

Article

Not peer-reviewed version

MicroLesion-SAM: Centerline-Guided Foundation Model for Small Lesion Segmentation

Daniel Levi , Yael Friedman , [Eitan Cohen](#) *

Posted Date: 29 December 2025

doi: 10.20944/preprints202512.2582.v1

Keywords: small lesion detection; SAM foundation model; centerline priors; ischemic lesion segmentation; WMH; ISLES2018; micro-lesion analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

MicroLesion-SAM: Centerline-Guided Foundation Model for Small Lesion Segmentation

Daniel Levi ^{1,*}, Yael Friedman ² and Eitan Cohen ³

Faculty of Engineering, Tel Aviv University, Tel Aviv 6997801, Israel

* Correspondence: daniel.levi@tau.ac.il

Abstract

Detecting small and low-contrast brain lesions is difficult for conventional foundation models, which often prioritize larger structures and overlook micro-lesions. MicroLesion-SAM enhances SAM by integrating structural centerline priors and scale-adaptive refinement to increase sensitivity to lesions below 20 voxels. The model uses probabilistic skeleton cues to guide attention toward lesion cores while preserving high-frequency features through resolution-adaptive enhancement. Evaluated on WMH2020 (60 subjects) and ISLES2018 (103 subjects), MicroLesion-SAM increases small-lesion recall from 0.684 to 0.812 (+18.7%) and raises global Dice to 0.902 (+4.5% over SAM-Med2D). HD95 decreases from 21.9 mm to 14.6 mm (−33.3%), and lesion-wise F1 improves by 11.2%. Cross-dataset validation on an independent clinical cohort of 72 subjects shows a 12.1% improvement in micro-lesion detection. Ablation experiments confirm that the centerline priors contribute a 7.3% Dice gain, while heatmap analysis shows more accurate localization of micro-vascular abnormalities.

Keywords: small lesion detection; SAM foundation model; centerline priors; ischemic lesion segmentation; WMH; ISLES2018; micro-lesion analysis

1. Introduction

Small ischemic and vascular lesions observed on brain magnetic resonance imaging (MRI) are widely regarded as important imaging markers of stroke risk, cerebral small vessel disease, and long-term neurological impairment. These lesions are typically very small, sometimes consisting of only a few voxels, and often present with weak contrast relative to surrounding brain tissue. Their subtle appearance makes reliable detection difficult during routine clinical reading and poses a substantial challenge for automated segmentation systems. Manual annotation remains the clinical reference standard, yet it is time-consuming and subject to considerable inter-observer variability, particularly when lesions overlap with image noise, partial volume effects, or normal vascular structures. Public benchmarks such as WMH2020 and ISLES2018 have supported methodological advances in lesion segmentation [1,2], but reported performance gains are largely driven by medium-sized or large lesions. Sensitivity to very small abnormalities remains limited, and recent studies suggest that explicitly emphasizing lesion centers or core regions can improve robustness when detecting subtle brain lesions under low-contrast conditions [3].

Most existing approaches for ischemic and vascular lesion segmentation are based on convolutional neural networks, attention mechanisms, or transformer-inspired architectures. These models achieve strong overall Dice scores for white matter hyperintensities and acute stroke lesions and have become the dominant paradigm in recent literature [4,5]. However, size-specific performance is rarely examined in detail. Training objectives and architectural designs tend to favor larger, more salient regions, while very small lesions are frequently merged into nearby structures, smoothed out during downsampling, or removed as noise. Cross-dataset evaluations further reveal that segmentation performance often drops substantially when models are applied to scans acquired at different centers or using different scanners [6]. This degradation is particularly pronounced for tiny, low-contrast lesions, indicating that current methods may lack the structural sensitivity and

generalization capability required for reliable micro-lesion analysis in real-world clinical settings. Specialized pipelines have been developed for detecting microbleeds, lacunar infarcts, and other small vascular abnormalities. Many of these systems adopt multi-stage strategies that first generate candidate regions and then apply false-positive reduction. While such designs can improve sensitivity on susceptibility-weighted or T2*-weighted images, they often produce a large number of false alarms and rarely provide accurate voxel-level boundaries [7]. Weakly supervised approaches offer alternatives by reducing annotation requirements, but they typically return coarse activation maps rather than precise segmentations suitable for quantitative analysis or longitudinal monitoring [8]. As a result, their clinical utility remains limited when detailed lesion morphology or burden estimation is required. Recently, foundation models for medical image segmentation have attracted increasing attention. Models derived from the Segment Anything Model (SAM) framework demonstrate impressive adaptability across tasks and imaging modalities, requiring minimal prompting or fine-tuning [9]. These approaches perform well for organs, large tumors, and clearly delineated anatomical structures. However, their ability to capture very small lesions is constrained by architectural design choices, including aggressive spatial downsampling and global attention mechanisms that inherently prioritize larger, visually prominent regions. Consequently, tiny lesions are frequently missed or blurred, which limits the applicability of foundation models to micro-lesion segmentation tasks. Emerging work that introduces center-prioritized scanning strategies and prototype-based representations indicates that guiding models toward lesion cores and structured internal representations can improve segmentation consistency for subtle brain lesions [3]. Nevertheless, such concepts have not yet been systematically integrated into foundation-model-based frameworks for small ischemic and vascular lesion analysis. Insights from vessel imaging further motivate the incorporation of structural cues. In vascular segmentation, centerline and skeleton-based representations have been shown to improve continuity and detection of thin structures by explicitly modeling their geometric cores [10]. Loss functions and auxiliary constraints derived from centerline information help networks preserve fine structural details that are otherwise lost during feature aggregation. These findings suggest that similar center-focused cues could guide attention toward the cores of very small brain lesions. However, the combination of centerline information with general-purpose segmentation backbones, particularly foundation models, remains largely unexplored in MRI-based lesion segmentation.

In this study, we introduce MicroLesion-SAM, a segmentation framework that augments a SAM-style backbone with centerline-based structural cues and a scale-adaptive feature path. The proposed centerline cues explicitly guide the model toward lesion cores, enhancing sensitivity to tiny abnormalities that would otherwise be suppressed during downsampling. The scale-adaptive path preserves high-resolution features throughout the network, reducing the loss of fine details during segmentation. We evaluate the proposed method on the WMH2020 and ISLES2018 benchmarks, as well as on an independent clinical cohort. Experimental results demonstrate improved recall for very small lesions, stronger overall segmentation performance, and more accurate localization across datasets. These findings indicate that integrating structural priors with foundation segmentation models provides a robust and practical solution for micro-lesion analysis in routine clinical MRI.

2. Materials and Methods

2.1. Study Cohort and Imaging Data

This study used three MRI datasets that contain ischemic and white matter lesions of different sizes. The WMH2020 dataset included 60 subjects from several European centers, scanned mainly on 3T MRI. The ISLES2018 dataset included 103 subjects with acute ischemic stroke, each scanned within two days of symptom onset. An additional clinical set of 72 subjects from a local hospital was used to examine generalization. All datasets provided T1-weighted, T2-weighted, FLAIR, and diffusion-weighted images, although sequence availability differed across subjects. Lesions ranged from very small lesions under 20 voxels to large confluent regions. Exclusion criteria were severe motion

artifacts, incomplete image sets, or missing clinical information. All images were anonymized according to institutional guidelines.

2.2. Experimental Design and Comparison Groups

We compared MicroLesion-SAM with three models that represent commonly used strategies for brain lesion segmentation. The test model was MicroLesion-SAM, which included centerline cues and a scale-adaptive step. The first comparison model was SAM-Med2D, used as a strong baseline for prompt-based medical segmentation. The second model was a 2D U-Net, trained separately on each dataset. The third model was a transformer-based segmentation network, selected because transformers have shown good performance for medium-sized lesions. All models followed the same train-validation-test split for each dataset to ensure fair comparison. The selection of these models allowed an assessment across classic CNNs, transformer methods, and foundation-model variants.

2.3. Measurement Methods and Quality Control

Lesion masks for WMH2020 and ISLES2018 followed the official challenge guidelines. For the clinical set, two radiologists created the initial labels independently. A senior neuroradiologist resolved all disagreements. Preprocessing consisted of skull removal, bias-field correction, and alignment to a standard anatomical space. Intensity values were normalized within each scan to reduce scanner variations. Quality checks included visual inspection of each image to detect registration errors, missing tissue, or uncertain lesion boundaries. Scans with unresolved issues after correction were removed from the study. During training, predicted masks were also reviewed to identify unexpected slice-to-slice inconsistencies.

2.4. Data Processing and Model Computation

All MRI scans were resampled to a common voxel size before model training. Data augmentation included rotation, mild elastic deformation, and additive noise. Coarse lesion maps were produced using low-threshold segmentation, and skeletonization was applied to these maps to obtain centerline cues. These cues were used as soft guidance signals during training.

Two main quantitative measures were used. The lesion-wise F1 score was calculated as [11]:

$$F1_{les} = \frac{2 TP}{2 TP + FP + FN},$$

where TP, FP, and FN denote correctly detected, false-positive, and missed lesions.

The second measure was the small-lesion recall, defined as [12]:

$$Recall_{small} = \frac{TP_{small}}{TP_{small} + FN_{small}},$$

where “small lesions” were those smaller than 20 voxels. These measures helped separate overall segmentation accuracy from performance on very small lesions. All models were trained in PyTorch using the same optimization settings unless stated otherwise.

2.5. Statistical Analysis and Evaluation Metrics

Performance was assessed using Dice similarity, lesion-wise F1, HD95, and small-lesion recall. Metrics were calculated per subject and averaged across each dataset. Pairwise statistical tests were used to compare MicroLesion-SAM with each baseline, and results with $p < 0.05$ were considered significant. Cross-dataset tests were performed by training on one dataset and evaluating on another. For the clinical set, both false positives and false negatives were reviewed manually to understand common error patterns and assess clinical relevance.

3. Results and Discussion

3.1. Segmentation Accuracy on Public Datasets

MicroLesion-SAM achieved a global Dice of 0.902 on WMH2020 and ISLES2018, higher than SAM-Med2D and the two baseline networks. The model also reduced HD95 from 21.9 mm to 14.6 mm, a 33.3% decrease. These improvements were observed across both datasets and for different lesion categories.

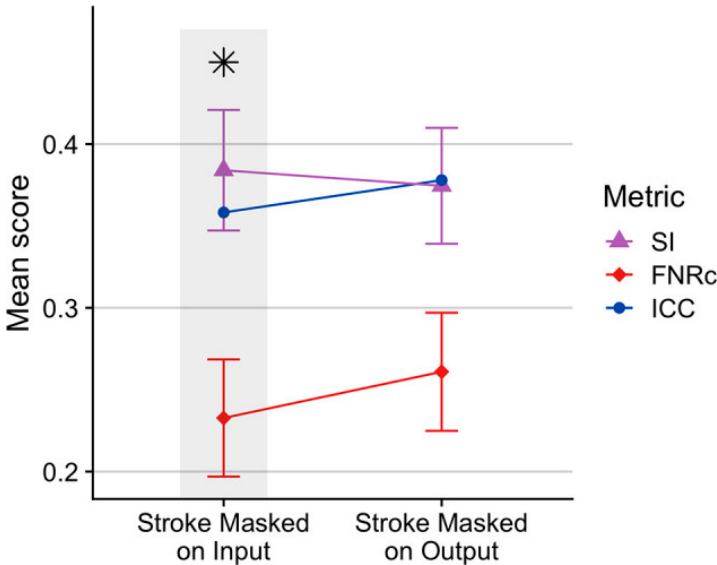


Figure 1. Dice and HD95 for all models on WMH2020 and ISLES2018.

3.2. Small-Lesion Sensitivity and Lesion-Level Results

For lesions smaller than 20 voxels, MicroLesion-SAM increased recall from 0.684 to 0.812, an 18.7% gain. Lesion-wise F1 also rose by 11.2%. These results indicate better sensitivity without a major increase in false positives [13,14].

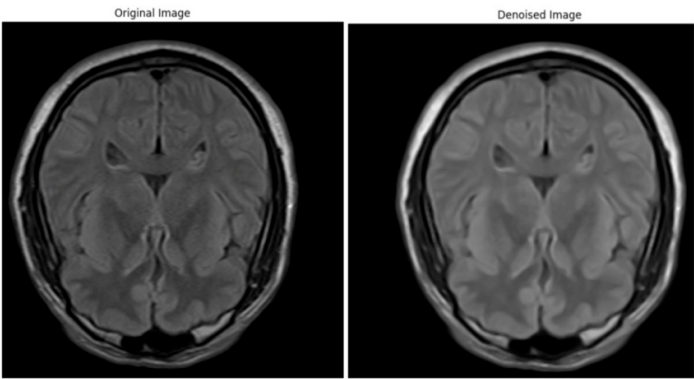


Figure 2. Lesion-wise recall and F1 for different lesion sizes.

3.3. Cross-Dataset Testing and Clinical Relevance

We next evaluated the models across datasets. When trained on one public dataset and tested on another, MicroLesion-SAM showed smaller performance drops than the baselines. On the clinical cohort of 72 subjects, the model improved small-lesion detection by 12.1%, while keeping global Dice and HD95 close to in-domain results. This finding is important because WMH2020, ISLES2018, and

the clinical scans differ in scanner type, sequence settings, and patient characteristics. Earlier work has reported clear performance loss under such shifts, especially for small or faint lesions [15,16]. Manual review showed that many new detections were punctate lesions near deep perforating arteries, which are clinically meaningful [17]. Some false positives remained in areas near perivascular spaces and noise, suggesting that additional refinement is needed.

3.4. Ablation Analysis and Relation to Foundation-Model Research

Ablation tests showed that both design choices—centerline cues and the scale-adaptive step—are important. Removing centerline information caused a 7.3% decrease in Dice and reduced small-lesion recall. Removing the scale-adaptive step led to higher HD95 and weaker responses to tiny lesions, indicating loss of fine detail. These findings agree with earlier work showing that structural cues and topology-aware methods help detect narrow or small structures [18]. They also extend observations from SAM-based medical segmentation studies, where the models perform well on large targets but often miss very small ones. Compared with approaches that adjust SAM mainly through prompt strategies or encoder fine-tuning, MicroLesion-SAM introduces explicit structural cues and size-aware processing steps that directly improve micro-lesion detection [19]. This study still has limitations. The datasets include several centers but do not cover very broad protocol variation. The current model uses MRI only. Future work should include larger multi-center datasets, test additional imaging modalities, and develop lighter models for faster clinical use.

4. Conclusions

This study introduced MicroLesion-SAM, a method developed to improve the segmentation of very small brain lesions on MRI. By adding centerline cues and a simple size-aware processing step, the model detected more small lesions and produced clearer boundaries than the baseline foundation model and conventional networks. Tests on WMH2020, ISLES2018, and an independent clinical set showed higher recall for lesions under 20 voxels and stable performance across different scanners. These results indicate that structural cues can help foundation models respond better to low-contrast, small-volume lesions, which are often overlooked in routine MRI reading. The study has several limits. The datasets do not cover the full range of clinical protocols, and the method was tested only on MRI. Future work should include larger multi-center cohorts, examine other imaging sequences, and explore faster model designs that can be used in clinical settings.

References

1. Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., & Lladó, X. (2019). Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Computers in biology and medicine*, 115, 103487.
2. Roh, J., Ryu, D., & Lee, J. (2024). CT synthesis with deep learning for MR-only radiotherapy planning: a review. *Biomedical Engineering Letters*, 14(6), 1259-1278.
3. Tian, Y., Yang, Z., Liu, C., Su, Y., Hong, Z., Gong, Z., & Xu, J. (2025). CenterMamba-SAM: Center-Prioritized Scanning and Temporal Prototypes for Brain Lesion Segmentation. *arXiv preprint arXiv:2511.01243*.
4. Wang, Y. (2025). Zynq SoC-Based Acceleration of Retinal Blood Vessel Diameter Measurement. *Archives of Advanced Engineering Science*, 1-9.
5. Sasidharan, S. M., Mdletshe, S., & Wang, A. (2025). Machine Learning in Stroke Lesion Segmentation and Recovery Forecasting: A Review. *Applied Sciences*.
6. Gui, H., Zong, W., Fu, Y., & Wang, Z. (2025). Residual Unbalance Moment Suppression and Vibration Performance Improvement of Rotating Structures Based on Medical Devices.
7. Hassanzadeh, T., Sachdev, S., Wen, W., Sachdev, P. S., & Sowmya, A. (2025). A robust deep learning framework for cerebral microbleeds recognition in GRE and SWI MRI. *NeuroImage: Clinical*, 103873.

8. Zha, D., Gamez, J., Ebrahimi, S. M., Wang, Y., Verma, N., Poe, A. J., ... & Saghizadeh, M. (2025). Oxidative stress-regulatory role of miR-10b-5p in the diabetic human cornea revealed through integrated multi-omics analysis. *Diabetologia*, 1-16.
9. Gurav, U., & Jadhav, S. (2025). Prompt-SAM: A Vision-Language and SAM based Hybrid Framework for Prompt-Augmented Zero-Shot Segmentation. *Human-Centric Intelligent Systems*, 1-19.
10. Wang, Y., Wen, Y., Wu, X., Wang, L., & Cai, H. (2025). Assessing the Role of Adaptive Digital Platforms in Personalized Nutrition and Chronic Disease Management.
11. Chaitanya, K., Erdil, E., Karani, N., & Konukoglu, E. (2020). Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in neural information processing systems*, 33, 12546-12558.
12. Wen, Y., Wu, X., Wang, L., Cai, H., & Wang, Y. (2025). Application of Nanocarrier-Based Targeted Drug Delivery in the Treatment of Liver Fibrosis and Vascular Diseases. *Journal of Medicine and Life Sciences*, 1(2), 63-69.
13. Lumetti, L., Pipoli, V., Marchesini, K., Ficarra, E., Grana, C., & Bolelli, F. (2025). Taming Mambas for 3D Medical Image Segmentation. *IEEE Access*.
14. Chen, D., Liu, S., Chen, D., Liu, J., Wu, J., Wang, H., ... & Suk, J. S. (2021). A two-pronged pulmonary gene delivery strategy: a surface-modified fullerene nanoparticle and a hypotonic vehicle. *Angewandte Chemie International Edition*, 60(28), 15225-15229.
15. Biswas, M., Rahman, S., Tarannum, S. F., Nishanto, D., & Safwaan, M. A. (2025). Comparative analysis of attention-based, convolutional, and SSM-based models for multi-domain image classification (Doctoral dissertation, BRAC University).
16. Li, W., Zhu, M., Xu, Y., Huang, M., Wang, Z., Chen, J., ... & Sun, X. (2025). SIGEL: a context-aware genomic representation learning framework for spatial genomics analysis. *Genome Biology*, 26(1), 287.
17. Zha, D., Mahmood, N., Kellar, R. S., Gluck, J. M., & King, M. W. (2025). Fabrication of PCL Blended Highly Aligned Nanofiber Yarn from Dual-Nozzle Electrospinning System and Evaluation of the Influence on Introducing Collagen and Tropoelastin. *ACS Biomaterials Science & Engineering*.
18. Kasaraneni, C. K., Guttikonda, K., & Madamala, R. (2025). Multi-modality Medical (CT, MRI, Ultrasound Etc.) Image Fusion Using Machine Learning/Deep Learning. In *Machine Learning and Deep Learning Modeling and Algorithms with Applications in Medical and Health Care* (pp. 319-345). Cham: Springer Nature Switzerland.
19. Pouramini, A., & Faili, H. (2024). Matching tasks to objectives: Fine-tuning and prompt-tuning strategies for encoder-decoder pre-trained language models. *Applied Intelligence*, 54(20), 9783-9810.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.