Quantitative Volumetric Analysis and Prediction Modeling of Hematoma Expansion in Patients with Intracerebral Hemorrhage

Lisa Guo, Jean Law, and Jun Won Park

Abstract

Intracerebral Hemorrhage (ICH) presents a significant health concern as the second most common cause of stroke, with a high mortality rate. Timely medical intervention is crucial to mitigate complications such as hematoma expansion. Computed Tomography (CT) stands as the primary diagnostic tool for trauma-related hemorrhage. While existing segmentation models like U-Net have shown to be effective, they often demand extensive data and computational resources, limiting their generalizability. In 2023, MedSAM emerged as a promising solution, a universal medical image segmentation model pre-trained on Meta AI's Segment Anything Model (SAM). Our study aims to leverage MedSAM's capabilities by fine-tuning LiteMedSAM, a knowledge-distilled variant, for segmenting hematomas in CT scans from 345 ICH patients. We employed Rank Stabilized Low-Rank Adaptation (rsLoRA) to refine LiteMedSAM's performance on pre-processed CT scans. Our proposed model surpasses established benchmarks such as Residual Attention U-Net (ResAttU-Net), MedSAM, and LiteMedSAM, achieving a notable improvement of 8% to 22% Sørensen–Dice index on test set. The segmentation results provide a basis for calculating hematoma volume, facilitating clinicians in diagnosing ICH patients with hematoma effectively.

Keywords: computed tomography, intracerebral hemorrhage, hematoma expansion, ResAttU-Net, MedSAM, LiteMedSAM, LoRA, rsLoRA

I. Background

Intracerebral Hemorrhage (ICH) refers to a medical condition caused by the tearing of a cerebral vessel [1]. This tearing allows blood to enter the brain tissue and the bleeding may collect and form a clot of blood called a hematoma [2]. As the hemorrhage progresses, more blood clots may form within the brain tissue, causing the hematoma to expand.

Hematoma expansion affects approximately $20\% - 40\%$ of patients with Intracerebral Hemorrhage (ICH), and research has shown that hematoma expansion serves as an important predictor of poor prognosis [3],[4]. Large volumes of hematoma expansions are associated with worse treatment outcomes and are shown to have increased mortality and disability [5]. As a consequence, assessing hematoma volume (HV) remains a vital task for clinicians as clinicians can use this information to identify patients at high risk and determine appropriate treatment strategies to improve treatment outcomes [6].

Hence, our team aims to develop a robust and accurate deep learning model to segment hematoma expansion in CT scans of patients with intracerebral hemorrhage which will allow us to assess hematoma volume.

Our model will help clinicians segment hematoma, improving operation success rates, and guiding treatment strategies for better outcomes, potentially reducing mortality and disability among patients.

II. Previous Work

A. History

Initially, the Fuzzy Clustering Method (FCM), also known as the Soft K-means, in which images are classified into a predefined number of clusters [7] was the most important clustering method for ICH segmentation. The random forest algorithm was also successfully used by Muschelli et al. in 2017 [8].

B. Dimensional Reduction U-Net (DR-UNet)

Yu et al. (2021) developed a robust deep learning segmentation model, DR-UNet, for efficient and accurate HV analysis using CT scans [9]. DR-UNet employs reduced dimensional residual convolution units, offering deeper architecture and larger receptive fields with fewer parameters. It was trained, validated, and tested on two datasets: one retrospective with $12,568$ CT slices from 512 ICH patients and one prospective with $1,257$ slices from 50 ICH patients. Additionally, the model's robustness was assessed on 13 irregular hematoma cases, 11 subdural and epidural hematoma cases, and 50 different HV cases divided into three groups.

The DR-UNet architecture is robust towards irregularly shaped hematomas, outperformed the FCM in every metric and also demonstrated significantly better performance than the traditional U-Net on most metrics.

C. Baseline Model

For our baseline, we have considered Yanting Yang's Residual Attention U-Net (ResAtt-UNet), a 2D segmentation model, trained on the ICH dataset (described in part IV.A). The Residual Attention U-Net combines residual connections and attention mechanisms, enhancing performance. Residual connections address vanishing gradients, while attention mechanisms improve focus on key input regions. The encoder path features repeated residual blocks with max pooling, and the decoder path utilizes convolution blocks and up-sampling. A concatenation layer connects encoding and decoding stages, with an attention module at the intermediate connection.

III. Method

The state-of-the-art models present in part II are inherently data-specific and require a large training dataset, rendering them incapable of generalizing to a different dataset of the same task. Foundational models like the Segment Anything Model (SAM) [10] address the challenge of generalization. Those models have the ability to perform well across a spectrum of tasks, due to their training on images encompassing different domains. Hence, we will be fine-tuning our dataset on a pre-trained model for better generalization capability.

A. MedSAM: Segment anything in medical images

MedSAM, fine-tuned on pre-trained SAM model, was introduced in 2023 as the first foundational model for universal medical image segmentation. Trained on one million medical image-mask pairs, 15 imaging modalities and 30 cancer types, MedSAM outperforms SOTA models like SAM and U-Net models on several metrics by 3.6% to 6.6% on different segmentation tasks [11]. The network architecture incorporates an image encoder, based on the Vision Transformer (ViT), responsible for extracting image features and image embedding, a prompt encoder, for integrating user interactions, and a mask decoder, for generating segmentation using the image embedding and prompt embedding.

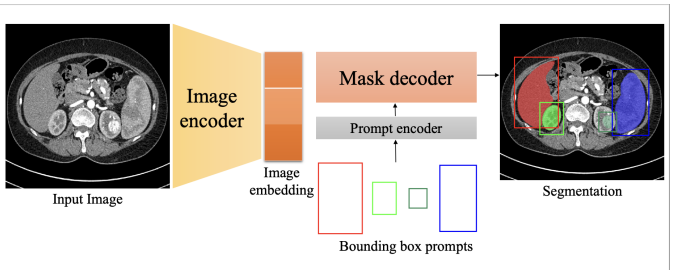

Fig. 1: MedSAM Architecture

B. LiteMedSAM

LiteMedSAM, a knowledge-distilled model from MedSAM, uses a smaller ViT instead of the base ViT, making it 10 times smaller and 10 times faster. However, internal testing showed an approximately 3% reduction in both Sørensen–Dice and Jaccard indices in comparison to MedSAM. Due to the computational limits, we will be using the pre-trained LiteMedSAM as our base model rather than MedSAM.

C. rsLoRA : Rank Stabilized Low Rank Adaptation

Fine-tuning on the target domain can be computationally expensive, especially if all the layers are used in the process. To address this limitation, we use Parameter-efficient fine-tuning (PEFT) and freeze the vast majority of the parameters to train only a reduced number of parameters. LoRA [12] was introduced as a simple yet effective parameter-efficient fine-tuning approach.

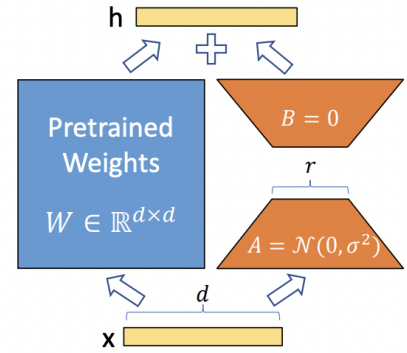

Fig. 2: LoRA Adapters Illustration

LoRA freezes the weights of the pre-trained base model and inserts trainable rank decomposition matricies into the desired transformer module [12], training only a subset of parameters. Another advantage of LoRA is that LoRA does not add any new vertical layers, allowing for more efficient inference than the traditional fine-tuning methods. Given some pre-trained weights $W \in \mathbb{R}^{d_1 \times d_2}$ and $b \in \mathbb{R}^{d_1}$, without the adapter, the model maps the input x_{in} to

$$x_{out} = Wx_{in} + b \tag{1}$$

Equation (1) is modified by rsLoRA, with parameters $r, \alpha \in \mathbb{N}$ and $A \in \mathbb{R}^{d_1 \times r}$, $B \in \mathbb{R}^{r \times d_2}$ and some scaling factor γ,

$$x_{out} = (W + \gamma AB)x_{in} + b \tag{2}$$

where the scaling factor is defined by

$$\gamma = \frac{\alpha}{\sqrt{r}}$$

which is theoretically proven to be optimal by Kalajdzievski [13].

IV. Experiments

A. Data Acquisition and Pre-processing

The dataset consists of 345 CT images of patients with ICH, in Neuro-imaging Informatics Technology Initiative (NIfTI) format. Each 3D image contains up to 64 grayscale slices along the axial axis, each with dimensions of 512×512. Initially, hematomas were not visible in the raw input data. Therefore, pre-processing techniques were applied before modeling. As our main task is segmentation and not object detection, we chose to remove small hematomas with less than 100 pixels, which accounts for less than 0.04% of total pixels, in the 2D slices. Moreover, as MedSAM and LiteMedSAM require a bounding box marking the region of interest, only the non-zero slices were considered. We applied intensity clipping to each of the CT images. The employed minimum and maximum intensities were 30 and 100 respectively, equivalent to the window width and level values of $(W : 70, L : 65)$. Subsequently, the intensity values were scaled to the range of $[0, 1]$, using min-max scaler. Then, bounding boxes were generated using the masks, which were considered as the ground truth. Finally, to meet LiteMedSAM's input requirements, the channel was repeated three times. Similarly, each slice, corresponding mask and box were resized to 256×256. These standardization processes ensured uniformity across all images. For training, the dataset was split into 9 to 1 ratio for training and validation respectively and an independent held-out test set was used.

B. Fine-tuning Setup

We fine-tuned our model on a single T4 16GB GPU. We used the Binary Cross Entropy with Logits Loss. Let l_i, g_i denote the logits and ground truth of pixel i, respectively. N is the number of pixels in the image.

$$L_{BCEL} = -\frac{1}{N}\sum_{i=1}^{N}[g_i log(\sigma(l_i)) + (1-g_i) log(1-\sigma(l_i))] \tag{3}$$

We fine-tuned the model using a cosine annealing learning rate scheduler with an initial learning rate of 0.0005. We also used AdamW optimizer with $\beta = (0.9, 0.999)$, weight decay of 0.01. A batch size 8 with gradient accumulation steps of 8 was employed. Dropout rates of $0, 0.1$, and 0.25 were tested. For rank (r) values of $4, 8, 16, 32$ and 64, corresponding α values of $8, 16, 32, 64$ and 128 were experimented with. Low-ranked matrices were integrated into transformer modules, more specifically, into the image encoder and the query and value matrices of the mask decoder.

C. Results

We have found that fine-tuning LiteMedSAM with LoRA is extremely efficient, as only 8% of the total number of parameters were trained. We achieved the best performing model with hyper-parameters rank $r = 64$, alpha $\alpha = 128$, and dropout rate of 0.1.

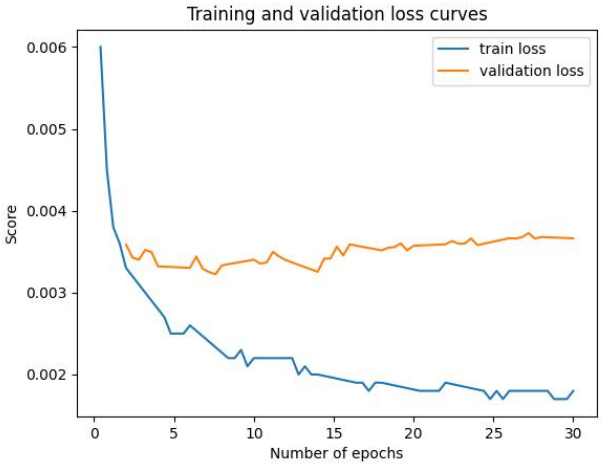

Fig. 3: Train and validation loss curves

Given the loss curves in Figure 3, the validation loss remains more or less constant at 0.0035 ± 0.0002. Hence, we retrieved the best checkpoint in training, based on the validation Sørensen–Dice and Jaccard indices, at epoch 24.

TABLE I: Validation and Test Performance

	Validation		Test	
Model	Dice	Jaccard	Dice	Jaccard
ResAtt-UNet	0.6178	0.4469	0.7070	0.5468
LiteMedSAM	0.7842	0.6755	0.8103	0.7127
MedSAM	0.8126	0.7005	0.8469	0.7499
LiteMedSAM + LoRA	**0.9307**	**0.8747**	**0.9249**	**0.8725**
DR-UNet*	-	-	0.874	0.794

*DR-UNet was tested on an external dataset.

LiteMedSAM with LoRA outperformed all the baseline models, ResAtt-UNet, LiteMedSAM and MedSAM, in all metrics in both test and validation datasets. For the test dataset, LiteMedSAM with LoRA achieved a high Sørensen–Dice index of 0.9249, which is 30% higher compared to ResAtt-UNet (0.7070) and $10\% - 15\%$

higher compared to LiteMedSAM (0.8103) and MedSAM (0.8469). It also appears to achieve a better performance than the state-of-the-art DR-UNet model that was trained for hematoma segmentation mentioned in Section II-B.

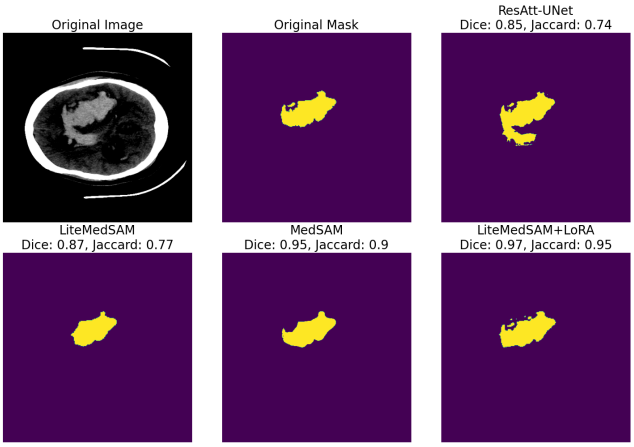

Fig. 4: Visualization of Segmentations across Models

LiteMedSAM with LoRA is able to accurately segment irregular-shaped hematomas, notably those with curvy and long borders that have traditionally posed challenges for computer vision models. In addition, LiteMedSAM with LoRA is more accurate in segmenting non-contiguous hematomas (i.e. with possible holes in the middle).

More generally, we have made the following observations:

1) ResAtt-UNet is capable of capturing segmentation details effectively but exhibits slight noise. Additionally, the model struggles with detecting small hematomas and occasionally over-segments, misclassifying negative regions that directly surround the hematoma.
2) LiteMedSAM and MedSAM are both consistent in detecting and segmenting the overall shape of the hematoma, but are unable to pick up on the details of the hematoma.
3) LiteMedSAM with LoRA identifies both the overall shape and details of the hematoma very effectively.

V. Discussion

Using our segmentation's results, we are able to compute the hematoma's volume of each patient, using 3D Slicer software's *Segment statistics* module. The values are computed by multiplying the number of voxels in the segment and the volume of a single voxel.

Due to computational resources' limitations, we were unable to fine-tune MedSAM, which would most likely yield better results given it is a bigger model. For the same reason, we have only attempted 2D segmentation, but 3D segmentation can also be performed. Moreover, we have also found that, as expected, a higher rank for the LoRA model generally leads to better performance but this improvement in performance decreases significantly at ranks 64 to 128 for which the Sørensen–Dice increased by only 0.003. We were unable to experiment with higher values of rank and alpha parameters, due to memory limits.

In terms of data pre-processing, additional pre-processing steps like skull stripping or de-noising might also further improve our results.

Finally, both MedSAM and LiteMedSAM present the limit of requiring bounding boxes to perform segmentation. In fact, we have attempted to fine-tune MedSAM for automatic segmentation but this resulted into poor performances. Hence, our final model is designed for clinicians, requiring manual bounding box for unlabeled inputs.

VI. Conclusion

We have explored the power of a pre-trained model for medical imaging segmentation. Unlike prior works that relied on variants of U-Net and training the model from scratch, we have used a parameter-efficient fine-tuning approach to effectively reduce the number of trainable parameters and training time. This simple, yet scalable, solution works very well, even surpassing the state-of-the-art models, while remaining cost-effective and easy to train.

While the initial results are promising, several challenges remain for clinical implementation. External validation is required to assess its effectiveness of generalization on diverse datasets. Nevertheless, our outcomes indicate the potential of our method. We anticipate further exploration and enhancement of our approach to enhance treatment outcomes for ICH patients with hematoma expansion.

Code availability. The pre-processing script, training script, and inference script have been made publicly available on Github.

Acknowledgments. We thank the Columbia University's Biomedical Engineering Department, the teaching team of BMENE4460, and Columbia University Irving Medical Center for providing the dataset. We would like to especially thank Dr. Jia Guo for advising our team, and Yanting Yang for providing his work. We also thank Meta AI and BoWangLab for making the source code publicly available.

REFERENCES

[1] L. Puy, A. R. Parry-Jones, E. C. Sandset, D. Dowlatshahi, W. Ziai, and C. Cordonnier, "Intracerebral haemorrhage," *Nature Reviews Disease Primers*, vol. 9, no. 1, p. 14, 2023.

[2] National Cancer Institute, "Hematoma." https://www.cancer.gov/publications/dictionaries/cancer-terms/def/hematoma. Accessed: April 1, 2024.

[3] S. Chen, B. Zhao, W. Wang, L. Shi, C. Reis, and J. Zhang, "Predictors of hematoma expansion predictors after intracerebral hemorrhage," *Oncotarget*, vol. 8, no. 51, p. 89348, 2017.

[4] Z. Li, Q. He, and B. Hu, "Hematoma expansion in intracerebral hemorrhage: an update on prediction and treatment," *Frontiers in neurology*, vol. 11, p. 537544, 2020.

[5] D. Dowlatshahi, A. Demchuk, M. Flaherty, M. Ali, P. Lyden, and E. Smith, "Defining hematoma expansion in intracerebral hemorrhage: relationship with patient outcomes," *Neurology*, vol. 76, no. 14, pp. 1238–1244, 2011.

[6] D. Haupenthal, S. Schwab, and J. B. Kuramatsu, "Hematoma expansion in intracerebral hemorrhage–the right target?," *Neurological Research and Practice*, vol. 5, no. 1, p. 36, 2023.

[7] P. Singh, V. Khanna, and M. Kamal, "Hemorrhage segmentation by fuzzy c-mean with modified level set on ct imaging," in *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 550–555, 2018.

[8] J. Muschelli, E. M. Sweeney, N. L. Ullman, P. Vespa, D. F. Hanley, and C. M. Crainiceanu, "Pitchperfect: primary intracranial hemorrhage probability estimation using random forests on ct," *NeuroImage: Clinical*, vol. 14, pp. 379–390, 2017.

[9] N. Yu, H. Yu, H. Li, N. Ma, C. Hu, and J. Wang, "A robust deep learning segmentation method for hematoma volumetric detection in intracerebral hemorrhage," *Stroke*, vol. 53, 10 2021.

[10] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.

[11] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, p. 654, 2024.

[12] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021.

[13] D. Kalajdzievski, "A rank stabilization scaling factor for fine-tuning with lora," *arXiv preprint arXiv:2312.03732*, 2023.