

CONDITIONAL GENERATIVE MODELING VIA LEARNING THE LATENT SPACE

Ramsinghe et al. ICLR 2021

Poster réalisé par LAFON Marc, BENOTSMANE Ismat et LAN Tianwei, M2 DAC Université Sorbonne Sciences

POINTS CLÉS

- Génération conditionnelle dans des espaces multimodaux et continus
- Optimisation de l'espace latent en apprentissage et en inférence
- Pas d'objectif adversaire
- Convergence plus rapide et plus stable

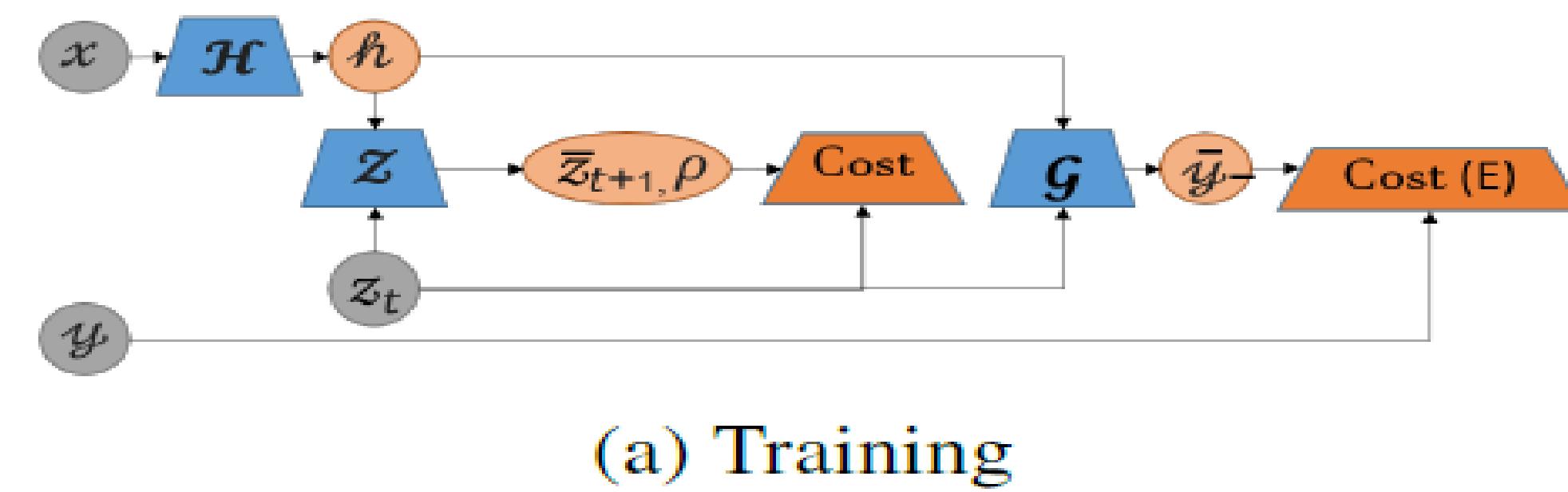
INTRODUCTION

Depuis leur introduction, les GAN et les GAN conditionnels (cGAN) suscitent un grand intérêt de part leur performance et les possibilités qu'ils ouvrent pour l'apprentissage non-supervisé. En revanche, ils présentent certains inconvénients : une grande instabilité lors de l'apprentissage, des prédictions déterministes et mono-modales en inférence et une utilisation d'un coût de reconstruction en contradiction avec l'objectif adversaire.

L'article de Ramasinghe et al. [2021] se place dans le cadre de la génération conditionnelle dans des espaces multi-modaux continus (CMM) et suggère de s'affranchir de l'objectif adversaire. À la place, les auteurs proposent d'apprendre une structure de l'espace latent permettant de trouver de "bons" vecteurs latents atteignables en un nombre limité d'itérations. Ce nouveau cadre, que les auteurs nomment cGML pour génération conditionnelle par modélisation de l'espace latent, améliore la stabilité, élimine la contradiction des loss adverses et de reconstruction et permet des résultats de meilleure qualité tout en assurant la diversité des prédictions.

MÉTHODOLOGIE

Considérons un dataset d'entraînement pour la génération multimodale conditionnelle constitué des conditions $\mathcal{X} = (x_1, \dots, x_J)$ et de cibles pour la génération $\mathcal{Y} = (y_1, \dots, y_J)$. Les cibles étant multimodales, pour $j \in J$, il existe K_j modes différents possibles : $y_j = (y_{j,1} \dots y_{j,K_j})$.



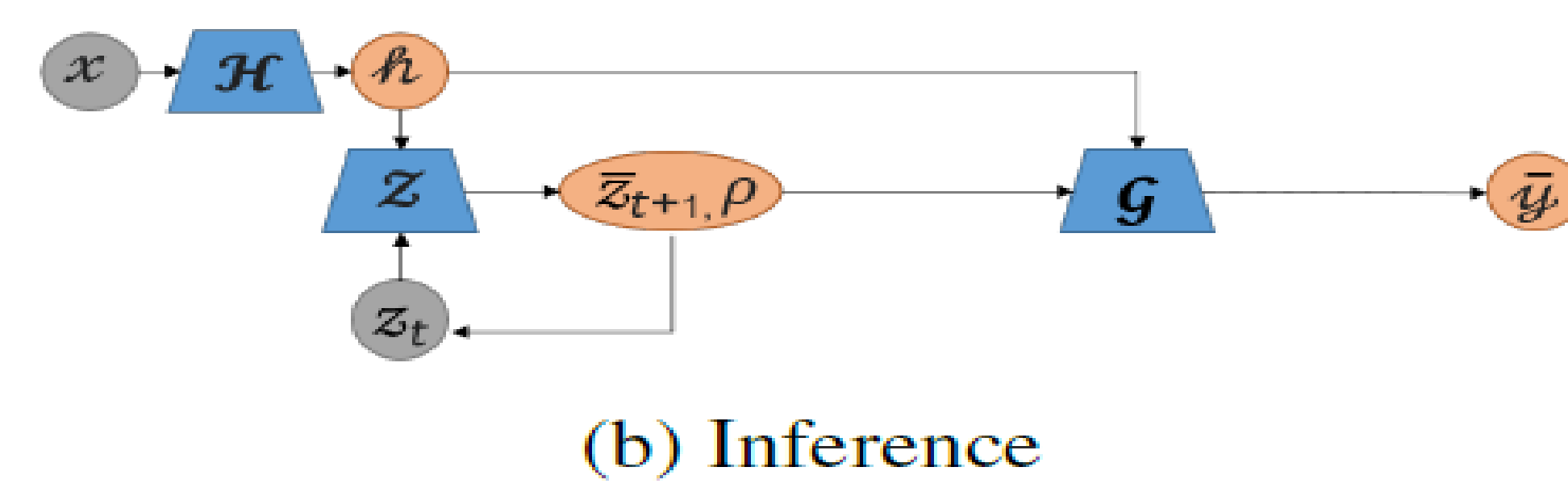
L'architecture proposée est constituée d'un encodeur \mathcal{H} prenant en entrée une condition $x \in \mathcal{X}$, d'un générateur conditionnel \mathcal{G} prenant en entrée la condition encodée $\mathcal{H}(x)$ et un vecteur latent z et un optimiseur \mathcal{Z} dont l'objectif est, à partir d'un vecteur latent z et de l'encodage $\mathcal{H}(x)$, de trouver un vecteur latent \hat{z} permettant au générateur de réaliser une meilleure génération que s'il avait utilisé le vecteur latent initial. Ce nouveau vecteur latent est mis à jour par descente de gradient avec momentum. Afin de pouvoir effectuer le même procédé d'optimisation de l'espace latent lors de l'inférence, le réseau \mathcal{Z} est entraîné à essayer de reproduire la même mise à jour.

Algorithm 1 Apprentissage

Input : $\mathcal{X} = (x_1, \dots, x_J)$, $\mathcal{Y} = (y_1, \dots, y_J)$, $\alpha, \beta, \gamma > 0$
for $k = 1, \dots, N$ **do**
 for $(x, y) \in \mathcal{X} \times \mathcal{Y}$ **do**
 $h = \mathcal{H}(x)$
 $z_t \sim \mathcal{U}_{B(0,1)}$
 $m_{t-1} = 0$
 for $p = 1, \dots, M$ **do** ▷ \mathcal{G}, \mathcal{H} figés
 $m_t = \gamma m_{t-1} + \beta \nabla_z \mathcal{L}_1(\mathcal{G}(h, z_t), y)$
 $z_{t+1} = z_t - m_t$
 $\rho = \|z_{t+1} - z_t\|$
 $w_z \leftarrow w_z - \alpha \nabla_{w_z} \mathcal{L}_1(\mathcal{Z}(h, z_t), (z_{t+1}, \rho))$ ▷ z figés
 $z_t = z_{t+1}$
 $w \leftarrow w - \alpha \nabla_w \mathcal{L}_1(\mathcal{G}(h, z_t), y)$ ▷ \mathcal{Z} et z figés

CORRECTION AVEC MOMENTUM

En inférence, le vecteur z_t traverse l'espace latent jusqu'à un vecteur optimal z^* , dès lors on a $z_{t+1} \approx z_t \approx z^*$. Cependant, dans le cas où la distribution des modes est parfaitement symétrique, il est possible qu'un vecteur latent z_t se trouve à équidistance entre deux modes de sorte qu'il est équiprobable de se diriger vers l'un ou l'autre. Ainsi, on aura $\mathbb{E}[z_{t+1}] \approx z_t$ et ce même loin de l'optimum. Afin d'éviter de converger vers ces *distributions fantômes*, le réseau \mathcal{Z} devra également prédire l'amplitude ρ de la mise à jour effectuée qui servira de coefficient de momentum pour la mise à jour lors de la prédiction en inférence.



Plus précisément, en inférence on itérera les mises à jours en utilisant le réseau \mathcal{Z} avec l'équation suivante :

$$\hat{z}, \hat{\rho} = \mathcal{Z}(\mathcal{H}(x), z_t)$$

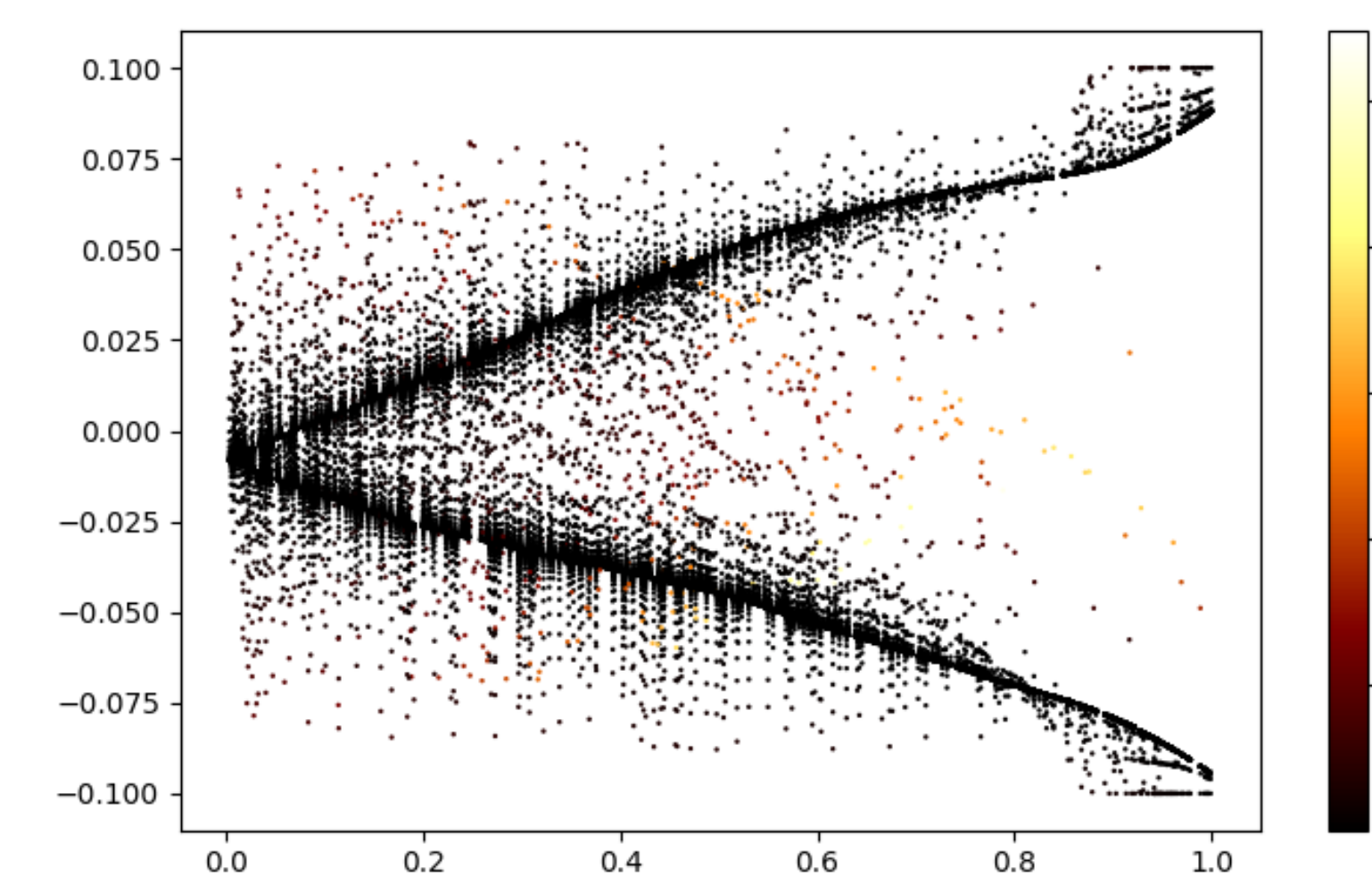
et

$$z_{t+1} = z_t + \hat{\rho} \frac{\hat{z} - z_t}{\|\hat{z} - z_t\|}$$

Une fois le vecteur latent optimal z^* obtenu, on peut générer une image en sortie avec le réseau \mathcal{G} .

CONTRAINTE LIPSCHITZ

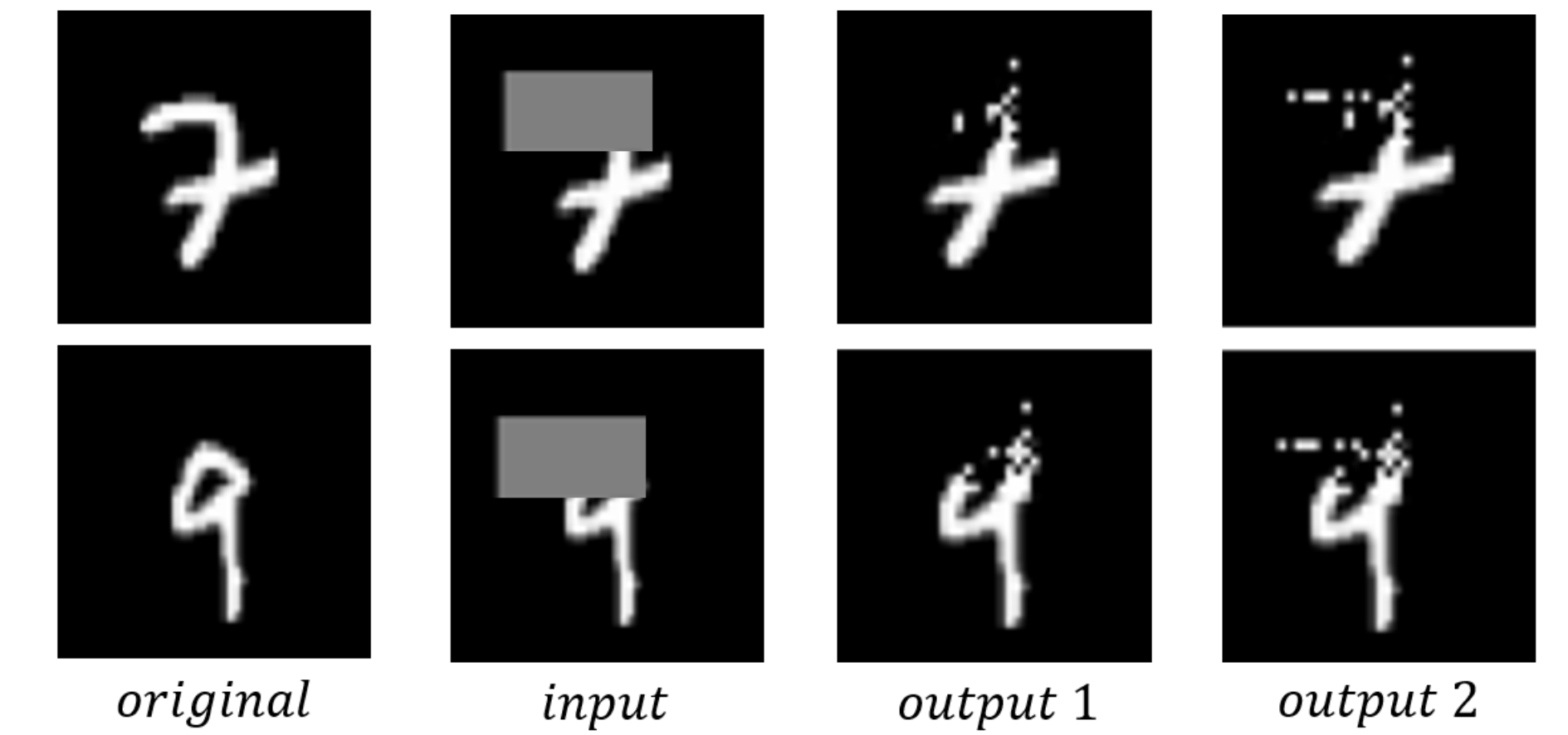
Afin d'assurer la stabilité des prédictions des réseaux \mathcal{G} et \mathcal{Z} lors de la mise à jours des vecteurs latents, les auteurs proposent d'imposer une contrainte de Lipschitz sur \mathcal{G} et \mathcal{Z} . Pour ce faire, les entrées des réseaux sont perturbées par un petit bruit aléatoire et les réseaux sont pénalisés pour des écarts importants entre les valeurs prédites. Cette contrainte permet ainsi de fournir à l'espace latent une structure pour laquelle la distance entre 2 vecteurs latents optimaux borne la différence entre deux modes optimaux générés à partir de ces vecteurs latents et d'une même condition x . On peut voir sur la figure ci-dessous la structure de l'espace latent (en ordonné) en fonction de la condition $x \in \mathcal{X}$ pour l'exemple jouet décrit ci-après.



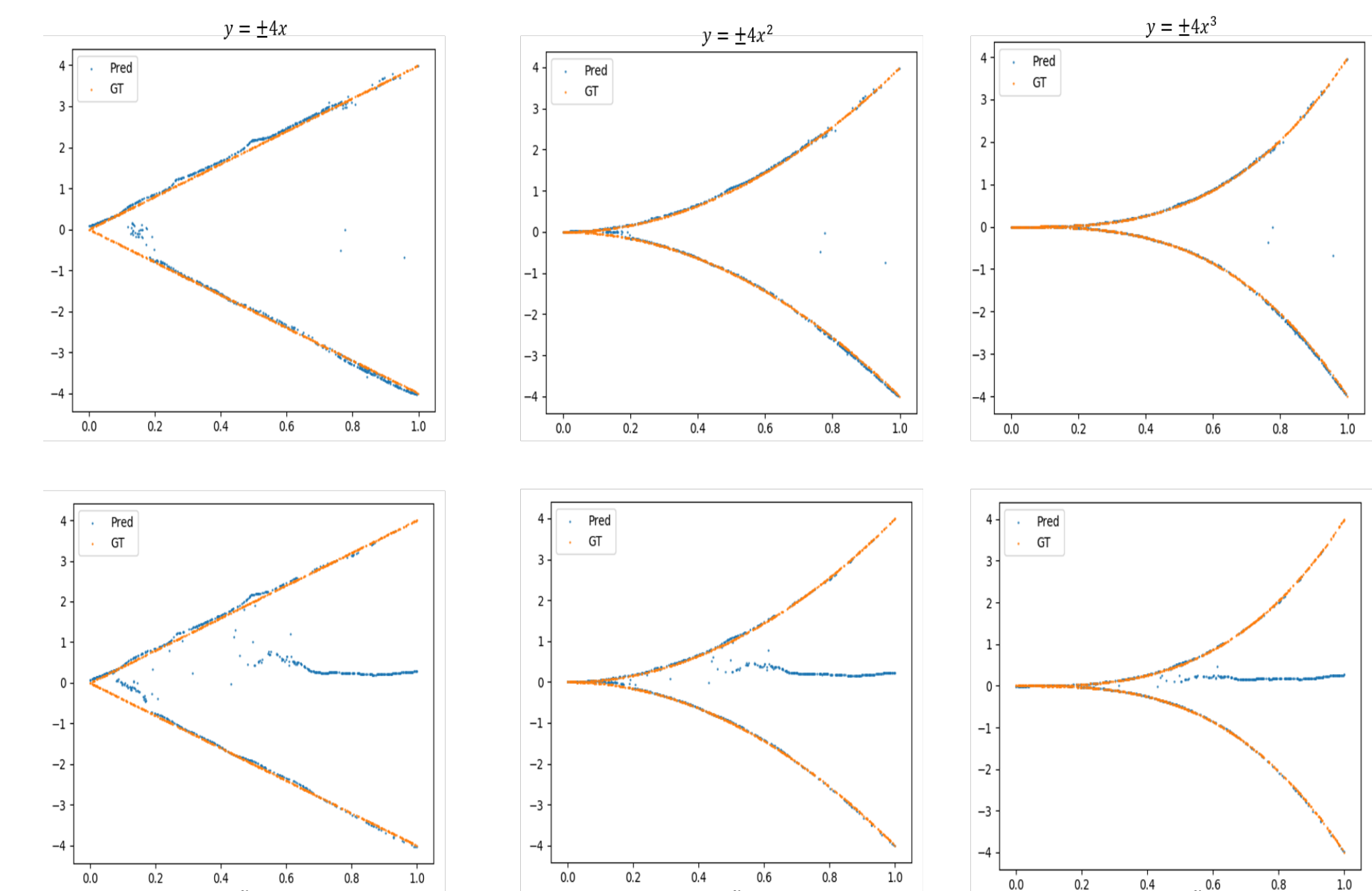
RÉSULTATS

Nous avons transformé le dataset MNIST original qui est mono-modal par défaut, en ajoutant un trait blanc à une position fixée sur 30% des données lors de l'apprentissage pour obtenir deux modes. Un mode qui représente les

données par défaut, et un autre mode qui représente les données avec un trait blanc. On peut voir ci-dessous que pour une même entrée, le modèle peut converger vers les 2 modes en sortie en fonction du vecteur latent initialement échantillonné.



Par ailleurs, nous avons également testé le modèle sur un exemple jouet pour lequel on considère un espace à deux modes en dimension 3 : $y = \pm(4x, 4x^2, 4x^3)$. On peut voir la première ligne du graphe ci-dessous que le modèle arrive bien à modéliser la distribution multimodale, on voit également un exemple de distribution fantôme sur la deuxième ligne.



REFERENCES

S. Ramasinghe, K. N. Ranasinghe, S. Khan, N. Barnes, and S. Gould. Conditional generative modeling via learning the latent space. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=VJnrYcnRc6>.