



Segment Anything In Medical Images

Minwoo Jung

mai-lab.net, *Medical Artificial Intelligence Laboratory*

Electrical and Electronic Engineering

Yonsei University



Medical Artificial
Intelligence Laboratory
At Yonsei University



Segment anything in medical images

Received: 24 October 2023

Accepted: 5 January 2024

Published online: 22 January 2024

 Check for updates

Jun Ma^{1,2,3}, Yuting He⁴, Feifei Li¹, Lin Han⁵, Chenyu You⁶ &
Bo Wang^{1,2,3,7,8} ✉

Medical image segmentation is a critical component in clinical practice, facilitating accurate diagnosis, treatment planning, and disease monitoring. However, existing methods, often tailored to specific modalities or disease types, lack generalizability across the diverse spectrum of medical image segmentation tasks. Here we present MedSAM, a foundation model designed for bridging this gap by enabling universal medical image segmentation. The model is developed on a large-scale medical image dataset with 1,570,263 image-mask pairs, covering 10 imaging modalities and over 30 cancer types. We conduct a comprehensive evaluation on 86 internal validation tasks and 60 external validation tasks, demonstrating better accuracy and robustness than modality-wise specialist models. By delivering accurate and efficient segmentation across a wide spectrum of tasks, MedSAM holds significant potential to expedite the evolution of diagnostic tools and the personalization of treatment plans.

Segmentation is a fundamental task in medical imaging analysis, which involves identifying and delineating regions of interest (ROI) in various medical images, such as organs, lesions, and tissues¹. Accurate segmentation is essential for many clinical applications, including disease diagnosis, treatment planning, and monitoring of disease progression^{2,3}. Manual segmentation has long been the gold standard for delineating anatomical structures and pathological regions, but

field of natural image segmentation have witnessed the emergence of segmentation foundation models, such as segment anything model (SAM)⁴ and Segment Everything Everywhere with Multi-modal prompts all at once⁵, showcasing remarkable versatility and performance across various segmentation tasks.

There is a growing demand for universal models in medical image segmentation: models that can be trained once and then applied to a

Segmentation

Med-SAM

Background & Goal

▶ Background

- Applicability of the segmentation foundation models to medical image segmentation remains limited due to the significant differences between natural images and medical images

▶ Goal

- Aim to fulfill the role of a foundation model for universal medical image segmentation

Method

► Data curation

- Curated publicly available medical image segmentation datasets with 1,570,263 medical image mask pairs

► Pre-processing

- Converted the 3D images to the widely used NifTI format & 2D images to the png format
- Performed intensity normalization across all images
- All images were resized to a uniform size of 1024 X 1024 X 3 (Bicubic interpolation for resizing images & Nearest-neighbor interpolation for resizing masks)

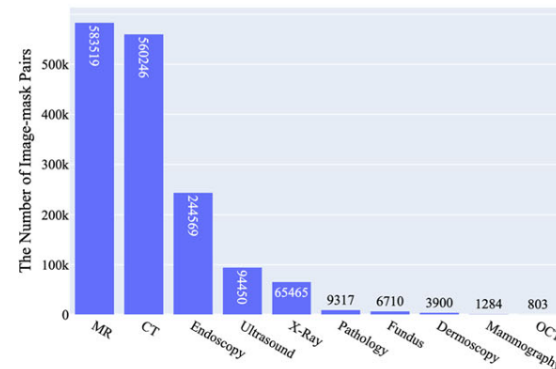


Fig 1. Overview of the modality distribution in the dataset

Method

► Network Architecture

- Vision Transformer (ViT)-based image encoder
- Prompt encoder for integrating user interactions(bounding boxes) → freeze
- Mask decoder that generated segmentation results and confidence scores

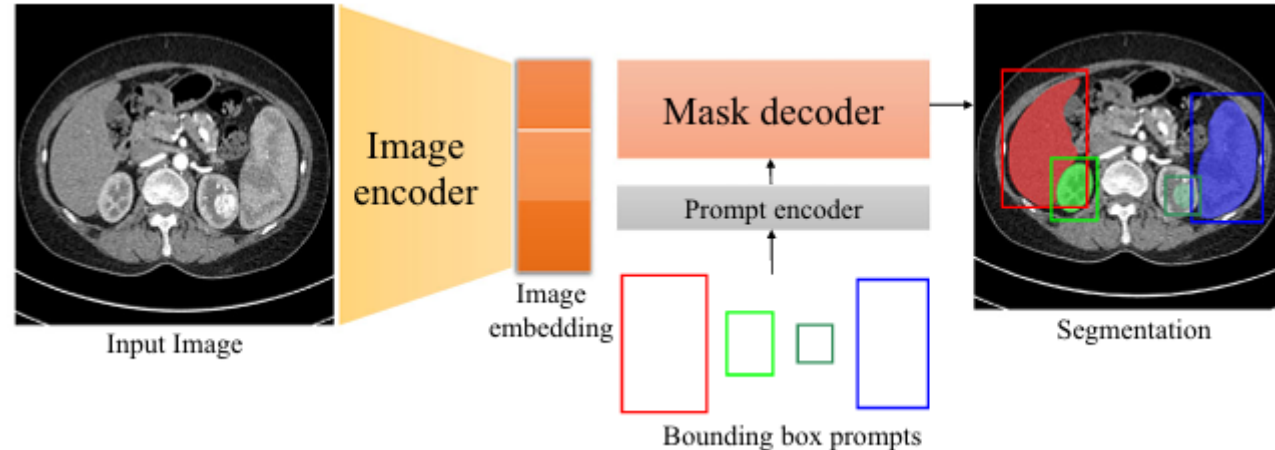


Fig 2. Overview of Med-SAM

Loss Functions

▶ Dice loss

- A loss function that is widely used in semantic segmentation where data has imbalanced characteristics

▶ Binary cross-entropy loss

- Measures the discrepancy between the true label and the probability that the two classes belong to as predicted by the model

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [g_i \log s_i + (1 - g_i) \log(1 - s_i)]$$

Eq 1. Binary cross-entropy loss

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N g_i s_i}{\sum_{i=1}^N (g_i)^2 + \sum_{i=1}^N (s_i)^2}$$

Eq 2. Dice loss

Evaluation Metrics

▶ Dice Similarity Coefficient

- Aiming to evaluate the region overlap between expert annotation masks and segmentation results

▶ Normalized Surface Distance

- Aiming to evaluate the boundary consensus between annotation masks and segmentation results(τ as 2)

$$\text{DSC}(G, S) = \frac{2|G \cap S|}{|G| + |S|}$$

Eq 3. Definition of DSC

$$\text{NSD}(G, S) = \frac{|\partial G \cap B_{\partial S}^{(\tau)}| + |\partial S \cap B_{\partial G}^{(\tau)}|}{|\partial G| + |\partial S|}$$

Eq 4. Definition of NSD

Results

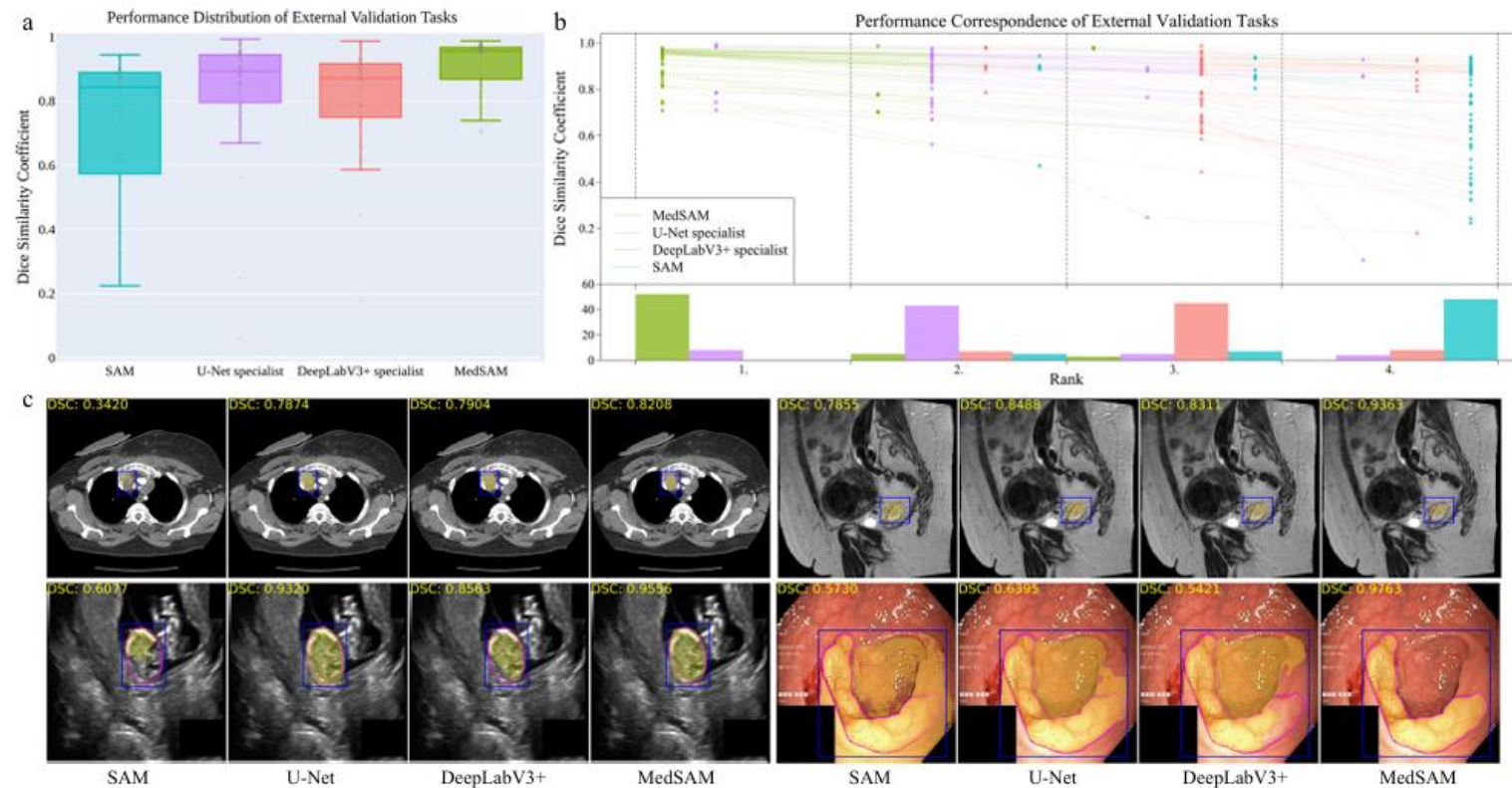


Fig 3. Quantitative and Qualitative evaluation results on the external validation set

Results

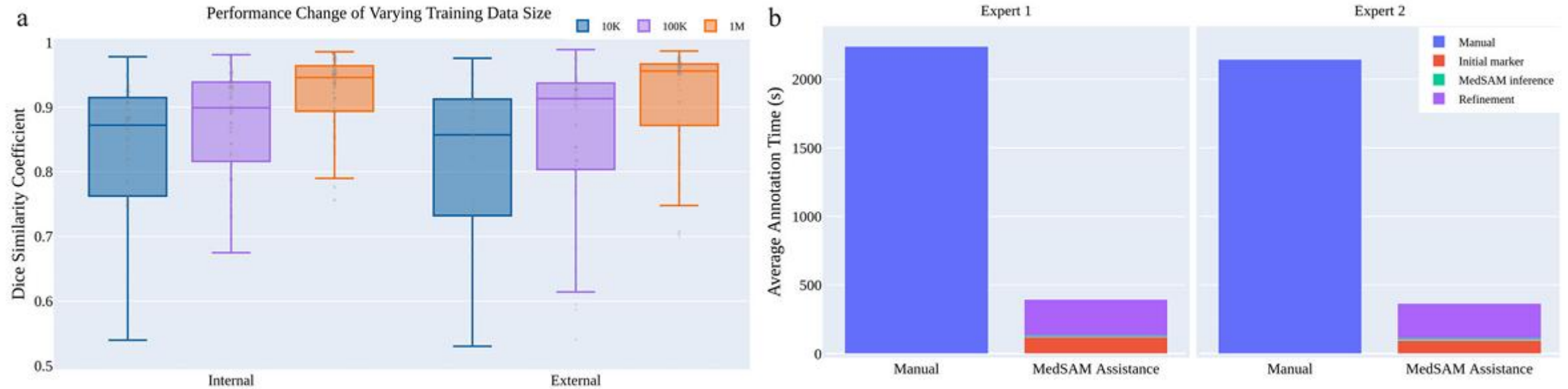


Fig 4. Effect of training dataset size and a user study of tumor annotation efficiency

Implications & Limitations

► Implications

- Introduce a deep learning-powered foundation model designed for the segmentation of a wide array of anatomical structures and lesions across diverse medical imaging modalities

► Limitations

- Modality imbalance in the training set(CT, MRI, and endoscopy images dominating the dataset)
- Difficulty in the segmentation of vessel-like branching structures because the bbox prompt can be ambiguous