

Morph-RAG-SAM: Few-Shot Medical Image Segmentation via Hybrid Semantic–Morphological Re-Ranking and Prompt-Guided Refinement

Lan-Phuong Nguyen Ngoc¹, Thanh-Danh Nguyen^{†2,3}, and Vinh-Tiep Nguyen^{2,3}

¹Bosch Global Software Technologies Vietnam, Ho Chi Minh City, Vietnam

²University of Information Technology, Ho Chi Minh City, Vietnam

³Vietnam National University, Ho Chi Minh City, Vietnam

phuong.nguyennngoclan2@vn.bosch.com, {danhnt, tiepvn}@uit.edu.vn, [†]corresponding author

Abstract—Recent few-shot medical image segmentation models have demonstrated remarkable capabilities to generate precise masks given minimal user guidance. However, this reliance on user-provided precise prompts for optimal performance represents a limitation in their applicability to automate high-throughput clinical workflows where manual intervention is impractical. Currently, retrieval-augmented approaches address this process by finding similar examples in a support set. Nonetheless, they face a critical challenge that the best semantic matching is often not the best morphological matching, leading to inaccurate prompt generation and poor segmentation. In this work, we propose Morph-RAG-SAM, a training-free retrieval-based framework for few-shot medical image segmentation. Our key contribution is a hybrid re-ranking module fusing a robust semantic similarity metric, i.e., Mahalanobis distance, with a morphological similarity score, i.e., dice coefficient, to select superior support exemplary samples. This exemplar is then used to generate a multi-point positive prompt for an initial coarse mask prediction. Then, this initial mask is analyzed to create a prompt constructed from positive and negative points to guide the final refinement step. We successfully demonstrate our contributions through extensive experiments on three public medical datasets: PH2, HAM10000, and ISIC2018. Our proposal significantly improves upon a strong semantic-only retrieval baseline, achieving an increase in Dice score of 15.64% on PH2, 13.42% on HAM10000, and 16.95% on ISIC2018. Code can be found at <https://github.com/danhntd/Morph-RAG-SAM>.

Index Terms—few-shot medical image segmentation, free-training, geometric prompting, morphological RAG.

I. INTRODUCTION

Medical image segmentation [1] is the cornerstone of modern clinical diagnostics, allowing quantitative analysis and treatment planning for a wide range of conditions. The dominant paradigm in recent years has been the use of deep learning models, particularly U-Net [2] and its variants [3]–[5], which are fine-tuned on large, expertly annotated datasets. Although these models can achieve expert-level performance, their reliance on extensive task-specific training data creates a significant bottleneck. The process of curating these datasets is notoriously expensive and time-consuming,

limiting the scalability of this approach across the diverse landscape of medical imaging tasks.

The recent advent of large-scale, pre-trained foundation models, such as the Segment Anything Model (SAM) [6] and its medically specialized successor [7] has introduced a powerful new paradigm: few-shot segmentation. These models, trained on massive, diverse general datasets, can segment novel objects without any task-specific fine-tuning, often guided by simple geometric prompts including points or boxes. Without expert, human-in-the-loop prompting, these models often struggle with the inherent ambiguities of medical imaging, such as low-contrast boundaries and complex anatomical structures, limiting their reliability in fully automated clinical workflows.

This work operates in the critical space between these two paradigms, with the aim of harnessing the power of foundation models within a training-free, few-shot framework. We build upon the retrieval-augmented approach, which uses a small, annotated support set to automatically provide context to a foundation model at inference time. Our central thesis is that the primary limitation of current retrieval-augmented methods is their reliance on simple similarity metrics that fail to distinguish between semantic content and morphological shape. A support image may be the best semantic match for a query but have a completely different size or orientation, leading to the generation of a flawed geometric prompt and a poor final segmentation.

To address this, we propose a novel retrieval and refinement framework, dubbed Morph-RAG-SAM, that intelligently guides a pre-trained SAMMed2D model [7] to address few-shot medical image segmentation. Our main contributions are threefold:

- **First, a Hybrid Re-ranking Module:** We introduce a sophisticated re-ranking mechanism that moves beyond simple similarity search. It fuses a distribution-aware semantic metric, i.e., Mahalanobis distance [8], with a direct measure of morphological similarity, i.e., Dice score [9], to

select a superior support exemplar that is a strong match in both semantic content and mask shape.

- Second, a Consensus-Based Coarse Prediction: To debias the reranking process, we first generate a stable coarse prediction for the query image using a consensus prompt derived from the top-5 similar semantic support images. These coarse masks serve as more reliable shapes proxy for subsequent morphological comparison.
- Third, a Fully Automated Training-Free Pipeline: Our complete framework operates entirely at inference time, requiring no further model fine-tuning. We demonstrate through experiments on three challenging skin lesion datasets that our method significantly outperforms a strong semantic-only retrieval baseline, providing a more accurate and robust solution for few-shot medical image segmentation. In detail, we achieve an increase over the baseline model measured in Dice score of 15.64% on PH2 [10], 13.42% on HAM10000 [11], and 16.95% on ISIC2018 [12].

II. RELATED WORKS

The development of large-scale foundation models has significantly impacted the field of medical image analysis with The Segment Anything Model (SAM) [6] and its medically-specialized successor [7].

SAMMed2D [7] comprehensively fine-tunes the encoder of SAM [6], utilising adapter layers for lower computational cost and decoder, presents a comprehensive study on applying SAM [6] to medical 2D images [13]. While effective for common organs and lesions, the model does not yet possess the full capability in the medical domain, and there remains a gap in training data volume and resolution compared to the original SAM [6].

MedSAM [14] demonstrates superior accuracy and robustness over the original SAM [6], even outperforming some modality-specific models. Its performance can be affected by imbalances in training data, and it exhibits difficulty in segmenting complex structures, leading to persistent challenges in creating a universal medical segmentation model.

RAG-SAM2 [15] leverages DINOv2 [16] and Segment Anything Model 2 (SAM2) [17] for medical image segmentation without requiring any retraining or finetuning on the target domain. However, a common limitation of these approaches is their reliance on simple similarity metrics as L2 distance and their use of only the retrieved sample to guide the segmentation lead to the support image that is the best semantic match may have a significantly different shape or size so that generation of a poor geometric prompt which can hinder the ability of DINOv2 [16] to effectively measure global similarity this remarkable generalization comes

at the cost of precision and struggles with small region segmentation and low-contrast vessel segmentation, a limitation attributed to its reliance on whole-image features and SAM2 inherent capabilities [17].

This has spurred research into few-shot segmentation methods that can leverage a small support set of annotated examples to guide the model automatically. Our work directly addresses these limitations by introducing a more robust re-ranking mechanism that considers both semantic and morphological similarity, thus enhancing the baseline logic of RAG-SAM2.

III. OUR PROPOSED METHOD - MORPH-RAG-SAM

We propose a training-free framework designed to enhance the performance of a pre-trained SAMMed2D [7] model. The framework consists of an offline indexing stage and a two-stage online inference pipeline comprising retrieval and refinement. The core of our contribution is a hybrid re-ranking module that selects a superior support exemplar by concurrently evaluating both semantic and morphological similarity. The complete workflow of our Morph-RAG-SAM is visualized in [Figure 1](#).

A. Stage 1: Initial Retrieval and Consensus-Based Coarse Segmentation

The initial stage of our pipeline is dedicated to generating a robust, coarse segmentation of the target region. To this end, we first employ a pre-trained DINOv2 [17] model as a universal feature extractor to generate a semantic embedding for every image in the support set. These embeddings are subsequently indexed into a FAISS [9] vector database to facilitate efficient similarity-based retrieval. Then we find initial candidate retrieval by extracting feature embedding of given query image using its DINOv2 [17] embed model and use FAISS [9] to retrieve the top-K most semantically similar support images based on L2 distance.

To mitigate the bias of relying on a single and potentially mismatched support image, we introduce a consensus-based prompting strategy. This involves aggregating information from the top-K retrieved support masks. Specifically, we generate bounding box and point prompts for each of these top-K masks and then fuse them by taking the union of the bounding boxes and the mean of the center points. Finally, this new, more stable consensus prompt is used to guide the SAMMed2D [14] predictor to generate an initial predicted mask. This resulting mask represents our initial hypothesis for the segmentation and serves as the primary input for the subsequent re-ranking stage.

B. Stage 2: Hybrid Re-ranking and Iterative Refinement

The second stage of our pipeline is designed to correct the potential morphological inaccuracies of the initial predicted mask generated in Stage 1. This is

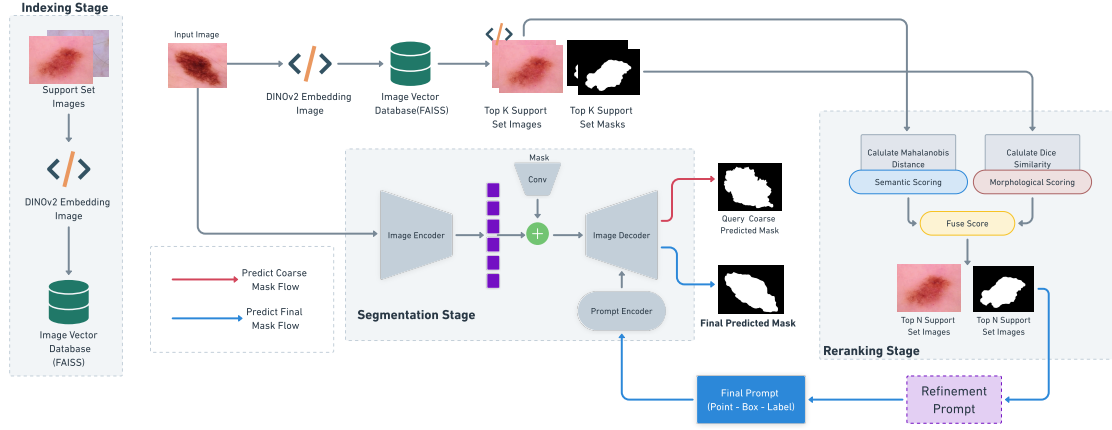


Fig. 1. Overview of our proposed Morph-RAG-SAM framework. During the indexing stage, a collection of support images is processed to build a reference. In the inference stage, a given input image is first converted into a DINOv2 embedding and segmentation pretrained model. This embedding is used to query the FAISS [9] database, retrieving the top-K most similar support set images and their corresponding masks. These masks will first be used to generate a coarse predicted mask. The coarse segmentation mask generated serves as the input for our hybrid re-ranking module. The support images are then re-ranked based on the fused score leads to a new prompt being generated from the top-ranked support exemplary samples. Then the input image is passed through an image encoder, and its output is combined with the final support masks via a convolutional layer before being fed into an image decoder. The decoder then generates the final predicted segmentation mask for the input image database.

achieved through a novel hybrid re-ranking mechanism followed by a two-step, prompt-based refinement process that leverages the geometric properties of the coarse mask itself to achieve a precise final segmentation sentence, as in Equation 1:

$$S_{\text{hybrid}}^{(i)} = (1 - \alpha) \cdot \text{Norm}(D_M(\mathbf{f}_q, \mathbf{f}_i))^{-1} + \alpha \cdot \text{Dice}(M_{\text{coarse}}, M_i), \quad (1)$$

where:

- M_{coarse} is the consensus-based coarse mask of the query image from Stage 1, and M_i is the ground-truth mask of the i -th support candidate.
- \mathbf{f}_q and \mathbf{f}_i are the DINOv2 [17] embeddings for the query and the i -th candidate, respectively.
- $D_M(\mathbf{f}_q, \mathbf{f}_i)$ is the Mahalanobis distance [18] between the embeddings, which accounts for the covariance of the feature distribution, offering a more robust semantic similarity metric than the L_2 distance. We use the inverse of the normalized distance so that higher scores are better.
- $\text{Dice}(M_{\text{coarse}}, M_i)$ is the Dice Similarity Coefficient [8], which serves as a direct measure of morphological or shape similarity.
- $\alpha \in [0, 1]$ is a hyperparameter balancing the contribution of semantic and morphological similarity.
- The candidate with the highest S_{hybrid} score is selected as the new top-1 exemplar, as it represents the optimal trade-off between semantic relevance and shape congruency.

1) *Exemplar-Guided Multi-Point Prompting*: Following the re-ranking process, the top-ranked support exemplar is utilized to generate a robust set of prompts for the final segmentation refinement. This multi-point prompting strategy is designed to be more resilient

to complex object morphologies compared to using a single centroid. The first phase of this process involves computing a minimal bounding box to tightly enclose the largest contour of the selected exemplar’s ground-truth mask, which provides a precise spatial constraint for the segmentation task.

In the second phase, we identify the most representative positive prompt points by employing a Euclidean distance transform on the exemplar’s binary mask. This transform calculates, for each foreground pixel, the distance to the nearest background pixel, effectively identifying the internal skeleton of the object. The points with the highest distance transform values are the most central to the object’s geometry. The distance transform $D(p)$ for any point p within the mask’s set of foreground points, denoted as Ω , is formally defined as in Equation 2:

$$D(p) = \min_{q \in \partial\Omega} \|p - q\|_2, \quad (2)$$

where:

- $\partial\Omega$ represents the boundary of the mask, (i.e., the set of all points on the contour)
- p is a random point on that boundary
- $\|p - q\|_2$ is the Euclidean distance between the interior point p and the boundary point q .

The min operation ensures that $D(p)$ is the shortest distance from p to any point on the object’s boundary. The points corresponding to the local maxima of this distance function are selected as the positive points for the prompt. This new, refined prompt, consisting of the bounding box and multiple positive points, is then used to guide the SAM [6] predictor to produce the final, high-fidelity segmentation mask.

2) *Self-Corrective Boundary Refinement*: While the intermediate mask represents a significant improvement, it may still exhibit ambiguous or imprecise boundaries. The final stage of our framework is a self-correction mechanism that leverages this mask to generate a highly specific set of positive and negative prompts, thereby honing the final segmentation boundary. First, to reinforce the model’s confidence in the core object area, the distance transform technique described previously is reapplied to the intermediate mask to select a new set of positive points.

Subsequently, to explicitly correct edge imperfections, we generate negative points targeted at the most ambiguous region: the boundary itself. This is achieved through morphological erosion. Let M_{inter} be the intermediate mask and K be a structuring element (e.g., a 5×5 kernel). We compute the eroded mask $M_{\text{eroded}} = M_{\text{inter}} \ominus K$. The set of candidate pixels for negative prompts, P_{neg} , is the difference between the original mask and its eroded version, as in Equation 3:

$$P_{\text{neg}} = M_{\text{inter}} \setminus M_{\text{eroded}} \quad (3)$$

where:

- M_{inter} represents the set of pixels in the intermediate mask
- M_{eroded} is the set of pixels in the mask after the erosion operation has removed the outermost layer of pixels

This set constitutes a pixel-wide band around the mask’s perimeter. We then randomly sample the number of negative points (e.g., 10) from P_{neg} .

Finally, these positive points-reinforcing the interior, and negative points-constraining the exterior, are combined into a comprehensive prompt. This final prompt guides the model one last time to produce a high-precision segmentation, effectively forcing it to disambiguate the boundary and leading to a significant improvement in accuracy.

IV. EXPERIMENTAL RESULTS

A. Implementation Details

We conducted experiments on three publicly available skin lesion datasets: PH2 [11], HAM10000 [9], and ISIC2018 [19]. For each dataset, we established a fixed split to form our support and query sets:

- PH2 [12]: The full dataset of 200 images was split into two 100-image sets of support and query.
- HAM10000 [19]: The dataset of 10,015 images was split into a 7,200-image support set and a 1,015-image query set, with the remainder reserved as a validation set, mirroring common benchmark practices.
- ICIS2018 [12]: The training dataset of 2,594 images was split into an 1,594-image support set and a 1,000-image query set.

The support set for each dataset was used exclusively for building the vector database, and all performance metrics were calculated on the unseen query set.

Our framework is constructed upon two off-the-shelf foundation models, utilized without any fine-tuning. For feature extraction, we employ the DINOv2 [17] ViT-S/14 model pre-trained on ImageNet [10]. All support set images are resized to 224×224 pixels and subsequently normalized using the standard ImageNet statistics [10] before being processed by the feature extractor. For segmentation, we use the officially pre-trained model of SAMMed2D [7]. The SAMMed2D predictor [7] receives the original, un-normalized query image to perform the final segmentation.

Efficient nearest-neighbor search is facilitated by the FAISS [9] Vector Database [12]. For the Mahalanobis distance [18] calculation, the covariance matrix is computed using the robust Ledoit-Wolf [10] estimator. In our hybrid re-ranking module, the semantic and morphological scores are combined using a simple weighted average, where both components are assigned an equal weight of 0.5. All experiments were conducted on a single NVIDIA RTX A5000 GPU.

B. Quantitative Evaluation

To validate the efficacy of our proposed method - Morph-RAG-SAM, the comprehensive results of our comparison across the three aforementioned datasets are summarized in Table I, with qualitative examples provided in Figure 2.

As illustrated in Table I, our proposed Morph-RAG-SAM method demonstrates a clear and consistent superiority over all other baseline methods across all three datasets, on every evaluation metric, including Dice Similarity Coefficient (DSC), Sensitivity (SE), Specificity (SP), and Accuracy (ACC). Our Morph-RAG-SAM achieves the highest scores, underscoring its robustness and effectiveness.

C. Qualitative Comparison

Qualitative analysis of our refinement strategy reveals that it successfully corrects two primary failure modes observed in the baseline model: under-segmentation and over-segmentation. Under-segmentation occurred when the baseline model, prompted by a small support mask, failed to capture the full extent of a large ground truth lesion. Our hybrid re-ranker addressed this by selecting a more appropriately sized support mask, leading to a more complete final segmentation. Conversely, over-segmentation was a problem for small lesions, where the baseline would sometimes choose a support image with a much larger lesion. This resulted in a broad, inaccurate segmentation. Our hybrid re-ranker mitigated this by incorporating the shape of the coarse prediction, allowing it to select a smaller, morphologically similar

TABLE I
STATE-OF-THE-ART COMPARISON OF SEGMENTATION METHODS ACROSS DATASETS: PH2 [11], HAM10000 [9], AND ISIC2018 [19].

Method	ICIS2018 [19]				PH2 [11]				HAM10000 [9]			
	DSC (\uparrow)	SE (\uparrow)	SP (\uparrow)	ACC (\uparrow)	DSC (\uparrow)	SE (\uparrow)	SP (\uparrow)	ACC (\uparrow)	DSC (\uparrow)	SE (\uparrow)	SP (\uparrow)	ACC (\uparrow)
RAG-SAM2 [15]	0.7142	0.6805	0.9339	0.8507	0.7203	0.6470	0.9119	0.8279	0.7188	0.6550	0.9612	0.8821
MEDSAM [14]	0.6586	0.5944	0.9534	0.7624	0.614	0.5267	0.9295	0.7009	0.6336	0.6027	0.9730	0.8113
SAMMed2D [7]	0.7990	0.8458	0.9471	0.8984	0.8393	0.8168	0.9642	0.8886	0.8434	0.8428	0.9751	0.9382
Morph-RAG-SAM (ours)	0.8484	0.8610	0.9675	0.9200	0.8767	0.8432	0.9806	0.9038	0.8883	0.8451	0.9898	0.9512

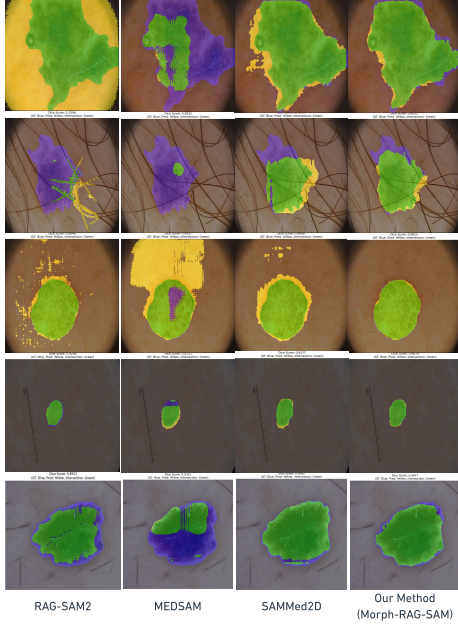


Fig. 2. Qualitative comparison of lesion segmentation results across different methods. From left to right: RAG-SAM2 [15], MEDSAM [14], SAMMed2D [7], and our proposed Morph-RAG-SAM. Green regions denote the intersection between ground truth and prediction, blue represents the ground truth (GT), and yellow indicates the predicted mask.

support image and producing a more precise final prediction. The comparison is visualized in Figure 3.

D. Ablation Study

Evaluation of Top-K retrieved samples. We evaluate the impact of the number of retrieved examples (Top-K) and the number of positive point prompts used for segmentation refinement. All experiments are performed across the HAM10000 [19], ISIC2018 [11], and PH2 [12] datasets. First, we investigated the optimal value for ‘K’ in the Top-K retriever, testing values of 20, 30, and 40. The results presented in Table II show that setting $K = 40$ consistently yields the best performance. On all three benchmarks, $K = 40$ achieved the highest Dice Similarity Coefficient (DSC), scoring 0.8858 on HAM10000 [19], 0.8485 on ISIC2018 [11], and 0.8736 on PH2 [12]. This suggests that a larger context of 40 retrieved examples provides a more robust foundation for the segmentation process.

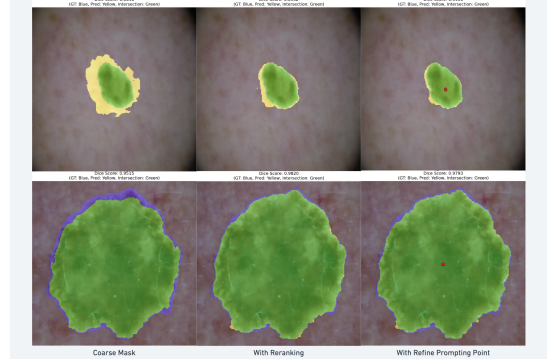


Fig. 3. Qualitative results showcasing the impact of hybrid re-ranking and prompt-based refinement. The first row illustrates the correction of over-segmentation, while the second row demonstrates the mitigation of under-segmentation.

Evaluation of the number of positive point prompts. We evaluated the influence of the number of positive point prompts during the refinement stage, while fixing the number of negative prompts to 10. We compared the use of 1, 3, and 5 positive points, which were selected from the top peaks of the distance transform inside the coarse mask, a strategy inspired by SAMMed2D [4]. As shown in Table III, we found a monotonic improvement in the Dice score as the number of positive prompts increased from 1 to 5. This trend was consistent across various organ sizes and shapes, with the most significant gains observed for elongated or multi-lobed structures. This can be attributed to positive points acting as strong inclusion anchors. Increasing the number of these points improves coverage of distinct subregions like lobes, tails, and cavities. This richer internal support helps prevent premature shrinkage and enhances recall, particularly for complex, non-convex shapes, thereby increasing the Dice score. While 5 positive points occasionally caused slight over-expansion into adjacent tissue for a few small/compact targets when the bounding box was loose, this was less common than the under-segmentation observed with 1-3 positives on irregular anatomy. The clustering of negative points could reduce boundary coverage, but overall, we did not observe performance degradation with 5 positive points.

TABLE II
DSC EVALUATION OF THE IMPACT OF TOP-K RETRIEVER WHEN
APPLYING JUST HYBRID RERANKING.

Top-K	HAM10000	ICSI2018	PH2
20	0.8835	0.8468	0.8721
30	0.8832	0.8290	0.8703
40	0.8858	0.8485	0.8736

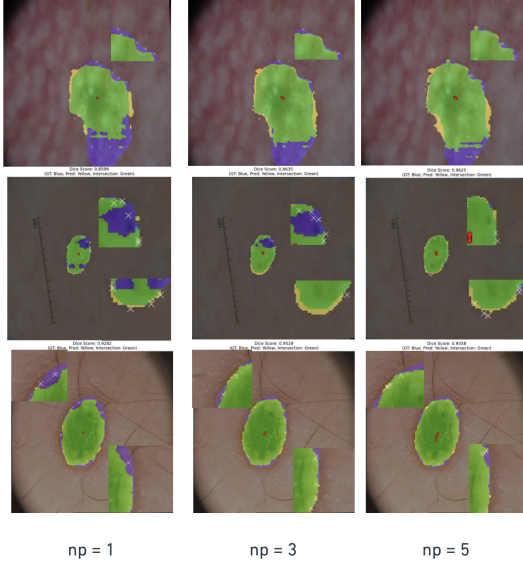


Fig. 4. Segmentation results with 1, 3, and 5 positive prompts (left to right). Ground truth is shown in blue, predictions in yellow, and overlaps in green; red dots indicate positive prompts and white crosses denote negative prompts. Zoomed views highlight boundary details. Increasing the number of prompts improves alignment, with 5-point prompting yielding the closest match to the ground-truth boundaries.

V. CONCLUSION

In this work, we have introduced Morph-RAG-SAM, a training-free retrieval-based framework for medical image segmentation given a few medical visual samples that addresses a key flaw in existing retrieval-augmented methods. Our key contribution is a hybrid re-ranking module fusing a robust semantic similarity metric with a morphological similarity score to select superior support exemplary samples. This exemplar is then used to generate a multi-point positive prompt for an initial coarse mask prediction. Finally, the initial mask is analyzed to create a prompt constructed from positive and negative points to guide the final refinement procedure. This approach generates more accurate prompts, leading to significant and consistent performance gains across three challenging datasets. Our future work may explore advanced feature fusion and the extension to 3D tasks.

VI. ACKNOWLEDGEMENTS

This research is funded by University of Information Technology - Vietnam National University Ho Chi Minh City under grant number D1-2025-31.

TABLE III
DSC EVALUATION OF THE IMPACT OF DIFFERENT COMPONENT
NUMBERS OF POINTS AS GEOMETRIC PROMPTING.

No. of Points	HAM10000	ICSI2018	PH2
1	0.8832	0.8290	0.8703
3	0.8970	0.8376	0.8823
5	0.9010	0.8418	0.8841

REFERENCES

- [1] W. Yao, J. Bai, *et al.*, “From cnn to transformer: A review of medical image segmentation models,” *Journal of Imaging Informatics in Medicine*, vol. 37, no. 4, pp. 1529–1547, 2024.
- [2] O. Ronneberger, P. Fischer, *et al.*, “U-net: Convolutional networks for biomedical image segmentation,” pp. 234–241, 2015.
- [3] N. Siddique, P. Sidike, C. Elkin, and V. Devabhaktuni, “U-net and its variants for medical image segmentation: theory and applications,” *arXiv preprint arXiv:2011.0111*, 2020.
- [4] W. Jiangtao, N. I. R. Ruhaiem, and F. Panpan, “A comprehensive review of u-net and its variants: Advances and applications in medical image segmentation,” *IET Image Processing*, vol. 19, no. 1, p. e70019, 2025.
- [5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *IEEE/CVF CVPR*, pp. 4700–4708, 2017.
- [6] A. Kirillov, E. Mintun, N. Ravi, H. Mao, and R. Girshick, “Segment anything,” in *IEEE/CVF CVPR*, pp. 4015–4026, 2023.
- [7] D. Cheng, Z. Qin, Z. Jiang, S. Zhang, Q. Lao, and K. Li, “Sam on medical images: A comprehensive study on three prompt modes,” *arXiv preprint arXiv:2305.00035*, 2023.
- [8] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [9] M. Douze, A. Guzhva, C. Deng, J. Johnson, *et al.*, “The faiss library,” *arXiv preprint arXiv:2401.08281*, 2024.
- [10] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *IEEE/CVF CVPR*, pp. 248–255, Ieee, 2009.
- [11] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, *et al.*, “Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic),” *arXiv preprint arXiv:1902.03368*, 2019.
- [12] T. Mendonça, M. Celebi, T. Mendonca, and J. Marques, “Ph2: A public database for the analysis of dermoscopic images,” *Dermoscopy image analysis*, vol. 2, 2015.
- [13] J. Ye, J. Cheng, J. Chen, Z. Deng, T. Li, *et al.*, “Sa-med2d-20m dataset: Segment anything in 2d medical imaging with 20 million masks,” *arXiv preprint arXiv:2311.11969*, 2023.
- [14] 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.
- [15] L. Zhao, X. Chen, E. Z. Chen, Y. Liu, T. Chen, and S. Sun, “Retrieval-augmented few-shot medical image segmentation with foundation models,” *IEEE Transactions on Neu. Net. and Lea. Sys.*, 2025.
- [16] W. Ehab and Y. Li, “Performance analysis of unet and variants for medical image segmentation,” *arXiv preprint arXiv:2309.13013*, 2023.
- [17] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, and P. Bojanowski, “Dinov2: Learning robust visual features without supervision,” *arXiv preprint arXiv:2304.07193*, 2023.
- [18] P. C. Mahalanobis, “On the generalized distance in statistics,” *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 80, pp. S1–S7, 2018.
- [19] P. Tschandl, C. Rosendahl, and H. Kittler, “The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific data*, vol. 5, no. 1, pp. 1–9, 2018.