

Dépasser les limites des traceurs en imagerie TEP/TDM : une segmentation universelle des lésions pour AutoPET III

Zacharia MESBAH^{1,2,3} Léo MOTTAY^{1,2,3} Su RUAN^{1,2} Pierre DECAZES^{1,2} Sébastien HAPDEY^{1,2} Sébastien THUREAU^{1,2}
Romain MODZELEWSKI^{1,2}

¹Centre Henri Becquerel, 76038 Rouen, France

²Laboratoire AIMS, Équipe QUANTIF, Université de Rouen Normandie, 76000 Rouen, France France

³Siemens Healthineers, 92400 Courbevoie, France

Résumé – Le challenge MICCAI AutoPET rassemble la communauté de l’imagerie médicale depuis trois ans autour d’un sujet d’actualité : la segmentation automatisée des lésions cancéreuses sur les scanners TEP (Tomographie par Émission de Positrons). Chaque année, un aspect différent du problème est abordé ; en 2024, la multiplicité des traceurs TEP était au cœur du défi. Plus précisément, l’édition de cette année vise à développer un algorithme entièrement automatique capable d’effectuer la segmentation des lésions sur un scanner TEP/TDM, sans connaître le traceur, qui peut être un traceur FDG ou PSMA. Les défis de la tâche de segmentation proviennent de la diversité des types d’images (FDG ou PSMA) ainsi que des types de lésions (lymphome, mélanome, cancer du poumon), qui sont inconnus a priori.

Dans cet article, nous présentons une approche pour la segmentation entièrement automatique des lésions à partir des images multimodales TEP/TDM en exploitant la puissance du framework nnUNetv2[1]. Nous proposons également un MIP-CNN permettant de sélectionner dynamiquement l’ensemble de modèles le plus adapté.

Abstract – For the last three years, the AutoPET competition gathered the medical imaging community around a hot topic: automated lesion segmentation on Positron Emitting Tomography (PET) scans. Each year a different aspect of the problem is wrestled with. In 2024 the multiplicity of existing and used PET tracers was at the core of the challenge. Specifically, this year’s edition aims to develop a fully automatic algorithm capable of performing lesion segmentation on a PET/CT scan, without knowing the tracer, which can either be a FDG or PSMA-based tracer. In this paper we describe how we used the nnUNetv2[1] framework to train two sets of 6 fold ensembles of models to perform fully automatic PET/CT lesion segmentation as well as a MIP-CNN to choose which set of models to use for segmentation.

1 Introduction

Pour la détection et le diagnostic du cancer, l’imagerie par tomographie par émission de positons (TEP) est extrêmement précieuse. Elle permet d’explorer des fonctions spécifiques de l’organisme, ce qui est particulièrement adapté à la détection du cancer, car les cellules cancéreuses présentent un comportement anormal. La TEP consiste à injecter un traceur radioactif dans l’organisme du patient et à observer sa distribution grâce aux rayons gamma qu’il émet.

Les traceurs sont conçus pour cibler une fonction spécifique de l’organisme. Le plus courant, le ^{18}F -FluoroDesoxyGlucose (^{18}F FDG), est un sucre qui s’accumule (c’est-à-dire est absorbé) dans les parties du corps qui consomment de l’énergie (c’est-à-dire celles à haute fonction). Les cellules cancéreuses, dans leur frénésie reproductive incontrôlée, ont besoin de beaucoup d’énergie. Elles sont généralement très visibles sur les TEP, ce qui fait de cette modalité d’imagerie la clé de voûte des protocoles de détection et de traitement du cancer.

La TEP est toujours associée à une tomodensitométrie (TDM), qui permet non seulement de corriger son atténuation mais aussi de fournir des informations anatomiques. La combinaison des informations anatomiques et fonctionnelles permet aux médecins de déterminer les captations malignes (lésions) et non malignes (captation physiologique). Le foie, le cerveau et les reins sont des exemples de captations non ma-

lignes du FDG. Le foie et les reins ont une captation élevée en raison de leur rôle nettoyant, tandis que le cerveau est toujours actif et consomme donc du sucre.

Au cours de la dernière décennie, une autre famille de traceurs a été de plus en plus utilisée. L’antigène membranaire spécifique de la prostate (PSMA) est une protéine généralement présente dans les tissus prostatiques sains et cancéreux. Grâce à cette propriété, le PSMA a été utilisé en médecine nucléaire pour le dépistage du cancer de la prostate dans le corps humain. Contrairement au FDG, cette famille de traceurs ne s’accumule pas dans les organes les plus actifs : un cerveau sain ne présente aucune activité lors d’une TEP au PSMA. Cependant, les organes d’élimination des déchets restent actifs, tout comme d’autres organes comme les glandes lacrymales et le tractus gastro-intestinal.

Les TEP au FDG et au PSMA s’intègrent progressivement dans la pratique clinique courante, ce qui implique une augmentation de la charge de travail des médecins en matière d’analyse TEP. La première étape consiste à détecter et à délimiter toutes les lésions malignes. Le nombre de lésions et leur localisation sont en effet essentiels pour établir le pronostic du patient et déterminer le traitement ultérieur.

Pour alléger cette tâche, le développement d’un outil fiable de segmentation automatique des lésions TEP serait extrêmement utile.

| | Entraînement | | Validation | Test | |
|---------------------|--------------|-----|------------|------|-----|
| | LMU | UKT | / | LMU | UKT |
| Nb. de Scans | 1611 | | 5 | 200 | |
| Répartition | 1014 | 507 | 5 | 100 | 100 |

TABLE 1 : Distribution du jeu de données. LMU : Munich, UKT : Tubingen

2 Compétition AutoPET

2.1 Édition 2024

L'édition de cette année, AutoPET III, a, comme les années précédentes, pour but de voir émerger un algorithme capable de segmenter automatiquement les lésions cancéreuses sur les examens TEP/TDM. La nouveauté de cette édition est la présence d'un deuxième traceur. Pour rappel, les éditions précédentes ont concentré les efforts sur les TEP au 18F-FDG. Cette année, des TEP au PSMA étaient également incluses. Lors du test, les algorithmes n'ont pas d'information sur le traceur utilisé pour un examen donné. Ils doivent donc soit construire un outil "universel", soit trouver une méthode pour identifier le traceur utilisé.

La base de données est composée de 1014 examens TEP/TDM FDG et 597 examens TEP/TDM PSMA, récoltés dans deux hôpitaux allemands :

- Hôpital universitaire de Tübingen (données FDG)
- Hôpital universitaire LMU, Munich (données PSMA)

Pour chacun des cas, deux modalités d'imagerie sont fournies au format NifTi : la **TomoDensitométrie** (TDM) ré-échantillonnée vers l'espace du TEP et La **TEP**, sur laquelle le calcul du SUV (Standardized Uptake Value) a été effectué. Les patients sélectionnés étaient soit des patients sains, soit atteints d'une maladie parmi lymphome, mélanome, cancer du poumon pour les TEP FDG ; et des patients atteints de cancer de la prostate pour les patients PSMA.

2.2 Séparation du jeu de données

Le jeu de données est séparé en trois parties :

- Le jeu d'entraînement, que les participants doivent utiliser pour développer et entraîner leurs algorithmes. Il contient les 1611 cas décrits ci-dessus.
- Le jeu de test préliminaire, qui contient uniquement 5 cas. Ce jeu de test est présent uniquement pour permettre aux participants de s'assurer que leur algorithme fonctionne dans l'environnement fermé d'exécution de *grand-challenge*.
- Le jeu de test final qui contient 200 scans. C'est le jeu de test sur lequel sont réellement évaluées les contributions de chaque équipe, afin de déterminer le classement de la compétition. La moitié (n=100) de ces examens suivent la même distribution que le jeu d'entraînement (FDG venant de Munich et PSMA venant de Tübingen) et l'autre moitié inversant les centres (PSMA venant de Munich et FDG venant de Tübingen).

2.3 Évaluation finale

L'évaluation des algorithmes sur le jeu de données de test final est effectuée sur chaque sous-ensemble de données (LMU-FDG, LMU-PSMA, UKT-FDG, UKT-PSMA). Les métriques utilisées sont :

- Dice Similarity Coefficient (DSC) : une métrique de similarité géométrique
- False Positive Volume (FPV) : Le nombre de voxels situés dans des composants connectés faussement identifiés comme des lésions
- False Negative Volume (FNV) : Le nombre de voxels situés dans des composants connectés appartenant à des lésions qui n'ont pas été identifiés.

Chaque métrique est moyennée séparément, puis les moyennes sont ajoutées avec une pondération de 0.5 pour le DSC et 0.25 chacun pour FPV et FNV.

3 Méthode proposée

3.1 nnUnetv2

nnUNetv2 est un framework d'apprentissage des réseaux de neurones pour la segmentation. Cet outil permet d'entraîner un modèle de segmentation sans avoir à écrire une ligne de code. Des outils de segmentation peuvent être construits en 2 ou 3 dimensions en utilisant nnUNetv2. L'architecture, ses paramètres, les paramètres d'entraînement sont déterminés automatiquement à l'aide d'heuristiques. Le framework nnUNetv2 est l'outil qui obtient des résultats à l'état de l'art comme analysé dans un récent article[2]. Il est également considéré comme une référence incontournable pour comparer toute nouvelle méthode de segmentation, bien que cela ne soit pas toujours le cas. Il est utilisé dans la plupart de solutions gagnantes pour les challenges MICCAI qui incluent de la segmentation.

3.2 Fenêtrage

Nous adaptons une méthode de fenêtrage sur TEP et TDM, inspirée de l'article de Yousefirizi et al. [5]. Au lieu d'entrer les 2 canaux uniquement dans le réseau entraîné par nnUNet, nous proposons de créer deux nouveaux canaux :

- le scan TEP tronqué entre 0 et 20 SUV
- le scan TDM tronqué entre -300 et 400 HU

Le but de cette méthode est de résoudre deux problèmes inhérents à la normalisation des données :

- En normalisant entre les valeurs maximale et minimale, les valeurs de SUV ou HU sont écrasées, ce qui signifie que des variations légères dans l'image peuvent être dégradés, voire devenir imperceptible. Ceci est dû en partie aux limites du format d'encodage des nombres décimaux, sur 16 bits, utilisé lors de l'entraînement des réseaux de neurones profonds. En normalisant les scans sur une échelle de valeurs réduite, nous limitons la perte d'information due à cet effet.

- Pour chaque patient, la valeur maximale des TEP peut grandement varier (en particulier avec le PSMA). Cela signifie qu'une fois les données normalisées, les valeurs des voxels n'ont plus la même signification. Cela retire aux scans TEP leur nature quantitative, ce qui peut nuire à la capacité du réseau à les comprendre. Le fenêtrage réduit également cet effet.

3.3 Labels supplémentaires

Nous employons une méthode décrite lors de la compétition AutoPET II en 2023. L'équipe ayant remporté la première place a réduit de manière importante le nombre de faux-positifs en ajoutant les contours d'autres organes comme tâche supplémentaire, pour encadrer l'entraînement. Il est possible d'interpréter cette baisse dans le nombre de faux positifs comme une conséquence du réseau de neurone étant mieux capable de discerner les zones de forte absorption du traceur dues aux lésions de celles liées au fonctionnement normal des organes. Nous avons défini une liste d'organes concernées par l'absorption normale de traceur dans le cadre de l'imagerie TEP/TDM FDG ou PSMA :

- Le cerveau, le cœur et l'aorte : consommation importante de FDG
- Le foie, les reins et la vessie : évacuation du traceur
- La rate : montre parfois une activité sur les TEP FDG
- Le système digestif, la prostate : activité élevée sur les TEP PSMA
- Le squelette, les poumons et le pancréas : repères anatomiques

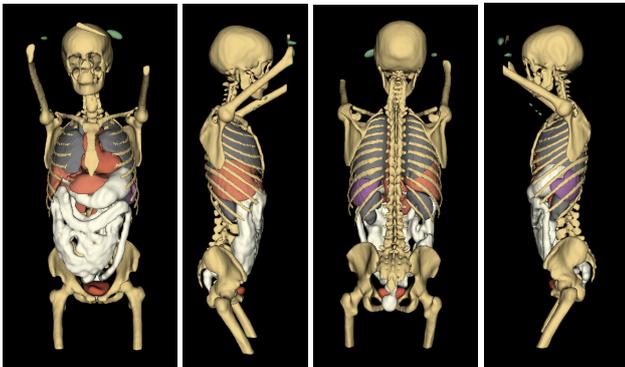


FIGURE 1 : Un exemple d'un patient avec des labels de lésions et d'organes sous différents points de vue

Nous avons utilisé l'outil TotalSegmentator[4] (TS) pour contourer tous ces organes sur les TDM. En regroupant plusieurs organes (par exemple les os dans la classe squelette), nous réduisons le nombre de canaux en sortie, ce qui réduit à son tour le temps de calcul et la mémoire utilisée lors de l'inférence.

En cas de recouvrement entre un organe tel que segmenté par TS et une lésion, la lésion était toujours prioritaire, puisqu'il s'agit du label correspondant à la tâche la plus importante.²

3.4 Classification des traceurs

Notre approche consiste à entraîner deux réseaux de neurones, chacun spécifique à un traceur. Cependant, comme le traceur utilisé n'est pas indiqué lors de l'évaluation, notre réseau doit obtenir cette information en se basant uniquement sur les images TEP/TDM. Pour résoudre ce problème, nous utilisons un réseau de neurones convolutif constitué de six couches de convolution 2D consécutives, suivies de cinq couches entièrement connectées et d'une activation sigmoïde. Les couches de convolutions et entièrement connectées utilisaient une fonction d'activation ReLU.

Pour passer nos volumes à travers ce réseau, nous effectuons le pré-traitement suivant :

- Ré-échantillonnage vers un espace de $3 \times 3 \times 3 \text{mm}^3$
- Calcul de la Maximal Intensity Projection (MIP)
- Zero-padding pour atteindre une taille unique de 224×224 si l'image est trop petite, troncature autour du centre de l'image si trop grande.

Pendant l'entraînement de ce modèle nous avons utilisé une fonction d'erreur entropie croisée binaire ainsi qu'un optimiseur AdamW avec un pas d'apprentissage de $1e^{-4}$ et durant 100 epochs avec arrêt anticipé pour réduire le sur-apprentissage. Les temps d'inférence sur l'ensemble des données, testés sur un processeur Intel i7-11850H, sont en moyenne de 2,18 secondes par patient et une validation croisée en 5 étapes a donné une précision de 99,64%.

3.5 Modèles de segmentation

Comme décrit précédemment, nous utilisons deux réseaux de segmentation, chacun spécifique à un traceur. Tous deux sont entraînés à l'aide du framework nnUNet, avec des paramètres/hyper-paramètres quasiment inchangés. Le seul changement concerne le paramétrage du modèle FDG. En effet, nnUNet utilise l'espacement médian comme espacement cible pour toutes les données d'entraînement. Cependant, l'espacement médian ($3 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$) pour les données FDG impliquait un temps d'inférence trop long, limité à 5 minutes par patient. Nous avons choisi de le remplacer par un espacement isotrope plus grand ($3,3 \text{mm} \times 3,3 \text{mm} \times 3,3 \text{mm}$), ce qui nous a permis d'utiliser différentes combinaisons d'assemblage de 6 modèles, avec des augmentations de données durant l'inférence (TTA) pour la plupart des patients. Nous fixons toujours un seuil en nombre de voxels à partir duquel nous supprimons une partie de TTA afin de respecter la limite de temps. Aucun post-traitement n'a été appliqué pour les deux modèles.

Deux modèles basés sur nnUNet ont été entraînés par les organisateurs pour servir de référence : le modèle **baseline** et le modèle **datacentric**. Le second modèle disposait d'une pipeline d'entraînement différente de celle du framework nnUNet.

4 Résultats

L'algorithme proposé améliore de manière importante les performances par rapport à la baseline proposée. Notre algorithme

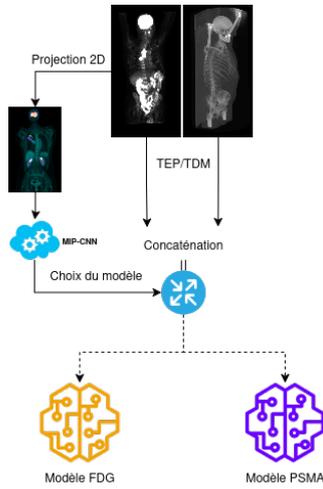


FIGURE 2 : Résumé de notre algorithme

| | DSC(%) | | FNV | | FPV | |
|--------------------|--------------|--------------|-------------|-------------|-------------|-------------|
| | psma | fdg | psma | fdg | psma | fdg |
| Baseline | 31.27 | 59.20 | 6.79 | 7.88 | 53.8 | 9.21 |
| Datacentric | 48.25 | 68.08 | 6.08 | 10.34 | 3.84 | 0.66 |
| Proposition | 57.47 | 67.25 | 6.26 | 4.51 | 2.44 | 5.38 |

TABLEAU 2 : Résultats de la compétition AutoPET III.



FIGURE 3 : Résultats des segmentations pour deux patients : en vert les contours proposés par notre algorithme. Les absorptions métaboliques sont évitées (reins, tube digestif, glandes, foie) et les absorptions malignes sont segmentées (principalement au niveau du bassin pour le patient de gauche, nombreuses métastases osseuses pour celui de droite)

parvient à discerner les lésions cancéreuses des organes montrant une absorption non pathologique du traceur. La plupart des lésions sont repérées correctement, cependant les limites des contours ne sont pas toujours fidèles à celles déterminées par les médecins.

Notre modèle nous a permis de prendre la huitième place sur dix-sept dans la compétition AutoPET III.

5 Conclusion

Le principal enseignement de notre participation à cette compétition est qu'un jeu de données contenant plus de 1000 volumes est un pas en avant dans la recherche d'un outil de contournage automatique des lésions cancéreuses sur TEP/TDM. La leçon supplémentaire de cette année est que différents traceurs peuvent être intégrés dans un même algorithme, soit en apprenant à discerner les traceurs depuis les images comme nous l'avons fait soit implicitement dans le modèle comme l'a montré l'équipe LesionTracer[3] qui a remporté la compétition dans la catégorie principale.

Si ils venaient à être adoptés en routine clinique, de tels algorithmes pourraient alléger la charge de travail des médecins en matière de segmentation TEP, leur permettant ainsi de consacrer leur temps à d'autres aspects de la routine et de la recherche clinique.

6 Bibliographie

Références

- [1] Fabian ISENSEE, Jens PETERSEN, Andre KLEIN, David ZIMMERER, Paul F. JAEGER, Simon KOHL, Jakob WASSERTHAL, Gregor KOEHLER, Tobias NORAJITRA, Sebastian WIRKERT et Klaus H. MAIER-HEIN : nnu-net : Self-adapting framework for u-net-based medical image segmentation, 2018.
- [2] Fabian ISENSEE, Tassilo WALD, Constantin ULRICH, Michael BAUMGARTNER, Saikat ROY, Klaus MAIER-HEIN et Paul F. JAEGER : nnu-net revisited : A call for rigorous validation in 3d medical image segmentation, 2024.
- [3] Maximilian ROKUSS, Balint KOVACS, Yannick KIRCHHOFF, Shuhan XIAO, Constantin ULRICH, Klaus H MAIER-HEIN et Fabian ISENSEE : From fdg to psma : A hitchhiker's guide to multitracer, multicenter lesion segmentation in pet/ct imaging. *arXiv preprint arXiv :2409.09478*, 2024.
- [4] Jakob WASSERTHAL, Hanns-Christian BREIT, Manfred T. MEYER, Maurice PRADELLA, Daniel HINCK, Alexander W. SAUTER, Tobias HEYE, Daniel T. BOLL, Joshy CYRIAC, Shan YANG, Michael BACH et Martin SEGEROTH : Totalsegmentator : Robust segmentation of 104 anatomic structures in ct images. *Radiology : Artificial Intelligence*, 5(5), septembre 2023.
- [5] Fereshteh YOUSEFIRIZI, Ivan S KLYUZHIN, Joo Hyun O, Sara HARSINI, Xin TIE, Isaac SHIRI, Muheon SHIN, Changhee LEE, Steve Y CHO, Tyler J BRADSHAW *et al.* : Tmtv-net : fully automated total metabolic tumor volume segmentation in lymphoma pet/ct images—a multicenter generalizability analysis. *European Journal of Nuclear Medicine and Molecular Imaging*, 51(7):1937–1954, 2024.