International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 11, No. 10, October, 2024, pp. 2817-2822
Journal homepage: http://iieta.org/journals/mmep

A Retinal Fundus Image Segmentation Approach Based on the Segment Anything Model

Bingyan Wei

International College, Krirk University, Bangkok 10220, Thailand

Corresponding Author Email: weibingyan66@gmail.com

https://doi.org/10.18280/mmep.111024

Received: 15 August 2024
Revised: 2 October 2024
Accepted: 10 October 2024
Available online: 31 October 2024

Keywords:
retinal fundus images, image segmentation, segment anything model, data augmentation

ABSTRACT

The segmentation of retinal blood vessels in fundus images is critical for the diagnosis and analysis of various ocular, circulatory, and neurological conditions. Accurate segmentation aids in the early detection of diseases such as diabetic retinopathy and glaucoma. Traditional automated segmentation methods often rely on extensive labeled datasets for model pre-training, which limits their generalization capacity. Recent advancements in artificial intelligence and computer vision have introduced foundational models, such as the Segment Anything Model (SAM), which demonstrate strong zero-shot segmentation performance and transferability in natural image processing. However, SAM's application to medical imaging, particularly in retinal vessel segmentation, has produced suboptimal results. This study proposes an improved approach to retinal fundus image segmentation by integrating a Generative Adversarial Network (GAN)-based data augmentation technique to enhance training data diversity. Additionally, a dynamic batch size mechanism was introduced, optimizing the loss function for mask prediction and allowing flexible control over slice selection during loss calculation. This dual enhancement aims to improve the precision of blood vessel segmentation in retinal fundus images, overcoming the limitations observed in previous applications of SAM to medical image segmentation. The proposed method demonstrates potential for advancing the accuracy of retinal vessel segmentation, providing a robust tool for clinical diagnosis.

1. INTRODUCTION

The retinal vascular system is widely recognized as an indispensable and vital element in the diagnosis of ophthalmic and cardiovascular diseases. For instance, in patients with glaucoma and diabetic retinopathy, examination of the retinal vascular system is the most direct method for disease detection [1]. Clinically, physicians typically collect fundus retinal images from patients using an ophthalmoscope and diagnose diseases by professionally analyzing the morphology of the retina.

Segmenting retinal vessels from fundus retinal images can facilitate better observation of diseases by physicians. Moreover, the segmentation of retinal vessel images also assists researchers in gaining a deeper understanding of the development and progression of eye diseases, providing crucial references for the formulation of treatment plans. The fundus vascular images are shown in Figure 1, where (a) is the fundus vascular image, and (b) is the segmented retinal vessel image after processing. Attributes of retinal vessels, including length, width, tortuosity, branching patterns, and angles, aid physicians in diagnosing and analyzing diseases.

With the advancement of deep learning technology, an increasing number of researchers are investigating segmentation methods for retinal images. The segmentation methods for color fundus retinal vascular images include both supervised and unsupervised learning-based approaches. Supervised learning-based retinal vessel segmentation methods require manual annotation to obtain retinal vascular information in advance, and then the model learns to find the optimal solution through continuous learning. The main deep learning models include Convolutional Neural Network (CNN) models [2], Fully Convolutional Network (FCN) models [3], and U-Net models [4]. Unsupervised learning methods primarily utilize adversarial learning approaches. Although many researchers have employed various methods to study the segmentation of fundus retinal images, there is a scarcity of publicly available datasets for fundus retinal images, leading to a lack of sample updates and excessive reliance on manual annotation.

In recent years, with the development of artificial intelligence technology and computing power, foundational models have become increasingly important in the field of natural language processing, such as Chat-GPT and GPT4.0 [5]. These large language models are gradually influencing the field of computer vision. Recently, Kirillov et al. proposed a foundational model for image segmentation: the SAM, which has achieved new breakthroughs in the field of computer vision [6]. The SAM possesses excellent zero-shot transferability and can segment any object in any image without any annotation, achieving good results in natural images.

Therefore, this paper conducts research on the segmentation performance of SAM in fundus retinal vascular images. We adopt a GAN-based method to increase the quantity of the training set in the fundus retinal vascular dataset, addressing the issue of low segmentation accuracy due to uneven sampling and enhancing the segmentation capability of SAM in fundus retinal vessels. By using dynamic batch size as the loss for mask prediction, this work flexibly control the slices for loss calculation to improve model performance.

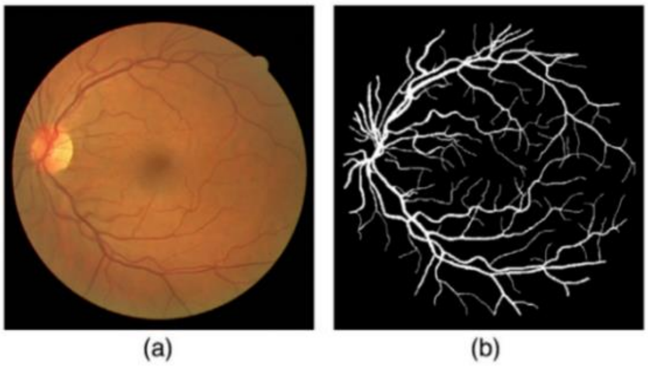

Figure 1. Vascular fundus images

2. RELATED WORK

2.1 SAM

In April 2023, Kirillov et al. [6] introduced a foundational model for image segmentation known as the SAM. SAM was designed and trained to be promptable, enabling zero-shot transfer to new image distributions and tasks, achieving instance segmentation without the need for any annotations, and demonstrating promising results in natural images [7].

SAM transforms the segmentation task into three main issues: tasks, models, and data. These three components are interwoven. Firstly, SAM defines a segmentation task that is general enough to provide robust pre-training objectives. SAM includes a prompt encoder and a mask decoder, combining these two information sources in a lightweight mask decoder that predicts segmentation masks. Subsequently, the model is trained using a diverse, large-scale dataset.

The future development directions and application scenarios for the SAM model are extensive. SAM can be utilized in computer vision-related applications. For instance, SAM can empower autonomous vehicles, accomplish industrial vision inspection tasks, annotate medical data, and in the medical field, assist physicians with pathology analysis and automatic diagnosis through the segmentation and annotation of medical images [8]. At the same time, SAM can effectively enhance the efficiency of image annotation and use the model and dataset for continuous iteration, improvement, and updates. In the field of image processing, it can be used for editing images or videos, such as removing unwanted objects or backgrounds from images and videos.

2.2 Research on medical image segmentation based on SAM

The advantages of SAM in the field of natural image segmentation are evident. The transferability and zero-shot segmentation capabilities of SAM are extremely important for medical images. If SAM can be applied to the field of medical image segmentation, it can effectively assist doctors in automatic disease diagnosis and screening. Some researchers have conducted studies on the role of SAM in medical image segmentation and have applied SAM to downstream segmentation tasks.

Zhang et al. [9] proposed a method that uses SAM as a medical image annotation tool, particularly for Multi-phase Liver Tumor Segmentation (MPLiTS). Wald et al. [10] assessed the segmentation capability of SAM on abdominal CT organs based on point and bounding box prompts, and the results indicated that SAM possesses zero-shot segmentation performance. Mohapatra et al. [11] conducted a comparative analysis of brain extraction techniques using the Brain Extraction Tool (BET) and SAM for various brain scan images with different image qualities, MRI sequences, and brain lesions affecting different brain regions. Zhou et al. [12] evaluated the performance of SAM in segmenting polyps using colonoscopy images without prompts. Experiments were conducted on five public datasets, and the results showed that there is still room for improvement when applying SAM to polyp segmentation tasks, and better segmentation effects can be achieved by fine-tuning the SAM model compared to no fine-tuning strategy.

The above studies indicate that SAM has great potential in medical image segmentation, being able to better integrate multimodal information. However, difficulties in segmenting vascular branching structures, such as those found in medical images, may be limited by the morphological imbalance in the training set. Vascular branching structure images are not the dominant dataset, so attempts can be made to improve the accuracy of medical images in different fields and targets by increasing the balance of the dataset.

2.3 Retinal vessel image segmentation methods

Retinal vessel segmentation methods encompass both supervised and unsupervised learning-based approaches. Supervised learning-based retinal vessel segmentation methods necessitate manual annotation to pre-obtain retinal vessel information, which is then used for continuous model learning to achieve the optimal solution. The primary deep learning models utilized by the majority of researchers include CNN models, FCN models, and U-Net models. Unsupervised learning methods mainly employ adversarial learning techniques.

Noh et al. [13] proposed a multi-scale residual CNN structure based on scale space approximation blocks. Ablation analysis indicated that the multi-scale residual CNN is indeed a significant factor in improving performance. To extract vascular contours of varying diameters, Guo and Peng [14] introduced a Bidirectional Symmetric Cascade Network (BSCN), which incorporates multi-scale feature representation of retinal vessels. This method was validated on four datasets: DRIVE, STARE, HRF, and CHASE_DB1, with an AUC above 0.98 for all. Irtaza Haider et al. [15] presented a lightweight CNN-based encoder-decoder deep learning model that utilizes stacking of two 3×3 convolutional layers without pooling layers, making the model more lightweight in terms of trainable parameters and computation time.

The aforementioned studies demonstrate that improving the loss function for retinal vessel segmentation in fundus images, by fitting the target with an appropriate loss function, can enhance the performance of retinal vessel segmentation.

3. METHODOLOGY

3.1 Generation of fundus retinal images based on adversarial learning

Data of fundus retinal vascular images are relatively scarce compared to other medical public datasets, and the cost of acquisition is high. In this paper, we employ a GAN, which is a method based on adversarial learning, to synthesize high-quality fundus retinal vascular images [16].

The basic structure of the GAN network is shown in Figure 2. The input to the generator is random noise, and the output is "fake" samples, which are in contrast to real samples. The real samples from the training set and the generated fake samples are then input together into the discriminator to obtain the final results.

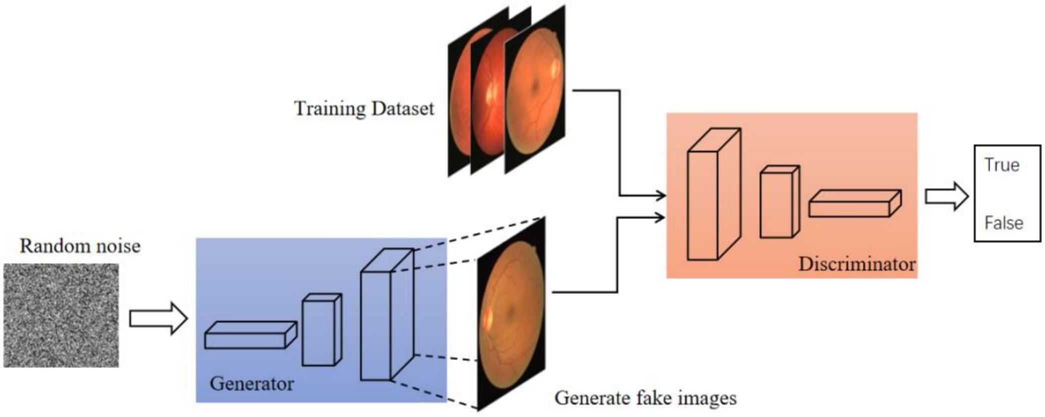

Figure 2. GAN network model structure

To ensure more stable training and to achieve high-quality fundus retinal vascular images, we introduce a feature matching loss here:

$$\mathcal{L}_{FM}(G,D_k)=\mathbb{E}_{(s,x)}\sum_{i}^{T}\frac{1}{N_i}[\| D_k^{(i)}(s,x)-D_k^{(i)}(s,G(s))\|_1] \quad (1)$$

where, $D_k^{(i)}$ is the i layer extracted from D_k, T is the total number of layers, N_i is the total number of elements in every layer. Consequently, the total objective function is given by:

$$\min_G\left(\left(\max_{D_1,D_2,D_3}\sum_{k=1,2,3}\mathcal{L}_{GAN}(G,D_k)\right)+\beta\sum_{k=1,2,3}\mathcal{L}_{FM}(G,D_k)\right) \quad (2)$$

where, β is the controlled the importance of the two loss functions, with D_k as a feature extractor only.

3.2 Retinal fundus vessel segmentation method based on SAM

This paper employs an image segmentation method based on the SAM. The original image, after passing through patch embedding, is input into the image encoder, which adopts a Transformer. Features are obtained through downsampling, reshaped to obtain the image embedding, and then CNN features are acquired through a 3×3 convolutional layer. To better integrate local and global features. Inspired by the study by Li et al. [17], a Feature Fusion Block is used between the CNN and Transformer features for feature correction, resulting in more representative and superior fused features that enhance the accuracy of segmentation results. The encoded prompt is passed into the Mask Decoder as input parameters, ultimately yielding the segmented mask. The model architecture is shown in Figure 3.

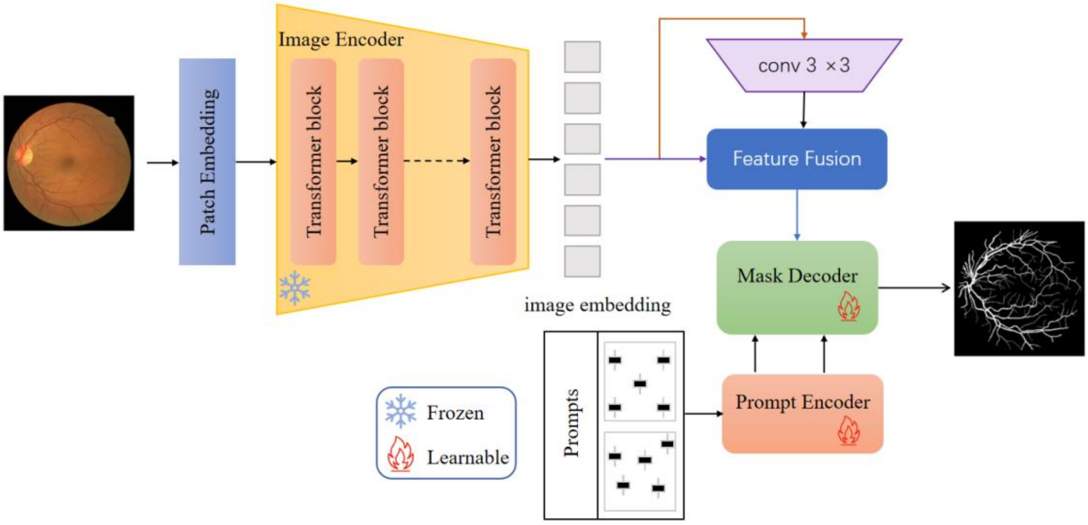

Figure 3. Model architecture

The encoder adopts the structure from the Transformer module, composed of multi-head attention mechanism modules and Multilayer Perceptron (MLP) modules. Layer Normalization (LN) and residual connections are used in each module. The encoder structure is depicted in Figure 4. The encoder is highly adaptable for processing high-resolution images.

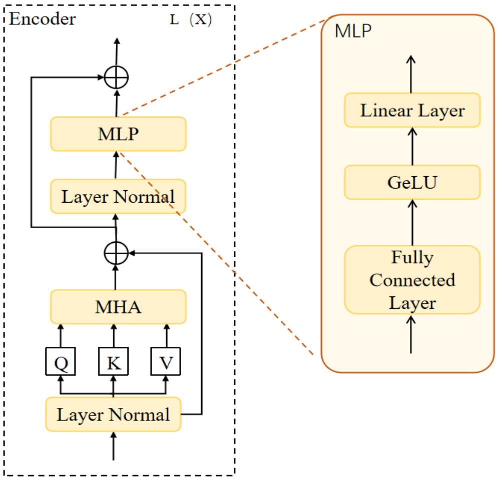

Figure 4. Encoder architecture

To improve image segmentation accuracy, this paper utilizes a batch-size-based dynamic Dice loss (BSD Loss), flexibly controlling the slices for loss calculation, effectively combining single-slice and multi-slice labels, and enhancing model performance.

To improve image segmentation accuracy, this paper utilizes a BSD Loss, flexibly controlling the slices for loss calculation, effectively combining single-slice and multi-slice labels, and enhancing model performance.

For each batch size, the mask can be represented as $\{\{mask_n^l\}_{l=1}^L\}_{n=1}^N$, and the segmentation result can be represented as $\{\{pred_n^l\}_{l=1}^L\}_{n=1}^N$, where $mask_n^l$ is the segmentation ground truth corresponding to the l label index of the n batch in the batch size. $pred_n^l$ is the segmentation prediction result of $mask_n^l$. N is the size of the batch size. L is the number of labels in the segmentation mask.

The dynamic Dice loss based on batch size can be expressed as:

$$loss\_BSD_i = \sum_{n=1}^{N}\sum_{l=1}^{L}\omega_l\left[1-Dice\left(mask_n^l, pred_n^l\right)\right] \quad (3)$$

where, $Dice(mask_n^l, pred_n^l)$ is the l label of the n batch of batch size corresponding to the segmentation ground truth and segmentation prediction result.

During the model fine-tuning, since clinicians are more concerned with the overall trajectory and branch characteristics of the fundus retinal vessels, this paper only uses Box prompts and inputs the generated prompts into the prompt encoder of SAM. Using SAM's mask decoder, combined with the input image and prompts, a multi-channel segmentation map is generated, which includes all categories. During the fine-tuning stage, only the parameters of the decoder are updated, while other parameters remain unchanged, employing a Parameter Efficient Fine-Tuning (PEFT) strategy to initiate the fine-tuning process.

4. EXPERIMENTS AND ANALYSIS

4.1 Datasets

In this work, a total of three public retinal datasets were utilized for image generation, super-resolution, and segmentation tasks. The DRIVE [18], STARE [19], and CHASE DB1 [20] datasets are employed, with Table 1 displaying detailed information on the datasets used in the experiments.

Table 1. Data set

Data Set	Format	Quantity
DRIVE	2D	40
STARE	2D	400
CHASE DB1	2D	28

4.2 Experimental process

The experimental environment is equipped with a workstation featuring NVIDIA 8 A100 GPUs, running Python 3.10.0, Pytorch 1.10.1, and CUDA 11.1 for local execution.

Initially, this paper trained a GAN using the STARE dataset. The network was constructed using TensorFlow, with 14,000 iterations, an Adam optimizer, a learning rate of 0.0002, and a momentum term beta1 of 0.5. Since the input consists of color images, the color dimension c_dimis set to 3.

Subsequently, the generated image dataset was input into SAM for training. The model was then tested using the DRIVE, STARE, and CHASE DB1 datasets to verify the accuracy of the segmentation results.

This paper employed PEFT technology to fine-tune the model, freezing the parameters of the encoder part (image and prompt encoders) and only updating the gradients of the decoder. This approach enhances model performance using limited data and computational resources, reducing the number of parameters that need to be optimized during training.

4.3 Evaluation metrics

This paper utilizes Pixel Accuracy (PA), Sensitivity (Se), Specificity (Sp), and other objective metrics to evaluate the results of medical image segmentation.

PA measures the proportion of correctly classified pixels (both vessel and non-vessel) relative to the total number of pixels in the image. The formula is given by:

$$PA = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Sensitivity (Se) measures the proportion of true positives (correctly classified vessel pixels) out of the total actual vessel pixels. The formula is given by:

$$\mathrm{Se} = \frac{TP}{TP+FN} \quad (5)$$

Specificity (Sp) measures the proportion of true negatives (correctly classified non-vessel pixels) out of the total actual non-vessel pixels. The formula is given by:

$$Sp = \frac{TN}{TN + FP} \tag{6}$$

In these formulas, *TP*, *TN*, *FP*, and *FN* represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

Additionally, the interrelationship between sensitivity and specificity is used to plot the Receiver Operating Characteristic (ROC) curve. The area under the ROC curve (AUC) reflects the performance of the segmentation method.

4.4 Experimental results and analysis

This paper compared the segmentation results of the original SAM model with U-net, U-Net++, and AF-Net on the STARE dataset, as shown in the table below. The results indicate that our segmentation method achieved better outcomes than the original SAM segmentation method. The results are displayed in Table 2.

The visualization results of this paper are shown in Figure 5. Figure 5(a) represents the original fundus retinal image; Figure 5(b) is the gold standard for segmentation, derived from manual annotations by physicians; Figure 5(c) is the result after segmentation using SAM with prompt points; Figure 5(d) is the segmentation result using the U-Net method; Figure 5(e) is the segmentation result of this paper. The results indicate that by increasing the diversity of the data, SAM can better learn the features of retinal images and has achieved good results in the three public datasets. However, the classification of small or minute vessels remains a challenge when applying SAM to vessel segmentation.

Table 2. Model segmentation results comparison

Method	Data Set	PA	SE	SP	AUC
U-net [21]	DRIVE	0.9548	0.7700	0.9829	0.9789
	STARE	0.9644	0.8251	0.9805	0.9832
	CHASE DB1	0.9630	0.8032	0.9789	0.9822
U-Net++ [22]	DRIVE	0.9580	0.8139	0.9818	0.9820
	STARE	0.9658	0.8372	0.9801	0.9837
	CHASE DB1	0.9667	0.8214	0.9811	0.9869
AF-Net [23]	DRIVE	0.9576	0.7918	0.9828	0.9798
	STARE	0.9712	0.7972	0.9779	0.9902
	CHASE DB1	0.9669	0.8194	0.9817	0.9867
SAM (auto)	DRIVE	0.7120	0.6050	0.7305	0.9614
	STARE	0.8523	0.7321	0.7250	0.8512
	CHASE DB1	0.6351	0.5300	0.7312	0.6973
Ours	DRIVE	0.9603	0.7803	0.9730	0.9890
	STARE	0.9605	0.8032	0.9856	0.9811
	CHASE DB1	0.9774	0.8103	0.9801	0.9846

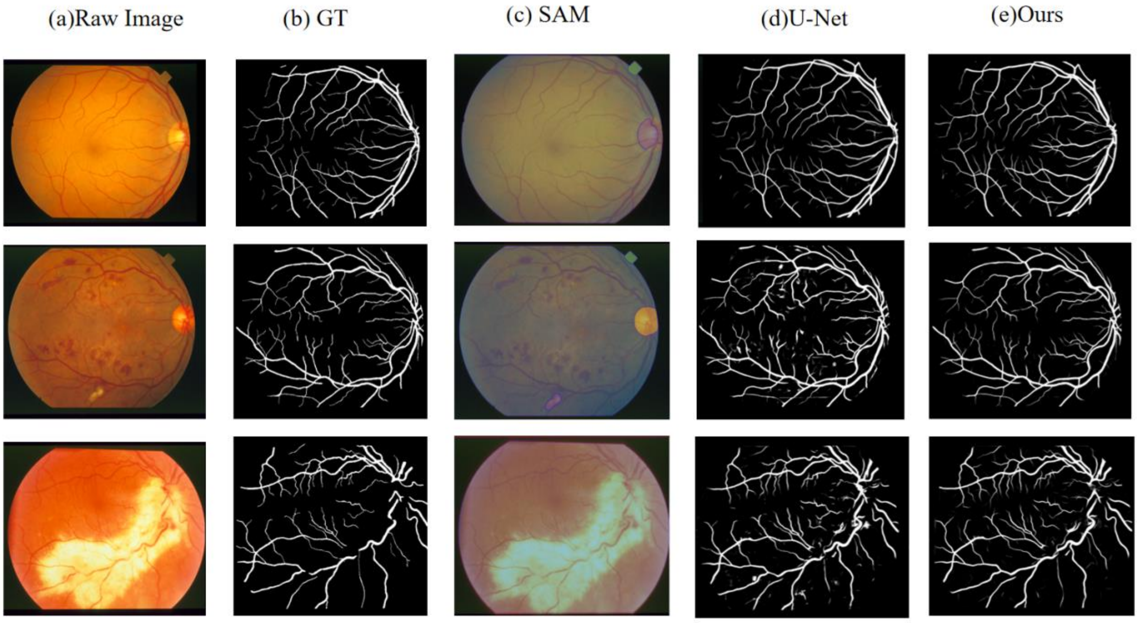

Figure 5. Segmentation results

5. CONCLUSIONS AND PROSPECT

This paper conducts research on the segmentation performance of SAM in fundus retinal vascular images. To enhance the segmentation accuracy of fundus retinal vessels, we adopt a BSD Loss on the original architecture, which allows for flexible control over the slices used in loss calculation and effectively combines single-slice and multi-slice labels to improve model performance.

In the future, further exploration can be conducted on utilizing SAM for multimodal segmentation results in the fundus, extending the 2D segmentation outcomes to 3D, thereby further enhancing segmentation accuracy. This could assist physicians in preoperative diagnosis and treatment of patients with brain tumors.

REFERENCES

[1] Handan, A., Huang, A.S., Francis, B.A., Sadda, S.R., Chopra, V. (2017). Retinal vessel density from optical coherence tomography angiography to differentiate early glaucoma, pre-perimetric glaucoma and normal eyes. PLoS ONE, 12(2): e0170476.

https://doi.org/10.1371/journal.pone.0170476

[2] Manthey, M. (2022). Artificial intelligence system. International Journal of Theoretical & Computational Physics.

[3] Azizpou, H., Razavian, A.S., Sullivan, J., Maki, A., Carlsson, S. (2016). Factors of transferability for a generic ConvNet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(9): 1790-1802. https://doi.org/10.1109/tpami.2015.2500224

[4] Ahmadi, M., Nia, M.F., Asgarian, S., Danesh, K., Irankhah, E., Lonbar, A.G., Sharifi, A. (2023). Comparative analysis of segment anything model and U-Net for breast tumor detection in ultrasound and mammography images. arXiv:2306.12510. https://doi.org/10.48550/arXiv.2306.12510

[5] Brown, T.B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. arXiv:2005.14165. https://doi.org/10.48550/arXiv.2005.14165

[6] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L. (2023). Segment anything. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, pp. 3992-4003. https://doi.org/10.1109/ICCV51070.2023.00371

[7] Ma, Z.H., Hong, X.P., Shangguan, Q.N. (2023). Can SAM count anything? An empirical study on SAM counting. arXiv:2304.10817. https://doi.org/10.48550/arXiv.2304.10817

[8] Zhang, Y.Z., Zhou, T., Wang, S., Liang, P.X., Chen, D.Z. (2023). Input augmentation with SAM: Boosting medical image segmentation with segmentation foundation model. arXiv:2304.11332. https://doi.org/10.48550/arxiv.2304.11332

[9] Zhang, Y.Z., Zhou, T., Wang, S., Liang, P.X., Chen, D.Z. (2023). When SAM meets medical images: An investigation of segment anything model (SAM) on multi-phase liver tumor segmentation. arXiv:2304.11332. https://doi.org/10.48550/arXiv.2304.11332

[10] Wald, T., Roy, S., Koehler, G., Disch, N., Rokuss, M. R., Holzschuh, J., Zimmerer, D., Maier-Hein, K. (2023). SAM.MD: Zero-shot medical image segmentation capabilities of the segment anything model. Medical Imaging with Deep Learning, Short Paper Track. https://openreview.net/forum?id=iilLHaINUW.

[11] Mohapatra, S., Gosai, A., Schlaug, G. (2023). SAM vs BET: A comparative study for brain extraction and segmentation of magnetic resonance images using deep learning. arXiv:2304.04738. https://doi.org/10.48550/arXiv.2304.04738

[12] Zhou, T., Zhang, Y.Z., Zhou, Y., Wu, Y., Gong, C. (2023). Can SAM segment polyps? arXiv:2304.07583. http://arxiv.org/abs/2304.07583.

[13] Noh, K.J., Park, S.J., Lee, S. (2018). Scale space approximation in convolutional neural networks for retinal vessel segmentation. arXiv:1806.09230. https://doi.org/10.48550/arXiv.1806.09230

[14] Guo, Y.F., Peng, Y.J. (2020). BSCN: Bidirectional symmetric cascade network for retinal vessel segmentation. BMC Med Imaging, 20: 20. https://doi.org/10.1186/s12880-020-0412-7

[15] Haider, S.I., Aurangzeb, K., Alhussein, M. (2022). Modified ANAM-Net based lightweight deep learning model for retinal vessel segmentation. Computers, Materials & Continua, 73(1): 1501-1526. https://doi.org/10.32604/cmc.2022.025479

[16] Tian, C.H., Yang, J., Li, P., Zhang, S.C., Mi, S.L. (2022). Retinal fundus image superresolution generated by optical coherence tomography based on a realistic mixed attention GAN. Medical Physics, 49(5): 3185-3198. https://doi.org/10.1002/mp.15580

[17] Li, M., Pi, D.C, Qin, S. (2023). An efficient single shot detector with weight-based feature fusion for small object detection. Scientific Reports, 13: 9883. https://doi.org/10.1038/s41598-023-36972-x

[18] Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging, 23(4): 501-509. https://doi.org/10.1109/TMI.2004.825627

[19] Hoover, A.D., Kouznetsova, V., Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Transactions on Medical Imaging, 19(3): 203-210. https://doi.org/10.1109/42.845178

[20] Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G.. (2012). An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Transactions on Biomedical Engineering, 59(9): 2538-2548. https://doi.org/10.1109/TBME.2012.2205687

[21] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[22] Zhou, Z.W., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.M. (2018). UNet++: A nested U-Net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham. https://doi.org/10.1007/978-3-030-00889-5_1

[23] Li, D.Y., Peng, L.X., Peng, S.H., Xiao, H.X., Zhang, Y.F. (2023). Retinal vessel segmentation by using AFNet. The Visual Computer, 39: 1929-1941. https://doi.org/10.1007/s00371-022-02456-8