Efficient automatic prompt generation for multi-organ segmentation with MedSAM2

Marie Friot (1), Mohamed Abderrahmane Lichir, Abdessalam Kouiss, Fatima Elsa Moukoudi Ntongono, Razmig Kechichian (1)
INSA-Lyon, Universite Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, LYON, France
(1) email addrress: first.last@insa-lyon.fr

Context

- SOTA for **3D multi-organ segmentation**: task-specific **U-Net** [1]
- **Foundation models**:
  - Good **generalization** capacities: zero-shot segmentation
  - Useful for **low-data regime**
  - Require **manual prompt**:
    - Previous automated methods for prompting:
      - Image registration [2] and learned detecor [3]

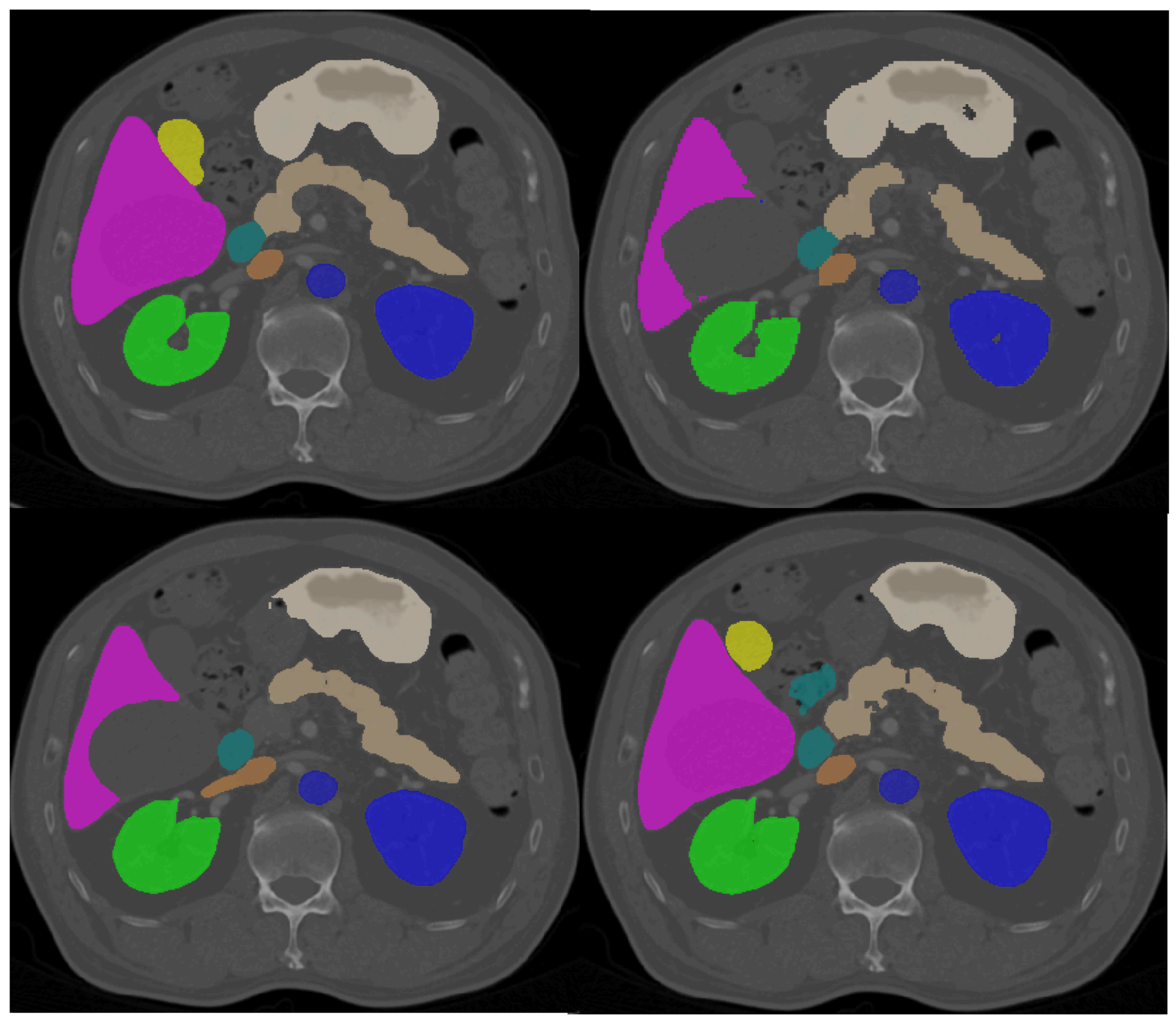

Figure 4- Segmentation results. Top: reference (left), U-Net (right). Bottom: NL (left), Yolo-n 3 boxes (right)

Methods

MedSAM2 [4]: slice-wise 2D inference for 3D scans

- **Prompting requirements**:
  - A **Prompting Region** (PR) per organ
  - One **2D bounding** box per organ within its PR.
- **Prompting strategies:**
  - Baseline: **A single box** at the center of the PR
  - **Three-box strategy**: Boxes at the 25%, 50%, and 75% quartiles of the PR (Figure 1)

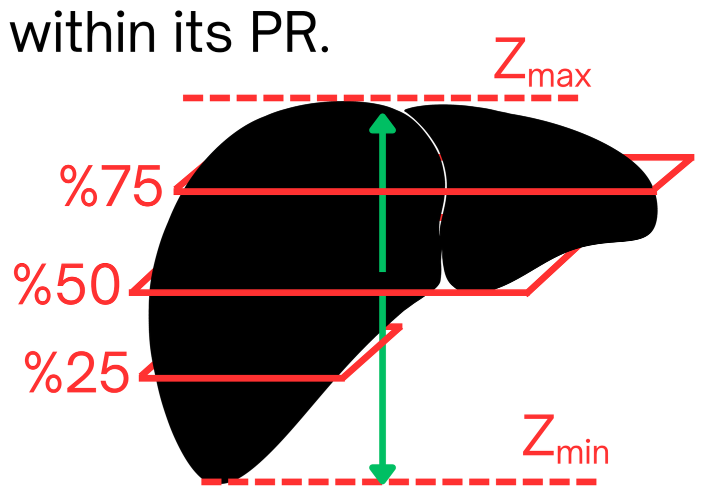

Figure 1 - Three-box prompting strategy

- **Automated prompting** methods
  a. No Learning prompting (NL) (Figure 2)
  b. Prompting with learned detector (Figure 3)
    - YOLO [5] models trained on the abdominal 3D CT dataset AMOS22 [6]

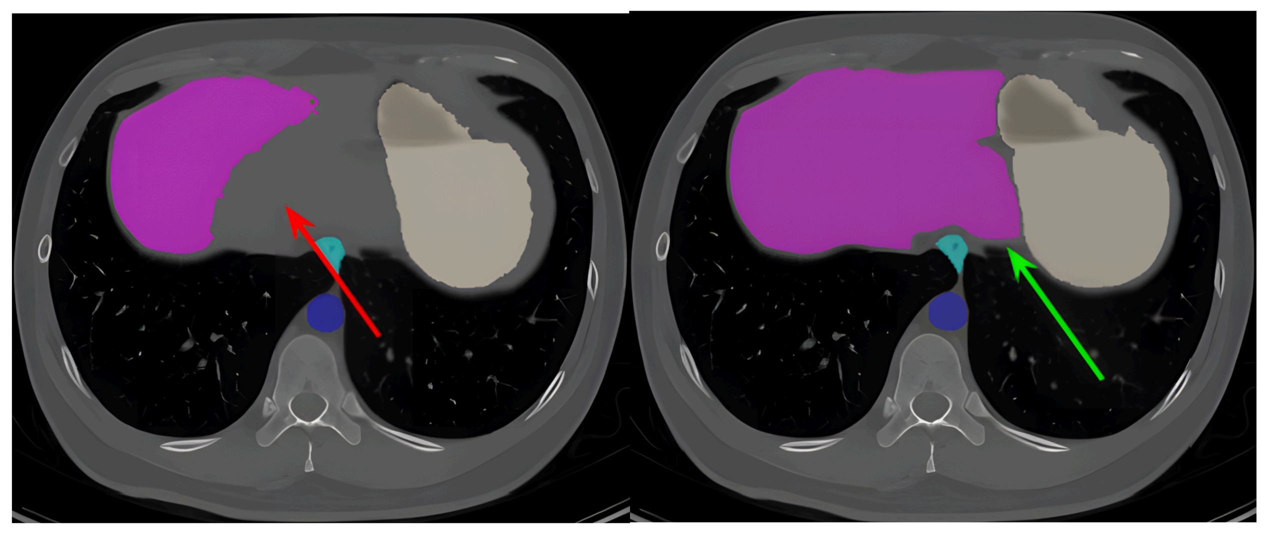

Figure 5- NL : Top: First pass segmentation with MedSAM2, Bottom : Second pass

Results

- Evaluation on 30 scans in the **AMOS22 validation set**.

p-value = 0.003 (Yolo-n 1 box vs Yolo-n 3 boxes)

U-Net	GT 3 boxes	Yolo-n 1 box	Yolo-n 3 boxes	Yolo-x 3 boxes	NL
0.82 ± 0.10	0.86 ± 0.09	0.79 ± 0.12	0.81 ± 0.12	0.82 ± 0.11	0.69 ± 0.17

Table 1 - Mean Dice scores and standard deviation calculated for 15 organs in all scans.

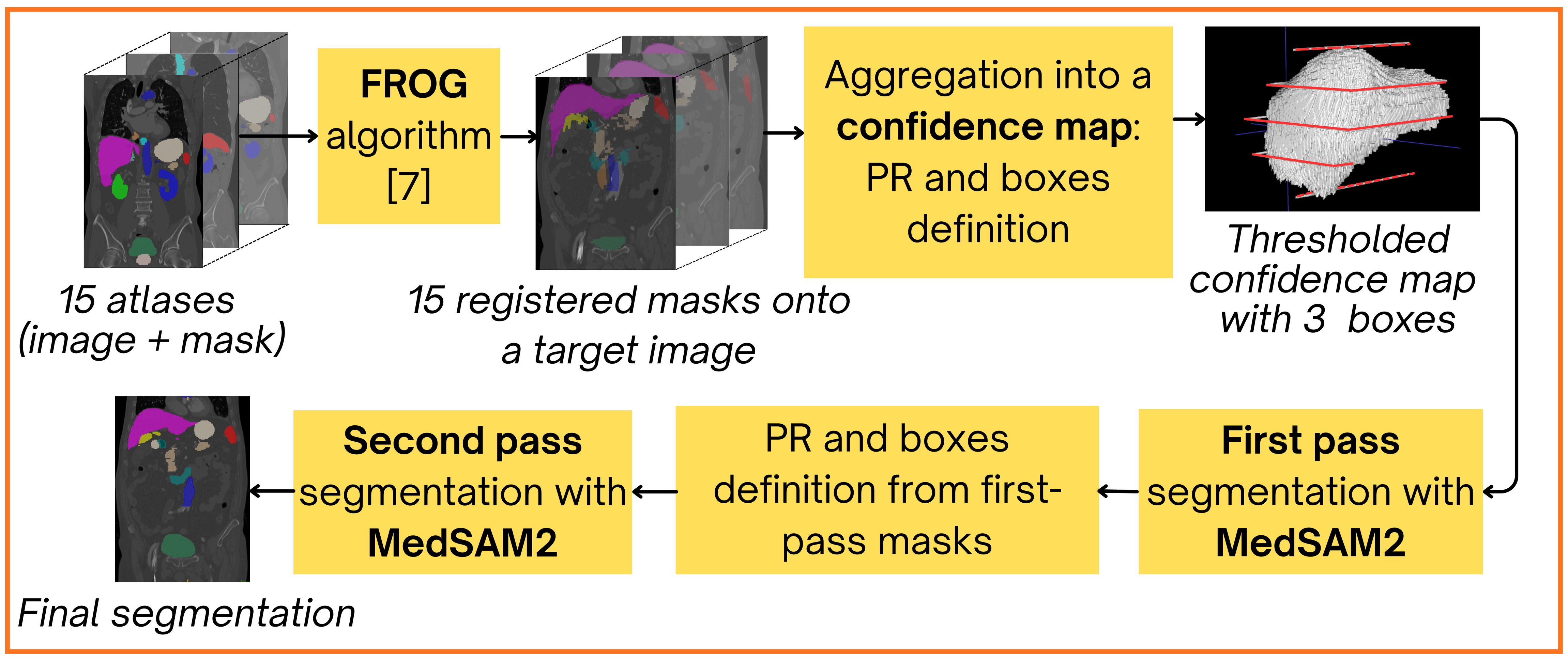

Figure 2 - No learning prompting method.

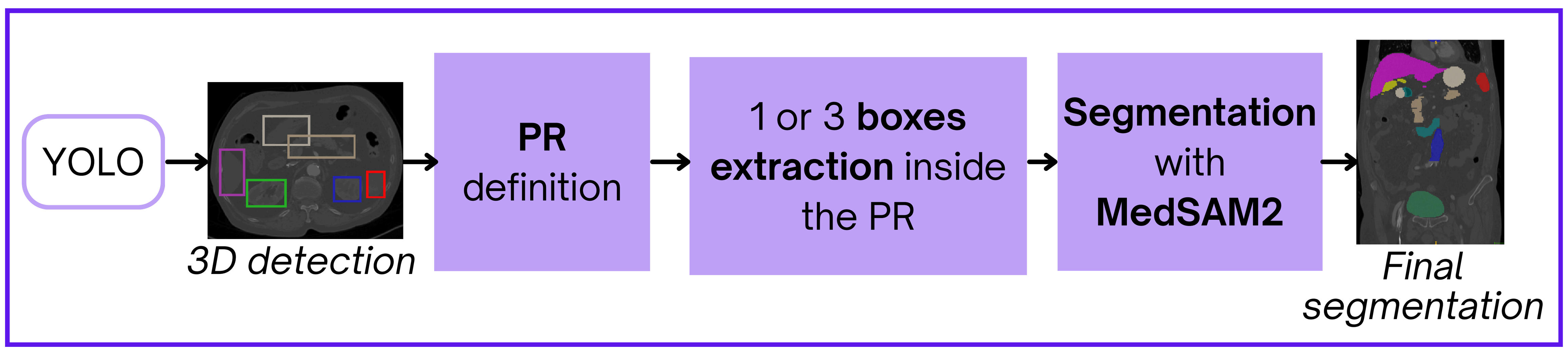

Figure 3 - Prompting with learned detector method.

Analysis

Prompting with learned detector

- **Three-box prompting** improves results
  - Notably for **highly-variable organs:** pancreas (Mean Dice 0.71 vs 0.75) and tubular structures: **postcava** (0.79 vs 0.82)
- **Training time**: U-Net 54 h, Yolo-n 5 h, Yolo-x 25 h
- **Inference speed** on GPU: U-Net 2.16 s, Yolo-n 3 boxes 22.5 s

No Learning prompting

- Registration struggles with **small, highly-variable** organs
- **16 mm localization MAE** for the PR center against **8 mm** for **Yolo-n**
- **Comparable** performance on **compact, stable organs**: right and left kidneys (Mean Dice, NL 0.94 vs U-Net 0.93)

Future Work

- Improve segmentation result with better prompting
- Enhance MedSAM2 robustness to prompt variability
- Improve inference speed

References

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

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON · Université Claude Bernard Lyon 1 · cnrs · Inserm · CREATIS · UNIVERSITÉ DE LYON