# AutoPET IV: Semi-automatic PET/CT across tracers

Zacharia Mesbah[1,2,4], Solène Perret[1,2,4], and Leo Mottay[1,2,4]

[1] INSA Rouen Normandie, Univ Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108, F-76000 Rouen, France
[2] Nuclear Medicine Department, Henri Becquerel Cancer Center, Rouen, France
[3] Radiotherapy Department, Henri Becquerel Cancer Center, Rouen, France
[4] Siemens Healthineers
zacharia.mesbah@chb-unicancer.fr

**Abstract.** For the last three years, the AutoPET competition gathered the medical imaging community around a hot topic: lesion segmentation on Positron Emitting Tomography (PET) scans. Each year a different aspect of the problem is presented; in 2024 the multiplicity of existing and used 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[**?**] 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.

**Keywords:** PET · Functional Imaging · Segmentation · Deep Learning

## 1 Introduction

When combined with Computed Tomography (CT), Positron Emission Tomography (PET) proves highly valuable, as it enables both the detection and monitoring of cancer by providing detailed metabolic and anatomical information. Currently, nuclear medicine physicians use only a subset of tumor lesions to evaluate tumor dynamics, from which they extract only one-dimensional information about the diameter. Precise lesion segmentation could allow for the extraction of a larger proportion of the tumor's morphological data. This could enable better assessment of cancer staging and more personalized treatment in the future. The problem is that manual segmentation is a time-consuming task, especially in cases of metastatic cancer. The solution lies in developing automatic lesion segmentation tools.

Automated segmentation continues to face considerable challenges, notably the variation in physiological uptake between different PET tracers, such as FDG and PSMA, hindering the development of a tracer-agnostic segmentation model. Models often struggle to distinguish between physiological and tumoral uptake.

One approach to address this issue is to include a human expert in the segmentation loop. The expert's role is to provide simple clicks to indicate whether a given pixel corresponds to a lesion or merely reflects physiological uptake. This method helps improve the model's performance by reducing false positives and missed segmentations. Moreover, the clicks required from the expert are quick and minimally time-consuming.

The autoPET challenge was created to provide researchers with a platform to directly tackle these issues. Now in its fourth edition, the challenge continues to evolve and address increasingly complex scenarios. The objective of the previous edition (autoPET III) was to develop algorithms capable of segmenting lesions in PET/CT scans acquired using either FDG or PSMA tracers, without prior knowledge of which tracer had been used. This year, the challenge introduces an interactive human-in-the-loop segmentation scenario, adding a new dimension to the task. Similar to the previous edition, the setting remains multi-tracer and multi-center, further reflecting the variability and complexity encountered in real-world clinical imaging.

This manuscript presents our proposed solution submitted to the autoPET IV challenge for the first task: single-staging whole-body PET/CT lesion segmentation.

## 2 Material and method

### 2.1 Dataset

**Images** The dataset consists of PET/CT images acquired using two distinct tracers: FDG and PSMA. The FDG cohort comprises 1,014 studies, including 501 patients with cancer and 513 healthy individuals. The PSMA cohort contains scans from 597 patients diagnosed with prostate cancer. All imaging data were collected from two clinical centers: University Hospital Tübingen, Germany (UKT), and LMU Hospital, LMU Munich, Germany (LMU). Three different PET/CT scanners were used for image acquisition: Siemens Biograph 64-4R TruePoint, Siemens Biograph mCT Flow 20, and GE Discovery 690. In addition to the imaging data, lesion segmentations were provided. These annotations were performed by different radiologists, depending on the originating medical center.

**Clicks** Furthermore, expert clicks—represented as sets of 3D coordinates—are included for each image. Each click is labeled as either background or tumor, indicating the corresponding region type. These clicks serve as a form of weak supervision to guide the network toward relevant anatomical or pathological areas during training.

### 2.2 Our Method

**Heatmaps** Heatmaps are generated based on the input clicks following the procedure below. Tumor clicks and background clicks are processed differently.

For tumor clicks, a zero-filled array named $heatmap_{tumor}$ is first created with the same dimensions as the PET image. For each tumour click, we create a zero-filled local array of the same size, then the voxel at the corresponding coordinates is set to 1 in the local array. A Gaussian filter is then applied to this local array, which is subsequently added to $heatmap_{tumor}$. Next, all voxels with an SUV value lower than the minimum between the SUV at the click location and 3 are set to zero. Finally, the values in $heatmap_{tumor}$ are clipped to the range [0, 1].

For background clicks, a zero-filled array named $heatmap_{background}$ is created with the same dimensions as the PET image. For each background click, the voxel at the corresponding coordinates is set to 1. Once all clicks are processed we apply a gaussian filter on the whole array. Then all voxels with a value greater than zero in the $heatmap_{tumor}$ array are set to zero in $heatmap_{background}$. The final heatmap is computed as $heatmap_{tumor} - heatmap_{background}$.

In the final heatmap, voxels with a positive value represent parts of the volume influenced by a positive click, likely to be part of a lesion. In the opposite, voxels with negative values represent parts of the volume influenced by a negative click. The network should learn to avoid classifying voxels in this area as part of a lesion.

**Tracer Discriminator** Building on last year's work, we also used a MIP-CNN network to automatically determine which tracer was used for the PET/CT scan. The output of this model defines the segmentation model which will be used to process the scan. This model is, as described in [**?**], a convolutional neural network which processes a 2D Maximum Intensity Projection (MIP) of the PET volume. Depending on the output of this discriminator network, we run the inference using either the PSMA or the FDG model. Both of these models run as a 5-folds ensemble in classic nnUNet fashion. The number of allowed mirroring axes is determined dynamically based on the processed volume's size (see below).

**Segmentation Models** We used the nnUNet framework to train two ResidualEncoderUNet models, more precisely using the L sized architecture. The patch size was set to $[192, 192, 192]$, with remaining experiment parameters unchanged. The FDG model was trained for 350 iterations per epoch, the PSMA model was trained for 250 iterations per epoch. Each model was trained for 1500 epochs. The trainers we used are modified versions of the ones proposed in [**?**], with the same data augmentation methods (including the misalignment data augmentation).

Last year's winning solution relied on pretraining the model on a large medical imaging dataset. The weights of the pretrained model were made publicly available by the team after last year's competition. We used these weights as the initial weights to train our models.

**Additional Organs Supervision** Using organs as additional labels has been shown to increase the performance and in particular reducing the amount of false

positives. We selected organs which exhibit tracer uptake without malignancy, namely the heart, brain, aorta, liver, spleen, prostate, parotid and submandibular glands as well as parts of the digestive system. We add the kidneys and bladder since they often contain high SUV values since they eliminate the tracer. Finally we include the lungs and skeleton as we hypothesize that including anatomical landmarks in the supervision makes it easier for the network to learn a coherent representation of human anatomy. We used a double headed model, with one segmentation head for lesions only and one for the organs' segmentation. At inference time, the organs segmentation head is disabled so we retrieve only a binary segmentation map: the lesions contours.

**Post-Processing** During the AutoPETII competition, Alloula et al.[**?**] showed that the trained segmentation models tend to output more small components in their segmentation than are initially present in the dataset. To deal with this inconsistency, they removed single connected components smaller than a fixed number of voxels. Through experiments they found the optimal threshold to be around 10 voxels. We adapted this small components removal method to PET scans by removing single connected components both smaller than 10 voxels and having a maximum SUV value lower than 4. This aims to further reduce the number of false positives.

**Dynamic Inference Depth** Inference must run in less than 15 minutes per case. Depending on the dimensions of the input image, our algorithm takes a different amount of time to run, mainly due to the fact that the number of forward passes needed depends on the number of patches needed to cover the volumes. Instead of playing it safe for all patients by removing mirroring altogether, we developed a dynamic algorithm which maximizes the number of allowed axes while respecting the time limit. For this algorithm, we looked for the optimal voxels number thresholds for the number of allowed mirroring axes, considering that each axis we remove roughly means a 2x speedup. We ran the inference algorithm with no mirroring for values for all dimensions ranging from 100 to 1000 voxels and measured the runtime for each combination. Considering the difference in performance between our GPU (Quadro P6000) and the one proposed by grand-challenge (Tesla T4), we chose to set the threshold for 1 mirroring axis at 90 millions voxels, 2 mirroring axes at 39M voxels and all 3 axes at 18M voxels.

## 3 Results

Using this method, we reached first place in the preliminary test set. More importantly, our method showed strong performance over all the used metrics: Dice Score, False Positive Volume and False Negative Volume.

# 4 Discussion

We designed a strong, robust automated PET/CT segmentation method which works in a human-in-the-loop scenario with input clicks.

## 4.1 Limitations

For post-processing we use a number of voxels instead of a real volume (e.g.: 200 cubic millimeters) as threshold.