# Med3D: Exploring the design space of nnUNet on human-in-the-loop tumor segmentation

Yang Xing[1], Yuxin Gong[1], and Kuang Gong[1]

University of Florida, Gainesville FL 32608, USA {yxing2, kgong}@ufl.edu

**Abstract.** This work presents our solution to Task 1 of the MICCAI 2025 AutoPET-IV Challenge, focusing on human-in-the-loop lesion segmentation from PET/CT images. To effectively incorporate sparse 3D user interactions in the form of clicks, we extend the 3D full-resolution nnU-Net framework by introducing an additional interaction channel. Each click, indicating tumor or background regions, is encoded as a Gaussian heat map and fused with the PET and CT modalities as network input. This enables the model to leverage both imaging information and expert guidance for accurate lesion delineation. The model was trained on 1,288 PET/CT studies and validated on 323 samples, achieving Dice scores of 0.7982 on the training set and 0.7946 on validation. Our approach demonstrates that augmenting a robust baseline segmentation network with human-in-the-loop interactions can improve accuracy while maintaining efficiency, aligning with the challenge objective of practical, expert-guided lesion segmentation.

**Keywords:** autoPET challenge · image segmentation · gaussian filter

## 1 Introduction

The AutoPET-IV challenge task 1 focuses on human-in-the-loop lesion segmentation from PET/CT images. In this setup, human interactions are provided in the form of clicks, which are sets of 3D coordinates identifying either tumor regions or background. Incorporating these sparse and discrete annotations effectively is crucial to improve segmentation performance while reducing annotation effort. In this work, we adapted nnUNet as the baseline segmentation framework and extended it by incorporating interaction-aware guidance maps derived from the human clicks. Specifically, we converted each set of clicks into a heat map representation, which was added as an auxiliary input channel alongside the PET and CT modalities. This approach enabled the network to leverage both imaging modalities and user-provided interaction cues for improved lesion detection and delineation.

## 2 Methods

### 2.1 Data

**Training and validation data** Dataset was obtained from MICCAI 2025 AutoPET-IV challenge[4]. The dataset comprises co-registered 3D PET and CT volumes of whole-body scans, typically ranging from the skull base to mid-thigh, with potential extensions to whole-body coverage depending on clinical relevance. It consisted of a total of 1,611 whole-body PET/CT studies, including 1,014 FDG-PET/CT scans drawn from 900 patients (501 with histologically confirmed malignant lesions and 513 negative controls), and 597 PSMA-PET/CT scans from 378 patients (of which 537 show PSMA-avid tumor lesions and 60 are negative). Dataset was split into training and validation data. Training data consisted of 1288 samples and validation data consisted of 323 samples.

### 2.2 Data pre-processing

Each human interaction consists of a set of sparse 3D coordinates (clicks) indicating tumor (foreground) or background regions. To incorporate this guidance into the segmentation model, we first resampled the original PET and CT volumes to the standard nnU-Net full-resolution grid and normalized intensities—CT values were clipped to a fixed Hounsfield Unit (HU) range and z-score normalized, while PET volumes were transformed into standardized uptake values (SUVs) and similarly normalized. Concurrently, each click was mapped into the resampled voxel grid and encoded as a localized Gaussian distribution:

$$G(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|^2}{2\sigma^2}\right),$$

centered at the click coordinate $\mathbf{c}$, where the standard deviation $\sigma$ (e.g., 3–5 voxels) controlled the spread. Foreground clicks generated positive peak Gaussians, while background clicks were represented using negative peaks, ensuring distinct spatial cues. All click-based Gaussians were superimposed to form a single-channel *interaction heat map* that aligned with the PET and CT voxel grids. This resulted in a three-channel input volume—PET, CT, and interaction heat map—that enabled the downstream 3D nnU-Net model to effectively leverage both imaging data and human-in-the-loop guidance for lesion segmentation.

### 2.3 Algorithm/model

For this task, we adopted the 3D full-resolution nnU-Net[3] as the baseline segmentation framework, following the AutoPET-II fine-tuning protocol, but extended it to integrate human-in-the-loop guidance. The base network was a 3D U-Net variant with residual encoder–decoder blocks, which improved gradient flow and stabilizes optimization in deep architectures.[2] Instead of using only the two imaging modalities as input, we augmented the network with an additional interaction channel derived from human clicks. Specifically, each set of foreground and background click coordinates was transformed into a Gaussian heat map, which is resampled to match the spatial dimensions of the PET/CT images. The final input to the network was therefore a 3-channel volume, consisting of the PET scan, the CT scan, and the click-derived heat map. This design allowed the model to leverage both imaging information and sparse human feedback simultaneously, guiding the segmentation process toward clinically relevant lesion regions. The network output a binary lesion segmentation map, with predictions optimized using a combination of Dice and cross-entropy losses, consistent with nnU-Net training defaults.

### 2.4 Data post-processing

Default post-processing methods of nnUNet[1] were included as the data posprocessing methods. It included removing small connected elements which are unlikely to be leisons and only keeping large connected elements.

### 2.5 Training and test parameters

The model was trained on 8 Nvidia B200 gpus which have a GPU memory of 180 GB. Therefore, to fully utilize the computation efficiency, batch size was set to 240. Also, the number of dataloader process was set to 28 to make sure the dataloading process was not the bottleneck. Learing rate was set to 1e-2 with polynomial decay and SGD with momentum was applied as optimizer.

### 2.6 Github repository

Link to Github repository: https://github.com/astlian9/AutoPETIV_solutions/

## 3 Results

5-fold cross validation was applied by using nnUNet default dataset processing methods. However, only fold 0 was used due to time limitation. Dice score of training data was 0.7982 and validation data was 0.7946.

## 4 Conclusion

This solution leverages nnUNet's robust baseline performance while extending it to incorporate human-in-the-loop interactions via heat map encoding of user clicks. By treating the interaction maps as an additional input channel, the model learns to align lesion predictions with sparse human guidance. The approach combines automation with expert feedback, aligning well with the challenge goal of efficient, accurate lesion segmentation.

## References

1. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z

**Table 1.** Algorithm details

| **Team name** | **algorithm name** (as submitted on grand-challenge) | **data pre-processing** | **data post-processing** | **training data augmentation** | **test time augmentation** | **ensembling** (e.g. cross-validation, model ensemble, ...) | **standardized framework?** (e.g. nnUNet, MONAI, ...) | **network architecture** (e.g. UNet (3D)) | **loss** | **training data** | **data/model dimensionality and size** (e.g. 2D: 128x128, 3D: 128x192x160, ...) | **use of pre-trained models** (public available or own developed) | **GPU hardware for training** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Med3D | nnUNet_redesigned | Normalization | - | Flipping, Random Rotation | - | - | nnUNet v2 (3D_fullres) | UNet (3D) | DSC + CE | 1014 FDG + 597 PSMA PET-CT of autoPET | 3D: Dx400x400, where D is the actual depth of sample | - | 8x Nvidia B200 |

2. Isensee, F., Maier-Hein, K.H.: Look ma, no code: fine tuning nnu-net for the autopet ii challenge by only adjusting its json plans (2023)
3. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation (2024), https://arxiv.org/abs/2404.09556
4. Küstner, T.: Automated lesion segmentation in whole-body pet/ct and longitudinal (autopet/ct iv) (Mar 2025). https://doi.org/10.5281/zenodo.15045096, https://doi.org/10.5281/zenodo.15045096