"Dall-e Brain" : A generative prompt model for synthetic healthy brain images.

Chantal MULLER (CREATIS), chantal.muller@insa-lyon.fr
Thomas GRENIER (CREATIS), thomas.grenier@insa-lyon.fr
Carole FRINDEL (CREATIS), carole.frindel@insa-lyon.fr

24 novembre 2023

Mots-clés : Medical Imaging, Generative AI, Stable diffusion, Foundation Model.

1 Contexte

Les bases de données de sujets sains en imagerie du cerveau sont de taille très limitée et insuffisante pour l'étape d'apprentissage inhérente aux méthodes basées sur le Deep Learning. L'acquisition d'images IRM nécessite du temps et des ressources coûteuses qui ne peuvent être dépensés pour des patients sains. De ce fait, les modèles sont toujours entraînés à partir d'images présentant des pathologies diverses : glioblastomes, lésions de sclérose en plaque, Alzheimer, etc. Disposer d'un système capable de générer des images de patients sains permettrait de constituer les grandes quantités de données indispensables à l'apprentissage des réseaux, de maîtriser l'étude des pathologies, en ajoutant aux données des marqueurs spécifiques temporellement variants afin de créer des séquences temporelles extrêmement rares en imagerie médicale. Les techniques de génération d'images médicales à partir de descriptions textuelles offrent par conséquent une alternative économique et non invasive pour produire des images de qualité clinique.

2 Description du projet

De nombreux travaux ont été proposés pour générer des images médicales synthétiques à partir d'images réelles [1]. Des modèles génératifs tels que les GAN (generative adversarial networks [2]), les VAE (variational autoencoders [3]) et les DM (diffusion models, DDPM[4], DDIM [5]) ont conduit à des images synthétiques réalistes et conformes à la distribution des images originales.
En 2021, le modèle CLIP d'OpenAI (Constrative Language-vision model [6]) a contribué à une avancée majeure en fournissant une représentation commune au langage naturel et aux images extraites du WEB. Dès lors, l'association de modèles génératifs et du modèle CLIP a ouvert la voie à une nouvelle génération de modèles basés sur le prompting, comme Dall-E2 [7], Imagen [8], Midjourney et Stable Diffusion [9]. Ces modèles disponibles en ligne se sont largement répandus et sont bien connus du public pour leurs résultats spectaculaires, leur capacité à transformer de manière efficace, pertinente et créative des prompts en images. Cependant, les images médicales sont peu représentées lors de l'apprentissage des modèles et de ce fait les générations d'images médicales sont peu réalistes [10].
Ce projet de Master propose de pallier ce défaut en spécialisant un modèle génératif de type Stable Diffusion [11, 12] sur la synthèse d'images IRM du cerveau. L'idée est de focaliser l'apprentissage de ce modèle sur des paires "descriptions textuelles anatomiques / images IRM".

2.1 Architecture du modèle

L'architecture d'un modèle génératif par prompt de type unCLIP à la base de Dall-e2 se décompose en deux parties (cf. Fig. 1) :

— la partie au-dessus des pointillés est chargée du process d'apprentissage pour CLIP par lequel une représentation commune (embedding) pour le texte et les images est apprise. Après cette

étape, le texte et les images représentent leurs informations dans le même espace (latent space). Pendant la phase d'apprentissage, les paramètres sont optimisés pour que la similarité entre les paires Texte / Image soit maximale et minimale pour les autres appariements (cf. Fig. 2).

— en-dessous des pointillés, l'embeddeding d'un texte est fourni à un modèle de diffusion 'prior' pour produire l'embedding d'une image. L'embeddding de cette image est ensuite fournie à un décodeur pour revenir à l'espace image et produire l'image finale.

.

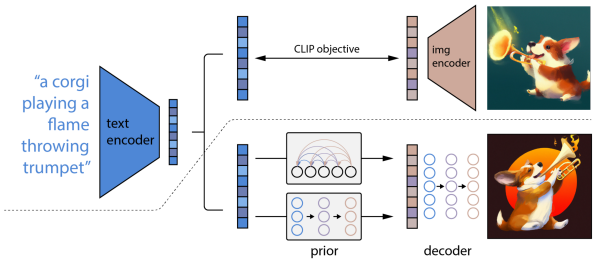

FIGURE 1 – Architecture du modèle unCLIP [7].

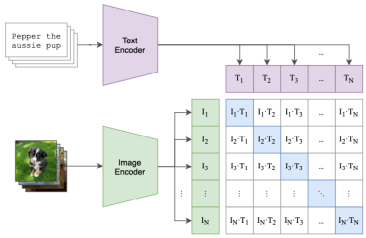

FIGURE 2 – Modèle CLIP - Contrastive pre-training

L'objet du Master est d'adapter le modèle à l'imagerie médicale comme illustré dans Fig. 3.

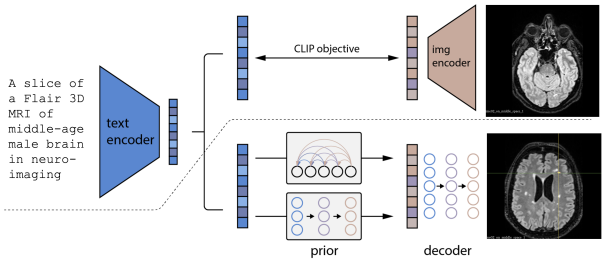

FIGURE 3 – Modèle Unclip spécialisé pour l'imagerie médicale.

2.2 Base de données

Le Master s'appuiera sur une base de données Images / Légendes préalablement créée à partir de bases d'images IRM publiques de sujets sains (OASIS [13], IXI [14], HCP [15]). L'ensemble des images sera normalisé et recalé avec un atlas cérébral. La description textuelle par mot-clés sera automatiquement générée à l'aide de l'atlas, des données patient, du volume de certaines structures, orientation de la coupe. Chaque image sera finalement associée à une légende en langage naturel créé à partir des mot-clés. Les données pourront être augmentées par diverses transformations géométriques.

2.3 Adaptation du modèle génératif

L'adaptation du modèle génératif est le cœur de ce travail de Master. Le travail s'articulera selon plusieurs phases :

— L'étude des modèles : CLIP, unCLIP, stable diffusion, modèles de diffusion dans l'espace latent.
— L'étude des solutions Open Source pour le modèle Stable Diffusion.
— La recherche des solutions pour spécialiser le modèle
  → L'enrichissement du modèle CLIP.
  → L'adaptation du modèle de diffusion dans l'espace latent.
  → L'ajustement fin par LoRA *Low-Rank Adaptation*, spécialisation de l'apprentissage de façon légère en s'attaquant aux couches d'attention croisée [16].

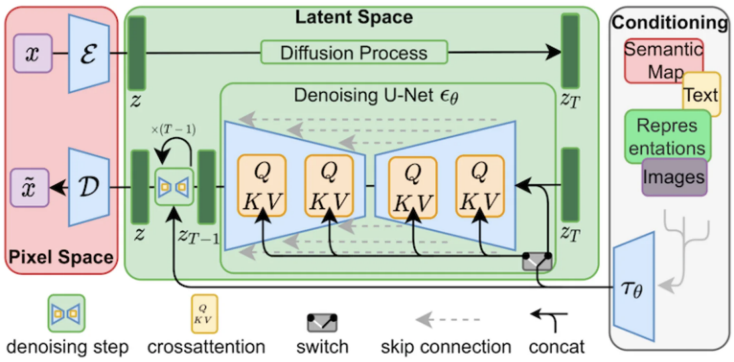

FIGURE 4 – *https://www.stablediffusion.blog/lora-stablediffusion*

2.4 Évaluation des images synthétiques

L'évaluation des images synthétiques sera menée selon plusieurs axes :

— qualitativement avec l'aide d'un expert médical (pertinence anatomique de l'image, conformité des contrastes avec la modalité visée).
— quantitativement (métriques pour comparer les images synthétiques avec les images réelles, similarité des distributions des niveaux de gris, analyse de texture).
— de manière exploratoire, dans l'espace latent pour mesurer la similarité entre les images réelles et synthétiques.

3 Résultats et perspectives

— Résultats :
  — Avoir des quantités de données suffisantes pour les modèles de fondations.
  — Résoudre le problème d'accès aux données, d'anonymisation, de confidentialité.
— Perspectives :
  — Obtenir une représentation d'un modèle normal.
  — Détecter les anomalies dans l'espace latent.
  — Maitriser l'étude des pathologies en rajoutant des marqueurs spécifiques.
  — Générer des séquences temporelles, avoir des modèles d'évolution et de prédiction.

4 Profil des candidats

— Etudiant(e) en dernière année de Master 2 ou école d'ingénieur.
— Solide formation en Deep learning et maitrise de Python ainsi que les librairies telles que Tensorflow/Pytorch.
— Bon niveau en anglais technique.
— Autonome, motivé et persévérant.

5 Informations générales

Pour postuler, envoyez vos CV + derniers bulletins de notes à l'adresse chantal.muller@creatis.insa-lyon.fr avec comme objet : **[Stage Dall-e Brain] Nom Prénom** .

Thème/Domaine : Machine Learning, traitement d'images médicales, AI.
Ville : Villeurbanne (69).
Lien : INSA Lyon, Laboratoire CREATIS.
Date de prise de fonction souhaitée : 2024-02-01.
Durée de la convention de stage : 6 mois.
Gratification : 614,26€.
Superviseurs : Chantal MULLER (CREATIS), Thomas GRENIER (CREATIS).

Références

[1] Aghiles Kebaili, Jérôme Lapuyade-Lahorgue, and Su Ruan. Deep learning approaches for data augmentation in medical imaging : A review. *Journal of Imaging*, 9(4), 2023.

[2] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans, 2016.

[3] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. *Advances in neural information processing systems*, 27, 2014.

[4] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.

[5] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.

[6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[7] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents, 2022. *URL https ://arxiv. org/abs/2204.06125*, 7, 2022.

[8] Su Wang, Chitwan Saharia, Ceslee Montgomery, Jordi Pont-Tuset, Shai Noy, Stefano Pellegrini, Yasumasa Onoe, Sarah Laszlo, David J Fleet, Radu Soricut, et al. Imagen editor and editbench : Advancing and evaluating text-guided image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18359–18369, 2023.

[9] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[10] Lisa C Adams, Felix Busch, Daniel Truhn, Marcus R Makowski, Hugo JWL Aerts, and Keno K Bressem. What does dall-e 2 know about radiology ? *Journal of Medical Internet Research*, 25 :e43110, 2023.

[11] https://github.com/Stability-AI/generative-models#readme.

[12] https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0.

[13] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies (oasis) : cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9) :1498–1507, 2007.

[14] http://brain-development.org/ixi-dataset/.

[15] https://www.humanconnectome.org/study/hcp-young-adult.

[16] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora : Low-rank adaptation of large language models, 2021.