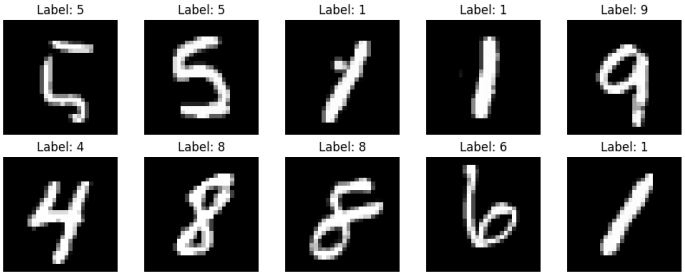
- 60,000 chiffres manuscrits
- 10 classes (0-9)
- Images 28×28 pixels
- Niveaux de gris

Prétraitement :

- Normalisation dans $[0, 1]$
- Division par 255
- Compatible avec Sigmoid

Split :

- 80% entraînement (48,000)
- 20% test (12,000)



Exemples d'images du dataset MNIST

Architecture :

Encodeur :

- Conv2D(32 filtres, 3×3, stride=2)
- Conv2D(64 filtres, 3×3, stride=2)
- Flatten + FC $\rightarrow (\mu, \log \sigma^2)$

Décodeur :

- FC $\rightarrow$ Reshape (64, 7, 7)
- Deconv2D(32 filtres, 3×3)
- Deconv2D(1 filtre, 3×3, Sigmoid)

Caractéristiques

- **59,525 paramètres** (compact)
- Préserve la structure spatiale
- Architecture symétrique
- Convolutions pour features locales

Avantages

- Efficacité computationnelle
- Préservation d'information spatiale
- Moins de paramètres que Dense

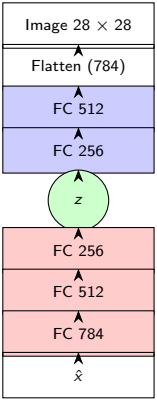
Architecture :

Encodeur :

- Flatten : $28 \times 28 \rightarrow 784$
- FC(512) + ReLU
- FC(256) + ReLU
- FC $\rightarrow (\mu, \log \sigma^2)$

Décodeur :

- FC(256) + ReLU
- FC(512) + ReLU
- FC(784) + Sigmoid
- Reshape $\rightarrow 28 \times 28$



Critère	VAE Conv	VAE Dense
Paramètres	59,525	1,068,820 (18×)
Structure spatiale	Préservée	Perdue
Reconstruction (BCE)	Plus élevée	Plus faible
Espace latent	Plus structuré	Plus de chevauchement
Vitesse d'entraînement	Rapide	Lente
Complexité	Faible	Élevée

Point clé

Le VAE Dense reconstruit mieux (BCE plus faible) mais nécessite **18 fois plus de paramètres** que le VAE Conv.

- **Conv** : meilleur compromis efficacité/performance
- **Dense** : meilleure fidélité de reconstruction

Hyperparamètres :

Paramètre	Valeur
Époques	20
Batch size	64
Optimizer	Adam
Learning rate	10^{-3}
Dimension latente	2 ou 4
β testé	10^{-3}, 1, 20

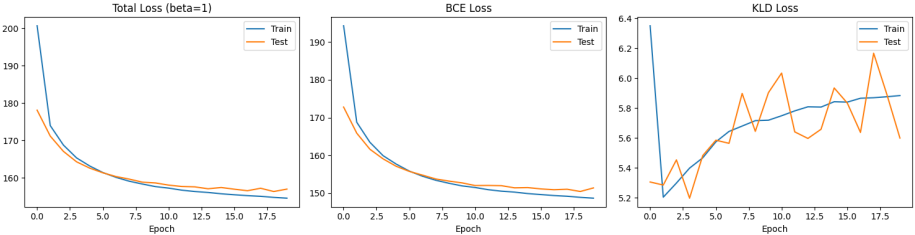
Métriques d'évaluation :

- **BCE** : qualité de reconstruction
- **KLD** : régularisation latente
- **Visualisations** :
  - Images reconstruites
  - Images générées
  - Espace latent 2D

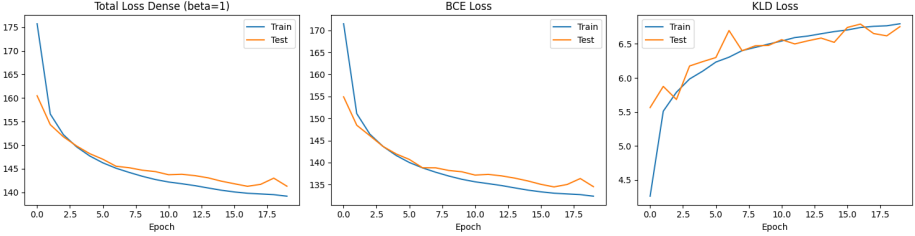
Objectif

Analyser l'impact de β sur le compromis reconstruction/régularisation pour les deux architectures (Conv et Dense).

VAE Convolutionnel



VAE Dense



VAE Convolutionnel

VAE Dense

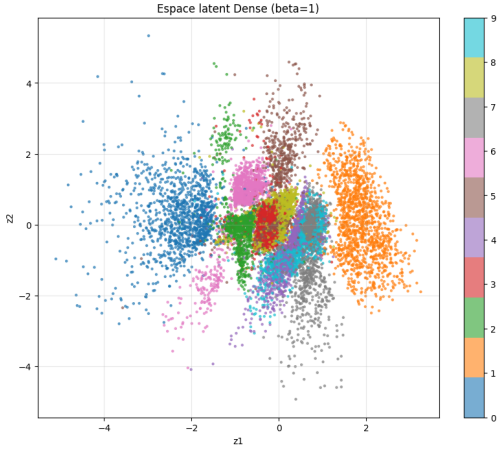
Analyse comparative

- **Dense** : reconstructions plus fidèles, détails mieux préservés
- **Conv** : légèrement plus floues mais structure spatiale préservée
- Bon équilibre reconstruction/régularisation pour $\beta = 1$

Espace Latent avec $\beta = 1$

VAE Convolutionnel

VAE Dense



VAE Dense

Images générées depuis l'espace latent Dense (beta=1)

VAE Convolutionnel

Images générées depuis l'espace latent

Impact de $\beta = 10^{-3}$: Régularisation Faible

Contexte théorique

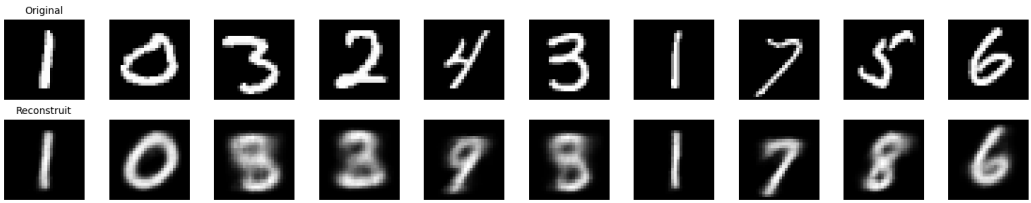
Avec $\beta = 10^{-3}$, la composante KLD devient négligeable : $\mathcal{L} \approx \text{BCE} + 0.001 \cdot \text{KLD}$

Conséquences attendues :

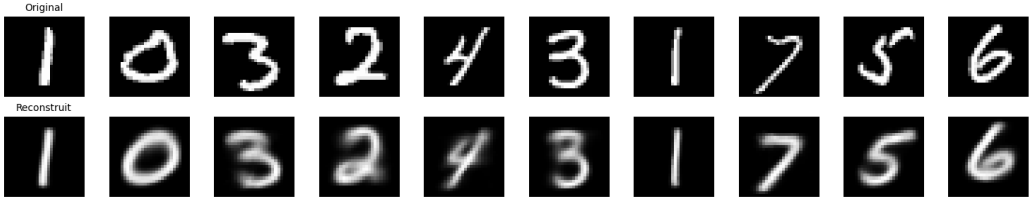
- Reconstructions excellentes (priorité à la fidélité)
- Espace latent dispersé (régularisation minimale)
- Génération moins contrôlée (perte de structure)

Analyse en 3 étapes : Reconstructions $\rightarrow$ Espace latent $\rightarrow$ Générations

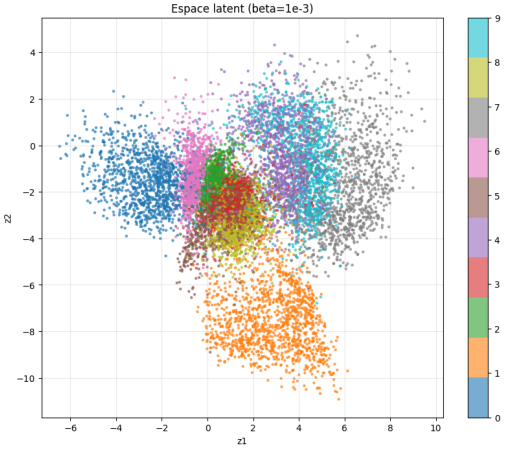
VAE Convolutionnel



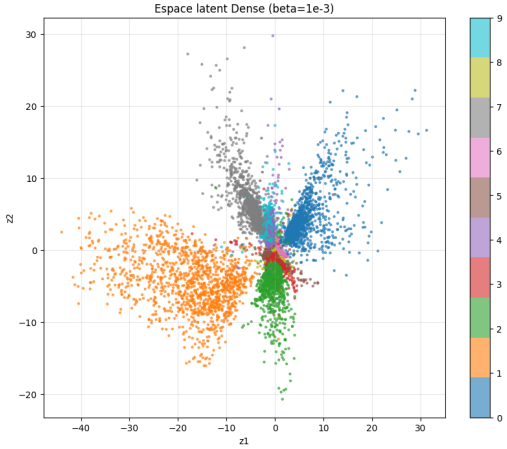
VAE Dense



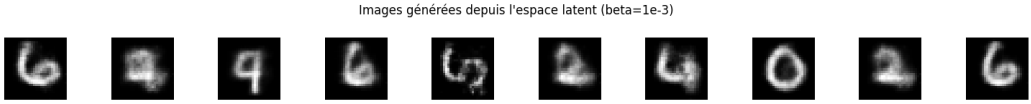
VAE Convolutionnel



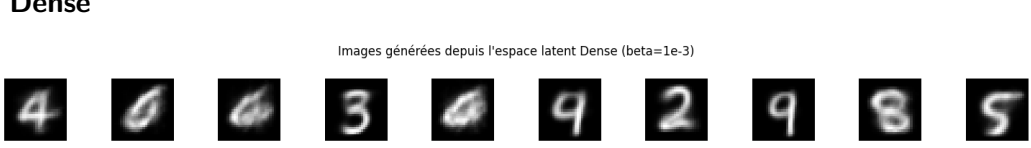
VAE Dense



VAE Convolutionnel



VAE Dense



Conséquence logique

Générations **variables et imprévisibles** dues à l'espace latent désorganisé. Confirme l'importance de la régularisation pour la génération.

Contexte théorique

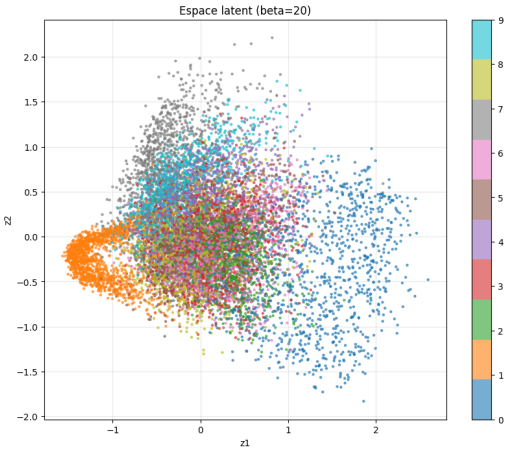
Avec $\beta = 20$, la régularisation domine : $\mathcal{L} = \text{BCE} + 20 \cdot \text{KLD}$

Conséquences attendues :

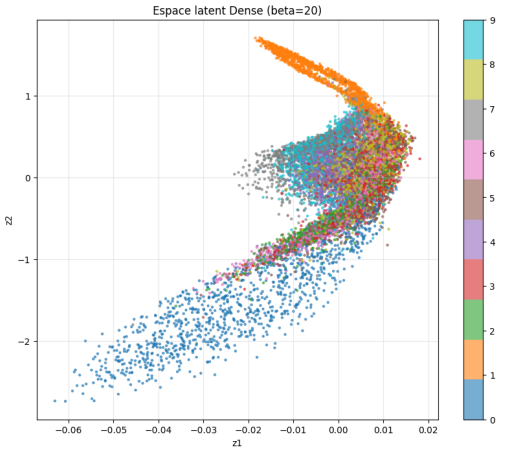
- Espace latent sur-compacté (contrainte excessive vers $\mathcal{N}(0, 1)$)
- Reconstructions dégradées (perte de détails)
- Générations lisses mais floues (manque de diversité)

Analyse en 3 étapes : Reconstructions $\rightarrow$ Espace latent $\rightarrow$ Générations

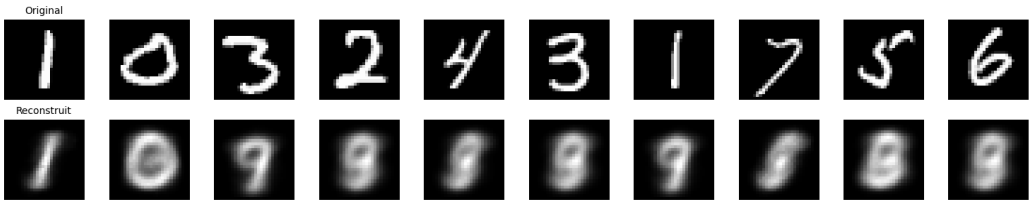
VAE Convolutionnel



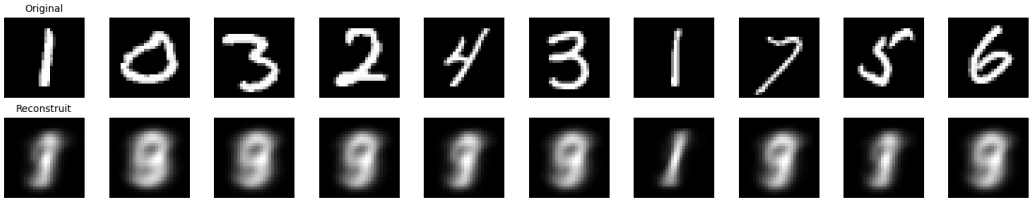
VAE Dense



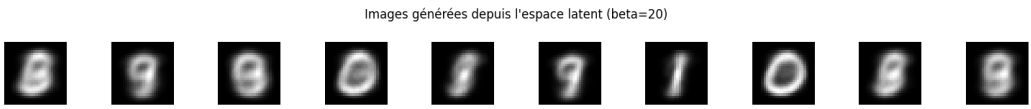
VAE Convolutionnel



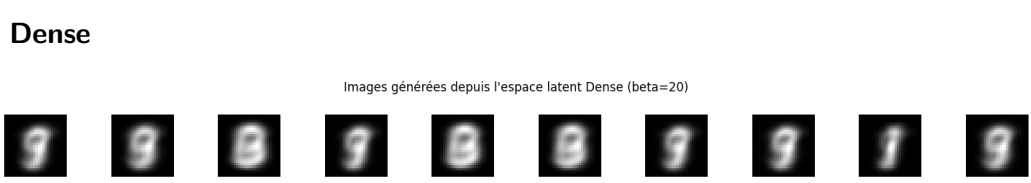
VAE Dense



VAE Convolutionnel



VAE Dense



Conséquence logique

Générations **uniformes et floues** - manque de diversité et de netteté dus à l'espace latent sur-contraint.

Critère	$\beta = 10^{-3}$	$\beta = 1$	$\beta = 20$
Reconstruction	Excellente	Bonne	Floue
Structure latente	Dispersée	Structurée	Trop compacte
Séparation classes	Faible	Bonne	Mauvaise
Images générées	Variable	Cohérentes	Lisses/floues
Cas d'usage	Compression	Équilibre	Régularisation excessive

Conclusion sur β

- β trop faible : privilégie la reconstruction au détriment de l'organisation latente
- β trop fort : sur-compacte l'espace et détériore les images
- $\beta = 1$ **est optimal** : compromis robuste entre reconstruction et régularisation

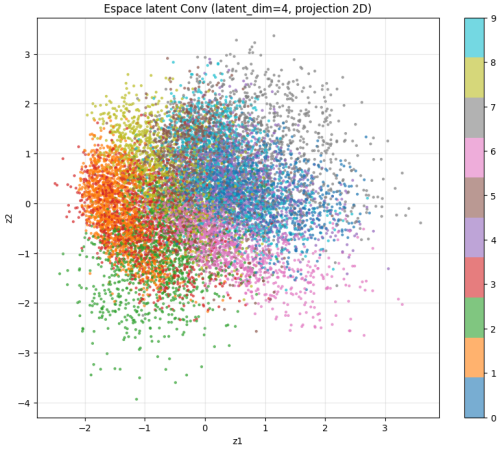
VAE Conv, $\beta = 1$, dimension latente = 4

Améliorations observées :

- Détails mieux préservés
- Moins de flou
- Netteté accrue

Résultat clé :

- Conv (dim=4) $\gg$ Dense (dim=2)
- Avec 18$\times$ moins de paramètres
- Meilleur compromis efficacité/performance



Projection 2D de l'espace latent 4D

Notre choix : VAE Convolutionnel

Nous recommandons le **VAE Convolutionnel** pour les raisons suivantes :

Avantages et inconvénients :

En dim=4 : qualité de reconstruction accrue.

18× moins de paramètres

Plus rapide à entraîner

Espace latent mieux structuré

Préserve la structure spatiale

Répartition des paramètres

- Conv : 59,525 (5.3%)
- Dense : 1,068,820 (94.7%)

Conclusion

Le VAE Conv offre le **meilleur compromis efficacité/performance** pour MNIST.

Applications générales :

1. **Génération** : créer de nouveaux chiffres réalistes
2. **Interpolation** : morphing entre classes
3. **Compression** : réduction de dimensionnalité
4. **Détection d'anomalies** : applications médicales

Principe de détection :

- Entraîner sur données **saines**
- Détecter anomalies par erreur de reconstruction élevée
- Applicable en imagerie médicale

Datasets disponibles :

- ChestX-ray14 (radiographies)
- ISIC (mélanomes)
- NIH Chest X-ray

Denoising Diffusion Probabilistic Models

Approche **différente des VAE** : génération par débruitage progressif

Principe :

1. **Forward** (diffusion directe) :
   Ajouter du bruit gaussien progressivement

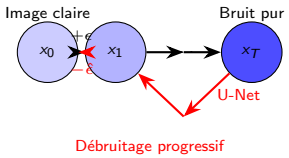
$$x_0 \to x_1 \to \cdots \to x_T$$

   Image claire $\to$ Bruit pur

2. **Reverse** (diffusion inverse) :
   Apprendre à retirer le bruit étape par étape

$$x_T \to x_{T-1} \to \cdots \to x_0$$

   Bruit pur $\to$ Image générée



Architecture :

- U-Net (3.3M paramètres)
- Prédit le bruit ϵ à chaque étape
- Conditionné sur le timestep t

Forward Process (diffusion directe) :

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

- Ajoute du bruit gaussien à chaque étape
- β_t contrôle la quantité de bruit
- Propriété : peut calculer x_t directement depuis x_0

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I)$$

avec $\alpha_t = 1 - \beta_t$ et $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

Reverse Process (génération) :

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

- U-Net prédit le bruit $\epsilon_\theta(x_t, t)$
- Retire le bruit progressivement
- Formule de sampling :

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z$$

Loss Function

$$\mathcal{L} = \|\epsilon - \epsilon_\theta(x_t, t)\|^2$$

MSE entre le bruit réel et le bruit prédit par le U-Net

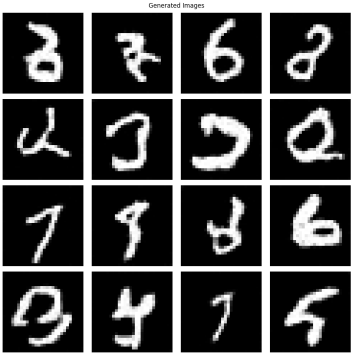
Hyperparamètres :

- 20 époques
- 4000 timesteps
- Cosine schedule pour β_t
- Batch size : 128
- Adam optimizer (lr=10^{-4})

Convergence :

- Train Loss : **0.0278**
- Test Loss : **0.0280**
- Convergence rapide (5 premières époques)
- Pas de surapprentissage

Images générées



Échantillons générés par DDPM

Qualité :

- Chiffres reconnaissables

Critère	VAE	DDPM
Principe	Encodeur/Décodeur	Débruitage progressif
Espace latent	Structuré (interprétable)	Pas d'espace latent explicite
Qualité génération	Bonne	Très bonne
Vitesse génération	Rapide (1 pass)	Lente (T passes)
Reconstruction	Oui	Non directement
Contrôle latent	Facile	Difficile
Entraînement	Stable	Stable
Applications	Compression, anomalies	Génération pure

Complémentarité

- **VAE** : idéal pour **analyse, compression, détection d'anomalies**
- **DDPM** : idéal pour **génération d'images de haute qualité**
- Les deux approches peuvent être **combinées** (Latent Diffusion Models)

Sur les VAE :

- β contrôle le trade-off reconstruction/régularisation
- $\beta = 1$ **est optimal** pour MNIST
- Dense reconstruit mieux (BCE↓) mais 18× plus lourd
- Conv plus efficace : structure spatiale + parcimonie
- Dim=4 améliore la reconstruction (Conv ≈ Dense)
- Applications médicales prometteuses

Sur les DDPM :

- Génération de haute qualité (FID=28.11)
- Convergence stable sans surapprentissage
- Approche complémentaire aux VAE
- Plus lent en génération (T passes)

Points clés

1. Le paramètre β est **crucial** pour équilibrer reconstruction et régularisation
2. Le VAE Conv offre le **meilleur compromis** efficacité/performance

Améliorations VAE :

- β-**annealing** : augmentation progressive de β
- Architectures plus profondes (ResNet-VAE)
- Datasets complexes :
  - Fashion-MNIST
  - CIFAR-10
  - CelebA (visages)
- Applications médicales réelles
- VAE conditionnels (contrôle de classe)

Améliorations DDPM :

- Plus d'époques (50-100) et timesteps (10000)
- **Attention mechanisms** dans U-Net
- **Latent Diffusion Models** (compression)
- **DDIM** pour accélération du sampling
- Génération conditionnelle (contrôle du chiffre)
- Applications :
  - Super-résolution
  - Inpainting (complétion d'images)

Approches Hybrides

Combiner VAE + DDPM : **Latent Diffusion Models**

- VAE compresse l'image dans un espace latent
- DDPM génère dans cet espace latent (plus rapide)

Questions ?

Merci de votre attention

BELHASSAN Adam — KEDDARI Bilel — LEGENDRE Antoine
Master 2 Data Science