CentraleSupélec

université PARIS-SACLAY

Rapport de Pôle Projet

Pôle Projet S7

Radiothérapie
Mise en place d'un modèle Transformers pour la dosimétrie

Réalisé par :
Marie Feller
Youssef Sellami
Adrian Demarcy
Mahdi Choura
Sami Rabinovitch

3 février 2025

Table des matières

1 Introduction

La radiothérapie, méthode essentielle dans le traitement de nombreux cancers, repose sur le principe de détruire l'ADN des cellules tumorales par irradiation, tout en minimisant les dommages aux tissus sains. Depuis ses débuts, cette discipline a connu d'importantes avancées technologiques. Historiquement, les cliniciens utilisaient des radiographies 2D pour localiser et estimer la position des tumeurs, une approche souvent approximative. Avec l'émergence de la dosimétrie précise et des simulations numériques basées sur les méthodes de Monte Carlo (MC), la radiothérapie est devenue une science de plus en plus précise.

Cependant, ces méthodes restent exigeantes en termes de temps de calcul et de ressources informatiques. Les modèles de deep learning, qui ont révolutionné des domaines tels que l'analyse d'images médicales, offrent des perspectives prometteuses pour accélérer les calculs et améliorer la qualité des traitements. Cet article vise à explorer les derniers développements dans ce domaine et à détailler une méthode intégrant des réseaux de neurones profonds pour la dosimétrie et les simulations.

2 Contexte

2.1 La radiothérapie

La radiothérapie joue un rôle central dans la prise en charge des patients atteints de cancer. Son principal objectif est de cibler avec une grande précision les tumeurs malignes afin de limiter les dommages aux tissus environnants. Les progrès technologiques, notamment l'imagerie en temps réel et la robotique, ont permis d'améliorer l'efficacité des traitements tout en réduisant les effets secondaires. Ces progrès reposent sur des techniques de dosimétrie précise et des algorithmes d'optimisation.

2.2 Dosimétrie

La dosimétrie consiste à mesurer, calculer et optimiser la dose de radiation absorbée par les tissus. Elle est essentielle pour garantir un traitement efficace et sécurisé. La précision de cette étape permet de minimiser les risques pour les organes à risque (OAR) tout en maximisant l'effet sur la tumeur. Les approches classiques reposent sur des calculs mathématiques et des modèles physiques, mais les simulations numériques comme celles basées sur Monte Carlo offrent une précision inégalée.

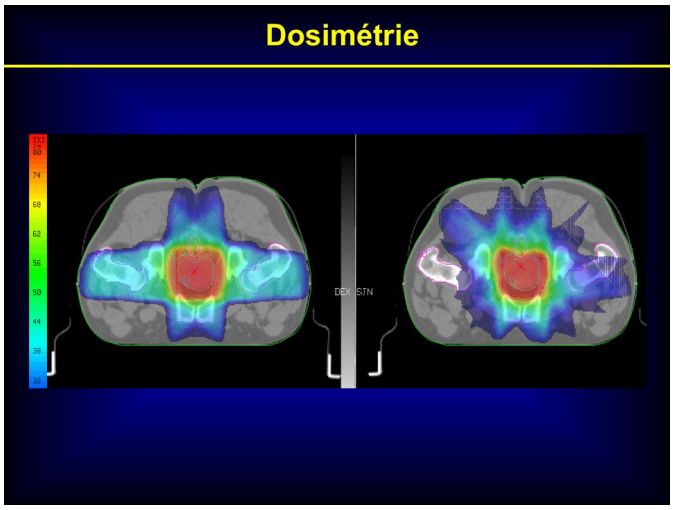

FIGURE 1 – Illustration de la dosimétrie dans un contexte clinique.

2.3 Simulation Monte Carlo

Les méthodes de Monte Carlo (MC) sont devenues une référence pour simuler les interactions des particules ionisantes avec les tissus biologiques. Elles permettent de créer des cartes de doses avec une grande précision, mais au prix de temps de calcul très élevés. Ces simulations se décomposent en trois étapes majeures :

1. **Modélisation du patient :** Une représentation 3D du patient est générée à partir d'images CT ou IRM. Les tissus sont divisés en voxels, chacun ayant des propriétés physiques spécifiques.
2. **Simulation des interactions :** Des millions, voire des milliards de particules, sont suivies dans leur interaction avec les voxels. Cela inclut les effets de diffusion, d'absorption et de transmission.
3. **Calcul de la dose :** Une carte 3D précise est produite, montrant la dose absorbée dans chaque voxel.

Malgré leur précision, ces simulations souffrent de limitations pratiques en raison de leur complexité computationnelle. Les méthodes hybrides et les approches basées sur le deep learning offrent des solutions pour accélérer ces processus.

3 Objectifs du Projet

Ce projet a pour but d'exploiter les modèles d'apprentissage profond pour générer des cartes de dose de haute résolution à partir de données sous-échantillonnées. Cela permettrait de généraliser les méthodes à divers contextes cliniques tout en réduisant les coûts de calcul. L'approche vise à combiner la précision des simulations MC avec la rapidité et l'efficacité des modèles de deep learning.

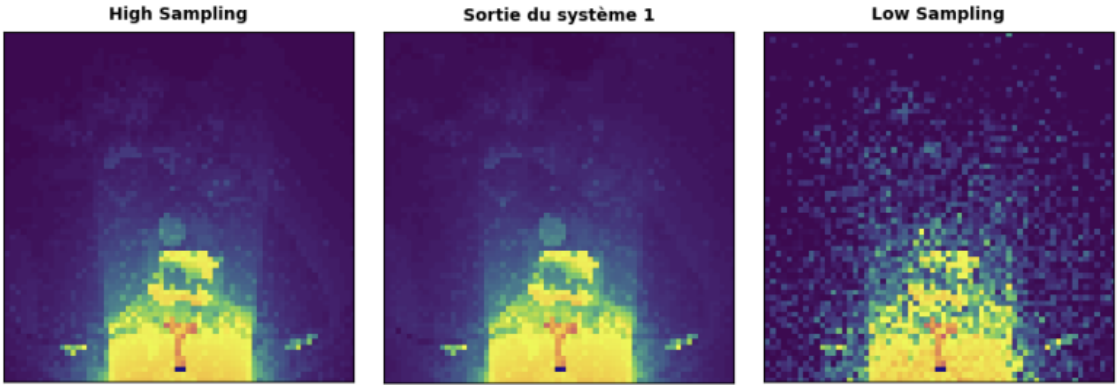

FIGURE 2 – Schéma représentant les objectifs principaux du projet.

4 État de l'art

4.1 Accélération Matérielle

Les cartes graphiques modernes (GPU) ont révolutionné la capacité à effectuer des simulations complexes, réduisant les temps de calcul de plusieurs jours à quelques heures.

4.2 Algorithmes Hybrides et Réduction de la Variance

Les techniques de réduction de variance (VRT) diminuent le bruit statistique des simulations MC sans compromettre leur précision. Elles sont combinées avec des approches hybrides pour optimiser les calculs.

4.3 Deep Learning

Les modèles d'apprentissage profond, comme U-Net, sont de plus en plus utilisés pour générer des cartes de dose à partir de données simplifiées. Cependant, leur dépendance vis-à-vis de données simulées pose des questions sur leur généralisation en clinique.

5 Méthode Proposée et U-Net

5.1 U-Net

Le réseau U-Net, initialement conçu pour la segmentation biomédicale, est idéal pour des tâches de réparation ou de complétion d'images. Son architecture, basée sur des couches convolutionnelles et des connexions de saut, préserve les détails importants tout en reconstruisant des images à haute résolution.

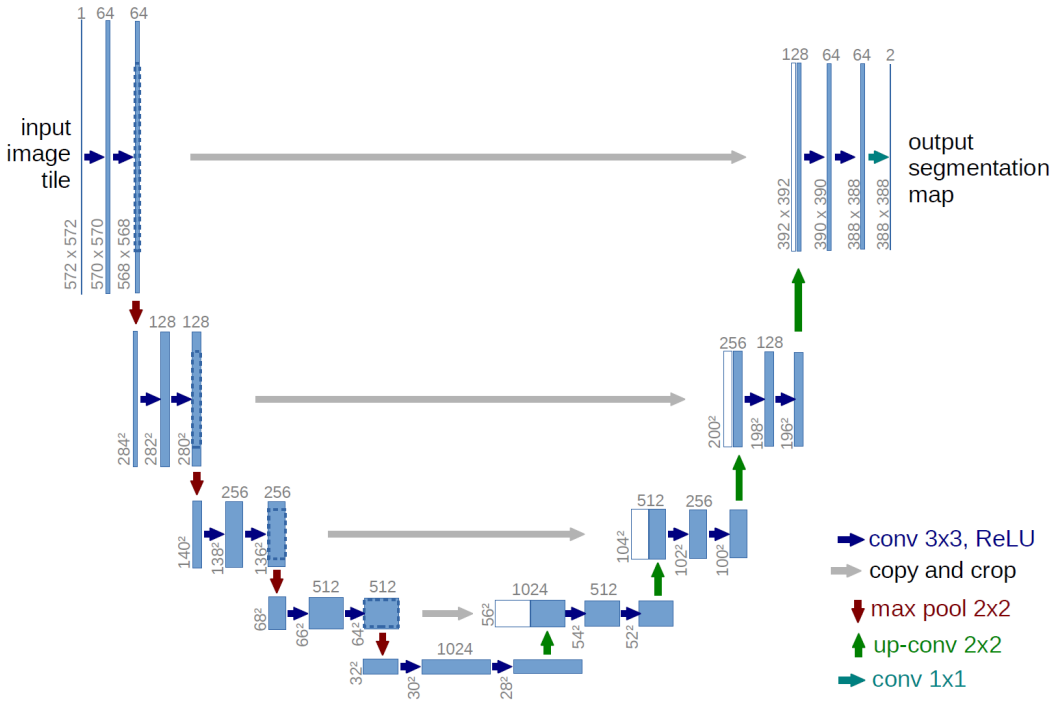

FIGURE 3 – Architecture du réseau U-Net.

5.2 Transformers

Les transformers, initialement développés pour le traitement du langage naturel, trouvent des applications innovantes en traitement d'images. Leur capacité à modéliser des relations à longue distance permet de capter des détails globaux dans les images médicales.

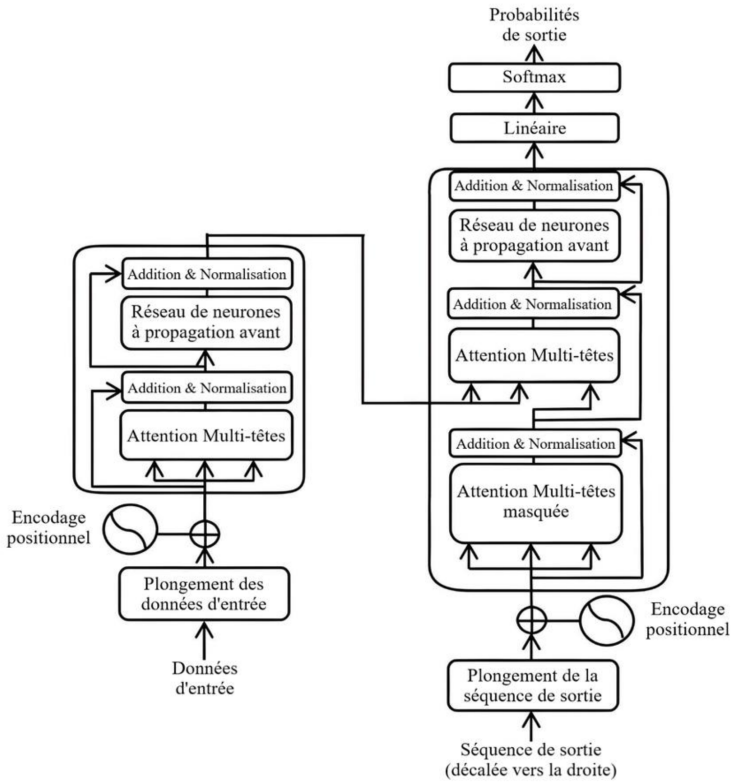

FIGURE 4 – Schéma d'un transformer appliqué aux images.

Dans notre approche, nous explorons l'utilisation d'une variante des Transformers adaptée aux images médicales, comme le Vision Transformer (ViT) ou les Swim Transformers (Swin-T). Ces modèles convertissent une image en une séquence de patchs, traités comme des tokens indépendants, puis appliquent des couches d'auto-attention pour extraire des informations pertinentes.

Le pipeline général de notre modèle basé sur les Transformers suit les étapes suivantes :

— 1. Découpage en patchs : L'image d'entrée (par exemple, une image CT ou une carte de dose de basse résolution) est divisée en petites régions carrées (patchs), qui sont ensuite vectorisées.

— 2. Encodage des patchs : Chaque patch est projeté dans un espace de dimension supérieure à l'aide d'une transformation linéaire et enrichi avec des informations de position (encodage positionnel).

— 3. Mécanisme d'auto-attention : À l'aide de la Self-Attention Multi-Tête (MSA), le modèle pondère l'importance de chaque patch par rapport aux autres, permettant ainsi une compréhension globale de la distribution des doses.

— 4. Génération de la carte de dose finale : Après plusieurs couches d'auto-attention et de normalisation, une carte de dose de haute résolution est reconstruite

6 Résultats avec le modèle actuel UNet

Dans un premier temps, nous utilisons le modèle UNet pour plusieurs applications :

— Les entrées sont le Low Sampling et le CT scan, et on cherche à avoir comme sortie le High Sampling reconstruit.

— L'entrée est le CT scan, et on cherche à avoir comme sortie le Mask CT reconstruit.

— Les entrées sont le Low Sampling, et on cherche à avoir comme sortie le High Sampling reconstruit.

Cas 1 : Entrée LS + CT, Sortie HS Reconstructed

- Paramètres : 200 epochs, batch size = 10, learning rate = 10^{-4}, MSE. - Résultats :

— PSNR : 44,02 dB.

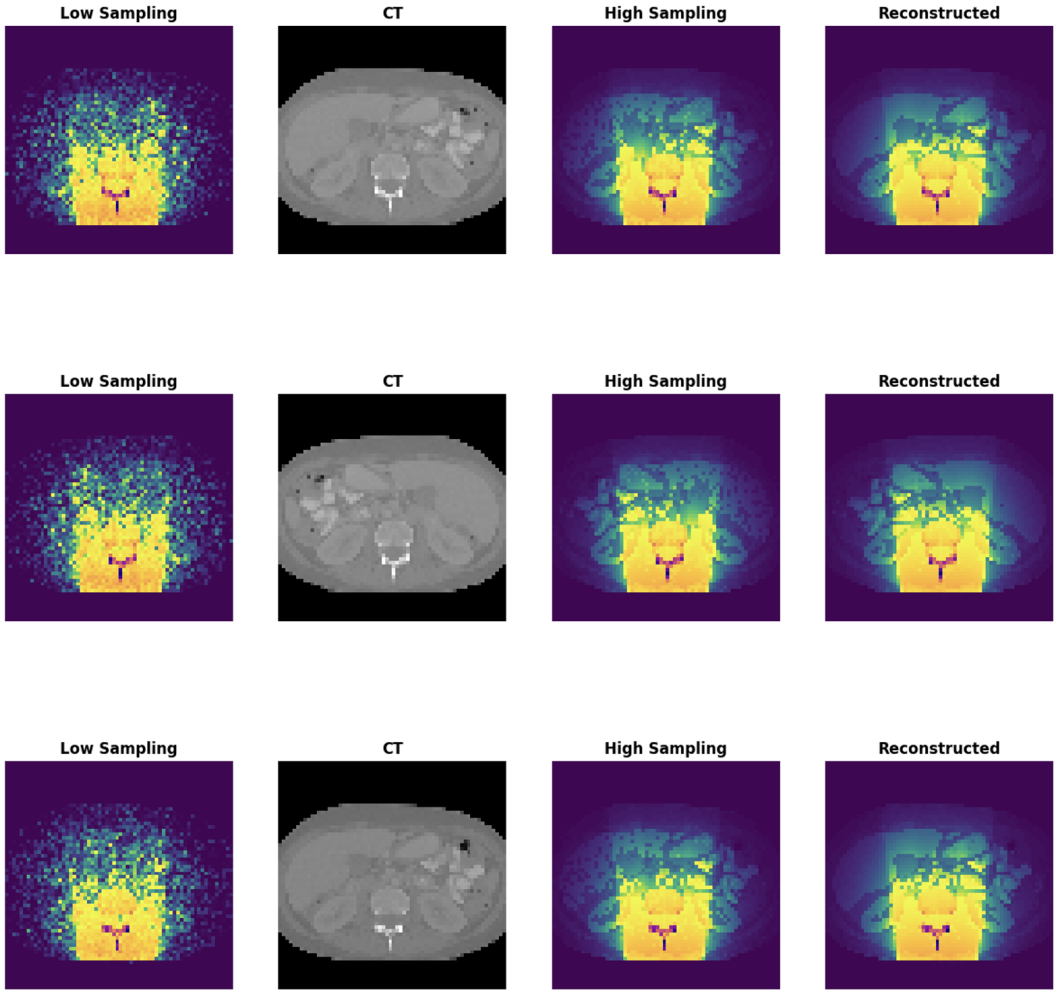

FIGURE 5 – Résultats pour l'entrée LS + CT et sortie HS reconstructed.

Analyse des résultats

Le PSNR de 44,02 dB indique une reconstruction de haute qualité pour ce cas. Cela montre que l'utilisation combinée des entrées LS et CT fournit des informations complémentaires qui améliorent la précision de la reconstruction HS.
La figure 8 montre une reconstruction visuellement satisfaisante, avec des détails bien préservés dans les zones critiques. Des artefacts peuvent être présents dans les régions à faible contraste, mais globalement, la qualité perçue reste élevée.

Cas 2 : Entrée CT scan, Sortie Mask CT Reconstructed

- Paramètres : 200 epochs, batch size = 10, learning rate = 10^{-4}, MSE. - Résultats :
— PSNR : 52,57 dB.

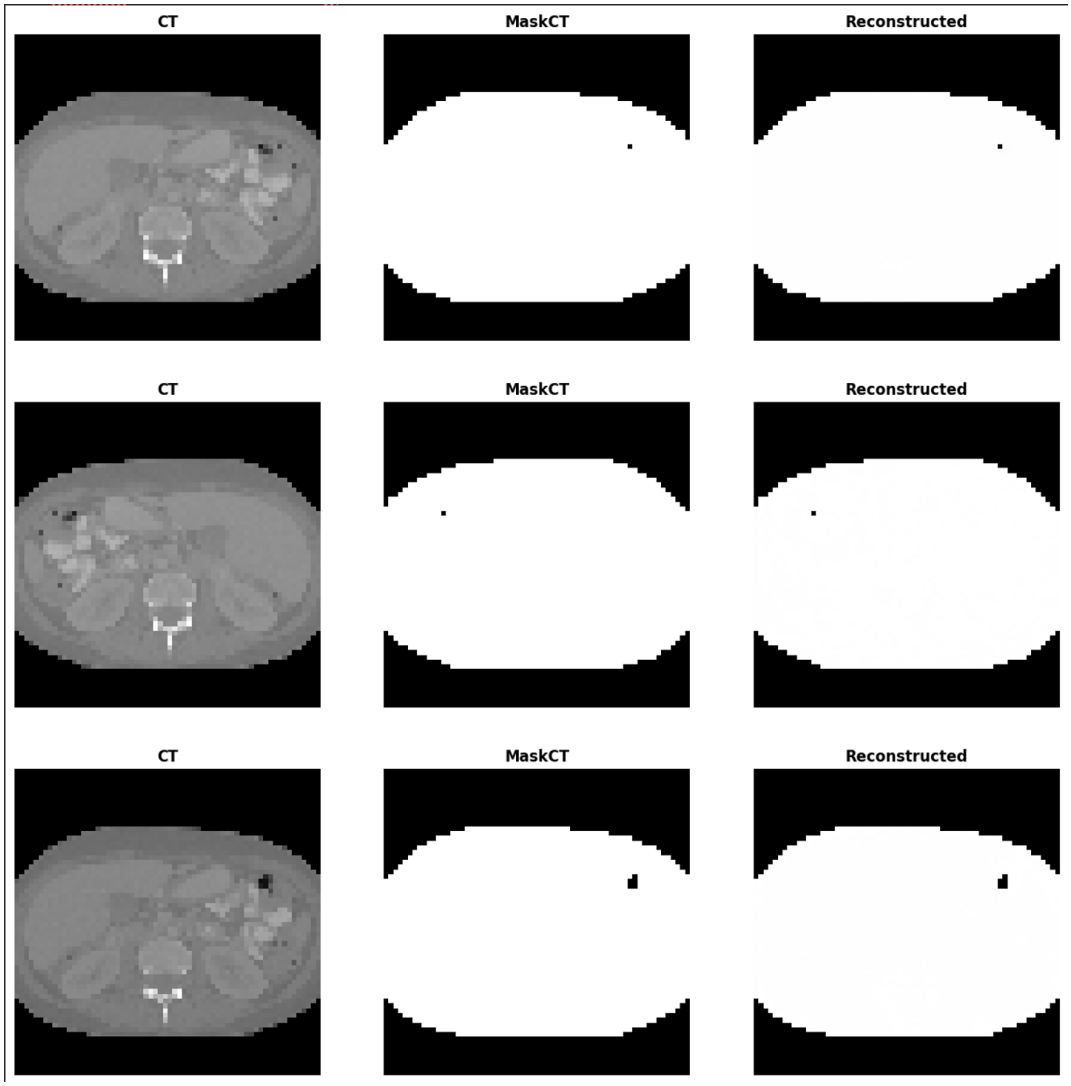

FIGURE 6 – Résultats pour l'entrée CT scan et sortie Mask CT reconstructed.

Analyse des résultats

Avec un PSNR de 52,57 dB, ce cas présente les meilleures performances parmi les scénarios analysés. La reconstruction du masque CT semble particulièrement efficace, probablement en raison des caractéristiques détaillées des scans CT qui facilitent une reconstruction précise. Ces résultats suggèrent que le modèle est bien adapté pour traiter des données CT et générer des reconstructions fiables.
Comme illustré dans la figure 6, les masques reconstruits présentent une précision remarquable dans les contours et les zones homogènes.

Cas 3 : Entrée LS, Sortie HS Reconstructed

- Paramètres : 200 epochs, batch size = 10, learning rate = 10^{-4}, MSE. - Résultats :
— PSNR : 37,65 dB.

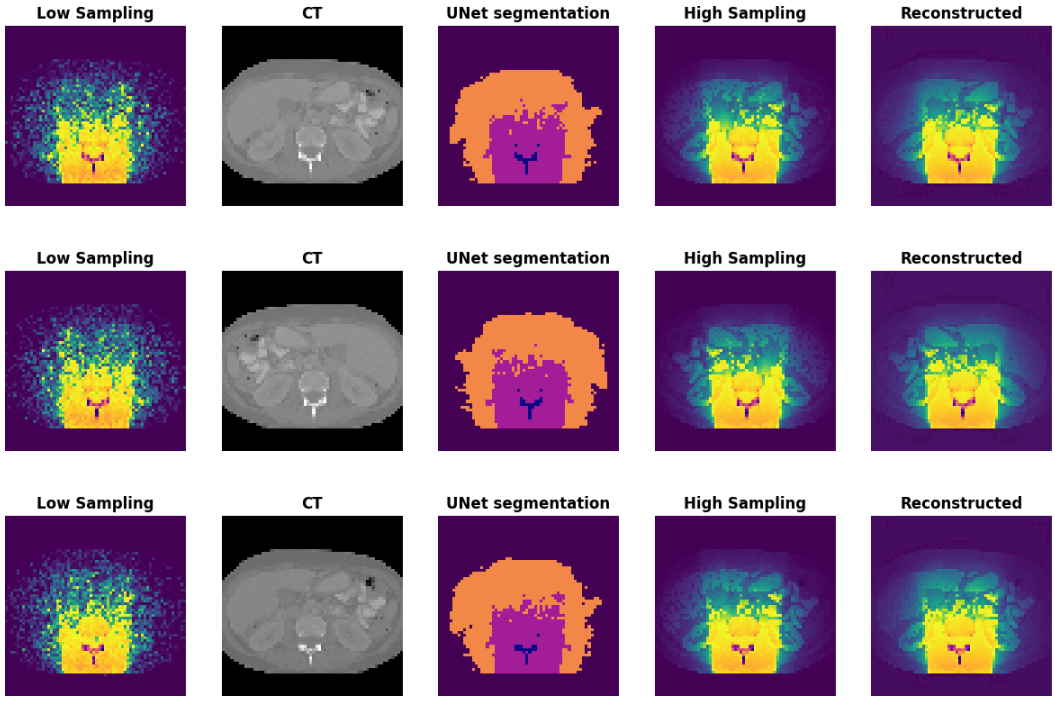

FIGURE 7 – Résultats pour l'entrée LS et sortie HS reconstructed.

Analyse des résultats

Le PSNR de 37,65 dB est le plus faible parmi les cas étudiés, ce qui indique une reconstruction moins précise. Cela pourrait être dû à l'absence d'informations provenant des scans CT, ce qui limite la capacité du modèle à capturer les détails nécessaires pour une reconstruction de haute qualité.
La figure 7 révèle des limitations visibles dans la reconstruction. Les zones à haute fréquence semblent manquer de détails, et des artefacts significatifs peuvent être observés. Le cas utilisant uniquement les scans CT comme entrée offre les meilleures performances, tandis que les scénarios basés uniquement sur LS présentent des limites notables.

7 Résultats avec le modèle UNet + Transformers

Dans ce modèle, nous avons ajouté un transformers au goulot d'étranglement du modèle de base UNet entre l'encoder et le decoder.

Le modèle `UNetWithTransformer` intègre un bloc Transformer dans la couche de bottleneck afin d'exploiter les capacités d'attention globale du Transformer en complément de l'encodage local réalisé par les convolutions. Ce bloc est implémenté à l'aide d'un `TransformerEncoder`, composé de plusieurs couches de `TransformerEncoderLayer`, où chaque couche applique une attention multi-tête suivie de mécanismes de normalisation et de feed-forward. L'entrée du bloc Transformer est obtenue après plusieurs étapes de pooling sur l'image d'entrée, réduisant ainsi sa résolution et permettant d'encoder des représentations de plus haut niveau.

Concrètement, le tenseur d'entrée est d'abord aplati spatialement puis transposé pour correspondre à la représentation attendue par le Transformer (`[H*W, batch, channels]`). Le Transformer traite alors cette représentation en capturant des relations non locales entre les pixels via des mécanismes d'attention. Après traitement, la sortie est réorganisée pour restaurer la structure spatiale originale du tenseur avant d'être transmise au décodeur du U-Net.

L'ajout du Transformer dans le goulot d'étranglement (bottleneck) améliore la capacité du modèle à capturer des relations à longue portée dans l'image, ce qui est particulièrement utile dans des tâches nécessitant une compréhension globale des structures, comme la segmentation sémantique.

Analyse des résultats

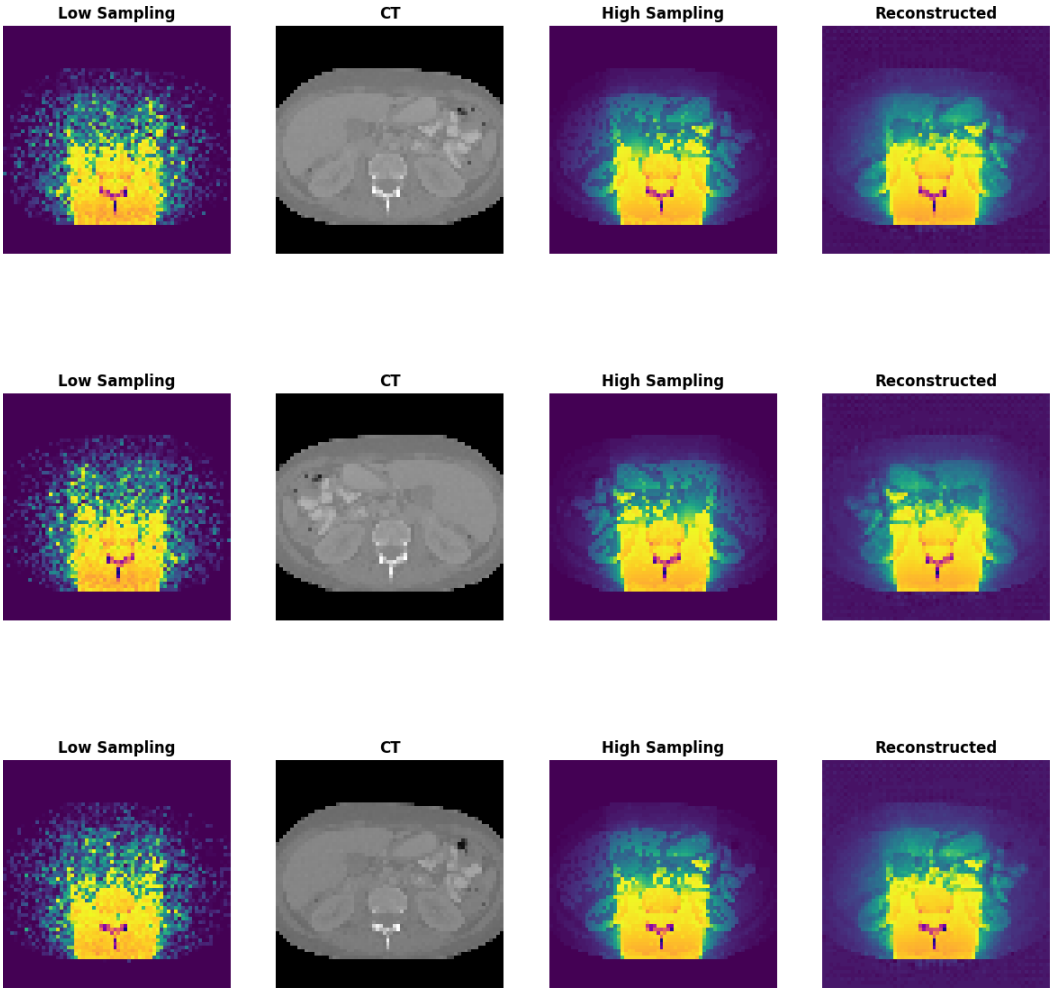

FIGURE 8 – Résultats pour l'entrée LS et sortie HS reconstructed.

Le PSNR de 44,02 dB indique une reconstruction de haute qualité pour ce cas. Cela montre que l'utilisation combinée des entrées LS et CT fournit des informations complémentaires qui améliorent la précision de la recons-

truction HS.
La figure 8 montre une reconstruction visuellement satisfaisante, avec des détails préservés dans au niveau des zones critiques. Des artefacts peuvent être présents dans les régions à faible contraste, mais globalement, la qualité perçue reste élevée. Cependant des points précis semblent différentier le High Sampling et l'image reconstruite.

8 Segmentation en zones

Dans cette partie, nous abordons une nouvelle approche qui consiste à segmenter chaque image en quatre zones selon l'absorption du corps et de procéder à l'apprentissage par zone, avant de reconstruire l'image.

Ainsi, dans un premier temps, nous avons séparer chaque image en quatre zones par la méthode GMM (Gaussian Mixture Model) à partir du Low Sampling et du CT Scan.

Le **modèle de mélange gaussien** (GMM, pour *Gaussian Mixture Model*) est une méthode statistique utilisée pour modéliser une distribution de données comme une combinaison de plusieurs distributions normales (gaussiennes). Dans le contexte de la segmentation d'images, le GMM permet de classifier les pixels en différents segments basés sur leurs intensités ou d'autres caractéristiques.

Le processus de segmentation par GMM peut être décrit en plusieurs étapes :

1. **Initialisation** : On spécifie le nombre de composantes gaussiennes K correspondant aux segments souhaités dans l'image. Les paramètres initiaux de chaque gaussienne, tels que la moyenne (μ_k), la variance (σ_k^2) et le poids (π_k), sont définis.
2. **Étape d'espérance (E-step)** : Pour chaque pixel, on calcule la probabilité qu'il appartienne à chaque composante gaussienne, en se basant sur les paramètres actuels des gaussiennes.
3. **Étape de maximisation (M-step)** : Les paramètres des gaussiennes sont mis à jour en maximisant la vraisemblance des données, en utilisant les probabilités calculées lors de l'étape précédente.
4. **Convergence** : Les étapes E et M sont répétées jusqu'à ce que les changements dans les paramètres soient infimes, indiquant que le modèle a convergé.

Une fois le modèle entraîné, chaque pixel de l'image peut être assigné au segment correspondant à la composante gaussienne pour laquelle il a la plus haute probabilité d'appartenance. Cette approche permet une segmentation efficace des images en identifiant des régions homogènes en termes d'intensité ou d'autres caractéristiques.

8.1 Analyse des résultats

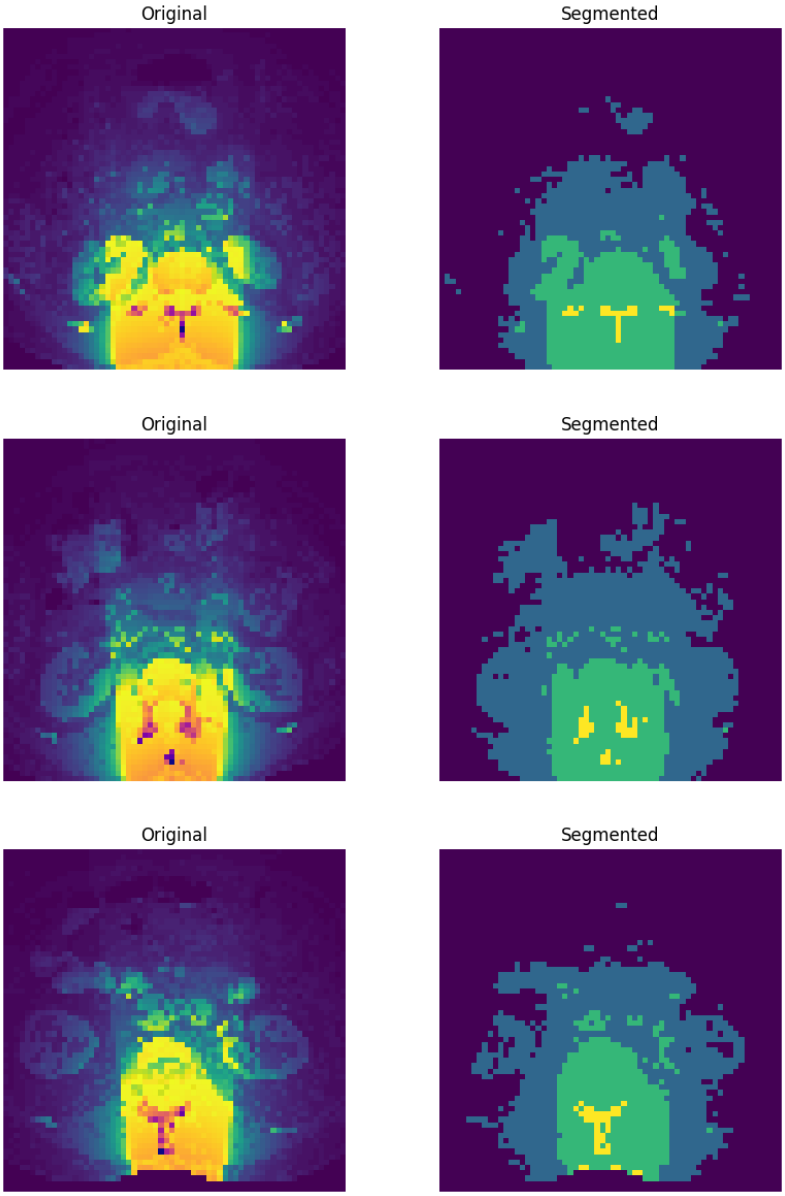

FIGURE 9 – Résultats pour l'entrée CT+LS et sortie des 4 zones

On retrouve des estimations très proches de celles obtenues par la méthode de high sampling. Toutefois, on pourrait penser que les prédictions du modèle semblent offrir une représentation encore plus réaliste de la dissipation de la radioactivité dans le corps. Cependant ce n

Cette amélioration peut être attribuée à l'utilisation du scan CT en complément du low sampling comme donnée d'entrée. En effet, le scan CT apporte des informations anatomiques précises qui permettent au modèle d'intégrer des variations structurelles spécifiques à chaque organe. Cette approche semble favoriser une meilleure fluidité dans les prédictions des niveaux de radioactivité au sein d'un même organe, alignant ainsi les résultats avec les dynamiques physiologiques réelles.

8.2 Ajout de ces zones dans notre modèle UNet

Par la suite, nous avons de nouveau entrainé notre modèle UNet mais avec comme entrée le Low Sampling, le CT Scan, ainsi que l'image segmentée, directement dans UNet, et non plus avec le GMM. On obtient en sortie le

High Sampling reconstruit.

Nous obtenons les résultats suivant :

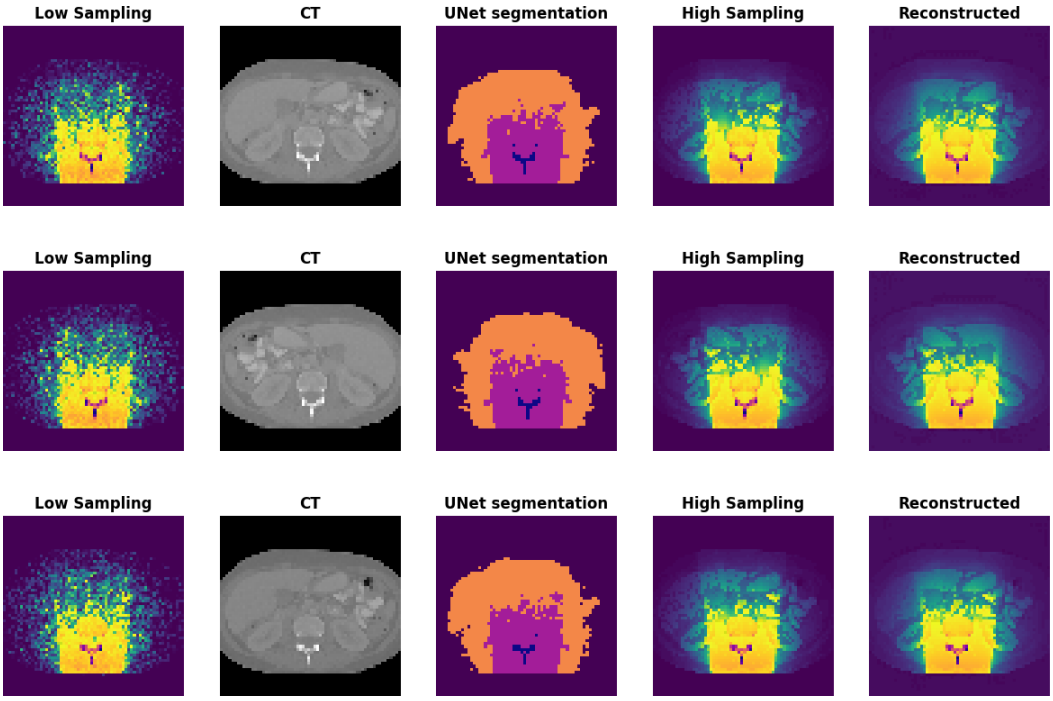

FIGURE 10 – Résultats pour l'entrée CT+LS et sortie des 4 zones

Nous obtenons alors un PSNR de 43,99, qui est donc plus élevé par rapport au cas où on avait pour entrées le Low Sampling et le Scan CT.

L'amélioration se reflète principalement dans l'augmentation du PSNR, qui atteint ici **43,99**, une valeur supérieure à celle obtenue avec les modèles utilisant uniquement le Low Sampling et le CT Scan en entrée. Cette hausse indique une réduction significative de l'erreur de reconstruction, démontrant que le modèle est capable de générer une image de haute qualité avec moins de bruit et une meilleure correspondance aux données de référence.

Visuellement, on observe une représentation plus fidèle des structures internes, avec une meilleure préservation des contours et des régions de forte intensité. En particulier, les points précis, représentant des zones de forte absorption ou de concentration spécifique, sont mieux représentés, ce qui témoigne d'une meilleure capture des détails fins par le modèle. Cette amélioration est probablement due à l'exploitation des informations de segmentation, qui contraignent le modèle à respecter des structures anatomiques et à mieux différencier les régions d'intérêt.

9 Conclusion

Ce projet a démontré l'intérêt des modèles de deep learning, et en particulier des Transformers, pour améliorer la dosimétrie en radiothérapie. L'intégration des Transformers dans l'architecture U-Net a permis d'exploiter le mécanisme d'attention afin de renforcer les relations entre les différentes régions des images médicales. Cette approche a été complétée par une adaptation de la segmentation, introduisant une entrée supplémentaire qui améliore la différenciation des zones anatomiques et optimise la reconstruction des cartes de dose.

Les performances élevées obtenues valident l'efficacité des méthodes mises en œuvre. La comparaison entre U-Net seul et l'approche hybride U-Net + Transformers met en évidence une meilleure précision dans la prédiction des doses, particulièrement dans les zones complexes où une modélisation fine des interactions est nécessaire.

Plusieurs axes d'amélioration et d'exploration peuvent être envisagés. Tout d'abord, il serait pertinent de tester différentes configurations de Transformers au-delà du simple bottleneck du U-Net et de comparer leurs performances selon diverses stratégies d'intégration. Ensuite, l'application des Transformers sur des données volumétriques en 3D permettrait de mieux capturer les relations spatiales entre les voxels, offrant ainsi une représentation plus fidèle des interactions de la dose au sein des tissus biologiques.

Références

[1] Automatants - CS Campus. Travaux pratiques sur u-net pour le débruitage d'images, 2025.

[2] Hong-Phuong Dang, Thibaut Wojdacki, Dimitris Visvikis, and Julien Bert. Apprentissage profond pour améliorer la qualité statistique et le temps de calcul de carte dosimétrique en radiothérapie. *LaTIM, University of Brest, INSERM UMR1101*, 2025. ECAM Rennes - Louis de Broglie, ENSAI & INSEE, Brest Hospital University.

[3] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+ : A full-scale connected unet for medical image segmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[4] Papers with Code. Vision transformer (vit), 2025.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015.

[6] Keiko Shibuya. Cutting edge research at yamaguchi university, 2025.