

文章编号 1004-924X(2024)18-2836-10

基于自然图像模型微调的小鼠脑部电镜图像 实例分割

承 骜¹, 赵国强², 张若冰², 王丽荣^{1*}

(1. 苏州大学 电子信息学院, 江苏 苏州 215000;

2. 中国科学院 苏州生物医学工程研究所, 江苏 苏州 215000)

摘要: 分割模型的准确性和鲁棒性是小鼠脑电镜图像处理中的主要考虑因素。针对电镜图像的技术特点, 提出了高度稳健的二维分割算法, 准确识别每张切片中各物体的形态结构。本文提出了基于大型自然图像模型的主干网络微调的大尺度二维电镜图像分割模型 EM-SAM, 用于脑部电镜图像中的实例分割。模型主干网络采用大型自然图像模型 SAM 中的已训练完成的图像编码器, 在电镜图像处理任务中最大化模型提取图像特征的能力。此外, 模型采用了 U 型的解码器设计, 并通过小鼠脑电镜图像分割任务进行微调。实验结果表明: 在公开数据集 SNEMI3D 中 A-Rand 可达到 0.054; 在公开数据集 MitoEM 中 AP-50 和 AP-75 分别可达到 0.883, 0.604。EM-SAM 在电镜图像神经分割任务中准确性高、鲁棒性强, 并且可针对不同任务进行微调。

关键词: 深度学习; 分割; 大模型; 电镜图像; 小鼠脑部

中图分类号: TP391 **文献标识码:** A **doi:** 10.37188/OPE.20243218.2836

Instance segmentation of mouse brain scanning electron microscopy images based on fine-tuning nature image model

CHENG Ao¹, ZHAO Guoqiang², ZHANG Ruobing², WANG Lirong^{1*}

(1. Soochow University, School of Electronic and Information, Suzhou 215000, China;

2. Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences,
Suzhou 215000, China)

* Corresponding author, E-mail: wanglirong@suda.edu.cn

Abstract: The accuracy and robustness of segmentation models are critical considerations in the processing of mouse brain electron microscopy images. We proposed a highly robust two-dimensional segmentation algorithm tailored to the technical characteristics of electron microscopy images, aiming to accurately delineate the morphological structure of cells in each slice. Aiming at accurately delineate the morphological structure of cells in each section, a highly robust two-dimensional segmentation algorithm based on natural image model tailored to the technical characteristics of electron microscopy images was proposed. EM-SAM was based on fine-tuning the backbone network of pre-trained large natural image model SAM for maximizing the capability of features extraction. The model employed the image encoder from the SAM

收稿日期: 2024-06-15; 修订日期: 2024-07-23.

基金项目: 国家自然科学基金资助项目 (No. 32271430); 国家重点研发计划资助项目 (No. 2023YFF0715904)

architecture, augmented with a U-shaped decoder, and was fine-tuned specifically for the segmentation of mouse brain electron microscopy images. Experimental results demonstrate that A-Rand achieves 0.054 on public dataset SNEMI3D. Additionally, AP-50 and AP-75 reach 0.883 and 0.604, respectively, on public dataset MitoEM. EM-SAM exhibits high accuracy and robustness in neural segmentation tasks of electron microscopy images, and it can be fine-tuned for different tasks.

Key words: computer vision; segmentation; large-scale models; electron microscopy images; mouse brain

1 引言

大脑神经组织是人体绝大多数生理行为与心理活动的主导,它是由近千亿个、包含了数百种不同的神经元细胞所构成的极为复杂的生物组织,是人类意识与思维形成的物质基础。探索脑的结构、理解脑的功能、阐明脑的工作原理是当今最具挑战性和前瞻性的科学问题。其中,解析哺乳动物脑组织的细胞层级微观结构在揭示神经作用、理解神经性疾病发病机制、发展人工智能等领域起到关键作用。

小鼠作为生物医学研究的主要模型,其脑部超微结构的研究具有重要意义。这些小型哺乳动物拥有类似于人类的神经系统结构,且其生殖周期短,易于实验控制,使它们成为研究神经科学的理想模型。

通过高效的分割算法,研究人员可以重建出如神经元、血管、细胞器等形态特征及分布。鉴于连续断层扫描电子显微镜成像技术的发展,如多电子束扫描电子显微镜的出现,图像采集的自动化和速度加快,导致如此大规模的连续断层扫描电子显微镜数据集的手动分割和标注非常耗时耗力。

2 相关工作

有关图像分割的工作一直是计算机视觉的重要分支之一,随着当前计算机硬件技术的进步与相关分割算法的发展,目前分割方法可以大体上可以分为两类:传统的图像分割方法和基于深度学习的图像分割方法。

2.1 传统的图像分割方法

传统的图像分割方法主要依赖于图像的底层特征,如灰度、颜色、纹理和边缘等。通过数学

和统计方法对这些特征进行分析和处理来达到分割的目的。

Frangakis 等人^[1]提出了一种利用像素之间的亲和力的标准(例如邻近性和灰度相似性),结合特征向量分析,将图像细分为对应于对象或有意义的区域。之后,Kylberg 等人^[2]提出了通过像素邻域的统计分析进行分割的方法,该方法基于对图像中所有像素的局部邻域的统计分析,然后进行对象宽度判别,并在最后对于细长对象进行边界细化操作。Liu 等人^[3]提出了一种利用分层结构和边界分类进行二维神经元分割的方法,即利用膜检测概率图构建分水岭合并树,用于表示分水岭算法中的分层区域合并,并使用非局部图像特征学习边界分类器来预测树中的每个潜在合并项,在此基础上,在一致性约束下做出合并决策以获取最终分割结果。Liu 等人^[4]进一步改进了该算法,提出了一种对神经元进行内部分割和外部分割的方法,构建分层合并树结构来表示多区域假设,并使用监督分类技术进行评估,使用具有一致性约束的合并树获得最终的神经元内部分割结果。此外,Karabag 等人^[5]提出了基于形态建模的方法,通过分析大小、位置以及与相邻切片中检测到的区域的对应关系来选择目标区域对应的超像素。此方法拥有较快的处理速度,能够高效处理大量数据。

总的来说,传统图像分割方法通常基于明确的数学原理和算法,易于理解和实现。并且分割方法的计算复杂度较低,能够在较短时间内处理图像,适用于实时或近实时应用。但多数传统方法需要根据具体图像特征和应用场景进行参数调整,例如,阈值分割需要设定阈值,边缘检测需要选择适当的算子和参数等。并且传统的分割方法对噪声较为敏感,高噪声的图像的分割效果通常不理想。

2.2 基于深度学习的图像分割方法

前文提到,传统的图像分割方法仍然存在其局限性,如依赖于预设的特征和参数,需要手工设计和选择特征,在不同应用场景和数据集之间迁移能力有限,通常需要针对具体任务进行重新调整和优化,对于高噪声的图片鲁棒性差等。深度学习技术作为机器学习的一个分支,旨在通过多层神经网络自动学习和提取数据中的复杂特征。在图像分割的领域主要是依赖卷积神经网络及其变体,如 U-Net 等,这些方法利用深度学习强大的特征提取能力和端到端学习的优势,使得图像分割在各个领域(如医学影像、自动驾驶、遥感影像等)取得了显著的进展和应用效果。

近年来,深度学习网络也被广泛应用在电镜图像分割任务中。Fakhry 等人^[6]提出了包含残差结构的二维全卷积网络分割电镜图像,通过简单高效的模型结构提供了密集预测挑战的解决方法,保留全分辨率预测并包含足够的上下文信息,并将此方法应用于电镜图像中三维神经突分割的开放挑战中,取得了较优成绩,证明了二维方法在连续电镜图像中的有效性。Xiao 等人^[7]提出了一种有效的深度学习方法来实现电镜图像堆栈中的神经元膜分割,该方法利用空间高效的残差网络和上下文线索的多级表示来实现准确的分割性能。Jiang 等人^[8]提出了一种有效的编码器解码器网络,通过改进的残差网络和金字塔模块提取神经细胞体和细胞核的分割特征,通过多层特征融合,合并多尺度上下文信息并生成有效的分割结果。Casser 等人^[9]提出了一种基于改进的 U-Net 架构的全自动线粒体检测器,实现了轻量化二维 U-Net 的高效电镜图像分割重建,可提供高精度和快速处理时间。Cao 等人^[10]受 DenseNet 的启发,提出了 DenseUNet,此网络在 U 型编码器解码器中加入了密集连接设计,缓解梯度消失问题,减少了参数数量,并使参数使用更加高效。Quan 等人^[11]将深度学习的最新进展(例如语义分割和残差神经网络),使用基于求和的跳跃连接与二维 U-Net 相结合,提出了 Fusion-Net,允许更深层次的网络架构,以实现更准确的神经元分割。Luo 等人^[12]采用了带有微型 U-Net 的两级网络架构,并且提出了一种新颖的软标签分解策略来利用手动标签中的形状知识,通过根

据线粒体的椭圆度将每个手动标签图分解为两个互补的标签图,并引入额外的子类别感知监督。Khadangi 等人^[13]提出了不同网络的集合 EM-stellar,并建立了电镜图像的大规模基准测试结果。Spiers 等人^[14]提出了三轴预测方法,解决二维 U-Net 无法处理三维空间数据的限制。Yuan 等人^[15]提出了引入了一种新颖的分层视图集成卷积(High-dimensional Convolution Embedded Convolution, HVEC),它是三维卷积的简单替代方案,通过分解和共享多视图信息,从而提高模型学习能力,使二维卷积更高效地学习三维空间的上下文信息。然而,这些分割网络由于忽略脑部语义信息的理解,增加了模型训练难度,降低了模型的表现。

近年自然图像处理领域也诞生并开源了一些大模型,例如 Dino^[16]和 SAM^[17]等,通过大量的无监督预训练和有监督训练,展现了极强的特征表征能力。本文设计了拥有强编码器的分割网络 EM-SAM,即通过大模型 SAM 主干网络微调的方法最大化模型特征学习效率。此网络的核心是利用预训练大模型强大的特征表征能力,增强网络对脑部语义信息的理解,降低网络训练所需成本,提升网络的准确性和泛化性。

3 基于大模型的分割算法 EM-SAM

EM-SAM 采用了基于视觉变换器模块(Vision Transformer, ViT)的主干网络,此主干网络为已训练完成的 SAM 模型的图像编码器,以 U 型解码结构作为整体网络框架,通过端到端的方式输出分割结果。

3.1 编码器结构

图 1 展示了 EM-SAM 的模型结构,使用了 SAM 模型基于 ViT 模块的图像编码器作为主干网络,即图中灰色背景的模块。ViT 是一种新型的神经网络架构,通过将图像块作为输入,并利用全局注意力机制,能够更好地捕捉图像的全局特征。从整体结构上,它是基于 MAE 预训练的标准 ViT,拥有 16 倍的补丁嵌入和 12 个 ViT 模块。与其他 ViT 模型不同的点在于,输入为 $1\,024 \times 1\,024$,意味着补丁嵌入后的特征图大小

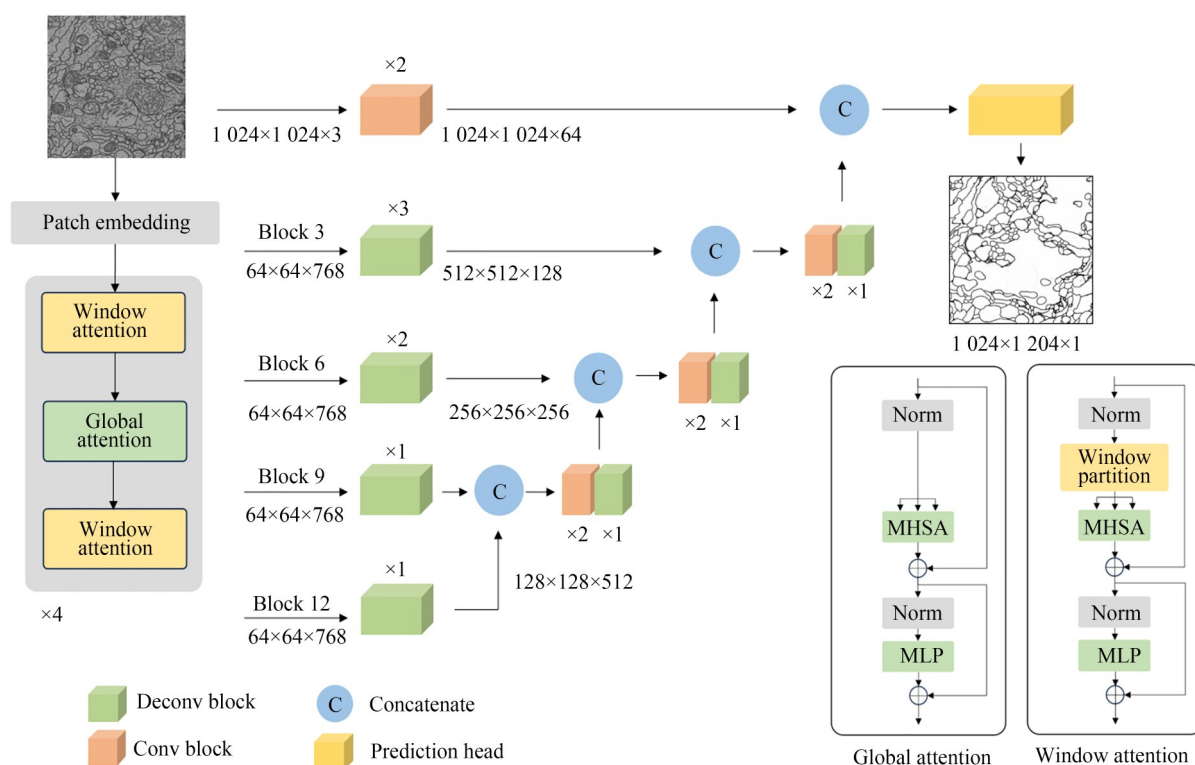


图1 EM-SAM模型结构示意图

Fig. 1 Illustration of EM-SAM.

为 64×64 , 需要大量的计算资源。由于运算量的限制, 它在 8 个 ViT 模块中采用 14×14 的窗口注意力, 在其余 4 个 ViT 模块中使用全局注意力。与^[18-19]不同, 本文从 SAM 的模型中删除了提示编码器 (Prompt Encoder) 和图像编码器的颈部 (Neck) 部分, 因为自动分割流程不涉及任何提示输入。图 1 的左侧展示了一个预训练的 SAM 模型中的补丁嵌入和注意力模块。由于预训练的 SAM 编码器采用的是 RGB 自然图像, 因此补丁嵌入层的输入为三通道。通过实验发现, 补丁编码层替换为单通道显著降低了模型的表现, 并增加了训练的不稳定性, 因此, EM-SAM 保留了三通道的补丁嵌入层。图像通过堆叠的方式变成三通道图像, 此外, 使用了 SAM 训练数据集提供的标准差和平均值进行归一化后, 再输入进编码器进行运算。如图 1 所示, 12 个 ViT 模块可划分成四个子模块, 每个子模块包含 2 个基于窗口注意力的 ViT 模块和 1 个基于全局注意力的 ViT 模块。

3.2 ViT 模块

视觉变换器模块是一种基于 Transformer 架构的图像处理模型, 它将自然语言处理领域的

Transformer 架构应用于计算机视觉任务, 在多项任务中都取得了优异的表现, 并具有很高的灵活性。ViT 的核心思想是将图像数据序列化以适应于 Transformer 的输入, 具体实现过程可以分为三步: (1) 图像分块: 将输入图像划分为固定大小的非重叠小块 (patches)。例如, 将一张 224×224 的图像划分为 16×16 的块, 每个块就变成 $16 \times 16 \times 3$ 的向量。(2) 线性投影: 每个图像块通过线性变换投影到一个固定长度的向量空间。这个过程类似于将图像块视为一维的“词”, 转换成 Transformer 模型所需的输入格式。(3) 位置编码: 与自然语言中的词一样, 图像块也需要位置编码以保留其在图像中的位置信息。ViT 将位置编码添加到每个图像块的向量中, 帮助模型理解图像块的相对位置。ViT 中也保留了标准的 Transformer 的自注意力机制。因此 ViT 能够捕捉到图像块之间的全局依赖关系, 而不仅限于局部信息, 进而能够高效学习图像的全局特征, 在图像任务中取得出色表现。

3.3 解码器结构

解码器结构采用 U 型设计, 经过 12 个 ViT

模块组成的编码器处理输入图像后,解码器分别提取四个子模块的特征图,即第 3,6,9 和 12 个 ViT 模块的输出,用于后续的解码。由于编码后的特征图与原图尺度差异过大,在解码过程中使用反卷积操作代替传统的上采样或者插值。这样的设计能提高解码运算之间的联系,利于梯度的反向传播,从而降低网络训练难度。反卷积后的特征将与更高维度的特征堆叠,然后输入到由残差卷积块构建的卷积块中。如图 1 所示,经过三次堆叠操作后,得到大小为 512×512 , 128 通道数的特征图,这些特征都是由 ViT 模块提取的。随后,使用两个卷积块对浅层特征进行进一步提取,以弥补编码过程中丢失的上下文信息。最终,经过预测头模块,获得大小为 $1\,024 \times 1\,024$ 的单通道预测结果。

4 实验结果

4.1 实验设定

本文对电镜数据集上主流的分割方法进行了解析,同时还对涉及自然图像挑战中提出的主流模型进行对比。本章详细考察了电镜数据集^[20, 21]上最先进的方法,这些方法包括使用 CNN 来预测二值图、亲和力图,以及额外的目标包括轮廓和距离。同时,还对自然图像挑战中主流基于 ViT 的模型进行对比^[22]。

4.2 数据集

本章采用上节所提训练数据集和测试数据集。训练时输入图像为整图,并使用相同的数据增强。测试时使用相同的测试增强,整图填充大小为模型输入大小的 $1/2$,填充完后截取补丁输入分割模型,截取的 x 和 y 方向步长设置为输入大小的 $1/2$ 。

4.3 模型及训练参数

模型解码器参数如图 1 所示,四个尺度分别为 $[512, 256, 128, 64]$ 。训练时为了确保与基于 ViT 的模型公平比较,本章对 ViT 使用与 EM-SAM 相同的解码器结构,对基于 CNN 的方法,本章使用了最先进的基于多任务学习 U-Net,并将其二维化,所有模型都采用相同的训练参数。

针对微调任务,首先,使用 Warmup Cosine Decay 学习率,预热迭代为 5 000,权重衰减为 0.05。其次,使用了 SGD 优化器通过更新所有

参数从而对预训练编码器进行微调。训练迭代总数设置为 200 000,且使用 Mini-batch size,大小为 2。最后,在训练中使用了二元交叉熵(Binary Cross-Entropy, BCE)损失,并在 2 张 NVIDIA 3090 GPU 上进行了训练。

与 CNN 网络不同,预训练模型的原始任务都是处理自然图像,因此输入为三通道的相同图像。

4.4 实验实施方案及评价指标

本章实验在 SNEMI3D^[21]和 MitoEM-R^[20]的数据集上分析了主流方法(U-Net),以及 ViT 的性能。本章将 EM-SAM、基于 U-Net 的 CNN 方法和 ViT 在不同形态特征的神经元分割任务中进行分析。分割实验使用 SNEMI3D 和 MitoEM-R 的训练数据集,训练真实电镜图像的分割模型。为了全面、公正地评估不同方法对分割性能的提升,本文在 SNEMI3D 数据集实验中采用原数据集使用的评价指标 Adapted Rand F-Score^[23](A-Rand)进行分析。A-Rand 即 $1 - \text{F-score}$, F-score 可以定义为:

$$\text{F-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (3)$$

其中:准确率(Precision)和召回率(Recall)的 TP , TN , FP 和 FN 对于像素分类分别为真阳性,真阴性,假阳性和假阴性。

MitoEM-R 数据集对比实验使用原文所提 mAP(Mean Average Precision),即各线粒体 AP(Average Precision)的平均值,其中 AP 是在不同 Recall 下的 Precision 值的均值,Recall 和 Precision 的定义分别在公式(2)和公式(3)。此外,AP-50 和 AP-75 代表 IoU 阈值为 0.5 和 0.75 时的 AP 测量值,即计算 TP 和 FP 时以 IoU 阈值作为标准,其中 IoU 表示某一类目标实际分割区域和真实标签区域的交集与其并集的比率。

4.5 SNEMI3D 对比实验

SNEMI3D 数据集是一个用于三维神经元图像分割的标准数据集。该数据集提供了神经元图像的原始数据和对应的手动标注分割结果,具有较高的空间分辨率,能清晰地显示神经元和突

触等细微结构。该数据集的任务是密集神经元的实例分割。本节的对比试验在 SNEMI3D 数据集上将所提出的 EM-SAM 与 U-Net 与 ViT 等先进的分割方法对比,实验设置遵循了原数据集提出的实验设置。实验指标如前所述采用 A-Rand,实验结果如表 1 所示。结果显示所提方法取得了最好的表现,对比 U-Net 提升最大,A-

Rand 降低了 36.47%,对于标准的 ViT 降低了 6%。这表明了所提 EM-SAM 在 SNEMI3D 数据集上取得了优于当前通用的最先进的分割模型的性能。

如图 2 所示,相比于 U-Net 方法,ViT 通过自注意力捕获长程依赖性的能力克服了分割目标物体尺度差异的分割挑战。此外,EM-SAM 方

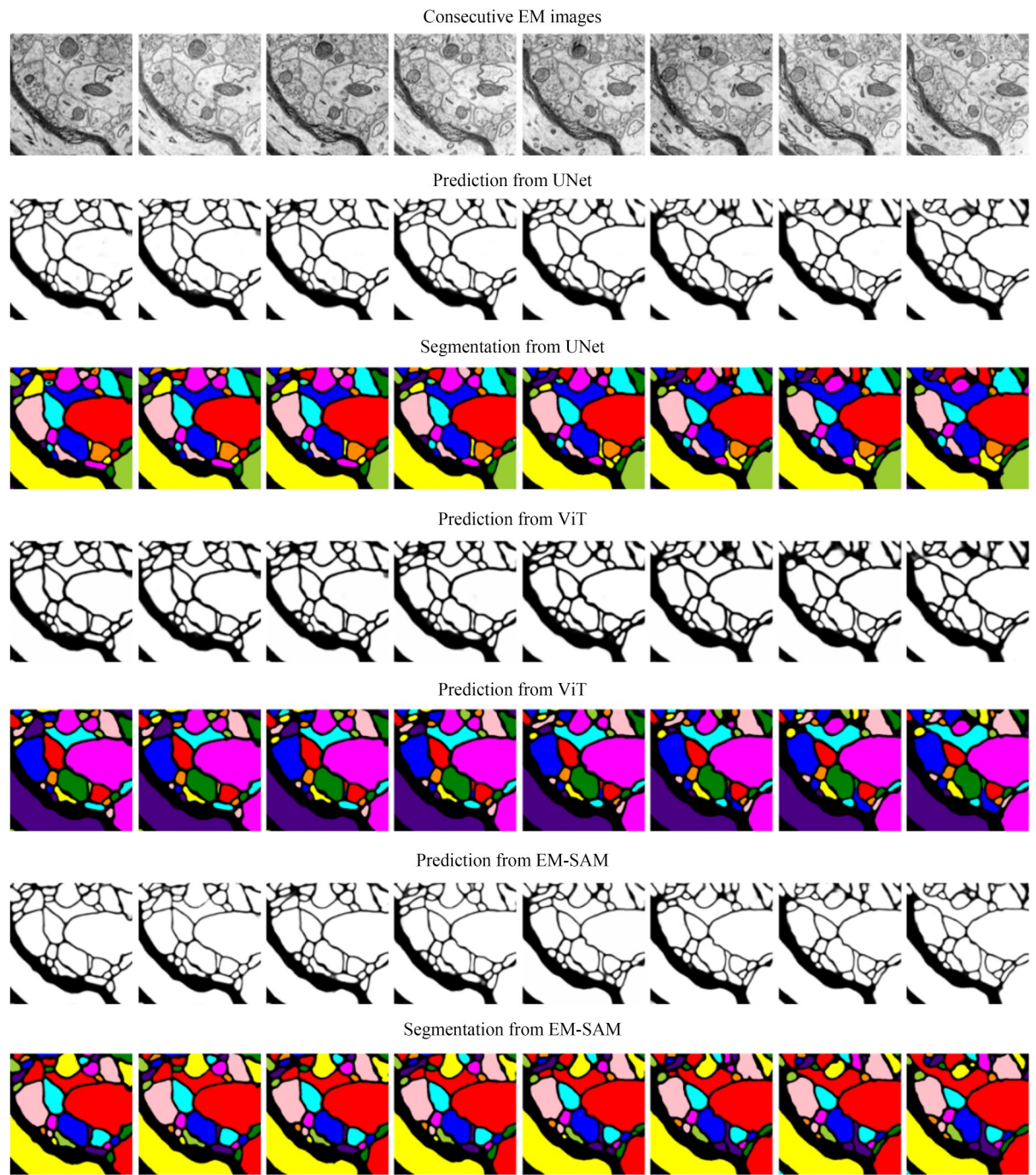


图 2 SNEMI3D 数据集实验结果

Fig. 2 Experimental results of SNEMI3D dataset

表 1 SNEMI3D 分割结果

Tab. 1 Segmentation results of SNEMI3D

方法	A-Rand
U-Net	0.085
ViT-B	0.058
EM-SAM	0.054

法由于较大的输入尺度($1\,024\times1\,024$),相比于 ViT 获得了更优的结果,细胞的边界预测更精准。

4.6 MitoEM-R 对比实验

本节实验使用 EM-SAM 在 MitoEM-R 数据集上进行对比实验,测试分割模型的有效性。该数据集的任务是线粒体的实例分割。本节对比实验按照原数据集提出的相同实验设置,使用二值图和实例轮廓作为网络预测目标。此外,所有

模型的后处理步骤均遵循默认配置,即通过二值图和轮廓预测生成最终的分割结果。最后,计算 MitoEM-R 验证数据集上的 mAP 值以展示分割模型的有效性。实验结果如图 2 和表 2 所示,在 AP-50 和 AP-75 指标中,所提的 EM-SAM 均表现出了最优的性能。在 AP-50 中,对比 U-Net 提升了 34%,对比 ViT-B 提升了 60%。同样的在 AP-75 中也有显著的提升,对比 U-Net 提升了 44.8%,对比 ViT-B 提升了 247%。因此,综合所提的 EM-SAM 在 MitoEM-R 数据集上 AP-75 和 AP-50 两个指标的优异表现,可以认为所提的模型对比 U-Net 和 ViT 拥有最好的性能。

如图 3 所示,ViT 的表现弱于 U-Net,这是由于在非密集预测任务中全局自注意力机制会引入大量背景区域进行计算,增加任务的难度。然而 EM-SAM 表现最优,说明大量预训练的全局

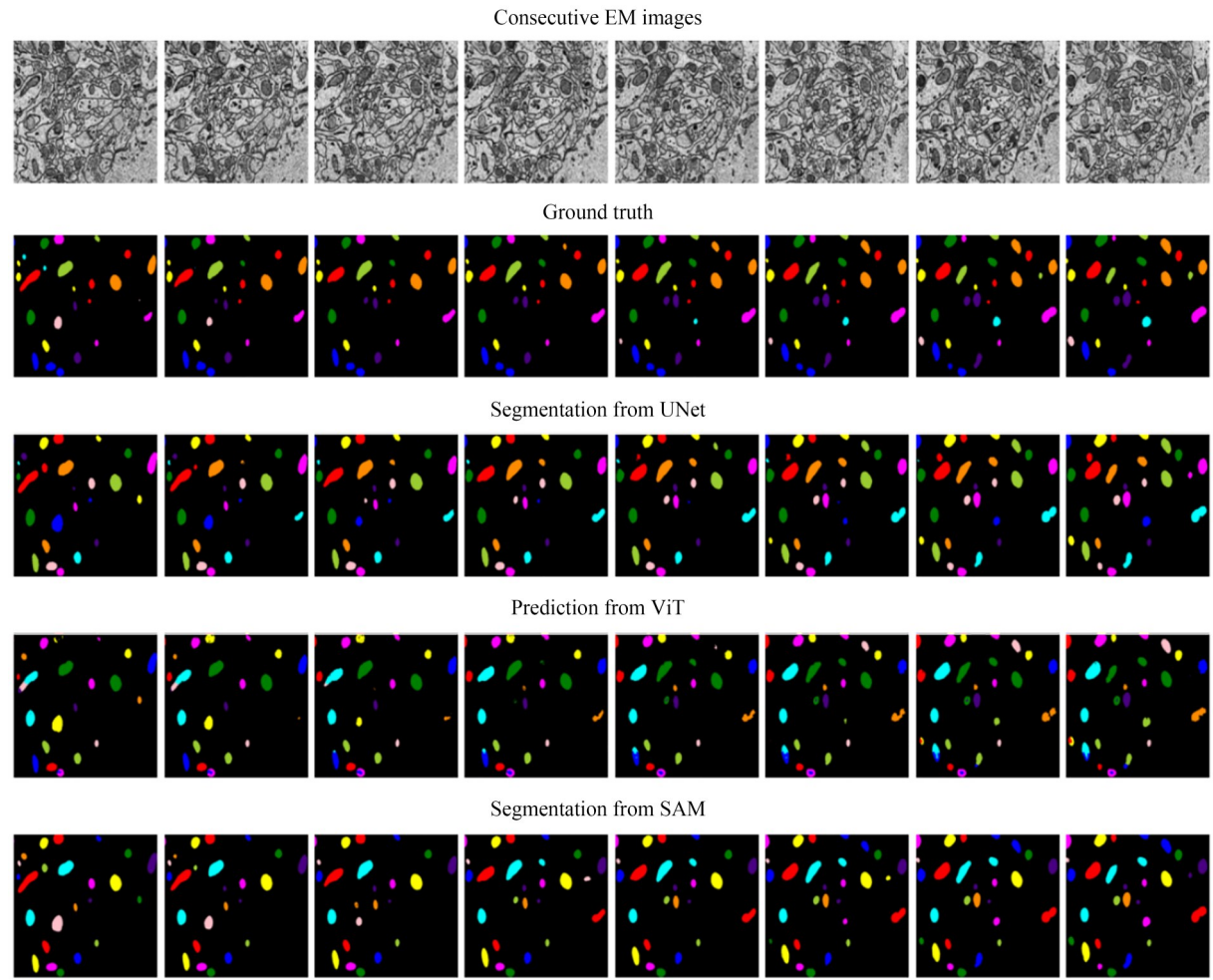


图 3 MitoEM-R 数据集实验结果
Fig. 3 Experimental results of MitoEM-R dataset

表 2 MitoEM-R 分割结果

Tab. 2 Segmentation Results of MitoEm-R

方法	AP-50	AP-75
U-Net	0.661	0.417
ViT-B	0.549	0.174
EM-SAM	0.883	0.604

自注意力机制模型通过其较强的特征提取能力,可以克服此类挑战。

4.7 消融实验

本节消融实验基于 SNEMI3D 数据集,通过对比随机初始化的 EM-SAM 和使用预训练 SAM 主干网络的 EM-SAM,证明微调方法的有效性。

如表 3 所示,随机初始化的 EM-SAM 获得了 0.076 的 A-Rand, 低于预训练 EM-SAM 的 0.054。此外,如图 4 所示,由于随机初始化的 EM-SAM 的特征提取能力较弱,因此预测结果中存在部分低置信度区域,导致分割准确性较低。

表 3 SNEMI3D 消融实验结果

Tab. 3 Ablation study result of SNEMI3D

方法	A-Rand
Random initial EM-SAM	0.076
EM-SAM	0.054

5 结 论

本章提出一个拥有强编码器的通用分割网络 EM-SAM,使用预训练的 SAM 大模型作为主干网络,利用了预训练模型的特征表征能力,增强网络对脑组织各目标的语义信息理解,并结合

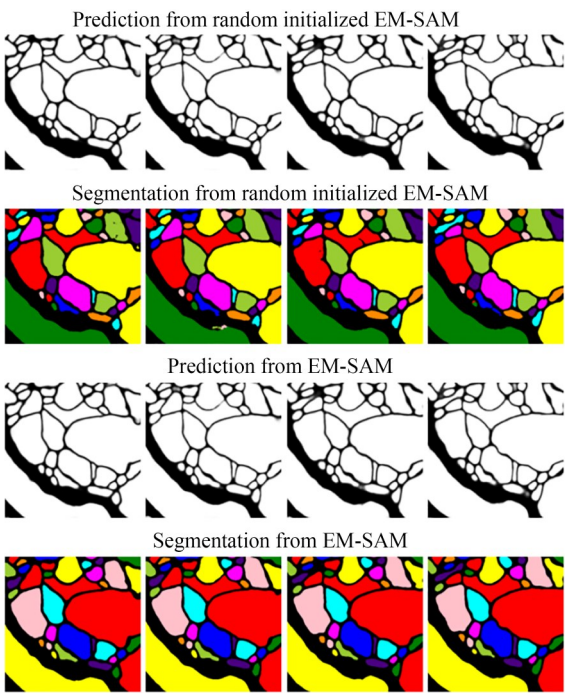


图 4 SNEMI3D 数据集消融实验
Fig. 4 Ablation study of SNEMI3D dataset

U 型解码器的设计,通过微调网络,最大化模型的特征学习效率,提高网络的准确性和泛化性,同时降低网络训练的成本。在 SNEMI3D 数据集上进行的对比实验中,EM-SAM 的 A-rand 指标对比 U-Net 和 ViT-b 分别降低了 36.47% 和 6%,在 MitoEM-R 数据集上的对比实验结果中提升同样显著,在 AP-50 中,对比 U-Net 提升了 34%,对比 ViT-B 提升了 60%。在 AP-75 中,对比 U-Net 提升了 44.8%,对比 ViT-B 提升了 247%。这充分证明了 EM-SAM 能够高效地分割小鼠脑部组织,并且相比其他主流方法,具有通用性和准确性,为大尺度电镜图像分割应用提供解决方法。此外,该微调方法的有效性证明了大型自然图像处理模型可以通过微调,自适应高通量的电镜图像处理任务。

参考文献:

[1] FRANGAKIS A S, HEGERL R. Segmentation of two- and three-dimensional data from electron microscopy using eigenvector analysis[J]. *Journal of Structural Biology*, 2002, 138(1/2): 105-113.

[2] KYLBERG G, UPPSTRÖM M, HEDLUND K O, et al. Segmentation of virus particle candidates in transmission electron microscopy images [J]. *Journal of Microscopy*, 2012, 245(2): 140-147.

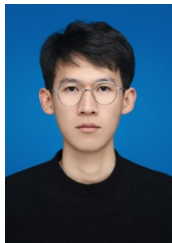
[3] LIU T, JONES C, SEYEDHOSSEINI M, et al. A modular hierarchical approach to 3D electron microscopy image segmentation[J]. *Journal of Neuroscience Methods*, 2014, 226: 88-102.

- [4] LIU T, JURRUS E, SEYEDHOSSEINI M, *et al.* Watershed merge tree classification for electron microscopy image segmentation[J]. *Proceedings of the IAPR International Conference on Pattern Recognition International Conference on Pattern Recognition*, 2012, 2012: 133-137.
- [5] KARABAĞ C, JONES M L, PEDDIE C J, *et al.* Segmentation and modelling of the nuclear envelope of HeLa cells imaged with serial block face scanning electron microscopy[J]. *Journal of Imaging*, 2019, 5(9): 75.
- [6] FAKHRY A, ZENG T, JI S W. Residual deconvolutional networks for brain electron microscopy image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2017, 36(2): 447-456.
- [7] XIAO C, LIU J, CHEN X, *et al.* Deep contextual residual network for electron microscopy image segmentation in connectomics[C]. 2018 *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Washington, DC, USA. IEEE, 2018: 378-381.
- [8] JIANG Y, XIAO C, LI L L, *et al.* An effective encoder-decoder network for neural cell bodies and cell nucleus segmentation of em images[C]. 2019 *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin, Germany. IEEE, 2019: 6302-6305.
- [9] CASSER V, KANG K, PFISTER H, *et al.* Fast mitochondria detection for connectomics [C]. *Proceedings of the Medical Imaging with Deep Learning*, 2020: 111-120.
- [10] CAO Y, LIU S G, PENG Y L, *et al.* DenseUNet: densely connected UNet for electron microscopy image segmentation[J]. *IET Image Processing*, 2020, 14(12): 2682-2689.
- [11] QUAN T M, HILDEBRAND D G C, JEONG W K. FusionNet: a deep fully residual convolutional neural network for image segmentation in connectomics [J]. *ArXiv e-Prints*, 2016: arXiv: 1612.05360.
- [12] LUO Z R, WANG Y, LIU S K, *et al.* Hierarchical encoder-decoder with soft label-decomposition for mitochondria segmentation in EM images[J]. *Frontiers in Neuroscience*, 2021, 15: 687832.
- [13] KHADANGI A, BOUDIER T, RAJAGOPAL V. EM-stellar: benchmarking deep learning for electron microscopy image segmentation[J]. *Bioinformatics*, 2021, 37(1): 97-106.
- [14] SPIERS H, SONGHURST H, NIGHTINGALE L, *et al.* Deep learning for automatic segmentation of the nuclear envelope in electron microscopy data, trained with volunteer segmentations[J]. *Traffic*, 2021, 22(7): 240-253.
- [15] YUAN Z M, MA X F, YI J J, *et al.* HIVE-Net: Centerline-aware hierarchical view-ensemble convolutional network for mitochondria segmentation in EM images[J]. *Computer Methods and Programs in Biomedicine*, 2021, 200: 105925.
- [16] OQUAB M, DARCET T, MOUTAKANNI T, *et al.* Dinov2: Learning Robust Visual Features without Supervision [EB/OL]. 2023: arXiv: 230407193.
- [17] KIRILLOV A, MINTUN E, RAVI N, *et al.* Segment anything[C]. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France. IEEE, 2023: 3992-4003.
- [18] MA J, HE Y T, LI F F, *et al.* Segment anything in medical images [J]. *Nature Communications*, 2024, 15: 654.
- [19] WU J, FU R, FANG H, *et al.* Medical Sam Adapter: adapting segment anything model for medical image segmentation [EB/OL]. 2023: arXiv:230412620.
- [20] WEI D L, LIN Z D, FRANCO-BARRANCO D, *et al.* MitoEM dataset: Large-scale 3D Mitochondria Instance Segmentation from EM images[C]. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 2020: 66-76.
- [21] LIN Z, WEI D, PETKOVA M D, *et al.* Nucmm Dataset: 3D neuronal nuclei instance segmentation at sub-cubic millimeter scale; proceedings of the medical image computing and computer assisted intervention-MICCAI [C]. 24th *International Conference, Strasbourg, France*, Springer, 2021.
- [22] HE K M, CHEN X L, XIE S N, *et al.* Masked

autoencoders are scalable vision learners[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA. IEEE, 2022: 15979-15988.

[23] NUNEZ-IGLESIAS J, KENNEDY R, PARAG T, *et al.* Machine learning of hierarchical clustering to segment 2D and 3D images[J]. *PLoS One*, 2013, 8(8): e71715.

作者简介:



承 骛(1997—),男,江苏苏州人,博士,2019年于英国曼彻斯特大学获得硕士学位,2024年于苏州大学获得博士学位,主要研究方向为医学图像处理。E-mail: 20204028006@stu.suda.edu.cn

通讯作者:



王丽荣(1966—),女,辽宁沈阳人,博士,教授,博士生导师,1989年毕业于吉林大学(原吉林工业大学)光电技术专业,2006年于吉林大学获得博士学位,主要研究方向为信息处理与自动控制的教学与科学研究。E-mail: wanglirong@suda.edu.cn