

Étude de la Fusion des Représentations Latentes Texte-Image en IA Générative pour l'Imagerie Médicale

Intitulé du doctorat
(spécialité) :

Traitement du signal et de l'Image

[Details du Sujet](#)

Informations de synthèse pour le laboratoire :

Mots clés :

Modèles de diffusion latente, IA générative multimodale, espace latent, LLM

Résumé :

Cette thèse porte sur l'étude et le développement d'un espace latent partagé texte-image en intelligence artificielle générative, appliqué à l'imagerie médicale. L'objectif est de comprendre comment combiner efficacement des descriptions textuelles et des images médicales dans un espace commun, afin de générer des images 3D réalistes, d'analyser ces images et de produire automatiquement des rapports explicatifs pour les cliniciens.

Le projet s'appuiera sur des approches telles que les AR-VAE Attribute-based Regularized Variational Autoencoders qui permettent de régulariser l'apprentissage d'un espace latent en contrôlant certains attributs spécifiques. Les travaux seront appliqués à des IRM cérébrales de sujets sains, avec des images annotées en 2D et en 3D, associées à des descriptions textuelles par coupe et par volume, sur les séquences T1, T2 et T2-FLAIR. Cette base de données constitue un support essentiel pour la construction et l'analyse de l'espace latent partagé.

La génération d'images médicales 3D se fera à l'aide de modèles de diffusion latente (LDM) adaptés aux contraintes spécifiques de l'imagerie médicale. L'étude cherchera à optimiser ces modèles pour produire des images tridimensionnelles cohérentes et réalistes, capables de simuler l'évolution de biomarqueurs dans des pathologies comme la sclérose en plaques (SEP). Cette modélisation temporelle permettra de suivre des processus pathologiques, comme l'apparition et la progression de lésions cérébrales.

L'explicabilité des modèles constituera un axe important de cette recherche. Comprendre l'impact des modalités texte et image sur les décisions des modèles est indispensable pour garantir leur fiabilité en contexte clinique. Cette partie du travail consistera à développer des outils permettant d'explorer l'espace latent, mais aussi à générer automatiquement des rapports textuels détaillant les choix effectués par le modèle. Ces rapports devront fournir des explications compréhensibles aux cliniciens : ils indiqueront, par exemple, pourquoi certaines zones ont été identifiées comme pathologiques, quelles descriptions textuelles ont influencé la décision et comment ces deux modalités interagissent dans les prédictions.

Le développement de ces rapports s'appuiera sur l'intégration de modèles NLP spécialisés, comme BiomedBERT et MedPaLM, capables de relier les observations visuelles aux descriptions textuelles pertinentes. Cette approche facilitera l'interprétation clinique et participera à l'amélioration de la confiance des praticiens dans les décisions prises par les modèles.

La validation des résultats sera menée en collaboration avec le Center for Neurological Imaging dirigé par le Dr Charles Guttman (Harvard Medical School), et avec le soutien méthodologique de Juan Carlos Prieto, co-encadrant et expert en IA appliquée à l'imagerie médicale. Les modèles développés seront testés sur des données réelles et synthétiques, afin d'évaluer leur pertinence clinique, notamment pour le suivi de l'évolution de la SEP.

Les applications potentielles incluent la création d'outils d'aide au diagnostic et au suivi longitudinal de biomarqueurs cérébraux. Les résultats feront l'objet de

publications dans des revues et conférences internationales spécialisées (MICCAI, MIDL, IEEE TMI).

Points Forts :

Cette étude vise à explorer la fusion des représentations latentes texte-image en IA générative appliquée à l'imagerie médicale. L'objectif est de comprendre comment ces deux modalités peuvent interagir dans l'espace latent pour répondre à divers besoins cliniques.

1. Fusion des représentations texte-image :

L'étude s'intéresse à la combinaison des modalités texte et image pour une compréhension approfondie des relations entre descriptions textuelles et caractéristiques visuelles. Cette approche permet de générer des images médicales 3D réalistes à partir de simples prompts textuels.

2. Génération d'images médicales tridimensionnelles :

L'utilisation de modèles de diffusion latente (LDM) et d'auto-encodeurs variationnels permet de créer des IRM 3D réalistes à partir de descriptions textuelles. Cette méthode favorise la création d'images médicales synthétiques sans nécessiter d'acquisitions coûteuses, et sera exploitée pour la modélisation de pathologies.

3. Génération automatique de rapports :

L'étude cherchera à produire des rapports textuels explicatifs sur le comportement du modèle et sur ses résultats. Le modèle devra associer les anomalies visuelles détectées à des descriptions compréhensibles pour les cliniciens, offrant ainsi un support pour l'interprétation médicale.

4. Identification et suivi de biomarqueurs :

L'apport de l'intégration texte-image sera exploité pour l'identification et le suivi de l'évolution de biomarqueurs médicaux. La thèse se concentre sur la détection de nouvelles lésions et la progression de l'atrophie cérébrale dans des maladies telles que la sclérose en plaques (SEP).

5. Modélisation temporelle des pathologies :

La génération de séquences d'images médicales au fil du temps permettra de modéliser la dynamique de certaines pathologies par simple description textuelle. Cette approche vise à améliorer la compréhension des mécanismes sous-jacents aux maladies chroniques.

Détails :

Collaboration(s)/partenariat (s) extérieur(s) éventuels :

Charles Guttman, MD
Director, Center for Neurological Imaging
Associate Professor of Radiology, Harvard Medical School
email: guttmann@bwh.harvard.edu

Juan Carlos Prieto
Research Assistant Professor
University of North Carolina, Chapel Hill.
email: jprieto@med.unc.edu

Domaine et contexte scientifiques :

L'intégration des modalités texte et image dans les systèmes d'intelligence artificielle a révolutionné de nombreux domaines [4]. Les modèles comme CLIP [7] ont montré leur efficacité dans des contextes généraux, tandis que MedCLIP [10] a été spécialement adapté pour les images médicales démontrant son utilité dans ce domaine. Ces avancées ont ouvert de nouvelles perspectives pour l'analyse et la génération d'images à l'imagerie médicale.

Cette thèse s'inscrit dans cette dynamique en étudiant la fusion des représentations latentes texte-image dans un espace commun, afin de développer des modèles capables de générer des images médicales tridimensionnelles à partir de descriptions textuelles. Cette recherche vise à améliorer la compréhension et l'interprétation des pathologies en intégrant de manière cohérente les informations visuelles et sémantiques, et en contribuant à l'explicabilité des modèles.

Les travaux récents dans ce domaine se concentrent sur plusieurs axes.

1. Les modèles de diffusion latente (LDM), comme celui proposé par Rombach et al [8] sont une avancée significative à la génération d'images par prompts textuels. Ces modèles utilisent des descriptions textuelles pour guider le processus de génération, offrant des résultats remarquables sur des jeux de données naturels. Toutefois, leur application aux données médicales reste limitée, notamment en raison de la nécessité de maintenir la cohérence anatomique et de s'adapter à la variabilité des pathologies.

2. La fusion multimodale constitue un autre axe d'intérêt. Xing et al. [11] ont montré que la combinaison de données textuelles et d'images permet de révéler des schémas pathologiques subtils, autrement difficilement détectables par des modèles unimodaux. Cependant, la manière dont ces représentations sont intégrées dans l'espace latent reste une question ouverte. Cette thèse cherchera à approfondir cette fusion pour améliorer la pertinence clinique des modèles génératifs, notamment en intégrant des descriptions textuelles spécifiques aux pathologies étudiées.

3. Les techniques d'apprentissage profond offrent des solutions efficaces pour la segmentation des structures médicales. Le Segment Anything Model (SAM) de Kirillov et al. [3] et sa version spécialisée, MedSAM de Ma et al. [5] démontrent des performances prometteuses pour segmenter des structures complexes à partir d'annotations minimales. Cependant, leur intégration dans des pipelines multimodaux reste complexe, en particulier lorsqu'il s'agit de générer des rapports textuels expliquant les décisions du modèle.

4. L'explicabilité des modèles est un critère indispensable pour leur application médicale. Barreto et al. [1] ont proposé des méthodes de génération de descriptions textuelles pour interpréter les résultats des modèles, mais leur impact réel sur la compréhension des décisions algorithmiques reste limité. Cette thèse vise à développer des mécanismes d'explicabilité qui non seulement fournissent des visualisations pertinentes, mais produisent également des rapports textuels clairs et exploitables par les cliniciens.

5. La structuration d'un espace latent partagé texte-image telle que le proposent les AR-VAE Attribute-based Regularized Variational Autoencoders décrits par Pati et al [6] sera explorée. Ces modèles introduisent une régularisation par attributs permettant de contrôler l'organisation de l'espace latent en fonction de caractéristiques cliniquement pertinentes [2, 9]. Cette approche offre un cadre pour mieux comprendre les interactions texte-image et générer des images médicales cohérentes et exploitables.

Malgré ces avancées, les interactions complexes entre le texte et l'image dans l'espace latent restent insuffisamment explorées. En particulier, la manière dont ces deux modalités interagissent et influencent les prédictions des modèles n'est pas encore totalement maîtrisée. Cette thèse s'attachera à modéliser et contrôler ces contributions respectives, afin d'améliorer la robustesse et la fiabilité des modèles génératifs appliqués à l'imagerie médicale.

Objectifs de la thèse :

Les objectifs de cette thèse sont les suivants :

1. Étudier la fusion des représentations latentes texte-image :

Comprendre comment les informations issues des modalités textuelle et visuelle peuvent être combinées dans un espace latent commun afin de générer des images médicales 3D réalistes et d'enrichir l'analyse d'imagerie.

2. Développer des modèles de génération d'images médicales tridimensionnelles :

Concevoir des modèles basés sur les modèles de diffusion latente (LDM) et d'autres architectures adaptées, capables de produire des images cohérentes et fidèles aux descriptions textuelles fournies.

3. Automatiser l'analyse et la segmentation d'images médicales :

Mettre en œuvre des techniques permettant d'identifier et de segmenter automatiquement les structures d'intérêt dans les images générées, avec la

production de rapports associés.

4. Améliorer l'explicabilité des modèles génératifs :

Élaborer des mécanismes permettant d'expliquer les décisions des modèles en générant des rapports textuels détaillés et des visualisations compréhensibles pour les cliniciens.

5. Simuler l'évolution temporelle des biomarqueurs médicaux :

Développer des algorithmes capables de générer des séquences temporelles d'images médicales afin de modéliser la progression de certaines pathologies, comme la SEP, et d'identifier des schémas pathologiques pertinents.

6. Contribuer à la création de bases de données médicales augmentées :

Générer des images synthétiques pour compléter les jeux de données existants, en tenant compte des contraintes d'anonymisation et de diversité clinique.

Verrous scientifiques :

1. Le premier verrou concerne la construction d'espaces latents contrôlés. Pati et al. [6] ont montré dans leurs travaux qu'il est possible de régulariser l'apprentissage des espaces latents en imposant des attributs spécifiques. Cette approche pourra être utilisée pour mieux contrôler la génération d'images médicales en fonction de caractéristiques cliniques précises, comme la taille des structures ou l'âge des patients.

2. Un deuxième verrou concerne la fusion cohérente des modalités texte et image dans un espace latent commun. Il s'agit de développer des techniques pour éviter les corrélations erronées et garantir une interprétation pertinente des informations textuelles et visuelles.

3. Un troisième verrou réside dans la génération d'images médicales tridimensionnelles conditionnée par le texte. Si les modèles de diffusion latente (LDM) ont démontré leur efficacité sur des images 2D, leur extension au domaine médical 3D reste limitée. La génération d'IRM respectant les structures anatomiques et les variations pathologiques, nécessitera des adaptations spécifiques.

4. La production de rapports médicaux explicatifs constitue un autre défi. Générer des textes pertinents et compréhensibles à partir des observations visuelles implique d'aligner correctement les concepts médicaux avec les résultats des modèles d'analyse. Cette tâche demande une intégration de modèles NLP spécialisés, tels que BiomedBERT ou MedPaLM, entraînés sur des corpus médicaux.

5. La thèse s'intéressera également à la modélisation temporelle de l'évolution des biomarqueurs. La génération de séquences d'images médicales permettant de suivre des évolutions pathologiques, comme celles observées dans la sclérose en plaques, nécessite de modéliser des processus biologiques dynamiques et de décrire textuellement ces trajectoires temporelles.

Figures :



Génération de coupes axiales IRM pondérées T1 et T2 à partir de deux prompts différents.

Contributions originales attendues :

Cette thèse vise à apporter des contributions originales dans le domaine de l'IA générative appliquée à l'imagerie médicale, en s'appuyant sur le développement et l'exploration d'un espace latent partagé texte-image. Cet espace commun doit permettre d'améliorer la génération d'images médicales tridimensionnelles, d'en faciliter l'analyse et d'enrichir l'interprétation clinique grâce à la production de rapports textuels explicatifs.

Les contributions attendues sont les suivantes :

1. Développement et exploration d'un espace latent partagé texte-image

Cette recherche repose sur la création d'un espace latent commun dans lequel les caractéristiques textuelles et visuelles sont intégrées de manière cohérente et

contrôlée. Cette approche doit permettre aux modèles de mieux comprendre les relations entre les descriptions textuelles et les motifs visuels associés, et d'exploiter cette compréhension pour générer des images médicales pertinentes. L'étude cherchera à identifier les mécanismes d'interaction entre ces deux modalités, à analyser leur contribution respective et à proposer des stratégies pour réguler leur influence dans l'espace latent. Cette étape fondamentale conditionnera la réussite des objectifs secondaires de la thèse.

2. Conception de modèles de génération d'images médicales tridimensionnelles guidés par le texte

Une fois l'espace latent partagé construit, la thèse visera à développer des modèles capables de générer des images médicales 3D réalistes à partir de prompts textuels. Ces modèles s'appuieront sur des architectures de diffusion latente (LDM) et d'auto-encodeurs variationnels, adaptées aux spécificités de l'imagerie médicale. L'objectif sera d'obtenir des images cohérentes sur le plan anatomique et suffisamment variées pour simuler différentes pathologies, comme celles observées dans la sclérose en plaques (SEP).

3. Régularisation et contrôle des espaces latents à l'aide d'attributs cliniques

Cette thèse proposera également des méthodes pour introduire des mécanismes de contrôle dans l'espace latent, inspirées des travaux de Pati et al. Il s'agira de contraindre l'apprentissage de l'espace latent en intégrant des attributs cliniquement pertinents, comme l'âge, la taille et la forme des lésions. Cette approche permettra de générer des images synthétiques spécifiques à certaines conditions pathologiques.

4. Production automatique de rapports textuels explicatifs

Un autre axe de la recherche consistera à exploiter l'espace latent partagé pour produire automatiquement des rapports médicaux associés aux images générées. Ces rapports devront décrire de manière compréhensible les anomalies détectées, les corrélations identifiées et les hypothèses du modèle. Contrairement aux approches existantes qui génèrent du texte de manière indépendante, cette contribution s'appuiera sur l'espace latent partagé pour assurer la cohérence entre les observations visuelles et leur description textuelle.

5. Modélisation temporelle et suivi longitudinal des biomarqueurs

L'étude se penchera sur la modélisation temporelle de l'évolution de certaines pathologies en utilisant des séquences d'IRM. En intégrant les informations textuelles décrivant les changements pathologiques dans l'espace latent, il sera possible de générer des séries d'images simulant l'apparition et la progression de biomarqueurs, comme les nouvelles lésions ou l'atrophie cérébrale progressive dans la SEP.

6. Validation et création de bases de données synthétiques pour l'imagerie médicale

Enfin, cette thèse proposera des critères d'évaluation pour valider la pertinence clinique des images et des rapports générés. L'espace latent partagé sera également exploité pour produire des bases de données synthétiques enrichies, destinées à améliorer l'apprentissage des modèles existants et à compenser le manque de données annotées.

Programme de recherche et démarche scientifique proposée :

L'objectif de cette thèse est de développer un espace latent partagé facilitant la génération d'images 3D réalistes à partir de descriptions textuelles et l'explicabilité des décisions des modèles.

Voici le programme de recherche pour les trois années

Année 1 : Construction de l'espace latent texte-image

Exploration des modèles existants (CLIP, MedCLIP, LDM) et adaptation à l'imagerie médicale.

Développement d'un premier espace latent partagé en utilisant les données disponibles : une base d'IRM cérébrales de sujets sains, avec des images 2D et 3D annotées par descriptions textuelles, par coupe et par volume sur les séquences T1, T2 et T2-FLAIR.

Premières expériences de génération d'images et analyse des interactions texte-image.

Année 2 : Optimisation et intégration de fonctionnalités avancées

Optimisation des modèles de diffusion latente pour générer des images médicales 3D cohérentes.

Implémentation de mécanismes de régularisation inspirés les AR-VAE pour contrôler l'espace latent en fonction de caractéristiques pathologiques.

Développement d'un module de segmentation et de production automatique de rapports textuels explicatifs.

Mobilité à l'Université de Chapel Hill dans le cadre du co-encadrement avec Juan Carlos Prieto.

Année 3 : Modélisation temporelle et validation clinique

Modélisation de l'évolution temporelle de biomarqueurs médicaux, en simulant des séquences IRM illustrant la progression de pathologies comme la SEP.

Validation clinique des images et des rapports textuels avec les experts médicaux.

Analyse de la généralisation du modèle à d'autres types d'imagerie et rédaction du manuscrit de thèse.

Les résultats attendus incluent un modèle performant de génération d'images médicales 3D, la production automatique de rapports explicatifs et une meilleure compréhension des relations texte-image dans un espace latent partagé.

Encadrement scientifique :

Nom :	Muller	Prénom :	Chantal
Rôle :	Directeur de Thèse	Labo :	CREATIS
Equipe / Dpt :	MYRIAD		
Compétences Scientifiques :	Apprentissage profond, IA générative, Traitement d'images, Segmentation		
Taux d'encadrement :	70 %		
Autres thèses co-dirigées et co-encadrées (% , année inscription) :	Aucun encadrement de thèse en cours.		

Nom :	PRIETO	Prénom :	Juan Carlos
Rôle :	Encadrant		
Labo :			
Equipe / Dpt / Nom labo (si autre) :	University of North Carolina, Chapel Hill.		
Compétences Scientifiques :	Data scientist, Artificial intelligence and medical imaging, 3D modeling		
Taux d'encadrement :	30%		
Autres thèses co-dirigées et co-encadrées (% , année inscription) :	Aucun encadrement de thèse en cours.		

Justification éventuelle :

Intégration au sein du (ou des) laboratoire(s) (Département/Equipe(s) impliquée(s)) (pourcentage du temps travail au sein de ce ou ces laboratoire(s)) :	100% CREATIS Myriad Mobilité de 3 mois pendant la 2e année à l'Université de Chapel Hill.
--	--

Informations complémentaires :

Financement de la thèse : **Bourse de l'école doctorale EEA pendant 3ans, 120 000€ salaire chargé.**

Profil du candidat recherché (prérequis) : **Le candidat recherché doit avoir un diplôme de niveau M2 et posséder une solide formation en intelligence artificielle, traitement d'images et apprentissage profond. Des compétences en modélisation de réseaux neuronaux (CNN, VAE, LDM) et en vision par ordinateur sont indispensables. Une maîtrise des outils de développement en Python et des frameworks comme PyTorch ou TensorFlow est requise. La**

connaissance de la librairie MONAI serait appréciée. Une expérience en imagerie médicale ou en fusion multimodale texte-image sera déterminante. Le candidat devra faire preuve de rigueur scientifique, d'autonomie et d'un esprit d'initiative, tout en ayant la capacité de collaborer au sein d'une équipe interdisciplinaire.

Objectifs de valorisation
des travaux de recherche :

Cette thèse a pour objectif de développer un espace latent partagé texte-image pour générer et analyser des images médicales tridimensionnelles. L'un des principaux cas d'application concernera la modélisation de la sclérose en plaques (SEP), avec l'ambition d'améliorer la compréhension et le suivi de cette pathologie.

Le projet bénéficiera de la collaboration avec le Dr Charles Guttman (Harvard Medical School, Center for Neurological Imaging), dont l'expertise en imagerie cérébrale et en suivi longitudinal de la SEP apportera un cadre clinique solide pour la validation des modèles développés. Juan Carlos Prieto, spécialiste en IA appliquée à l'imagerie médicale, co-encadrera ces travaux et contribuera à l'optimisation des modèles, notamment pour ce qui concerne leur explicabilité et leur fiabilité.

Les résultats de cette recherche seront valorisés par :

- La publication d'articles dans des revues et conférences internationales reconnues, comme MICCAI, MIDL et IEEE Transactions on Medical Imaging.
- Le développement de modules logiciels intégrés dans des outils open-source, comme 3D Slicer, afin de rendre ces avancées accessibles à la communauté médicale.
- Des échanges et des validations cliniques en partenariat avec le Center for Neurological Imaging, pour adapter et tester les modèles sur des données réelles et évaluer leur pertinence dans le suivi de la SEP et d'autres pathologies cérébrales.

Cette recherche devrait permettre d'améliorer l'interprétation des biomarqueurs cérébraux et d'enrichir les pratiques d'imagerie médicale en s'appuyant sur des outils d'IA générative à la fois explicables, fiables et adaptés aux besoins cliniques.

Compétences qui seront
développées au cours du
doctorat :

Au cours de cette thèse, le doctorant développera des compétences avancées dans le domaine de l'IA générative appliquée à l'imagerie médicale, en se concentrant sur la fusion des représentations latentes texte-image.

Sur le plan méthodologique, il maîtrisera la conception et l'entraînement de modèles d'apprentissage profond, notamment les modèles de diffusion latente (LDM) et les auto-encodeurs variationnels VAE et les AR-VAE pour la régularisation et le contrôle des attributs dans l'espace latent.

Le doctorant développera une expertise en fusion multimodale texte-image, en travaillant sur des architectures comme CLIP et MedCLIP. Il étudiera les mécanismes d'attention pour analyser les interactions entre les modalités et comprendra comment ces modèles exploitent les représentations sémantiques et visuelles.

Il sera amené à manipuler des données IRM cérébrales 2D et 3D de sujets sains et pathologiques, avec des descriptions textuelles associées, sur des séquences pondérées T1, T2 et T2-FLAIR. Il mettra en œuvre des méthodes de prétraitement, d'annotation et d'augmentation de ces données, appliquera des fonctions de perte et des métriques d'évaluation adaptées.

Il concevra et testera des algorithmes basés sur des modèles NLP spécialisés (BiomedBERT, MedPaLM) pour produire des rapports expliquant les prédictions et les corrélations texte-image.

Il développera des compétences en modélisation et développera des modèles capables de simuler l'évolution de biomarqueurs, notamment dans le cadre de la sclérose en plaques.

Le travail inclura également une collaboration étroite avec des praticiens spécialisés en neuroimagerie pour valider l'efficacité clinique des modèles développés. Cette interaction sera nécessaire pour ajuster les algorithmes et garantir la pertinence des images ou des rapports textuels générés.

Perspectives
professionnelles après le
doctorat :

La thèse bénéficiera d'un partenariat international avec le Center for Neurological Imaging dirigé par le Dr Charles Guttman (Harvard Medical School), ainsi que de l'expertise en IA appliquée à l'imagerie médicale de Juan Carlos Prieto. Ce partenariat permettra de tester et de valider les modèles sur des bases de données réelles et d'assurer une diffusion internationale des résultats obtenus.

Ces travaux seront menés en utilisant des outils et frameworks spécialisés tels que PyTorch, TensorFlow, MONAI et 3D Slicer, et donneront lieu à des publications dans des conférences et revues spécialisées (MICCAI, MIDL, IEEE TMI).

Les compétences acquises au cours de cette thèse offriront au doctorant diverses opportunités professionnelles.

Dans le milieu académique, il pourra poursuivre sa carrière en tant que chercheur ou enseignant-chercheur dans des laboratoires spécialisés en imagerie médicale et intelligence artificielle. Les connaissances développées sur la fusion texte-image et la génération d'images médicales 3D permettront d'intégrer des équipes de recherche en France ou à l'international, notamment grâce au partenariat établi avec le Center for Neurological Imaging de la Harvard Medical School.

Le secteur privé offre également de nombreuses possibilités, en particulier dans les entreprises qui conçoivent des outils d'IA pour la santé. Le doctorant pourra contribuer à des projets visant à améliorer les outils de diagnostic et de suivi médical grâce à l'IA générative, que ce soit pour optimiser l'analyse d'IRM, modéliser l'évolution de biomarqueurs ou générer des rapports automatisés pour les cliniciens.

Les compétences en vision par ordinateur et en modélisation d'espaces latents ouvrent aussi la porte à des postes en recherche et développement dans d'autres secteurs, comme la robotique, les véhicules autonomes ou les systèmes industriels.

Enfin, le doctorant pourra envisager de créer sa propre entreprise, en s'appuyant sur les résultats de la thèse pour développer des solutions logicielles destinées à la recherche médicale et au suivi clinique.

Références
bibliographiques sur le
sujet de thèse :

- [1] Artur Gomes Barreto, Juliana Martins de Oliveira, Francisco Nauber Bernardo Gois, Paulo Cesar Cortez, and Victor Hugo Costa de Albuquerque. A new generative model for textual descriptions of medical images using transformers enhanced with convolutional neural networks. *Bioengineering*, 10(9):1098, 2023.
- [2] I Kalinin, G Makshakov, and E Evdoshenko. The impact of intracortical lesions on volumes of subcortical structures in multiple sclerosis. *American Journal of Neuroradiology*, 41(5): 804–808, 2020.
- [3] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015– 4026, 2023.
- [4] Chunyuan Li, Zhe Gan, Zhengyuan Yang, Jianwei Yang, Linjie Li, Lijuan Wang, Jianfeng Gao, et al. Multimodal foundation models: From specialists to general-purpose assistants. *Foundations and Trends® in Computer Graphics and Vision*, 16(1-2):1–214, 2024.
- [5] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.
- [6] Ashis Pati and Alexander Lerch. Attribute-based regularization of latent spaces for variational auto-encoders. *Neural Computing and Applications*, 33:4429–4444, 2021.
- [7] 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.
- [8] 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.
- [9] Constantina A Treaba, Tobias E Granberg, Maria Pia Sormani, Elena Herranz,

Russell A Ouellette, Céline Louapre, Jacob A Sloane, Revere P Kinkel, and Caterina Mainero. Longitudinal characterization of cortical lesion development and evolution in multiple sclerosis with 7.0-t mri. *Radiology*, 291(3):740–749, 2019.

[10] Zifeng Wang, Zhenbang Wu, Dinesh Agarwal, and Jimeng Sun. Medclip: Contrastive learning from unpaired medical images and text. arXiv preprint arXiv:2210.10163, 2022.

[11] Xiaodan Xing, Junzhi Ning, Yang Nan, and Guang Yang. Deep generative models unveil patterns in medical images through vision-language conditioning. arXiv preprint arXiv:2410.13823, 2024.

Fiche créée le 12/02/2025 17:29:53

Dernière mise à jour le 14/02/2025 19:50:55