Synthèse d'images basée sur le VAE Hamiltonien pour l'amélioration de la segmentation de tumeurs

Aghiles KEBAILI Jérôme LAPUYADE-LAHORGUE Pierre VERA Su RUAN

Univ Rouen Normandie, INSA Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108, Quantif, F-76000 Rouen, France

Résumé – Malgré l'utilisation croissante de l'apprentissage profond dans la segmentation d'images médicales, l'acquisition de grande base de données étiquetée reste un défi dans le domaine médical. En réponse à cela, des techniques d'augmentation de données ont été proposées. Cependant, générer des images médicales diverses et réalistes ainsi que leurs masques correspondants reste une tâche difficile lorsqu'on travaille avec des ensembles de données insuffisants. Pour résoudre ces limitations, nous présentons une nouvelle architecture basée sur l'auto-encodeur Hamiltonien (HVAE), qui offre une représentation d'espace latent plus expressive et une meilleure approximation de la distribution postérieure par rapport aux auto-encodeurs variationnels (VAE) vanilla. De plus, nous introduisons une régularisation discriminative pour améliorer davantage la qualité des images générées. Nous avons testé notre méthode sur la base de données BRATS 2020. Les résultats ont démontré des améliorations en termes de diversité d'image et de qualité de représentation du masque tumoral par rapport aux VAEs et aux réseaux antagonistes génératifs par moindre carré (LSGAN) sur de petites bases de données.

Abstract – Despite the increasing use of deep learning in medical image segmentation, acquiring enough training data remains a challenge in the medical field. In response to this, data augmentation techniques have been proposed; However, generating diverse and realistic medical images and their corresponding masks remains a challenging task, when working with insufficient training sets. To address these limitations, we present a new architecture based on the Hamiltonian variational autoencoder (HVAE), which provides a more expressive latent space model and a better approximation of the posterior distribution compared to vanilla variational autoencoders (VAEs). In addition, we introduce a discriminative regularization to further improve the quality of generated images. We conducted experiments on the publicly available dataset Brain Tumor Segmentation challenge, where our proposed methods demonstrate improvements in image diversity and the quality of tumor mask representation when compared to VAEs and Least Squares generative adversarial network (LSGAN) under a data-scarce regime.

1 Introduction

Les tâches de segmentation médicale ont récemment bénéficié des avancées de l'apprentissage profond, montrant des résultats convaincants dans diverses modalités, comme l'imagerie par résonance magnétique (IRM) [12]. Cependant, le manque de données médicales constitue un défi majeur, en particulier lors de la nécessité de contourner manuellement les tumeurs, une tâche fastidieuse et chronophage pour les médecins. L'augmentation de données est une technique permettant de générer de nouveaux échantillons artificiels pour enrichir l'ensemble d'apprentissage. Les techniques avancées d'augmentation de données basées sur l'apprentissage profond, telles que les Modèles Génératifs Antagonistes (GAN) [3], ont été proposées pour générer des échantillons réalistes. Cependant, les GAN présentent des limites, notamment l'instabilité de l'apprentissage, les difficultés de convergence et le phénomène de l'effondrement des modes (*mode collapse*) où le générateur ne peut produire qu'une plage limitée d'échantillons possibles. En alternative, les VAEs [7] ont suscité l'intérêt en tant qu'approche d'augmentation de données surpassant les GAN en termes de diversité et de couverture des échantillons. Diverses méthodes ont été proposées afin d'améliorer les performances des VAEs, en réponse à leur propension à générer des images floues. Ces méthodes peuvent être classées en trois approches. La première approche implique l'utilisation de distributions a priori plus complexes que la distribution normale standard [11]. La deuxième consiste à améliorer l'approximation de la distribution postérieure dans la borne inférieure de l'évidence (ELBO), ce qui conduit à une meilleure approximation de la vraie distribution des données [6]. La troisième approche, qui a été moins explorée jusqu'à présent, elle consiste à modéliser l'espace latent en utilisant des géométries non-Euclidiennes [5]. Peu d'études ont exploré l'utilisation des VAEs pour l'augmentation de tâches de segmentation, car cela nécessite la généralisation à la fois de l'image et du masque correspondant. Le travail de Huo et al. [4] est l'un des rares à avoir exploré cette problématique en utilisant un *Progressive Adversarial* VAE, bien qu'il se limite à la génération de régions tumorales insérées naïvement sur des images authentiques pour produire des échantillons synthétiques.

Dans cette étude, nous présentons une nouvelle approche d'augmentation basée sur l'architecture du VAE Hamiltonien [1]. Cette architecture, qui intègre la dynamique Hamiltonienne, est plus expressive que les VAEs traditionnels. Elle offre une borne plus étroite sur la log-vraisemblance des données, ce qui permet une meilleure sélection de modèle et une estimation plus précise de la distribution réelle des données. Notre approche est spécifiquement adaptée à la segmentation des tumeurs, une tâche fastidieuse dans l'imagerie médicale. Ainsi, la génération d'images et de masques de tumeurs associés constitue une solution efficace pour augmenter la quantité de données disponible. Pour améliorer la qualité des échantillons générés, nous incorporons des termes de régularisation

supplémentaires dans la fonction de perte. Cette stratégie de régularisation améliore la qualité visuelle et la netteté des échantillons générés, tout en préservant la capacité du HVAE à produire une diversité d'images et à éviter un effondrement des modes. Les contributions de cette étude sont les suivantes :

- Une approche générative combinant un HVAE avec une méthode d'apprentissage antagoniste.
- Une méthode permettant la génération simultanée d'images médicales et des masques de segmentation correspondants.
- Une amélioration significative de la segmentation des tumeurs par rapport aux méthodes de l'état de l'art.

2 Méthode proposée

Notre approche exploite les HVAE afin d'améliorer la capacité de génération d'images tout en conservant la compacité de la représentation latente. Nous avons introduit une régularisation discriminative supplémentaire dans la fonction de perte, combinant ainsi les avantages des VAE et de l'apprentissage antagoniste pour améliorer la qualité des images générées. De plus, nous avons incorporé une perte perceptuelle pour améliorer la réalité perceptive des images. L'architecture de notre approche est présentée dans la Figure 1.

2.1 L'autoencodeur variationnel Hamiltonien

Le HVAE est une modification du VAE originel qui permet une meilleure approximation de la distribution a posteriori $p_\theta(z|x)$ de l'espace latent, conduisant à une meilleure qualité d'échantillonnage. Le HVAE est basé sur la dynamique d'Hamilton, qui consiste à traiter les variables latentes comme des particules se déplaçant dans un espace de grande dimension en fonction des principes de la conservation de l'énergie et de la quantité de mouvement [1]. L'idée derrière le HVAE est d'introduire une variable auxiliaire $\rho \sim \mathcal{N}(0, \mathbf{M})$ appelée *momentum*, et d'utiliser la dynamique Hamiltonienne pour la génération. L'Hamiltonien dans HVAE est défini selon :

$$\mathcal{H}(z, \rho) = -\log p_\theta(z|x) + \frac{1}{2}\rho^T \mathbf{M}^{-1}\rho \tag{1}$$

Où le premier terme est l'énergie potentielle et le second, l'énergie cinétique. $\mathbf{M}$ peut s'interpréter comme une masse. Le HVAE utilise une technique d'échantillonnage de Monte Carlo pour exploiter la dynamique d'Hamilton et explorer plus efficacement la distribution a posteriori. L'idée est d'échantillonner (z, ρ) en utilisant la dynamique, ce qui conduit à la création d'une chaîne de Markov ergodique et réversible dans le temps pour z, dont la distribution stationnaire correspond à la distribution cible. ρ et z sont mis à jour en utilisant :

$$\frac{\partial \rho}{\partial t}(t) = -\nabla_z \mathcal{H}, \quad \frac{\partial z}{\partial t}(t) = \nabla_\rho \mathcal{H}, \tag{2}$$

où t est l'étape de la chaîne de Markov et ∇_u le gradient respectif selon u. Les détails de la discrétisation des équations (2) peuvent être consultés dans [1]. Une proposition finale z_K, où K est le nombre total d'étapes, est générée et acceptée en utilisant l'algorithme de rejet-acceptation avec le taux de Metropolis-Hastings [1], qui est utilisé pour équilibrer l'exploration entre l'état z_K et l'exploitation de l'état courant z_{K-1}.

De manière similaire aux VAEs, le HVAE repose également sur une technique d'inférence variationnelle qui estime la distribution postérieure $p_\theta(z|x)$ en maximisant la borne inférieure de l'évidence (ELBO) à l'aide d'une distribution plus simple $q_\phi(z|x)$. θ et ϕ étant les paramètres de la distribution postérieure et de la distribution variationnelle. En introduisant la méthode de Monte-Carlo Hamiltonien (HMC) [1], certaines composantes de la fonction sont réévaluées analytiquement, ce qui conduit à la formulation suivante du terme ELBO :

$$\begin{aligned}\mathcal{L}^{\text{H}}_{\text{ELBO}} = \mathbf{E}_{z_K \sim q_\phi(z|x)}[&\log p_\theta(x, z_K) \\ &- \frac{1}{2}\rho_K^T \mathbf{M}^{-1}\rho_K - \log q_\phi(z_K|x)]\end{aligned} \tag{3}$$

Ici, $q_\phi(z_K|x) \sim \mathcal{N}(z_K; \mathbf{0}, \mathbf{I})$ représente la log-vraisemblance du vecteur latent z_K.

2.2 Régularisation par reconstruction de caractéristiques

Les fonctions de reconstruction par pixels offrent des avantages en maintenant les structures globales, mais ne parviennent pas à reproduire fidèlement les subtilités de la perception humaine et peuvent entraîner des pertes de détails. Pour résoudre ces problèmes, un terme de régularisation sur les caractéristiques est ajouté à la perte globale. Il est basé sur l'activation ReLU des 16 couches d'un réseau de type VGG pré-entraîné pour extraire les caractéristiques intermédiaires, et encourager la sortie de l'image à avoir des représentations similaires. Aussi appelée régularisation perceptuelle, elle permet d'obtenir des détails de texture pour une meilleure identification de la tumeur :

$$\mathcal{L}_{feature}(\hat{x}, x) = \sum_j^L \left(\phi_j(\hat{x}) - \phi_j(x)\right)^2 \tag{4}$$

Où L est le nombre de couches total, $\phi_j(x)$ est la carte de caractéristiques de la j-ème couche du réseau ϕ pour une image d'entrée x, et $\phi_j(\hat{x})$ pour l'image reconstruite.

2.3 Régularisation discriminative

Nous proposons également d'intégrer les avantages des VAEs et de l'apprentissage antagoniste en utilisant un terme de régularisation discriminatif pour améliorer la qualité visuelle et la netteté des images générées. Ceci permet de résoudre partiellement les problèmes liés à la génération floue des VAEs, tout en évitant un *mode collapse*. Le terme de régularisation discriminatif est défini comme suit :

$$\begin{aligned}\mathcal{L}_{\text{disc}} &= \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[\log D_\theta(x)\right] \\ &+ \mathbb{E}_{z \sim p(z)}\left[\log\left(1 - D_\theta(G_\phi(z))\right)\right]\end{aligned} \tag{5}$$

Où D_θ représente le discriminateur, G_ϕ correspond à la fonction du HVAE qui agit comme générateur. Les paramètres des deux modèles sont notés θ et ϕ respectivement. $p_{\text{data}}(x)$ représente la distribution des échantillons de données réelles, tandis que $p(z)$ désigne la distribution a priori des variables latentes. Pour atténuer les instabilités de l'apprentissage antagoniste, nous proposons d'inclure un coefficient de régularisation,

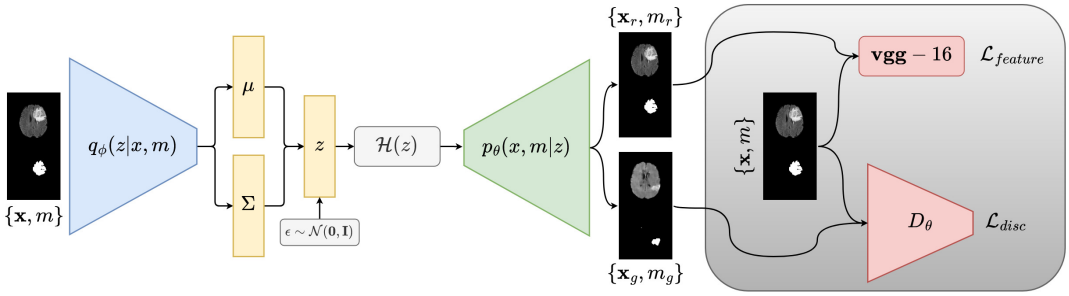

FIGURE 1 : L'architecture proposée se compose d'un d'encodeur q_ϕ et de décodeur p_θ, d'un discriminateur D_θ qui fait office de régularisateur, et d'un VGG pré-entraîné à 16 couches pour la reconstruction de caractéristiques. Cette architecture prend en entrée une image médicale et son masque de tumeur correspondant, noté $\mathbf{x}, m$, et la reconstruit en $\mathbf{x}_r, m_r$. Les images nouvellement générées sont notées $\mathbf{x}_g, m_g$.

permettant de modéliser le processus comme un apprentissage traditionnel visant à minimiser la perte du HVAE, notée $\mathcal{L}^{\text{H}}_{\text{ELBO}}$ (référence à l'article original [1]). Cela assure une stabilité d'entraînement. La fonction de perte globale peut alors être résumée comme suit :

$$\mathcal{L}_{\text{globale}} = \mathcal{L}^{\text{H}}_{\text{ELBO}} + \lambda_1 \mathcal{L}_{\text{disc}} + \lambda_2 \mathcal{L}_{\text{feature}} \tag{6}$$

Les valeurs de λ_1 et λ_2 ont été déterminées par une expérimentation extensive et ont été toutes les deux fixées à 0,01.

3 Expériences

Nous avons effectué une analyse comparative entre notre architecture, désignée sous le nom de dHVAE (pour *discriminative* HVAE), et plusieurs modèles génératifs de l'état de l'art. Ces modèles incluent le VAE, le HVAE et le LSGAN [8]. À l'heure actuelle, notre méthode se limite à la génération d'images 2D, mais son extension à des images 3D est possible grâce à une puissance de calcul accrue. L'évaluation a été menée sur l'ensemble de données BRATS. Les performances de chaque modèle ont été évaluées en utilisant le coefficient de similarité de Dice (DSC) pour la tâche de segmentation. La qualité des images générées a été évaluée à l'aide de deux mesures : le rapport signal sur bruit maximal (PSNR) (Tableaux 1 et 2).

3.1 Jeu de données

Le jeu de données BRATS [9] comprend 1258 sujets, avec des scans IRM et des masques de segmentation contenant 3 étiquettes de tumeurs. Les images ont une dimension de 240×240 pixels. Dans notre étude, nous évaluons l'efficacité de notre modèle en utilisant uniquement la modalité FLAIR. Un ensemble de base de 100 patients a été sélectionné de manière aléatoire à partir de l'ensemble des patients pour l'apprentissage de notre modèle.

3.2 Paramètres d'apprentissage

Les modèles génératifs proposés sont conditionnés par les masques de tumeur, ce qui est réalisé en concaténant naïvement les masques aux images médicales correspondantes, formant ainsi une entrée à plusieurs canaux. Bien qu'il soit possible de conditionner le modèle au niveau de l'espace latent pour plus de liberté de génération, nous avons choisi cette approche en raison de sa facilité de mise en œuvre. Afin de reproduire des scénarios concrets où les données sont limitées, nous avons restreint l'ensemble d'entraînement à seulement 100 patients. Il est important de noter qu'un U-Net [10] n'obtient un score DSC acceptable qu'à partir de 500 patients. L'évaluation a été réalisée sur 100 autres patients, et le U-Net a été entraîné sur chaque configuration expérimentale mentionnée dans le Tableau 1 et les résultats de la DSC moyenne et de l'écart type sont rapportés sur 10 exécutions indépendantes.

3.3 Évaluation quantitative et visuelle

Nous commençons par présenter la DSC de référence obtenue en entraînant un U-Net sur l'ensemble de base composé de 100 images réelles. Ensuite, nous avons utilisé des modèles génératifs profonds de l'état de l'art, y compris l'architecture proposée, pour générer des images et des masques synthétiques correspondants. Plus précisément, nous avons généré 100, 200, 300 et 500 images synthétiques, que nous avons ensuite mélangées aux 100 images de l'ensemble d'entraînement réel. Cette partie correspond au processus d'augmentation.

TABLE 1 : Performance quantitative évaluée en termes de DSC, en utilisant l'ensemble BRATS. La première ligne correspond au score DSC (%) de référence, sur les lignes suivantes on fait varier le nombre d'images synthétiques ajoutées au données d'apprentissage.

	LSGAN	VAE	HVAE	dHVAE (notre)
Réf.		0.633		
+ 100	0.680±0.03	0.751±0.01	0.770±0.03	0.773±0.05
+ 200	0.703±0.04	0.760±0.01	0.781±0.02	0.802±0.08
+ 300	0.685±0.02	0.752±0.04	0.803±0.07	**0.819**±0.01
+ 500	0.667±0.08	0.726±0.06	0.760±0.10	0.796±0.01

Notre architecture améliore de manière significative le DSC sur la base de données BRATS, avec une amélioration de 29,3% par rapport à la référence. De plus, l'écart-type de la méthode proposée est très faible (∼ 1% DSC), confirmant sa robustesse contre la variabilité des données. En termes de PSNR et de SSIM, notre architecture dépasse également les méthodes de l'état de l'art, avec un PSNR de 13,709 dB et un SSIM de

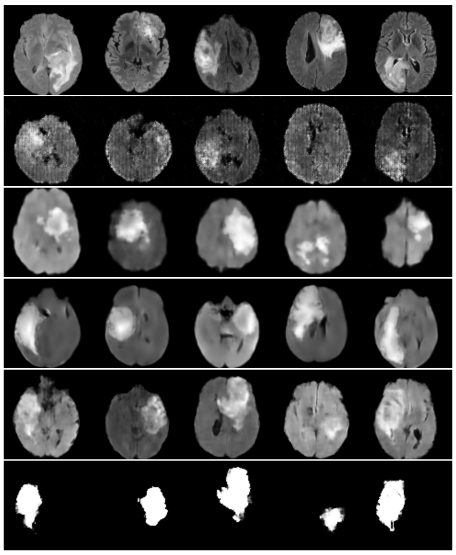

FIGURE 2 : Comparaison de différentes architectures pour la génération d'IRMs synthétiques. Les images de la première ligne sont les images originales, suivies des images synthétisées avec les architectures LSGAN, VAE, HVAE et dHVAE. La dernière ligne présente les masques de segmentation générées par le dHVAE.

TABLE 2 : Performance quantitative évaluée en termes de PSNR et SSIM entre l'ensemble de test et les ensembles générés.

Architecture	PSNR (dB)	SSIM (%)
LSGAN	12.932	0.613
VAE	13.322	0.835
HVAE	13.611	**0.844**
dHAVE (notre)	**13.709**	**0.844**

84,4% calculés entre 100 images synthétiques et réelles. La méthode proposée permet de générer des images médicales entières avec leurs masques de tumeurs correspondants, obtenant des résultats de meilleure qualité, comme illustré dans la Figure 2. On y constate une amélioration de la qualité en passant du VAE au dHVAE, ainsi qu'une représentation plus précise des zones tumorales sur les IRMs. On remarque également que l'architecture GAN est particulièrement affectée en raison de ses exigences élevées en quantité de données. Ces observations confirment l'hypothèse sur le potentiel des VAEs face aux ensembles de données insuffisants. Nous observons également sur le tableau 1 une diminution du score DSC au-delà de l'ajout de 500 images synthétiques. Cette baisse peut s'expliquer par le déséquilibre entre le nombre d'images réelles et synthétiques, où le modèle favorise l'optimisation sur les images synthétiques, impactant ainsi négativement le score DSC sur l'ensemble de test.

4 Conclusion

Dans cette étude, nous avons présenté une nouvelle architecture hybride combinant une régularisation discriminative avec le modèle HVAE pour l'augmentation des données dans les tâches de segmentation d'images médicales. Les résultats expérimentaux ont démontré que notre méthode proposée surpassait les modèles génératifs de l'état de l'art en termes DSC sur des petits ensembles de données. Nos orientations futures consistent à étendre l'approche proposée aux images médicales en 3D et à explorer la modélisation de l'espace latent en tant que variété Riemannienne.

Références

[1] A. L CATERINI, A. DOUCET et D. SEJDINOVIC : Hamiltonian variational auto-encoder. *Advances in Neural Information Processing Systems*, 31, 2018.

[2] C. CHADEBEC, E. THIBEAU-SUTRE, N. BURGOS et S. ALLASSONNIÈRE : Data augmentation in high dimensional low sample size setting using a geometry-based variational autoencoder. *IEEE TPAMI*, 2022.

[3] I. J GOODFELLOW, J. POUGET-ABADIE, M. MIRZA, B. XU, D. WARDE-FARLEY, S. OZAIR, A. COURVILLE et Y. BENGIO : Generative adversarial networks. *arXiv preprint arXiv :1406.2661*, 2014.

[4] J. HUO, V. VAKHARIA, C. WU, A. SHARAN, A. KO, S. OURSELIN et R. SPARKS : Brain Lesion Synthesis via Progressive Adversarial Variational Auto-Encoder. *In International Workshop on Simulation and Synthesis in Medical Imaging*, pages 101–111. Springer, 2022.

[5] A. KEBAILI, J. LAPUYADE-LAHORGUE et S. RUAN : Deep learning approaches for data augmentation in medical imaging : A review. *Journal of Imaging*, 9, 2023.

[6] D. P KINGMA, T. SALIMANS, R. JOZEFOWICZ, X. CHEN, I. SUTSKEVER et M. WELLING : Improved variational inference with inverse autoregressive flow. *Advances in neural information processing systems*, 29, 2016.

[7] D. P KINGMA et M. WELLING : Auto-encoding variational Bayes. *arXiv preprint arXiv :1312.6114*, 2013.

[8] X. MAO, Q. LI, H. XIE, R. YK LAU, Z. WANG et Stephen PAUL S. : Least squares generative adversarial networks. *In Proceedings of the IEEE ICCV*, pages 2794–2802, 2017.

[9] B. H MENZE, A. JAKAB, S. BAUER, J. KALPATHY-CRAMER, K. FARAHANI, J. KIRBY, Y. BURREN, N. PORZ, J. SLOTBOOM, R. WIEST *et al.* : The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE TMI*, 34(10):1993–2024, 2014.

[10] O. RONNEBERGER, P. FISCHER et T. BROX : U-net : Convolutional networks for biomedical image segmentation. *In MICCAI*, pages 234–241. Springer, 2015.

[11] J. TOMCZAK et M. WELLING : VAE with a VampPrior. *In International Conference on Artificial Intelligence and Statistics*, pages 1214–1223. PMLR, 2018.

[12] T. ZHOU, P. VERA, S. CANU et S. RUAN : Missing Data Imputation via Conditional Generator and Correlation Learning for Multimodal Brain Tumor Segmentation. *Pattern Recognition Letters*, 158:125–132, 2022.