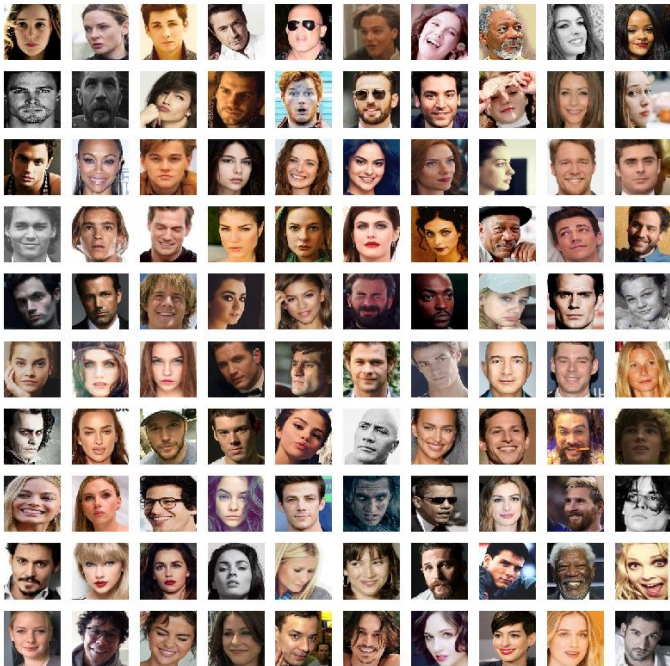


Traitement et Analyse d'Images

Modèles génératifs Auto-encodeur variationnel

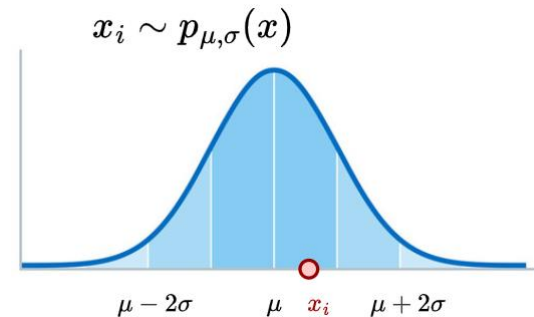
Quel est l'intérêt des modèles génératifs ?

► Comment générer des visages synthétiques ?



En modélisant la distribution correspondante $x \sim p_{\theta}(x)$!

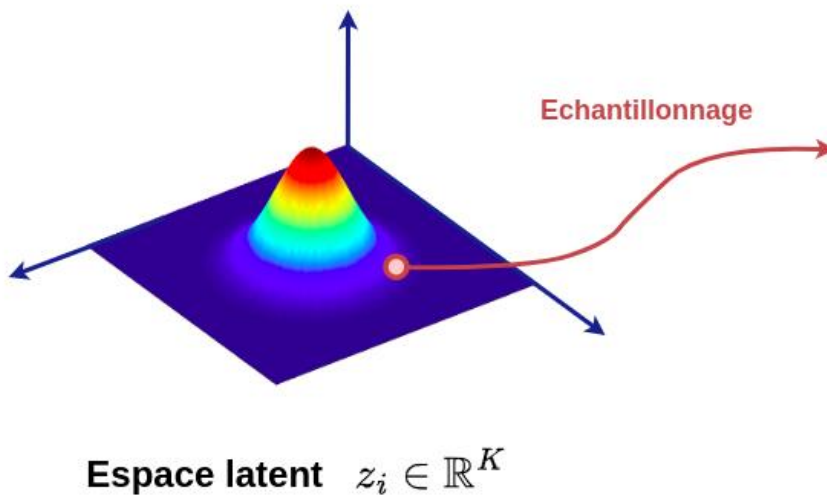
→ Est ce que des distributions classiques restent pertinentes ?



Quel est l'intérêt des modèles génératifs ?

► Comment modéliser des distributions complexes ?

On projette l'espace image (modélisé par $p(x)$) dans un espace de représentation cachée, plus efficace, appelé espace latent (modélisé par $p(z)$)

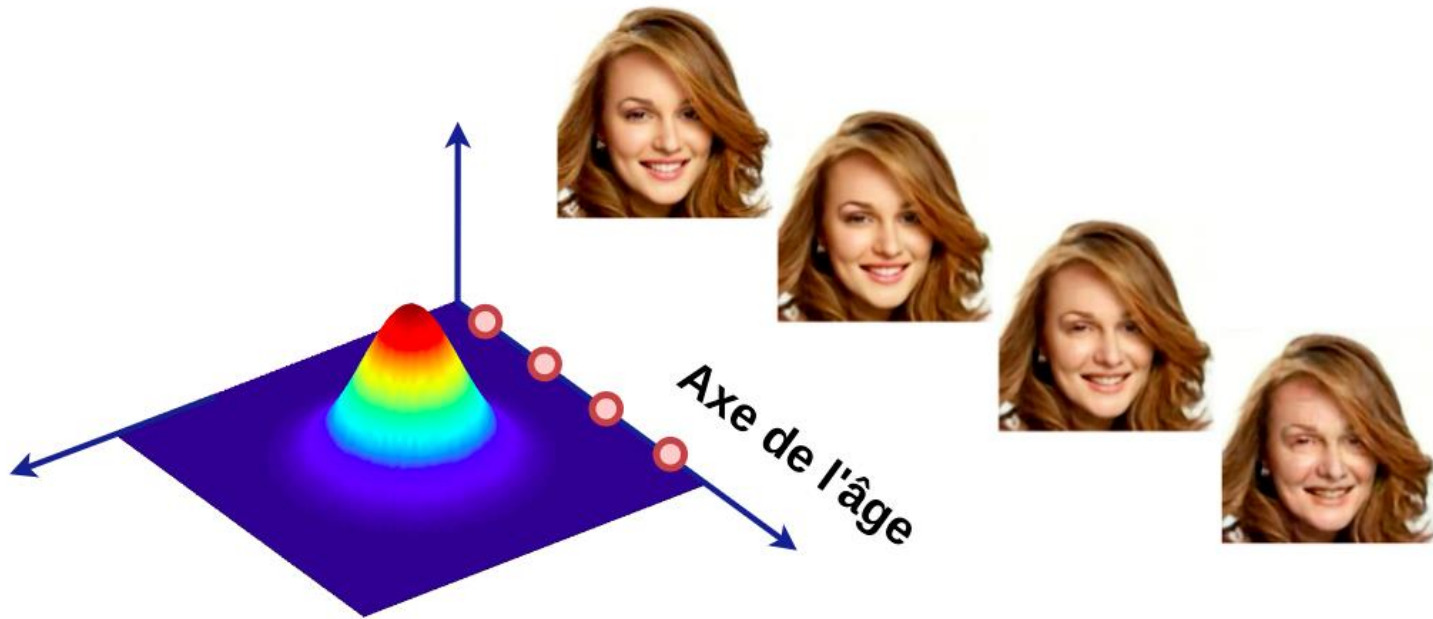


Distribution des visages

Quel est l'intérêt des modèles génératifs ?

► Pour quel but ?

Une obsession : maîtriser l'espace latent !!

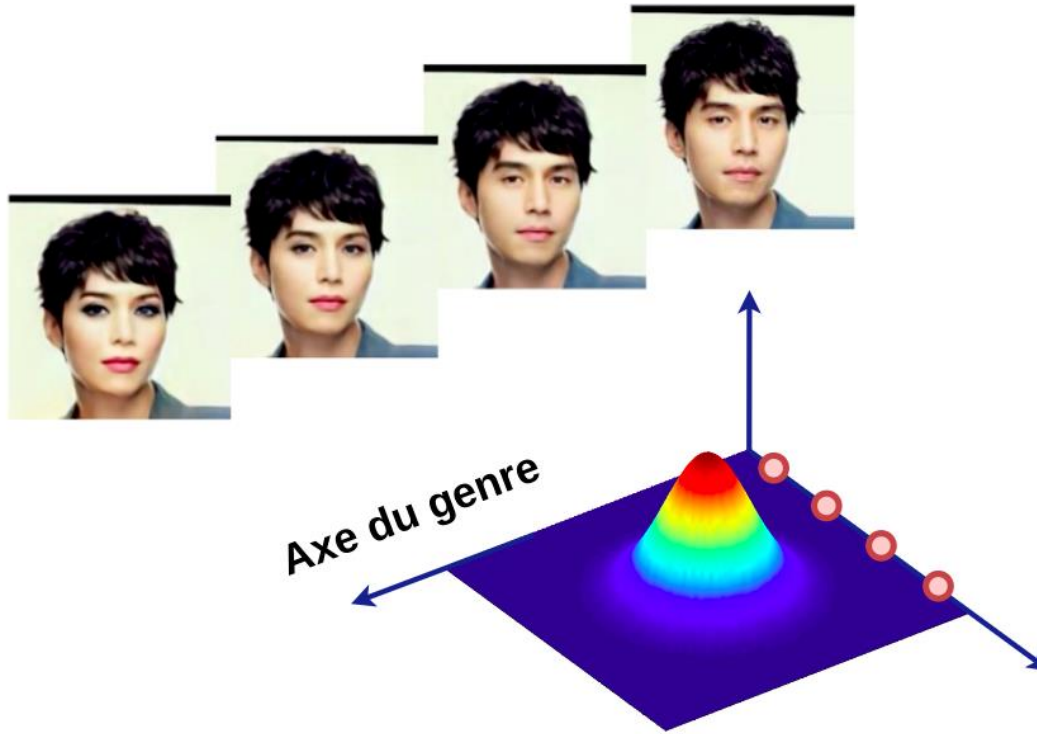


Espace latent $z_i \in \mathbb{R}^K$

Quel est l'intérêt des modèles génératifs ?

► Pour quel but ?

Une obsession : maîtriser l'espace latent !!



Espace latent $z_i \in \mathbb{R}^K$

Auto-encodeurs

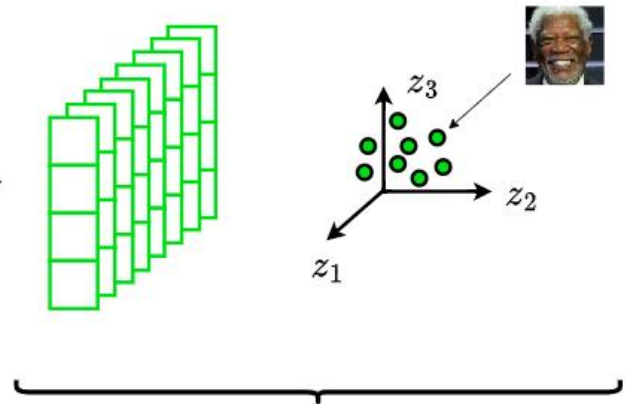
Comment apprendre une distribution ?

- Projection dans un espace de représentation plus efficace et de dimension inférieure



Espace d'entrée $x_i \in \mathbb{R}^{N \times M}$

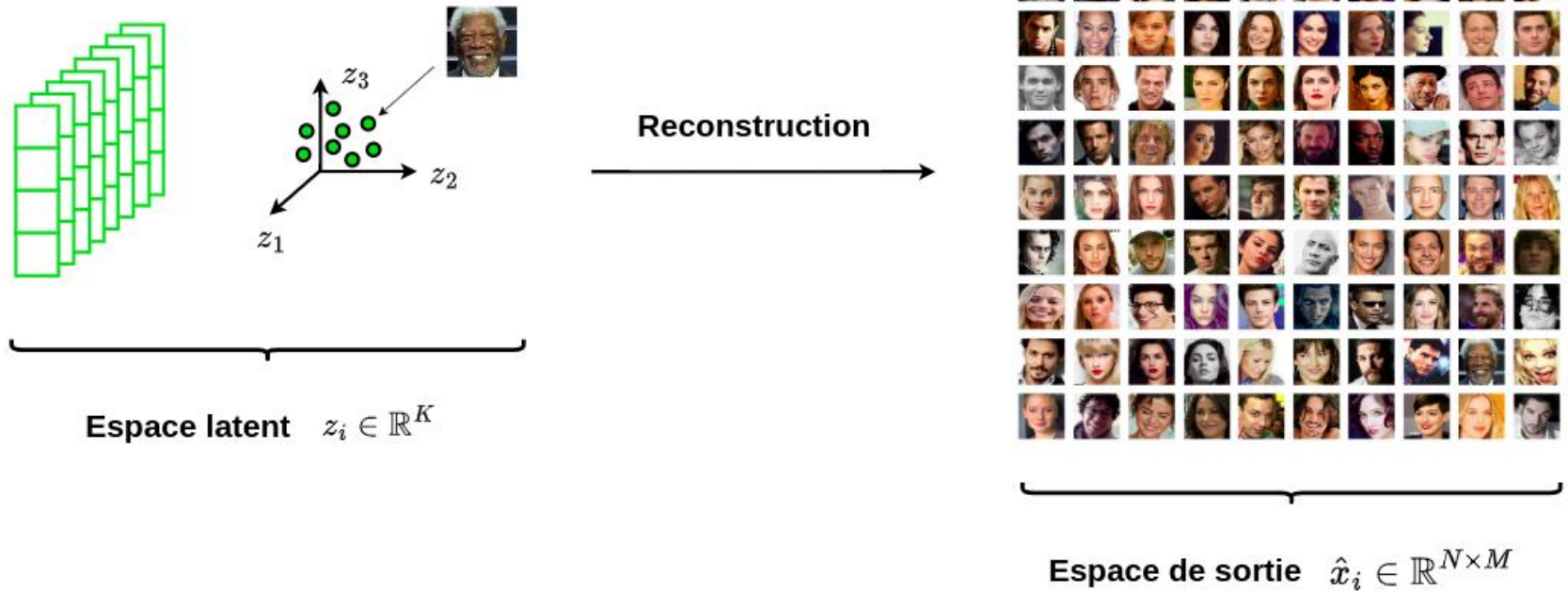
Projection



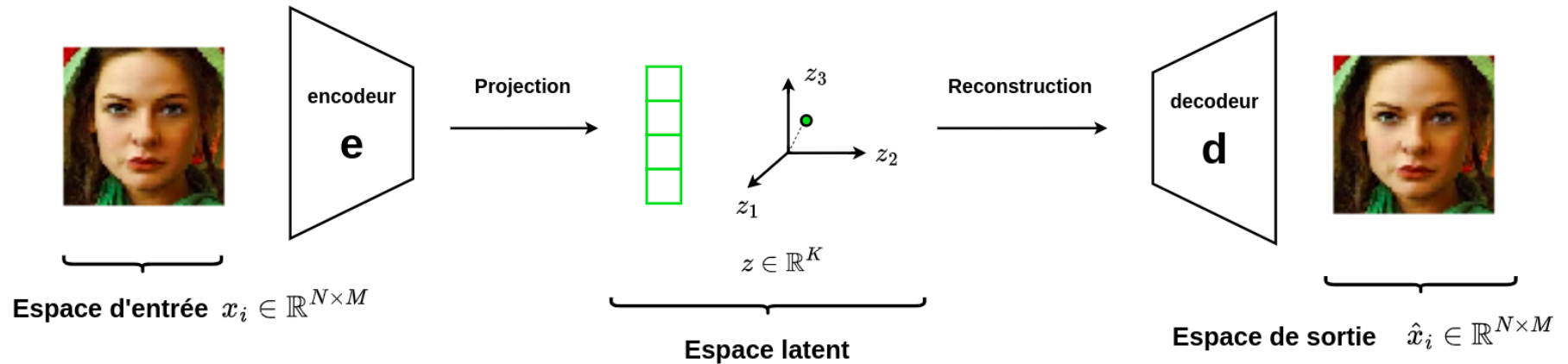
Espace latent $z_i \in \mathbb{R}^K$

Comment apprendre une distribution ?

- ▶ Comment disposer d'un espace de représentation pertinent ?



► Architectures standards

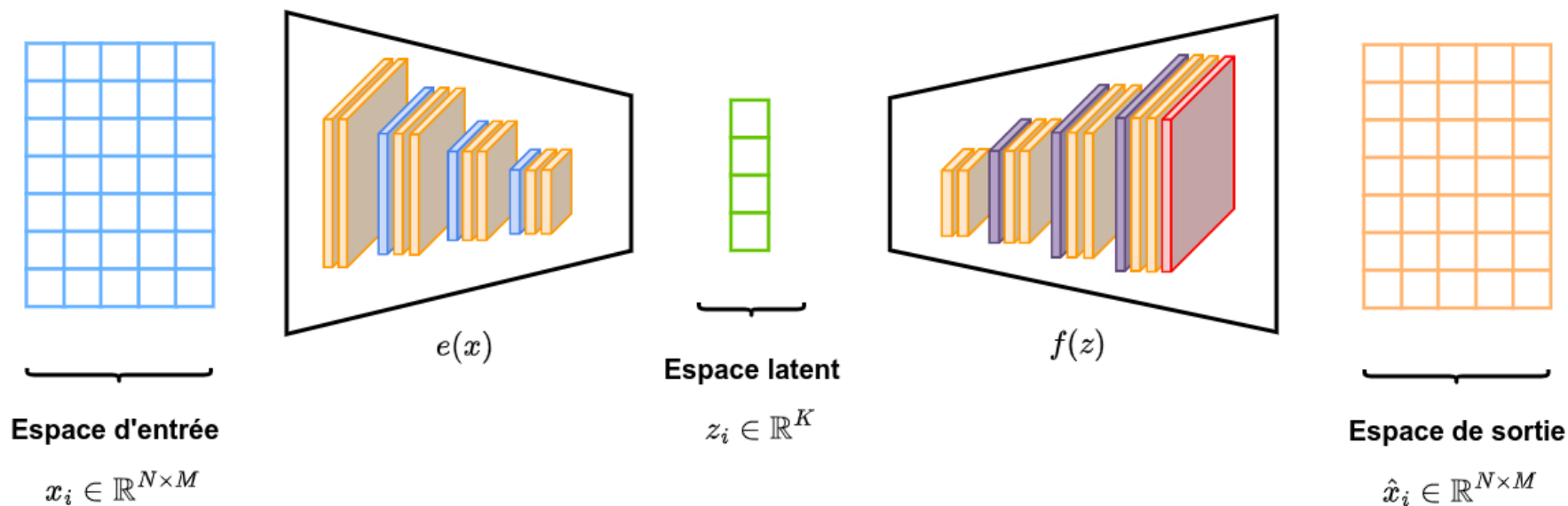


► Exemple de fonction de coût

$$\text{loss} = \|x - \hat{x}\|^2$$

Implémentation par apprentissage profond

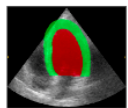
- ▶ Encodeur / Décodeur modélisé par des réseaux neuronaux (convolutionnels)



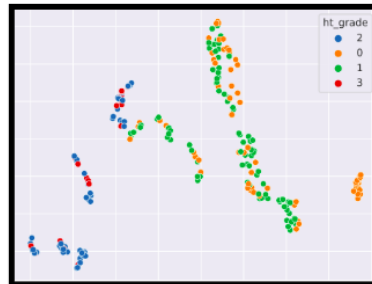
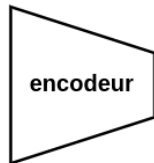
$$\text{loss} = \|x - f(e(x))\|^2$$

► Auto-encodeur ? Pour quoi faire ?

► Représentation de données

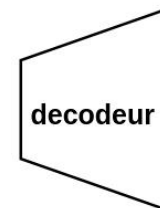


Patients



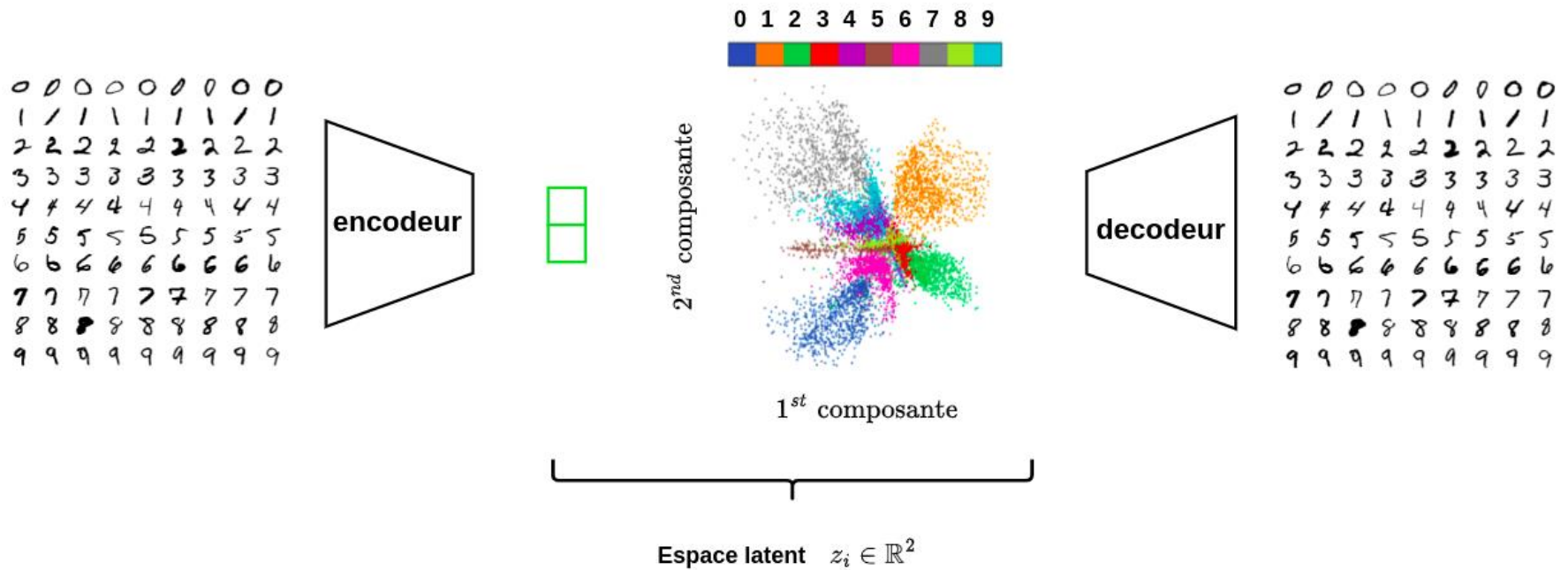
Representation de population

► Modèles génératifs

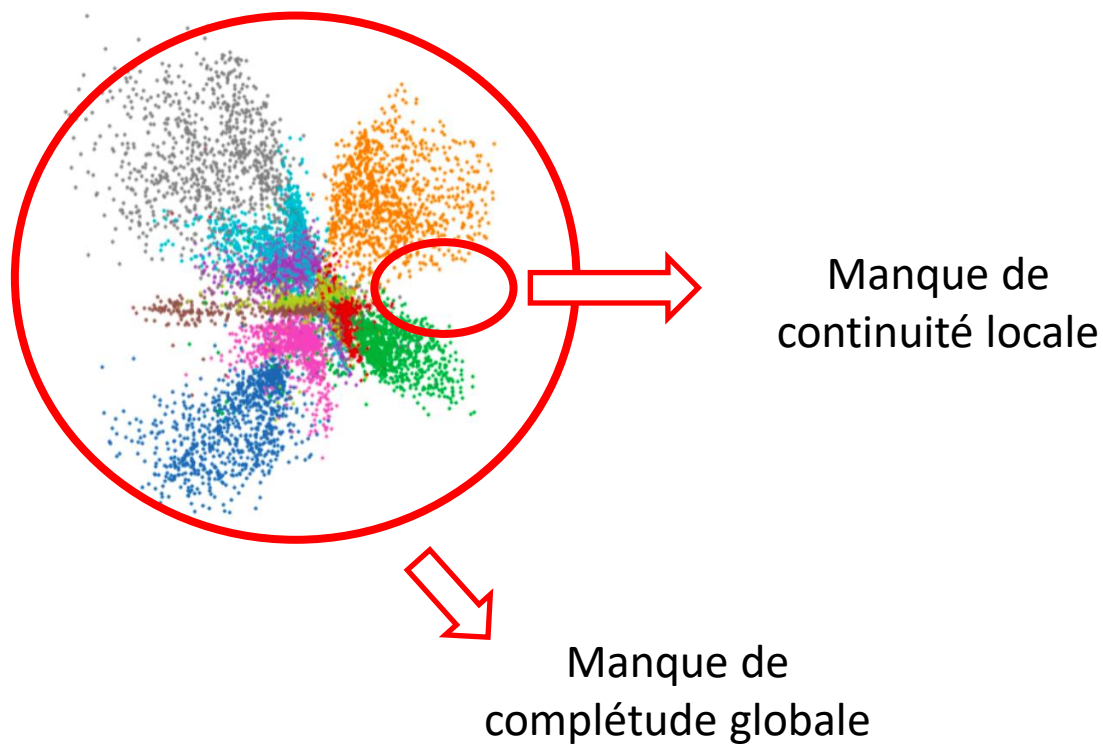


0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9

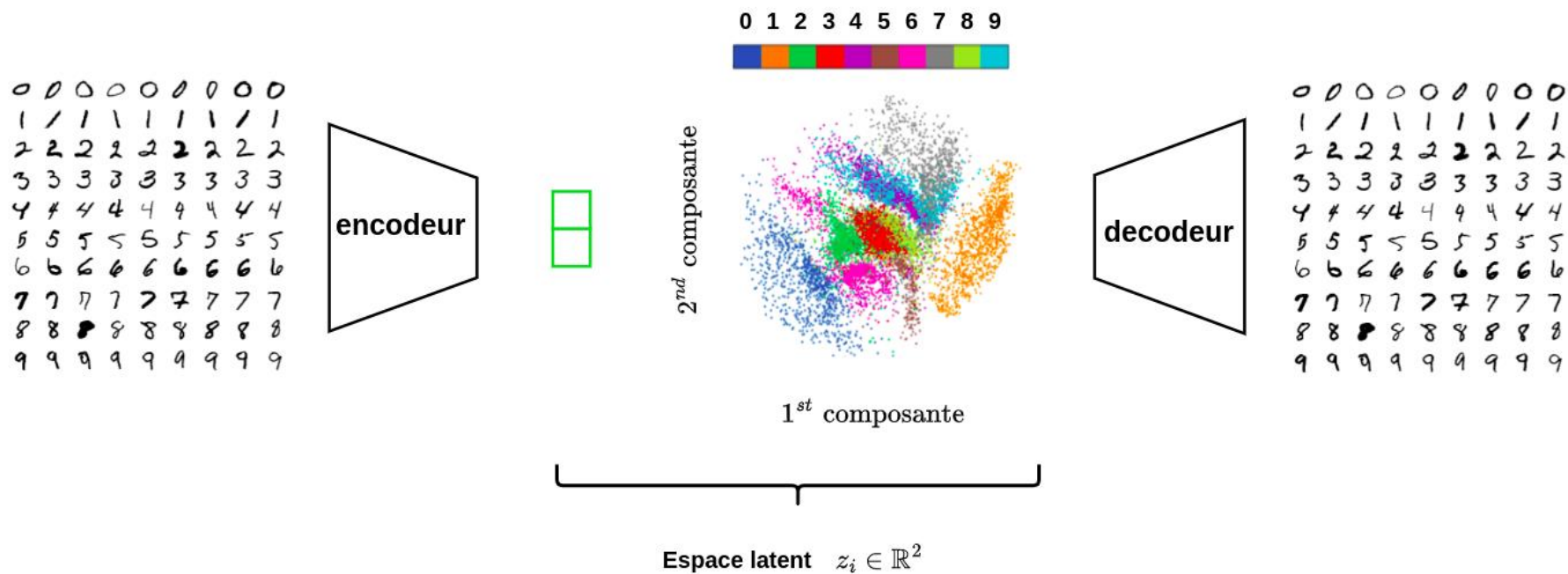
► Modèles génératifs



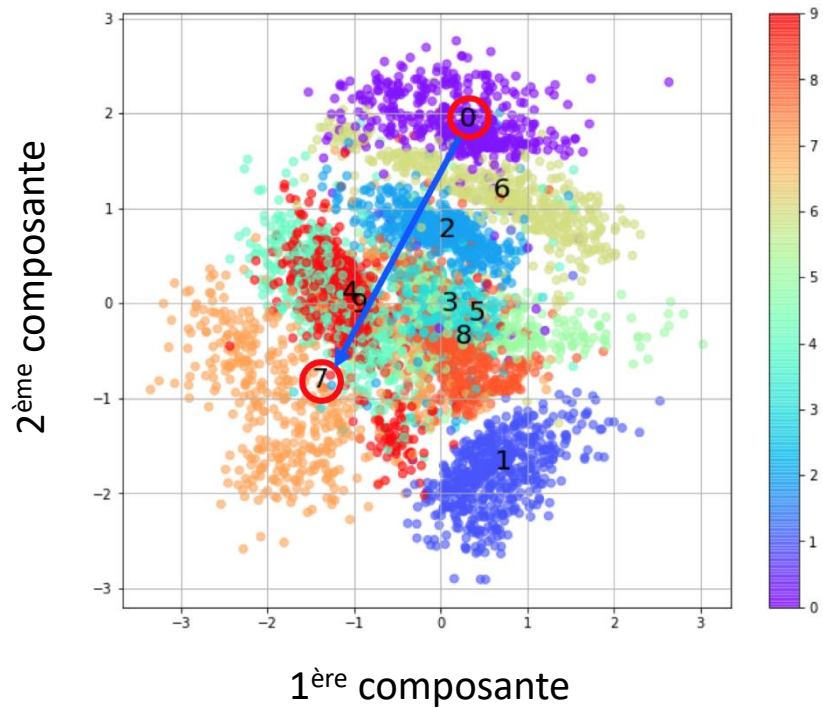
- Nécessité de mieux contrôler la structure de l'espace latent



- ▶ Modèle génératif avec de meilleures propriétés grâce au *cadre variationnel*

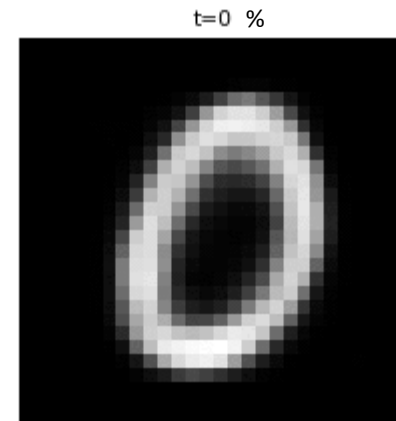


► Modèle génératif avec cadre variationnel



Interpolation linéaire dans l'espace latent

$$t \cdot z_0 + (1 - t) \cdot z_7, \quad 0 \leq t \leq 1$$



Les auto-encodeurs variationnels

L'ensemble des mathématiques sont décrites dans le blog suivant

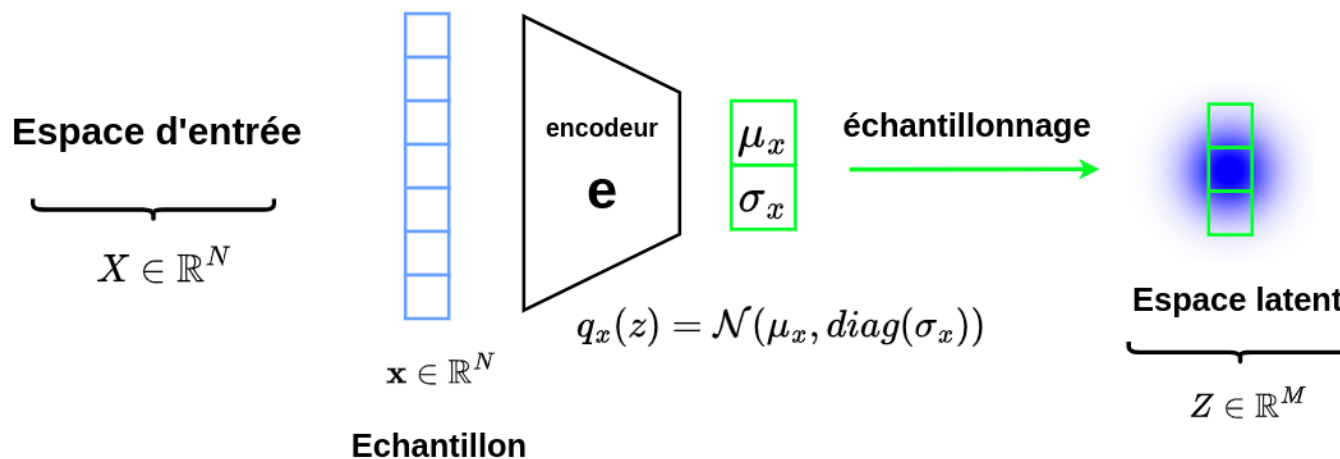
<https://creatis-myriad.github.io/tutorials/2022-09-12-tutorial-vae.html>

- ▶ Renforcement d'un espace latent structuré
 - Au travers d'un cadre probabiliste
 - En imposant des contraintes de continuité
 - En imposant des contraintes de complétude

► Cadre probabiliste: *continuité*

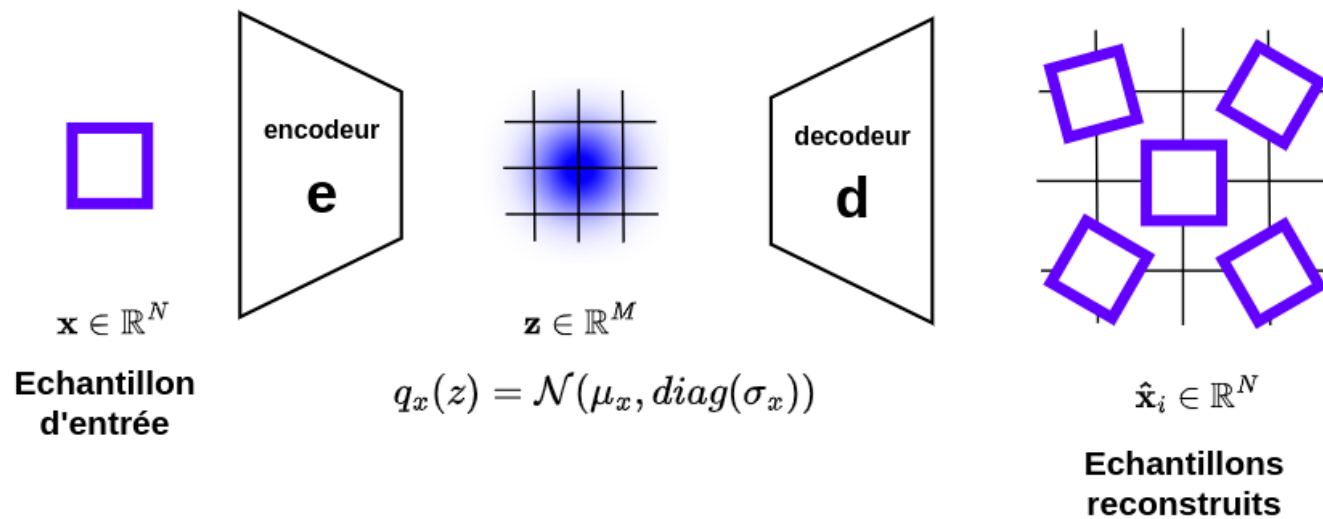
→ Introduction de régularisations locales de l'espace latent

→ Chaque donnée d'entrée x est encodée sous forme d'une distribution gaussienne $q_x(z) = \mathcal{N}(\mu_x, \text{diag}(\sigma_x))$



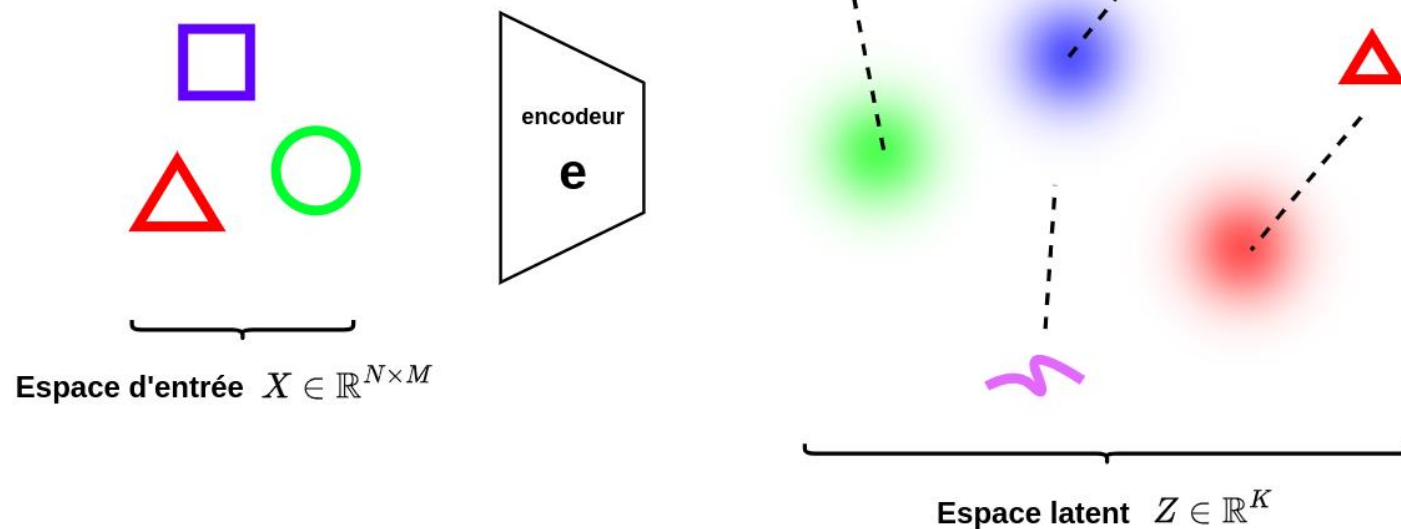
► Cadre probabiliste: *continuité*

→ L'échantillonnage d'une région locale de l'espace latent produit des résultats proches



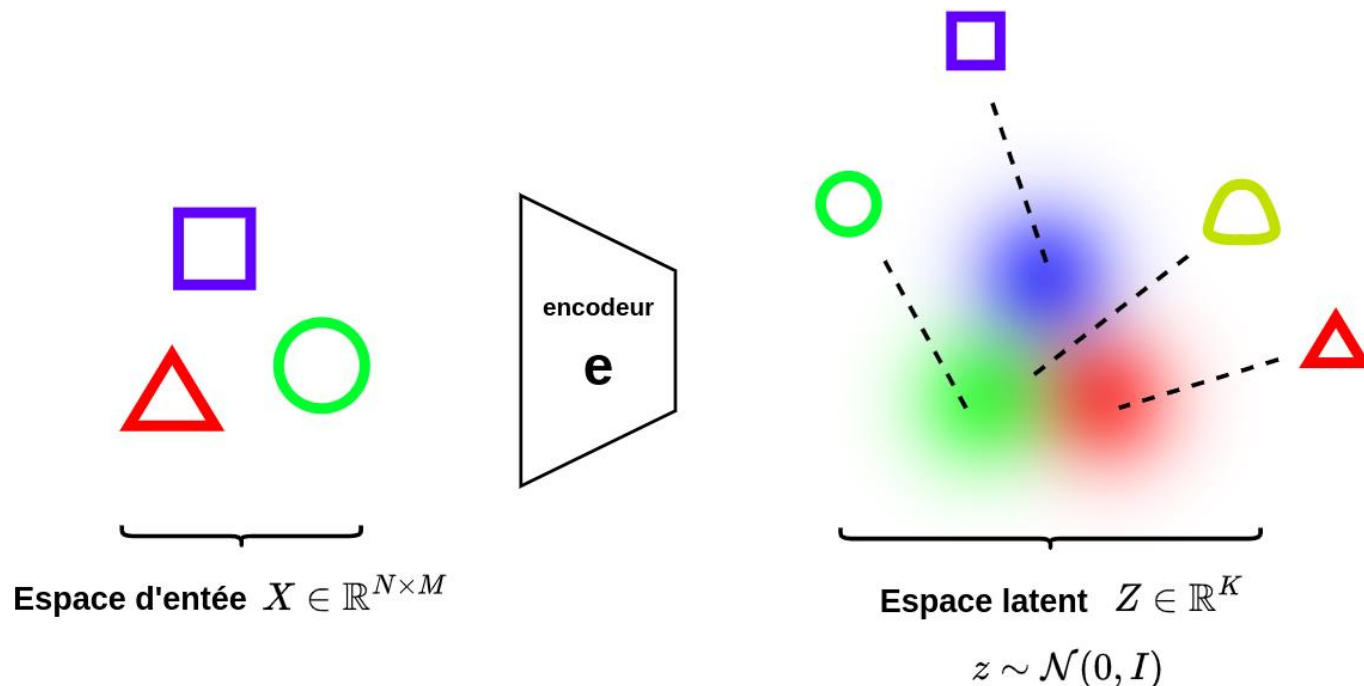
► Cadre probabiliste: *complétude*

→ Favoriser que tout point reconstruit dans l'espace latent donne des résultats cohérents



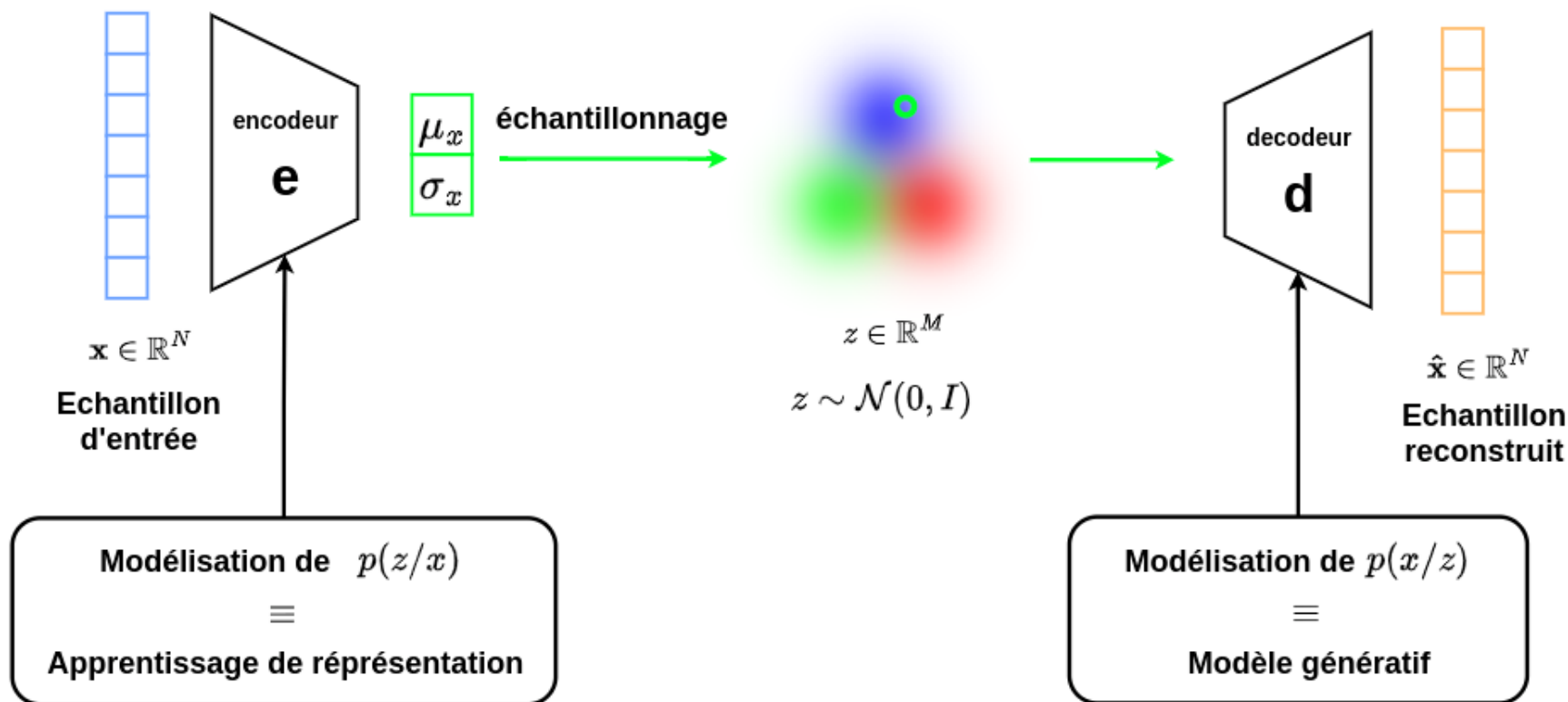
► Cadre probabiliste: *complétude*

- Imposer que toutes les distributions $q_x(z)$ soient proches d'une distribution normal standard $\mathcal{N}(0, I)$
- Variances proches de 1 \Rightarrow limite la génération de distributions ponctuelles
- Moyennes proches de 0 \Rightarrow favorise des distributions proches les unes des autres



► Cadre probabiliste: *continuité & complétude*

→ Architecture des VAE



Les auto-encodeurs variationnels

Formulation mathématique

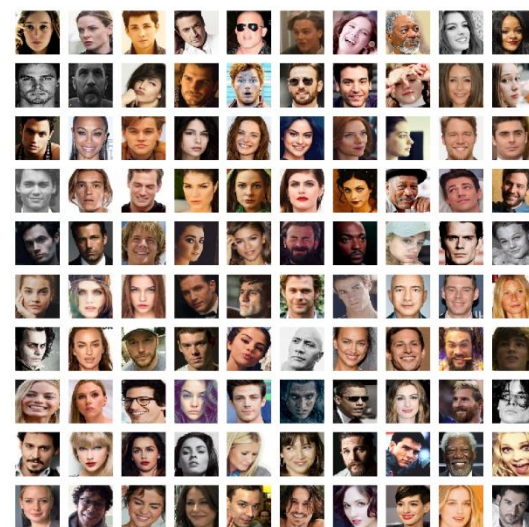
Quel est l'intérêt des modèles génératifs ?

► Comment générer des visages synthétiques ?

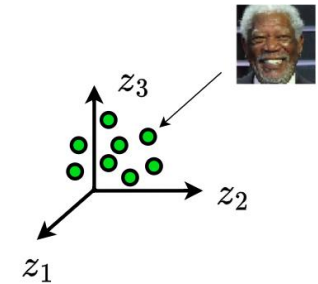
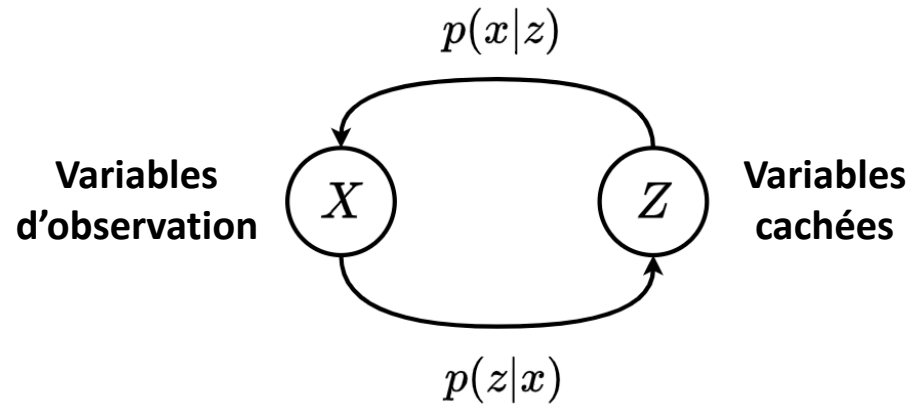
- Soit $p(\cdot)$ la distribution qui représente les visages humains
- On souhaite trouver un modèle f qui génère des échantillons x dont la probabilité $p(x)$ ou soit maximale

$$f^* = \arg \max_f p(x) \quad \text{avec } x \text{ généré par un modèle } f(\cdot)$$

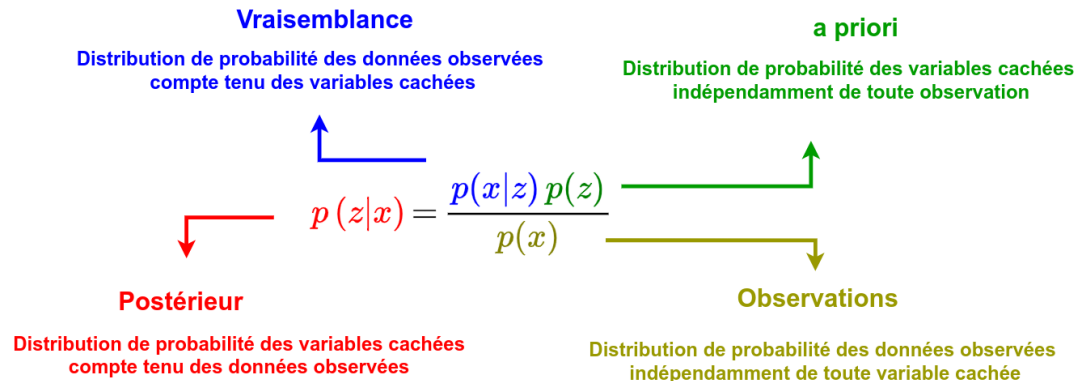
- Dans ce cas, les échantillons générés s'apparentent aux visages humains de la base de données d'entraînement



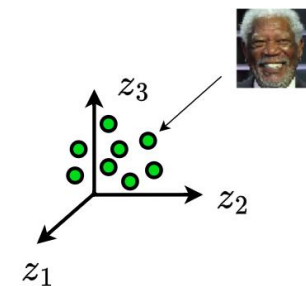
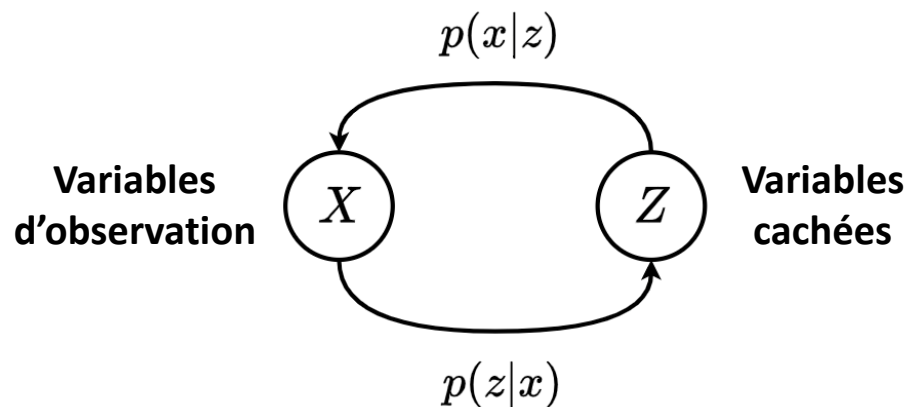
- Modélisation d'une variable caché z pour réduire la complexité du problème



→ Rappel du théorème de Bayes



► Formulation mathématique



→ La distribution $p(z|x)$ est généralement complexe à modéliser

Approximation de $p(z|x)$ par une fonction $q(z|x)$ simple et calculable et qui permettra d'échantillonner efficacement z

▶ Inférence variationnelle

- Technique d'approximation statistique de distributions complexes, ici $p(z|x)$
- Définition d'une famille paramétrée de distributions
 - ▶ ex. famille des gaussiennes de paramètres μ_x, σ_x modélisées par des fonctions à déterminer
- Rechercher la meilleure approximation de la distribution cible dans cette famille
- Le meilleur élément de la famille minimise une mesure d'erreur d'approximation entre deux distributions
 - ▶ Fonction de divergence de Kullback-Liebler est souvent utilisée

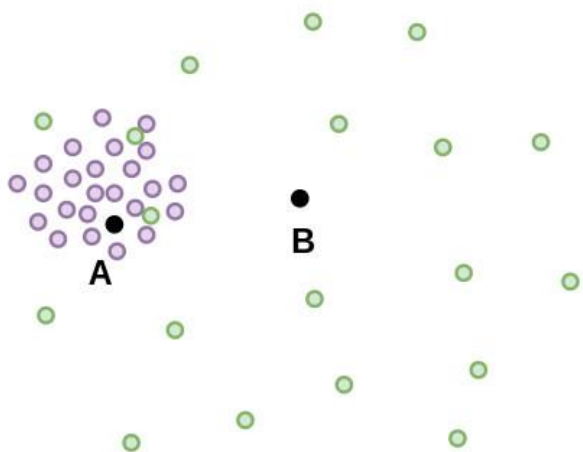
► Fonction de divergence de Kullback-Liebler

→ Mesure de distances entre deux distributions via l'entropie relative

$$D_{KL}(p \parallel q) = \int p(x) \cdot \log \left(\frac{p(x)}{q(x)} \right) dx$$

→ D_{KL} est une mesure toujours positive $D_{KL}(p \parallel q) \geq 0$

→ D_{KL} est une mesure non symétrique $D_{KL}(p \parallel q) \neq D_{KL}(q \parallel p)$



- Pour la distribution en violet, la distance AB est grande
- Pour la distribution en vert, la distance AB est modérée
- La notion de distance est différente en fonction des distributions

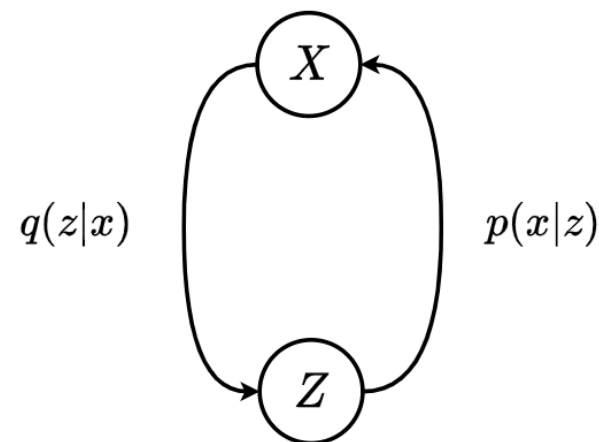
► Inférence variationnelle

- $p(z|x)$ est approximé par une famille de fonctions $q(z|x)$
- $q(z|x)$ est modélisée par une distribution gaussienne alignée sur les axes

$$q(z|x) = \mathcal{N}(\mu_x, \sigma_x) = \mathcal{N}(g(x), \text{diag}(h(x)))$$

- $g(x)$ et $h(x)$ sont des fonctions qui représentent les moyennes μ_x et les covariances σ_x
- Mesure d'approximation entre les deux distributions $p(z|x)$ et $q(z|x)$

$$(g^*, h^*) = \arg \min_{(g, h)} D_{KL}(q(z|x) \parallel p(z|x))$$



► Inférence variationnelle

- En jouant sur les expressions de $p(x)$, il est possible de retrouver les définitions et relations suivantes

$$\log p(x) \geq \int q(z|x) \log \left(\frac{p(x, z)}{q(z|x)} \right) dz$$

$$\log p(x) \geq ELBO$$

$$\log p(x) = ELBO + D_{KL}(q(z|x) || p(z|x))$$

- ELBO est une borne inférieure de $\log p(x)$
- Maximiser ELBO revient à maximiser $\log p(x)$
- Si on maximise $\log p(x)$, alors on minimise $D_{KL}(q(z|x) || p(z|x))$

► Processus d'optimisation

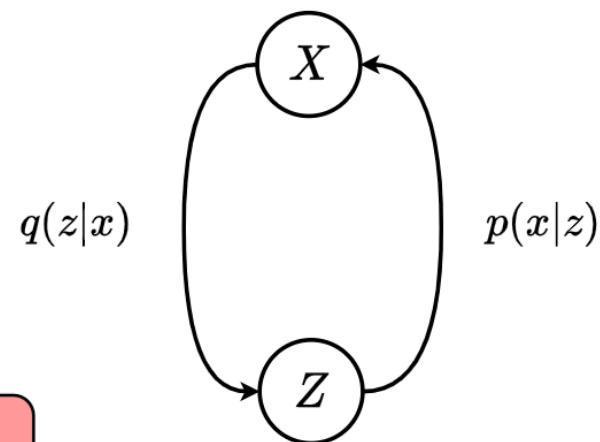
→ La borne inférieure ELBO peut être reformulée

$$\mathcal{L} = \mathbb{E}_{z \sim q(z|x)} [\log(p(x|z))] - D_{KL}(q(z|x) \parallel p(z))$$

→ En exploitant l'hypothèse gaussienne suivante

$$p(x|z) = \mathcal{N}(f(z), cI)$$

$$\mathcal{L} \propto \mathbb{E}_{z \sim q(z|x)} [-\alpha \|x - f(z)\|^2] - D_{KL}(q(z|x) \parallel p(z))$$



► Processus d'optimisation

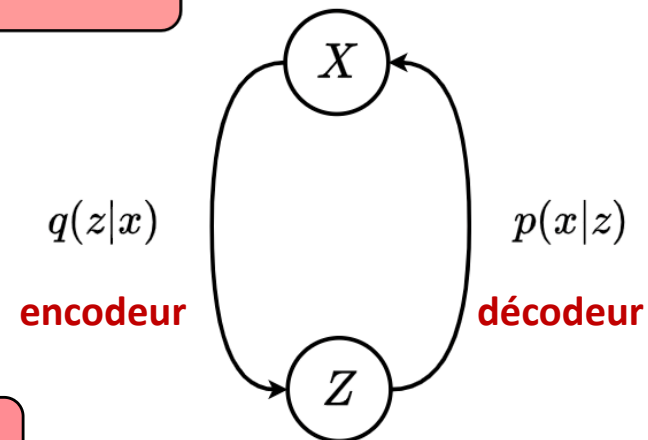
$$(f^*, g^*, h^*) = \arg \min_{(f, g, h)} (\mathbb{E}_{z \sim q(z|x)} [\alpha \|x - f(z)\|^2] + D_{KL}(q(z|x) \parallel p(z)))$$

► Fonction de perte en apprentissage profond

$$\text{loss} = \alpha \|x - f(z)\|^2 + D_{KL}(\mathcal{N}(g(x), \text{diag}(h(x))), \mathcal{N}(0, I))$$

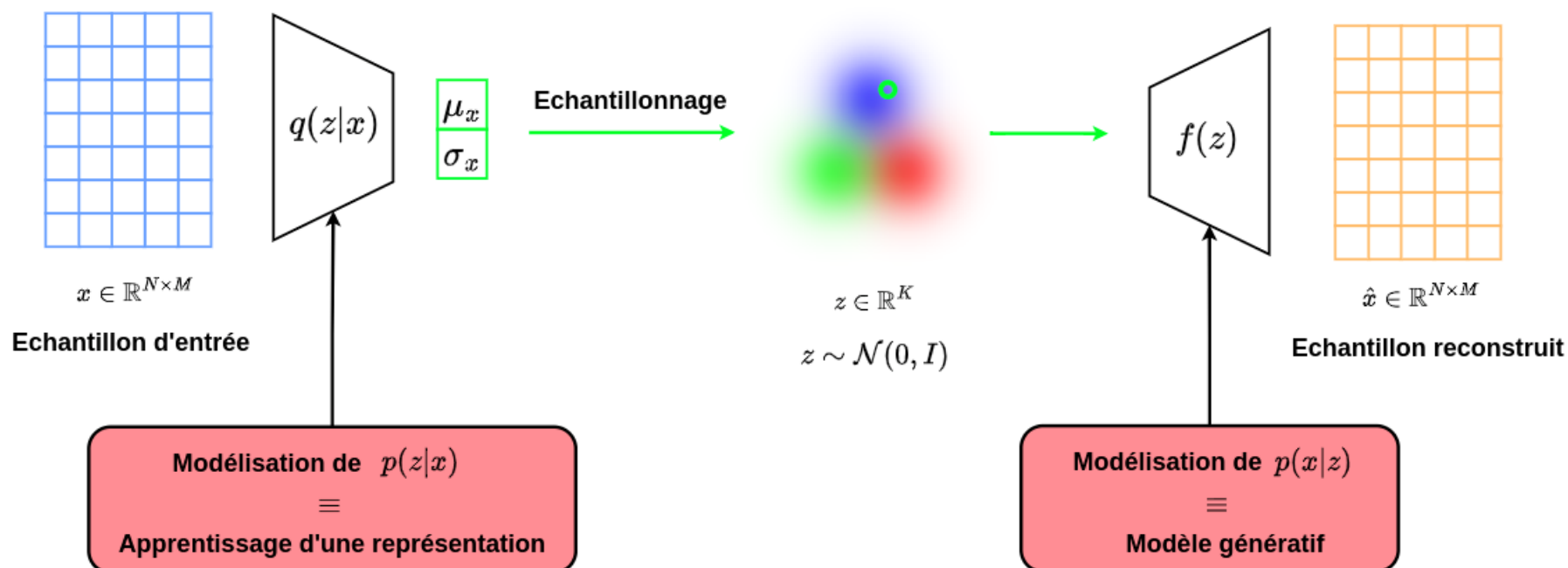
→ $g(\cdot)$ et $h(\cdot)$ sont modélisés par un encodeur

→ $f(\cdot)$ est modélisée par un décodeur



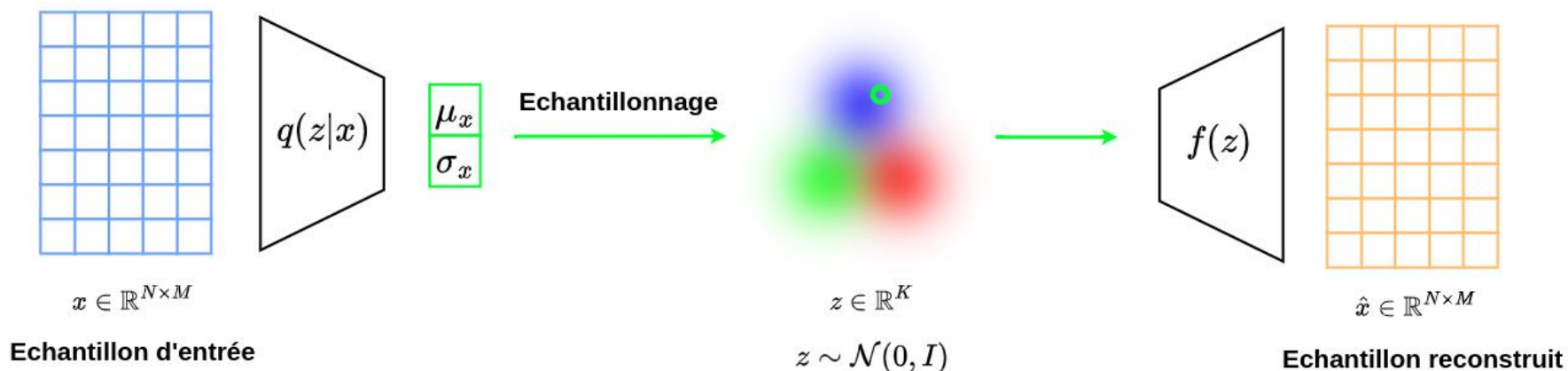
► Interprétation de la fonction de perte

$$\text{loss} = D_{KL}(\mathcal{N}(g(x), \text{diag}(h(x))), \mathcal{N}(0, I)) + \alpha \|x - f(z)\|^2$$



► Interprétation de la fonction de perte

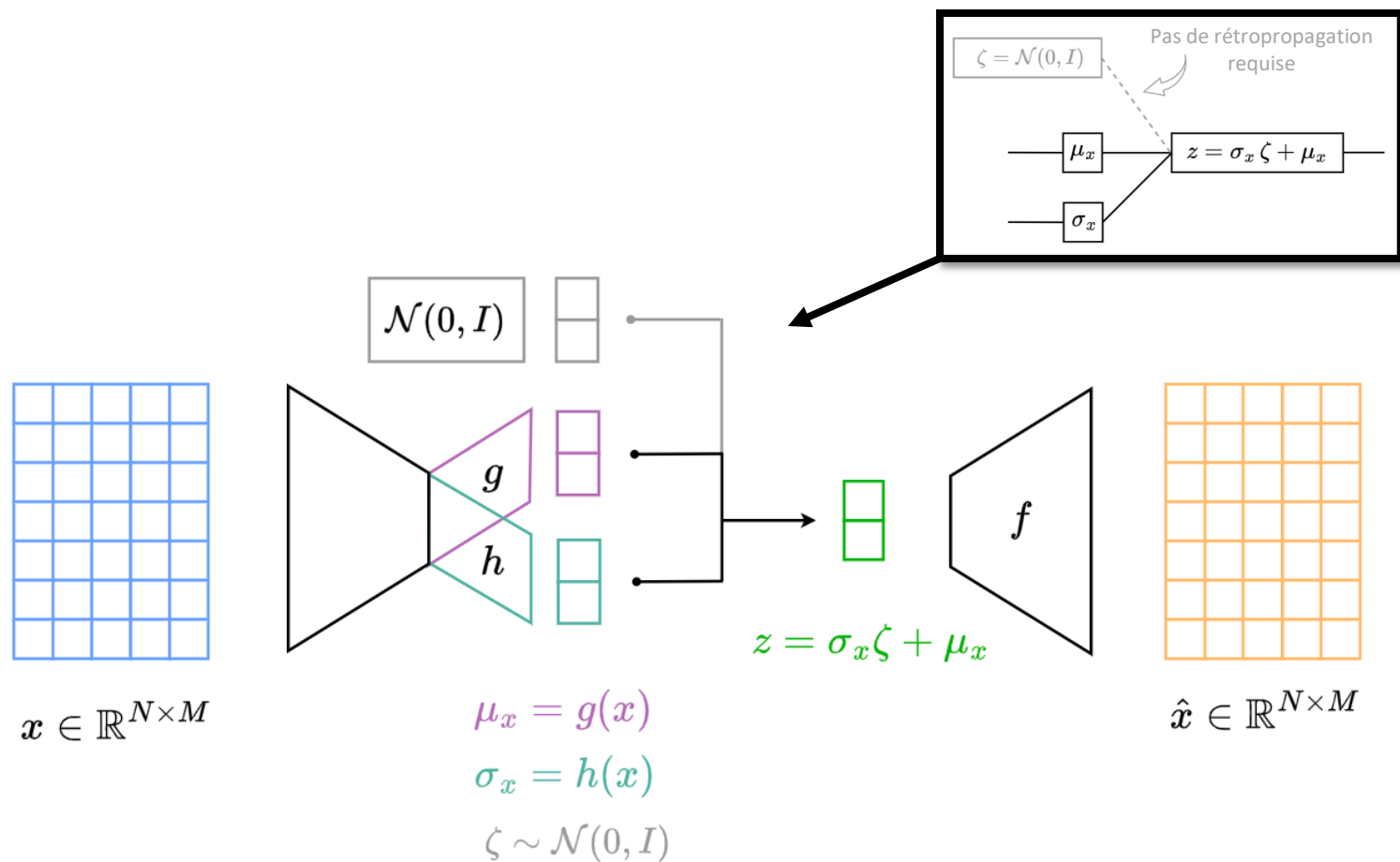
$$\text{loss} = D_{KL}(\mathcal{N}(g(x), \text{diag}(h(x))), \mathcal{N}(0, I)) + \alpha \|x - f(z)\|^2$$



→ $\mathcal{N}(g(x), h(x))$ impose une contrainte de **continuité** locale

→ $\mathcal{N}(\cdot, \mathcal{N}(0, I))$ impose une contrainte de **complétude** globale

► Astuce de reparamétrage



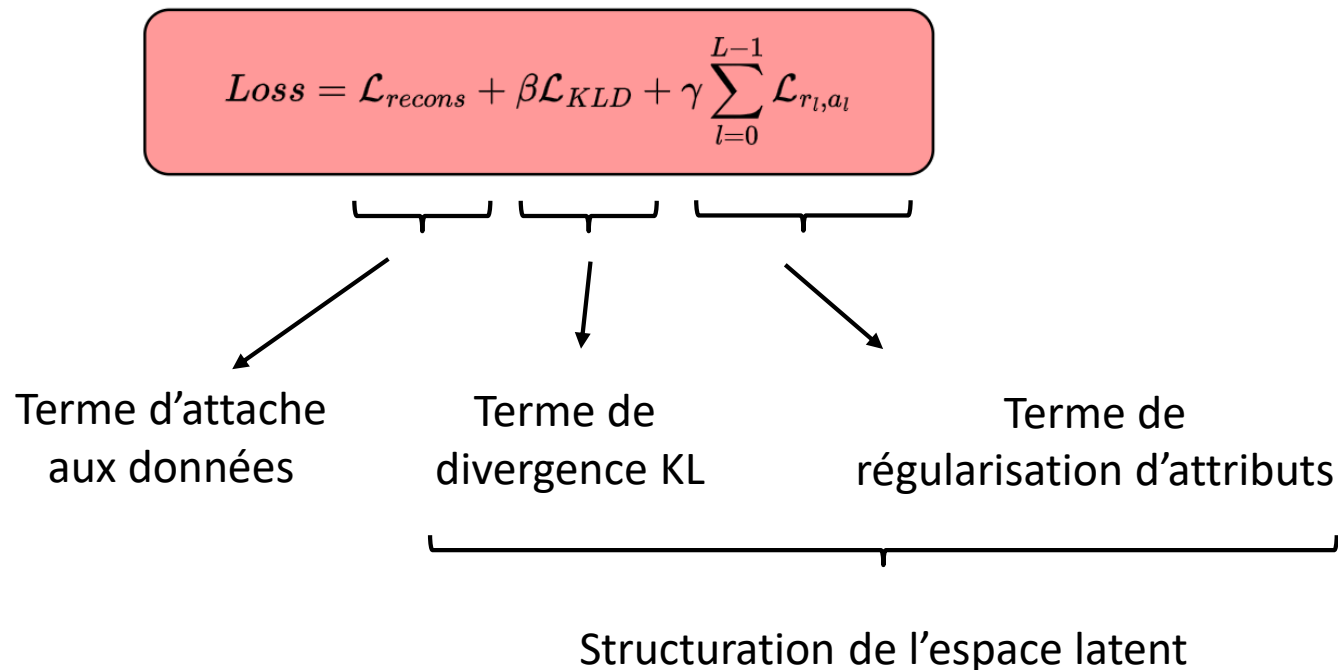
Application pratique

L'obsession est de maîtriser l'espace latent !!

► Régularisation de l'espace latent des VAE basée sur des attributs images

- Génération d'un espace latent structuré

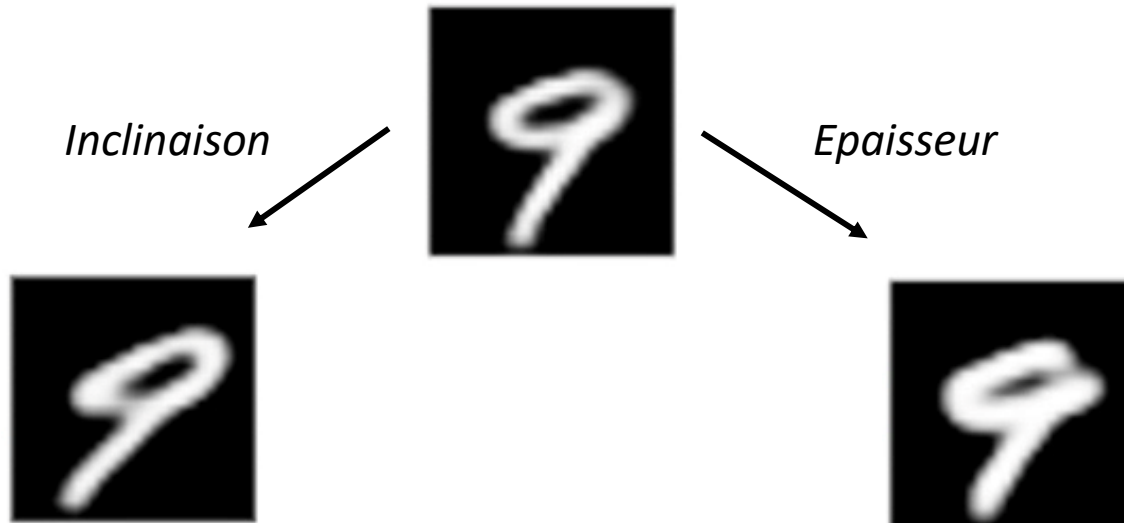
→ Les attributs spécifiques à valeur continue doivent être codés selon des dimensions spécifiques



► Terme de régularisation d'attributs

- C'est quoi un attribut ?

- Mesure effectuée dans l'espace image qui permet de caractériser un objet cible
- Ex: écriture de chiffres manuscrits (base de données MNIST)
 - Attributs: épaisseur du trait, inclinaison, longueur, surface, ...
- Mesures des attributs images avant apprentissage, utilisés comme données d'entrée



► Terme de régularisation d'attributs

- Lors de la phase d'apprentissage

→ Calcul pour chaque attribut d'une matrice de distance $D_a \in \mathbb{R}^{m \times m}$ à partir des m images $\{x_i\}_{1 \leq i \leq m}$ présentes dans le batch courant

$$D_a(i, j) = a(x_i) - a(x_j) \quad \text{avec} \quad i, j \in [0, m)$$

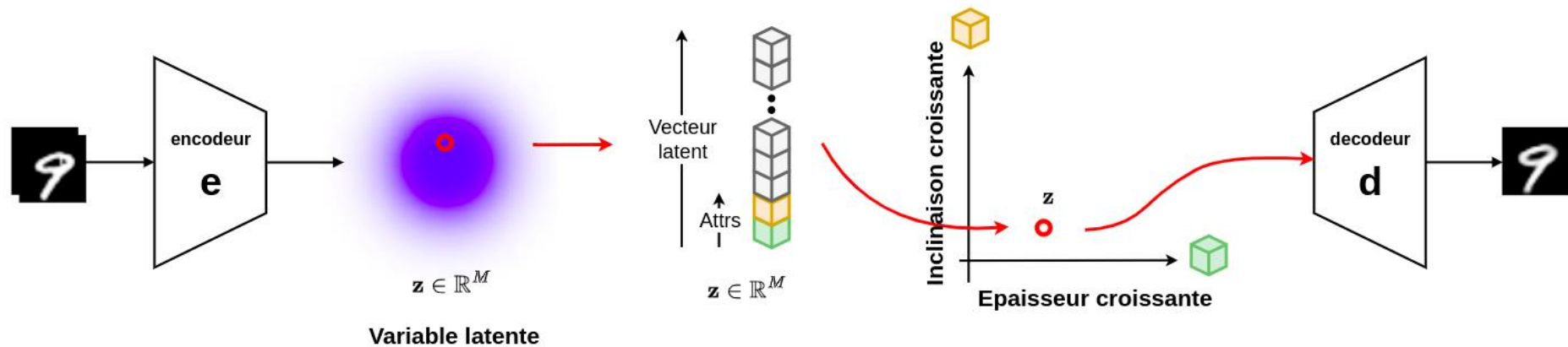
→ Calcul pour chaque attribut r d'une matrice de distance $D_r \in \mathbb{R}^{m \times m}$ à partir des m vecteur latents $\{z_i\}_{1 \leq i \leq m}$ correspondant aux images du batch courant

$$D_r(i, j) = z_i^r - z_j^r \quad \text{avec} \quad i, j \in [0, m)$$

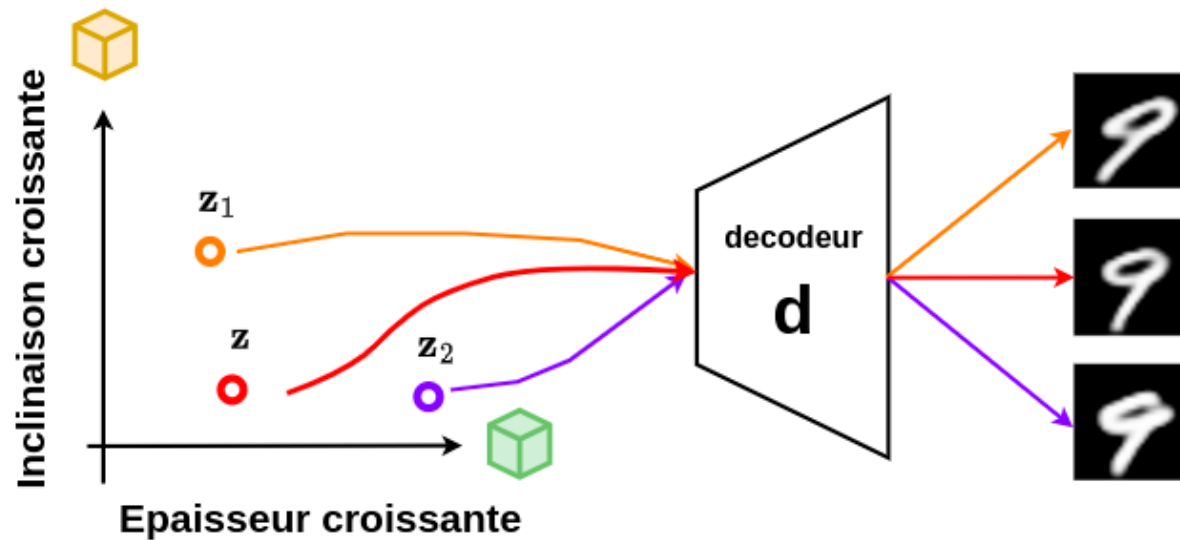
→ Introduction du terme de loss suivant

$$\mathcal{L}_{r,a} = MAE(\tanh(D_r) - \text{sign}(D_a))$$

- Génération d'un espace latent structuré suivant des attributs



- ▶ Génération d'un espace latent structuré suivant des attributs
 - Echantillonnage de l'espace latent structuré

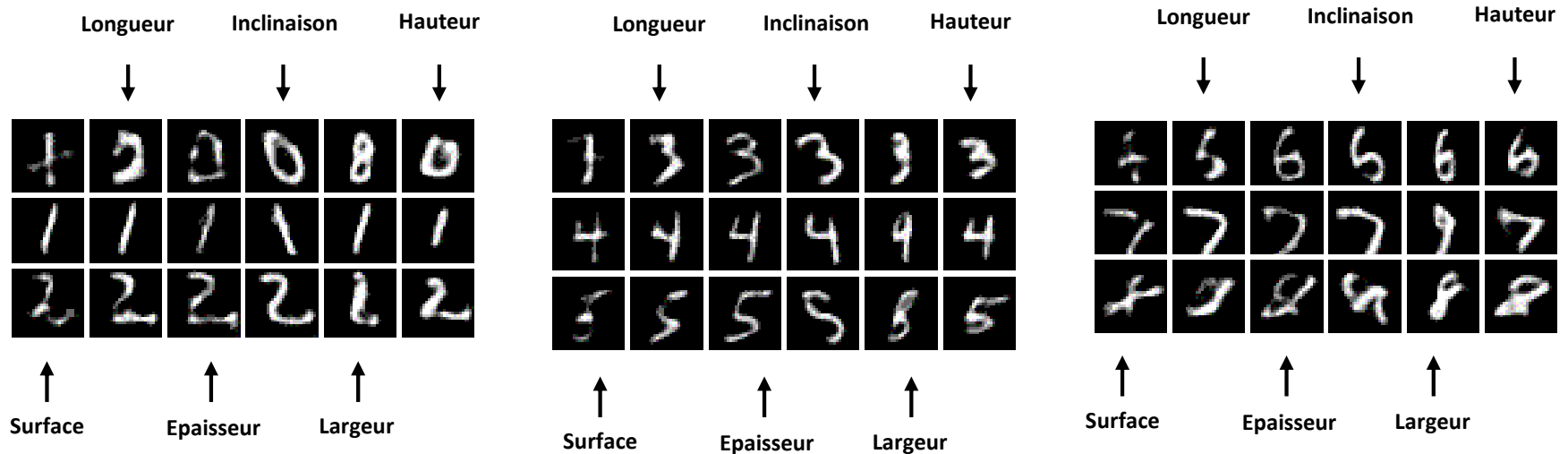


► Génération d'un espace latent structuré suivant des attributs

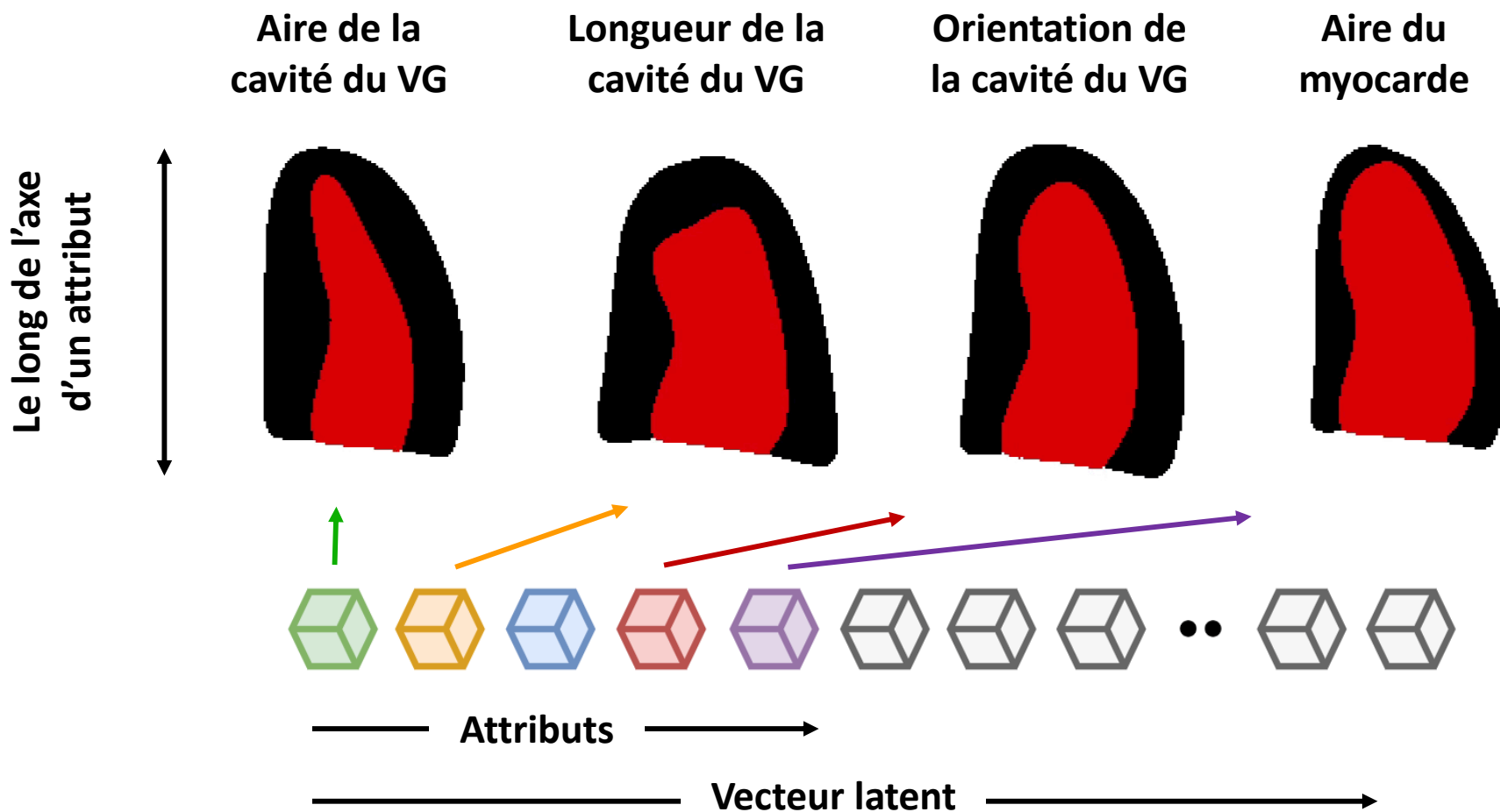
- Echantillonnage de l'espace latent structuré

- ➔ Attribut spécifique: surface, longueur, épaisseur, inclinaison, largeur, hauteur

- ➔ Chaque colonne correspond à une traversée le long d'une dimension régularisée



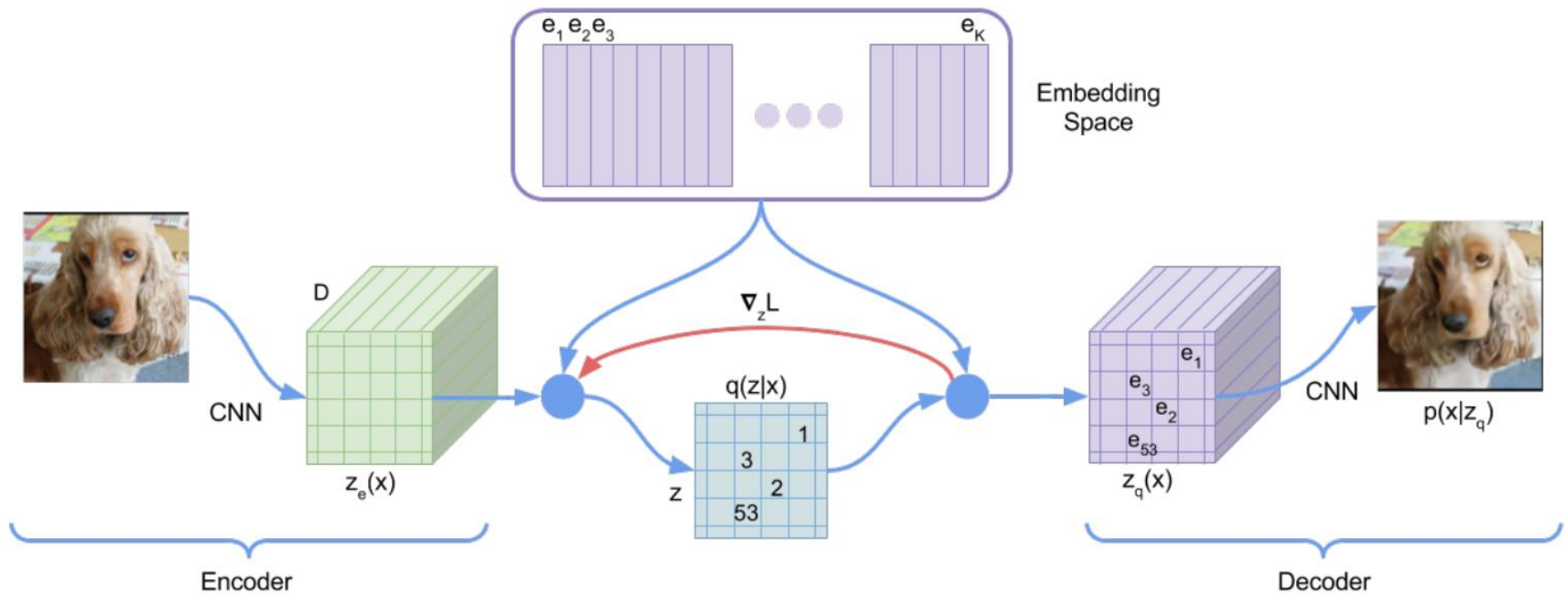
Structuration de l'espace latent: AR-VAE



Les auto-encodeurs variationnels avec quantification vectorielle

Autre méthode inspirée des VAE: VQ-VAE

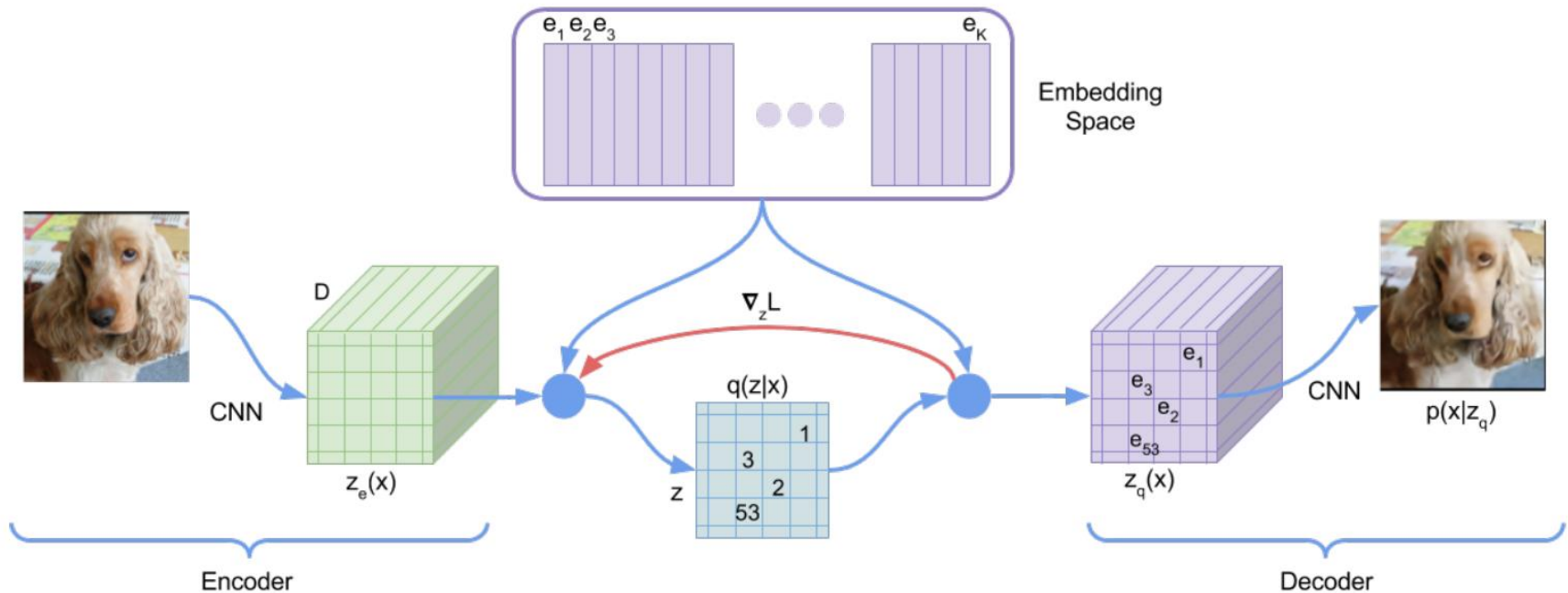
- Apprentissage conjoint d'un auto-encodeur et d'une représentation discrète de l'espace latent



- L'espace latent est défini par l'ensemble des vecteurs $\{e_i\}_{i \in [1, K]}$ qui sont appris

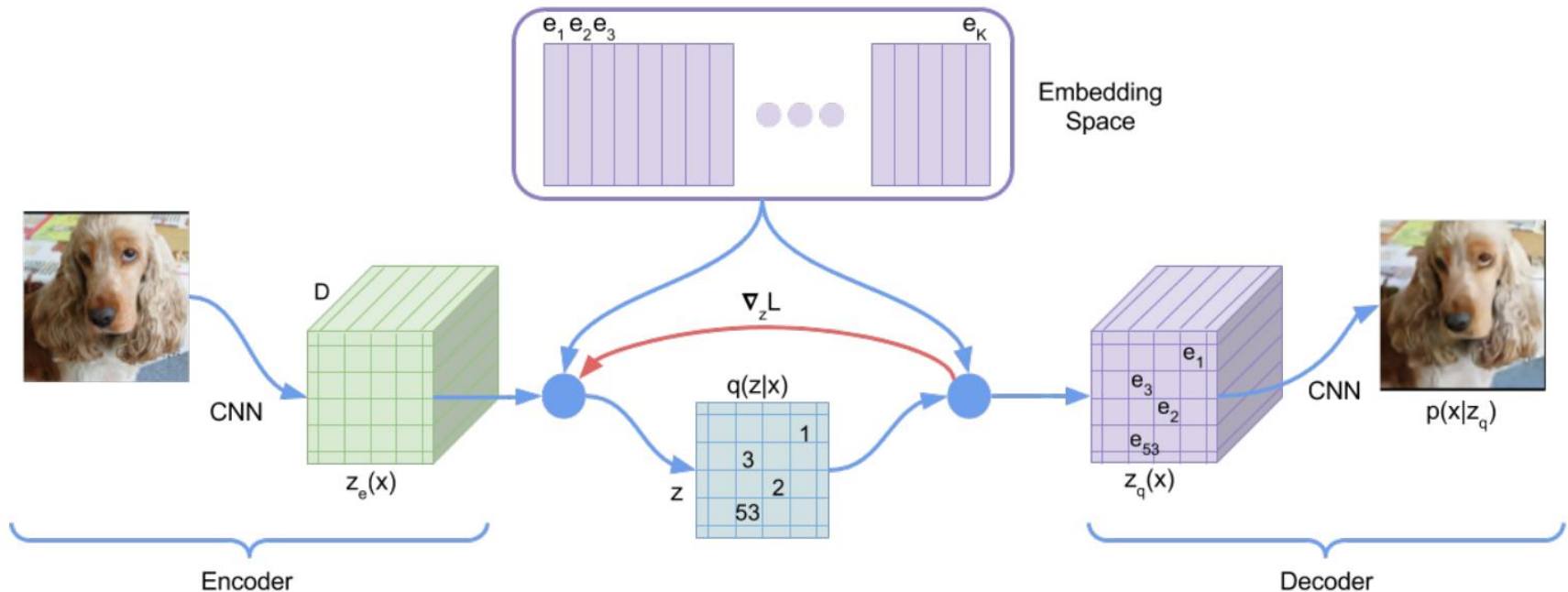
Autre méthode inspirée des VAE: VQ-VAE

- L'encodeur produit en sortie une matrice de taille $[M \times M \times D]$ qui correspond à $[M \times M]$ vecteurs de taille D



Autre méthode inspirée des VAE: VQ-VAE

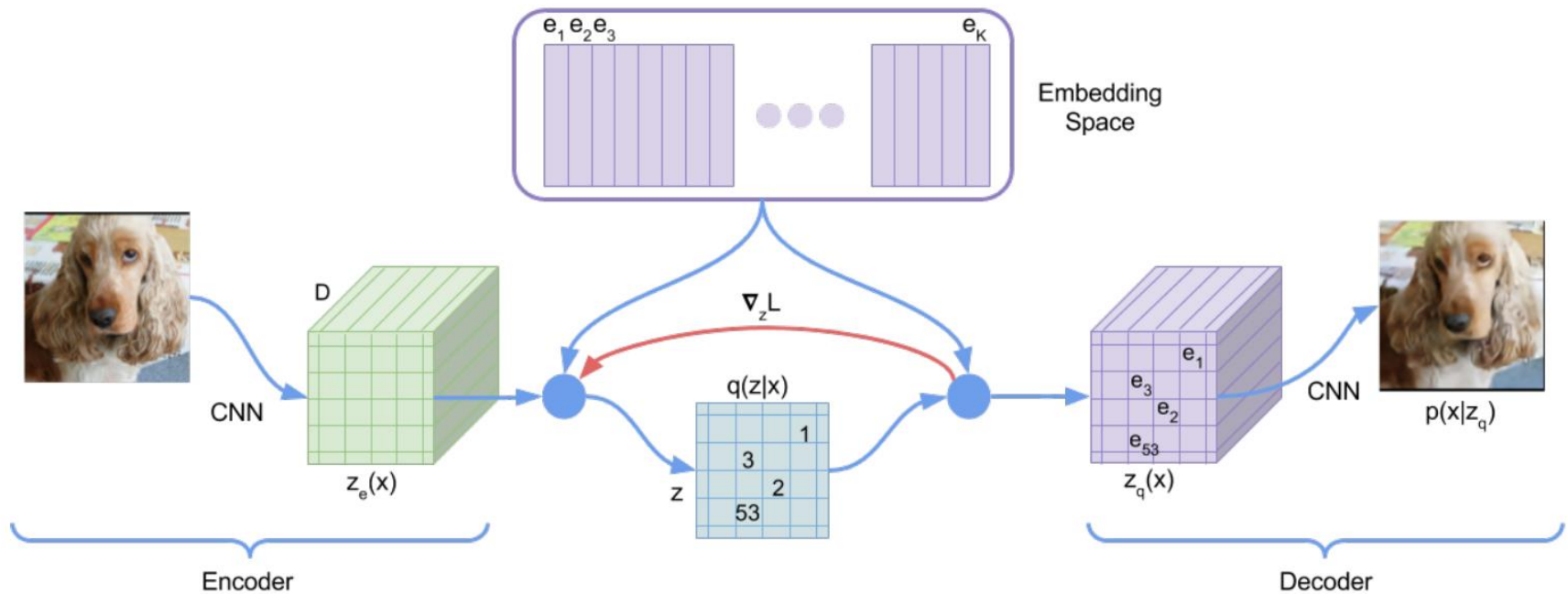
- Chaque vecteur de l'encodeur est comparé aux vecteurs de l'espace latent, et le numéro du vecteur le plus proche est affecté dans l'espace discret $q(z|x)$



Autre méthode inspirée des VAE: VQ-VAE

- La fonction de perte à minimiser est la suivante

$$\mathcal{L} = \log(p(x|z_q)) + \|sg[z_e(x)] - e\|_2 + \beta \|z_e(x) - sg[e]\|_2$$



$sg[\cdot]$: Identité en forward et nulle en backward

That's all folks

► Représentation de données

