Evidential Uncertainty-Guided Mitochondria Segmentation for 3D EM Images

Ruohua Shi[1], **Lingyu Duan**[1,2], **Tiejun Huang**[1,3], **Tingting Jiang**[1,4*]
[1]National Engineering Research Center of Visual Technology, National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University
[2]Peng Cheng Laboratory
[3]Beijing Academy of Artificial Intelligence
[4]National Biomedical Imaging Center, Peking University
{shiruohua,lingyu,tjhuang,ttjiang}@pku.edu.cn

Abstract

Recent advances in deep learning have greatly improved the segmentation of mitochondria from Electron Microscopy (EM) images. However, suffering from variations in mitochondrial morphology, imaging conditions, and image noise, existing methods still exhibit high uncertainty in their predictions. Moreover, in view of our findings, predictions with high levels of uncertainty are often accompanied by inaccuracies such as ambiguous boundaries and amount of false positive segments. To deal with the above problems, we propose a novel approach for mitochondria segmentation in 3D EM images that leverages evidential uncertainty estimation, which for the first time integrates *evidential uncertainty* to enhance the performance of segmentation. To be more specific, our proposed method not only provides accurate segmentation results, but also estimates associated uncertainty. Then, the estimated uncertainty is used to help improve the segmentation performance by an uncertainty rectification module, which leverages uncertainty maps and *multi-scale information* to refine the segmentation. Extensive experiments conducted on four challenging benchmarks demonstrate the superiority of our proposed method over existing approaches.

Introduction

Positioned at the heart of cellular metabolism, mitochondria serve a key role in powering life through massive and varied metabolic functions (Annesley and Fisher 2019; Bock and Tait 2020). Thanks to the Electron microscopy (EM) technique, high-resolution images of mitochondria and other cellular structures are now available, making them a valuable resource for studying cellular biology and connectomics (Casser et al. 2020; Wei et al. 2020; Lucchi et al. 2011). The utilization of deep learning algorithms in mitochondria segmentation has shown significant progress, as demonstrated by state-of-the-art (SOTA) methods (Luo et al. 2021; Peng, Yi, and Yuan 2020; Peng and Yuan 2019; Yuan et al. 2021). Most of these techniques employ the U-Net (Ronneberger, Fischer, and Brox 2015) architecture or its variations (Casser et al. 2020; Mekuč et al. 2020) to address the unique challenges posed by EM image segmentation. Recently, transformer and self-attention (Franco-Barranco, Muñoz-Barrutia, and Arganda-Carreras 2022) have also shown advantages in mitochondria image segmentation (Franco-Barranco, Muñoz-Barrutia, and Arganda-Carreras 2022; Yuan et al. 2020, 2021).

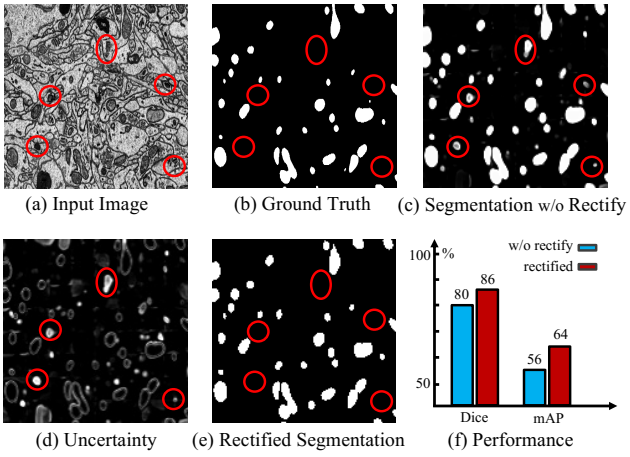

Figure 1: Illustration of the effectiveness of using uncertainty information to modify segmentation results.

Despite the advancements made by deep learning (DL) models, they remain plagued by considerable uncertainty within their predictions. This uncertainty originates from various sources (Guo et al. 2017), including out-of-domain inputs, data quality issues, and nuances of the training conditions. Within the EM images, this issue is exacerbated by the variations in data quality caused by artifacts or deformations during the image acquisition process. Consequently, DL models might produce overconfident but erroneous predictions. These predictions output by the current mitochondria segmentation framework impose limitations on the practical applicability of these models, particularly in 3D cell reconstruction and subsequent functional analysis. Therefore, it makes uncertainty estimation essential to prevent potentially disastrous decisions based on segmentation results. This brings three key questions: How to represent the uncertainty in EM image segmentation? What methodologies can be employed to accurately estimate this uncertainty? How can this uncertainty estimation be harnessed to enhance the segmentation performance?

For the representation of the uncertainty, there are two

*Corresponding author

main types of uncertainty in DL method (Kendall and Gal 2017): epistemic (data) uncertainty arising from the inherent randomness or variability in the data itself, and aleatoric (model) uncertainty arising from the limitations and lack of knowledge in the model to learn the data. This paper, along with many previous works, focuses on aleatoric uncertainty.

For the estimation of aleatoric uncertainty, three mainstream methods are now available: *dropout-based* (Gal and Ghahramani 2016; Mobiny et al. 2021), *ensemble-based* (Lakshminarayanan, Pritzel, and Blundell 2017), and *evidential-based* (Sensoy, Kaplan, and Kandemir 2018) methods. Among them, evidential-based methods, relying on the Dempster-Shafer Evidence Theory (Dempster 1968), have shown more robust results with lower computational costs (Zou et al. 2022) compared to the other two methods. Notably, while evidential-based methods have been used in segmenting natural and medical images, they have not been explored for EM images. In this paper, we intend to use evidential-based methods to estimate the aleatoric uncertainty for EM image segmentation.

Then, for the utilization of the estimated uncertainty, previous studies have explored various approaches, such as generating pseudo labels for unlabeled data (Peng, Yi, and Yuan 2020) and incorporating uncertainty-based weights to fuse predictions from diverse sources (Basir and Yuan 2007). However, in EM segmentation tasks, limited attention has been given to the potential of directly rectifying errors using uncertainty information. During our mitochondria segmentation experiments, as an example illustrated in Figure 1, we observe that the areas of high uncertainty are prone to erroneous segmentation predictions. By using the uncertainty estimation (d) to rectify the original probabilistic prediction (c), the performance of the rectified prediction (e) demonstrated a notable enhancement, with a 7.5% improvement in Dice and a 14.3% enhancement in mAP. Building upon these observations, we advocate harnessing the estimated uncertainty as a strategy to effectively rectify segmentation errors.

In this study, we introduce a novel segmentation method, named Evidential Uncertainty-guided Mitochondria Segmentation for 3D EM Images (EUMS-3D), which is illustrated in Figure 2. EUMS-3D enables both uncertainty estimation and segmentation rectification by taking advantage of evidential deep learning (Sensoy, Kaplan, and Kandemir 2018). Specifically, EUMS-3D initially predicts the probabilities of semantics (inner part of the objects) and boundaries for the mitochondria by the backbone network. Then, an *Evidential Estimation Module* (EEM) is incorporated to model the uncertainty at the voxel level for all probability predictions. Subsequently, these predictions are rectified through the attention mechanism-based *Uncertainty Rectification Module* (URM), which integrates the uncertainty information from EEM and the multi-scale information from the designed *Feature Aggregation Module* (FAM). Our experimental results demonstrate the effectiveness of incorporating evidential uncertainty estimation to enhance 3D mitochondria segmentation, as EUMS-3D outperforms existing methods on four benchmark datasets: MitoEM-R (Wei et al. 2020), MitoEM-H (Wei et al. 2020), Kasthuri++(Casser et al. 2020), and Lucchi++(Casser et al. 2020). Ablation studies further confirm the contributions of each designed module in improving segmentation performance.

In summary, the contributions of this paper are as follows.

- To our best knowledge, this is the first evidential uncertainty-guided 3D mitochondria segmentation network for EM images.
- The uncertainty rectification module is proposed to enhance the segmentation performance by leveraging estimations of associated uncertainty and incorporating multi-scale features using the attention mechanism.
- Effectiveness of our method is verified by extensive experiments on four challenging benchmarks and on different backbone models.

Related Works

Mitochondria Segmentation. Recently, the field of 3D mitochondria segmentation has witnessed significant advancements. Numerous approaches have been proposed to tackle this challenging task. Traditional methods (Jorstad and Fua 2015; Lucchi, Li, and Fua 2013; Vazquez-Reina et al. 2011; Lucchi et al. 2012) often rely on manual or semi-automatic techniques, struggling with the complexity and variability of mitochondria structures in large-scale datasets. In response to these limitations, DL approaches have gained substantial attention. Convolutional neural networks and their variants have shown remarkable success in various image segmentation tasks. Recent studies have explored the adaptation and development of deep learning architectures for 3D mitochondria segmentation, including the use of U-Net (Ronneberger, Fischer, and Brox 2015), Mask R-CNN (Liu et al. 2018), and their 3D extensions. Additionally, advanced techniques such as attention mechanisms (Franco-Barranco, Muñoz-Barrutia, and Arganda-Carreras 2022), and generative adversarial networks (GANs) (Zhang et al. 2022) have also been investigated to enhance the accuracy and robustness of mitochondria segmentation.

Uncertainty estimation methods. Researchers have introduced a spectrum of uncertainty estimation techniques. These include Bayesian neural network (BNN) (Hinton and Van Camp 1993; MacKay 1992), ensemble-based (Lakshminarayanan, Pritzel, and Blundell 2017), dropout-based (Gal and Ghahramani 2016; Lakshminarayanan, Pritzel, and Blundell 2017; Kendall, Badrinarayanan, and Cipolla 2015), and evidential-based methods (Sensoy, Kaplan, and Kandemir 2018; Tsiligkaridis 2021; Tong, Xu, and Denoeux 2021). Classical BNNs model uncertainty by assimilating the weight distribution of the network, and approximate the integral of parameters using variational inference or Laplace approximation to gauge the posterior prediction distribution (Tsiligkaridis 2021). However, it is complex to train BNN due to the explicit representation of model parameters, which limits their scalability in terms of architecture and data size (Gawlikowski et al. 2021). In contrast, learning an ensemble of deterministic networks (Lakshminarayanan, Pritzel, and Blundell 2017; Mehrtash et al. 2020) and introducing Monte Carlo dropout (Gal and Ghahramani 2016) are more intuitive and simple, referred to as ensemble-based and dropout-based methods, respectively.

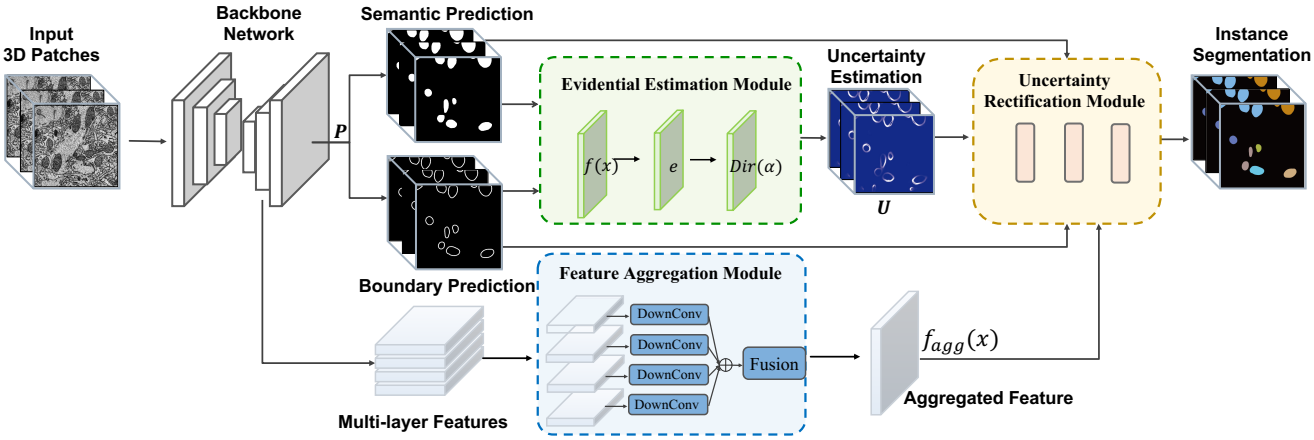

Figure 2: Overview of the proposed method. Initially, taking 3D EM image patches as input, an encoder-decoder network produces both a semantic mask and an instance boundary in parallel. Then, a feature aggregator fuse multi-layer features from the decoder. Subsequently, the evidential estimation module is employed to determine the uncertainty of each voxel for two inputs: semantic prediction, boundary prediction. Following this, the uncertainty rectification module integrates the two uncertainty maps and the aggregated multi-layer features to refine segmentations and obtain the final mitochondria instances.

Although been widely used, the ensemble-based methods require training multiple models, leading to high computational costs, and dropout-based methods may produce inconsistent outputs (Mobiny et al. 2021). Instead, the evidential-based method has shown more reliable performance in uncertainty estimation. Moreover, they demonstrate more robust results with lower computational costs compared to ensemble-based and dropout-based methods (Yager and Liu 2008). Although they have been utilized in natural and medical image segmentation tasks, their application in EM image segmentation, particularly for mitochondria segmentation, requires further investigation.

Method

Preliminary of Evidential Uncertainty Estimation

The evidential uncertainty estimation method (EDL) is a generalization of Bayesian theory to subjective probability. As Figure 2 illustrates, it assigns *belief masses* to each possible class label, and then the *belief distribution* of DST in the framework can be formalized as a *Dirichlet distribution* by Subjective Logic (SL) (Dempster 1968; Jøsang 2016).

Specifically, if we treat the segmentation as a voxel-wise K-class classification problem, for a voxel i, the EDL regards the classification task as giving a multinomial subjective opinion in a $K-$dimensional domain$\{1, \ldots, K\}$. The subjective opinion is expressed as a triplet $\omega = (\mathbf{b}, u, \mathbf{a})$, where $\mathbf{b} = \{b_1, \ldots, b_K\}$ is the *belief mass*, u represents the uncertainty, and $\mathbf{a} = \{a_1, \ldots, a_K\}$ is the base rate distribution. For any $k \in [1, 2, \ldots, K]$, the probability mass of a multinomial opinion is defined as $p_k = b_k + a_k u$. To enable the probability meaning of p_k, i.e., $\sum_k p_k = 1$, the base rate a_k is typically set to $1/K$ and the subjective opinion is constrained by $u + \sum_{k=1}^{K} b_k = 1$. For a K-class setting, the probability mass $\mathbf{p} = [p_1, p_2, \ldots, p_K]$ is assumed to follow a Dirichlet distribution parameterised by a K-dimensional Dirichlet strength vector $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_K\}$. The total strength of the Dirichlet is defined as $S = \sum_{k=1}^{K} \alpha_k$.

According to the evidence theory, the term evidence is introduced to describe the amount of supporting observations for classifying the voxel i into a class. Let $\mathbf{e} = \{e_1, \ldots, e_K\}$ be the evidence for K classes, $e_k = \alpha_k - 1$. In this way, the Dirichlet evidence can be mapped to the subjective opinion by setting the following: $b_k = \frac{e_k}{S}$, and $u = \frac{K}{S}$.

Therefore, we can see that if the evidence e_k for the k-th class is predicted, the corresponding expected class probability can be rewritten as $p_k = \alpha_k / S$, and the predictive uncertainty u can be determined after α_k is obtained.

Overview of the Architecture

Here we introduce the evidential uncertainty-based mitochondria segmentation in 3D EM Images (EUMS-3D) method. As illustrated in Figure 2, EUMS-3D consists of four modules: the *Backbone Network* for feature learning, the *Feature Aggregation Module* (FAM) for multi-layer feature combining, the *Evidential Estimation Module* (EEM) for the uncertainty estimation, and the *Uncertainty Rectification Module* (URM) for rectifying predictions. In the following sections, we describe the four modules, respectively.

Backbone Network

Recently, U-Net and its variants (Siddique et al. 2021) have shown remarkable performance in segmenting biomedical images. Building on this success, transformer-based architectures, originally popularized in natural language processing (NLP) tasks, have emerged as promising alternatives. In our work, we have modified the skip-connected encoder-decoder architecture known as 3D UX-Net (Lee et al. 2023) as our backbone network to predict the probability of semantic and boundary for each input 3D patch simultaneously. As a result, the output of the backbone is the concatena-

tion of the two probabilistic predictions, denoted as $\mathbf{P} \in \mathcal{R}^{K\times H\times W\times D}$, and p_{ik} is the k-th class prediction for voxel i. In this framework, $k \in \{1, ..., K\}, K = 2$. Compared to other SOTA models, 3D UX-Net, with lightweight volumetric ConvNet using hierarchical Transformers, demonstrates stable voxel segmentation performance across various challenging public datasets. It has proven its effectiveness in handling complex image segmentation tasks. It is essential to highlight that our framework is highly flexible, enabling designers to freely choose different backbones, such as Res-Unet3D (Li et al. 2022) and TransBTS (Lin et al. 2022), among others. This adaptability further enhances the versatility and applicability of our approach.

Feature Aggregation Module

Mitochondria in EM images are often with small size and ambiguous boundaries, necessitating details with higher resolution for enhanced differentiation. To effectively handle small objects while preserving the lightweight attribute of the system, we introduce the Feature Aggregation Module (FAM). FAM incorporates a multi-layer aggregation mechanism (Zheng et al. 2021), where the intermediate output of the decoder is concatenated together to produce a mask feature map $f_{agg(x)}$. To better leverage the information from the original resolution, after upsampling the mask feature map to the original resolution, we concatenate it with the original image and use another 3D convolution to fuse the information and generate the final mask.

Evidential Estimation Module

Based on the evidential uncertainty modeling method illustrated in Section 3.1, we facilitate the quantification of classification uncertainty by jointly modeling the probability maps $\mathbf{P}$ output by the backbone model. Figure 2 illustrates the process, where the output of the backbone network $f(x)$ undergoes an activation function layer (softplus) to ensure non-negative values and gain the evidence. Subsequently, the subjective logic offers a belief mass function, enabling the model to calculate the segmentation result uncertainties for different classes, resulting in the uncertainty estimation $\mathbf{U} \in \mathcal{R}^{H\times W\times D}$ as the output.

Uncertainty Rectification Module

To refine the predictions $\mathbf{P}$, the *Uncertainty Rectification Module* (URM) leverages both the uncertainty estimation $\mathbf{U}$ and the aggregated multi-layer features $f_{agg}(x)$ as depicted in Figure 3. Since predictions with high uncertainty are prone to erroneous results, URM first divides the uncertainty maps $\mathbf{U}$ and probabilistic prediction $\mathbf{P}$ into *certain* and *uncertain* by a threshold τ. Then, for the voxel i, two strategies are applied for the k-th class prediction p_{ik} to get the rectified prediction p_{ik}^{rec} by using the $u_i \in \mathbf{U}$, which is the uncertainty estimation of the voxel i.

a) *Certain* case ($u_i \leq \tau$): the uncertainty u_i is regarded simply as a weight: $p_{ik}^{rec} = p_{ik} \cdot (1 - u_i)$.

b) *Uncertain* case ($u_i > \tau$): when the prediction of the model is uncertain, mere reliance on the model's self-contained prediction may not suffice for effective rectification. To address this challenge, an approach involves incorporating contextual information from the surrounding voxels to aid in decision-making, thereby preventing over-correction. To achieve this, the segmentation prediction is subjected to a convolutional operation with a kernel size of $7 \times 7 \times 7$, resulting in $\mathbf{P}^{conv}$. p_{ik}^{conv} is the probability that the i-th voxel is classified into the k-th class output, which is then modulated by u_i as a weighting factor, leading to the refined prediction denoted as $p_{ik}^{rec} = p_{ik}^{conv} \cdot (1 - u_i)$.

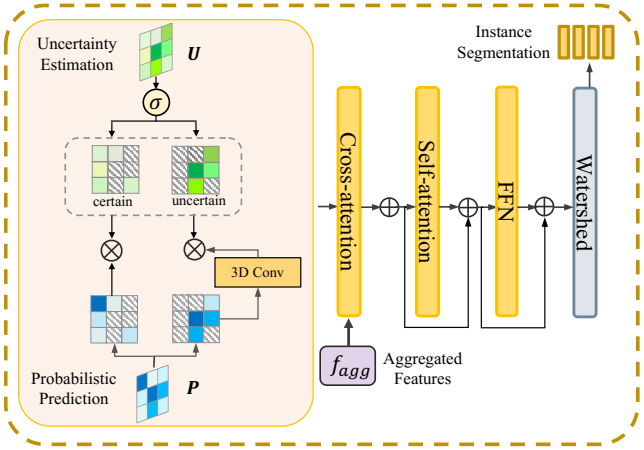

Figure 3: Illustration of the uncertainty rectification module.

Furthermore, to mitigate potential overly arbitrary modifications caused by uncertainty estimates, we introduce a mask decoder that incorporates multi-scale information, adapted from (Cheng et al. 2022). As illustrated in Figure 3, the attention mechanism is applied to leverage the aggregated multi-layer features obtained from FAM. This process leads to the prediction of new uncertainty-aware masks, denoted as $\mathbf{P}^{final}$. Subsequently, we utilize the watershed (Shafarenko, Petrou, and Kittler 1997) algorithm to generate instance results from the uncertainty-aware masks.

Loss Functions

Three loss functions are utilized in voxel-wise manner. First, we use the EDL loss function modified from cross-entropy loss proposed in (Sensoy, Kaplan, and Kandemir 2018; Zou et al. 2022) for the outputs of EEM. For voxel i, y_{ik} and p_{ik} are the label and predicted probability for class k. $\psi(\cdot)$ denotes the *digamma* function. S_i is the total strength of a Dirichlet distribution parameterized by α_k.

$$\mathcal{L}_{EDL}^{i} = \sum_{k=1}^{K} y_{ik}(\psi(S_i) - \psi(\alpha_{ik})). \tag{1}$$

Second, the KL divergence loss function is introduced, where $\Gamma(\cdot)$ is the *gamma* function. $\widetilde{\alpha}_{ik} = y_{ik} + (1 - y_{ik}) \odot \alpha_{ik}$ denotes the adjusted parameters of the Dirichlet distribution, which aims to ensure that ground truth class evidence is not mistaken for 0.

$$\begin{aligned}\mathcal{L}_{KL}^{i} = \quad & \log\left(\frac{\Gamma(\sum_{k=1}^{K} \widetilde{\alpha}_{ik})}{\Gamma(K)\sum_{k=1}^{K}\Gamma(\widetilde{\alpha}_{ik})}\right) \\ & + \sum_{k=1}^{K}(\widetilde{\alpha}_{ik} - 1)[\psi(\widetilde{\alpha}_{ik}) - \psi(\textstyle\sum_{k=1}^{K} \widetilde{\alpha}_{ik})].\end{aligned} \tag{2}$$

Methods	Dataset / Metric	MitoEM-H		MitoEM-R		Kasthuri++		Lucchi++	
		Dice	mAP	Dice	mAP	Dice	mAP	Dice	mAP
ConvNets -based	Xiao (Xiao et al. 2018)	0.798	0.812	0.830	0.903	0.947	0.900	0.882	0.910
	Peng and Yuan (Peng and Yuan 2019)	0.757	0.793	0.802	0.848	0.909	0.833	0.893	0.806
	U3D-BC (Wei et al. 2020)	0.746	0.773	0.775	0.844	0.889	0.831	0.880	0.753
	Zhili (Li et al. 2021)	0.765	0.787	0.794	0.870	0.935	0.909	0.865	0.811
	HIVE-Net (Yuan et al. 2021)	0.825	0.856	0.851	0.901	0.862	0.904	0.893	0.841
	nnU-Net (Isensee et al. 2021)	0.807	0.830	0.825	0.864	0.859	0.872	0.856	0.829
	Res-Unet3D (Li et al. 2022)	0.783	0.828	0.815	0.917	0.943	0.892	0.889	0.769
Transformer -based	nnFormer (Zhou et al. 2021)	0.787	0.830	0.824	0.813	0.919	0.874	0.889	0.825
	SwinUNETR (Hatamizadeh et al. 2022a)	0.779	0.822	0.803	0.867	0.904	0.861	0.869	0.874
	DSTUnet (Lin et al. 2022)	0.762	0.799	0.836	0.824	0.916	0.867	0.896	0.802
	TransBTS (Wang et al. 2021)	0.801	0.827	0.866	0.909	0.931	0.899	0.902	0.896
	UNETR (Hatamizadeh et al. 2022b)	0.775	0.809	0.813	0.836	0.855	0.858	0.872	0.801
	3D UX-Net (Lee et al. 2023)	0.816	0.845	0.859	0.901	0.950	0.917	0.902	0.910
EUMS-based	EUMS-3D (HIVE-Net)	<u>0.837</u>	<u>0.871</u>	0.865	0.911	0.895	0.909	0.907	0.869
	EUMS-3D (Res-Unet3D)	0.818	0.848	0.839	<u>0.923</u>	<u>0.951</u>	<u>0.915</u>	0.900	0.806
	EUMS-3D (SwinUNETR)	0.803	0.841	0.824	0.882	0.930	0.879	0.918	<u>0.924</u>
	EUMS-3D (TransBTS)	0.823	0.870	<u>0.878</u>	0.917	0.945	0.914	<u>0.919</u>	0.913
	EUMS-3D (3D UX-Net)	**0.845**	**0.901**	**0.890**	**0.928**	**0.972**	**0.931**	**0.937**	**0.953**

Table 1: Quantitative results of methods on MitoEM-H, MitoEM-R, Kasthuri++ and Lucchi++ datasets. The methods are grouped into three types: ConvNets, Transformer, and EUMS based. For EUMS-based methods, the bracket shows the backbone model. The bold values indicate the best performance and the underlined values indicate the second best.

Third, we use a soft Dice loss (Milletari, Navab, and Ahmadi 2016) $\mathcal{L}_{Dice}$ to optimize the network, where $p_{ik}^{final} \in \mathbf{P}^{final}$ is the uncertainty-aware prediction in URM.

$$\mathcal{L}_{Dice}^{i} = \sum_{k=1}^{K}(1 - \frac{2y_{ik}p_{ik}^{final}}{y_{ik} + p_{ik}^{final}}). \tag{3}$$

The overall loss function can be defined as follows: $L_{total} = L_{Dice} + \lambda_1 L_{EDL} + \lambda_2 L_{KL}$. λ_1 and λ_2 are balance factors.

Experiments

Experimental Settings

We compare our proposed method with several SOTA segmentation methods. These include the current ConvNets-based and Transformer-based methods on image segmentation in volumetric settings. For ConvNets-based methods, we choose Res-Unet3D (Li et al. 2022), HIVE-Net (Yuan et al. 2021), Zhili (Li et al. 2021), Peng and Yuan (Peng and Yuan 2019), Xiao (Xiao et al. 2018), Res-Unet3D (Li et al. 2022), HIVE-Net (Yuan et al. 2021), and nnU-Net (Isensee et al. 2021). For Transformer-based methods, we choose DSTUnet (Lin et al. 2022), TransBTS (Wang et al. 2021), UNETR (Hatamizadeh et al. 2022b), nnFormer (Zhou et al. 2021), SwinUNETR (Hatamizadeh et al. 2022a), and 3D UX-Net (Lee et al. 2023). The λ_1 and λ_2 in the loss function are set to be 1 and 0.5 following (Zou et al. 2022).

Datasets. We evaluate our method on four datasets: **MitoEM-R** (Wei et al. 2020), **MitoEM-H** (Wei et al. 2020), **Kasthuri++** dataset (Casser et al. 2020) and **Lucchi++** (Casser et al. 2020). MitoEM is a dense mitochondria instance segmentation dataset from ISBI 2021 challenge, including two subsets collected from an adult human and an adult rat with volumes (30 μm^3) of resolution of 8×8×30 nm. Each volume has 500 annotated grayscale images of resolution (4096×4096), out of which 400 for training and 100 for testing. Kasthuri++ has 85 image slices of size 1643×1613 for training and 75 slices of size 1334×1553 images for testing. Lucchi++ is a sparse mitochondria semantic segmentation dataset with the training and testing volume size of $165 \times 1024 \times 768$.

Experimental Results

We evaluate the methods following the ISBI 2021 challenge (Wei et al. 2020), including mean 3D Average Precision (mAP) and Dice scores at the instance level. Based on the quantitative results in Table 1 and the visualization results shown in Figure 4, our proposed EUMS-3D algorithm, utilizing 3D UX-Net as the backbone, demonstrates superior performance in all segmentation tasks and maintains a moderate parameter count compared to other models.

Evaluation on MitoEM. The quantitative results highlight EUMS-3D(3D UX-Net) superior performance in precise mitochondria segmentation. Across both datasets, it shows the highest scores for both Dice and mAP, showcasing its excellent performance in accurately segmenting mitochondria. Specifically, it achieved approximately 84.5%/90.1% (Dice/mAP) on the human dataset and 89.0%/92.8% on the rat dataset, respectively. Notably, the comparison between Transformer-based and ConvNets-based methods indicates no significant difference in effect, while our EUMS-3D achieves SOTA performance. Several segmentation examples are presented in Figure 4, highlighting our method's proficiency in capturing the morphology of mitochondria.

Evaluation on Kasthuri++ and Lucchi++. Considering the relatively smaller size of the two datasets, the segmentation boundary plays a critical role in influencing the scores. Remarkably, our method EUMS-3D (3D UX-Net) achieves a performance of 97.2% in Dice coefficient and 93.1% in

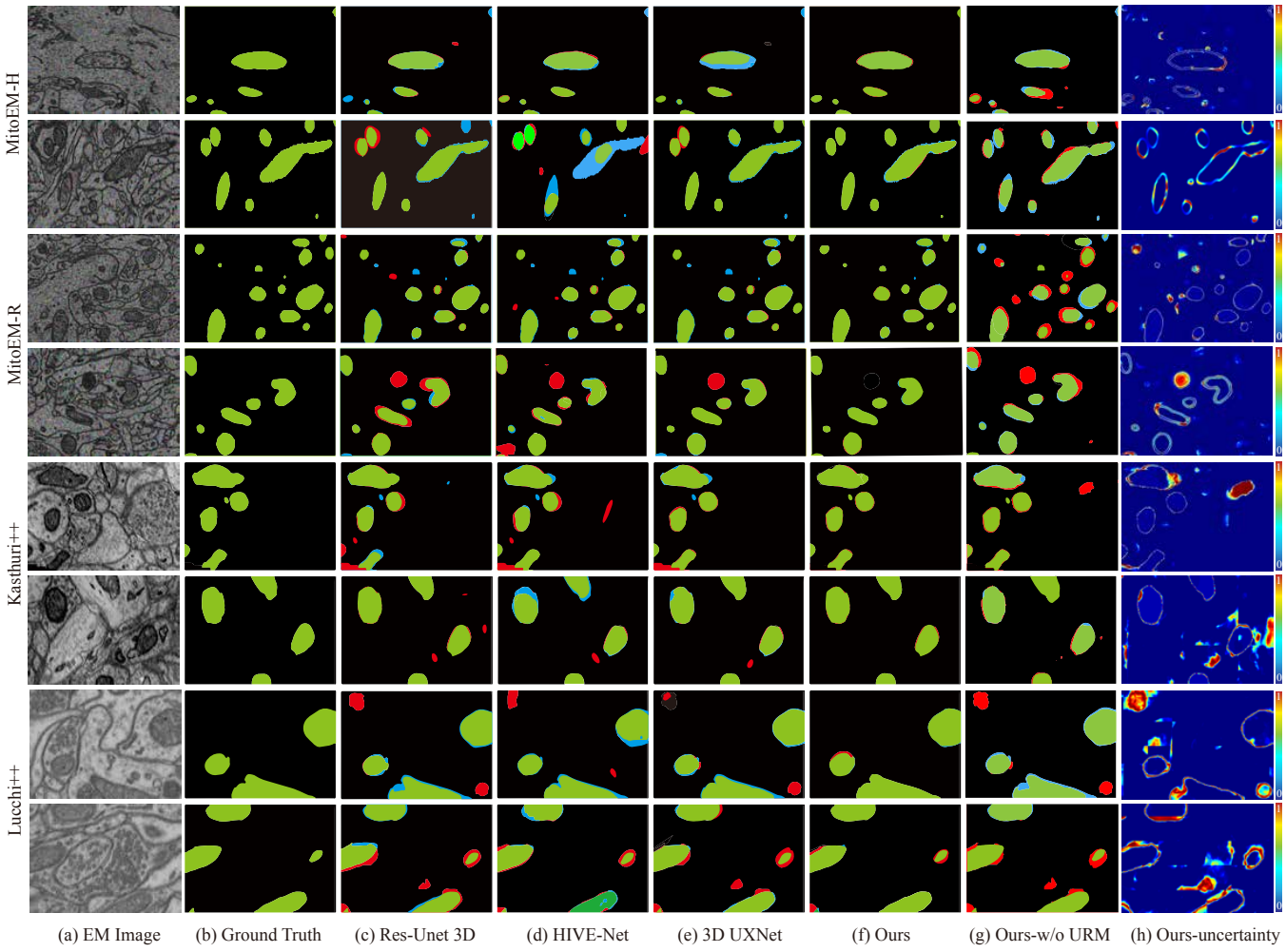

Figure 4: The visualization of segmentation results on MitoEM-H, MitoEM-R, Kasthuri++, and Lucchi++ datasets. (a,b) The EM image with the ground truth. (c-e) The predictions of the three compared methods. (f) The prediction of our EUMS-3D (3D UXNet). (g) The prediction of the ablation for EUMS-3D (3D UXNet) by removing URM module. (h) The uncertainty map u within EUMS-3D (3D UXNet). For all the segmentation figures, true positive samples are highlighted in green, false positive in red, and false negative in blue.

mAP on the Kasthuri++ dataset. Similarly, on the Lucchi++ dataset, the method achieves a Dice score of 93.7% and an mAP score of 95.3%, effectively closing the gap toward human-level benchmarks. The visual illustrations of segmentation instances, elucidated in Figure 4, reinforce and substantiate the prowess of our approach, adeptly and accurately refining the mitochondrial boundaries.

Exploring Uncertainty Estimation Methods

To further investigate the effectiveness of different uncertainty estimation methods, we compare the evidential-based method with two other methods, namely *dropout* (Mukhoti et al. 2021) and *ensemble* (Lakshminarayanan, Pritzel, and Blundell 2017) by replacing the uncertainty estimation module. All reported results are based on the 3D UXNet backbone model, and we use the two MitoEM datasets for evaluation. The Dice and mAP scores of the three methods are shown in Table 2. It indicates that the evidential uncertainty estimation method outperforms the other methods with at least 8% improvement in mAP for the two MitoEM datasets.

Dataset / Metric	MitoEM-H		MitoEM-R	
	Dice	mAP	Dice	mAP
dropout	0.835	0.873	0.867	0.895
ensemble	0.839	0.880	0.868	0.902
evidential	**0.845**	**0.901**	**0.890**	**0.928**

Table 2: Performance of different uncertainty estimation modules on MitoEM-R and MitoEM-H datasets.

Besides, we compare the calibration performance across the three methods. As the estimation of uncertainty plays a key role in refining segmentations, it becomes imperative to possess a well-calibrated model. Such a model should

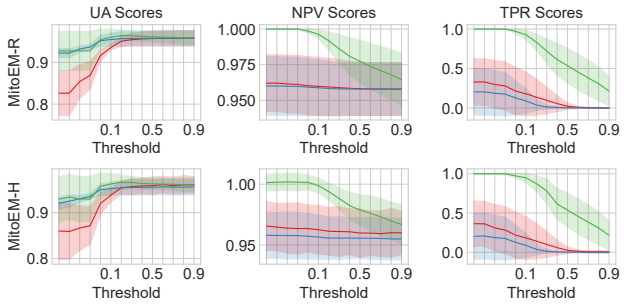

Figure 5: The calibration scores of different uncertainty estimation methods. Red/blue/green: ensemble/dropout/evidential. Solid line/color region: mean/std.

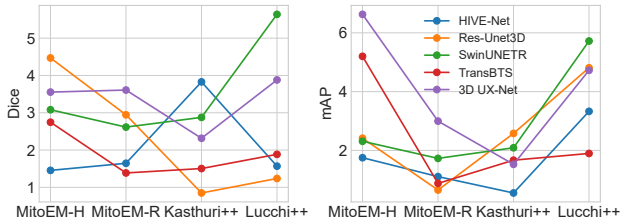

Figure 6: Ablation study for various backbone networks. The scores show the improvements achieved by our approach when applied to different backbones across two MitoEM datasets.

be confident in its predictions when being accurate, and be uncertain about inaccurate ones (Houlsby et al. 2011). The evaluation metrics are proposed in (Mobiny et al. 2021), including *negative predictive value* (NPV), *true positive rate* (TPR), and *uncertainty accuracy* (UA). Higher values indicate better calibration of the model, which was dependent on the uncertainty threshold. As shown in Figure 5, the evidential-based method has the highest scores across three metrics. This validates that the evidential-based method could be more effective to rectify the erroneous predictions.

Ablation Study

Backbone Models. We compare various backbone networks to investigate their influence on the efficacy of our methodology. The results demonstrate the improvements achieved by our approach across two MitoEM datasets when applied to different backbones, including two ConvNets-based (Res-Unet3D, HIVE-Net) and three Transformer-based (TransBTS, SwinUNETR, 3D UXNet). The results in Table 1 and Figure 6 indicate the performance augmentation exhibited by all five original backbone models. Notably, the performance of SwinUNETR attained a commendable 5.6% increase in Dice score on the Lucchi++ dataset. Additionally, the 3D UX-Net exhibited a noteworthy 6.6% improvement in Dice score on the MitoEM-H dataset. It should be highlighted that our proposed module is lightweight to gain performance enhancement by adding only 2.2M parameters.

URM and FAM. To investigate the effectiveness of the two modules, we conducted an ablation study by training the network without the two modules, denoted as w/o URM and w/o FAM. The quantitative results presented in Table 3 demonstrate that leveraging the URM with FAM can improve performance by effectively utilizing multi-scale information and uncertainty estimation to correct segmentation errors. Specifically, it shows an increment of 2.2% in mAP score across both datasets with the URM, and 1.4% in mAP with FAM. Several visualization results are shown in Figure 4, the uncertainty map has high sensitivity in contour areas. Particularly for regions with morphological similarities to mitochondrial structures that are not annotated in the ground truth, the model may make mistakes in these zones without URM. By using URM, the mistakes are effectively rectified. We also use dilated convolutions (dilation rate of 2) to replace 3D convolutions. It shows a slight decrease by using Dilated Conv compared to original 3D Conv. This marginal decrease may suggest that the broader receptive field potentially introduces additional noise.

Dataset & **Metric**	**MitoEM-H**		**MitoEM-R**	
	Dice	mAP	Dice	mAP
w/o URM	0.827	0.869	0.871	0.906
w/o FAM	0.840	0.884	0.883	0.914
w Dilated Conv	0.843	0.900	0.884	0.925
EUMS-3D	**0.845**	**0.901**	**0.890**	**0.928**

Table 3: Ablation study of the modules of URM and FAM on MitoEM-R and MitoEM-H dataset.

Threshold τ	**0**	**0.3**	**0.5**	**0.7**	**1**
Dice	0.880	0.884	**0.890**	0.887	0.887
mAP	0.918	0.923	**0.928**	0.924	0.923

Table 4: Ablation study for different threshold in URM on MitoEM-R dataset.

Parameter in URM. In URM, we use a threshold τ to divide the uncertainty estimation and followed by the *certain* and *uncertain* strategies. To explore the influence of τ, we train the model using various threshold values $\tau = 0, 0.25, 0.5, 0.75, 1$ on MitoEM-R dataset, respectively. Notably, when $\tau = 0$, URM treats all predictions as *uncertain* case, whereas $\tau = 1$ corresponds to *certain* case. The results in Table 4 demonstrate that the best performance is achieved when τ is set to 0.5, indicating that dividing the predictions by uncertainty can help improve segmentation performance.

Conclusion

This research, for the first time, presents a novel 3D instance segmentation method for trustworthy segmentation of mitochondrion on EM images. To our best knowledge, for the first time, we employ evidential-based uncertainty estimation and neighborhood information to modify segmentation outcomes and generate reliable fusion. We conduct comprehensive experiments on three benchmark datasets to validate the efficacy of our approach in improving segmentation results and uncertainty estimation.

Acknowledgments

This work was partially supported by the Natural Science Foundation of China under contract 62088102. We also acknowledge High-Performance Computing Platform of Peking University for providing computational resources.

References

Annesley, S. J.; and Fisher, P. R. 2019. Mitochondria in health and disease. *Cells*, 8(7): 680.

Basir, O.; and Yuan, X. 2007. Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory. *Information Fusion*, 8(4): 379–386.

Bock, F. J.; and Tait, S. W. 2020. Mitochondria as multifaceted regulators of cell death. *Nature Reviews Molecular Cell Biology*, 21(2): 85–100.

Casser, V.; Kang, K.; Pfister, H.; and Haehn, D. 2020. Fast mitochondria detection for connectomics. In *Proceedings of the Third Conference on Medical Imaging with Deep Learning*, volume 121, 111–120.

Cheng, B.; Misra, I.; Schwing, A. G.; Kirillov, A.; and Girdhar, R. 2022. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1290–1299.

Dempster, A. P. 1968. A generalization of Bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2): 205–232.

Franco-Barranco, D.; Muñoz-Barrutia, A.; and Arganda-Carreras, I. 2022. Stable deep neural network architectures for mitochondria segmentation on electron microscopy volumes. *Neuroinformatics*, 20(2): 437–450.

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, 1050–1059.

Gawlikowski, J.; Tassi, C. R. N.; Ali, M.; Lee, J.; Humt, M.; Feng, J.; Kruspe, A.; Triebel, R.; Jung, P.; Roscher, R.; et al. 2021. A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342*.

Guo, C.; Pleiss, G.; Sun, Y.; and Weinberger, K. Q. 2017. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, 1321–1330.

Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H. R.; and Xu, D. 2022a. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 272–284.

Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; Landman, B.; Roth, H. R.; and Xu, D. 2022b. UNETR: Transformers for 3D medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 574–584.

Hinton, G. E.; and Van Camp, D. 1993. Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the 6th Annual Conference on Computational Learning Theory*, 5–13.

Houlsby, N.; Huszár, F.; Ghahramani, Z.; and Lengyel, M. 2011. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*.

Isensee, F.; Jaeger, P. F.; Kohl, S. A.; Petersen, J.; and Maier-Hein, K. H. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2): 203–211.

Jorstad, A.; and Fua, P. 2015. Refining mitochondria segmentation in electron microscopy imagery with active surfaces. In *European Conference on Computer Vision Workshops*, 367–379.

Jøsang, A. 2016. *Subjective logic*, volume 4. Springer.

Kendall, A.; Badrinarayanan, V.; and Cipolla, R. 2015. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv:1511.02680*.

Kendall, A.; and Gal, Y. 2017. What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, volume 30.

Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30.

Lee, H. H.; Bao, S.; Huo, Y.; and Landman, B. A. 2023. 3D UX-Net: A large kernel volumetric ConvNet modernizing hierarchical transformer for medical image segmentation. In *the Eleventh International Conference on Learning Representations*.

Li, M.; Chen, C.; Liu, X.; Huang, W.; Zhang, Y.; and Xiong, Z. 2022. Advanced deep networks for 3D mitochondria instance segmentation. In *IEEE 19th International Symposium on Biomedical Imaging*, 1–5.

Li, Z.; Chen, X.; Zhao, J.; and Xiong, Z. 2021. Contrastive learning for mitochondria segmentation. In *Annual International Conference of the IEEE Engineering in Medicine Biology Society*, 3496–3500.

Lin, J.; Ge, H.; Wan, Y.; and Zhang, R. 2022. Depth Swin transformer Unet for serial section biomedical image segmentation. In *International Conference on Signal and Image Processing*, 452–457.

Liu, J.; Li, W.; Xiao, C.; Hong, B.; Xie, Q.; and Han, H. 2018. Automatic detection and segmentation of mitochondria from SEM images using deep neural network. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 628–631.

Lucchi, A.; Li, Y.; and Fua, P. 2013. Learning for structured prediction using approximate subgradient descent with working sets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1987–1994.

Lucchi, A.; Li, Y.; Smith, K.; and Fua, P. 2012. Structured image segmentation using kernelized features. In *European Conference on Computer Vision*, 400–413.

Lucchi, A.; Smith, K.; Achanta, R.; Knott, G.; and Fua, P. 2011. Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features. *IEEE Transactions on Medical Imaging*, 31(2): 474–486.

Luo, Z.; Wang, Y.; Liu, S.; and Peng, J. 2021. Hierarchical encoder-decoder with soft label-decomposition for mitochondria segmentation in EM images. *Frontiers in Neuroscience*, 15: 687832.

MacKay, D. J. 1992. A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4(3): 448–472.

Mehrtash, A.; Wells, W. M.; Tempany, C. M.; Abolmaesumi, P.; and Kapur, T. 2020. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE Transactions on Medical Imaging*, 39(12): 3868–3878.

Mekuč, M. Ž.; Bohak, C.; Hudoklin, S.; Kim, B. H.; Kim, M. Y.; and Marolt, M. 2020. Automatic segmentation of mitochondria and endolysosomes in volumetric electron microscopy data. *Computers in Biology and Medicine*, 119: 103693.

Milletari, F.; Navab, N.; and Ahmadi, S.-A. 2016. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *International Conference on 3D Vision*, 565–571.

Mobiny, A.; Yuan, P.; Moulik, S. K.; Garg, N.; Wu, C. C.; and Van Nguyen, H. 2021. Dropconnect is effective in modeling uncertainty of Bayesian deep networks. *Scientific Reports*, 11(1): 1–14.

Mukhoti, J.; van Amersfoort, J.; Torr, P. H.; and Gal, Y. 2021. Deep deterministic uncertainty for semantic segmentation. *arXiv preprint arXiv:2111.00079*.

Peng, J.; Yi, J.; and Yuan, Z. 2020. Unsupervised mitochondria segmentation in EM images via domain adaptive multitask learning. *IEEE Journal of Selected Topics in Signal Processing*, 14(6): 1199–1209.

Peng, J.; and Yuan, Z. 2019. Mitochondria segmentation from EM images via hierarchical structured contextual forest. *IEEE Journal of Biomedical and Health Informatics*, 24(8): 2251–2259.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems*, volume 31.

Shafarenko, L.; Petrou, M.; and Kittler, J. 1997. Automatic watershed segmentation of randomly textured color images. *IEEE Transactions on Image Processing*, 6(11): 1530–1544.

Siddique, N.; Paheding, S.; Elkin, C. P.; and Devabhaktuni, V. 2021. U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9: 82031–82057.

Tong, Z.; Xu, P.; and Denoeux, T. 2021. Evidential fully convolutional network for semantic segmentation. *Applied Intelligence*, 51: 6376–6399.

Tsiligkaridis, T. 2021. Information Aware max-norm Dirichlet networks for predictive uncertainty estimation. *Neural Networks*, 135: 105–114.

Vazquez-Reina, A.; Gelbart, M.; Huang, D.; Lichtman, J.; Miller, E.; and Pfister, H. 2011. Segmentation fusion for connectomics. In *International Conference on Computer Vision*, 177–184.

Wang, W.; Chen, C.; Ding, M.; Yu, H.; Zha, S.; and Li, J. 2021. Transbts: Multimodal brain tumor segmentation using transformer. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 109–119.

Wei, D.; Lin, Z.; Franco-Barranco, D.; Wendt, N.; Liu, X.; Yin, W.; Huang, X.; Gupta, A.; Jang, W.-D.; Wang, X.; et al. 2020. MitoEM dataset: Large-scale 3D mitochondria instance segmentation from EM images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 66–76.

Xiao, C.; Chen, X.; Li, W.; Li, L.; Wang, L.; Xie, Q.; and Han, H. 2018. Automatic mitochondria segmentation for EM data using a 3D supervised convolutional network. *Frontiers in Neuroanatomy*, 12: 92.

Yager, R. R.; and Liu, L. 2008. *Classic Works of the Dempster-Shafer Theory of Belief Functions*, volume 219.

Yuan, Z.; Ma, X.; Yi, J.; Luo, Z.; and Peng, J. 2021. HIVE-Net: Centerline-aware hierarchical view-ensemble convolutional network for mitochondria segmentation in EM images. *Computer Methods and Programs in Biomedicine*, 200: 105925.

Yuan, Z.; Yi, J.; Luo, Z.; Jia, Z.; and Peng, J. 2020. EM-Net: Centerline-aware mitochondria segmentation in EM images via hierarchical view-ensemble convolutional network. In *IEEE International Symposium on Biomedical Imaging*, 1219–1222.

Zhang, Y.; Liu, J.; Li, Z.; Guo, J.; and Han, H. 2022. CFDA-M: Coarse-to-fine domain adaptation for mitochondria segmentation via patch-wise image alignment and online self-training. In *IEEE International Conference on Bioinformatics and Biomedicine*, 1810–1815.

Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P. H.; et al. 2021. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6881–6890.

Zhou, H.-Y.; Guo, J.; Zhang, Y.; Yu, L.; Wang, L.; and Yu, Y. 2021. nnFormer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*.

Zou, K.; Yuan, X.; Shen, X.; Wang, M.; and Fu, H. 2022. TBraTS: Trusted brain tumor segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 503–513.