AUTOMATED MEDICAL IMAGE SEGMENTATION VIA YOLO-NAS AND MEDSAM2 WITH POST-TRAINING COMPRESSION

Thi - Nguyen Ngoc Lan

University of Information Technology
HCMC, Vietnam

What ?

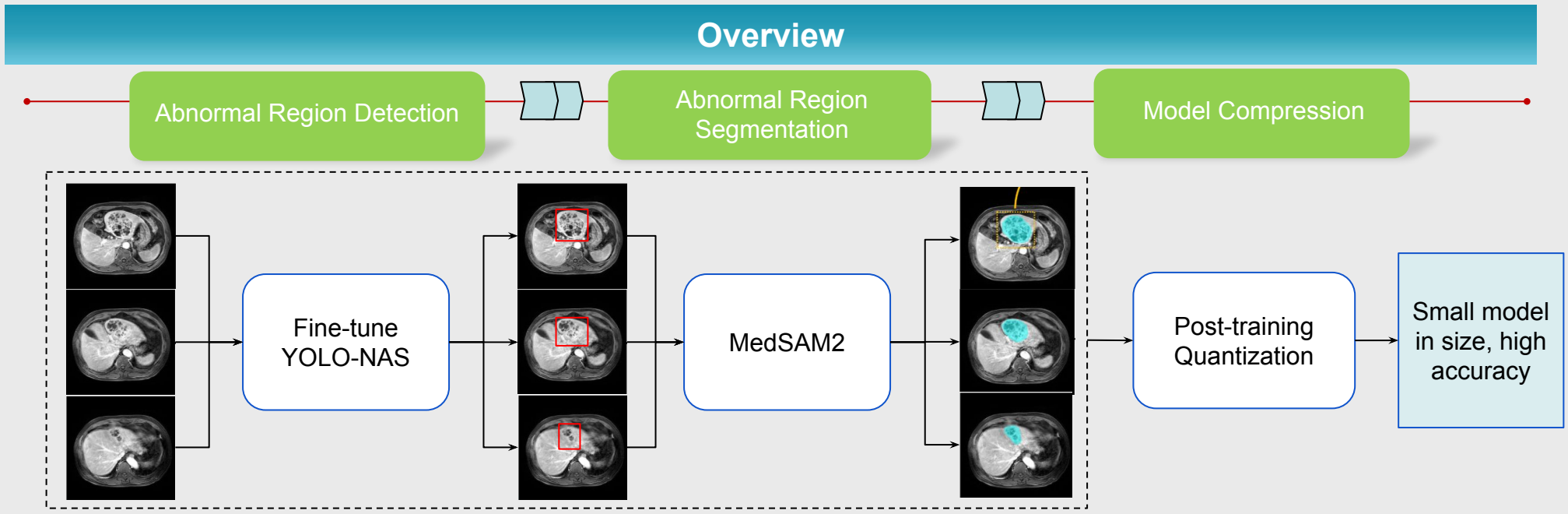
We introduce a pipeline to process and segment medical images and videos, in which we have:

- Proposed automated medical image segmentation pipeline.
- Applied quantization to reduce model size.
- Evaluated several medical image segmentation methods.

Why ?

- MedSAM2 offer high accuracy but rely on manually provided prompts (bounding boxes).
- Existing models are large in size.

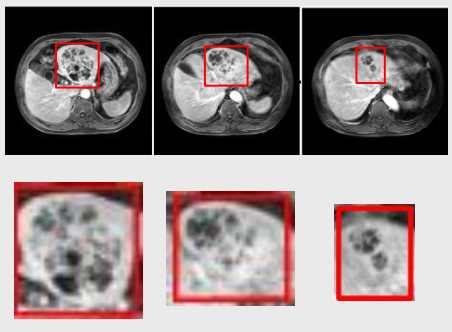
Overview



Description

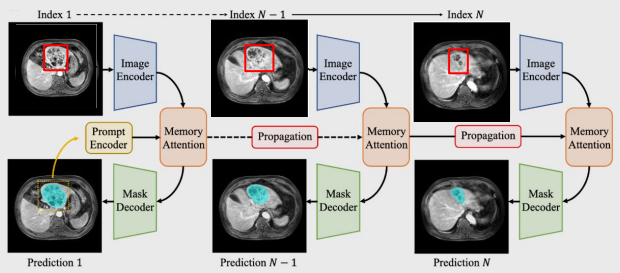
1. Abnormal Region Detection

- The YOLO-NAS model was trained on preprocessed medical images to detect suspected pathological regions using bounding boxes. Training was performed on an RTX 3080 GPU with 10GB VRAM over 100 epochs, using a batch size of 4 and a learning rate of 1e-3.
- The output consists of a list of bounding boxes [x,y,width,height] along with their corresponding confidence scores.



2. Abnormal Region Segmentation

- From the bounding boxes output by YOLO-NAS, the MedSAM2 model (a version of SAM2 fine-tuned on medical data) uses this information as prompts to perform detailed segmentation of the pathological regions.
- The predicted segmentation is then compared to the original manual prompt to evaluate consistency.
- The output is a segmentation mask with the same resolution as the input image (640×640).



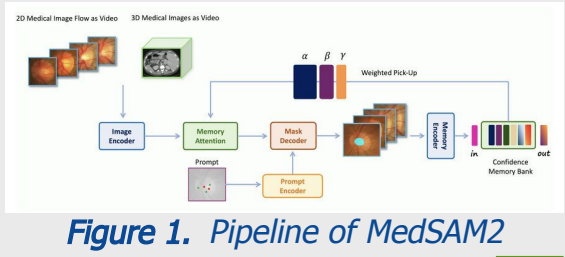


Figure 1. *Pipeline of MedSAM2*

3. Model Compression

- Post-Training Quantization (INT8) is used to reduce model size and resource consumption while maintaining high accuracy.

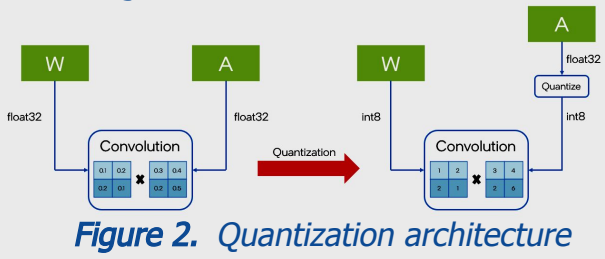
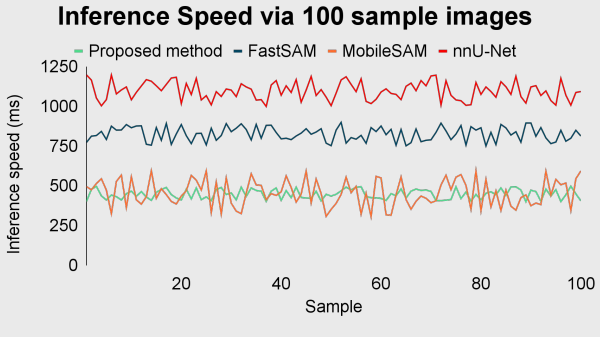


Figure 2. *Quantization architecture*



NII

Thi Nguyen Ngoc Lan – University of Information Technology, HCMC
TEL: +843 8 213 2704 Email: thinnl.19@grad.uit.edu.vn