# Test-Time Adaptation with SaLIP: A Cascade of SAM and CLIP for Zero-shot Medical Image Segmentation

Supplementary Material

## 7. Qualitative Results

In this section, we comprehensively present the qualitative results of SaLIP to demonstrate its effectiveness. X-ray Masks and Labels [19] dataset has two regions of interest i.e. left and right lungs. Fig. 7 shows that SaLIP yields correct predictions for both lungs in a zero-shot manner without any domain expertise or prompt engineering.

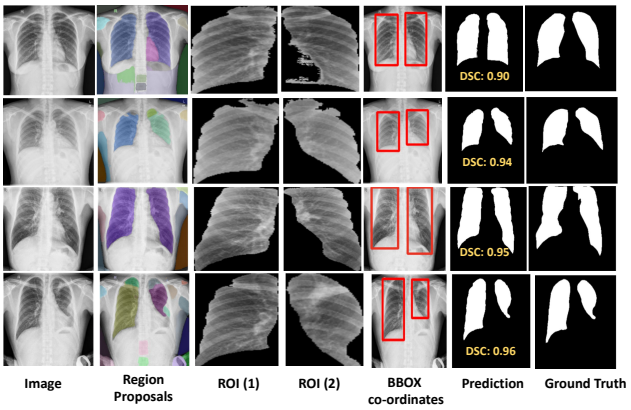


Figure 7. SaLIP qualitative results on X-ray labels and masks [19] dataset.

Fig. 8 demonstrates that SaLIP accurately predicts fetal head positions with zero-shot learning. Despite the variation in fetal head size, SaLIP accurately segments them.

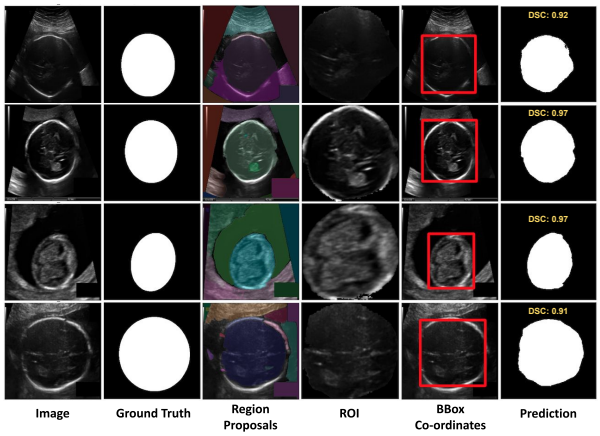


Figure 8. SaLIP Qualitative results on HC18 [28] dataset.

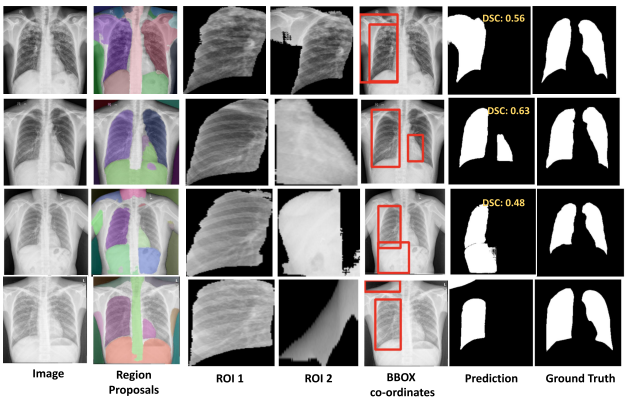


Figure 9. Visualization of failure cases on X-ray labels and masks dataset [19]. First two rows: SAM generates masks for both the left and right lung, but CLIP retrieves only the right lung mask correctly. While rows three and four show cases where SAM does not generate a mask for the left lung, eventually CLIP does not recognize the correct crop for the left lung in such cases.

### 7.1. Limitations

In this section, we highlight the limitations of our method. We report the limitations both at SAM and CLIP levels as shown in Fig. 9. The first row shows the case multiple mask generations by SAM leads to the misclassification of ROIs by CLIP. Specifically, CLIP erroneously predicts the SAM region mask for the lungs, thereby impacting the final prediction. The second row shows the case where SAM correctly generates masks for both ROIs, however, CLIP misclassifies the crop for the left lung. The third and fourth rows illustrate scenarios where SAM fails to predict the ROI for the left lung, consequently leading to CLIP retrieving the incorrect crop region. We also report the limitations in the fetal head dataset as shown in Fig. 10. The first and second rows show instances where $SAM_{EM}$ fails to generate the relevant mask for the fetal head ROIs, which leads to incorrect ROI recognition by CLIP. The third and fourth rows showcase which $SAM_{EM}$ generates multiple masks for fetal head ROIs, resulting in miss-identification by CLIP. Such misidentifications lead to inaccurate predictions for the fetal head, consequently lowering the DSC.

## 8. Visual prompting

In this section, we provide details of different visual prompts around all the SAM's generated masks to improve lung segmentation. We applied three distinct visual prompts

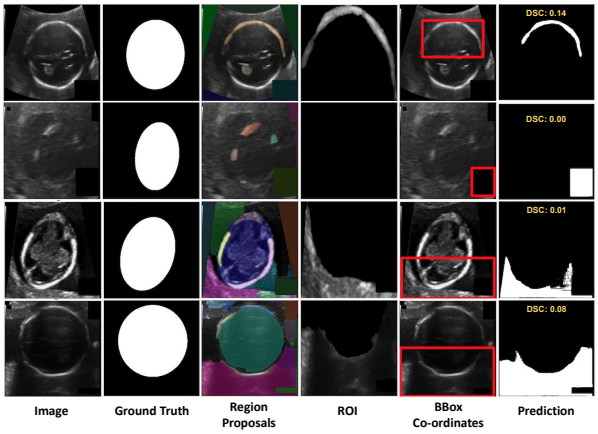


Figure 10. Visualization of failure cases for HC18 [28]. First two rows: $SAM_{EM}$ does not generate a mask for the fetal head; therefore, CLIP fails to identify the correct crop in these instances. While row three and four show instances where $SAM_{EM}$ successfully generates a mask for the fetal head, whereas CLIP does not retrieve the correct mask.

(VPTs) to all SAM-generated masks on the input image: a red bounding box marker, contour delineation, and a reverse gray box highlighting the masked area while blurring the rest of the image [31]. However, VPTs do not improve performance for medical imaging datasets as depicted in Fig. 11. While in certain instances a single ROI is predicted correctly, the overall recognition is inconsistent across the dataset, indicating scalability issues.

## 9. GPT-3.5 Prompts

### 9.1. Lungs

In this study, we use a two-step prompting strategy to identify relevant characteristics that aid in the detection of crops having the desired organ, following the approach [15].

Example of the first prompt for the lung dataset:

*I am creating a zero-shot lung recognition system in a chest x-ray. Write me 20 attributes to describe the lungs in the chest x-ray.*

Creating a zero-shot lung recognition system involves identifying and describing features that can uniquely identify and describe the lungs in a chest X-ray. Here are 20 attributes that could be relevant for such a task:

*Lung Size: The overall size of the lungs, noting whether they are expanded, normal, or show signs of constriction or atelectasis. Lung Shape: The general shape of the lungs, including any deviations from the typical butterfly shape. Opacity: The degree of darkness observed in the lung fields, indicating the presence of fluids, masses, or other abnormalities. Texture: The internal texture of the lung fields,*

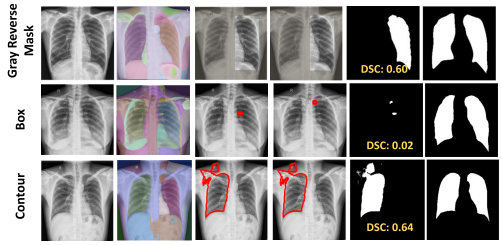


Figure 11. VPT results on X-ray labels and masks [19] dataset: Although SAM successfully generates masks for both lungs, VPTs do not facilitate CLIP in accurately retrieving ROIs.

*whether it appears smooth, granular, or shows specific patterns such as reticular, nodular, or cystic formations. Symmetry: The comparative symmetry between the right and left lungs, noting any significant differences. Lung Borders: The clarity and integrity of the lung borders, particularly against the heart and diaphragm. Hilum Position and Appearance: The location and appearance of the lung hila, which may be displaced or show signs of enlargement. Pleural Line: The visibility and condition of the pleural line, including any thickening, plaques, or pneumothorax (air in the pleural space). Costophrenic Angles: The sharpness and clarity of the costophrenic angles, noting any blunting which may suggest pleural effusion. Vascular Markings: The visibility and pattern of vascular markings within the lung fields. Air Bronchograms: The presence of air-filled bronchi (tubular structures) against a background of opaque lung, indicating consolidation. Tracheal Position: The position of the trachea, noting if it is centered or displaced. Diaphragm Appearance: The contour and position of the diaphragm, including any signs of elevation or abnormal curvature. Cardiac Silhouette: The size and shape of the heart shadow as it relates to the lungs. Mediastinal Width: The width of the mediastinum, noting any widening that may suggest mediastinal masses or lymphadenopathy. Interstitial Markings: The presence of lines and markings that may indicate interstitial lung disease. Lung Apex: The condition and appearance of the lung apex, looking for masses, fibrosis, or other abnormalities. Presence of Calcifications: The presence and distribution of calcifications within the lung fields, which may suggest old infections or tumors. Pleural Effusion: The presence of fluid in the pleural space, visible as an area of increased opacity. Foreign Bodies or Devices: The presence of any foreign bodies, surgical clips, pacemaker leads, or other medical devices. These attributes are critical for a detailed analysis and recognition of lung health and disease in chest X-rays. Each attribute can help differentiate between normal lung anatomy and various pathologies, aiding in the development of a comprehensive zero-shot lung recognition system.*

Example of the second prompt for the lung dataset which

uses GPT-3.5's response for the first prompt as part of the prompt:

*Now write me sentence about each of these attributes with "lung" in the sentences and return the answer as a Python dictionary.*

*["The X-ray shows the shape of the lungs, whether they appear rounded or irregular.", "The X-ray indicates the relative size of the lungs within the chest cavity, providing insights into lung health.", "The X-ray helps assess the proper position of the lungs within the chest cavity.", "The X-ray reveals the density of the lung areas, indicating the presence of air or tissue within the lungs.", "The X-ray evaluates how clear the lung fields appear by assessing the transparency of the lungs.", "The X-ray shows the texture of the lungs, whether they appear smooth or have irregularities.", "The X-ray provides insights into the vascular pattern within the lungs, aiding in the assessment of lung circulation.", "The X-ray examines the borders of the lungs, assessing for normal anatomy or potential abnormalities.", "The X-ray checks the presence and integrity of lung fissures, the thin lines separating lung lobes.", "The X-ray evaluates structures at the lung hilum, such as lymph nodes and bronchi, for any abnormalities.", "The X-ray assesses the contour of the diaphragm where it meets the lungs, ensuring proper lung function.", "The X-ray checks the angles formed by the diaphragm and the chest wall, assessing for effusion or pneumothorax in the lungs.", "The X-ray examines the appearance and distribution of bronchi and blood vessels within the lungs.", "The X-ray detects fine linear or reticular opacities within the lung fields, indicating interstitial lung disease.", "The X-ray assesses the space between the lung and the chest wall for any effusion or pneumothorax.", "The X-ray evaluates the deviation of the trachea from its normal position, indicating potential mediastinal shift or mass effect on the lungs.", "The X-ray examines the appearance of ribs surrounding the lungs, checking for integrity and potential fractures.", "The X-ray shows the outline of the heart within the chest, which can influence the appearance of the lung fields.", "The X-ray assesses the structures within the mediastinum, providing insights into their impact on the adjacent lungs." ]*

## 9.2. Separate prompts for left and right lung

Example of first prompt for the lung dataset:

*I am creating a zero shot right and left lung recognition system in a chest x-ray. Write me 1o attributes to describe the left and right lungs in the chest x-ray.*

Attributes of the Left Lung:

["Position: Situated on the left side of the thoracic cavity." "Size: Slightly smaller than the right lung to accommodate the heart." "Lobes: Consists of two lobes - the superior and inferior lobes, separated by the oblique fissure." "Function: Main function is to exchange oxygen and carbon dioxide with the blood during respiration." "Medical Significance: Annotated images showing the left lung are essential in medical imaging for diagnosing various respiratory conditions." "Blood Circulation: Receives oxygenated blood from the pulmonary veins and sends deoxygenated blood to the pulmonary arteries." "pH Regulation: Plays a role in regulating blood pH through the removal of carbon dioxide." "Development: Undergoes morphological changes during development, including the formation of bronchial buds and bronchopulmonary segments." "Susceptibility: Susceptible to infections such as pneumonia and tuberculosis, which can cause inflammation and compromise respiratory function."]

Attributes of the Right Lung:

["Position: Located on the right side of the thoracic cavity." "Size: Slightly larger than the left lung due to the presence of the heart." "Lobes: Divided into three lobes - the superior, middle, and inferior lobes, separated by the horizontal and oblique fissures." "Function: Plays a crucial role in oxygenating the blood and removing carbon dioxide from the body." "Medical Significance: Medical annotations on images showing the right lung aid in diagnosing diseases such as pneumonia, lung cancer, and tuberculosis." "Blood Circulation: Receives deoxygenated blood from the systemic circulation via the pulmonary arteries and returns oxygenated blood to the heart through the pulmonary veins." "Aspiration Risk: More prone to injury from aspiration due to its more vertical orientation of the main bronchus compared to the left lung." "Diseases: Susceptible to diseases such as emphysema, bronchitis, and pulmonary embolism, which can impair respiratory function." "Development: Undergoes branching morphogenesis during embryonic development to form the bronchial tree and alveoli." "Gas Exchange: Has a larger capacity for gas exchange compared to the left lung due to its larger surface area."]

Example of the second prompt for the lung dataset which uses GPT-3.5's response for the first prompt as part of the prompt:

*Now write me a sentence about each of these attributes with "left lung" and "right lung" in the sentences of each respective class and return the answer as a Python dictionary.*

{ "left lung": [ "A left lung is situated on the left side of the thoracic cavity.", "It is slightly smaller than the right lung to accommodate the heart.", "The left lung consists of two lobes: the superior and inferior lobes, which are separated by the oblique fissure.", "Its main function is to exchange oxygen and carbon dioxide with the blood during

respiration.", "Annotated images showing the left lung are essential in medical imaging for diagnosing various respiratory conditions.", "The left lung receives oxygenated blood from the pulmonary veins and sends deoxygenated blood to the pulmonary arteries.", "In addition to its respiratory function, the left lung also plays a role in regulating blood pH through the removal of carbon dioxide.", "Certain diseases, such as pulmonary edema and pulmonary fibrosis, can affect the left lung's ability to function properly.", "During development, the left lung undergoes morphological changes, including the formation of bronchial buds and the establishment of bronchopulmonary segments.", "The left lung is susceptible to infections, such as pneumonia and tuberculosis, which can cause inflammation and compromise its respiratory function." ],
"right lung": [ "A right lung is located on the right side of the thoracic cavity.", "It is slightly larger than the left lung due to the presence of the heart.", "The right lung plays a crucial role in oxygenating the blood and removing carbon dioxide from the body.", "Medical annotations on images showing the right lung aid in diagnosing diseases such as pneumonia, lung cancer, and tuberculosis.", "The right lung is divided into three lobes: the superior, middle, and inferior lobes, separated by the horizontal and oblique fissures.", "It receives deoxygenated blood from the systemic circulation via the pulmonary arteries and returns oxygenated blood to the heart through the pulmonary veins.", "The right lung is more prone to injury from aspiration due to its more vertical orientation of the main bronchus compared to the left lung.", "Diseases affecting the right lung include emphysema, bronchitis, and pulmonary embolism, which can impair its respiratory function.", "During embryonic development, the right lung undergoes branching morphogenesis to form the bronchial tree and alveoli.", "The right lung has a larger capacity for gas exchange compared to the left lung due to its larger surface area." ] }

### 9.3. Fetal Head

Example of the first prompt for the fetal head segmentation:

*I am creating a fetal head recognition system from the transthalamic and transventricular ultrasound planes; write me 20 attributes to describe the fetal head.*

*Head Circumference: Measurement of the circumference of the fetal head, indicating overall size. Biparietal Diameter (BPD): Distance between the two parietal bones, a key indicator of fetal head size and gestational age. Head Shape: Describing the overall shape of the fetal head, which can vary from round to more elongated. Skull Density: Density of the fetal skull bones, influencing ultrasound image contrast. Fontanelle Size: Size and appearance of the fontanelles, indicating the stage of skull ossification. Cerebral Ventricles: Presence and size of cerebral ventricles within the fetal brain. Cranial Bone Integrity: Assessment of the integrity of the cranial bones, looking for any fractures or abnormalities. Midline Structures: Identification of midline structures such as the falx cerebri and cavum septi pellucidi. Cerebellar Vermis: Presence and size of the cerebellar vermis, important for assessing brain development. Third Ventricle: Measurement and appearance of the third ventricle, indicative of normal brain development. Cerebral Hemispheres: Assessment of the size and symmetry of the fetal cerebral hemispheres. Occipitofrontal Diameter (OFD): Measurement from the occipital bone to the frontal bone, providing another indicator of fetal head size. Nuchal Fold Thickness: Thickness of the nuchal fold, which can indicate chromosomal abnormalities. Facial Features: Identification and characterization of facial features such as the nose, lips, and orbits. Fetal Neck: Assessment of the fetal neck for any abnormalities or masses. Eyes and Eye Orbit: Identification of the eyes and assessment of the eye orbits for normal development. Temporal Bone Anatomy: Examination of the temporal bones for any abnormalities or ossification issues. Brain Symmetry: Assessment of symmetry in brain structures, including the hemispheres and ventricles. Brain Parenchyma Texture: Evaluation of the texture and appearance of the fetal brain parenchyma. Fetal Presentation: Determination of the fetal presentation (cephalic, breech, transverse), which can impact the segmentation process.*

Example of the second prompt for the fetal head dataset which uses GPT-3.5's response for the first prompt as part of the prompt:

*Now write me sentences about each prompt of these attributes with "fetal head" in the sentences, and return the answer as a python dictionary.*

The response of the second prompt for fetal head as an example:

*["Fetal head circumference is a crucial metric measured to assess overall growth and development.", "Biparietal diameter (BPD) provides insight into fetal head size and gestational age.", "The fetal head's shape varies, ranging from round to more elongated forms.", "Skull density contributes to the clarity and contrast of structures within the fetal head.", "Fontanelle size offers information about the fetal head's ossification stages and cranial development.", "Cerebral ventricles within the fetal head are assessed to ensure normalcy and appropriate size.", "Cranial bone integrity is essential for the protection and development of the fetal head.", "Midline structures, such as the falx cerebri, contribute to the overall architecture of the fetal head.",*

*"The cerebellar vermis, located within the fetal head, influences motor coordination and balance.", "The third ventricle within the fetal head is examined to evaluate central nervous system development.", "Symmetry of the cerebral hemispheres is crucial for the overall morphology of the fetal head.", "Occipitofrontal diameter (OFD) measurement aids in assessing the proportionality of the fetal head.", "Nuchal fold thickness measurement provides insights into potential abnormalities affecting the fetal head.", "Facial features within the fetal head region contribute to the overall assessment of anatomical development.", "Evaluation of the fetal neck ensures proper alignment and structural integrity of the fetal head and neck.", "Eyes and eye orbits within the fetal head region are assessed to evaluate ocular development.", "Temporal bone anatomy within the fetal head is evaluated to detect any abnormalities or ossification issues.", "Brain symmetry within the fetal head is assessed to ensure proper development and alignment of structures.", "Texture and appearance of the brain parenchyma provide insights into overall health and development within the fetal head.", "Fetal presentation, whether cephalic, breech, or transverse, influences the positioning and segmentation of the fetal head during imaging." ]*