Advanced Deep Learning 2024
Assignment 4

Dengke Chen

June 14, 2024

1 Segment Anything

1.1 1.Answer:

The SAM is designed to perform segmentation tasks on a wide variety of images. It has been trained on a vast and diverse dataset containing over 1 billion masks on 11 million images. The model's key strength lies in its generalization capability, allowing it to segment objects in images it has never seen before with minimal user input.

Training of SAM – The SAM model was trained using a large-scale dataset that encompasses various image types and segmentation scenarios. The training process focused on: Diverse Data: The model was exposed to a vast array of images, including those from natural scenes, objects, people, and animals, to enhance its ability to generalize. Interactive Segmentation: SAM can take user inputs in the form of points, boxes, or masks to refine its segmentation results, making it highly flexible and interactive..

Evaluation Results–The evaluation of SAM demonstrated its robustness and flexibility in handling a variety of segmentation tasks, outperforming many existing models in general benchmarks. The key metrics for evaluation included: Mask Quality: The precision and accuracy of the segmentation masks produced. Generalization Ability: The model's performance on unseen and diverse datasets..

Application to Medical X-ray Segmentation. Medical x-ray images present a unique challenge for segmentation models due to their specialized nature and the subtle differences in the anatomical structures they depict. These images are grayscale, lack the diversity seen in natural images, and often require domain-specific knowledge to interpret accurately. Given SAM's training background and evaluation performance, there are several points to consider regarding its expected performance on medical x-ray segmentation tasks:.

Domain Differences: Medical x-ray images are significantly different from the types of images SAM was primarily trained on. X-rays are grayscale, have lower contrast, and contain specific anatomical structures that require expert knowledge to interpret correctly. The features and patterns in medical images are quite distinct from natural images, which might pose a challenge for a model trained on diverse but non-medical data..

Generalizability: Given SAM's training on a highly diverse dataset, it is expected to have some level of adaptability to new domains, including medical imaging. The model's ability to segment based on various prompts (e.g., points, boxes) can potentially be advantageous in identifying anatomical regions within x-rays, provided the prompts are well-defined..

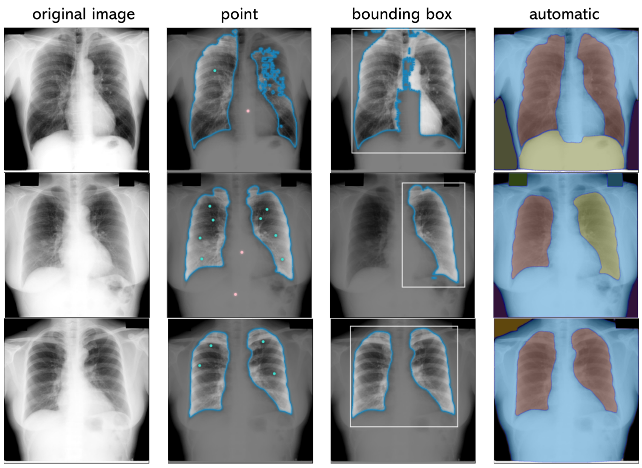
Limitations: Despite its versatility, SAM might struggle with the nuanced and subtle variations present in medical images that require domain-specific knowledge. The segmentation of medical x-rays often demands high precision, and the general-purpose nature of SAM might not match the accuracy of models specifically trained on medical datasets..

SAM's robustness to various segmentation tasks suggests that it could potentially provide reasonable segmentation masks for x-rays, especially with well-defined prompts. The lack of domain-specific training data in SAM's development means it may not capture the fine details necessary for precise medical diagnosis without additional tuning or prompts..

Based on the training and evaluation results presented in the SAM paper, it is reasonable to expect that SAM could perform reasonably well on medical x-ray segmentation tasks to some extent. However, it may not match the performance of specialized medical segmentation models due to the unique characteristics of medical images and the absence of such data in its training set.

1.2 2.Answer:

I used point, bounding boxes, and automatic mode to interactively segment the lungs on the three additional lung x-ray images on SAM's demo website and tried several times in each mode. Here are some examples from the run:



Points Prompting–Pros: Provides a high degree of control over the segmentation process. Allows for refinement by adding multiple points to improve the segmentation accuracy. Can help in guiding the model to specific areas of interest within the lung regions. Cons: Requires manual effort and precision in selecting points. The initial segmentation might be less accurate if points are not well-placed. Might require multiple iterations to achieve satisfactory results.

Bounding Box Prompting–Pros: Easier to specify larger areas of interest compared to point prompts. Faster to define regions, especially for larger anatomical structures like lungs. Can lead to more accurate initial segmentation as the bounding box provides a clearer context. Cons: Might include unwanted regions within the bounding box, requiring further refinement. Less precise than point prompting in delineating exact boundaries of complex shapes. Can sometimes over-segment or under-segment if the bounding box is not well-placed.

Automatic Mode–Pros: No manual effort required, making it the fastest method. Useful for a quick, rough segmentation to get an initial idea of the region. Can be a good starting point before applying more precise prompts. Cons: Likely less accurate than point or bounding box prompts for complex images like x-rays. May not handle the subtleties and fine details of the lung boundaries well. High chance of including non-relevant areas, leading to poor segmentation quality.

General Suitability of SAM for Lung X-ray Segmentation: SAM demonstrates flexibility and robustness, showing potential in adapting to the x-ray domain despite being trained on non-medical images. The interactive modes (point and bounding box) allow for fine-tuning and adjustments, which can enhance the segmentation quality for medical images. The model's generalization capability means it can be a useful tool for preliminary segmentation tasks in medical imaging. The lack of domain-specific training might lead to less precise segmentation compared to specialized medical models. The unique characteristics of x-ray images (e.g., grayscale, low contrast, specific anatomical structures) can pose challenges for a general-purpose model like SAM. High dependency on user input for accurate segmentation, especially in complex cases.

2 Prompt Engineering in Python

2.1 3. Answer

I selected the points prompt in this task, four foreground points ([75, 100], [80, 75], [180, 75], [185, 100]) and one background point ([135, 150]).

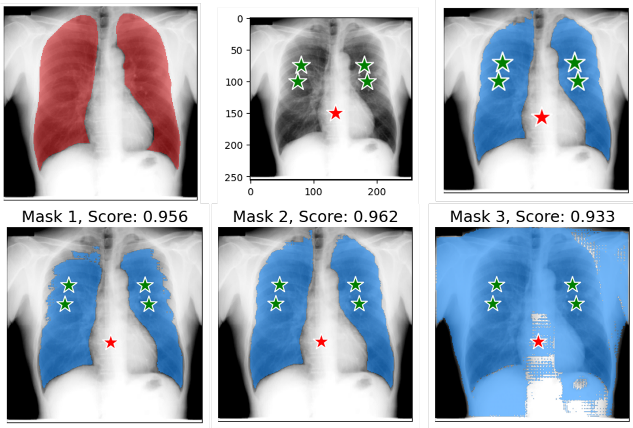


Figure 1: Results visualization

2.2 4. Answer

In order to use the best mask from the SAM model as input to further refine the segmentation results, I set the mask prediction steps to first get the mask with the highest model score by a preliminary prediction and then use the best mask as input for further predictions.

The mean of the F1 score: 0.9084, This score shows that the lung segmentation using the SAM model works very well and the model is able to accurately predict the lung region in most cases. The standard deviation of the F1 score: 0.0247, This standard deviation is low, indicating that the performance of the model on different images is relatively stable and the consistency of the prediction results is high.

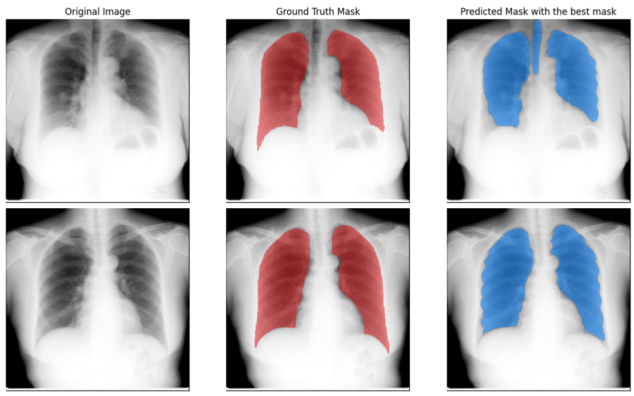


Figure 2: The results on two selected example images.

As can be seen from the image, the predicted mask of the SAM model is very close to the true mask, indicating that the model has high accuracy. The predicted mask covers most of the lung area, which has important clinical significance for the diagnosis and treatment of lung diseases. The prediction results for the two example images are consistent, indicating that the model has stable performance on different images. The model can adapt to different lung morphologies and image qualities and shows good robustness. In some image boundary regions, the predicted mask may be slightly biased. This can be improved by further optimizing the choice of cue points or using more cue points. Especially in the top and middle areas of the lungs, try adding more cue points to capture these details. Although there are errors in some detail areas, the overall performance is still very good and can effectively segment the lung area. Through these analyses, the feasibility and effectiveness of using the SAM model for lung X-ray image segmentation were confirmed.

3 Dynamic Prompting

3.1 5. Answer

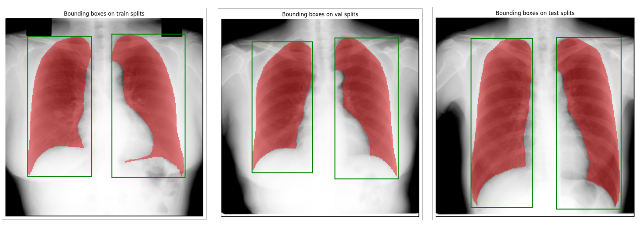


Figure 3: Examples–two bounding boxes from the ground-truth lung segmentation masks for the train, val and test splits.

Approach: For each lung mask, find contours (the boundaries of connected components). For each contour, compute the bounding rectangle that encloses the contour. This is done using OpenCV's cv2.boundingRect() function. Convert the bounding rectangle format from [x, y, w, h] to [x0, y0, x1, y1]. Implementation: The get_bounding_boxes() function takes a binary mask as input and returns bounding boxes for each connected component. The generate_bounding_boxes_for_dataset() function applies get_bounding_boxes() to each mask in the dataset (train, validation, and test sets).

3.2 8. Answer

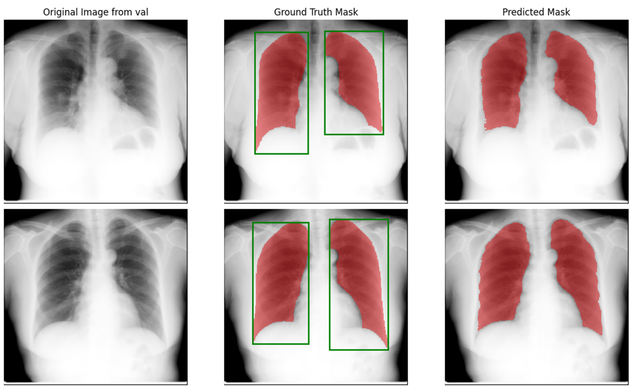


Figure 4: Two example segmentations from bounding box prompts method on val splits:

Point prompts: The mean of the F1 score: 0.9084,The standard deviation of the F1 score: 0.0247;

Bounding box prompts: The mean of the F1 score: 0.9177, The standard deviation of the F1 score: 0.0248.

In both prompts methods, I both used the best mask in the SAM model as the input to further refine the segmentation results, First, the mask with the highest score in the model is obtained through preliminary prediction, and then the best mask is used as the input for further prediction. I used 'Logical or' action (np.logical_or) in the bounding box prompt method to combine multiple prediction masks (left and right lungs) to get a comprehensive final mask that can be used to accurately evaluate and compare model performance. See the attachment of the code for specific implementation.

Mean F1 Score: The Mean F1 score for bounding box prompts (0.9177) is slightly higher than that for point prompts (0.9084). This indicates that bounding box prompts provide a marginal improvement in segmentation accuracy compared to point prompts. The improvement, though small, suggests that bounding boxes may help the SAM model to better localize and segment the lung regions by providing more precise spatial information.

Standard Deviation: The standard deviation of the F1 score for both methods is quite similar (0.0247 for point prompts and 0.0248 for bounding box prompts). This implies that the variability in segmentation performance across different images is almost the same for both methods. The low standard deviations in both cases suggest that both methods provide consistent segmentation performance across the validation dataset.

3.3 7. Answer

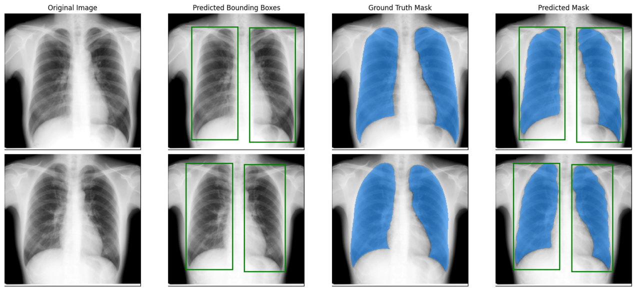


Figure 5: Visualize the results for two example images (YOLOv8+SAM)

In this task, I trained a YOLOv8 model for 250 epochs using the ground-truth bounding boxes generated in Task 5.

Mean IoU on the test set: 0.9364, The mean of the F1 score: 0.9233, The standard deviation of the F1 score: 0.0280.

U-net from Assignment 1: The mean of the F1 score: 0.6073; The standard deviation of the F1 score: 0.0901.

Efficacy of YOLOv8 + SAM: YOLOv8 effectively identifies the regions of interest (lungs) with high precision, providing accurate bounding boxes. SAM, when prompted with accurate bounding boxes, can produce high-quality segmentation masks that closely match the ground-truth. The approach shows consistent performance across different images, as indicated by the low standard deviation in F1 scores. The combination of YOLOv8 for detecting bounding boxes and SAM for segmenting the lungs has proven to be highly effective: High Mean IoU (0.9369) indicates that the YOLOv8 model accurately predicts bounding boxes that closely match the ground-truth boxes. High Mean F1 Score (0.9233) reflects a high degree of overlap between the predicted and ground-truth masks, indicating that SAM, when prompted with accurate bounding boxes, performs very well. Low Standard Deviation (0.0280) suggests that the performance of the YOLOv8 + SAM approach is consistent across different test images.

Comparison with Points Prompts: The mean F1 score for the points prompts method (0.9084) is slightly lower than that of the YOLOv8 + SAM approach (0.9233). This suggests that using bounding boxes provides a more precise and reliable prompt for SAM compared to points prompts. The standard deviation for points prompts (0.0247) is similar to that of YOLOv8 + SAM (0.0280), indicating that both methods have consistent performance. However, the slightly higher consistency of points prompts may be due to the simplicity of the prompt, though it doesn't capture the lung regions as accurately as bounding boxes. Comparison with U-Net from Assignment 1: The U-Net model's mean F1 score (0.6073) is significantly lower than both the YOLOv8 + SAM approach and the points prompts method. This indicates that the U-Net model struggles to achieve the same level of segmentation accuracy. The standard deviation of the U-Net model (0.0901) is much higher compared to both the YOLOv8 + SAM approach and points prompts method. This suggests that the U-Net model's performance is less consistent, with greater variability across different test images.

The results clearly demonstrate the superiority of the YOLOv8 + SAM approach over both the points prompts method and the U-Net model from Assignment 1. The use of an object detection model to provide precise bounding boxes as prompts for SAM significantly enhances the segmentation accuracy.