Mathematical Biosciences and Engineering

MBE, 22(3): 585–603.
DOI: 10.3934/mbe.2025022
Received: 03 December 2023
Revised: 07 February 2024
Accepted: 01 March 2024
Published: 25 February 2025

http://www.aimspress.com/journal/MBE

Research article

MAET-SAM: Magneto-Acousto-Electrical Tomography segmentation network based on the segment anything model

Shuaiyu Bu[1,2,3], Yuanyuan Li[2,3,*], Guoqiang Liu[2,3] and Yifan Li[4]

[1] State Grid Beijing Electric Power Company, Beijing 100031, China
[2] Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing 100190, China
[3] School of Electronic Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[4] China Railway Communication and Signal Survey & Design Co., Beijing 100036, China

* **Correspondence:** Email: lyy@mail.iee.ac.cn.

Abstract: Magneto-Acousto-Electrical Tomography (MAET) is a hybrid imaging method that combines advantages of ultrasound imaging and electrical impedance tomography to image the electrical conductivity of biological tissues. In practical applications, different tissue or disease organization display various conductivity traits. However, the conductivity map consists of overlapping signals measured at multiple locations, the reconstruction results are affected by noise, which results in blurred reconstruction boundaries, low contrast, and irregular artifact distributions. To improve the image resolution and reduce noise of MAET, a dataset of conductivity maps reconstructed from MAET was established, dubbed MAET-IMAGE. Based on this dataset, we proposed a MAET tomography segmentation network based on the Segment Anything Model (SAM), termed as MAET-SAM. Specifically, we froze the encoder weights of SAM to extract rich feature information of image and design, an adaptive decoder with no prompts. In the end, an end-to-end segmentation model for specific MAET images with MAET-IMAGE was proposed. Qualitative and quantitative experiments demonstrated that MAET-SAM outperformed traditional segmentation methods and segmentation models with initial weights in terms of MAET image segmentation performance, bringing new breakthroughs and advancements to the field of medical imaging analysis and clinical diagnosis.

Keywords: MAET; image segmentation; segment anything model (SAM)

1. Introduction

The electrical properties of biological tissues are closely related to the structure, function, physiology, and pathology of tissues, which holds significant implications for medical diagnosis [1]. When biological tissues undergo early pathological changes and have not yet exhibited alterations in morphological structure, the electrical properties of the tissues at the affected sites also change [2]. Thus, electrical property imaging method are expected to be a promising imaging technique for the early lesion detection. As a form of electrical property imaging, MAET leverages the benefits of high contrast from electrical impedance tomography and high resolution from ultrasound imaging, enabling non-invasive imaging of the electrical properties of biological tissues [3]. Biological tissues are excited by both static magnetic field and ultrasound beam generated by acoustic probe. Ions of target sample are subjected to a Lorenz force in the presence of static magnetic field and acoustic field. The current distribution inside tissues varies with the propagation of ultrasound. By measuring this current by electrodes, the distribution of conductivity can be deduced. Despite significant research efforts into the theoretical aspects of MAET, challenges persist due to the low conductivity of biological tissues, the non-uniformity of the static magnetic field, reflected acoustic waves around the sound field, etc. These factors result in MAET images characterized by a low signal-to-noise ratio [4]. Additionally, the influence of both system and environmental noise hinder the accurate localization of pathological tissue in MAET images. Therefore, it is necessary to address these issues by applying image processing techniques.

Advancements in medical image processing techniques have significantly enhanced the capabilities of prevention, diagnosis, and treatment of disease [5]. As a pivotal medical image processing technique, medical image segmentation aims to precisely segment distinct tissues and lesion regions from medical images. This task enables healthcare professionals to obtain clearer insights into structures, features, and alterations of tissues, thus laying the foundation for subsequent identification and analysis of crucial regions [6]. Specifically, image segmentation methods classify all pixels in medical images through manual or adaptive computation, which divides the image into various distinct and meaningful regions. Within these regions, there is no overlap between any two regions, and each region possesses certain similar characteristics. Segmenting medical images allows healthcare professionals to accurately localize and quantify target regions, thus providing more precise information for disease diagnosis and treatment [7].

We can significantly enhance disease prevention and diagnosis capabilities by utilizing MAET for imaging the electrical properties of biological tissues, along with the application of medical image segmentation techniques for locating lesion regions. Since the MAET image is produced by superimposing signals measured at multiple positions, the reconstruction result is affected by noise superposition, which leads to few samples, blurred boundary, and irregular artifacts. Traditional segmentation methods cannot accurately segment conductivity regions. To address the mentioned issues, we proposed a deep learning network MAET-SAM to segment MAET images. Specifically, we froze the encoder weights of SAM [8] to extract richer feature information of image and design an adaptive decoder with no prompts. Due to the lack of publicly available datasets for MAET images, we measured multiple sets of conductivity distribution maps and combined them with simulated conductivity distribution maps for training. Qualitative and quantitative experiments demonstrated that MAET-SAM outperforms traditional segmentation methods and segmentation network models with initial weights in terms of MAET image segmentation performance. The major contributions of this

paper are threefold:

1) Established a dataset of conductivity maps reconstructed from MAET, named MAET-IMAGE. This dataset comprises 2000 pairs of simulated and 750 pairs of real-measured conductivity-mask images.

2) Proposed a MAET tomography segmentation method based on Segment Anything Model (SAM), termed as MAET-SAM, which is an end-to-end segmentation model for specific MAET images.

3) A prompt-free adaptive decoder was designed for MAET-SAM, which consists of multiple convolutional layers and transposed convolutional layers.

2. Related works

Wen et al. [3] first proposed MAET, also known as HEI (Hall Effect Imaging), in 1998. MAET leverages the benefits of high contrast from electrical impedance tomography and high resolution from ultrasound imaging, enabling non-invasive imaging of the electrical properties of biological tissues. Throughout the development of MAET, numerous scholars and research teams have made significant contributions. Montalibet et al. [9] derived the measurement formula for MAET and utilized Wiener inverse filtering to extract conductivity parameters. Haider et al. [10] introduced the reciprocity theorem in the reconstruction algorithm and obtained high spatial resolution images of current density. Grasland-Mongrain et al. [11] demonstrated that the detected voltage is proportional to the convolution with the piezoelectric conductivity. Guo et al. [12,13] improved logarithmic reconstruction algorithm of MAET in coil detection which is noninvasive conductivity imaging modality with high resolution. Kunyansky et al. [14] rotated object and used two pairs of electrodes, which were immersed into saline surrounding the object, to reconstruct conductivity image. A novel MAET with a chirp signal was investigated by Dai which could detect electrical properties of phantom effectively [15]. Then Dai et al [16] applied linear interpolation algorithm to increase the smoothness of conductivity images. Yu et al. [17] combined MAET with sinusoid-Barker coded excitation to improve the spatial resolution of MAET results. Multi-angle MAET with image rotation method was proposed by Sun to discern irregularly-shaped tumors and improve quality of reconstructed images [18]. In 2022, Deng et al. [19] explored the sensitivity of coded-excitation MAET in experiments, which was about 0.16 S/m. Rotary-scanning-based MAET is employed to improve image distortion [20]. Li et al. [21] discussed mathematical model of MAET with nonuniform static magnetic field and verified the theory with 0.2 S/m phantom in experiments.

Medical image segmentation task can extract vital information from reconstructed images of specific tissues or lesion, serving as a basis for subsequent disease diagnosis and assessment. Traditional medical image segmentation methods mostly include three categories: Threshold-based segmentation algorithms [22], edge detection-based segmentation algorithms [23], and region-based segmentation algorithms [24]. However, medical images often exhibit characteristics such as low contrast, complex tissue textures, and blurred boundary regions, which greatly limit the effectiveness and applicability of such image segmentation algorithms. In recent years, with significant advancements in image processing techniques based on deep learning, the performance of medical image segmentation has greatly improved [25,26]. Segmentation models built upon backbone networks such as AlexNet [27], VGG [28], ResNet [29], DenseNet [30], EfficientNet [31], ViT [32], Swin Transformer [33], and others can learn rich semantic feature representations from medical images. Subsequently, to restore the features of images into the same size as the input images, additional

upsampling modules are incorporated after the backbone network to achieve pixel-level classification and prediction. Long et al. [34] proposed Fully Convolutional Networks (FCNs) for segmenting medical image, where the FCN structure applies several convolution blocks composed of convolution, activation, and pooling layers on the encoder to capture semantic representations. Similarly, it employs convolution layers and upsampling operations in the decoder to achieve pixel-level predictions. Based upon FCNs, Ronneberger et al. developed the U-Net [35] model, which is more suitable for biomedical image segmentation. Subsequently, various segmentation models emerged as variations of U-Net [36]. DPS-Net [37] proposes a modified U-Net for automatic LVEF assessment using 2D echocardiography images across various heart disease phenotypes and different echocardiographic systems, demonstrating high performance in segmentation and diagnostic accuracy. Moreover, the success of Transformer [38] in natural language processing has propelled the development of computer vision. Many medical image segmentation methods based on Transformer were proposed. TransUNet [39] combined the strengths of Transformer and U-Net for improved medical image segmentation by leveraging global context extraction and precise localization. Swin-Unet [40] integrated pure Transformer architecture into the U-Net framework, enabling effective local-global semantic feature learning for medical image segmentation tasks. DS-TransUNet [41] inserted a hierarchical Swin Transformer into both the encoder and decoder of U-Net, facilitating non-local dependency modeling and multiscale context enhancement for medical image segmentation. NAG-Net [42] proposed a nested attention-guided deep learning model and incorporated task-related clinical domain knowledge. Recently, Meta trained a general segmentation model SAM [8] on the large-scale natural image dataset SA-1B, comprising over 11 million images and 1 billion masks. This model can accurately segment images based on human-provided prompts. However, since SAM has not been trained on medical images, its segmentation performance in this field, particularly for MAET images, is not satisfactory. Therefore, the focus of this study is to develop a more suitable segmentation dataset for MAET and propose an adaptive model for effective MAET segmentation.

3. Methods

3.1. Theory of MAET

The principle of MAET conductivity reconstruction is shown in Figure 1. The target sample, whose electrical conductivity is σ, is under the combined excitation of the static magnetic field B_0 and the ultrasound generated by the ultrasonic transducer. The direction of sound wave propagation is perpendicular to the direction of the static magnetic field. Ions q in the target sample vibrate under the action of static magnetic field and ultrasound whose speed is v, and vibrating ions will be subjected to the Lorentz force and generate charge separation. Since the Lorentz force in positive ions and negative ions is equal in magnitude but opposite in direction, the distributed current in the target sample can be expressed as:

$$J_e = \sigma v \times B_0, \tag{1}$$

where J_e is the current density of the equivalent current source. The voltage signal detected by electrodes attached to the imaging body is:

$$U(t) = \frac{\alpha W}{\rho_0} \int_l \frac{\partial \sigma}{\partial x} M B_0 \, dx, \tag{2}$$

where α is the proportional constant representing the current detected by the acquisition system, W is the width of the ultrasonic beam, ρ_0 is the density distribution of the target object, and M is the ultrasonic momentum. After obtaining the voltage signal, the sound source $H(r) = \nabla \cdot (J_2(r) \times B_0(r))/\rho_0$ at any point r in the target sample can be solved by time reversal. When the position of ultrasonic transducer changes, the detected voltage signal will also change accordingly. The conductivity imaging can be realized by detecting voltage signals at multiple positions, and the results can reflect internal electrical characteristics of the target sample.

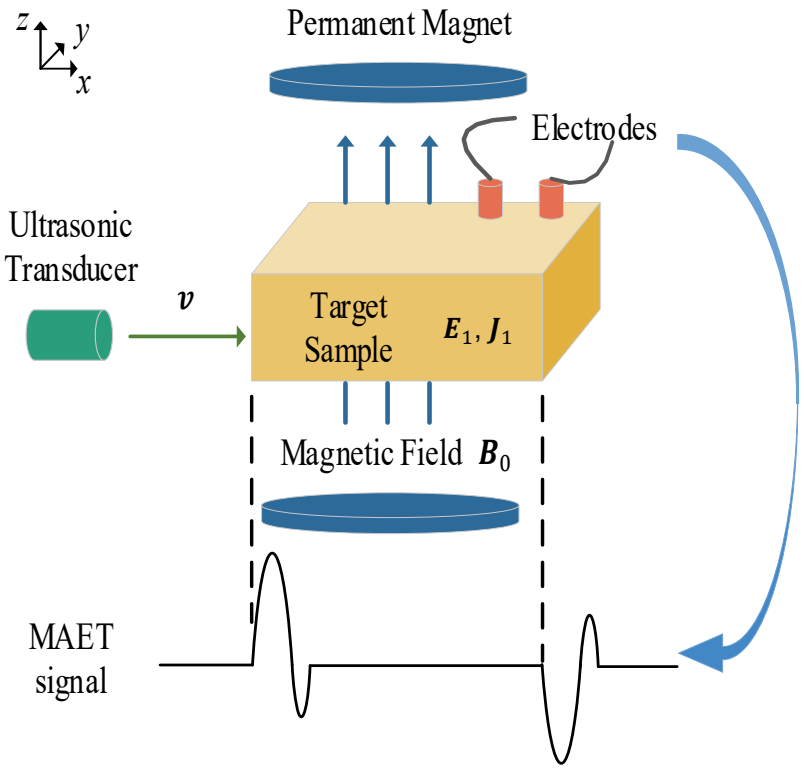

Figure 1. Schematic diagram of MAET imaging.

3.2. MAET image dataset MAET-IMAGE

In order to train a segmentation network model for MAET image segmentation with supervised learning, we established a set of MAET image dataset, termed MAET-IMAGE. This dataset contains 2000 pairs of simulated conductivity-mask images and 750 pairs of real measured conductivity-mask images. Due to the time-consuming and high-cost nature of real measured conductivity images, simulated conductivity images were included as part of the MAET-IMAGE dataset.

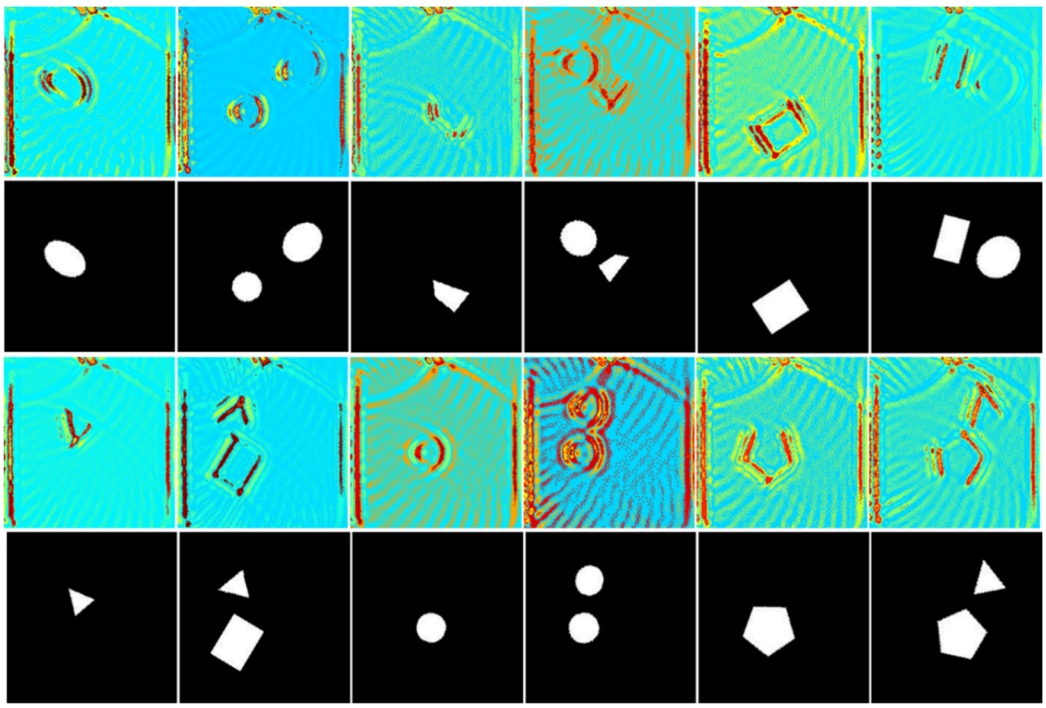

Figure 2. A portion of simulated conductivity-mask images.

Figure 2 shows a portion of simulated conductivity-mask images. These images are generated by COMSOL software. When simulating conductivity images, it is necessary to pre-define the shape of harmful tissues. Therefore, we create 2000 diagrams with geometric models, including circular, elliptical, rectangular, and triangular, in different quantities, sizes, positions, and directions. These diagrams are later directly used as mask images.

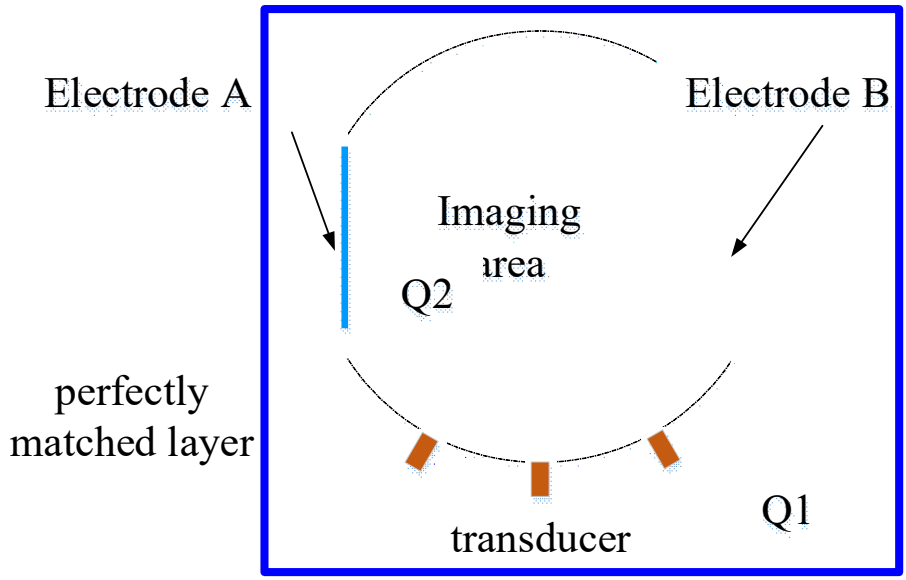

Figure 3. The simulation setup of MAET images.

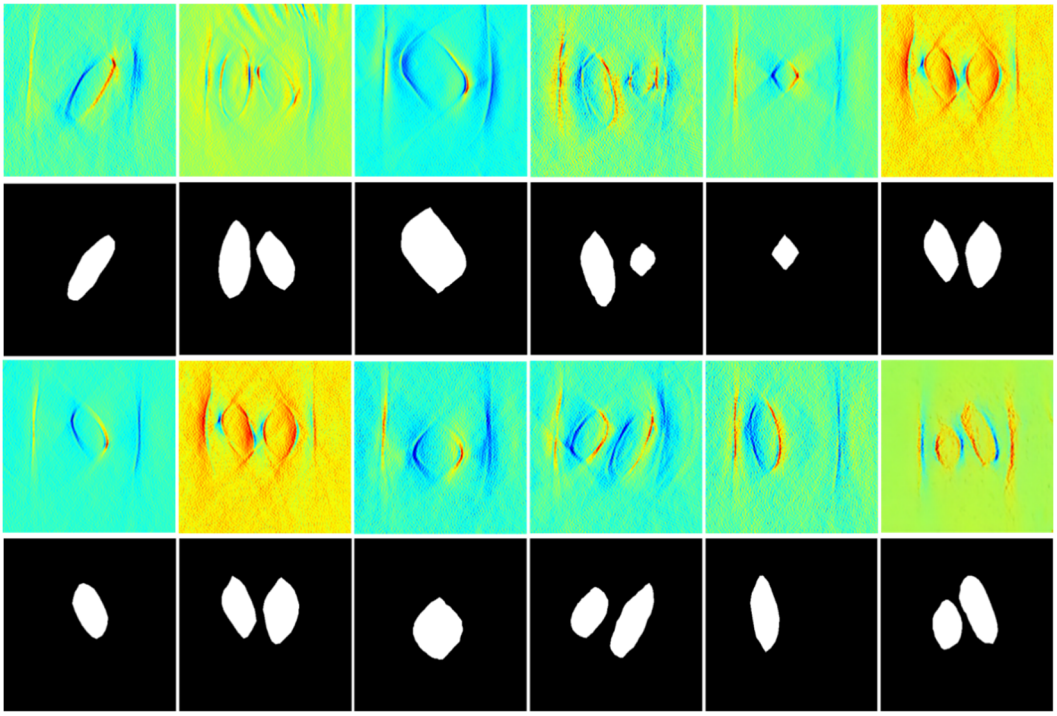

Figure 4. A portion of real measured conductivity-mask images.

The MAET process is shown in Figure 3. A circular area with a radius of 2.5 cm is set up, and the outermost layer of the solution area acted as the perfectly matched layer (the blue box in Figure 3) which is used to absorb sound waves to prevent ultrasound reflection Target sample is a 1.5 cm × 2 cm rectangle with a conductivity of 0.2 S/m, which is used to simulate the normal biological tissue. The target sample contained isolated tissue with different shapes, and the conductivity of the isolated part is 0.5 S/m. The ultrasonic probe is placed at a distance of 1.7 cm from the center of the target sample, rotating around the center of target sample. During the scanning process, the ultrasonic probe emits sound waves every 5°, and the scanning range is 120° in total. Ions inside the target sample vibrated under the action of ultrasound and 0.3 T static magnetic field to generate current source. With AC/DC module, electrodes collect voltage data every 4×10^{-8} seconds. After detecting voltage signals, the time reversal algorithm is initially used to calculate the sound source at any point within the tissues, followed by the use of the Newton iteration algorithm to reconstruct the internal conductivity distribution.

In real measurement, the experimental platform includes the 300 mT magnetic field, the ultrasound excitation source, amplifier, acquisition module, and the transducer rotation control platform. A single pulse signal with a center frequency of 0.5 MHz is applied to the transducer, and the generated ultrasonic wave propagated through the phantom along the direction of the transducer. The voltage signal is then generated with the static magnetic field and measured by electrodes. After 56 dB amplification, the signal is finally collected by the NI data acquisition card. A portion of real measured conductivity-mask images are shown in Figure 4.

To obtain the experimental phantom, a certain amount of NaCl is dissolved in water to achieve 0.2 S/m conductivity, which is measured through a Zurich Instruments MFIA instrument. Then, the agar is mixed with the liquid at a ratio of 1 g/100 ml, the mixture is heated to boiling and cooled until solidified. The isolated bodies in the phantom included circles, ellipses, rectangular, and triangular. Multiple conductivity maps are obtained by changing the size, number, direction, and position of the isolated

bodies. After measuring 50 pairs of images, the measurement data was augmented to 750 pairs through operations such as cropping, rotating, flipping, scaling, and translating.

The mask images in the dataset are annotated by medical professionals through the Labelme labeling software to ensure the accuracy.

3.3. MAET-SAM

In this subsection, we illustrated the structure of MAET-SAM in detail. First, we reviewed the main modules and training strategies of SAM briefly. Then, we introduced the framework of MAET-SAM. Eventually, we dived into the adaptive decoder for MAET image segmentation.

3.3.1. SAM

SAM is a deep learning-based general image segmentation model that can generate masks for any object in any image. To train this model, a data engine was constructed that iteratively uses the model to assist in data collection and improves the model with newly collected data. This led to the creation of a massive segmentation dataset, known as SA-1B, which includes over a billion masks and 11 million images. SAM primarily consists of an image encoder, a prompt encoder, and a mask decoder.

The image encoder is used to embed the input image into a feature vector. SAM employs the Vision Transformer (ViT) backbone network as its image encoder:

$$x^l = \mathrm{MLP}\left(\mathrm{MSA}\left(x^{l-1} + x_{pos}^{l-1}\right) + x^{l-1}\right), l = 1, \cdots L. \tag{3}$$

Specifically, the image encoder takes the input image and partitions it into patches with a size of 16 × 16. These patches are projected into patch embeddings x through MLPs. Then, the patch embeddings, adding the positional embeddings x_{pos}, are fed into the multi-head attention operation MSA and MLP to extract high-dimensional features. The positional embedding provides positional information for each patch in the image, enabling the model to understand the spatial information in the image and perform image segmentation more accurately. SAM utilizes multiple ViT blocks to process these features, with each block enhancing feature representation while preserving sequence information. Ultimately, we obtain a high-dimensional image embedding (e.g., with a dimension of 64 × 64 × 256) containing rich semantic information and detailed features. This embedding is then channel-adjusted for subsequent segmentation tasks. Meta offers three scale image encoders, ViT-H, ViT-L, and ViT-B, providing a balance between time and accuracy.

The prompt encoder is responsible for embedding various types of prompts into feature vectors. SAM supports sparse and dense prompts. Sparse prompts include point, box, and text. Point and box prompts are directly embedded into features, while text prompts are embedded using the Contrastive Language-Image Pretraining (CLIP) encoder. Dense prompts involve masks, which are embedded through convolution and then element-wise summed with the image embedding.

The mask decoder maps image embeddings and prompt embeddings to segmentation masks. It consists of multiple attention mechanisms and Multi-Layer Perceptron (MLP) layers to interact between prompt embeddings and image embeddings. The image embeddings then upsample through several layers and are ultimately mapped to a linear classifier.

3.3.2. Architecture of MAET-SAM

The main characteristic of SAM is its ability to perform zero-shot learning and few-shot learning in a new image segmentation domain. However, due to the absence of specialized medical images in the dataset, SAM exhibits suboptimal performance in medical image segmentation. We train MAET-SAM using the established dataset MAET-IMAGE to learn capability for segmenting MAET images. Furthermore, due to the challenge of providing precise prompts in most medical image segmentation cases, we freeze the image encoder of the SAM, remove the prompt encoder, and adjust the mask decoder. This modification enables the model to produce highly accurate segmentation results without any prompts.

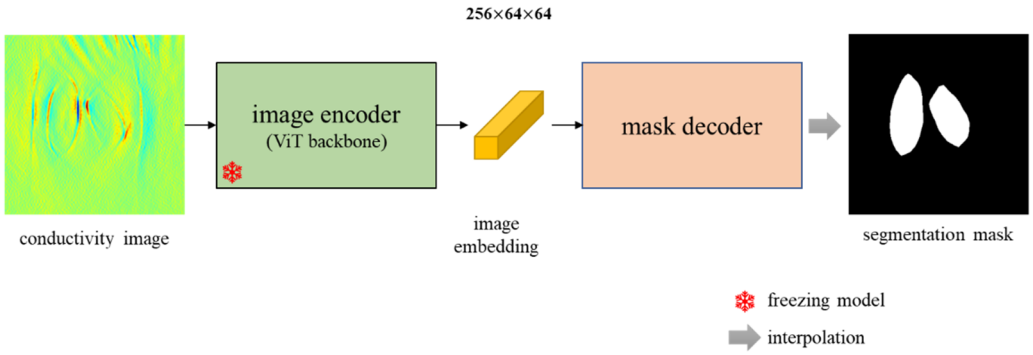

Figure 5. The framework of MAET-SAM. Froze the image encoder of the SAM, removed the prompt encoder, and adjusted the mask decoder for generating the masks of MAET images.

Figure 5 shows the overview of MAET-SAM. Since MAET image segmentation distinguishes only two regions with different conductivity, there is no need for additional prompt information. Additionally, the original mask decoder in SAM also needs to be adjusted for generating the masks of MAET images. Specifically, the MAET image with any resolution can be fed into the frozen image encoder to extract a 256 × 64 × 64 image embedding. Then, the adaptive mask decoder converts the image embedding into a 2 × 256 × 256 feature matrix. The detailed description of the adaptive mask can be found in Section 3.3.3.

3.3.3. Adaptive decoder of MAET-SAM

In the mask decoder of MAET-SAM shown in Figure 6, the image embedding is initially passed through a convolutional network to enhance feature dimensions. Subsequently, we use a transpose convolutional layer to reduce feature dimensions and increase spatial dimensions:

$$I^l = \text{ReLU}(\text{BN}(\text{TransConv}(\text{ReLU}(\text{BN}(\text{Conv}(I^{l-1}))))))), l = 1, \cdots, 3, \quad (4)$$

where I indicates the image embedding, Conv is the convolution operation, TransConv is the transpose convolution operation, BN is the batchnorm, and ReLU is the activation function. The purpose of this design is to ensure that the decoder takes into account both the global and fine-grained details of the image during feature fusion, thereby enhancing the accuracy of image segmentation. After repeating the aforementioned steps in 2 times, the dimension of the image embedding

becomes 256 × 256 × 256. Our aim is to obtain a 2-dimensional feature matrix, with each dimension representing scores for foreground and background. This step can be achieved using convolutional layers to change the dimensionality of the features. We can obtain a mask by selecting the indices of the max score. Finally, the segmentation mask is recovered to the size of the input using the bilinear interpolation. The bilinear interpolation estimates unknown pixel positions by averaging the weighted values of the four nearest neighboring pixels in the horizontal and vertical directions.

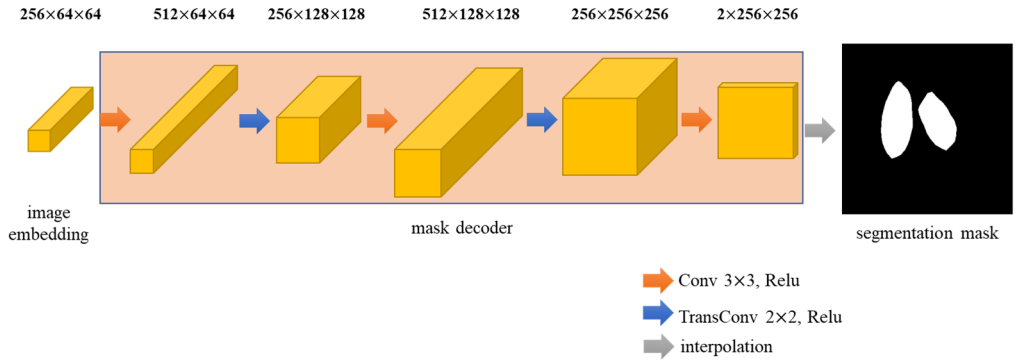

Figure 6. Schematic illustration of the decoder of MAET-SAM.

3.3.4. Loss function

In terms of the loss function, image segmentation can be viewed as a pixel-wise classification task. Therefore, we employed cross-entropy as the loss function:

$$L_{CE} = -\frac{1}{N}\sum_{i=0}^{N-1} y_i \, log(\hat{y}_i), \tag{5}$$

where $y = [\hat{y}_0, \ldots, \hat{y}_{N-1}]$ is a probability distribution, and each element $\hat{y}_i$ represents the probability of a sample belonging to class i. $y = [y_0, \ldots, y_{N-1}]$ is the sample label, and y_i indicates the sample belongs to class i. N represents the segmentation classes in the image. Additionally, to address the issue of imbalance between positive and negative samples (foreground and background) in the segmentation images, we introduced the dice loss function:

$$L_{\text{Dice}} = 1 - \frac{2\sum_{i=0}^{N-1} y_i \hat{y}_i}{\sum_{i=0}^{N-1}(y_i)^2 + \sum_{i=0}^{N-1}(\hat{y}_i)^2}. \tag{6}$$

The final loss function L is:

$$L = L_{CE} + L_{Dice}. \tag{7}$$

4. Experiment

We conducted quantitative and qualitative comparisons on the MAET-IMAGE dataset among the OTSU segmentation algorithm, Region-growing segmentation algorithm, U-Net segmentation model, DeepLabV3+ segmentation model, SAM pre-trained segmentation model, and the MAET-SAM segmentation model proposed in this paper. The OTSU and Region-growing segmentation algorithm belong to traditional segmentation methods, while the U-Net and DeepLabV3+ segmentation model

belong to deep learning methods trained from initial weights. SAM falls under the category of pre-trained deep learning methods. Besides, we also validated the significance of the established dataset and the framework of MAET-SAM through ablation experiments.

4.1. Baselines

1) OTSU [43]: The OTSU segmentation algorithm is an adaptive thresholding technique used to divide an image into foreground and background components. It relies on the image's grayscale histogram to find the optimal threshold globally, minimizing intra-class variance or maximizing inter-class variance between the two segmented categories.

2) Region-growing [44]: The region-growing segmentation algorithm is an image segmentation method based on pixel similarity. This technique starts from defined seed points or regions and gradually merges neighboring pixels that are similar to the current region, continuing until the similarity condition is no longer met.

3) U-Net [35]: U-Net is a widely used deep learning architecture for image segmentation, which consists of an encoder with down-sampling operations and a decoder with up-sampling operations. Its distinctive U-shaped design and skip connection mechanism allow U-Net to effectively utilize both local and global information of images, resulting in strong feature representation capabilities.

4) DeepLabV3+ [45]: DeepLabV3+ is an effective segmentation model designed to capture multi-scale information from images. The model's encoder consists of convolutional neural networks with dilated convolutions and spatial pyramid pooling modules. In the decoder, lower-level features are fused with higher-level features to further improve segmentation accuracy.

5) TransUNet [39]: TransUNet combines the strengths of Transformer and U-Net by using the CNN-Transformer hybrid as the image feature extractor. The CNN-Transformer hybrid consists of ResNet-50 and ViT models pre-trained on ImageNet. Furthermore, the cascaded upsampler with skip-connections enables feature aggregation at different resolution levels. In our experiments, we select the R50-ViT-B_16 as the pre-trained encoder.

6) Swin-Unet [40]: Swin-Unet is an Unet-like pure Transformer. The encoder is a hierarchical Swin Transformer with shifted windows, and the decoder is a symmetric Swin Transformer-based structure. During the training period, the weights pre-trained on ImageNet are used to initialize the model parameters.

7) SAM [8]: SAM is a promptable general model designed for image segmentation. It exhibits strong generalization capabilities of zero-shot and few-shot learning. In our experiments, we adopt points as the prompt for the SAM model. We first obtain the arbitrary pixel coordinates of harmful tissues manually using the setMouseCallback function in OpenCV, and then the SAM model takes them as inputs.

4.2. Dataset

In the experiments, we adopted the established MAET-IMAGE as our experimental dataset. We divided the real measurement images into training, validation, and testing sets with 8:1:1 ratio. Finally, we evaluated the model that performed the best on the validation set using the testing set, and the metrics were averaged over the testing data.

4.3. Experimental setups

To ensure fairness, we employed a standardized experimental platform to compare all methods. The experimental platform includes the Ubuntu 18.04 operating system, Intel(R) Xeon(R) Gold 6240 CPU, and a 32 GB Tesla V100 GPU. Traditional segmentation methods were implemented using OpenCV library, while segmentation neural network models were constructed using PyTorch.

During network training, we utilized the Adam optimizer and ReduceLROnPlateau learning rate decay strategy. The initial learning rate was set to 5e-4, weight decay to 1e-4, and the batch size to 32. A total of 120 epochs were trained. We loaded the pre-trained ViT-B encoder of the SAM model as the image feature extraction module. To address the limited sample issue, we applied data augmentation techniques such as cropping, rotation, flipping, scaling, and translation to improve the generalization capability of the model.

To evaluate the MAET image segmentation results, we selected pixel-wise segmentation accuracy (Acc), Intersection over Union (IoU), Dice, and 95% Hausdorff Distance (HD95) as the segmentation metrics. Specifically, the model classifies pixels in mask images as 1 for the foreground area (harmful tissues) or 0 for the background area. Then, True Positive (TP) indicates the number of foreground pixels classified as foreground, False Positive (FP) indicates the number of background pixels misclassified as foreground, False Negative (FN) indicates the number of foreground pixels misclassified as background, and True Negative (TN) indicates the number of background pixels classified as background. Acc represents the correct positive and negative predictions divided by the total number of predictions:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}. \tag{8}$$

Due to the imbalance between the foreground and background areas in MAET segmentation images, the Acc metric is significantly influenced by the background area. Therefore, we introduce IoU and Dice metrics. IoU calculates the ratio of the intersection between the predicted segmentation region and the ground truth to the union region:

$$IoU = \frac{TP}{TP+FP+FN}. \tag{9}$$

Dice measures the ratio of the intersection between the predicted segmented region and the ground truth to the total region:

$$Dice = \frac{2TP}{2TP+FP+FN}. \tag{10}$$

The difference between IoU and Dice metrics is that the IoU penalizes under- and over-segmentation more than Dice. HD95 evaluates the edge accuracy of the segmentation result.

4.4. Quantitative analysis

Table 1 lists the segmentation results of the eight methods on the MAET-IMAGE dataset. We can observe that traditional segmentation methods have lower segmentation performance, whereas deep learning methods exhibit higher segmentation accuracy and better semantic understanding on these

images. This highlights the advantage of deep learning on complex image segmentation tasks. Compared to initial weight-trained segmentation models (U-Net and DeepLabV3+), the general segmentation model SAM, trained on a large set of natural images, demonstrates certain segmentation advantages. This suggests that the segmentation capabilities of the network are significantly influenced by the model architecture and dataset size. Since the initial weights of TransUNet and Swin-Unet are pre-trained on ImageNet, their segmentation performance surpasses that of U-Net and DeepLabV3+. However, the size of the MAET-IMAGE dataset limits the capability of Transformer. Specifically, the proposed segmentation method MAET-SAM outperforms the SAM across various metrics. This indicates that the established MAET-IMAGE dataset and the adaptive decoder are effective for the specific medical image segmentation.

Table 1. The segmentation results on MAET-IMAGE. mIoU represents the average intersection over union between predicted and ground truth images. Fore-IoU and Back-IoU represent the intersection over union for the foreground and background segmentation, respectively.

Methods	Metrics (%)					
	Acc↑	mIoU↑	Fore-IoU↑	Back-IoU↑	Dice↑	HD95↓
OTSU thresholding	49.5	28.1	8.3	47.9	43.9	81.5
Region growing	59.6	41.3	12.8	69.8	58.5	82.3
U-Net	84.4	77.5	59.9	95.2	87.3	31.7
DeepLabV3+	89.2	83.0	69.4	96.6	90.7	26.1
TransUNet	91.4	86.5	76.5	96.5	93.2	21.5
Swin-Unet	91.8	87.9	78.9	96.9	93.4	20.3
SAM (ViT-B)	92.7	88.5	79.3	97.7	93.9	21.6
MAET-SAM	**97.1**	**92.7**	**86.9**	**98.5**	**96.2**	**16.3**

4.5. Qualitative analysis

Figure 7 shows the segmentation results of the eight methods on four real measurement images. We can observe it due to the low signal-to-noise ratio of MAET images, segmentation methods based on global threshold OTSU and pixel similarity Region growing struggle to differentiate between the foreground and background. Although U-Net and DeepLabV3+ segmentation models can distinguish the foreground and background of the images, the limited training samples lead to significant errors in the segmentation results. TransUNet and Swin-Unet exhibit similar segmentation results, but it is hard to distinguish multi-regions. Since SAM is a large model trained on millions of natural images, it can roughly localize lesions regions in MAET images. However, it may not achieve optimal results at the edges of lesions. Furthermore, the SAM requires pixel coordinates as prompts, which is quite inconvenient for MAET image segmentation tasks. The proposed MAET-SAM effectively improves the segmentation performance of detailed regions on a small-scale dataset.

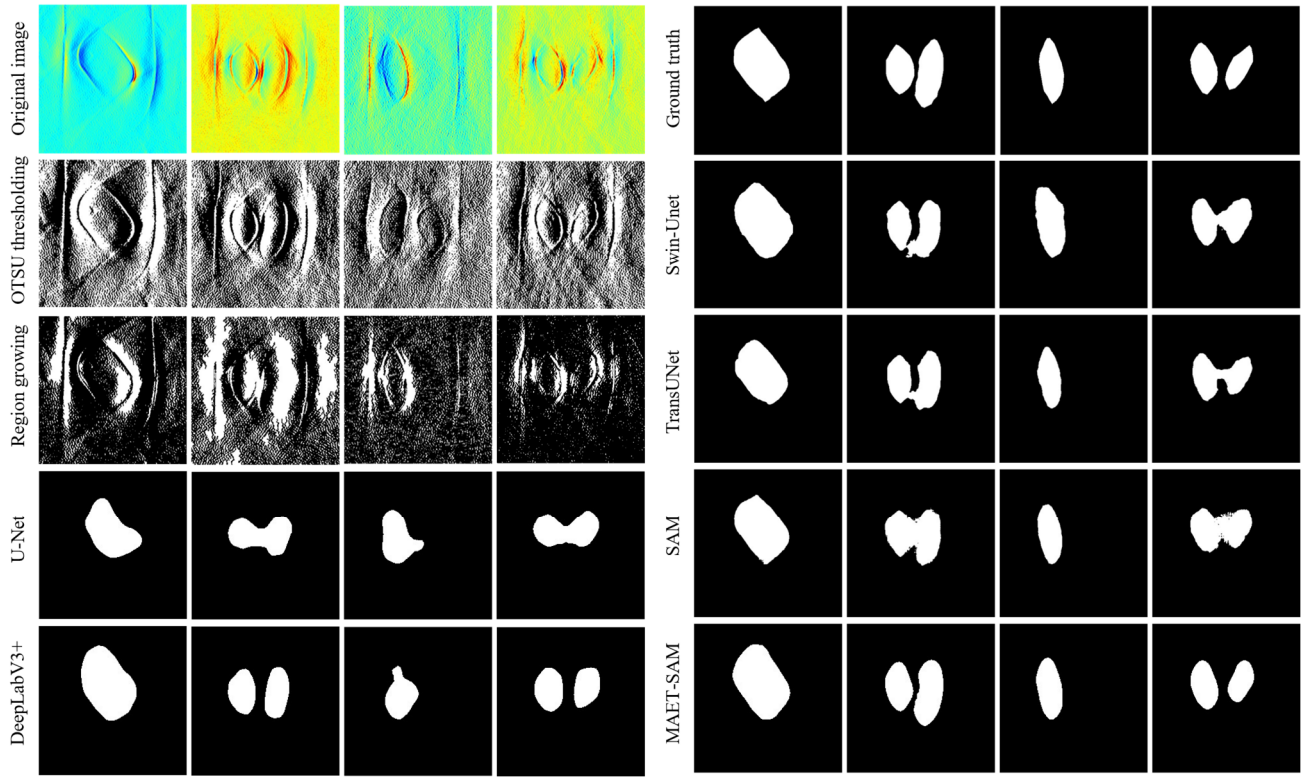

Figure 7. The visualization of the segmentation results of the eight methods on four real measurement images.

4.6. Ablation study

We conducted ablation experiments to further validate the effectiveness of the established dataset and the framework of MAET-SAM for MAET image segmentation. Unless otherwise specified, the experimental settings remain consistent with those previously described.

4.6.1. Components of dataset

In this experiment, we employed four different datasets to train MAET-SAM, as illustrated in Table 2. Without dataset utilization, MAET-SAM achieved a segmentation accuracy of only 45.3% on MAET images. The accuracies obtained through individual training on the simulated dataset, or the test dataset were inferior to the results achieved by training with both datasets simultaneously. This observation verifies the effectiveness of the dataset constructed in this study.

Table 2. The influence of different dataset training strategies on MAET image segmentation.

Datasets		Metric (%)
2000 simulated images	750 measured images	mIoU
×	×	45.3
√	×	85.1
×	√	90.4
√	√	**92.7**

4.6.2. Backbone

We employ the pre-trained encoder to extract embeddings of the conductivity images, while the embeddings are crucial for downstream segmentation tasks. To validate the rationality of the MAET-SAM network architecture, we use different pre-trained encoders for the MAET image segmentation task. The pre-trained encoders are presented in Table 3. We can observe that convolution-based pre-trained encoders exhibit comparable accuracy in MAET image segmentation. However, the ViT-B encoder outperforms others by a large margin. This is partly attributed to the superior feature learning capability of the attention mechanism in the ViT-B pre-trained encoder. Additionally, ViT-B is trained on millions of segmentation images, unlike other pretrained encoders trained on ImageNet-1k. Considering the task consistency and dataset scale, the utilization of ViT-B for the backbone of MAET-SAM is the optimal choice.

Table 3. The influence of different backbones on MAET image segmentation.

Backbones	ResNet152 [29]	DenseNet201 [30]	EfficientNet-b4 [31]	ViT-B [8]
mIoU (%)	75.5	78.6	79.6	92.7

5. Conclusions

Electrical characteristics of biological tissues can reflect their structure, function, physiology, and pathology. Thus, the MAET, which can reflect the conductivity distribution of biology tissues, holds significant importance for current clinical early inspection, prevention and treatment. However, the process of MAET imaging is deeply influenced by system and environmental noise, resulting in issues like blurred boundary, low contrast, and irregular artifacts. The focus of current research is how to accurately extract the conductivity information of the lesion area from MAET images to improve image quality and accurately assess tissue conditions. Despite the fact that medical image segmentation techniques are well developed, there is currently no method specifically for segmenting MAET images.

In this paper, we established a dataset called MAET-IMAGE, composed of conductivity maps reconstructed by MAET. The dataset contains 2000 pairs of simulated and 750 pairs of real measured conductivity-mask maps, which can be used for training segmentation network models. Given the limited sample size of the MAET-IMAGE dataset, a deep learning network MAET-SAM were proposed. In this model, we froze the encoder parameters trained on a dataset of millions. This approach can extract features and semantic information from MAET images effectively and provide richer information for downstream segmentation tasks. Additionally, since the MAET images requires only distinguishing between foreground and background, there is no need for prompts. Therefore, we propose a prompt-free decoder. The decoder is composed of multiple convolutional layers and transposed convolutional layers, enabling it to output the mask of the MAET image without the need for any prompts. In the experiments, we compared five baseline methods with our proposed MAET-SAM. The results showed that traditional segmentation methods do not perform well in segmenting MAET images, which was due to the low signal-to-noise ratio. Segmentation networks trained with initial weights can form a regional shape segmentation effects. Due to the limited number of dataset samples, these segmentation models have low segmentation accuracy. The large model SAM can roughly cover the lesion area, however, there are significant segmentation errors in detail. By training MAET-SAM using the MAET-IMAGE dataset constructed in this paper, MAET-SAM significantly

improved the inability of SAM to accurately segment details. We also validated the significance of the established dataset and the framework of MAET-SAM through ablation experiments. Therefore, the proposed MAET-SAM can accurately extract the lesion area and provide more comprehensive conductivity segmentation results, bringing new breakthroughs and progress in areas such as clinical diagnosis and treatment decision-making.

While MAET-SAM performs well in conductivity image segmentation, it is not perfect. Due to the time-consuming and high-cost nature of real measured conductivity images, the MEAT-IMAGE dataset contains parts of simulated conductivity images. Obtaining more real measured conductivity images could further improve the segmentation accuracy. Additionally, the MAET-SAM network can currently only segment reconstructed images with two different conductivities. In future research, we plan to increase the number of different conductivity tissues in the reconstructed images and output the segmentation results for these tissues.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant Nos 52377227, 52007182 and 51937010.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. X. Song, Y. Xu, F. Dong, R. S. White, An instrumental electrode configuration for 3-D ultrasound modulated electrical impedance tomography, *IEEE Sens. J.*, **17** (2017), 8206–8214. https://doi.org/10.1109/JSEN.2017.2706758
2. P. Grasland-Mongrain, C. Lafon, Review on biomedical techniques for imaging electrical impedance, *IRBM*, **39** (2018), 243–250. https://doi.org/10.1016/j.irbm.2018.06.001
3. H. Wen, J. Shah, R. S. Balaban, Hall effect imaging, *IEEE Trans. Biomed. Eng.*, **45** (1998), 119–124. https://doi.org/10.1109/10.650364
4. H. Wen, R. S. Balaban, The potential for hall effect breast imaging, *Breast Dis.*, **10** (1998), 191–195. https://doi.org/10.3233/BD-1998-103-418
5. T. M. Deserno, H. Handels, K. H. Maier-Hein, S. Mersmann, C. Palm, T. Tolxdorff, Viewpoints on medical image processing: from science to application, *Curr. Med. Imaging*, **9** (2013), 79–88. https://doi.org/10.2174/1573405611309020002
6. D. L. Pham, C. Xu, J. L. Prince, Current methods in medical image segmentation, *Ann. Rev. Biomed. Eng.*, **2** (2000), 315–337. https://doi.org/10.1146/annurev.bioeng.2.1.315

7. R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, A. K. Nandi, Medical image segmentation using deep learning: A survey. *IET Image Process.*, **16** (2022), 1243–1267. https://doi.org/10.1049/ipr2.12419
8. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, et al., Segment anything, in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, (2023), 4015–4026. https://doi.org/10.1109/ICCV51070.2023.00371
9. A. Montalibet, J. Jossinet, A. Matias, Scanning electric conductivity gradients with ultrasonically-induced Lorentz force, *Ultrason. Imaging*, **23** (2001), 117–132. https://doi.org/10.1177/016173460102300204
10. S. Haider, A. Hrbek, Y. Xu, Magneto-acousto-electrical tomography: A potential method for imaging current density and electrical impedance, *Phys. Meas.*, **29** (2008), S41. https://doi.rog/10.1088/0967-3334/29/6/S04
11. P. Grasland-Mongrain, J. M. Mari, J. Y. Chapelon, C. Lafon, Lorentz force electrical impedance tomography, *IRBM*, **34** (2013), 357–360. https://doi.org/10.1016/j.irbm.2013.08.002
12. L. Guo, G. Liu, H. Xia, Magneto-acousto-electrical tomography with magnetic induction for conductivity reconstruction, *IEEE Trans. Biomed. Eng.*, **62** (2015), 2114–2124. https://doi.org/10.1109/TBME.2014.2382562
13. L. Guo, G. F. Liu, Y. J. Yang, G. Q. Liu, Vector based reconstruction method in Magneto-Acousto-Electrical Tomography with magnetic induction, *Chin. Phys. Lett.*, **32** (2015), 094301. https://doi.org/10.1088/0256-307X/32/9/094301
14. L. Kunyansky, C. P. Ingram, R. S. Witte, Rotational magneto-acousto-electric tomography (MAET): theory and experimental validation, *Phys. Med. Biol.*, **62** (2017), 3025. https://doi.org/10.1088/1361-6560/aa6222
15. M. Dai, X. Chen, M. Chen, H. Lin, F. Li, S. Chen, A novel method to detect interface of conductivity changes in Magneto-Acousto-Electrical tomography using chirp signal excitation method, *IEEE Access*, **6** (2018), 33503–33512. https://doi.org/10.1109/ACCESS.2018.2841991
16. M. Dai, X. Chen, M. Chen, H. Lin, F. Li, S. Chen, A 2D magneto-acousto-electrical tomography method to detect conductivity variation using multifocus image method, *Sensors*, **18** (2018), 2373. https://doi.org/10.3390/s18072373
17. Z. F. Yu, Y. Zhou, Y. Z. Li, Q. Y. Ma, G. P. Guo, J. Tu, Performance improvement of magneto-acousto-electrical tomography for biological tissues with sinusoid-Barker coded excitation, *Chin. Phys. B*, **27** (2018), 094302. https://doi.org/10.1088/1674-1056/27/9/094302
18. T. Sun, X. Zeng, P. Hao, C. T. Chin, M. Chen, J. Yan, et al., Optimization of multi-angle magneto-acousto-electrical tomography (MAET) based on a numerical method, *Math. Biosci. Eng.*, **17** (2020), 2864–2880. https://doi.org/10.3934/mbe.2020161
19. D. Deng, T. Sun, L. Yu, Y. Chen, X. Chen, M. Chen, et al., Image quality improvement of magneto-acousto-electrical tomography with Barker coded excitation, *Biomed. Signal Process. Control*, **77** (2022), 103823. https://doi.org/10.1016/j.bspc.2022.103823
20. P. Li, W. Chen, G. Guo, J. Tu, D. Zhang, Q. Ma, General principle and optimization of magneto-acousto-electrical tomography based on image distortion evaluation, *Med. Phys.*, **50** (2023), 3076–3091. https://doi.org/10.1002/mp.16317
21. Y. Li, S. Bu, X. Han, H. Xia, W. Ren, G. Liu, Magneto-acousto-electrical tomography with nonuniform static magnetic field, *IEEE Trans. Instrum. Meas.*, **72** (2023), 1–12. https://doi.org/10.1109/TIM.2023.3244814

22. P. K. Sahoo, S. Soltani, A. K. C. Wong, A survey of thresholding techniques, *Comput. Vis. Graph Image Process.*, **41** (1998), 233–260. https://doi.org/10.1016/0734-189X(88)90022-9
23. J. K. Udupa, S. Samarasekera, Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation, *Graph Models Image Process.*, **58** (1996), 246–261. https://doi.org/10.1006/gmip.1996.0021
24. P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **33** (2010), 898–916. https://doi.org/10.1109/TPAMI.2010.161
25. D. Shen, G. Wu, H. I. Suk, Deep learning in medical image analysis, *Ann. Rev. Biomed. Eng.*, **19** (2017), 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442
26. Q. Huang, H. Tian, L. Jia, Z. Li, Z. Zhou, A review of deep learning segmentation methods for carotid artery ultrasound images, *Neurocomputing*, **545** (2023), 126298. https://doi.org/10.1016/j.neucom.2023.126298
27. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems*, 2012.
28. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv:1409.1556.
29. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770–778.
30. G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. https://doi.org/10.1109/CVPR.2017.243
31. M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, preprint, arXiv:1905.11946.
32. A. Dosovitskiy, An image is worth 16 × 16 words: Transformers for image recognition at scale, preprint, arXiv:2010.11929.
33. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, Swin transformer: Hierarchical vision transformer using shifted windows, in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, (2021), 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00986
34. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015), 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965
35. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
36. N. Siddique, S. Paheding, C. P. Elkin, V. Devabhaktuni, U-Net and its variants for medical image segmentation: A review of theory and applications, *IEEE Access*, **9** (2021), 82031–82057. https://doi.org/10.1109/ACCESS.2021.3086020
37. X. Liu, Y. Fan, S. Li, M. Chen, M. Li, W. K. Hau, et al., Deep learning-based automated left ventricular ejection fraction assessment using 2-D echocardiography, *Am. J. Physiol-Heart Circ. Physiol.*, **321** (2021), H390–H399. https://doi.org/10.1152/ajpheart.00416.2020
38. A. Vaswani, Attention is all you need, *Adv. Neural Inf. Process. Syst.*, **2017** (2017).
39. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, TransUNet: Transformers make strong encoders for medical image segmentation, preprint, arXiv: 210204306.

40. H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, et al., Swin-Unet: Unet-like pure transformer for medical image segmentation, in *Computer Vision–ECCV 2022 Workshops*, (2022) 205–218. https://doi.org/10.1007/978-3-031-25066-8_9
41. A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, D. Zhang, DS-TransUNet: Dual swin transformer U-Net for medical image segmentation, *IEEE Trans. Instrum. Meas.*, **71** (2022), 1–15. https://doi.org/10.1109/TIM.2022.3178991
42. Q. Huang, L. Zhao, G. Ren, X. Wang, C. Liu, W. Wang, NAG-Net: Nested attention-guided learning for segmentation of carotid lumen-intima interface and media-adventitia interface, *Comput. Biol. Med.*, **156** (2023), 106718. https://doi.org/10.1016/j.compbiomed.2023.106718
43. N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man. Cybern.*, **9** (1979), 62–66. https://doi.org/10.1109/TSMC.1979.4310076
44. R. Adams, L. Bischof, Seeded region growing, *IEEE Trans. Pattern Anal. Mach. Intell.*, **16** (1994), 641–647. https://doi.org/10.1109/34.295913
45. L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 801–818.

AIMS Press