UTILIZING THE 'SEGMENT ANYTHING MODEL' FOR MAMMOGRAPHY IMAGE ANALYSIS

Nikola Jovišić [1] [ORCID 0009-0003-3000-2920], *Velibor Ilić* [1] [ORCID 0000-0001-5010-1377], *Slobodan Ilić* [1] [ORCID 0000-0001-7771-6128], *Ivana Šešić* [2] [ORCID 0000-0002-7563-6834], *Aleksandra Dragičević* [3] [ORCID 0000-0003-2606-9703]

[1] *The Institute for Artificial Intelligence Research and Development of Serbia - IVI, Fruškogorska 1, 21000 Novi Sad, Serbia*
[2] *Radiology Specialist Doctor's Office DOUTORA, Belgrade, Serbia*
[3] *University of Belgrade, Faculty of Mechanical engineering, Department for Biomedical Engineering, Kraljice Marije 16, 11120 Belgrade, Serbia*

Abstract: *Mammography, considered the "gold standard" for breast mass detection, boasts approximately 90% sensitivity. However, the process of mammography reading and interpretation is time-consuming and susceptible to a significant number of false negatives and false positives. Compounding these issues is a notable shortage of radiologists specializing in breast imaging, highlighting a need for innovative solutions. This paper proposes the use of an artificial intelligence (AI) system, the Segment Anything Model (SAM), to address these challenges. SAM, a versatile image segmentation model, can generate segmentation masks in response to various prompts such as points, bounding boxes, masks, or textual descriptions. It employs a transformer-based architecture and utilizes a large-scale dataset for training. Within mammography, SAM could be deployed to automatically segment suspicious areas in the breast, thereby providing invaluable support to radiologists during diagnostic decision-making. Furthermore, SAM can be fine-tuned and trained on specific medical image datasets to enhance its performance. Our results underscore SAM's efficacy in producing accurate segmentation masks, demonstrating its potential for integration into human-in-the-loop systems for annotation and decision support in medical image analysis. Nevertheless, further research and training are essential before implementing SAM in real-world medical scenarios.*

Key words: image segmentation, semantic segmentation, computer vision, mammography

1. INTRODUCTION

Mammography, widely recognized as the "gold standard" for breast mass detection, offers an approximate sensitivity of 90% (Boot and Irshad, 2020; Jha et al., 2023) Each mammography examination generates X-ray images from two distinct angles for each breast, which are then carefully evaluated by a radiologist. This powerful diagnostic tool aids in the early detection of breast abnormalities by facilitating the identification of micro and macro calcifications, architectural distortions, as well as benign and malignant lesions (Liu et al., 2021; Singh et al., 2020). These diagnostic images and their subsequent interpretation are further guided by the Breast Imaging Reporting and Data System (BI-RADS). This classification system, proposed by the American College of Radiology, is a valuable tool utilized in interpreting and categorizing findings from mammograms, as well as breast ultrasounds and MRIs. With six categories that range from normal or benign to highly suspicious of malignancy, BI-RADS streamlines the decision-making process and directs subsequent follow-up recommendations (Magny et al., 2023). Despite the effectiveness of mammography, the interpretation process is time-consuming and lengthy, with a propensity for a significant number of false negatives and false positives. Compounding this issue is the shortage of radiologists specializing in breast imaging, underscoring the need for innovative solutions. The development of deep learning-based Computer-Aided Detection (CAD) systems has brought about considerable improvements in the detection, classification, and segmentation of Regions of Interest (ROI) within mammograms, and their use is highly recommended to support radiologists (Chanda and Sarkar, 2020; Jha et al., 2023; Kumar and Ramadevi, 2022; Salama and Aly, 2021).

The process of segmentation in mammogram images involves the division of these images into separate components, aiding in the identification of masses and facilitating the extraction of the ROI (Ali et al., 2019). It's important to remove pectoral muscles and artifacts before segmentation to avoid misguiding classification algorithms (Michael et al., 2021). One solution to these challenges is the application of artificial intelligence, particularly the Segment Anything Model (SAM) (Kirillov et al., 2023). SAM is capable

of autonomously determining suspicious tissue regions through segmentation, or it can do so with minimal expert intervention. Additionally, SAM can function within an active learning loop for annotation, potentially enhancing the accuracy and efficiency of mammographic diagnostics.
Building on the aforementioned challenges and potential solutions, this paper aims to illustrate the effectiveness of incorporating a contemporary artificial intelligence model, such as SAM, to establish a framework that enhances the field of mammography. We explore the possibility of augmenting radiologists' expertise with an automated system capable of segmenting and identifying areas of concern, thereby potentially improving diagnostic accuracy and efficiency. The integration of such a model offers the prospect of an active learning loop for annotation, which may advance the overall quality of mammographic diagnostics.

2. METHODS

To concretize the theory and test the effectiveness of SAM in a practical setting, we have employed specific mammography datasets in our research. To ensure the scientific robustness of our approach, we have made use of open mammography datasets from RSNA and CBIS-DDSM. Given that these datasets provide ground truth masks for some images, we can objectively evaluate the quality of segmentation achieved by the SAM model. This approach is designed to replicate real-world scenarios, where images would typically be loaded through communication with a Picture Archiving and Communication System (PACS) or from historical data kept in Digital Imaging and Communications in Medicine (DICOM) formats, as in our case. Our focus remains on the extensive preprocessing techniques necessary to adapt images to model-expected input, including normalization techniques, Contrast Limited Adaptive Histogram Equalization (CLAHE), image resizing, artifacts removal, boundaries cropping, panning, and replicating grayscale through all three RGB channels (Punitha and Perumal, 2019). These steps are critical to accurately and effectively utilizing AI models like SAM in real-world medical imaging.

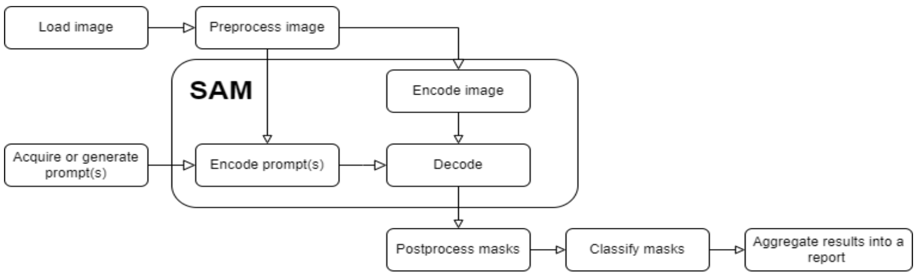

Figure 1 - Data flow chart of mammography image processing based on SAM

The Segment Anything model itself is a foundational model for image segmentation. The version used in this paper is an out-of-the-box one, intending to use it in a zero-shot regime and explore its efficacy in such a constellation. SAM introduces three novelties new task, segmentation model, and dataset (with segmentation annotations). The defined task is to create a segmentation mask based on an image as input and a specific prompt. Prompts can be point, bounding box, mask, and textual. In addition, there are three levels for mask generation, which the authors call object, part, and subpart, so if the user, for example, clicks on a suspicious mass on a mammogram, the model could theoretically generate three levels of segmentation - suspicious mass mask, suspicious mass mask with surrounding tissue and a mask of the whole breast, and the user chooses which mask corresponds to what he imagined.
In addition to the multi-mask mode, the model can also work in a single-mask mode if the user still wants to let the model generate only one mask, the one in which the model is most confident. There is also a multi-input mode, where the user can enter several different types of prompts, such as defining a bounding box around the breast on the mammogram and a point located in the suspicious mass but label the point as negative. In that case, the model should segment the entire breast except for the suspicious mass. As for the model itself, it contains three components: an image encoder, a prompt encoder, and a decoder. They are built on transformer models for computer vision with specific modifications such that the model can work in real-time.
A version of the Vision Transformer (ViT) model minimally adapted so that it can process high-resolution inputs in real-time is used as the image encoder. It runs once per image and can be run in advance before

the prompt is established. This part is usually frozen when fine-tuning, which is shown in the medical image processing example (Bagchi et al., 2020; Ma and Wang, 2023). The prompt encoder handles sparse and dense prompts differently. Sparse prompts include dots, frames, and text, and dense prompts include masks. Dots and frames are encoded by positional encoding and learned encoding, as well as off-the-shelf text encoder, and masks by convolution and summation with the input image, as shown in Figure 1. The decoder uses a transformer architecture to map the co-encoded image and prompt to a segmentation map. The dataset it was trained on consists of about 11 million images with about 1.1 billion segmentations. In the case of mammograms, a mask generator was used, which in the background performs the segmentation for the grid of input points as prompts to the model. The segmentation in which the model is the most certain is selected with its post-processing, i.e., for which the IoU metric is large enough. Another possibility of applying this model to this problem is the following: use the SAM model in prompt mode.

Depending on the type of prompt that would be used, there are several options for how to exploit this, so, for example, theoretically the text prompt "malignant tumor" or "benign calcification" should work on such images and find exactly what it is searching for. This certainly, even if we ignore the fact that the text mode of operation is not yet available, we could not expect it to work out-of-the-box if the model is not already trained on medical images of that type. It would be particularly challenging to extract a BI-RADS (Breast Imaging-Reporting and Data System) classification (Spak et al., 2017) through such a constellation because "BI-RADS 3" or "BI-RADS 4" is certainly not a textual prompt that a universally trained model would expect. What, however, exists as a possibility is to use this mode of operation in the future when it becomes available by performing fine-tuning on the model whereby some of the model's awareness of classes of interest from initial training could be used, but the model could additionally learn to accurately distinguish malignant and benign tumors, as well as benign and malignant calcifications from healthy tissue. Another type of prompt, as a representative of dense prompts, are masks. A mask of the breast itself could be used as a prompt mask. The breast mask can be automatically extracted through a series of steps - image binarization, contour finding, and keeping the inside of the largest contour as a binary mask corresponding to the breast. Such a mask found fully automatically, could be directly used as a model prompt. Additionally, for use in the human-in-the-loop systems, active learning (Ilić and Tadić, 2021), and help with annotation, which is certainly as interesting as the complete automation of decision-making (whether the subject is the system which will remain in this mode where it is left partly dependent on expert expertise or only on a temporary solution that will be useful in the stage of data set preparation - in the annotation), usage of an initial mask provided by an expert - radiologist could be exploited. An expert could mark the parts of the tissue that he considers suspicious, and the model would refine his masks into a more accurate segmentation. Boxes and points as prompts are also an option. Boxes can be manually created, or this coarse-level localization could also be automatized (Liu et al., 2020). Properly placed points labeled as 1 (foreground – suspicious mass) and 0 (background – healthy tissue) may be sufficient for the model to determine correct segmentations. This can be seen from the example with the generation of masks automatically, because in that example, in the background, nothing but multiple inference was performed on the model with prompts in the form of points.

It is also interesting to mention that prompts can be combined, so you can, for example, frame a certain group of calcifications and thus provide a box prompt, and additionally mark a certain number of points that are, and a certain number of points that are not, within the calcifications. This is how you get a complex prompt that will result in the model finding calcifications within the given region and correctly labeling them as foreground.

Masks acquired from a model could be further processed to filter out masks by size, shape, etc. As the SAM is a foundation model which is not particularly intended only for one type of segmentation, the type of masks differ according to the type of used prompts. If the pipeline is configured in such a way for the model to return instance masks of individual suspicious parts of the tissue, each mask could be further classified by severity, e.g., in a standard BI-RADS manner using an image classifier.

3. RESULTS

Figures 2 and 3 show the application of automatic mask generation using a grid of point prompts where parameters were refined to get from the segmentation on the left to the one on the right.

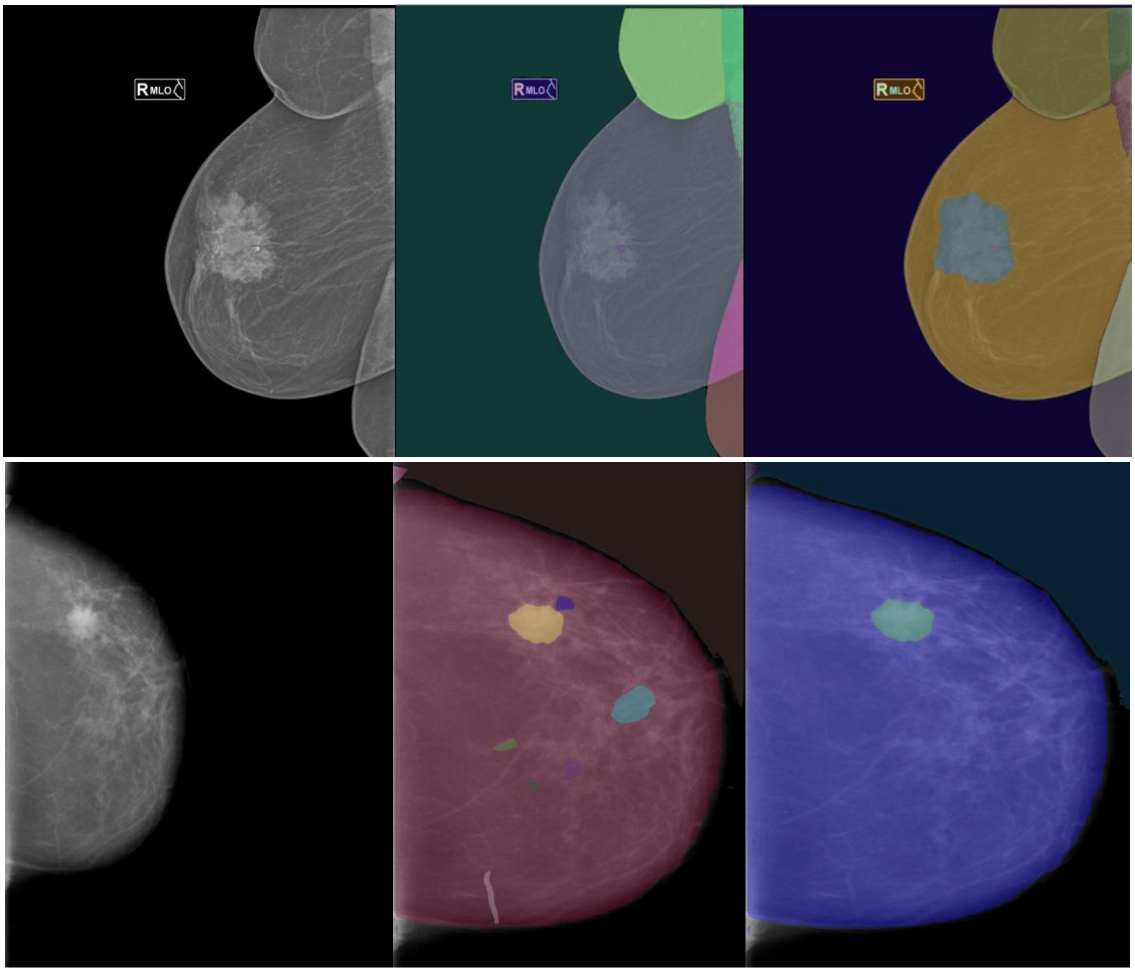

Figure 2 - Mammogram images processed with the SAM model with different settings, parameters tuned to accurately find the lesion

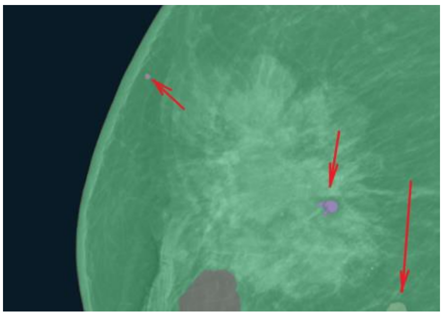

Figure 3 - Mammogram images processed with the SAM model with parameters tuned for finding calcifications

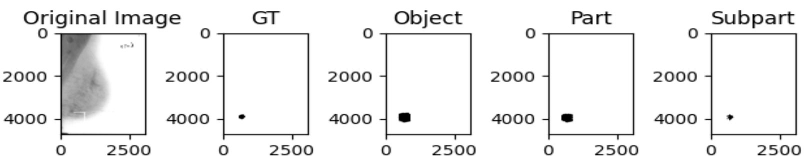

Figure 4 - The original mammogram image, the ground truth mask, the detected object and the part of the detected object and its subpart

Figure 4 shows the application of the SAM with a bounding box prompt artificially generated as an annotation frame. In this form, this has no practical application, but it serves to simulate a human-in-the-loop usage and demonstrate the general awareness of this model about the objects on the mammogram, and examine how capable it is, given the approximate location of what it is looking for, to segment it. The fact that the division into object-part-subpart in the case of finding suspicious masses on a mammographic image does not have clear semantics for a human, it can be expected that it exists in some way for the model. Thus, for prompts from the figure it can be seen that the subpart category best matches the ground truth mask - it is clear that the IoU (intersection over union) metric is the best in the case of the subpart mask and reaches values of the above 70% on this particular dataset *(CBIS-DDSM)*.

4. DISCUSSION

The greatest significance of the usage of SAM in any type of segmentation, therefore also in mammography, is its ability to produce very high-quality masks in a zero-shot state. Traditional segmentation methods like otsu thresholding (Fazilov et al., 2022) are not able to produce such high-quality masks. Usual segmentation methods in medical image processing like U-net (Baccouche et al., 2021; Ronneberger et al., 2015) are able to learn to make high-quality masks, but require a vast amount of data to be trained on and lack a foundation model. The results seen from the usage of SAM are comparable to those methods even in its zero-shot state, and for the fatty breast images where mass boundaries are easier to identify, this multi-purpose model performs even better than specifically trained models. In addition, multiple interfaces offer help in adapting the model to a wider range of usages, so users are not limited only to one type of segmentation, one type of prompt etc.

5. CONCLUSIONS

As shown, SAM could certainly be used in some form in processing mammography images. It could be used for human-in-the-loop systems for providing help in segmentation or help in the annotation phase in its prompt mode or could even be used in the automatic mask-generating mode in some cases, even in zero-shot. Further training of the model would, however, be highly recommended if any type of medical image analysis were to be seriously considered.

6. REFERENCES

Ali, M. T., Lahmood, F. M., & Chisab, R. F. (2019). *Mammogram image segmentation for improving the diagnosis of dense breast issues*. *International Journal of Engineering and Technology*, *8*(1), 44-52.

Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., Elmaghraby, A. S. (2021). Connected-UNets: A deep learning architecture for breast mass segmentation. Npj Breast Cancer, 7(1), 151.

Bagchi, S., Tay, K. G., Huong, A., & Debnath, S. K. (2020). *Image processing and machine learning techniques used in computer-aided detection system for mammogram screening-A review*. *International Journal of Electrical and Computer Engineering (IJECE)*, *10*(3), 2336-2348.

Boot, T., & Irshad, H. (2020). *Diagnostic assessment of deep learning algorithms for detection and segmentation of lesion in mammographic images*. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23* (pp. 56-65). Springer International Publishing.

Chanda, P. B., & Sarkar, S. K. (2020). *Detection and classification of breast cancer in mammographic images using efficient image segmentation technique*. In *Advances in Control, Signal Processing and Energy Systems: Select Proceedings of CSPES 2018* (pp. 107-117). Springer Singapore.

Fazilov, S. K., Yusupov, O. R., & Abdiyeva, K. S. (2022). *Mammography image segmentation in breast cancer identification using the otsu method*. *Web of Scientist: International Scientific Research Journal*, *3*(8), 196-205.

Ilić V., Tadić J. M. (2021) "*Active Learning using a Self-Correcting Neural Network (ALSCN)*", Applied Intelligence, The International Journal of Research on Intelligent Systems for Real Life Complex Problems

Jha, S., Ahmad, S., Arya, A., Alouffi, B., Alharbi, A., Alharbi, M., & Singh, S. (2023). *Ensemble Learning-Based Hybrid Segmentation of Mammographic Images for Breast Cancer Risk Prediction Using Fuzzy C-Means and CNN Model*. Journal of Healthcare Engineering, 2023, 1–18.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). *Segment anything*. arXiv preprint arXiv:2304.02643.

Kumar, M. A., & Ramadevi, Y. (2022). *A study: mammogram image segmentation and classification based on ABC algorithm and artificial neural networks*. *Advanced Engineering Services (54/02)*.

Liu, K., Shen, Y., Wu, N., Chłędowski, J., Fernandez-Granda, C., & Geras, K. J. (2021). *Weakly-supervised high-resolution segmentation of mammography images for breast cancer diagnosis*. Proceedings of machine learning research, 143, 268.

Liu, Q., Liu, Z., Yong, S., Jia, K., & Razmjooy, N. (2020). *Computer-aided breast cancer diagnosis based on image segmentation and interval analysis. Automatika*, *61*(3), 496-506.

Ma, J., & Wang, B. (2023, April 24). *Segment anything in Medical Images*. arXiv.org.

Magny, S. J., Shikhman, R., Keppke, A. L. (2023). *Breast Imaging Reporting and Data System*. In StatPearls. StatPearls Publishing

Michael, E., Ma, H., Li, H., Kulwa, F., & Li, J. (2021). *Breast Cancer Segmentation Methods: Current Status and Future Potentials*. BioMed Research International, 2021, 1–29.

Punitha, M., & Perumal, K. (2019). *Hybrid segmentation and feature extraction approach to detect tumour based on fuzzy rough-in mammogram images*. Procedia Computer Science, 165, 478-484.

Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation, *arXiv.org*

Salama, W. M., Aly, M. H. (2021). *Deep learning in mammography images segmentation and classification: Automated CNN approach*. Alexandria Engineering Journal, 60(5), 4701–4709.

Singh, V. K., Rashwan, H. A., Romani, S., Akram, F., Pandey, N., Sarker, Md. M. K., Saleh, A., Arenas, M., Arquez, M., Puig, D., Torrents-Barrena, J. (2020). *Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network*. Expert Systems with Applications, 139, 112855.

Spak D.A., Plaxco J.S., Santiago L., Dryden M.J., Dogan B.E.,, BI-RADS® fifth edition: *A summary of changes, Diagnostic and Interventional Imaging*, Volume 98, Issue 3, 2017, Pages 179-190, ISSN 2211-5684