

Autoencoders for brain tumor segmentation

Nicolas Melaerts
253882

Janvier 2024

1 Abstract

La segmentation des tumeurs cérébrales permet d'améliorer le diagnostic, le suivi, et le traitement des patients atteints de pathologies neurologiques. Pour ce faire, les auto-encodeurs (AEs) sont très utiles dans ce contexte d'imagerie médicale. Ces modèles sont composés de deux parties principales : un encodeur qui extrait les caractéristiques importantes des images, et un décodeur qui reconstruit les images originales, comme les régions d'intérêt dans les images médicales. Les auto-encodeurs sont en effet très efficaces pour la segmentation d'images.

Ce projet vise à exploiter des auto-encodeurs pour effectuer la segmentation de tumeurs cérébrales à partir de scans CT. En utilisant des architectures avancées, comme Unet, LinkNet, FPN, PSPNet et DeeplabV3, et en s'appuyant sur une base de données disponible sur Kaggle, ce travail utilise ces différentes architectures pour voir si elles sont performante pour l'analyse de scan CT. Ces techniques de radiologie assistée par intelligence artificielle ont pour but d'apporter de l'aide aux médecins dans l'analyse de ce type d'image.

2 Introduction

Pour ce projet, l'objectif est d'explorer et de comparer différentes architectures d'auto-encodeurs appliquées à la segmentation de tumeurs cérébrales à partir de scans CT. La première étape consiste à utiliser l'architecture **Unet**, étudiée lors du TP. Une version améliorée sera également mise en oeuvre, combinant Unet avec un modèle préentraîné **Resnet-18** comme encodeur.

Ensuite, plusieurs autres architectures avancées seront explorées, notamment **Linknet**, **FPN** (Feature Pyramid Network), **PSPNet** (Pyramid Scene Parsing Network) et **DeepLabV3**, toutes utilisant **Resnet-18** comme encodeur.

Ici, l'objectif va être d'analyser les résultats obtenus avec ces différentes architectures, notamment avec la métrique appelée **IoU** qui mesure le rapport entre l'intersection et l'union des zones prédites et des zones réelles dans la segmentation. Ce qui permet d'évaluer la correspondance spatiale entre les deux. Il y aura aussi des comparaisons entre le temps d'entraînement pris par les différentes architectures, ainsi que la taille des modèles et le nombre de paramètres qu'ils contiennent.

3 Related work

La segmentation des tumeurs cérébrales à l'aide d'architectures d'auto-encodeurs a fait l'objet de nombreuses recherches ces dernières années. Notamment grâce aux avancées en apprentissage profond et à l'accès plus facile aux données médicales. Dans cette section, nous verrons brièvement quatre travaux récents qui exploitent des auto-encodeurs pour des tâches relativement similaires.

Le premier article, **"U-Net: Convolutional Networks for Biomedical Image Segmentation"** présente l'architecture U-Net comme une approche innovante pour la segmentation d'images biomédicales. U-Net repose sur un réseau convolutif avec un chemin contractant pour capter le contexte global et un chemin expansif pour une localisation précise des objets dans les images. Cette architecture est particulièrement adaptée aux situations où il y a peu d'images annotées disponibles, comme en imagerie biomédicale. En utilisant une augmentation des données par déformations élastiques et une stratégie de tiling pour des images de grande taille, l'U-Net surmonte les limites liées à la mémoire GPU tout en offrant des résultats de segmentation précis. L'approche a permis à U-Net de remporter plusieurs défis de segmentation, notamment dans le cadre de la compétition ISBI, avec des résultats supérieurs aux méthodes précédentes.

Un deuxième article intéressant appelé **"A Method of Image Semantic Segmentation Based on PSPNet"** présente une méthode améliorée de segmentation sémantique d'images basée sur le modèle PSPNet (Pyramid Scene Parsing Network), utilisant le pooling pyramidal spatial pour améliorer la précision de la segmentation. L'optimisation des paramètres du modèle est réalisée grâce à des calculs distribués sur GPU, permettant d'augmenter les performances du modèle. Les résultats expérimentaux, montrent une amélioration significative par rapport aux autres modèles existants dans le domaine de la segmentation sémantique. Cette méthode est particulièrement utile dans des applications comme la conduite autonome, la robotique, et l'analyse d'images médicales.

Le troisième article que j'ai choisi appelé **"Brain tumor image segmentation based on improved FPN"** propose une méthode de segmentation d'images de tumeurs cérébrales basée sur un réseau neuronal convolutif amélioré utilisant les réseaux de pyramides de caractéristiques (FPN). Le gliome est le type de tumeur cérébrale le plus courant, et nécessite une segmentation précise pour aider au diagnostic et au traitement. Les approches basées sur l'apprentissage profond, telles que le modèle U-Net, ont montré des résultats prometteurs, mais souffrent de limitations dans la gestion des détails fins et des relations entre pixels. L'amélioration proposée dans cet article combine les avantages des réseaux FPN et de l'architecture U-Net pour capturer des informations à plusieurs échelles, permettant une segmentation plus fine des bords des tumeurs, particulièrement utile pour les tumeurs de tailles variables. Ce modèle permet ainsi une meilleure extraction des caractéristiques des images IRM de tumeurs cérébrales, augmentant la précision de la segmentation tout en réduisant les erreurs liées aux bords.

Le dernier article **"Multi-Modal Semantic Segmentation Model using Encoder Based Link-Net Architecture for BraTS 2020 Challenge"** aborde la segmentation des tumeurs cérébrales, en particulier les gliomes, à partir d'images multimodales d'IRM. Une tumeur comme le gliome, présente un défi majeur en raison de la diversité de son apparence, de sa taille et de son emplacement dans le cerveau. L'article propose un modèle de segmentation basé sur un réseau Linknet. Ce modèle améliore la segmentation des tumeurs en exploitant les relations complexes entre les différentes modalités d'IRM. Comparé aux modèles traditionnels, il atteint une précision impressionnante de 99,2%, démontrant son efficacité pour le diagnostic des gliomes et sa robustesse dans des contextes cliniques.

3.1 Architectures Utilisées

Dans ce projet, nous utilisons plusieurs architectures d'auto-encodeurs avancées pour la segmentation des tumeurs cérébrales :

3.1.1 U-Net

L'architecture U-Net est un modèle de réseau de neurones convolutifs souvent utilisé pour la segmentation d'images biomédicales. Elle se distingue par sa structure en "U", avec un encodeur qui extrait les caractéristiques et un décodeur qui reconstruit l'image segmentée. Ce modèle

utilise des connexions résiduelles entre les couches correspondantes de l’encodeur et du décodeur, ce qui permet de préserver les informations spatiales et de mieux gérer les détails fins lors de la segmentation, ce qui est particulièrement important avec des images médicales.

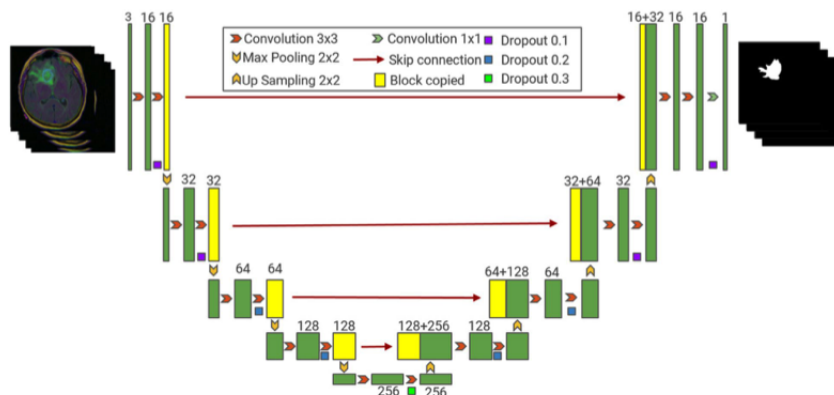


Fig. 1 Network structure of U-Net model

3.1.2 Linknet

Linknet est une architecture de segmentation d'images qui se distingue par son approche simplifiée et efficace. Contrairement aux architectures traditionnelles comme U-Net, LinkNet utilise des encodeurs et décodeurs connectés par des liens directs, permettant ainsi une propagation plus fluide des informations entre les couches. Cette approche réduit la complexité tout en maintenant une bonne précision. Linknet est particulièrement efficace pour des tâches de segmentation en temps réel, en raison de son efficacité computationnelle.

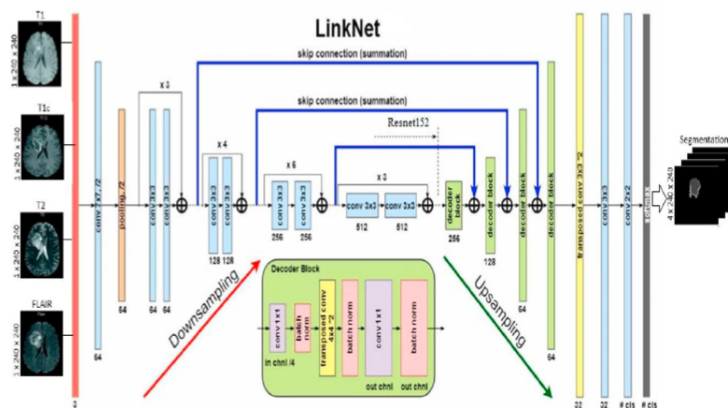


Fig. 2. Link-Net Architecture

3.1.3 FPN (Feature Pyramid Network)

Le réseau FPN (Feature Pyramid Network) est une architecture qui se concentre sur l'extraction de caractéristiques à différentes échelles. Il améliore les performances des modèles de détection et de segmentation d'objets en construisant une pyramide de caractéristiques, où chaque niveau contient des informations provenant de différentes résolutions d'image. Cette capacité à intégrer des informations à différentes échelles améliore la précision de la segmentation, en particulier pour les objets de tailles variées, et est largement utilisée dans les tâches de détection d'objets et de segmentation sémantique.

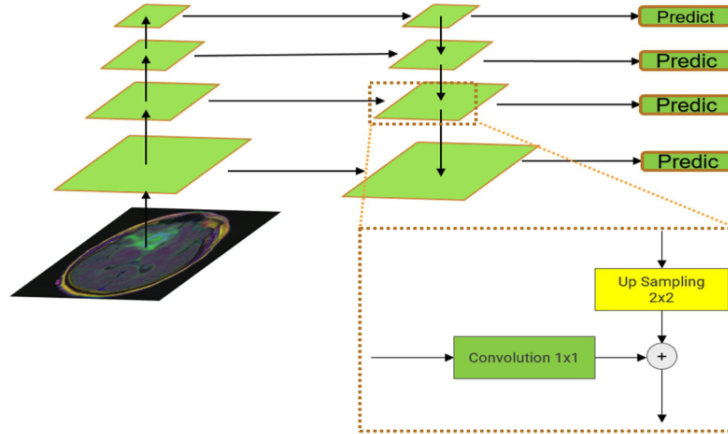


Fig. 2 Network structure of FPN model

3.1.4 PSPNet (Pyramid Scene Parsing Network)

Le PSPNet (Pyramid Scene Parsing Network) est un modèle avancé de segmentation sémantique qui se base sur une approche pyramidale pour analyser les scènes à différentes échelles. En combinant des caractéristiques provenant de plusieurs niveaux de la pyramide, PSPNet capture les informations globales et locales de l'image, ce qui permet une segmentation précise. Cette architecture est particulièrement utile pour segmenter des scènes complexes avec des variations d'échelle, comme celles rencontrées dans les images de villes ou de paysages.

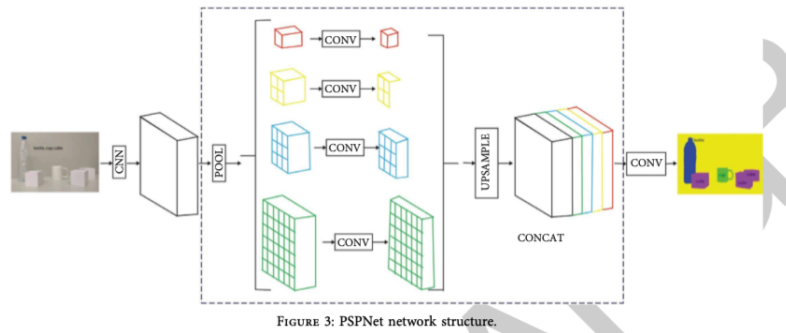


FIGURE 3: PSPNet network structure.

3.1.5 DeepLabV3

DeepLabV3 est un modèle de segmentation sémantique qui intègre des techniques de pointe telles que la convolution dilatée pour capturer de grandes zones contextuelles sans perte de résolution. Il utilise également une approche d'apprentissage multi-échelle, permettant de traiter efficacement des objets de différentes tailles. DeepLabV3 est connu pour sa capacité à segmenter des images complexes. Cette architecture est largement utilisée dans des applications de vision par ordinateur telles que la segmentation d'images satellites et la reconnaissance de scènes urbaines.

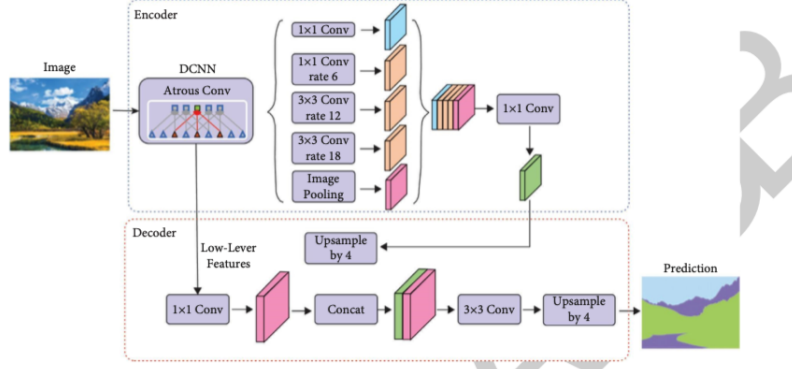


FIGURE 1: Encoder decoder network structure of DeepLab v3.

(toutes les images présentant les architectures proviennent des articles)

4 Proposed Approach

Pour expérimenter les différentes architectures de réseaux de neurones j'ai utilisé la base de données Kaggle. L'objectif était de comparer la performance de ces architectures en termes de **temps d'entraînement**, **taille du modèle**, **accuracy** et **IoU** (Intersection over Union). Pour ce faire, j'ai utilisé un sous-ensemble de 1000 images, avec un batch size de 16 et un learning rate de 0.0001, sur une durée d'entraînement de 30 epochs. J'ai dans un premier temps implémenté l'architecture Unet fournie lors du tp, et ensuite j'ai expérimenté avec l'architecture Unet de base et l'encoder resnet18. Ensuite, respectivement Linknet, FPN, PSPNet et DeeplabV3 toutes avec resnet18 également. À noter que les poids d'entraînement viennent de Imagenet et que j'utilise une fonction sigmoïd comme fonction d'activation.

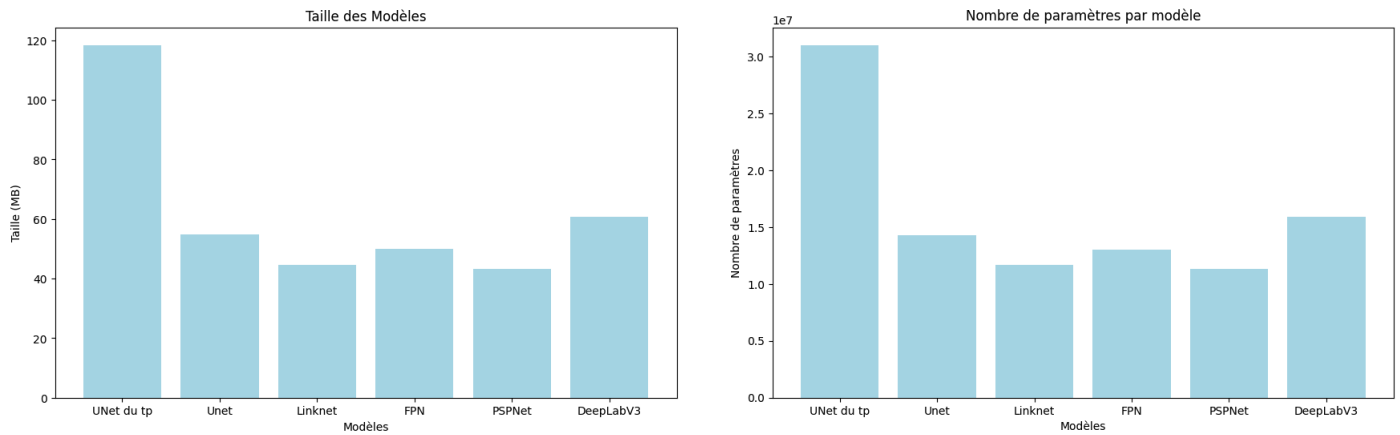
Il est important de signifié que le temps d'entraînement peut avoir été réduit à cause des conditions durant l'entraînement. Pour cause, j'arrête l'entraînement si la valeur de l'IoU ne s'améliore après 5 époques consécutives. Cela permet de réduire le temps d'entraînement si celui-ci ne s'améliore plus.

Et puis j'ai fait un dernier entraînement avec Unet de base et l'encoder resnet101 pour voir si ça améliorerait les performance par rapport à resnet18 et cela pourrait ouvrir la voie pour tester d'autres encore plus conséquent comme resnet152 ou efficientnet-b7 mais cela nécessiterait plus de mémoire GPU que ce à quoi j'ai accès avec Google Colab.

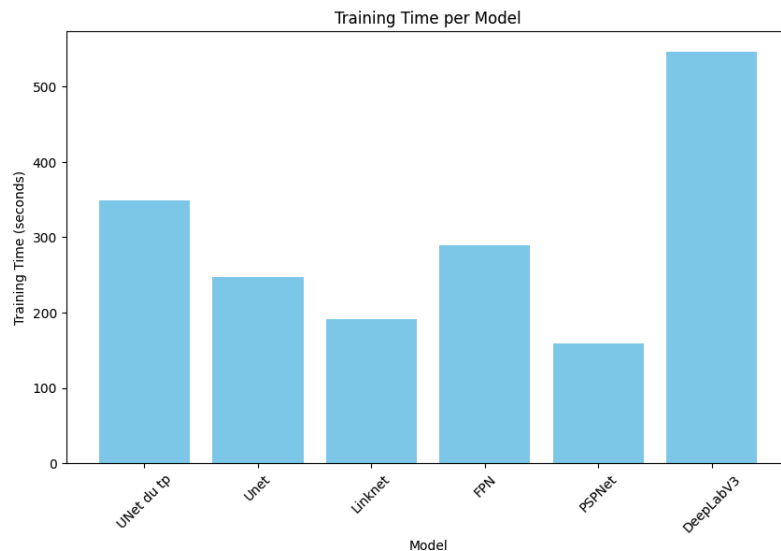
5 Experimental results

Cette section présente une analyse des résultats expérimentaux obtenus avec différentes architectures de modèles pour la tâche de segmentation d'images.

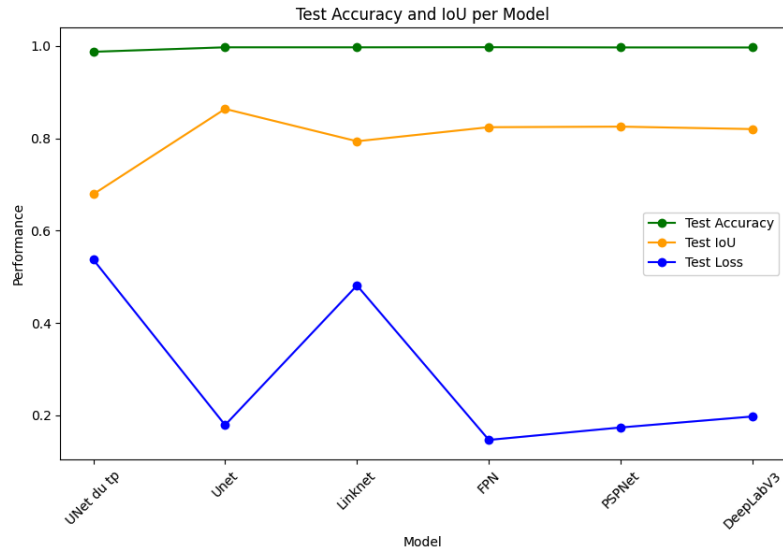
La version de U-Net utilisée dans le TP présente une taille plus de deux fois supérieure à celle des autres architectures. Cette taille plus élevée peut avoir un impact sur les performances en termes de consommation de mémoire et de temps d'entraînement, bien qu'elle puisse aussi potentiellement améliorer la qualité des prédictions en raison de sa capacité à apprendre des caractéristiques plus complexes. Mais on remarque que ce n'est pas le cas. On remarque aussi que la taille des modèles est directement proportionnelle au nombre de paramètres.



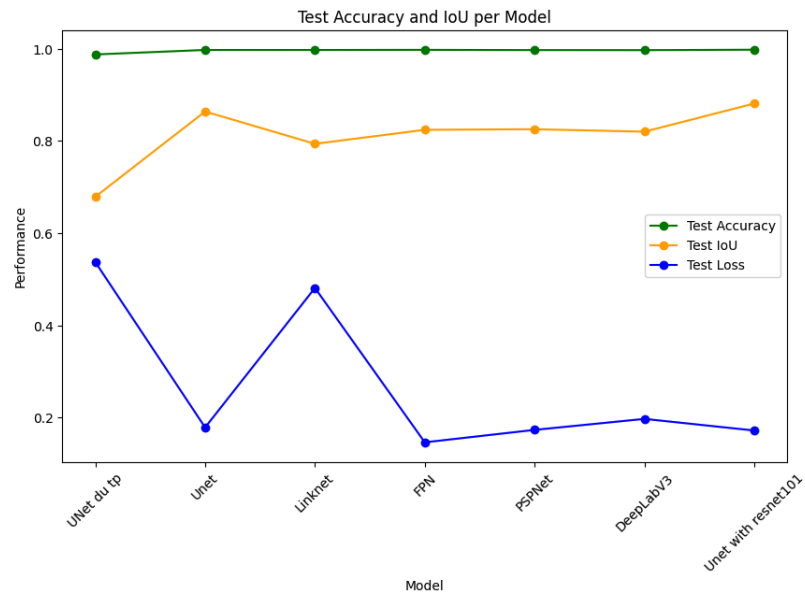
Comme on peut le constater, DeeplabV3 nécessite davantage de temps d'entraînement en raison de la complexité de son architecture, qui inclut des composants comme les convolutions dilatées, augmentant ainsi la charge computationnelle. Il va être intéressant dans le graphique suivant d'examiner si ce temps d'entraînement améliore la précision du modèle ou non.



L'IoU des différentes architectures, sauf U-Net du TP, dépasse les performances de ce dernier. Ce résultat pourrait être dû à la conception du modèle, qui ne bénéficie pas autant que les autres architectures de l'usage d'un modèle pré-entraîné comme ResNet18. Cela montre que sans optimisation fine, la performance des modèles converge, mais qu'une optimisation des hyperparamètres pourrait offrir un gain significatif.



Il est intéressant d'observer ici que l'IoU qui pour tous les architectures va être meilleures que Unet du tp. C'est en partie grâce à l'utilisation d'un modèle pré-entraîné comme resnet18. Il est intéressant de constater que les modèles pour cette tâche se valent (avec un petit avantage quand même pour Unet) lorsqu'il n'y a aucune optimisation des hyperparamètres.



L'utilisation d'un encodeur plus puissant comme ResNet101 a effectivement amélioré l'IoU, ce qui suggère qu'un encodeur plus complexe permet d'extraire des caractéristiques plus précises et donc d'améliorer la qualité des prédictions.

Ce résultat montre clairement que l'architecture du modèle et l'optimisation des hyperparamètres jouent un rôle clé dans la performance globale du modèle. Il reste donc une marge de progression importante.

6 Conclusion

Ce projet a permis d'explorer et de comparer différentes architectures d'auto-encodeurs avancées pour la segmentation de tumeurs cérébrales à partir de scans CT, notamment U-Net, LinkNet,

FPN, PSPNet et DeepLabV3. Cependant, aucune des approches évaluées ne peut encore être choisie comme solution optimale.

En effet, le choix d’une architecture adaptée dépend de multiples facteurs qui méritent une exploration plus approfondie, tels que :

Les contraintes en milieu clinique : le temps d’exécution, la complexité du modèle, et l’efficacité énergétique jouent un rôle clé dans le déploiement pratique. La précision requise : certains modèles, bien que performants globalement, peuvent manquer de finesse dans des cas spécifiques, comme la détection des contours d’une petite tumeur. Ce projet souligne donc la nécessité de mener davantage d’expérimentations. Cela inclut l’élargissement des jeux de données, le test de nouvelles métriques spécifiques au domaine médical, et l’exploration de modèles plus récents ou hybrides combinant les forces de plusieurs architectures.

Enfin, le choix d’une architecture en milieu médical doit être nuancé. Par exemple, Linknet pourrait être privilégié pour des tâches nécessitant une précision immédiate et des ressources limitées, tandis que DeepLabV3 conviendrait mieux dans des environnements où la puissance de calcul est disponible et où la capture des détails à grande échelle est essentielle.

7 Annexes

7.1 Base de données utilisée

La base de données utilisée pour ce projet est disponible sur Kaggle : <https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation>.

7.2 Articles consultés

Les articles suivants ont été utilisés pour approfondir les connaissances nécessaires :

- *U-Net: Convolutional Networks for Biomedical Image Segmentation* <https://arxiv.org/pdf/1505.04597>
- *A Method of Image Semantic Segmentation Based on PSPNet* <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2022/8958154>
- *Brain tumor image segmentation based on improved FPN* <https://link.springer.com/content/pdf/10.1186/s12880-023-01131-1.pdf>
- *Multi-Modal Semantic Segmentation Model using Encoder Based Link-Net Architecture for BraTS 2020 Challenge* <https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050923X00027/1-s2.0-S1877050923000534/main.pdf>

7.3 Documents pour la compréhension des modèles

Les ressources suivantes ont été consultées pour mieux comprendre les modèles étudiés :

- U-Net : <https://datasciencetoday.net/index.php/en-us/deep-learning/228-unet>
- LinkNet : <https://codeac29.github.io/projects/linknet/>
- Feature Pyramid Networks (FPN) :
 - <https://paperswithcode.com/method/fpn>
 - <https://jonathan-hui.medium.com/understanding-feature-pyramid-networks-for-object->
- Pyramid Scene Parsing Network (PSPNet) :
 - <https://paperswithcode.com/method/pspnet>

- <https://developers.arcgis.com/python/latest/guide/how-pannet-works/>
- DeepLabV3 :
 - <https://www.ikomia.ai/blog/understanding-deeplabv3-image-segmentation>
 - <https://www.kaggle.com/code/balraj98/deeplabv3-resnet101-for-segmentation-pytorch>