EFFICIENT AUTOMATIC PROMPT GENERATION FOR MULTI-ORGAN SEGMENTATION WITH MEDSAM2

Marie Friot, Mohamed Abderrahmane Lichir, Abdessalam Kouiss, Fatima Elsa Moukoudi Ntongono, Razmig Kechichian

INSA-Lyon, Universite Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, LYON, France

Introduction

Abdominal organ segmentation is crucial for clinical decision. While U-Net [1] is the state-of-the-art for 3D CT multi-organ segmentation, foundation models such as SAM [2] are gaining popularity due to their good generalization capacity for 0 shot segmentation, which however requires manual prompts. Automated methods for SAM prompting include image registration [3] and learned detector-based approaches [4], however they are largely restricted to 2D segmentation due to the original SAM design. We use MedSAM2 [5], a foundation model for 3D clinical scans, to develop an automated segmentation framework requiring little to no training, useful for low-data regimes, which attains performance on par with U-Net.

Methods

MedSAM2 processes 3D scans via slice-wise 2D inference. It requires a prompting region and at least one 2D axial bounding box prompt for each organ in that region. The model was originally trained with one box in the center of the prompting region, but we also test a three-box strategy, by placing prompts at the 25%, 50%, and 75% quartiles of the prompting region.

No-Learning prompting (NL): We use the FROG algorithm [6] to register 15 atlases (image + mask pairs) onto a target image. Registered masks are aggregated into a confidence map to define the prompting region which is processed with 3 bounding boxes to create a first segmentation. Final segmentation is obtained with a cascaded strategy: first-pass masks serve as refined prompts for a second MedSAM2 pass, where we extend the prompting region and boxes to remedy initial under-segmentation caused by poor first-stage prompts.

Prompting with learned detectors: We trained a lightweight ImageNet-pretrained YOLO [8] (Yolo-n) on the abdominal 3D CT dataset AMOS22 [7] for 5 h, along with a larger one (Yolo-x) trained from scratch for 25 h to explore performance bounds of the method. The detectors perform 3D organ detection, defining their prompting regions, from which 1 or 3 boxes are extracted to guide MedSAM2.

Results

We evaluate both methods on 30 scans in the AMOS22 validation set. Mean Dice scores calculated for 15 organs in all images are given in Table 1. Figure 1 gives qualitatives segmentation results. We assess statistical significance with one-sided paired t-tests.

U-Net	Yolo-n 1 box	Yolo-n 3 boxes	Yolo-x 3 boxes	NL
0.82 ± 0.10	0.79 ± 0.12	0.81 ± 0.12	0.82 ± 0.11	0.69 ± 0.17

Table 1: Mean Dice scores and standard deviation.

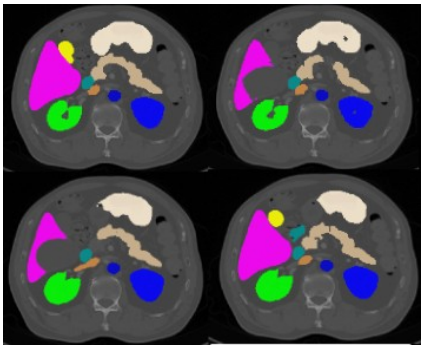

Figure 1: Segmentation results. Top: reference (left), U-Net (right). Bottom: NL (left), Yolo-n 3 boxes (right)

Discussion

Three-box prompting significantly improves results, notably for highly-variable organs such as the pancreas (Mean Dice: 0.71 vs 0.77) and tubular structures such as the postcava (0.79 vs 0.82) by reducing spatial ambiguity and providing a shape prior via quartile boxes. Yolo-x gives significantly better results than Yolo-n with 3 boxes. There is no significant difference between YOLO and U-Net models, however the latter requires longer training: 54 h. Thus, our approach gives similar performance with markedly lower training time than current SOTA. Yet, this training efficiency trades off inference speed: while U-Net infers in 2.16 s on GPU, our lightest strategy (Yolo-n, 3 boxes) requires 30 s per scan. Our NL method gives an overall lower mean score as registration struggles with small, highly-variable organs leading to a 16 mm localization mean average error for the prompting zone center (8 mm for Yolo-n). Nevertheless, it has comparable performance on compact, stable organs such as right and left kidneys (Mean Dice, NL: 0.94 vs U-Net: 0.93). Future work will address improving the robustness of our method for variable organs and its speed of inference.

Références

1. Çiçek, Ö. et al, MICCAI, 424–432, 2016
2. Kirillov, A. et al, ICCV, 4015–4026, 2023.
3. Xu, J. et al, preprint arXiv:2411.17363, 2024.
4. Yin, D. et al, Expert Syst Appl, 127048, 2025.
5. Ma, J. et al, preprint arXiv:2504.03600, 2025
6. Ji, Y. et al, preprint arXiv:2206.08023, 2022.
7. Agier, R. et al, Med. Image Anal., 101564, 2020.
8. Jocher, G. et al, Ultralytics YOLO11 v11.0.0, 2024