MitoStructSeg: A Comprehensive Platform for Mitochondrial Structure Segmentation and Analysis

Xinsheng Wang[1], Buqing Cai[1], Zhuo Jia[1], Yuanbo Chen[1],
Shuai Guo[2], Zheng Liu[2], Xiaohua Wan[1*], Fa Zhang[1*] Bin Hu[1]
[1]Beijing Institute of Technology, Beijing, China
[2] Southern University of Science and Technology, Shenzhen, China
Corresponding authors: Xiaohua Wan(wanxiaohua@bit.edu.cn), Fa Zhang(zhangfa@bit.edu.cn)

Mitochondria are crucial organelles within cells, playing key roles in various biological processes, particularly in energy conversion, cell death, and signal transduction. Mitochondria consist of an outer membrane and inner cristae, the latter being essential for energy conversion. Consequently, quantitative analysis of the inner cristae not only reveals the functional state of mitochondria but also highlights its role in cellular metabolism and pathological conditions. However, there is currently a shortage of effective tools. This paper introduces MitoStructSeg, a deep learning-based platform for the segmentation and quantitative analysis of mitochondrial structures. Among these, the AMM-Seg model is proposed for mitochondrial structure segmentation, surpassing current state-of-the-art (SOTA) methods. Quantitative analysis of segmentation results elucidates the relationship between mitochondrial health and cristae structure. In addition, a user-friendly open source tool is available.

Mitochondria occupy a pivotal position in cell biology, often metaphorically referred to as the "powerhouses of the cell." This designation arises from their fundamental role as the main source of energy for most cells, achieved through the efficient generation of adenosine triphosphate[1,2] via the oxidative phosphorylation process. This provides essential power for various cellular activities. However, the significance of mitochondria extends well beyond mere energy production. They play an indispensable role in the regulation of programmed cell death[3–6], the generation of reactive oxygen species (ROS), the transmission of calcium signals, and numerous metabolic pathways.

The capacity of mitochondria to execute these complex functions can be attributed to their distinctive structure—comprising an inner and an outer membrane, which encircle the matrix, with the inner membrane further folding into cristae[7,8]. Investigating mitochondrial structure not only aids in understanding how they efficiently conduct energy conversion, but also unveils their role in broader functions, including cell death and survival signaling regulation. Furthermore, abnormalities in mitochondrial structure are linked with numerous diseases[9–11], making the study of their structure essential for the development of new therapeutic methods. The researchers capture images of mitochondria using scanning electron microscopy[11,12] and analyze these images to gather structural information. However, traditional research methods are heavily based on manual observation, which is time consuming and labor intensive, and can lead to inaccuracies due to the lack of uniform evaluation standards and subjective biases.

With the widespread application and development of deep learning technology in various fields, its potential in biomedical image analysis has become increasingly apparent. Image segmentation technology[13,14], a key application of deep learning, has become a powerful tool for extracting and identifying structures of interest in medical images, enabling automatic differentiation and localization of specific cellular structures or lesions from complex biological tissue images. CDeep3M[15] is an innovative image segmentation tool specifically designed for biomedical image processing, offering significant technical advantages. Its core lies in the deep convolutional neural network model implemented based on cloud computing, allowing users to directly utilize prebuilt models for image segmentation without the need for complex environment configuration and model training. This feature significantly simplifies the application of deep learning in the image segmentation process. DeepContact[16] is a deep learning model specifically designed for quantifying membrane contact sites (MCSs) in electron microscopy imaging. Its features include high-precision automatic segmentation and the ability to accurately handle small and complex MCSs. It adopts the architecture of a convolutional neural network, including an input layer, multiple convolutional layers, activation function, pooling layer, fully connected layer, and output layer. These layers work together to effectively extract and analyze MCSs from electron microscope images. PHILOW[17] is a novel deep learning platform tailored for image analysis. This system utilizes a human-in-the-loop algorithm to meticulously analyze large volumes of data from focused ion beam-scanning electron microscopy (FIB-SEM). PHILOW has been instrumental in elucidating the intricate three-dimensional nanostructures of the inner and outer membranes of mitochondria. It offers detailed quantitative insights into the structural attributes of cristae in numerous mitochondria. Although the above methods can identify and segment the complete structure of mitochondria, demonstrating their ability to recognize the overall morphology, they have obvious shortcomings in segmenting the more detailed internal structure of mitochondria. In particular, they cannot precisely segment the outer membrane and inner cristae of mitochondria. Furthermore, although they provide statistical analysis of

mitochondrial structural information, such as surface areas and volumes, they lack comparative analysis between the outer membrane and the inner cristae. Thus, there is a pressing need for advanced segmentation techniques that can accurately and independently analyze these critical substructures of mitochondria.

In this study, we introduce MitoStructSeg, a pioneering all-in-one software platform for mitochondrial structure analysis Fig.1. MitoStructSeg integrates a mitochondrial segmentation algorithm module, a mitochondrial statistical analysis module, and an interface and interaction module, providing a comprehensive solution to meet the diverse needs of researchers. The mitochondrial segmentation algorithm module is based on our proposed AMM-Seg network architecture, achieving efficient segmentation by dividing data into labeled and unlabeled domains, applying two-channel image processing, and utilizing data augmentation and texture enhancement techniques. The AMM-Seg network extracts and integrates feature representations to enhance texture features and capture dynamic characteristics. The statistical analysis module employs advanced algorithms for fast, accurate, and quantitative assessments of mitochondrial structural information, offering immediate analytical insights. The 3D visualization module utilizes 3D Slicer software to reconstruct and visualize mitochondrial structures in three dimensions, supporting multi-angle and multi-level exploration. Our main contributions include the introduction of AMM-Seg for precise segmentation of the mitochondrial outer membrane and inner cristae, the development of MitoStructSeg as a comprehensive analysis platform, validation of its efficacy on real-world datasets, and its provision as open-source software with a user-friendly GUI. In summary, these features establish MitoStructSeg as a powerful and adaptable platform, significantly enhancing mitochondrial research across both fundamental scientific studies and clinical applications.

Results

Benchmark against existing state-of-the-art methods

The AMM-Seg model uses 200 labeled images (800x800 pixel) from Patient 1 as the source domain data, while 200 unlabeled images (800x800 pixel) are employed as the target domain data for training. The target domain is derived from other patients' data to more effectively evaluate the model's generalization ability on different types of data. Subsequently, the model segmentation performance is validated on 40 labeled images (800x800 pixel) from the target domain. We utilize two distinct metrics for evaluation: (1) the F1 score, a harmonic mean of precision and recall, combines precision and recall to assess the model's accuracy in identifying true positives. (2) Intersection-over-union (IoU) measures the overlap between the predicted segmentation and the true segmentation. It is calculated by dividing the area of the intersection of the predicted and true regions by the area of the union.

We conducted a comparative analysis of the segmentation performance of AMM-Seg against six benchmark methods, including MEDIAR[18,19], DA-ISC[20], UNet[21], Swin-UNet[22], and ViT-UNet, a hybrid method integrating Vision Transformer (ViT)[23] and UNet architectures. These benchmark methods utilize advanced segmentation techniques, incorporating multimodal data processing, domain adaptation strategies, and novel neural network architectures to improve accuracy and performance in medical image analysis. The results of the comparative experiment are shown in Fig.2a. Overall, AMM-Seg demonstrates superior performance in terms of the F1 score in three datasets, achieving high and stable results. MEDIAR and DA-ISC follow closely, whereas UNet, Swin-UNet, and ViT-UNet have lower median scores and a wider range of distribution, indicating less stable performance.

Original data were acquired through scanning electron microscopy imaging, which exhibits a high degree of continuity between adjacent images. Therefore, when using a neural network model to segment mitochondria, it is essential not only to accurately segment the inner cristae and outer membranes of mitochondria but also to ensure the continuity of segmentation results between adjacent images. Fig.2b shows the continuity results of adjacent image segmentation results for each method in three different datasets. AMM-Seg and DA-ISC demonstrate exceptional performance in maintaining the continuity of adjacent slices, even surpassing the level of human annotation. MEDIAR also performs commendably, albeit slightly inferior to AMM-Seg and DA-ISC, yet still exceeding human annotation. However, the performance of UNet, ViT-UNet, and Swin-UNet is not satisfactory, exhibiting low and highly variable IoU values, indicating significant deficiencies in maintaining the continuity of adjacent images in the SEM imaging data.

Fig.3 compares the visual segmentation results of AMM-Seg with those of the best baseline method(DA-ISC) and human annotation, by analyzing three datasets of slice images (labeled a, b, and c). The AMM-Seg method demonstrates excellent boundary recognition accuracy, clearly delineating the structure of the mitochondrial outer membrane and inner cristae, with reduced instances of over-segmentation or under-segmentation. It also demonstrates an exceptional ability to preserve intricate details in complex biological structures, crucial for medical diagnostic accuracy. This is particularly evident in part a and b, where textures and layers are portrayed more effectively. overall, the segmentation results obtained by the algorithm reach the level of human experts annotation.

Comparison of segmentation results in different resolutions

To further verify the effectiveness of our model, we specifically perform segmentation prediction analysis on im-

ages with varying resolutions. The experimental steps are as follows: we start with a selected original image and successively reduce it to 1/2, 1/4, and 1/8 of its original size to simulate different resolution levels. During the image reduction process, we choose the LANCZOS filter, which effectively reduces the aliasing effects that may be introduced by the reduction operation. Subsequently, to ensure processing consistency, we use nearest-neighbor interpolation to restore all reduced images to a unified 800x800 resolution. These specially processed images are then used for model training and prediction to assess model adaptability to inputs of different resolutions. As illustrated in Fig.4, we display images processed by different reduction factors along with their corresponding model prediction results. From the results, it is observable that although the image details gradually blur due to the reduction in resolution, the prediction output of the model does not show significant degradation. The model continues to accurately identify and segment the outer membrane and inner cristae of mitochondria. This consistency demonstrates the high generalization ability of our model in processing images of different resolutions.

Additionally, we compared the performance of our model with the DA-ISC method under these varying resolutions. The results, illustrated in Fig.4, indicate that while both models show a decline in performance with decreasing resolution, AMM-Seg consistently outperforms DA-ISC. This suggests that AMM-Seg is better equipped to handle information loss due to resolution reduction, which is critical for practical applications where high-resolution images may not always be available. In terms of sensitivity, which measures the true positive rate, AMM-Seg demonstrated superior capability in identifying relevant features in the images, regardless of the resolution. This aspect is crucial for medical applications where missing a critical feature can have significant consequences. The consistent sensitivity across resolutions highlights the model's reliability and its potential in clinical settings.

Overall, our experiments demonstrate the adaptability and reliability of the AMM-Seg model in processing images of varying resolutions. Even when image details are blurred by reduced resolution, AMM-Seg continues to effectively identify and segment key mitochondrial structures, a critical capability for ensuring accurate and reliable mitochondrial statistical quantification.

Quantitative statistical analysis of mitochondria in healthy and damaged states

We selected 80-100 mitochondria from each of three patient datasets for quantitative analysis. The statistical results are presented in Tables 1, 2, and 3, respectively. The results indicate significant differences in the outer membrane and inner cristae surface areas between healthy and damaged states. In the case of patient 1, the mean outer and inner surface areas of mitochondria were 1.85 μm^2 and 10.42 μm^2 in the healthy state, compared to 0.82 μm^2 and 3.88 μm^2 in the damaged state, respectively. This significant reduction in the area of mitochondria suggests structural damage. The average ratio of outer membrane to inner cristae surface area was 0.23 in the healthy state and 0.66 in the damaged state. This suggests more irregular morphology or structural alterations in mitochondria during the damaged state. Further data analysis revealed varied mitochondrial response patterns across patients. Similar trends were observed in patients 2 and 3, where the outer membrane and inner cristae surface areas were significantly reduced in the damaged state. This consistent pattern underscores the strong correlation between mitochondrial damage and disease progression. Additionally, the increased ratio of outer membrane to inner cristae surface in the damaged state corresponded with observed morphological changes in mitochondria, such as the collapse of the inner cristae. The statistical analysis demonstrated that quantitative evaluations of mitochondrial structures provided further evidence of the correlation between alterations in intramitochondrial cristae and mitochondrial damage.

Interface and interaction

For researchers, a user-friendly interactive interface is important. Fig.5 illustrates several key user interfaces of the system. The system provides an efficient and user-friendly interface. Users simply need to input the required initial files into area 1 of the interface. After the initialization process, users are prompted to select the workflow for image segmentation, as shown in Figure 5Fig.5(a) (area 2). The system automatically preprocesses the images and then utilizes advanced deep learning techniques and image processing algorithms to segment the outer membrane and inner ridges of mitochondria in 2D images. The processed images are displayed in real time on the interface, as depicted in Fig.5(b) (area 3). Users can also interact with the interface to fine-tune and confirm the segmentation results, as shown in Fig.5(b) (area 4). Subsequently, users can choose the computation workflow, as shown in Fig.5(c) (area 5), to obtain the lengths of the outer membrane and inner ridges of each mitochondrion in a single image and the total sums of the outer membrane and inner ridges of each mitochondrion in all images. The results of the calculation can be viewed on the right side of the main interface, as illustrated in Fig.5(c) (area 6). Based on the interactive design, users can click to select a mitochondrion, such as those shown in Fig.5(d) (area 7), to view its data information in Fig.5(d) (area 8).

Discussion

Mitochondria are vital organelles within cells, primarily responsible for energy production and playing pivotal roles in various biological processes including cell death, cell proliferation, and metabolic regulation. The structure of mitochondria is complex, comprising an outer

TABLE 1: Patient1 dataset mitochondria properties.

		Outer surface (μm^2)	Inner surface (μm^2)	Surface (μm^2)	Outer/Inner surface	Volume (μm^3)	MCI
Max	Health	6.85	60.08	66.93	0.53	1.59	0.99
	Demage	2.94	21.26	24.20	1.01	0.54	0.97
Min	Health	0.19	0.36	0.56	0.11	0.01	0.18
	Demage	0.23	0.23	0.46	0.47	0.01	0.38
Median	Health	1.72	8.91	1.93	0.20	0.24	0.58
	Demage	0.44	0.87	1.15	0.55	0.26	0.55
Mean	Health	1.85	10.42	12.27	0.23	0.27	0.63
	Demage	0.82	3.88	3.37	0.66	0.11	0.60
SD	Health	1.16	9.25	10.41	0.09	0.24	0.27
	Demage	0.85	8.13	6.71	0.16	0.17	0.18

TABLE 2: Patient2 dataset mitochondria properties.

		Outer surface (μm^2)	Inner surface (μm^2)	Surface (μm^2)	Outer/Inner surface	Volume (μm^3)	MCI
Max	Health	3.38	12.84	15.47	0.45	0.47	0.92
	Demage	3.19	9.88	13.07	0.73	0.33	0.89
Min	Health	0.62	1.37	2.00	0.20	0.05	0.42
	Demage	0.68	0.94	1.63	0.32	0.05	0.48
Median	Health	1.14	4.51	5.66	0.31	0.14	0.56
	Demage	1.28	2.98	4.23	0.45	0.12	0.71
Mean	Health	1.54	5.60	7.14	0.31	0.19	0.73
	Demage	1.63	4.07	5.70	0.48	0.15	0.69
SD	Health	0.84	3.52	4.31	0.07	0.13	0.42
	Demage	0.85	3.05	3.89	0.13	0.09	0.17

membrane and inner cristae, each with unique biological functions. Accurate segmentation of mitochondrial outer membrane and inner cristae is crucial for thoroughly understanding their functions and associated pathological states. Existing mitochondrial segmentation technologies fail to accurately segment the outer membrane and inner cristae and fall short in quantitative analysis of mitochondrial structure, lacking a comprehensive suite of analysis and evaluation mechanisms.

In this study, we developed a deep learning-based MitoStructSeg platform, which contains three integral components designed to optimize mitochondrial analysis: the Adaptive Multi-Domain Mitochondrial Segmentation (AMM-Seg) module, the mitochondrial statistical analysis module, and the Interface and Interaction module.

The AMM-Seg module, a cornerstone of this platform, employs a sophisticated deep learning model, specifically developed to accurately segment both the outer membranes and the inner cristae of mitochondria. Currently, the model enhances its segmentation ability on unlabeled data through a domain adaptation strategy. During processing, AMM-Seg first receives the original images from the source and target domains and conducts rigorous data preprocessing to produce enhanced and textured versions of these images. Subsequently, the encoder accurately extracts unique features from both images. These features are then inputted into the adaptive fusion module, which integrates the two set of characteristics to create a comprehensive and enhanced feature representation. Finally, the decoder utilizes this enhanced feature representation to generate accurate mitochondrial segmentation results. In this process, the source and target domain images share network parameters, and are alternately trained, enabling the model to effectively transfer learning between domains, thereby enhancing its generalization capabilities and practicality in new scenarios. Compared to several widely-used image-based segmentation methods, AMM-Seg offers higher segmentation accuracy, greater adaptability to variations in image quality. Furthermore, AMM-Seg's domain adaptation capabilities enable it to perform effectively even on datasets with limited or unlabeled data, significantly reducing dependency on extensive annotated datasets and facilitating more scalable and flexible applications in medical imaging and other fields. Following segmentation, the mitochondrial statistical analysis module quantitatively assesses the results. It provides comprehensive metrics, including surface area and volume, among others, enabling precise evaluations of mitochondrial morphology. This module is crucial for deriving meaningful insights from the complex data generated by AMM-Seg. Together, these modules integrate seamlessly with the Interface and Interaction module, designed to enhance user engagement through intuitive controls and efficient data management. This component ensures easy system navigation, data manipulation, and the extraction of actionable insights from mitochondrial analysis.

Although the MitoStructSeg platform offers more ac-

TABLE 3: Patient3 dataset mitochondria properties.

		Outer surface (μm^2)	Inner surface (μm^2)	Surface (μm^2)	Outer/Inner surface	Volume (μm^3)	MCI
Max	Health	2.16	9.58	11.72	0.63	0.30	0.90
	Demage	0.61	1.08	1.69	1.06	0.06	0.68
Min	Health	0.43	0.69	1.13	0.22	0.04	0.36
	Demage	0.22	0.21	0.43	0.50	0.01	0.36
Median	Health	1.08	3.57	4.80	0.31	0.13	0.64
	Demage	0.38	0.68	1.06	0.61	0.03	0.41
Mean	Health	1.23	4.20	5.43	0.36	0.14	0.63
	Demage	0.40	0.67	1.07	0.67	0.03	0.47
SD	Health	0.56	2.82	3.36	0.12	0.09	0.15
	Demage	0.13	0.27	0.40	0.18	0.02	0.12

curate segmentation capabilities and more comprehensive quantitative analysis functions—significantly advancing mitochondrial research—it still confronts challenges such as data variability, limited adaptability to various imaging technologies, and the necessity to process unlabeled data. To address these challenges, we can enhance the model's self-learning capability by training with large-scale unlabeled datasets and integrating pre-trained models optimized for specific fields. These measures will significantly enhance the model's generalization capabilities in practical applications, enabling it to better adapt to evolving application environments and data conditions.

Methods

Imaging of myocardial slices

The myocardial tissue samples were sourced from patients infected with COVID-19, and the process for obtaining the data required for model training is as follows. Myocardial biopsy samples from COVID-19 patients were rapidly fixed in 4% paraformaldehyde (PFA) for 24 hours. After fixation, the tissues were dehydrated in a graded alcohol series (70%, 80%, 95%, and 100%), cleared in xylene, and embedded in paraffin. The paraffin blocks were quickly cooled to solidify, and sections of 5-6 µm thickness were cut using a microtome, spread in 45°C warm water, mounted on slides, dried at 45°C, and stored at room temperature. For deparaffinization and rehydration, the sections were baked at 60°C for 2 hours, followed by immersion in xylene (two changes for 5 minutes each), then 100% ethanol, 95% ethanol, 80% ethanol, and 70% ethanol (each for 5 minutes), and finally washed in ddH_2O. The sections were stained with Weigert's iron hematoxylin for 5-10 minutes, washed thoroughly, and differentiated with acid alcohol if overstained. They were then blued with Masson's bluing solution for 3-5 minutes, washed with distilled water for 1 minute, and finally stained with Ponceau-Fuchsin for 5-10 minutes. Sections were stained with hematoxylin for 1 minute, washed with ddH2O, differentiated in 1% hydrochloric acid for 10 seconds to remove excess stain, blued in 1% ammonia water for 5 seconds, and counterstained with eosin for 2 minutes. Samples were cut into 1mm x 1mm pieces and pre-fixed in a mixture of 2% paraformaldehyde and 2.5% glutaraldehyde (GA) at pH 7.2 for 12 hours. For high-pressure freezing (HPF), samples were loaded into copper carriers with cryoprotectant hexadecene, rapidly frozen under high pressure (approximately 2000 bar), and then transferred to a freeze substitution device. Freeze substitution was performed in a solution containing osmium tetroxide and uranyl acetate at -90°C, with gradual warming to 0°C over several hours, followed by washing with acetone. Samples underwent OTO (osmium-thiocarbohydrazide-osmium) block staining. They were fixed in 1% osmium tetroxide with potassium ferrocyanide for 1 hour on ice, washed in phosphate buffer, then ddH2O, treated with 1% thiocarbohydrazide, and post-fixed in 1% osmium tetroxide again. After uranium acetate staining and lead citrate treatment, samples were dehydrated through graded ethanol and acetone series and infiltrated with Epon 812 resin for embedding. Ultrathin sections (70 nm) were cut using an ultramicrotome, collected on copper grids, and stained with 3% uranyl acetate and lead citrate. Sections were imaged using a 120 kV electron microscope.

According to the above process, we obtained the data from three patients. The data sizes are 6144x4096x800 voxels for Patient 1, 6144x4096x800 voxels for Patient 2, and 6500x5200x800 voxels for Patient 3. The pixel size is 5x5nm and the slice interval is 10nm. Initially, we cropped two regions (800x800x100 voxels) from the data of Patient 1 and manually annotated them to create the source domain dataset (with labels). Subsequently, we cropped a region (800x800x200 voxels) from the data of Patients 1, 2, and 3 respectively to form the target domain dataset (without labels). Additionally, we cropped a region (800x800x40 voxels) from the data of these three patients to create the test set for evaluating the segmentation accuracy and generalization performance of AMM-Seg.

AMM-Seg

An overview of the model is depicted in Fig.1b.

Preliminaries

In the task of mitochondrial segmentation, most public datasets are unlabeled. There are almost no public datasets for the more refined mitochondrial structure segmentation task (simultaneously segmenting the mitochondrial outer membrane and cristae). Therefore, we want to design a domain adaptation learning algorithm that can not only segment mitochondrial structures, but also migrate to other datasets to fill the gap in mitochondrial structure segmentation tasks. In the source domain, we have electron microscopy (EM) image data $\boldsymbol{X}^S = \{\boldsymbol{x}_i^S\}_{i=1}^M$, where M is the number of sequential images. And we have a label set $\boldsymbol{L}^S = \{l_i^S \in \{0,1\}\}_{i=1}^M$ corresponding to $\boldsymbol{X}^S$, where 0 represents the foreground and 1 represents the background. In the target domain, we only have EM image data $\boldsymbol{X}^T = \{\boldsymbol{x}_i^T\}_{i=1}^N$, where N is the number of sequential images.

Augmented image preprocess. In our study, we implemented a data augmentation module for the mitochondrial dataset to address overfitting and improve the generalization capability of the model across diverse morphological features and imaging conditions. Due to the distinct data distributions in the source and target domains, it is necessary to adopt different augmentation strategies for each domain to boost the model's overall generalization performance. Specifically, the augmentation strategy for the source domain is designed to prevent overfitting, whereas the strategy for the target domain is focused on enhancing the model's generalization ability. Therefore, tailored augmentation approaches are essential for both the source and target domains.The weak geometrical augmentation $\mathcal{A}_g$, as shown in Table **??**, includes standard resizing, cropping, and flipping operations. We apply only $\mathcal{A}_g$ to the source domain as follows:

$$\boldsymbol{a}_i^S = \mathcal{A}_{\mathrm{g}}(\boldsymbol{x}_i^S), \tag{1}$$

$$\boldsymbol{a}_{i+m}^S = \mathcal{A}_{\mathrm{g}}(\boldsymbol{x}_{i+m}^S). \tag{2}$$

Furthermore, we design a random intensity-based augmentation, denoted by $\mathcal{A}_r$, which only apply on the target domain as follows:

$$\boldsymbol{a}_i^T = \mathcal{A}_{\mathrm{r}}(\boldsymbol{x}_i^T), \tag{3}$$

$$\boldsymbol{a}_{i+n}^T = \mathcal{A}_{\mathrm{r}}(\boldsymbol{x}_{i+n}^T). \tag{4}$$

Textured image preprocess The structure of mitochondria is mainly composed of the outer membrane and the inner membrane, where the inner membrane further forms a complex folded structure called cristae. Therefore, to accurately segment mitochondrial structures, it is crucial to extract effective texture characteristics.

Based on the first step of extracting the mitochondrial region, the mitochondrial texture information is further extracted using a high-pass filter. Therefore, we can obtain the texture features of mitochondria in the source domain as follows:

TABLE 4: List of various image transformations in the weak geometrical augmentation and random intensity-based augmentation.

	Weak Geometrical Augmentation
Flip	Randomly flip images horizontally or vertically.
Rotate	Rotation of images at random angles.
Crop	Randomly crops a region from the image.
	Intensity-based Augmentation
Contrast	Randomly adjust the contrast of the image.
MotionBlur	Randomly motion blurring the image.
Sharpness	Randomly adjusts the sharpness of the image.
Elastic	Randomly elastically deform the image.
Brightness	Randomly adjust the brightness of the image.

$$\boldsymbol{f}_i^S = \mathrm{Filter}(\boldsymbol{s}_i^S), \tag{5}$$

$$\boldsymbol{f}_{i+m}^S = \mathrm{Filter}(\boldsymbol{s}_{i+m}^S). \tag{6}$$

where Filter$(\cdot)$ is the high-pass filter used to enhance the details of the image, such as Canny, Sobel, etc. Simlarity, we can obtain texure features $\boldsymbol{f}_i^T$ and $\boldsymbol{f}_{i+n}^T$ in the target domain.

Encoder. Since there is a changing relationship between the 2D images of the scanning electron microscope images, continuous image features are spliced during the downsampling process to capture the changing features between the 2D images. The detail of Encoder as follows:

$$\mathcal{D}(\boldsymbol{a}_i^S, \boldsymbol{a}_{i+m}^S) = \mathrm{Down}(\mathrm{Down}(\boldsymbol{a}_i^S) \oplus \mathrm{Down}(\boldsymbol{a}_{i+m}^S)), \tag{7}$$

where $\oplus$ is the concatenation operation and Down$(\cdot)$ is the downsampling operation.Similarity, we can obtain encoder $\mathcal{D}(\boldsymbol{f}_i^S, \boldsymbol{f}_{i+m}^S)$ for texture features.

Adaptive Fusion. In order to more effectively integrate the original features and texture features, we adopt a adaptive fusion strategy to combine them together. This enables the network to simultaneously consider both the basic information of the image and higher-level texture information, and to achieve dynamic feature selection, thereby adaptively retaining important feature information. As shown below:

$$\boldsymbol{Q}^S = W_q \cdot \mathcal{D}(\boldsymbol{a}_i^S, \boldsymbol{a}_{i+m}^S), \tag{8}$$

$$\boldsymbol{K}^S = W_k \cdot \mathcal{D}(\boldsymbol{f}_i^S, \boldsymbol{f}_{i+m}^S), \tag{9}$$

$$\boldsymbol{V}^S = W_v \cdot \mathcal{D}(\boldsymbol{f}_i^S, \boldsymbol{f}_{i+m}^S), \tag{10}$$

where W_q, W_k, W_v are learnable parameter matrix. Next, the self-attention mechanism[24] is used to fuse the original features and texture features.

$$\boldsymbol{e}_i^S = \mathcal{D}(\boldsymbol{a}_i^S, \boldsymbol{a}_{i+m}^S) + \mathrm{softmax}\left(\frac{Q^S(K^S)^T}{\sqrt{d_k}}\right) V^S. \tag{11}$$

Similarly, we can obtain the encoding result $\boldsymbol{e}_i^T$ in the target domain. It is worth noting that the encoder of

the original and texture features adopts a weight-sharing strategy to reduce model complexity and improve computational efficiency.

Decoder. In the encoder section, to capture the continuous changing features of the image, we encode the concatenated features. Therefore, in the decoder section, we used two decoders for different targets, one for decoding the segmentation results and the other for decoding the results of continuous image changes.Specifically, we upsample the encoder results separately as follows:

$$\boldsymbol{p}_i^S = \mathrm{Up}_{\mathrm{seg}}(\boldsymbol{e}_i^S, \mathrm{Down}(\boldsymbol{a}_i^S), \mathrm{Down}(\boldsymbol{f}_i^S)), \tag{12}$$

$$\boldsymbol{p}_{i+m}^S = \mathrm{Up}_{\mathrm{seg}}(\boldsymbol{e}_{i+m}^S, \mathrm{Down}(\boldsymbol{a}_{i+m}^S), \mathrm{Down}(\boldsymbol{f}_{i+m}^S)), \tag{13}$$

$$\boldsymbol{r}_i^S = \mathrm{Up}_{\mathrm{diff}}(\boldsymbol{e}_i^S, \boldsymbol{e}_{i+m}^S), \tag{14}$$

where $\mathrm{Up}_{seg}(\cdot)$ and $\mathrm{Up}_{diff}(\cdot)$ are upsampling operations, but they differ in network structure. $\boldsymbol{p}_i^S$ and $\boldsymbol{p}_{i+m}^S$ is the segmentation result and $\boldsymbol{r}_i^S$ is the feature of image variation. Similarity allows us to obtain $\boldsymbol{p}_i^T$, $\boldsymbol{p}_{i+n}^T$, and $\boldsymbol{r}_i^T$ for the target domain.

Loss Function

In this section, we introduce the loss functions used during model training. For the source domain dataset, where true labels are available, the loss function can be calculated directly. However, for the target domain dataset, which lacks true labels, we construct pseudo-labels to compute the loss function.

Source Domain Loss. After obtaining the predicted results, we need to consider the model's training approach from two perspectives: segmentation performance and the continuity of image changes. Therefore, the loss function is formulated to constrain these two aspects, as follows:

$$\mathcal{L}_{S1} = CE(\boldsymbol{p}_i^S, \boldsymbol{l}_i^S) + CE(p_{i+m}^S, l_{i+m}^S), \tag{15}$$

$$\mathcal{L}_{S2} = CE(\boldsymbol{r}_i^S, \mathrm{XOR}(\boldsymbol{l}_i^S, \boldsymbol{l}_{i+m}^S)) \tag{16}$$

where $\mathrm{CE}(\cdot)$ is the cross-entropy loss function and XOR is the exclusive OR operation.

Target Domain Loss. Due to the lack of real labels in the target domain, we are unable to directly calculate the loss function. To solve this problem, we generate pseudo-labels based on the predicted results of the model that are similar to the real labels. Through this approach, we can overcome the challenge of missing target domain labels and improve the model's generalization ability and performance in the target domain. The specific details are as follows:

$$\boldsymbol{l}_i^T = |\arg\max(\boldsymbol{p}_{i+n}^T) - \arg\max(\boldsymbol{r}_i^T)| \tag{17}$$

$$\boldsymbol{l}_{i+n}^T = |\arg\max(\boldsymbol{p}_i^T) - \arg\max(\boldsymbol{r}_i^T)| \tag{18}$$

$$\mathcal{L}_T = CE(\boldsymbol{p}_i^T, \boldsymbol{l}_i^T) + CE(p_{i+n}^T, l_{i+n}^T) \tag{19}$$

where $\boldsymbol{l}_i^T$ and $\boldsymbol{l}_{i+n}^T$ are artificially constructed pseudo labels.

Overall Loss. The overall loss contains the two domain losses, so we jointly optimize:

$$\mathcal{L} = \mathcal{L}_{S1} + \mathcal{L}_{S2} + \lambda_1 \mathcal{L}_T, \tag{20}$$

where λ_1 is a hyperparameter.

References

[1] Jianxing Song. Adenosine triphosphate energy-independently controls protein homeostasis with unique structure and diverse mechanisms. *Protein Science*, 30(7): 1277–1293, 2021.

[2] Weihong Guo, Zhian Chen, Xiaoli Feng, Guodong Shen, Huilin Huang, Yanrui Liang, Bingxia Zhao, Guoxin Li, and Yanfeng Hu. Graphene oxide (go)-based nanosheets with combined chemo/photothermal/photodynamic therapy to overcome gastric cancer (gc) paclitaxel resistance by reducing mitochondria-derived adenosine-triphosphate (atp). *Journal of nanobiotechnology*, 19(1):146, 2021.

[3] Florian J Bock and Stephen WG Tait. Mitochondria as multifaceted regulators of cell death. *Nature reviews Molecular cell biology*, 21(2):85–100, 2020.

[4] Anna Picca, Riccardo Calvani, Hélio José Coelho-Junior, and Emanuele Marzetti. Cell death and inflammation: The role of mitochondria in health and disease. *Cells*, 10(3):537, 2021.

[5] Esmee Vringer and Stephen WG Tait. Mitochondria and cell death-associated inflammation. *Cell Death & Differentiation*, 30(2):304–312, 2023.

[6] Zahra Al Amir Dache and Alain R Thierry. Mitochondria-derived cell-to-cell communication. *Cell Reports*, 42(7), 2023.

[7] Werner Kühlbrandt. Structure and function of mitochondrial membrane protein complexes. *BMC biology*, 13:1–11, 2015.

[8] Till Stephan, Christian Brüser, Markus Deckers, Anna M Steyer, Francisco Balzarotti, Mariam Barbot, Tiana S Behr, Gudrun Heim, Wolfgang Hübner, Peter Ilgen, et al. Micos assembly controls mitochondrial inner membrane remodeling and crista junction redistribution to mediate cristae formation. *The EMBO journal*, 39(14):e104105, 2020.

[9] Sabzali Javadov, Andrey V Kozlov, and Amadou KS Camara. Mitochondria in health and diseases, 2020.

[10] Giacomo Monzio Compagnoni, Alessio Di Fonzo, Stefania Corti, Giacomo P Comi, Nereo Bresolin, and Eliezer Masliah. The role of mitochondria in neurodegenerative diseases: the lesson from alzheimer's disease and parkinson's disease. *Molecular neurobiology*, 57:2959–2980, 2020.

[11] Andrea G Marshall, Steven M Damo, and Antentor Hinton. Revisiting focused ion beam scanning electron microscopy. *Trends in Biochemical Sciences*, 48(6):585–586, 2023.

[12] Tanmoy Saha, Chinmayee Dash, Ruparoshni Jayabalan, Sachin Khiste, Arpita Kulkarni, Kiran Kurmi, Jayanta Mondal, Pradip K Majumder, Aditya Bardia, Hae Lin Jang, et al. Intercellular nanotubes mediate mitochondrial trafficking between cancer and immune cells. *Nature nanotechnology*, 17(1):98–106, 2022.

[13] Pengshuai Yin, Rui Yuan, Yiming Cheng, and Qingyao Wu. Deep guidance network for biomedical image segmentation. *IEEE access*, 8:116106–116116, 2020.

[14] Philipp D Lösel, Thomas van de Kamp, Alejandra Jayme, Alexey Ershov, Tomáš Faragó, Olaf Pichler, Nicholas Tan Jerome, Narendar Aadepu, Sabine Bremer, Suren A Chilingaryan, et al. Introducing biomedisa as an open-source online platform for biomedical image segmentation. *Nature communications*, 11(1):5577, 2020.

[15] Matthias G Haberl, Christopher Churas, Lucas Tindall, Daniela Boassa, Sébastien Phan, Eric A Bushong, Matthew Madany, Raffi Akay, Thomas J Deerinck, Steven T Peltier, et al. Cdeep3m—plug-and-play cloud-based deep learning for image segmentation. *Nature methods*, 15(9):677–680, 2018.

[16] Liqing Liu, Shuxin Yang, Yang Liu, Xixia Li, Junjie Hu, Li Xiao, and Tao Xu. Deepcontact: High-throughput quantification of membrane contact sites based on electron microscopy imaging. *Journal of Cell Biology*, 221(9): e202106190, 2022.

[17] Shogo Suga, Koki Nakamura, Yu Nakanishi, Bruno M Humbel, Hiroki Kawai, and Yusuke Hirabayashi. An interactive deep learning-based approach reveals mitochondrial cristae topologies. *PLoS Biology*, 21(8):e3002246, 2023.

[18] Jun Ma, Ronald Xie, Shamini Ayyadhury, Cheng Ge, Anubha Gupta, Ritu Gupta, Song Gu, Yao Zhang, Gihun Lee, Joonkee Kim, et al. The multimodality cell segmentation challenge: toward universal solutions. *Nature methods*, pages 1–11, 2024.

[19] Tongle Fan, Guanglei Wang, Yan Li, and Hongrui Wang. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020.

[20] Wei Huang, Xiaoyu Liu, Zhen Cheng, Yueyi Zhang, and Zhiwei Xiong. Domain adaptive mitochondria segmentation via enforcing inter-section consistency. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 89–98. Springer, 2022.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[22] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

[23] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[24] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

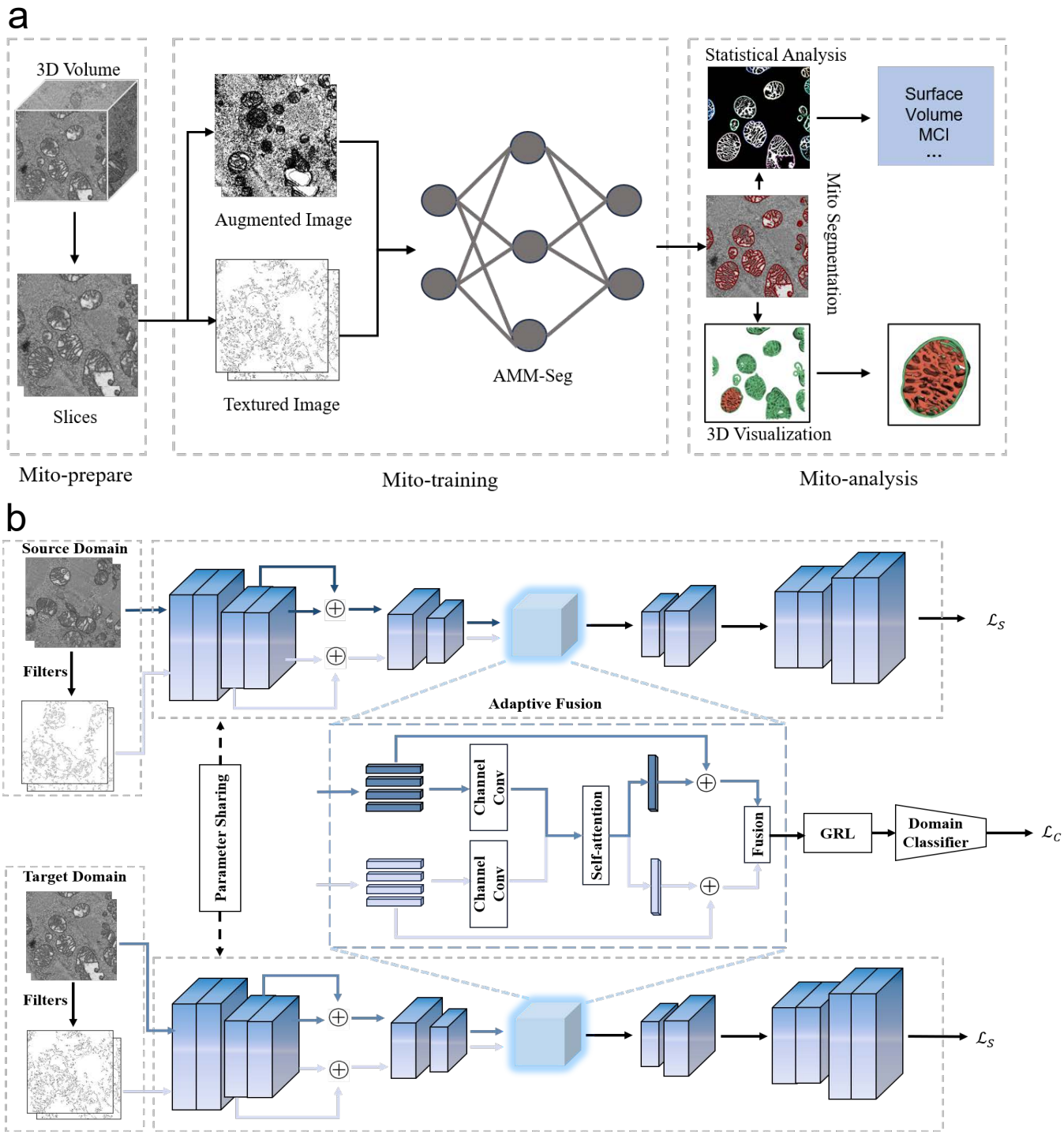

Fig. 1 Overview of MitoStructSeg. **a** Workflow of MitoStructSeg. **b** Summary of the AMM-Seg architecture.

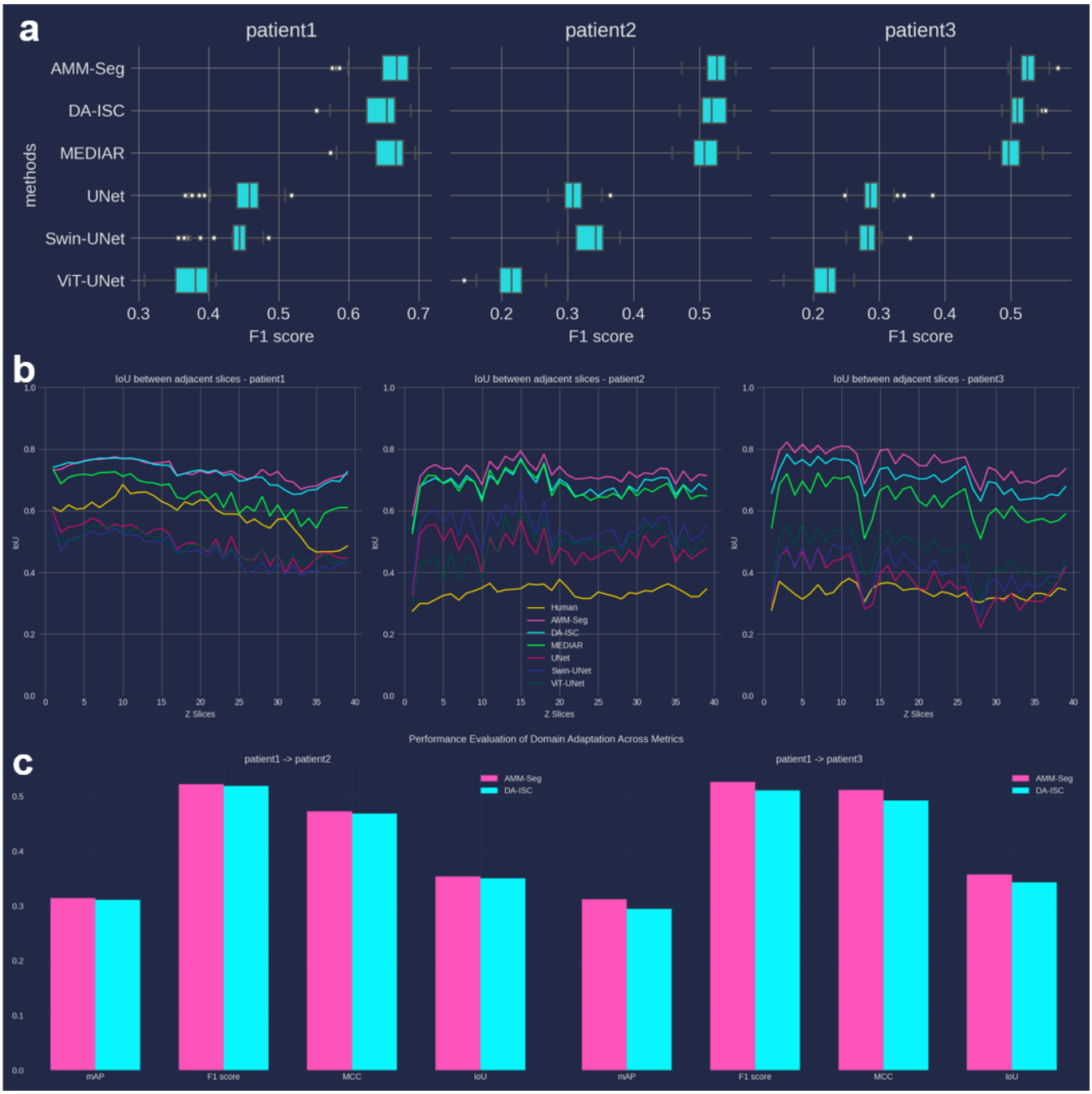

Fig. 2 Performance evaluation of image segmentation method. **a,** Comparison of F1 Scores Across Different Patient Data. Comparison of AMM-Seg with five other image segmentation methods using the correlation benchmark. Each 2D image used for evaluation contributes one data point to the correlation presented in the boxplot. AMM-Seg achieved significantly higher segmentation accuracy than the other methods across three datasets. **b,** IoU Performance Evaluation Between Adjacent Slices. Compare the continuity of 2D image segmentation results between AMM-Seg and six other methods, including human annotation. Overall, AMM-Seg surpasses the other methods in preserving continuity across images within three datasets. **c,** Performance Evaluation of Domain Adaptation Across Metrics. The effects of two domain adaptation methods, AMM-Seg and DA-ISC, are compared using multiple evaluation indicators. On two cross-domain datasets (patient1 to patient2, patient1 to patient3), AMM-Seg outperforms DA-ISC across multiple evaluation indicators.

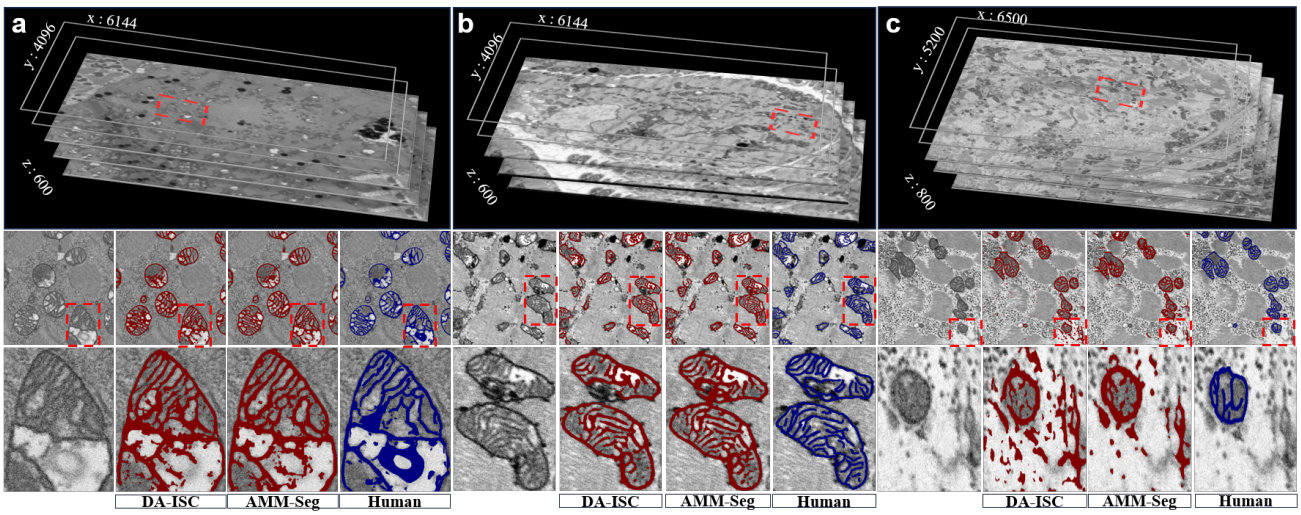

Fig. 3 Visualization of segmentation results. a, Patient1 dataset. FIB-SEM volume acquired at 5nm×5nm voxelsize (6144x4096x600 voxel). We focus on the data within the red dotted box (800x800x600 voxel) for predictive analysis. The segmentation results of AMM-Seg and DA-ISC, including human annotations, are visualized to evaluate the actual effect of segmentation. **b,** Patient2 dataset. **c,** Patient3 dataset.

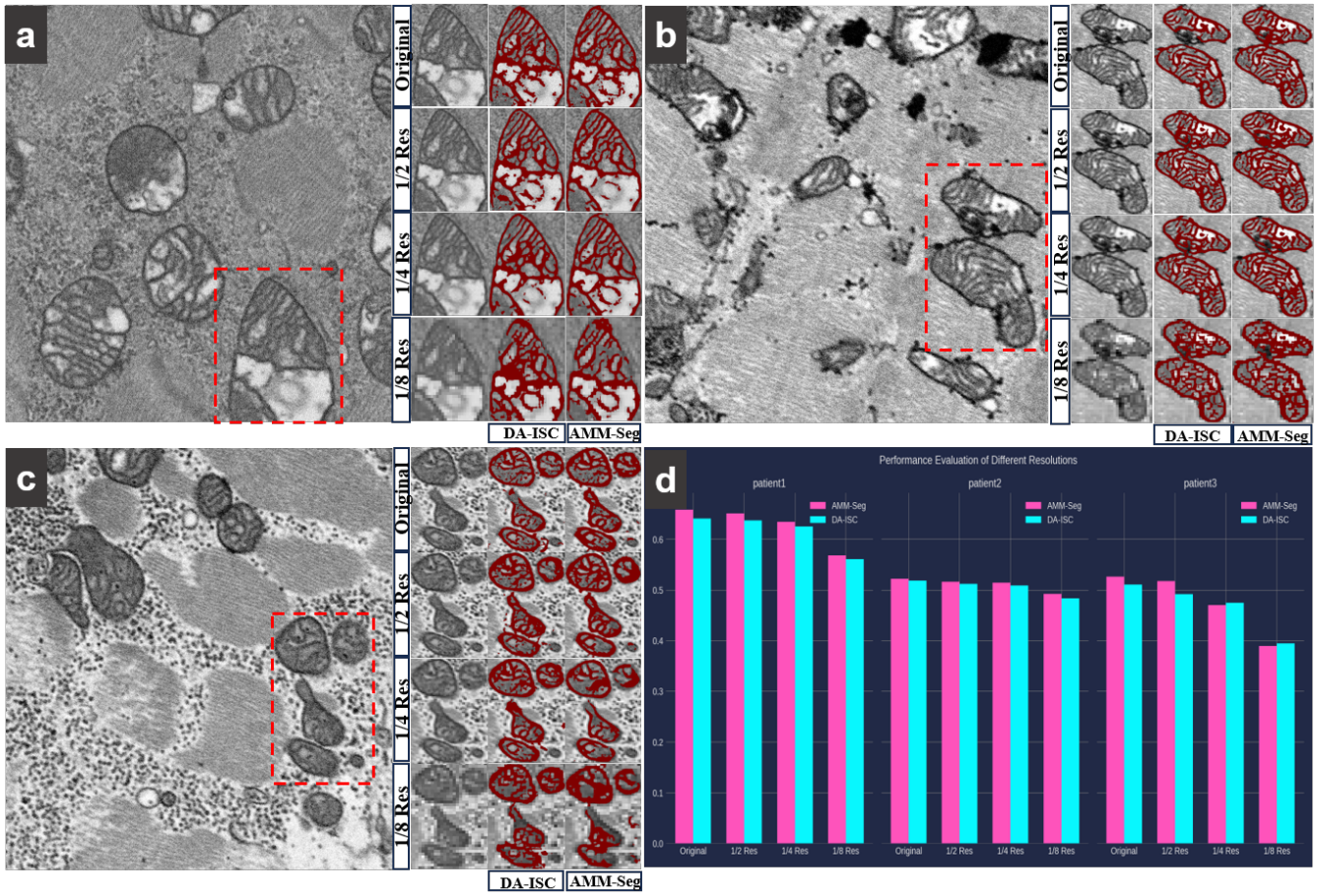

Fig. 4 Comparison of segmentation results at different resolutions. a, Patient1 dataset. **b,** Patient2 dataset. **c,** Patient3 dataset. **d,** The bar chart illustrates the performance evaluation of two segmentation methods, AMM-Seg and DA-ISC, in three patients at different resolutions: original, 1/2 resolution, 1/4 resolution and 1/8 resolution.

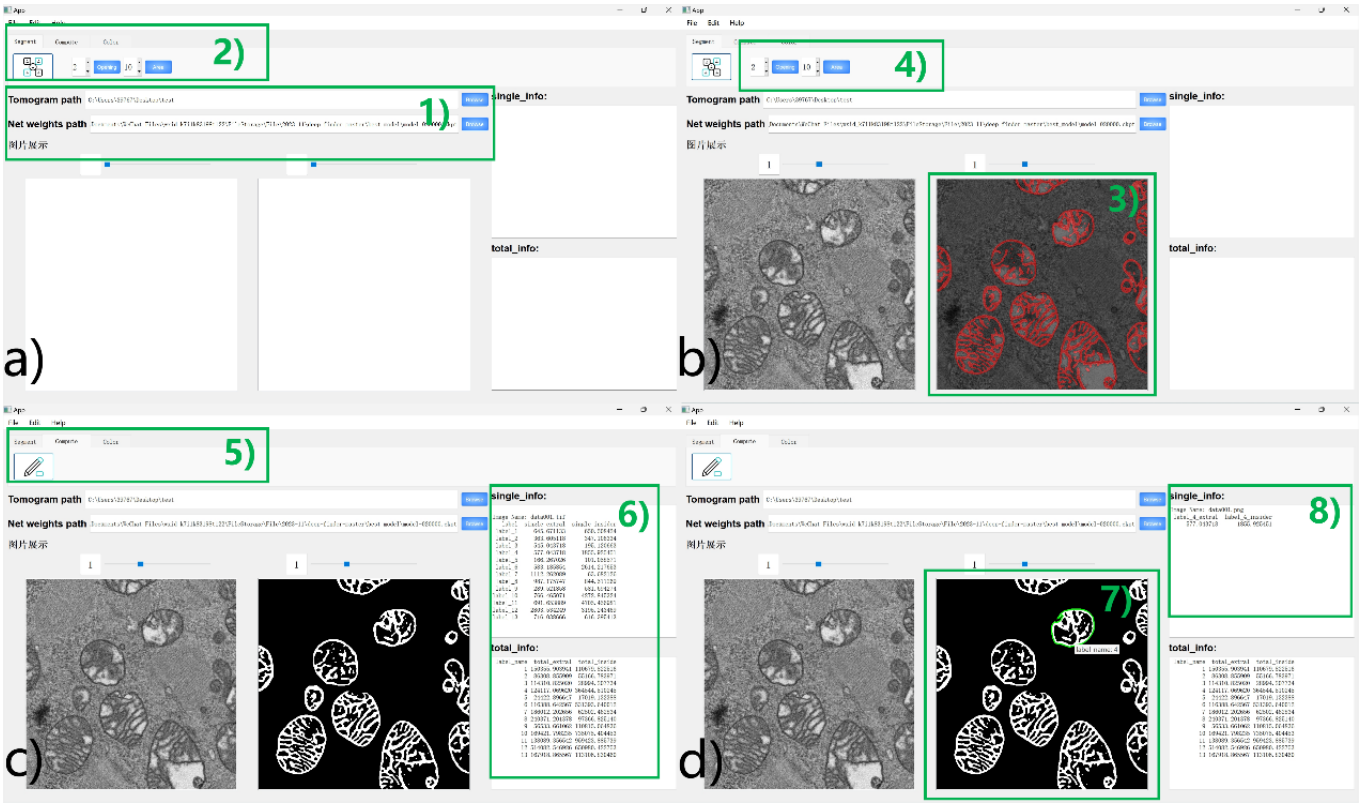

Fig. 5 Illustration of the user interface of system. a, Image Segmentation Module: Perform image segmentation based on the data provided by the user. **b,** Visualization and Post-processing: Visualize the segmented results and allow the user to perform post-processing operations as needed. **c,** Membrane structure calculation module: Based on the results of segmentation, calculate the outer membrane and inner cristae of mitochondria and display the results in real time. **d,** Interactive design: Mouse-click on a specific mitochondrion to obtain data.