


CONESCAPANHONDURAS2025paper136.pdf

 Institute of Electrical and Electronics Engineers (IEEE)

Document Details

Submission ID

trn:oid:::14348:477757937

Submission Date

Jul 31, 2025, 10:09 PM CST

Download Date

Aug 12, 2025, 6:30 PM CST

File Name

CONESCAPANHONDURAS2025paper136.pdf

File Size

435.9 KB

6 Pages





3,817 Words

21,801 Characters




20% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

-  **34 Not Cited or Quoted 13%**
Matches with neither in-text citation nor quotation marks
-  **7 Missing Quotations 2%**
Matches that are still very similar to source material
-  **5 Missing Citation 5%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 17%  Internet sources
- 18%  Publications
- 0%  Submitted works (Student Papers)

Integrity Flags





0 Integrity Flags for Review

No suspicious text manipulations found.




Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

-  **34 Not Cited or Quoted** 13%
Matches with neither in-text citation nor quotation marks
-  **7 Missing Quotations** 2%
Matches that are still very similar to source material
-  **5 Missing Citation** 5%
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted** 0%
Matches with in-text citation present, but no quotation marks

Top Sources

- 17%  Internet sources
- 18%  Publications
- 0%  Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	arxiv.org	8%
2	Internet	github.com	2%
3	Publication	Bowei Xue, Han Cheng, Qingqing Yang, Yi Wang, Xiaoning He. "Adapting Segmen...	1%
4	Publication	Gajraj Singh, Anand D. Darji, Jignesh N. Sarvaiya, Suprva Patnaik. "Preprocessing ...	1%
5	Internet	export.arxiv.org	<1%
6	Internet	www.tobaccoinaustralia.org.au	<1%
7	Internet	www.ejournal.org.cn	<1%
8	Internet	www.ieee-jas.net	<1%
9	Internet	www.neurodata.io	<1%
10	Publication	Thi Tan Tien Nguyen, Quoc Bao Bui, Cuong Pham. "A data-constrained approach f...	<1%

11	Publication	Nie, Yuqi. "Applications of Machine Learning in Time Series and Finance", Princet...	<1%
12	Internet	epub.uni-luebeck.de	<1%
13	Internet	eandv.biomedcentral.com	<1%
14	Publication	Aleksandar Miladinović, Alessandro Biscontin, Andrea Bonini, Francesco Bassi et ...	<1%
15	Publication	Iman Gandomi, Mohammad Vaziri, Mohammad Javad Ahmadi, M Reyhaneh Hadi...	<1%
16	Internet	radum.ece.utexas.edu	<1%
17	Internet	www.mdpi.com	<1%
18	Internet	www.tcsae.org	<1%
19	Publication	Banerjee, Soumyanil. "Enhancing Healthcare Informatics Through Deep Learning ...	<1%
20	Publication	Loris Nanni, Daniel Fusaro, Carlo Fantozzi, Alberto Pretto. "Improving Existing Se...	<1%
21	Publication	Mahani, Golnar Khalili Zadeh. "Interpretability and Annotation Scarcity in Deep M...	<1%
22	Publication	Wu, Chen. "Backdoor Attacks and Defenses in Federated Machine Learning", The ...	<1%
23	Publication	Yuanyuan Wang, Zheng Ding, Jiange Liu, Kexiao Wu, Md Sharid Kayes Dipu, Ting...	<1%
24	Internet	mftp.mmcheng.net	<1%

25	Internet	
www.geeksforgeeks.org		<1%
26	Internet	
www.iaarc.org		<1%
27	Publication	
Al-Marri, Maryam Salem M. A.. "Q-SAM: A Hybrid Quantum-Classical Segment Any...		<1%
28	Publication	
Truong, Thanh-Dat. "Towards Robust and Fair Vision Learning in Open-World Envi...		<1%

Few-Shot Cataract Detection via Feature Density Learning: Evaluating SAM Models and Backbone Embeddings

Abstract—Medical image segmentation has seen rapid advances in recent years, driven by the development of universal models such as Meta’s Segment Anything Model (Segment Anything Model (SAM)). Originally designed for general-purpose segmentation across a wide range of domains, SAM and its variants have shown potential for medical applications, particularly when combined with prompt-based guidance and zero-shot generalization. In this work, we present a comprehensive benchmark of multiple SAM-based models for cataract detection using anterior segment images—a domain that remains under-explored compared to fundus-based approaches. Our evaluation encompasses key SAM variants (SAM-Vision Transformer (ViT)-H, SAM-ViT-L, SAM-ViT-B), enhanced versions like HQ-SAM and MobileSAM, and widely used backbone networks, including ResNet-18, ResNet-34, and ViT-B/16. Furthermore, we incorporate feature density analysis using Kernel Density Estimation (KDE) to support few-shot learning (Few-Shot Learning (FSL)) scenarios and assess model robustness under potential out-of-distribution (Out-Of-Distribution (OOD)) conditions. The results provide new insights into the effectiveness of general-purpose segmenters in the ophthalmic domain and highlight the value of integrating density-based methods for improving diagnostic relevance in resource-constrained environments.

Index Terms—SAM, Medical Image Segmentation, Cataract Detection, Feature Density, Ophthalmology Artificial Intelligence (AI), FSL, OOD

I. INTRODUCTION

Cataract is a leading cause of reversible blindness worldwide, especially in underserved regions where access to specialists is limited [1]. While automated segmentation tools offer a scalable solution, ensuring generalization and accuracy in medical image analysis remains challenging. Meta’s Segment Anything Model (SAM) [2] introduced a prompt-based, zero-shot segmentation framework applicable across domains. Though effective in general contexts, its application to ophthalmology—particularly using anterior eye images—has been limited, as most studies focus on fundus imaging [3]. Variants like HQ-SAM [4] and MobileSAM [5] aim to improve segmentation quality and deployment efficiency, respectively. In this study, we benchmark three SAM variants (SAM-ViT-H, L, B) and their enhanced versions across different backbones—ResNet-18, ResNet-34 [6], [7], and ViT-B/16 [8]. Additionally, we propose a feature density analysis pipeline using Kernel Density Estimation (KDE) [9], suitable for few-shot learning (FSL) [10] and robust under out-of-distribution (OOD) conditions [11]. Our work delivers a systematic evaluation of foundation models for cataract detection, insights on

backbone effectiveness, and recommendations for lightweight, real-world deployments.

II. RELATED WORK

In the context of deep learning, the term *backbone* typically refers to the core feature extraction network within a larger model. Its primary function is to process raw input data and convert it into meaningful feature representations that can be utilized by subsequent layers. Commonly adopted backbones are based on well-established convolutional neural network (Convolutional Neural Network (CNN)) architectures, which have demonstrated strong performance on simpler tasks such as image classification. Due to their robustness and versatility, these architectures are frequently employed in more complex pipelines, including object detection, semantic segmentation, and medical image analysis [12].

In this study, we adopt ResNet-18, ResNet-34, and ViT-B-16 as backbone networks. ResNet-18, a lightweight residual network with 18 layers, has proven effective in content-based medical image retrieval, achieving an accuracy of 92% and a mean average precision (mean Average Precision (mAP)) of 0.90 on modalities such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) [6]. Similarly, ResNet-34, which provides deeper feature extraction capabilities, has been used successfully in segmentation tasks like chronic wound analysis, reaching an Intersection over Union (IoU) of 0.973 and a Dice score of 0.986 [7]. Additionally, transformer-based backbones such as ViT-B/16 have gained popularity for their ability to model long-range dependencies and have been integrated into architectures like TransUNet for improved multi-organ segmentation accuracy [8], [13].

Complementary to architectural advancements, the field has also seen increased interest in training paradigms that enhance generalization. One such area is *Out-of-Distribution* (OOD) detection, which focuses on managing inputs that differ from the training distribution. As shown by De Silva et al. [11], incorporating small amounts of OOD data can help improve model robustness. However, this relationship is non-linear, as excessive OOD data can ultimately degrade performance, emphasizing the need for careful management of such data.

In parallel, **Few-Shot Learning** (FSL) offers an alternative approach by enabling models to generalize from only a few labeled examples. Unlike traditional deep learning methods that require extensive labeled datasets, FSL relies on meta-learning strategies or prior knowledge to reduce both data and

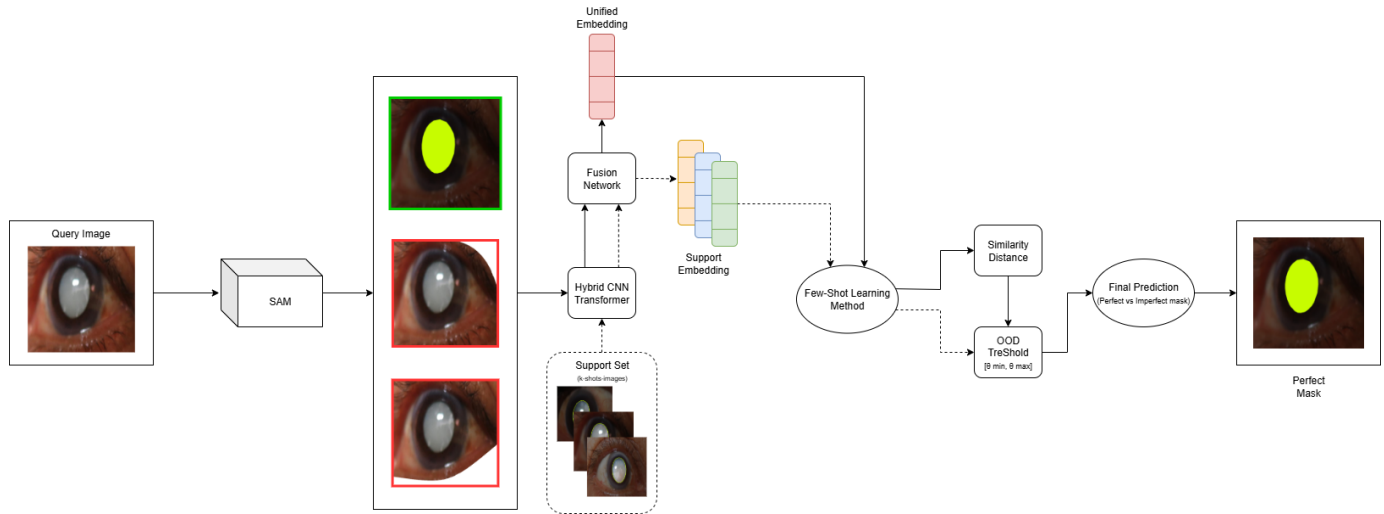


Fig. 1. High-level pipeline that pairs SAM-generated mask proposals with few-shot classification, incorporating a hybrid backbone and OOD thresholding.

computational demands. This makes it especially valuable in domains like medical imaging, where annotated data can be expensive and time-consuming to acquire [10].

Recent efforts in image segmentation have focused on developing general-purpose models that require minimal domain-specific adaptation. One prominent example is the *Segment Anything Model* (SAM), introduced as a universal segmentation framework capable of operating with various prompt types—such as points, boxes, or free-form text. SAM features a modular architecture that includes a heavyweight image encoder and a lightweight prompt encoder and mask decoder. This separation allows for efficient reuse of the image embeddings across different prompts, enabling near real-time segmentation with processing speeds of approximately 50 milliseconds per image in browser environments [2]. A core strength of SAM lies in its promptable design: it can generate valid segmentation masks even in ambiguous scenarios [3]. Its versatility extends across multiple tasks, including instance segmentation, edge detection, and object proposal generation. Furthermore, its zero-shot generalization capabilities allow it to adapt to new domains without retraining [3]. In the medical imaging domain, SAM has shown promising results when segmenting organs with clear boundaries in CT and MRI. However, its performance decreases when dealing with amorphous or irregular lesions and in scenarios where prompts are not provided. Studies have demonstrated that manual guidance, such as the use of point clicks or bounding boxes, improves its effectiveness [14]. For instance, in brain extraction tasks, SAM has outperformed traditional tools like BET [14]. To expand the capabilities of SAM, several pre-trained variants have been released, each based on different sizes of the Vision Transformer (ViT) backbone. These include SAM-ViT-H, SAM-ViT-L, and SAM-ViT-B, offering different trade-offs between computational efficiency and segmentation accuracy. Recognizing the need to improve mask precision, High-Quality SAM (HQ-SAM) was proposed as an enhanced

version of the original model [4]. HQ-SAM introduces a learnable High-Quality Output Token within the mask decoder, which helps refine the segmentation boundaries without compromising promptable inference or zero-shot generalization. The model is available in three backbone variants: HQ-SAM-ViT-BASE, HQ-SAM-ViT-BASE-TINY, and HQ-SAM-ViT-BASE-HUGE. In addition, MobileSAM was developed to address the demand for lightweight models suitable for mobile and edge devices [5]. It preserves the core prompt-based framework of SAM while significantly reducing size and latency. MobileSAM can process an image in approximately 10 milliseconds—split between 8 ms in the image encoder and 4 ms in the mask decoder—making it about five times faster and seven times smaller than FastSAM. It also demonstrates better segmentation performance [5], [15]. These improvements are achieved by leveraging recent advances in lightweight ViTs and applying a decoupled distillation strategy, making MobileSAM well-suited for next-generation segmentation applications in resource-constrained environments.

III. METHODOLOGY

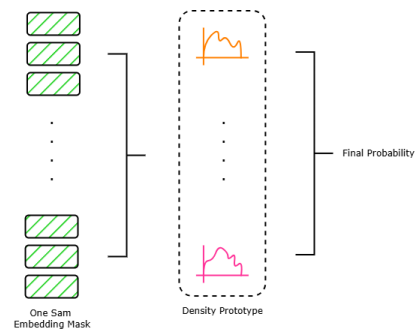


Fig. 2. Illustration of density estimation and prototype creation per class.

Feature Density Method: We perform the *object of interest* detection by sampling a small support set of size

$k \in \{3, 6, 9, \dots, 42\}$ from a directory of *object of interest* mask images, and evaluating on a query set that contains the remaining *object of interest* masks, plus all *background* mask images. Rather than training a full classifier, we model the embedding distribution of cataract masks via per-coordinate KDE and decide by thresholding the log-density score (see Figure 2 for density estimation and prototype creation).

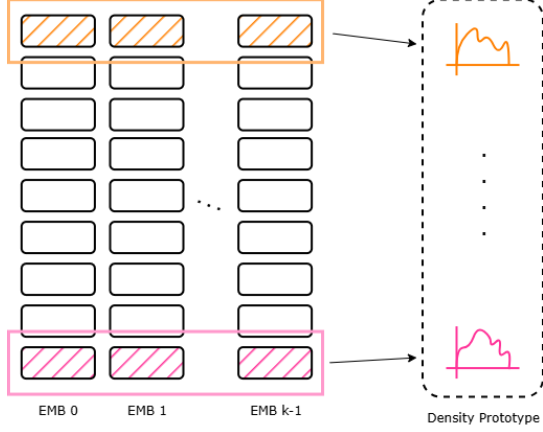


Fig. 3. Inference pipeline for scoring query embeddings against class density prototypes.

Each support or query image is resized 224×224 , scaled to $[0, 1]$, and normalized with ImageNet mean and standard deviation. We extract embeddings $z \in \mathbb{R}^d$ using a hybrid CNN-Transformer backbone (ResNet-50 + ViT-B/16 fused into a 768-dimensional vector).

At training time, given the k support embeddings $\{z_i\}_{i=1}^k$, we fit an independent one-dimensional Gaussian KDE for each embedding dimension j :

$$\hat{p}_j(u) = \frac{1}{k h} \sum_{i=1}^k \exp \left[-\frac{1}{2} \left(\frac{u - z_{ij}}{h} \right)^2 \right], \quad h = k^{-1/5}.$$

To compute the decision thresholds for the log-likelihood score computed later during inference time, we use the OOD thresholding method.

Therefore, at inference time, each query image is preprocessed identically and embedded to z^* (see Figure 3 for the inference pipeline). We compute its log-density $\ell^* = \sum_j \log \hat{p}_j(z_j^*)$ and classify as

$$\hat{y} = \begin{cases} \text{cataract}, & \theta_{\min} \leq \ell^* \leq \theta_{\max}, \\ \text{normal}, & \text{otherwise.} \end{cases}$$

Accuracy Calculation: The accuracy metric evaluates the proportion of correctly classified samples over the total number of predictions made. For instance, if the dataset contains 10 images of cataracts (ground truth), and the model correctly identifies only 3 of them as cataract, the accuracy would be 30%. In general, it is computed using the following formula:

$$\text{Accuracy} = \left(\frac{\# \text{ correct predictions}}{\# \text{ total predictions}} \right) \times 100.$$

IV. DATABASE

We conducted our experiments using the Cataract-SEG dataset [16], a public image segmentation dataset specifically designed for cataract detection. The dataset contains a total of 719 annotated anterior segment images, each paired with a binary mask highlighting the cataract region. The data is divided into three splits: 630 images for training (88%), 59 for validation (8%), and 30 for testing (4%).

Prior to training, all images were preprocessed using automatic orientation correction and resized to a fixed resolution of 640×640 pixels via stretching. Data augmentation was applied to enhance generalization, with each training image producing three augmented samples. The augmentations included random horizontal and vertical flips, rotations between -15 and $+15$, saturation adjustments between -30% and $+30\%$, and mild Gaussian blur (up to 0.5px). These transformations help simulate real-world variations and improve the robustness of the segmentation models.

V. EXPERIMENTS

Our experimental framework was designed to systematically evaluate the performance of SAM-based models and traditional backbones for cataract segmentation, with a particular focus on the feature-density approach under varying few-shot conditions. All experiments were conducted in Google Colaboratory using L4 Graphics Processing Unit (GPU) instances to ensure consistent hardware acceleration across trials.

Experimental Protocol: The study followed a structured evaluation protocol divided into two main phases. First, we assessed segmentation quality across different model architectures, including three SAM variants (SAM-ViT-H, SAM-ViT-L, SAM-ViT-B), their enhanced counterparts (HQ-SAM, MobileSAM), and traditional backbones (ResNet-18, ResNet-34, ViT). Second, we implemented our feature-density classifier under progressively increasing support-set sizes from $k=3$ to $k=42$ in increments of 3, repeating each configuration five times with different random seeds to ensure statistical reliability. **Model Implementation:** The SAM variants were initialized using their respective public checkpoints, with only the mask decoder undergoing fine-tuning while keeping the image encoders frozen. For traditional architectures (ResNet-18, ResNet-34, ViT-B/16), we initialized weights from ImageNet pretraining, fine-tuning the encoder components while training a new two-layer convolutional decoder from scratch. All models shared identical data augmentation pipelines during training, including random flips, small-angle rotations (± 15), saturation variations ($\pm 30\%$), and mild Gaussian blurring ($\sigma = 0.5\text{px}$). **Few-Shot Evaluation Protocol:** The feature-density experiments employed a rigorous k -shot learning paradigm. For each trial run, we randomly sampled k cataract masks from the training set to form the support set, using the remaining cataract masks and all background samples as queries. This procedure was repeated across 14 different support-set sizes ($k \in \{3, 6, \dots, 42\}$), with each configuration evaluated through five independent trials using different random seeds. The KDE classifier was then applied to measure how well

TABLE I

PERFORMANCE OF THE RESNET-18 BACKBONE AS A FEATURE EXTRACTOR ACROSS DIFFERENT SEGMENTATION ARCHITECTURES. AS EXPLAINED IN SECTION II.

Rank	Architecture	Max Acc (%)	Std. Max	Mean Acc (%)	Acc Range (%)
1	sam-vit-b	64.67 (K=21)	4.99	53.95	42.0–64.7
2	hq-sam-vit-h	63.33 (K=21)	7.60	53.52	44.7–63.3
3	hq-sam-vit-b	60.67 (K=6)	6.80	54.90	44.7–60.7
4	sam-vit-h	60.00 (K=21)	9.66	52.24	45.3–60.0
5	hq-sam-vit-tiny	60.00 (K=21)	2.98	47.29	37.3–60.0
6	sam-vit-l	59.33 (K=21)	6.46	52.48	45.3–59.3
7	mobileSAM	56.67 (K=21)	6.32	43.05	34.7–56.7

TABLE II

PERFORMANCE OF THE RESNET-34 BACKBONE AS A FEATURE EXTRACTOR ACROSS DIFFERENT SEGMENTATION ARCHITECTURES. AS EXPLAINED IN SECTION II.

Rank	Architecture	Max Acc (%)	Std. Max	Mean Acc (%)	Acc Range (%)
1	hq-sam-vit-h	65.33 (K=21)	8.06	56.76	41.3–65.3
2	sam-vit-b	64.00 (K=21)	6.46	55.67	39.3–64.0
3	sam-vit-h	64.00 (K=21)	8.27	55.76	40.0–64.0
4	sam-vit-l	64.00 (K=30)	8.27	55.67	39.3–64.0
5	hq-sam-vit-b	62.67 (K=21)	6.80	57.38	42.0–62.7
6	hq-sam-vit-tiny	58.67 (K=30)	5.42	49.24	34.7–58.7
7	mobileSAM	51.33 (K=30)	4.52	42.38	36.7–51.3

TABLE III

PERFORMANCE OF THE ViT-B-16 BACKBONE AS A FEATURE EXTRACTOR ACROSS DIFFERENT SEGMENTATION ARCHITECTURES. AS EXPLAINED IN SECTION II.

Rank	Architecture	Max Acc (%)	Std. Max	Mean Acc (%)	Acc Range (%)
1	hq-sam-vit-h	65.33 (K=27)	8.06	53.24	40.0–65.3
2	sam-vit-h	65.33 (K=27)	8.06	53.29	40.0–65.3
3	hq-sam-vit-b	65.33 (K=27)	6.53	55.48	42.0–65.3
4	sam-vit-b	65.33 (K=27)	6.53	53.95	39.3–65.3
5	hq-sam-vit-tiny	65.33 (K=27)	8.59	48.95	34.0–65.3
6	sam-vit-l	65.33 (K=27)	8.59	53.71	40.0–65.3
7	mobileSAM	64.00 (K=27)	7.12	42.76	33.3–64.0

the model could generalize from these limited examples, with decision thresholds calibrated on the validation set, as shown in Figure 1. **Computational Environment:** All experiments were conducted in a controlled computational environment using Google Colaboratory with L4 GPU acceleration. This standardization ensured consistent evaluation metrics across model variants while maintaining practical constraints representative of real-world deployment scenarios. Training times varied significantly between architectures, with MobileSAM completing epochs approximately $5\times$ faster than the largest SAM-ViT-H variant, demonstrating the practical trade-offs between model size and computational efficiency.

VI. RESULTS

The results of our evaluation across three different backbone architectures—ResNet-18, ResNet-34, and ViT-B/16—are presented in Tables I, II, and III, respectively. Each table ranks the best-performing SAM-based variants according to maximum accuracy, mean accuracy, and stability across varying support set sizes.

For the **ResNet-18** backbone, the best-performing configuration was the sam-vit-b architecture, reaching a maximum accuracy of 64.67% at K=21 and the highest mean accuracy of 53.95%, with relatively low variance (Std. Max = 4.99).

Interestingly, while hq-sam-vit-b achieved a slightly lower maximum (60.67%), it maintained a higher mean accuracy (54.90%), suggesting greater consistency in performance with smaller support sizes. Other architectures such as hq-sam-vit-h and sam-vit-h performed comparably, but exhibited higher standard deviations and less stable accuracy across K values. Overall, ResNet-18, despite being a lightweight model, provided competitive results when paired with strong mask generators. In the case of the **ResNet-34** backbone, hq-sam-vit-h emerged as the top performer, attaining the highest maximum accuracy of 65.33% at K=21 and a mean accuracy of 56.76%. This variant also showed the widest performance range (41.3–65.3%) across support sizes from K=12 to K=21, indicating its sensitivity to the size of the support set. On the other hand, hq-sam-vit-b recorded the best mean accuracy of 57.38%, despite a lower peak (62.67%). These findings suggest that deeper convolutional backbones like ResNet-34 are able to extract richer features that complement the precision of HQ-SAM models, especially in few-shot setups. With the **ViT-B/16** backbone, a notable pattern emerged: six out of seven architectures—excluding MobileSAM—reached the same maximum accuracy of 65.33%, though at varying support set sizes (all at K=27). Among them, hq-sam-vit-b achieved the highest mean accuracy (55.48%) with the lowest

standard deviation among top models (Std. Max = 6.53), indicating stable performance across trials. Architectures like sam-vit-h and hq-sam-vit-h showed competitive results, but with slightly lower mean accuracy and higher variability. The consistent high performance observed with ViT-B/16 confirms the advantages of transformer-based embeddings when paired with high-quality segmentation heads, particularly in few-shot learning contexts where generalization is critical.

In general, across all three backbones, the HQ-SAM variants frequently outperformed their standard SAM counterparts in terms of mean accuracy and robustness. Moreover, MobileSAM, while computationally efficient, consistently ranked lower in performance, reinforcing the trade-off between speed and segmentation precision. Finally, the results validate that increasing the support set size up to $K=21$ or $K=27$ tends to yield the highest accuracy, after which improvements may plateau or fluctuate depending on model and backbone combinations. These findings provide a clear indication that careful pairing of segmentation architectures with appropriate backbones is essential for maximizing performance in few-shot medical image classification tasks.

VII. DISCUSSION & CONCLUSION

This work presents the first benchmark of SAM-based models for cataract detection using feature density analysis. Despite the task's complexity, especially under few-shot conditions, our results show promising accuracy levels—up to 65.33%—across different architectures and backbones. The maximum accuracy of 65.33% marks a strong starting point, particularly given the small support sizes ($k = 3$ to $k = 42$) and the challenge of segmenting cataracts in anterior eye images. HQ-SAM variants consistently outperformed standard ones, especially when paired with ResNet-34 or ViT-B/16 backbones. While MobileSAM achieved lower accuracy, its speed and size make it an attractive option for deployment in constrained environments. Our KDE-based density approach proved effective for few-shot learning, with optimal performance typically between $k = 21$ and $k = 27$. The ViT-B-16 backbone yielded robust results across nearly all architectures, suggesting transformer-based extractors offer stronger generalization in low-data settings. **Limitations and Future Work.** The dataset size (719 images) limits generalizability, and current accuracy is below clinical deployment standards. Future work should expand datasets, explore improved feature modeling, and validate performance across diverse populations and imaging setups. Integrating uncertainty estimation and domain adaptation may further improve real-world applicability. **Clinical Potential:** Though not ready for deployment, this approach offers value for early screening in under-resourced areas. The few-shot capabilities reduce reliance on large annotated datasets, and lightweight variants like MobileSAM open the door for mobile-based applications. The method also holds promise for extending to other ophthalmic conditions. **Conclusion.** We demonstrate the viability of combining SAM-based models with KDE-driven feature density analysis for cataract detection. Our benchmark across multiple architectures sets

a foundation for future research and clinical translation. The proposed framework contributes toward accessible and scalable ophthalmic AI systems, addressing real-world constraints and global health needs.

REFERENCES

- [1] Y. Zeng, Y. Liu, X. Chen, R. Rong, and X. Xia, "Global, regional, and national burden of blindness and vision loss attributable to smoking from 1990 to 2021, and forecasts to 2030: findings from the global burden of disease study 2021," *BMC Public Health*, vol. 25, p. 440, 2025.
- [2] 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*, 2023, pp. 4015–4026.
- [3] C. Zhang, L. Liu, Y. Cui, G. Huang, W. Lin, Y. Yang, and Y. Hu, "A comprehensive survey on segment anything model for vision and beyond," *arXiv preprint arXiv:2305.08196*, 2023.
- [4] L. Ke, M. Ye, M. Danelljan, Y. Liu, Y.-W. Tai, C.-K. Tang, and F. Yu, "Segment anything in high quality," *arXiv preprint arXiv:2306.01567*, 2023.
- [5] C. Zhang, D. Han, Y. Qiao, J. U. Kim, S.-H. Bae, S. Lee, and C. S. Hong, "Faster segment anything: Towards lightweight sam for mobile applications," *arXiv preprint arXiv:2306.14289*, 2023.
- [6] S. Ayyachamy, V. Alex, M. Khened, and G. Krishnamurthi, "Medical image retrieval using resnet-18," in *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*. SPIE, 2019.
- [7] M. Alabdulhafith, A. S. Ba Mahel, N. Abdel Samee, N. F. Mahmoud, R. Talaat, M. S. A. Muthanna, and T. M. Nassef, "Automated wound care by employing a reliable u-net architecture combined with resnet feature encoders for monitoring chronic wounds," *Frontiers in Medicine*, vol. 11, 2024.
- [8] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Le, A. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.
- [9] S. Wang, J. Wang, and F.-I. Chung, "Kernel density estimation, kernel methods, and fast learning in large data sets," *IEEE Transactions on Cybernetics*, vol. 44, no. 1, pp. 1–20, 2014.
- [10] A. Parnami and M. Lee, "Learning from few examples: A summary of approaches to few-shot learning," 2022. [Online]. Available: <https://arxiv.org/abs/2203.04291>
- [11] A. De Silva, R. Ramesh, C. Priebe, P. Chaudhari, and J. T. Vogelstein, "The value of out-of-distribution data," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 7366–7389. [Online]. Available: <https://proceedings.mlr.press/v202/de-silva23a.html>
- [12] E. Zvornicanin, "What does backbone mean in neural networks?" *Baeldung*, 2025.
- [13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.
- [14] Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen, S. Liu, H. Chi, X. Hu, K. Yue, L. Li, V. Grau, D.-P. Fan, F. Dong, and D. Ni, "Segment anything model for medical images?" *Medical Image Analysis*, vol. 92, p. 103061, 2024.
- [15] X. Zhao, W. Ding, Y. An, Y. Du, T. Yu, M. Li, M. Tang, and J. Wang, "Fast segment anything," *arXiv preprint arXiv:2306.12156*, 2023.
- [16] M. Risma, "Cataract-seg: Cataract segmentation dataset," <https://universe.roboflow.com/muhammad-risma/cataract-seg/dataset/2>, 2024.

ACRONYMS

AI	Artificial Intelligence. 1
CNN	Convolutional Neural Network. 1, 3
CT	Computed Tomography. 1, 2
FSL	Few-Shot Learning. 1

GPU Graphics Processing Unit. 3, 4

KDE Kernel Density Estimation. 1, 3

mAP mean Average Precision. 1

MRI Magnetic Resonance Imaging. 1, 2

OOD Out-Of-Distribution. 1–3

SAM Segment Anything Model. 1–5

ViT Vision Transformer. 1–5