Unlocking SAM's Potential in Medicine

A Literature Review on Fine-Tuning the Segment Anything Model for Medical Image Segmentation

Cassandra Wallace
Computer Science
University of Cape Town
Cape Town, Western Cape, South Africa
WLLCAS004@myuct.ac.za

ABSTRACT

The Segment Anything Model's (SAM) potential for medical image segmentation, crucial for improved healthcare, necessitates bridging the gap between its 2D architecture and 3D medical data complexities. This review analyses fine-tuning approaches, highlighting full fine-tuning and Parameter-Efficient Fine-Tuning (PEFT) with MedSAM as a benchmark. However, a hybrid PEFT and framework modification approach is suggested for practicality. Automatic prompting methods like EviPrompt show promise for future scalability. Medical imaging modality significantly impacts SAM, with well-defined structures (X-ray, MRI, CT) as a good starting point. Tailored approaches are needed for domains with greater complexity. This review emphasizes the need for further research on hybrid fine-tuning, automatic prompting, and domain-specific strategies to unlock SAM's full potential in medical image segmentation and advance healthcare.

KEYWORDS

Computer Vision, Machine Learning, Deep Learning, Image Segmentation, Medical Imaging, Medical Image Segmentation

1 INTRODUCTION & MOTIVATION

Medical image segmentation plays a pivotal role in modern healthcare, enabling accurate diagnosis, treatment planning, and disease monitoring [33]. However, traditional segmentation methods, while established, can be time-consuming and prone to human error [21].

Deep learning offers a promising alternative, as it can be trained directly from data to achieve complex and accurate segmentation, reducing manual intervention. The Segment Anything Model (SAM), a recent image segmentation model from 2023, produced by Meta AI [25], demonstrates impressive object segmentation capabilities, even without specific object-based training. However, it was designed for natural images and requires adaptation for the specific challenges of medical images, which is introduced due to inherent image differences like 3D data and noise levels.

This review focuses on recent advancements in "fine-tuning" SAM for medical image segmentation. Fine-tuning leverages a pre-trained model's knowledge (like SAM's) and adapts it to a new domain by modifying specific parts of its architecture or training process [13]. This holds immense potential for improving SAM's effectiveness in medical image segmentation, a domain with diverse image characteristics.

We aim to critically analyse existing research on different fine-tuning strategies for SAM in medical image segmentation, like those used in pioneering works such as MedSAM [33] and AdaptiveSAM [34], alongside more recent advancements like ProMISe [28]. By exploring the advantages and limitations of these approaches, we aim to inform the development of our project, SAMSeg. This review will bridge the gap in our current knowledge and lay the foundation for further research in this domain.

2 DEEP LEARNING FOR MEDICAL IMAGE SEGMENTATION: PRE-SAM

This section explores earlier established image segmentation approaches and their role in medicine before the emergence of SAM.

While deep learning has become the dominant force in medical image segmentation, it is valuable to understand prior methods that laid the foundation which it builds upon. While classical techniques, such as thresholding, offered computational efficiency, their performance suffered due to noise and complex shapes [22]. Similarly, machine learning approaches, such as clustering-based algorithms, provided a foundation for later advancements, but were limited in handling intricate image features [15].

Deep learning architectures, often consisting of an encoder-decoder structure with millions of parameters, have revolutionised image segmentation due to their ability to learn complex relationships within image data. This has led to remarkable performance in various tasks, including surgical scene segmentation [34]. However, deep learning techniques can be inherently domain-specific, requiring vast amounts of labelled medical data, which can be scarce due to privacy concerns and annotation complexity. In addition, their complexity can lead to overfitting and challenges in handling inherent variability in medical images.

Despite these limitations, deep learning remains a powerful tool for medical image segmentation. By acknowledging these challenges and employing appropriate strategies like data augmentation and transfer learning, researchers are continuously working to improve the generalisability and robustness of deep learning models in this domain.

2.1 U-Net

Introduced in 2015 by Ronneberger et al. [36], U-Net is a widely used Convolutional Neural Network (CNN) architecture designed for biomedical image segmentation. Its U-shaped structure consists of two main paths: a contracting path that captures the context of the input image, and a symmetrically expanding path for precise pixel localisation within the input image. This allows for accurate segmentation of complex shapes with less training data compared to previous state-of-the-art methods.

Despite its success, U-Net has limitations. Studies report struggles with class imbalances in medical image datasets, where some classes (e.g., rare diseases) have far fewer pixels compared to others (e.g., background) [34]. In addition, U-Net's generalisability on unseen datasets can be limited, requiring significant retraining for new applications [33].

Despite these drawbacks, U-Net remains a widely used and influential architecture in medical image segmentation. Researchers often use it as a baseline model for comparison with newer approaches. This has led to the development of an abundance of U-Net variations in the medical realm, such as nnU-Net.

2.1.1 nnU-Net. Building upon U-Net [36] and addressing its limitations, nnU-Net, produced by Isensee et al. [21], offers a robust and self-adapting framework for both 2D and 3D medical image segmentation. It incorporates features like leaky ReLU activation and strided convolutions for downsampling, alongside extensive data augmentation strategies. Notably, it can automatically configure itself for different tasks, eliminating the need for extensive architecture design. This, combined with automated pre-processing and post-processing pipelines, significantly contributed to its success.

nnU-Net achieved remarkable results in numerous medical image segmentation challenges, ranking first in accuracy in the 2018 Medical Decathlon Challenge [3]. This established nnU-Net as a new state-of-the-art method and foundational framework for further advancements in medical image segmentation.

2.2 Transformer-Based: Visual Transformers (ViTs)

Before 2021, medical image segmentation relied heavily on Convolutional Neural Network (CNN) architectures. On the other hand, transformer-based architectures were primarily only used in natural language processing. However, in 2021, Dosovitskiy et al. [12] introduced Visual Transformers (ViTs), a pure transformer-based architecture that achieved impressive performance on image segmentation tasks, even outperforming state-of-the-art CNNs, particularly when pre-trained on very large datasets.

ViTs offer advantages over CNNs, including lower computational cost, which can be crucial when dealing with large medical image datasets. In addition, they have weaker inductive image biases, i.e. they make significantly fewer image-structure assumptions, making them adaptable to diverse medical image characteristics. They achieve high accuracy even with large datasets due to their core component, the Transformer encoder, which leverages both linear and positional embeddings to capture both local and global image features.

While ViTs offer advantages, they are a relatively new approach in medical image segmentation. This means established best practices for medical image segmentation tasks might still be under development for ViT architectures. In addition, for highly complex segmentation tasks, ViTs might require more training data compared to CNNs.

ViT has various sizes, specifically the ViT-Base (with 12 transformational layers and 91 million parameters), the VIT-Large, and the VIT-Huge (with 32 transformational layers and 636 million parameters). While ViT-Base is the most common, the larger models offer increased capacity but require significantly more training data.

The introduction of ViTs opened doors for exploring Transformer-based architectures in medical image segmentation, leading to improved performance and broader applicability.

3 CURRENT ERA: THE SAM REVOLUTION

On April 5, 2023, Meta AI introduced the Segment Anything Model (SAM) [25]. SAM gained significant attention for its ability to generate accurate object segmentations through both automation and user interaction. This research marked a shift towards prompt-driven models, opening doors to explore previously unknown capabilities in image segmentation.

However, SAM was designed for natural images. While researchers are exploring its potential in various domains beyond its original training data, limitations emerge in specific use cases. For instance, studies on camouflaged object detection [39] showed promise for natural objects like spiders but struggle with more complex camouflage. Similarly, investigations in agriculture [23] revealed success in pest and disease monitoring, but limitations in whole-pest detection and crop segmentation, likely due to inherent biases in the training data. Likewise, studies in manufacturing [23] suggest its effectiveness for prominent anomalies but highlight the need for expert guidance for subtle defect detection. Finally, research on building and road extraction [23] indicates proficiency with regular-shaped buildings but challenges with smaller structures and diverse road characteristics. These use cases highlight both the promise and limitations of SAM, particularly when applied to domains with significant visual differences from its training data. This is where fine-tuning SAM for medical image segmentation becomes crucial.

This section delves deeper into SAM's inner workings, exploring its transformer-based architecture and the key advantages that make it suitable for medical image segmentation tasks.

3.1 A Powerful Transformer-Based Model

Specifically, SAM is a powerful transformer-based model, with an architecture that consists of three main components: an image encoder, a prompt encoder, and a mask decoder.

The **image encoder** processes the input image into a high-dimensional image embedding space. It utilises a standard Vision Transformer [12] pre-trained with the Masked Auto-Encoder (MAE) training scheme [16]. The ViT efficiently captures and extracts essential features and relationships within the image. As mentioned in Section 2, Meta AI provides three different pre-trained SAM models corresponding to the three ViT sizes. However, their research suggests that using larger ViT models as the backbone of SAM's image encoder offered only marginal improvements in accuracy, whilst resulting in significantly higher computational demands [25].

SAM's **prompt encoder** translates user prompts into internal representations that the model can understand. These prompts can be in the form of a point (positive or negative, to indicate the foreground and background of the target respectively), a bounding box (to surround the spatial region of the target object), or even textual descriptions. SAM also offers a special '*Everything'* mode, where the model segments all potential objects in the entire image.

The **mask decoder** iteratively combines the image embeddings from the image encoder and the prompt embeddings from the prompt encoder to segmentation masks, reflecting both the input image's visual features and the user's specified target through the prompt. It is designed as a modified Transformer-decoder block, utilising techniques such as prompt self-attention, which allows the prompt embeddings to interact and refine their representations, and cross-attention in two directions, which facilitates information exchange between the image embedding and the prompt embeddings. For ambiguous prompts, the decoder can produce multiple ranked output masks, allowing SAM to prioritise the most likely segmentation result.

3.2 Training Data and Target Domain

SAM was trained on a massive dataset, called **SA-1B**, which contained over 11 million 2D natural images consisting of over 1 billion masks. This carefully curated dataset encompasses a wide variety of natural objects and scenes, providing a broad understanding of visual concepts. It is important to note that the dataset only includes natural images, such as everyday objects (cars, animals, etc.), where annotated image-mask pairs are readily available.

3.3 Advantages of SAM

Several advantages of SAM make it suitable for various image segmentation tasks, including those in the medical domain.

As discussed, SAM leverages several **prompting methods**. This offers flexibility in the segmentation process, allowing users to leverage different prompts to direct SAM to specific structures. This is crucial for adapting SAM to the diverse and complex imagery in medical domains.

More significantly, a key strength of SAM is its capability to perform **zero-shot learning**. This means that it has the potential to segment unseen medical structures without extensive retraining. By providing high-quality prompts, SAM can be instructed on entirely new segmentation tasks, reducing the burden of data-specific training commonly required for medical image analysis models. This significantly reduces the need for data-specific training for medical tasks, although fine-tuning can still improve performance for highly complex medical images.

4 RECENT ADVANCEMENTS: APPLYING SAM TO MEDICAL IMAGES

While SAM shows promise, its application to medical image segmentation faces challenges due to inherent differences from natural images [11, 18, 33]. Fortunately, several efforts have successfully adapted SAM for this domain, achieving performance improvements (as seen in Figure 1).

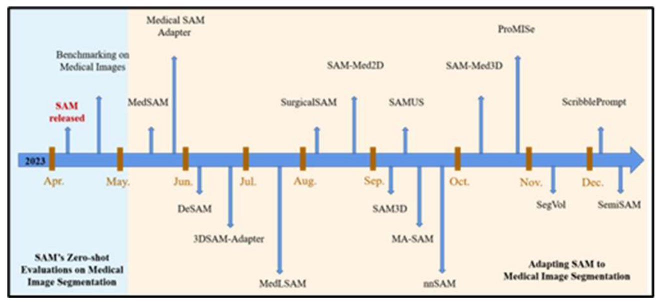

Figure 1: A brief chronology of SAM and its variants for Medical Image Segmentation (2023-2024) [50].

This section explores three key areas crucial for this adaptation: fine-tuning approaches, prompting methods, and dataset considerations.

4.1 Fine-Tuning Methods

Some research has found that fine-tuning strategies are most effective in improving SAM for medical image segmentation [33, 43], in comparison to training an entirely new SAM from scratch, due to the model's already-perfected weight initialisations.

4.1.1. Full Fine-Tuning. The most straightforward approach involves fine-tuning all of SAM's components (image encoder, prompt encoder, and mask decoder) on a medical dataset. This leverages SAM's pre-trained weights as a starting point, allowing the model to adapt to the new domain.

SkinSAM, by Hu et al [19], used this approach for skin cancer segmentation. Their model demonstrated substantial improvement from 81.25% to 88.79% Dice scores, with the highest performance on vascular lesions. Similarly, an approach by Li et al. [30], **Polyp-SAM**, fully fine-tuned SAM for polyp segmentation, resulting in outstanding results with Dice scores consistently above 88% for several datasets.

Slightly different, **MedSAM**, by Ma et al. [33], fully fine-tuned SAM's image encoder and mask decoder, whilst leaving the prompt encoder frozen, with SAM's original initialised weights. This allows MedSAM to learn how to fuse medical prompts and image embeddings correctly. This model outperformed SAM by a large margin on medical image segmentation tasks, even exceeding the performance of traditional specialist models. The features of MedSAM's segmentations contain more semantic information than their U-Net [36] baseline model, even on challenging target objects. Specifically, it achieved Dice scores of 94%, 94.4%, 81.5%, and 98.4% for the segmentation tasks involving intracranial haemorrhage CT, glioma MRI, pneumothorax CXR, and polyp-endoscopy images respectively, demonstrating MedSAM's ability to learn complex anatomical structures and strong generalisability.

As shown through these existing attempts, a full fine-tuning approach of SAM for medical image segmentation can achieve high Dice scores. However, its drawback is that it requires significant training time and is computationally expensive.

4.1.2 Parameter-Efficient Fine-Tuning (PEFT). PEFT addresses the drawbacks of full fine-tuning by fine-tuning only a subset of SAM's parameters, striking a balance between leveraging pre-trained knowledge and adapting to medical images.

Med-SA, by Wu et al. [43], and **SAMed**, by Zhang and Liu [46], both added Low-Rank Adaption (LoRA) modules to SAM's image and prompt encoders, whilst keeping SAM's original pre-trained weights frozen. Both models achieved Dice scores above 80% across several datasets, outperforming both previous state-of-the-art methods and MedSAM, while fine-tuning less of SAM's parameters.

However, Paranjape et al. [34] argue that the approach used by both **Med-SA** [43] and **SAMed** [46] can be memory-intensive due to the additional adapter layers added to every transformer block, increasing computational overhead, alongside the many gradient computations required for these parameters.

Instead, for **AdaptiveSAM**, Paranjape et al. [34] proposed taking advantage of SAM's already-proven capability to capture distinctive features in the initial layers of the network by freezing the image encoder and prompt encoder. Then, they performed a simple yet efficient method for fine-tuning, called Bias-Tuning, where they essentially only modified the biases of the encoders. Since SAM is already trained with one billion masks, it can already inherently segment general objects with its weight matrices. So, they simply added a trainable shift parameter to the outputs of each transformational layer. Only these shift parameters were trainable, as all other weights and biases were frozen, resulting in the model more easily learning the intricacies of medical images. Their findings showed that AdaptiveSAM outperforms state-of-the-art image segmentation methods on the following: surgical datasets (focused on medical instruments in robotic surgeries), abdominal ultrasounds, and chest x-rays. However, careful consideration of data imbalance issues was needed, as AdaptiveSAM's performance deteriorated massively when data was imbalanced.

As shown through these models, PEFT approaches require significantly less training time and computational resources, as compared to full fine-tuning. However, it may need additional strategies for limited data, due to the fewer parameters being specifically tuned for the data.

4.1.3 Framework Modification. This approach either refines SAM's architecture or integrates SAM with existing frameworks for optimal performance for medical image segmentation.

For example, Li et al. [31] introduced **nnSAM**, an integration of SAM as a plug-and-play module with nnU-Net [21], previously mentioned in Section 2. nnSAM achieved more accurate medical image segmentation in comparison to both raw SAM and raw nnU-Net models. In other words, nnSAM leveraged the strengths of both models.

Subsequently, an approach by Lin et al. [32] is **SAMUS**, which introduces a parallel CNN branch that injects local features into SAM's image encoder for enriched feature representation. Notably, SAMUS is designed for adapting SAM to small-size inputs, making it potentially more suitable for healthcare deployments. Evaluations demonstrated that SAMUS outperforms both **MedSAM** [33] and **Med-SA** [43] across various modalities, while significantly reducing computational cost.

Through these framework modification efforts to adapt SAM to the medical domain, one can deduce that these models either aim to improve performance (**nnSAM** [31]) or reduce computational resources (**SAMUS** [32]). However, a disadvantage is that it may introduce additional complexity.

4.1.4 Efficient Annotation Learning Facilitation. Due to the high cost of medical image annotation, some methods have explored how SAM can be utilised for efficient learning with limited labelled data.

To produce reliable pseudo-labels, Li et al. [29] leveraged SAM to conduct model predictions consistent with generated confidence-based pseudo-labels. They then worked with a subset with the highest confidence to further boost the existing semi-supervised segmentation model. Their evaluations achieved between 6.84% and 10.76% improvement, using a dataset with only 5% of labelled data.

However, instead of generating pseudo labels, Zhang et al. [47] introduced a semi-supervised framework called **SemiSAM**, that leverages a pre-trained segmentation model with domain knowledge to guide SAM. In other words, this is a form of transfer learning, where the pre-trained segmentation model acts as a teacher for SAM. Their evaluation on a dataset of heart MRIs demonstrated that SemiSAM achieves significant improvements, particularly when labelled data is extremely limited.

These annotation facilitation approaches reduce the need for extensive labelled data by leveraging pre-trained knowledge. However, this is a relatively new attempt in the medical SAM community with minimal data, so further research and experimentation are needed.

4.2 Prompting Methods

While fine-tuning strategies improve SAM for medical images, high-quality prompts are crucial for accurate segmentation. However, creating them can be challenging due to noisy annotations and the need for medical expertise.

Prompting methods offer a promising approach to improve the performance and practicality of SAM for medical image segmentation tasks. By automating prompt generation, reducing reliance on expert input, and integrating uncertainty estimation, researchers are paving the way for more robust and reliable medical image segmentation with SAM. While this literature review excludes natural language processing in the form of text prompting, this section explores various prompting methods that attempt to enhance SAM's capabilities in medical image segmentation.

4.2.1 Manual Prompting Methods. Most fine-tuning approaches utilised bounding boxes or points as prompts. However, while effective, they require manual intervention for each image, hindering their real-world practicality due to the bottleneck created by required medical expertise [34].

Ma et al. [33] report that with the point-based prompting method, segmentations can suffer from ambiguity and often require several human interventions to correctly segment the target object. Conversely, SAM's fully automated Everything prompting mode, while not requiring human intervention, may not provide enough focus, often producing segmentations that lack practical utility. This is because medical professionals tend to only work with a specific region-of-interest in a medical image, such as a particular organ.

Ultimately, for manual prompting methods, Ma et al. [33] argue that the bounding box prompting method is the best as it can specify the target object with the minimum amount of human intervention, whilst reducing ambiguity. However, they acknowledge that their model, **MedSAM**, experienced poor performance in segmenting vessel-like branches, such as arteries and veins, because the bounding box can have a high level of ambiguity in this context. This is due to when enclosing the entire complex structure in the box, other irrelevant regions-of-interest are often captured simply due to the nature of the target object's shape.

Therefore, the effectiveness of using SAM's manual prompting methods can be observed. However, it is hindered by its impracticality of requiring user input for each image and its inherent ambiguity from human error.

4.2.2 Automatic Prompt Generation. To address these challenges, research has explored automatic prompt generation methods that establish a feedback loop between the fine-tuned SAM and the prompt generator, leading to more reliable segmentation outcomes by mitigating the fluctuations in SAM's performance based on prompts [13].

For example, Lei et al. [26] employed a **few-shot localisation process** for their **MedLSAM** model. This was used to identify volumetric bounding boxes for anatomical structures in 3D medical images, using the assumption that locally similar pixel distributions correspond to the same region-of-interest. It then derived 2D boxes for each slice to guide SAM in automatically segmenting the target anatomy. However, MedLSAM achieved comparable performance to SAM, with minimal improvements in the case of head and neck organs.

Alternatively, Anand et al. [2] proposed a **one-shot localisation process** with a segmentation framework, guided together by a singular template image that was used to prompt SAM. This method utilised pre-trained ViT-based foundation models to extract dense features from the template image, achieving Dice scores between 62-90% across various modalities, using just one image as a reference point.

4.2.2.1 Integrating Uncertainty. Integrating uncertainty estimation of the generated prompts further enhances SAM's prompt robustness. It helps to identify potential segmentation inaccuracies, ultimately improving reliability and aiding medical professionals in real-world applications.

For instance, Xu et al. introduced **EviPrompt** [26], a training-free prompt-generation method based on uncertainty estimation that does not need expert interaction or training. It only requires a single medical image-annotation pair as a reference. Their experiments found that an EviPrompt-trained MedSAM outperformed the standard MedSAM model [33].

Similarly, Zhang et al. introduced the **Uncertainty Rectified SAM (UR-SAM)** framework [48], which aimed to enhance the robustness and reliability of auto-prompting medical image segmentation by estimating uncertainty maps. In other words, their framework utilises the uncertainty of a prompt to rectify the possible error, improving the final segmentation result. Their experiments on two 3D medical datasets demonstrated that using the UR-SAM framework can improve performance with up to 10.7% and 13.8% in Dice scores.

While automatic prompt generation addresses the challenges of manual prompting, models that utilise it have been observed to underperform, such as **MedLSAM** [26]. However, improved performance and reliability can be gained from the integration of uncertainty estimation, through employing certain frameworks, such as **EviPrompt** [44] or **UR-SAM** [48].

4.3 Learnable Prompts

An alternative approach to automatic prompt generation involves learnable prompts, where a separate component is trained to generate them, eliminating the need for any pre-defined prompts.

For example, for **AutoSAM**, Shaharabany et al. [38] used a separate, trainable CNN-based head that extracts features from the input image to create conditional prompts, while keeping the image encoder's weights frozen. What makes their model "auto" is that it involves training this external prompt encoder to generate "surrogate prompts" without any further fine-tuning of SAM, allowing the model to go beyond typical prompts in a fully auto-prompted manner. It demonstrated state-of-the-art results across various medical benchmarks.

Similarly, an approach for Cui et al.'s **All-in-SAM** [10] was to utilise SAM as a foundational model to generate pixel-level annotations from weak prompts, such as bounding boxes. This allowed the model to not require high-quality manual prompts and was able to achieve competitive performance in nuclei segmentation in comparison to state-of-the-art methods while reducing the annotation time.

These learnable prompt methods eliminate the need for high-quality prompts by training a separate component to generate them, achieving state-of-the-art performance, such as with **AutoSAM** [38] and **All-in-SAM** [10]. However, it does introduce additional complexity through the addition of this external trainable component.

4.3 Dataset Curation: Across Diverse Medical Modalities

The success of fine-tuning SAM for medical image segmentation hinges heavily on the quality and diversity of the curated dataset that the model is trained on. SAM's pre-training on inherent 2D architectures can lead to suboptimal results when directly applied to 3D medical image segmentation.

4.3.1 2D vs. 3D Medical Images. In deciding between catering for 2D versus 3D medical images, Ma et al. [33] (**MedSAM**) argue practically: to focus on 2D data, as 3D images can be processed as a series of 2D slices.

However, as argued by Zhang et al. [49], this method discards important depth-related spatial context. This is extremely important for the identification of certain volumetric medical objects in real-world applications, such as accurately quantifying the volume of tumours.

Bui et al.'s **SAM3D** [6] takes a different approach to 2D slice segmentation. They process each 2D slice individually using SAM, generating slice embeddings. Then, they utilise a specialised, lightweight 3D decoder to produce the final segmentation result, ensuring minimal loss of depth information. However, this requires careful design to ensure that the dataset used is compatible with the 3D decoder.

Ultimately, the choice between 2D and 3D data curation for SAM-based medical image segmentation depends on the application's specific needs for detail and efficiency. While 2D approaches benefit from leveraging the existing infrastructure of SAM, 3D methods offer a more comprehensive understanding of spatial relationships within medical images.

4.3.2 Adapting SAM to 3D Images. For 3D medical images, several methods have been proposed to bridge the gap between SAM's 2D architecture and the complexities of 3D data.

Firstly, Wu et al. [43] (**Med-SA**) proposed a technique called **Space-Depth Transpose (SD-Trans),** where a bifurcated attention mechanism is employed. One branch captures spatial correlations, while the other focuses on depth correlations. In other words, Med-SA adapted 2D SAM to 3D medical images by transposing spatial depth by adding self-attention blocks.

Both **3DSAM-Adapter**, by Gong et al. [14], and **Modality-Agnostic SAM (MA-SAM)**, by Chen et al. [7], utilised **3D adaptors** within each ViT [12] block of SAM's image encoder. These 3D adapters were fine-tuned alongside SAM's mask decoder, providing them with the capability of handling both 2D and 3D volumetric data. 3DSAM-Adapter achieved significant outperformance over nnU-Net [21] on tumour datasets, while MA-SAM consistently surpassed both 3DSAM-Adapter and nnU-Net, even when using SAM's *Everything* prompting mode.

Finally, Li et al. [28] proposed the **Prompt-driven 3D Medical Image Segmentation** model **(ProMISe),** which integrates **lightweight adaptors** to extract depth information without modifying SAM's pre-trained weights. Through their evaluations, ProMISe showed superior performance in comparison with state-of-the-art methods, particularly in colon and pancreas tumour segmentation datasets.

4.3.3 Medical Image Diversity. The sheer variety of medical image modalities requires diverse datasets for effective fine-tuning of SAM for medical image segmentation tasks, each presenting its own challenges. The following will briefly cover some of the significant medical imaging types and their challenges. However, in practice, these challenges also vary depending on the particular region-of-interest being segmented.

X-ray images, vital for fracture detection, face difficulties in identifying small or faint structures, such as fractures, against complex, overlapping anatomical backgrounds, such as ribs. Common datasets include ChestX-Ray8 [42], MIMIC-CXR-JPG [24] and CheXpert [20].

Computed Tomography (CT) scans, providing detailed cross-sectional views, often encounter label imbalance issues, where there is a large background region, such as a lung, and a small target, like a tumour. Certain organs, such as the liver, can have large size variations between patients, further affecting segmentation accuracy. The Cancer Imaging Archive (TCIA) and the Lung Database Consortium (LIDC) [4] are valuable resources for CT datasets.

Magnetic Resonance Imaging (MRI) scans, providing detailed images of organs through strong magnetic fields and radio waves, can be challenging due to the high precision needed when segmenting small structures, like the hippocampus in the brain. Organ size variation can also be problematic, especially with limited training datasets and high patient variability. Datasets like Ischemic Stroke Lesion (ISLES) [17], fastMRI [45], and the Brain Tumour Segmentation Challenge (BraTS) [5] are commonly used for MRI segmentation research.

Ultrasound images, utilised for real-time visualisation of internal organs through sound waves, often have lower resolution and greater noise, making the segmentation of fine-grained structures, such as nerves, more challenging. In addition, speckle patterns and organ deformations due to probe pressure introduce further complexity. Breast Ultrasound Images [1] and Ultrasound Nerve Segmentation (UNS) [51] are common ultrasound datasets.

Positron Emission Tomography (PET) scans, using radioactive tracers to measure metabolic activity in tissues, often have a lower spatial resolution compared to CTs and MRIs. In addition, a PET scan's signal intensity varies depending on the radiopharmaceutical used, introducing challenges for consistent segmentation. Due to this, there are not many high-quality PET datasets used for medical image segmentation research.

Endoscopy images, performed using a thin, flexible tube with a camera to examine internal organs, present challenges due to varying illumination conditions and blurriness caused by camera motion. In addition, the presence of various fluids and instruments deteriorates image quality, complicating segmentation. Kvasir [35] and EndoScene [41] are commonly used endoscopy datasets.

Dermoscopy uses a magnified view of the skin to aid in the diagnosis of skin cancer. Segmenting skin lesions can be challenging due to variations in colour, texture, and shape. A commonly used skin lesion dataset is the HAM10000 [40].

5 ETHICAL CONSIDERATIONS: AI IN MEDICAL DIAGNOSIS & DECISION-MAKING

A SAM-based medical image segmentation model, such as SAMSeg, promises enhanced healthcare outcomes by aiding in diagnosis and decision-making. However, ethical considerations demand careful attention throughout development and deployment, including rigorous validation, **explainability** of outputs, and open communication [37].

SAMSeg's performance could be susceptible to **biases** present from training data. To mitigate this, curating diverse datasets and monitoring for bias through standardised fairness metrics are essential to minimise bias amplification [8, 9].

For example, Ma et al. [33] compared MedSAM's performance with six different human experts in a prostate image segmentation task. Their findings showed that MedSAM exceeded a third of the experts, whilst performing comparably with the others. This highlights the potential for medical image segmentation as a real tool in clinical practice. However, SAMSeg is envisioned as a tool to **augment** human capabilities, not replace them. Automating tasks like image segmentation can free up valuable time for doctors to focus on complex cases that require human judgment.

Guided by the **FUTURE-AI principles** [27], SAMSeg's development and deployment should prioritise *Fairness, Universality, Traceability, Usability, Robustness,* and *Explainability*. Through this, we can ensure that SAMSeg is a force for good, improving healthcare outcomes while minimising potential ethical concerns.

6 DISCUSSION & CRITICAL COMPARISON

Having explored various approaches for leveraging the Segment Anything Model (SAM) in medical image segmentation tasks, this section delves into the key findings of this literature review and their implications for our project, SAMSeg.

Firstly, this review confirms SAM's potential for medical image segmentation tasks due to its zero-shot learning and domain adaption capabilities [33]. However, the degree of improvement may vary depending on the medical domain and fine-tuning approach. While SAM is promising, some efforts haven't surpassed existing, specialised methods like nnU-Net [21] due to their tailored architectures.

The primary focus of this literature review was to investigate successful fine-tuning approaches for SAM in medical image segmentation tasks, and for which specific medical image modalities were these methods effective.

Full fine-tuning, while effective and accurate (MedSAM [33]), is computationally expensive and time-consuming. SAMSeg can benefit from this approach, particularly if high accuracy is prioritised, but careful consideration must be given based on available computational resources.

Parameter-Efficient Fine-Tuning (PEFT), used in Med-SA [43] and Adaptive SAM [34], offers a balance between leveraging pre-trained knowledge and adapting to medical images, requiring less training time and resources. However, PEFT might require additional strategies for limited data scenarios due to fewer tuneable parameters. SAMSeg can leverage PEFT for faster training and explore techniques like bias-tuning for data scarcity.

Modifying SAM's framework can improve performance (nnSAM [31]) or reduce computational costs (SAMUS [32]). While potentially beneficial for SAMSeg, it can introduce additional complexity. SAMSeg can explore these modifications to achieve goals (e.g., performance boost or resource efficiency) while carefully considering the introduced complexity.

Facilitating efficient annotation learning, used by methods by Li et al. [29], is a new approach that can be valuable for SAMSeg, especially for medical domains with limited labelled data. These methods leverage pre-trained knowledge or transfer learning to generate pseudo-labels, reducing the need for extensive manual annotation.

In terms of prompting methods, manual prompting, while effective, requires user input for each image, hindering practicality. SAMSeg should explore automatic prompting methods to reduce reliance on manual intervention. However, it has been found to underperform, as seen in MedLSAM [26]. Integrating uncertainty estimation by using a framework like EviPrompt [44] or UR-SAM [48], can enhance reliability, making it a promising direction for SAMSeg.

Alternatively, methods that make use of learnable prompts, such as AutoSAM [38], achieve state-of-the-art results but introduce additional complexity. SAMSeg should investigate the feasibility of this approach, considering the trade-off between performance gains and added complexity.

Finally, in terms of dataset curation, the choice between 2D or 3D medical data depends on SAMSeg's needs. While 2D approaches are efficient, 3D methods offer a more comprehensive understanding. Therefore, SAMSeg should prioritise 3D data if spatial relationships of our chosen medical domain are especially crucial.

This review has presented several methods to bridge the gap between SAM's 2D architecture and 3D data complexities, such as those used in Med-SA [43] and 3DSAM-Adapter [14]. SAMSeg can benefit from these advancements to effectively handle 3D medical images, particularly techniques that don't require significant modification to pre-trained weights, such as ProMISe [28].

In terms of the vast array of medical imaging modalities, careful consideration of the specific challenges associated with each one should be prioritised when curating our dataset for SAMSeg, such as noise in ultrasounds, varying illumination in endoscopy, and many more.

By carefully considering the advantages and disadvantages of these approaches, we can select the most suitable fine-tuning strategies, prompting methods, and dataset considerations for developing a robust and effective SAMSeg model.

7 CONCLUSIONS

This review investigated how the Segment Anything Model (SAM) can be leveraged for medical image segmentation. SAM's zero-shot learning and domain adaptation capabilities offer promise, albeit with effectiveness contingent on fine-tuning methods and medical domains.

Our findings highlight the success of fine-tuning approaches, especially full fine-tuning and Parameter-Efficient Fine-Tuning (PEFT) methods. MedSAM [33], a benchmark in the field, achieves state-of-the-art performance across diverse medical image datasets. However, its full fine-tuning approach may pose computational challenges.

For SAMSeg's development, a hybridised fine-tuning approach is recommended, combining the efficiency of PEFT [13, 34, 43, 46] with targeted modifications to the SAM framework for medical image segmentation, specifically integrating with a nnU-Net model [21], as seen in nnSAM [31]. We also suggest using MedSAM [33] as a benchmark for achieving high accuracy segmentations, while considering our computational constraints.

In terms of SAMSeg's prompting method, focusing on manual prompting with bounding boxes is a good starting point due to its simplicity for tasks requiring specific region-of-interest segmentation [33]. However, despite its effectiveness, it is impractical in real-world applications due to needing user input on every image. Therefore, SAMSeg should consider exploring automatic prompt generation techniques, such as EviPrompt [44] or UR-SAM [48], reducing manual effort for long-term development.

Regarding data dimensionality, the choice between 2D or 3D data hinges on SAMSeg's prioritisation. If spatial relationships are crucial, 3D data is preferred. However, for an initial development phase, we recommend focusing on 2D data due to efficiency. As SAMSeg progresses, incorporating techniques from methods like 3DSAM-Adapter [14] can enable the effective handling of 3D medical images.

Similarly, the chosen medical imaging modality should consider the unique challenges associated with each type. During SAMSeg's dataset curation, we recommend starting with modalities that have clear and consistent characteristics, such as dermoscopy or CT scans. Success with these can pave the way for handling more complex modalities like endoscopy and ultrasounds.

In conclusion, future research should focus on hybrid fine-tuning and automatic prompt generation to further enhance SAM's capabilities across diverse domains, while acknowledging the need for tailored approaches for domains with greater complexity. By implementing these conclusions, developing our robust and efficient SAMSeg model is possible, unlocking the full potential of SAM for medical image segmentation tasks.

REFERENCES

[1] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. 2020. Dataset of breast ultrasound images. *Data in Brief* 28, (February 2020), 104863. https://doi.org/10.1016/j.dib.2019.104863

[2] Deepa Anand, Gurunath Reddy M, Vanika Singhal, Dattesh D. Shanbhag, Shriram KS, Uday Patil, Chitresh Bhushan, Kavitha Manickam, Dawei Gui, Rakesh Mullick, Avinash Gopal, Parminder Bhatia, and Taha Kass-Hout. 2023. One-shot Localization and Segmentation of Medical Images with Foundation Models. (2023). https://doi.org/10.48550/ARXIV.2310.18642

[3] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, AnnetteKopp-Schneider, Bennett A. Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M. Summers, Bram van Ginneken, Michel Bilello, Patrick Bilic, Patrick F. Christ, Richard K. G. Do, Marc J. Gollub, Stephan H. Heckers, Henkjan Huisman, William R. Jarnagin, Maureen K. McHugo, Sandy Napel, Jennifer S. Goli Pernicka, Kawal Rhode, Catalina Tobon-Gomez, Eugene Vorontsov, Henkjan Huisman, James A. Meakin, Sebastien Ourselin, Manuel Wiesenfarth, Pablo Arbelaez, Byeonguk Bae, Sihong Chen, Laura Daza, Jianjiang Feng, Baochun He, Fabian Isensee, Yuanfeng Ji, Fucang Jia, Namkug Kim, Ildoo Kim, Dorit Merhof, Akshay Pai, Beomhee Park, Mathias Perslev, Ramin Rezaiifar, Oliver Rippel, Ignacio Sarasua, Wei Shen, Jaemin Son, Christian Wachinger, Liansheng Wang, Yan Wang, Yingda Xia, Daguang Xu, Zhanwei Xu, Yefeng Zheng, Amber L. Simpson, Lena Maier-Hein, and M. Jorge Cardoso. 2022. The Medical Segmentation Decathlon. *Nat Commun* 13, 1 (July 2022), 4128. https://doi.org/10.1038/s41467-022-30695-9

[4] Samuel G. Armato, Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng Zhao, Denise R. Aberle, Claudia I. Henschke, Eric A. Hoffman, Ella A. Kazerooni, Heber MacMahon, Edwin J. R. Van Beeke, David Yankelevitz, Alberto M. Biancardi, Peyton H. Bland, Matthew S. Brown, Roger M. Engelmann, Gary E. Laderach, Daniel Max, Richard C. Pais, David P. Y. Qing, Rachael Y. Roberts, Amanda R. Smith, Adam Starkey, Poonam Batrah, Philip Caligiuri, Ali Farooqi, Gregory W. Gladish, C. Matilda Jude, Reginald F. Munden, Iva Petkovska, Leslie E. Quint, Lawrence H. Schwartz, Baskaran Sundaram, Lori E. Dodd, Charles Fenimore, David Gur, Nicholas Petrick, John Freymann, Justin Kirby, Brian Hughes, Alessi Vande Casteele, Sangeeta Gupte, Maha Sallamm, Michael D. Heath, Michael H. Kuhn, Ekta Dharaiya, Richard Burns, David S. Fryd, Marcos Salganicoff, Vikram Anand, Uri Shreter, Stephen Vastagh, and Barbara Y. Croft. 2011. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys* 38, 2 (February 2011), 915–931. https://doi.org/10.1118/1.3528204

[5] Ujjwal Baid, S. Ghodasara, M. Bilello, S. Mohan, E. Calabrese, E. Colak, K. Farahani, Jayashree Kalpathy-Cramer, F. Kitamura, Sarthak Pati, L. Prevedello, J. Rudie, C. Sako, R. Shinohara, Timothy Bergquist, R. Chai, James A. Eddy, Julia Elliott, W. Reade, T. Schaffter, Thomas Yu, Jiaxin Zheng, BraTS Annotators, C. Davatzikos, John T. Mongan, C. Hess, S. Cha, J. Villanueva-Meyer, J. Freymann, J. Kirby, B. Wiestler, P. Crivellaro, Rivka R.Colen, Aikaterini Kotrotsou, D. Marcus, Mikhail Milchenko, Arash Nazeri, H. Fathallah-Shaykh, R. Wiest, A. Jakab, M. Weber, A. Mahajan, Bjoern H. Menze, A. Flanders, and S. Bakas. 2021. The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. *ArXiv* (July 2021). Retrieved March 21, 2024 from https://www.semanticscholar.org/paper/The-RSNA-ASNR-MICCAI-BraTS-2021-Benchmark-on-Brain-Baid-Ghodasara/94be3df88de438fa8326b6ef766f538ccbebcacb

[6] Nhat-Tan Bui, Dinh-Hieu Hoang, Minh-Triet Tran, Gianfranco Doretto, Donald Adjeroh, Brijesh Patel, Arabinda Choudhary, and Ngan Le. 2023. SAM3D: Segment Anything Model in Volumetric Medical Images. (2023). https://doi.org/10.48550/ARXIV.2309.03493

[7] Cheng Chen, Juzheng Miao, Dufan Wu, Zhiling Yan, Sekeun Kim, Jiang Hu, Aoxiao Zhong, Zhengliang Liu, Lichao Sun, Xiang Li, Tianming Liu, Pheng-Ann Heng, and Quanzheng Li. 2023. MA-SAM: Modality-agnostic SAM Adaptation for 3D Medical Image Segmentation. https://doi.org/10.48550/arXiv.2309.08842

[8] Richard J. Chen, Judy J. Wang, Drew F. K. Williamson, Tiffany Y. Chen, Jana Lipkova, Ming Y. Lu, Sharifa Sahai, and Faisal Mahmood. 2023. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat. Biomed. Eng* 7, 6 (June 2023), 719–742. https://doi.org/10.1038/s41551-023-01056-8

[9] Mildred K. Cho. 2021. Rising to the challenge of bias in health care AI. *Nat Med* 27, 12 (December 2021), 2079–2081. https://doi.org/10.1038/s41591-021-01577-2

[10] Can Cui, Ruining Deng, Quan Liu, Tianyuan Yao, Shunxing Bao, Lucas W. Remedios, Yucheng Tang, and Yuankai Huo. 2023. All-in-SAM: from Weak Annotation to Pixel-wise Nuclei Segmentation with Prompt-based Finetuning. https://doi.org/10.48550/arXiv.2307.00290

[11] Ruining Deng, Can Cui, Quan Liu, Tianyuan Yao, Lucas W. Remedios, Shunxing Bao, Bennett A. Landman, Lee E. Wheless, Lori A. Coburn, Keith T. Wilson, Yaohong Wang, Shilin Zhao, Agnes B. Fogo, Haichun Yang, Yucheng

Tang, and Yuankai Huo. 2023. Segment Anything Model (SAM) for Digital Pathology: Assess Zero-shot Segmentation on Whole Slide Imaging. https://doi.org/10.48550/arXiv.2304.04155

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/arXiv.2010.11929

[13] Weijia Feng, Lingting Zhu, and Lequan Yu. 2023. Cheap Lunch for Medical Image Segmentation by Fine-tuning SAM on Few Exemplars. (2023). https://doi.org/10.48550/ARXIV.2308.14133

[14] Shizhan Gong, Yuan Zhong, Wenao Ma, Jinpeng Li, Zhao Wang, Jingyang Zhang, Pheng-Ann Heng, and Qi Dou. 2023. 3DSAM-adapter: Holistic Adaptation of SAM from 2D to 3D for Promptable Medical Image Segmentation. https://doi.org/10.48550/arXiv.2306.13465

[15] J. A. Hartigan and M. A. Wong. 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 1 (1979), 100–108. https://doi.org/10.2307/2346830

[16] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollar, and Ross Girshick. 2022. Masked Autoencoders Are Scalable Vision Learners. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2022), 15979–15988. https://doi.org/10.1109/CVPR52688.2022.01553

[17] Moritz R. Hernandez Petzsche, Ezequiel de la Rosa, Uta Hanning, Roland Wiest, Waldo Valenzuela, Mauricio Reyes, Maria Meyer, Sook-Lei Liew, Florian Kofler, Ivan Ezhov, David Robben, Alexandre Hutton, Tassilo Friedrich, Teresa Zarth, Johannes Bürkle, The Anh Baran, Björn Menze, Gabriel Broocks, Lukas Meyer, Claus Zimmer, Tobias Boeckh-Behrens, Maria Berndt, Benno Ikenberg, Benedikt Wiestler, and Jan S. Kirschke. 2022. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Sci Data* 9, 1 (December 2022), 762. https://doi.org/10.1038/s41597-022-01875-5

[18] Chuanfei Hu, Tianyi Xia, Shenghong Ju, and Xinde Li. 2023. When SAM Meets Medical Images: An Investigation of Segment Anything Model (SAM) on Multi-phase Liver Tumor Segmentation. https://doi.org/10.48550/arXiv.2304.08506

[19] Mingzhe Hu, Yuheng Li, and Xiaofeng Yang. 2023. SkinSAM: Empowering Skin Cancer Segmentation with Segment Anything Model. (2023). https://doi.org/10.48550/ARXIV.2304.13973

[20] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng. 2019. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence*, July 17, 2019. 590–597. https://doi.org/10.1609/aaai.v33i01.3301590

[21] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein. 2018. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. https://doi.org/10.48550/arXiv.1809.10486

[22] Ashish Issac, M. Partha Sarathi, and Malay Kishore Dutta. 2015. An adaptive threshold based image processing technique for improved glaucoma detection and classification. *Computer Methods and Programs in Biomedicine* 122, 2 (November 2015), 229–244. https://doi.org/10.1016/j.cmpb.2015.08.002

[23] Wei Ji, Jingjing Li, Qi Bi, Tingwei Liu, Wenbo Li, and Li Cheng. 2023. Segment Anything Is Not Always Perfect: An Investigation of SAM on Different Real-world Applications. (2023). https://doi.org/10.48550/ARXIV.2304.05750

[24] Alistair E. W. Johnson, T. Pollard, Nathaniel R. Greenbaum, M. Lungren, Chih-ying Deng, Yifan Peng, Zhiyong Lu, R. Mark, S. Berkowitz, and S. Horng. 2019. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. January 21, 2019. . Retrieved March 21, 2024 from https://www.semanticscholar.org/paper/MIMIC-CXR-JPG%2C-a-large-publicly-available-database-Johnson-Pollard/3d29ce781f297dc543e44dfb39990baff3a3acca

[25] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023. 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371

[26] Wenhui Lei, Xu Wei, Xiaofan Zhang, Kang Li, and Shaoting Zhang. 2023. MedLSAM: Localize and Segment Anything Model for 3D CT Images. June 26, 2023. . Retrieved from https://www.semanticscholar.org/paper/MedLSAM%3A-Localize-and-Segment-Anything-Model-for-3D-Lei-Wei/9cea0a7f5d6be5e9fa97d6b8ce9f76ae02aabcf1

[27] Karim Lekadira, Richard Osuala, C. Gallin, Noussair Lazrak, Kaisar Kushibar, G. Tsakou, Susanna Auss'o, Leonor Cerd'a Alberich, K. Marias, Manolis Tskinakis, S. Colantonio, Nickolas Papanikolaou, Zohaib Salahuddin, H. Woodruff, P. Lambin, and L. Mart'i-Bonmat'i. 2021. FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging. *ArXiv* (September 2021). Retrieved March 21, 2024 from https://www.semanticscholar.org/paper/FUTURE-AI%3A-Guiding-Principles-and-Consensus-for-in-Lekadira-Osuala/c4885564498a31fd2fe57a007a89bbc2807b4f31?sort=is-influential

[28] Hao Li, Han Liu, Dewei Hu, Jiacheng Wang, and Ipek Oguz. 2023. Promise:Prompt-driven 3D Medical Image Segmentation Using Pretrained Image Foundation Models. (2023). https://doi.org/10.48550/ARXIV.2310.19721

[29] Ning Li, Lianjin Xiong, Wei Qiu, Yudong Pan, Yiqian Luo, and Yangsong Zhang. 2024. Segment Anything Model for Semi-supervised Medical Image Segmentation via Selecting Reliable Pseudo-labels. In *Neural Information Processing*, 2024. Springer Nature, Singapore, 138–149. https://doi.org/10.1007/978-981-99-8141-0_11

[30] Yuheng Li, Mingzhe Hu, and Xiaofeng Yang. 2023. Polyp-SAM: Transfer SAM for Polyp Segmentation. (2023). https://doi.org/10.48550/ARXIV.2305.00293

[31] Yunxiang Li, Bowen Jing, Zihan Li, Jing Wang, and You Zhang. 2023. nnSAM: Plug-and-play Segment Anything Model Improves nnUNet Performance. (2023). https://doi.org/10.48550/ARXIV.2309.16967

[32] Xian Lin, Yangyang Xiang, Li Zhang, Xin Yang, Zengqiang Yan, and Li Yu. 2023. SAMUS: Adapting Segment Anything Model for Clinically-Friendly and Generalizable Ultrasound Image Segmentation. (2023). https://doi.org/10.48550/ARXIV.2309.06824

[33] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. 2023. Segment Anything in Medical Images. https://doi.org/10.48550/arXiv.2304.12306

[34] Jay N. Paranjape, Nithin Gopalakrishnan Nair, Shameema Sikder, S. Swaroop Vedula, and Vishal M. Patel. 2023. AdaptiveSAM: Towards Efficient Tuning of SAM for Surgical Scene Segmentation. https://doi.org/10.48550/arXiv.2308.03726

[35] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler, and Pal Halvorsen. 2017. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. https://doi.org/10.1145/3193289

[36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.48550/arXiv.1505.04597

[37] Rohan Shad, John P. Cunningham, Euan A. Ashley, Curtis P. Langlotz, and William Hiesinger. 2021. Designing clinically translatable artificial intelligence systems for high-dimensional medical imaging. *Nat Mach Intell* 3, 11 (November 2021), 929–935. https://doi.org/10.1038/s42256-021-00399-8

[38] Tal Shaharabany, Aviad Dahan, Raja Giryes, and Lior Wolf. 2023. AutoSAM: Adapting SAM to Medical Images by Overloading the Prompt Encoder. (2023). https://doi.org/10.48550/ARXIV.2306.06370

[39] Lv Tang, Haoke Xiao, and Bo Li. 2023. Can SAM Segment Anything? When SAM Meets Camouflaged Object Detection. (2023). https://doi.org/10.48550/ARXIV.2304.04709

[40] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data* 5, 1 (August 2018), 180161. https://doi.org/10.1038/sdata.2018.161

[41] David Vázquez, Jorge Bernal, F. Javier Sánchez, Gloria Fernández-Esparrach, Antonio M. López, Adriana Romero, Michal Drozdzal, and Aaron Courville. 2017. A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images. *Journal of Healthcare Engineering* 2017, (2017), 1–9. https://doi.org/10.1155/2017/4037190

[42] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. 2017. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017), 3462–3471. https://doi.org/10.1109/CVPR.2017.369

[43] Junde Wu, Wei Ji, Yuanpei Liu, Huazhu Fu, Min Xu, Yanwu Xu, and Yueming Jin. 2023. Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation. https://doi.org/10.48550/arXiv.2304.12620

[44] Yinsong Xu, Jiaqi Tang, Aidong Men, and Qingchao Chen. 2023. EviPrompt: A Training-Free Evidential Prompt Generation Method for Segment Anything Model in Medical Images. (2023). https://doi.org/10.48550/ARXIV.2311.06400

[45] Jure Zbontar, F. Knoll, Anuroop Sriram, Matthew Muckley, M. Bruno, Aaron Defazio, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, H. Chandarana, Zizhao Zhang, M. Drozdzal, Adriana Romero, Michael G. Rabbat, Pascal Vincent, James Pinkerton, Duo Wang, N. Yakubova, Erich Owens, C. L.

Zitnick, M. Recht, D. Sodickson, and Y. Lui. 2018. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *ArXiv* (November 2018). Retrieved March 21, 2024 from https://www.semanticscholar.org/paper/fastMRI%3A-An-Open-Dataset-and-Benchmarks-for-MRI-Zbontar-Knoll/65d3e7fa7b452611aa729ac789d4373d6ee505b0

[46] Kaidong Zhang and Dong Liu. 2023. Customized Segment Anything Model for Medical Image Segmentation. (2023). https://doi.org/10.48550/ARXIV.2304.13785

[47] Yichi Zhang, Yuan Cheng, and Yuan Qi. 2023. SemiSAM: Exploring SAM for Enhancing Semi-Supervised Medical Image Segmentation with Extremely Limited Annotations. (2023). https://doi.org/10.48550/ARXIV.2312.06316

[48] Yichi Zhang, Shiyao Hu, Sijie Ren, Chen Jiang, Yuan Cheng, and Yuan Qi. 2023. Enhancing the Reliability of Segment Anything Model for Auto-Prompting Medical Image Segmentation with Uncertainty Rectification. (2023). https://doi.org/10.48550/ARXIV.2311.10529

[49] Yichi Zhang, Qingcheng Liao, Le Ding, and Jicong Zhang. 2022. Bridging 2D and 3D segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5D solutions. *Computerized Medical Imaging and Graphics* 99, (July 2022), 102088. https://doi.org/10.1016/j.compmedimag.2022.102088

[50] Yichi Zhang, Zhenrong Shen, and Rushi Jiao. 2024. Segment anything model for medical image segmentation: Current applications and future directions. *Computers in Biology and Medicine* 171, (March 2024), 108238. https://doi.org/10.1016/j.compbiomed.2024.108238

[51] Ultrasound Nerve Segmentation. Retrieved from https://kaggle.com/competitions/ultrasound-nerve-segmentation