计算机科学
Computer Science
ISSN 1002-137X,CN 50-1075/TP

《计算机科学》网络首发论文

题目：基于深度学习的图像分割综述
作者：黄雯珂，滕飞，王子丹，冯力
收稿日期：2023-09-01
网络首发日期：2023-11-20
引用格式：黄雯珂，滕飞，王子丹，冯力．基于深度学习的图像分割综述[J/OL]．计算机科学. https://link.cnki.net/urlid/50.1075.TP.20231118.1805.004

基于深度学习的图像分割综述

黄雯珂[1] 滕飞[1] 王子丹[1] 冯力[1]

1 西南交通大学计算机与人工智能学院 成都 611756)
（huangwenke321@163.com）

摘 要 图像分割是计算机视觉中的一项基本任务，其主要目的是从图像输入中提取有意义和连贯的区域。多年来，在图像分割领域已经开发了各种各样的技术，包括基于传统方法，以及利用卷积神经网络的最新图像分割技术。随着深度学习的发展，更多的深度学习算法也被应用到图像分割任务中。尤其地，近两年来学者对于深度学习的兴趣高涨，涌现了许多应用于图像分割任务的深度学习算法，然而大部分新的算法还没有被归纳分析，这将不利于后续研究的进行。在本综述中，提供了近两年发表在文献中基于深度学习的图像分割研究的全面回顾。首先，本文对图像分割的常用数据集进行了简要介绍，接着阐明了基于深度学习的图像分割的新分类，最后，讨论了现有的挑战，并对今后的研究方向进行了展望。

关键词： 图像分割；语义分割；深度学习；网络结构；监督学习

中图法分类号 TP391

Image Segmentation Based on Deep Learning: A Survey

HUANG Wenke[1], TENG Fei[1], WANG Zidan[1] and FENG Li[1]

1 School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756

Abstract Image segmentation is a fundamental task in computer vision and its main purpose is to extract meaningful and coherent regions from the image input. Over the years, a wide variety of techniques have been developed in the field of image segmentation, including those based on traditional methods, as well as more recent image segmentation techniques utilizing convolutional neural networks. With the development of deep learning, more deep learning algorithms have been applied to image segmentation tasks. In particular, there has been a surge of scholarly interest in deep learning over the past two years, and many deep learning algorithms have emerged for image segmentation tasks. However, most of the new algorithms have not been summarized or analyzed, which will hinder the progress of subsequent research. In this review, we provide a comprehensive review of deep learning-based image segmentation research published in the literature in the past two years. First, we briefly introduce common datasets for image segmentation. Next, we clarify new classifications for image segmentation based on deep learning. Finally, we discuss existing challenges and look forward to future research directions.

Key words Image segmentation; Semantic segmentation; Deep learning; Network structure; Supervised learning

1 引 言

图像分割在计算机视觉中起着至关重要的作用，是计算机视觉最关键的领域之一，它也是物体识别、医学成像、自动驾驶和图像编辑等广泛应用的基础。图像分割的主要目标是提取连贯的和有意义的部分或区域，为了实现这一目标，多年来发展了大量而广泛的技术，包括基于传统的技术，如阈值分割、边缘检测和区域增长，以及最近在深度学习中利用包括卷积神经网络（CNN）或深度学习模型（GAN、Transformer）的技术。尽管已经取得了巨大的进步，图像分割仍然面临着准确描绘复杂对象边界、优化网络结构和处理多域数据等方面的挑战。

本文所引用的文献大部分来自 CVPR、ICCV、AAAI 等相关会议，其他文献来自包括 TMI、Nature 等期刊或会议，文献的来源分布如图 1 所示。

图像分割领域的学者已经探索了各种方法来解决多领域数据、弱监督、实例分割和全景分割

到稿日期：2023-09-01 返修日期：2023-11-15
通信作者：滕飞（fteng@home.swjtu.edu.cn）

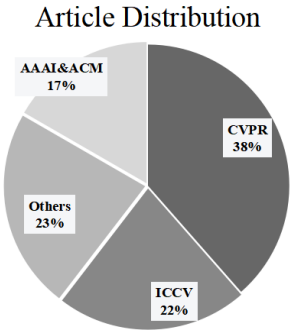

图 1 文献分布

Fig.1 Article distribution

等挑战，也越来越多地使用深度学习方法，其中使用全卷积神经网络（FCN）的方法引起了人们的极大兴趣，并在领域内广为普及使用。在深度学习方法中，监督学习仍然是一个引人注目的研究领域，与之相反的是，最近的研究表明，人们对传统方法的兴趣减弱了。以自监督、无监督和半监督学习方法作为获取真值(GT)标签的解决方案已经开始被学者们关注并加以利用，且在图像分割的研究中这些方法的利用已经形成了一个日益增长的趋势。基于 Transformer 的模型和自监督学习技术因其在提高分割精度方面的有效性而引起了人们的广泛关注。包括利用动态学习、生成对抗网络和循环架构在内的，用以提高分割性能和效率的网络结构或学习策略的图像分割方法，也逐渐开始出现在人们的视野中。

基于深度学习的图像分割技术在近几年已经逐渐成为计算机视觉的研究热点，Saeid Asgari Taghanaki 等人[1]对图像分割技术进行了全面的综述，介绍了语义分割在自然和医学图像中的进展，并将语义分割文献按照其贡献程度分为了六类：基于架构的改进、基于优化函数的改进、基于数据合成的改进、弱监督模型、序列模型和多任务模型。Manar Aljabri 等人[2]总结了在医学图像分割领域的深度学习模型，强调不同深度学习架构在医学图像分割应用中的特殊贡献，同时展示了一些医学图像分割的应用场景。包括 Jayashree Moorthy and Usha Devi Gandhi[3]、Mohamed T. Bennai 等人[4]都对医学图像分割提出了不同的分类方式，并综述了当时最有效、最先进的图像分割方法，图像分割在医学图像领域早已是一项非常重要的技术。同时可以发现，图像分割技术的发展迅速，所以对最新研究进展的回顾和总结是十分必要的。

与基于基本方法论（如监督学习、半监督学习或无监督学习）总结相关工作的综合性的综述相比，本文增加了对所研究文章的优化动机的关注，主要目的是探讨热点研究问题，综合最佳思路，为解决相关问题提供方向指导。因此，我们将基于深度学习的图像分割的技术进展分为三大类：基于网络结构的改进，基于数据的改进，基于评价机制的改进。所有的分类如图 3 所示，其中如果单分类下存在两个分支的方法，分支上方代表这是基于视频进行分割的，下方代表是基于图像进行分割的。本文将通过综合考虑研究的动机和研究所完成的具体任务来总结最近的工作，为解决相关问题提供指导。

本文第 2 节简要介绍了常见的数据集,第 3～5 节阐明新的分类——基于网络结构的改进，基于数据的改进，基于评价机制的改进；第 6 节讨论了图像分割的现有挑战；第 7 节总结全文并提出对未来研究方向的展望。

2 常见数据集

为了对最新的图像分割研究进行更加清晰的

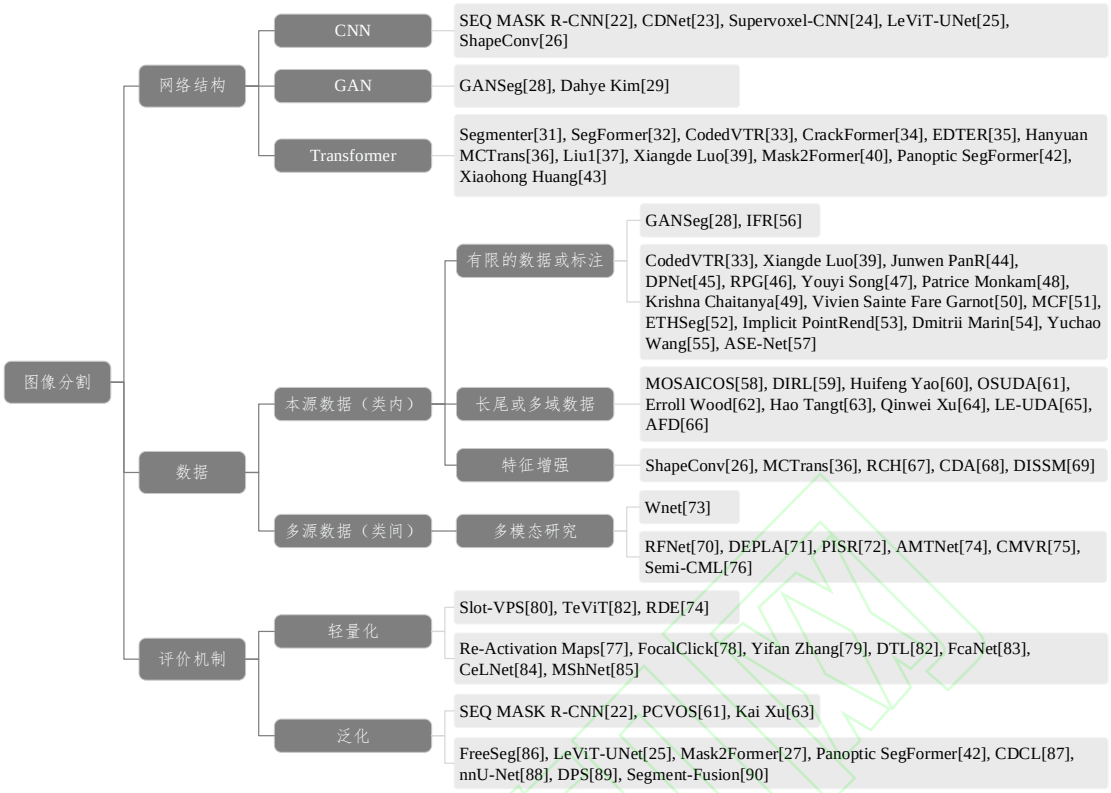

图 2 基于深度学习的图像分割分类

Fig. 2 Classifications of image segmentation methods based on deep learning

阐述，本文对领域内常用的数据集进行了广泛调研，并对最新研究中常用的几个数据集进行了简要介绍。图像分割领域中著名的公共数据集包括 COCO[5]、Cityscapes[6]、ADE20K[7]、Pascal Context[8]、BDD100K[9]、CamVid[10]、PASCAL VOC[11]、ScanNet v2[12]、YouTubeVOS[13]、DAVIS16/17[14]、CUB-200-2011[15]、SemanticKITTI[16]、LVIS[17]、LIP[18]和 GrabCut[19]。其中在最新研究中常用的是 COCO、Cityscapes、ADE20K、PASCAL VOC 这 4 个数据集。

COCO 数据集由微软构建，包含几种类型的标注：物体检测、关键点检测、实例分割、全景分割、图片标注，COCO 每个类别中包含的实例数量相较 PASCAL VOC 会更多，2015 年累积版本包含总共 165482 张训练图像、81208 张验证图像和 81434 张测试图像。

Cityscapes 数据集是常见的语义分割数据集，包含来自 50 个不同欧洲城市街道场景的图像、视频数据，其中有 5000 张城市环境驾驶场景的像素级精细标注图像（2975 张训练图像、500 张验证图像、1525 张测试图像），和 20000 张粗略标注图像，精细标注图像主要用于强监督学习，粗略标注图像主要用于弱监督语义分割。

ADE20K 数据集在图像分割中常用与场景解析，包含了 25000 张复杂的日常场景图，其中有 20210 张训练图像、2000 张验证图像、3000 张测试图像。数据集的标注主要是用于场景理解的像素级注释，包含离散对象、有无定型背景区域的物体、对象三种视觉概念。

PASCAL VOC 数据集来自 PASCAL VOC 挑战赛，主要用于分类识别和目标检测，也可以用于图像分割、动作识别和人体布局等任务，是目标检测技术重要的基准之一。数据集包含 11540 张用于分类识别和目标检测任务训练和验证的图像，

6929 张用于图像分割任务的图像。

3 基于网络结构的改进

基于网络结构的改进旨在增强基础网络架构，提升网络模型的性能和泛化能力，并在网络成本、准确率、速度之间取得平衡。本文将各种基础神经网络方法按照其所提出模型的网络结构进行了分类，这些网络结构包括 CNN、GAN、Transformer。

3.1 CNN

卷积神经网络（CNN）是具有三维神经元排列的神经网络，包括宽度、高度、深度 3 个维度，它的网络结构可以分为以下几个部分：卷积层，激活函数、池化层和全连接层。其中卷积层产生了主要的计算量，所以也被称作 CNN 的核心，它相当于图像处理中的滤波器，能够捕捉到局部的特征点，例如图像的边缘或斑点等。卷积层的输出要经过一个激活函数，让输出转换为非线性，同时允许模型学习更复杂的特征。池化层中的下采样用于降低输出的维度，防止过拟合，也能够减少参数量。经典的 CNN 网络中经过两次卷积、池化后的输出进入全连接层，在全连接层中映射到样本标记空间，其中每一层相当于序列神经元的平铺。

2012 年基于 CNN 改进的 AlexNet[20]成功面世并引起了巨大的反响，在 AlexNet 网络中包含 5 个卷积层、3 个池化层和 2 个全连接层，它的网络更加复杂，但也意味着它能够更加准确地处理图像。由此，CNN 在图像分割中具有的出色特征提取和表达能力被人们发现，并引起了学者们极大的兴趣。CNN 在语义分割模型中的应用有着巨大的多样性，为了克服 CNN 具有固有归纳偏差、分割利用率低等限制，发展出了如 R-CNN、FCN、UNet、DeepLab 等经典模型，这些模型在相当长的一段时间里占据了语义分割模型的重要地位。AliF.Khalifa 等人[21]在阐述图像分割在医学图像领域中的重要性的同时重点介绍了 CNN 的架构改进，针对于 CNN 的大量图像分割技术仍然在不断涌现。

Huaijia Lin 等人[22]提出了一种新的范式——Propose-Reduce 和一个 Seq Mask R-CNN 框架，通过一步生成输入视频的完整序列，他们进一步在现有的图像级实例分割网络上建立了一个序列传播头来进行长期传播。Seq Mask R-CNN 通过在 Mask R-CNN 上增加一个额外的序列传播头，建立帧间的时间关系。Hongliang He 等人[23]提出了一种新的核实例分割的向心方向网络（CDNet）。具体来说，他们将向心方向特征定义为一类指向核中心的相邻方向，来表示核内像素之间的空间关系。为了使网络更加轻量化，研究人员引入了超体素的概念，即将具有相似特征的体素云聚类来减少三维空间的计算量，Shi-Sheng Huang 等人[24]探索了一种基于超体素的深度学习解决方案，形成一个有效的基于超体素的卷积神经网络（Supervoxel-CNN），Supervoxel-CNN 可以有效地融合在线三维重建过程中投影在超体素上的多视图 2D 特征和 3D 特征，期望在效率和分割精度之间取得平衡。Guoping Xu 等人[25]提出了一种将 LeViT Transformer 模块集成到 UNet 架构中的 LeViT-Unet，用于进行快速而准确的医学图像分割，可以有效地重用特征映射的空间信息，且更好地平衡了分割的准确性和效率。Jinming Cao 等人[26]介绍了一个形状感知卷积层（ShapeConv）处理深度特征，深度特征首先分解成形状组件和基本组件，下两个可学习的权重与他们独立合作，最后一个卷积应用于这两个组件的重加权组合。CNN 网络对图像特征的有效提取在以上工作中起了关键支撑作用，非线性的处理能力让 CNN 在不同的维度中提取图像特征时显得非常优越。

3.2 GAN

生成对抗网络（GAN）[27]属于生成式模型，该类模型在无监督深度学习方面占据着重要位置。GAN 的提出是受到博弈论中的“二人零和博弈”的启发，它通过框架中的生成模型网络和鉴别模型网络不断互相博弈学习来获得更好的分布收敛，二者一方进行伪造一方进行侦辨，使得模型在这种博弈过程中互相训练，两个模型不断竞争对抗，直到达到纳什均衡。

虽然 GAN 能够在自我博弈中获取较高的性能，但像素间的关系易被忽略而造成特征的缺失，分割结果也可能因此而不够连续或产生较大形变。为了实现提升 GAN 进行图像分割的性能、补全全局关系、加强卷积分割网络等目标，GAN 的变体 SegAN、SCAN、PAN 等网络模型应运而生。

Xingzhe He 等人[28]提出了一种基于 GAN 的分割方法 GANSeg，它可以生成基于潜在掩膜条件的图像，减轻了现有的自动编码网络中对图像对和预定义图像变换的依赖，从而减轻了以往方法所需要的完整或弱注释。Dahye Kim 等人[29]提出了一种无监督图像分解算法，以获得一种基于乘法图像模型的内在表示，该表示对不良偏差场具有良好的鲁棒性。分割模型进一步设计为在 GAN 框架中加入几何约束，使分割函数分布与先验形状分布之间的差异最小化。

3.3 Transformer

深度架构的改进已经成为一个研究热点，Transformer 也逐渐成为基于深度学习的图像分割方法中的主流架构。Transformer 最早由 Vaswanid 等人[30]提出，由于其编解码器的独特设计，它能够轻易地处理不确定的输入和捕获长距离依赖关系，Transformer 也因此逐渐成为自然语言处理等人工智能领域的主要深度学习模型。以 Transformer 为基础衍生出了各种强大的分割模型，且这些模型几乎都能达到目前最先进的水平，基于 Transformer 的图像分割方法在最新的研究中也受到了高度的重视。

原始的 Transformer 由编码器和解码器组成，在编解码器中都使用了自注意力机制加上前馈神经网络的结构，在 Transformer 中还使用了多层的自注意力机制构成的多头注意力机制，提升了模型对不同信息之间相关度的捕获，这种特性很有利于分析上下文信息。Robin Strudel 等人[31]提出的 Segmenter 是一种用于语义分割的 Transformer 模型，与基于卷积的方法相比，它允许在第一层和整个网络中对全局上下文进行建模。SegFormer[32]反思了 Segmenter 其中出现的一些问题，比如需要位置编码而失去了图像大小的宽容性，而在大尺度训练中这一点会成为计算复杂度上较大的障碍，SegFormer 则做到了架构轻量化和特征图层次化。

文献如[33-39]都是基于问题而提出的 Transformer 图像分割模型，其中，Mengyang Pu 等人[35]提出了一种基于 Transformer 的边缘检测器，边缘检测变压器（EDTER），同时利用全图像上下文信息和详细的局部线索，提取清晰清晰的物体边界和有意义的边缘；Hanyuan Liu 等人[37]提出了一种创新的基于 Transformer 的框架，即基于高级特征和低级线索来识别虚线曲线。这两个工作都聚焦于对象边界检测和分割问题，在这类问题中，被分割目标的边界通常由背景的混合组成，而在其他情况下，分割目标是整个图像中的小部分。此外，边界可以表现出可变性和小尺度特征，这些特征导致目标边界检测和分割的精度较低。Transformer 的上下文依赖的特点能够很好的将边界特征利用起来，取得很好的分割效果。文献[40, 41]则在 Transformer 的帮助下获得了通用性的提升，本质上是利用了 Transformer 架构的稳定性和与其他网络模型结合的互补性来增

强模型的稳定性，例如文献[42]就是 Transformer 与高分辨率 CNN 特征图相结合，以实现精确的图像补丁定位。Xiaohong Huang 等人[43]也借鉴了 CNN 的思路，基于 U 型结构重新设计了 Transformer 块中的前馈网络，命名为 ReMix-FFN，它通过重新整合局部上下文和全局依赖关系来探索全局依赖关系和局部上下文，从而更好地识别特征。

4 基于数据的改进

本节介绍了图像分割方法基于数据层面的提升，通过对具体任务的区分，本文将基于数据的改进分为本源数据和多源数据两类，这两类可以分别试做专注于类内和聚焦于类间的两个不同的数据处理视角，这将有助于此后研究数据时的方向侧重。

4.1 本源数据（类内）

本源数据指出现问题的数据本身，或任务着重处理的数据并不与其他数据发生交互。根据这种分类标准，最近的研究主要以三个问题为导向：有限的数据或标注、长尾或多域数据、特征增强。

4.1.1 有限数据或标注

用于训练的数据集对于深度学习来说是至关重要的，当数据或标注不足时可能会使得模型过拟合或是无法较好地收敛，导致模型表现很差，但实际情况是这些数据或标注并不容易获得，或是十分昂贵。围绕有限数据或标注问题展开研究的工作中，共同面临着的挑战是如何在有限的样本中获得更好的模型效果，研究的方向主要是提升数据或标签的利用率。Junwen Pan 等人[44]提出了一种标签高效混合监督框架，该框架单独考虑每个弱注释实例，并在强注释实例的梯度方向引导下学习其权重。Binjie Mao 等人[45]提出了一种双原型网络（DPNet）来处理小样本语义分割，通过查询图像中前景特征的伪原型来挖掘样本中有用的信息。Constantin Marc Seibold 等人[46]提出了一种半监督语义分割的监督生成方法，共享标记图像和未标记图像之间的视觉相似区域的语义。Youyi Song 等人[47]提出了一种交替估计重要权值和更新网络的迭代算法，根据外部数据与内部数据之间的分布差异来估计重要性权重，并施加约束来驱动网络，比不使用外部数据的网络能更有效地学习。Patrice Monkam 等人[48]提出 SegMix 利用分段粘贴混合概念，基于一些手动获取的标签生成大量带注释的样本，同时提出了一种新的框架，该框架能够在超声（US）图像分割中部署深度学习方法，并且只需要非常有限的手动注释样本。Krishna Chaitanya 等人[49]提出了一种局部对比损失，通过利用从未标记图像的伪标签中获得的语义标签信息以及带有真值（GT）标签的有限注释图像来学习对分割有用的良好像素级特征。以上这些工作都是为了提高有限数量数据或标签的利用率。

Vivien Sainte Fare Garnot 等人[50]提出了第一个端到端、单阶段用于卫星图像时间序列（SITS）全光分割的方法。该模块可以与我们基于时间自注意的图像序列编码网络相结合，提取丰富、自适应的多尺度时空特征。Zunlei Feng 等人[51]提出了一种用于病理图像中核检测和分割的相互互补框架 MCF，它的两个分支以相互互补的方式进行训练，其中检测分支补充了分割分支的伪掩膜，而逐步训练的分割分支通过计算预测掩膜与检测结果之间的掩膜残差来补充缺失的核模板。Lingteng Qiu 等人[52]介绍了一个新的问题实例级垃圾分割，在 X 射线图像中进行智能垃圾检查，并贡献了一个由 5038 个 X 射线图像（总共 30881 废物项目）与高质量的注释（即垃圾类别、对象盒和实例级遮罩）组成的真实的数据集作为这个问题的基准。Bowen Chen 等人[53]提出了一种点注释方案来收集弱监督实例分割，除了边界框

之外，他们还为每个边界框内均匀采样的一组点收集了二进制标签。他们表明，现有的为全掩膜监督开发的实例分割模型可以通过他们的方案收集的基于点的监督进行无缝训练。

半监督语义分割的关键是为未标记图像的像素分配足够的伪标签。一种常见的做法是选择高度自信的预测作为伪真值，但它会导致一个问题，即大多数像素可能由于其不可靠性而没有被使用。Dmitrii Marin 等人[54]提出了一种鲁棒信任区域方法来实现正则化损失，提升在标注很少的样本中的弱监督学习效果，这也可以看作是经典链规则的高阶推广。Yuchao Wang 等人[55]认为每个像素都对模型训练很重要，即使它的预测也是模糊的。直观地说，一个不可靠的预测可能会在顶级类（即那些概率最高的类）之间发生混淆，然而，它应该确信该像素不属于其余的类。因此，这样的像素可以被令人信服地视为那些最不可能类别的负面样本。基于这一见解，他们开发了一个有效的管道来充分利用未标记的数据。为了更好地探索视频数据的半监督分割问题，Jiafan Zhuang 等人[56]提出了一个半监督的视频语义分割任务，在这项任务中他们观察到，在训练视频中有标记的帧和未标记的帧之间的过度提升非常严重，尽管它们在风格和内容上非常相似。这被称为内部视频过度提升，它实际上会导致较差的性能。为了解决这个问题，他们提出了一种新的帧间特征重建（IFR）技术，利用真值标签来监督未标记帧上的模型训练。Tao Lei 等人[57]提出了一种基于动态卷积的对抗自集成网络(ASE-Net)，可以同时计算不同数据扰动下未标记数据的像素级和图像级一致性，从而提高标注的预测质量，并且他们设计了一个基于动态卷积的双向注意分量(DyBAC)，能够提高 ASE-Net 的特征表示能力，降低网络的过拟合风险。

4.1.2 长尾或多域数据

长尾数据中，对象类的分布严重偏向于少数的主导类，而大量的类只有有限数量的实例，这些非主导类别称为尾类。在典型的对象检测任务中，数据集通常被认为是良好平衡的，每个类的实例数量大致相等。然而，在现实世界中，如监测、医学成像或稀有物种识别，物体类别的分布可能是高度不平衡。这是因为在系统中个体的尺度相差巨大，头和尾的尺度可能相差了几个数量级，同时也没有一个合适的数据规模来限制头数据带来的影响。长尾数据在学习中主导类往会掩盖尾类，导致尾类被忽视或表现不佳，这将使模型过拟合，严重影响模型的泛化性。这类问题在出现时往往伴随着其他的数据问题，最常见导致的就是在模型进行域的迁移时的表现差异大。Zhang Cheng 等人[58]提出了把以对象为中心的图像拼接为以场景为中心图像（MOSAICOS），能够减轻域之间的差异，在应对长尾对象检测的挑战时非常有效。

多域数据问题是指源域与目标域的图像风格、图像内容等信息存在差异，使得在源域上训练好的模型在目标域上的表现并不理想。其中源域是训练时使用的数据集，目标域则是在应用时使用的数据集。多域数据的概念与两个关键主题密切相关：域自适应（DA）和域泛化（DG）。DA 专注于解决将在一个域上训练的模型应用于不同但相关的域时模型性能下降的问题。在 DA 中，源域是指训练模型的域，而目标域是模型需要执行良好的域。源域和目标域通常具有数据分布的差异，例如照明条件、视点或成像模态的变化。DA 的目标是通过减少域偏移来调整模型，使其在目标域上表现良好，DG 解决了一个更具挑战性的场景，即模型需要很好地推广到与训练期间看到的所有域不同的看不见的目标域。

在相关研究中，Qi Xu 等人[59]提出了用于域泛化的域不变表示学习（DIRL），该方法利用特征敏感性作为特征，然后指导模型泛化能力的增强。Huifeng Yao 等人[60]提出了一种新的半监督域广义医学图像分割算法。Xinyi Wu 等人[61]解决了语义分割的一次性无监督域适应(OSUDA)问题，即分割者在训练过程中只看到一个未标记的目标图像。

研究人员试图通过数据混合、领域适应和领域对抗训练来弥补域间的差距，但 Errooll Wood 等人[62]表明，以最小的域差距合成数据是可能的，因此在合成数据上训练的模型可以推广到真实的野外数据集。Hao Tang 等[63]提出了一种基于原型网络的少镜头医学图像分割的新框架。他们的创新之处在于设计了两个关键模块： (1)上下文关系编码器（CRE），它使用相关性来捕获前景和背景区域之间的局部关系特征；(2)循环掩膜细化模块，重复使用 CRE 和原型网络来重新获取上下文关系的变化，并迭代细化分割掩膜。Qinwei Xu 等人[64]介绍了一种基于傅里叶的域泛化观点，主要的假设是傅里叶相位信息包含高级语义，不容易受域移的影响。Ziyuan Zhao 等人[65]提出了一个标签高效无监督域自适应通用框架（LE-UDA），为两个领域之间的知识转移构建了自组装一致性，并构建了一个自组装对抗性学习模块，以实现更好的 UDA 特征对齐，可以有效地利用有限的源标签来提高跨域分割性能。Jie Wang 等人[66]提出了一种新的基于特征解耦（FD）的无监督域自适应（UDA）方法,对抗性特征解耦（AFD），通过特征分离和图像转换来解耦内容相关特征和风格相关特征，以便于学习域不变特征。

4.1.3 特征增强

特征增强是指从对象中提取多尺度特征或从一系列嵌套的区域中提取特征，通过这些特征提高网络中卷积特征的质量，改进图像分割的预测性能。在关于特征增强的研究中，Hualiang Wang[67]提出了辅助表示校准头（RCH），它由图像解耦、原型聚类、误差校准模块和度量损失函数组成，以校准这些容易出错的特征表示，以获得更好的类内一致性和分割性能。Yukun Su 等人[68]提出了一种上下文解耦增强（CDA）方法，以改变对象出现的固有上下文，从而驱动网络消除对象实例与上下文信息之间的依赖关系，使模型更加关注对象的特征。Ashwin Raju 等人[69]提出了深度隐式统计形状模型（DISSMs），这是一种将深度网络的表示能力与 SSM 的好处结合起来的新方法。DISSMs 使用隐式表示来产生紧凑和描述性的深度表面嵌入，允许解剖方差的统计模型。Yuanfeng Ji 等人通过提出一个多复合 Transformer（MCTrans）网络，来解决对不同像素的跨尺度依赖关系、不同标签的语义对应关系以及特征表示和语义嵌入的一致性的学习，它将丰富的特征学习和语义结构挖掘集成到一个统一的框架中。此外，还引入了一种可学习的代理嵌入，分别利用自注意和交叉注意对语义关系和特征增强进行建模。

4.2 多源数据（类间）

与本源数据相对应，多源数据指的是有多个来源的数据，包括但不限于来自各种网络、传感器、媒体等不同信息源的数据。近两年来，人们越来越关注多模态的研究，多模态的研究也逐渐渗透到各个领域当中，图像分割技术也不例外。在本节中，本文将多模态数据作为多源数据的主要组成部分，也只对围绕或使用多模态数据的图像分割研究进行总结和分析，但这并不代表这两者是一一对应的关系。

人类社会被广泛认为是由多模态信息构成的世界，各种实义信息的准确传达通常都需要多种

模态信息的参与，所以多模态数据的研究也是推动人工智能和认知科学的关键性研究。模态是信息的一种来源或传递形式，在事情的发生、经历、结束的过程中获取、使用、传递信息的各种方式都可以被称作是模态。模态通常以人类的感官为基础来划分，其中视觉对应能够看到的图像、视频、文字信息，听觉对应能够听见的音频信息，触觉对应各种能够触摸到的压力、温度等信息，嗅觉对应能够闻见的气味信息等。在深度学习的背景下，模态通常是指不同类型的数据，如图像、文本、音频或传感器数据。

在多模态数据研究中，Yuhang Ding 等人[70]提出了一种区域感知融合网络（RFNet），能够自适应、有效地利用多模态数据的不同组合进行肿瘤分割。Shuyang Sun 等人[71]提出了 DEPLA，在给定两层要聚合的特征的前提下，首先检测并识别一层需要更新的位置和内容，然后将识别位置的信息替换为另一层的特征，能够避免不一致的模式，同时在合并的输出中保留有用的信息。

在现有的工作中，通常使用共享的主干来提取事物（可数类，如车辆）和东西（不可数类，如道路）的特性。然而，这并不能捕捉到它们之间丰富的关系，但这些关系可以用来提高视觉理解和分割性能。Shubhankar Borse 等人[72]提出了一种新的泛视、实例和语义关系（PISR）模块来利用这种上下文关系。音频引导视频对象分割是视觉分析和编辑中一个具有挑战性的问题，它根据参考的音频表达式自动将视频序列中的前景对象与背景分离。然而，由于缺乏对视听交互内容的语义表示建模，现有的参考视频对象分割工作主要集中在基于文本的引用表示的指导上。Wenwen Pan 等人[73]基于端到端的方法，从去噪编解码器网络学习的角度考虑了音频引导视频语义分割问题，他们提出了基于小波的编码器网络来学习跨模态表示，用于表示具有音频形式查询的视频内容，并且取得了不错的成效。Shenhai Zheng 等人[74]提出了一种新的用于三维医学图像分割的自动化多模态变换器网络（AMTNet），该网络是一个建模良好的 U 型网络架构，在特征编码、融合和解码部分进行了许多有效和显著的更改，此外，设计了一个基于 Transformer 的自适应通道交织 Transformer 特征融合模块，以完全融合不同模态的特征。Wenjing Zhang 等人[75]提出了一种用于指称分割的跨模态注意力引导视觉推理模型（CMVR），他们所设计的跨模态注意力模块自适应地将表达式中的信息关键字与输入图像中的重要区域对齐，并促进不同模态的特征之间的匹配。Shuo Zhang 等人[76]提出了一种半监督对比互学习（Semi-CML）分割框架，利用跨模态信息和不同模态之间的预测一致性来进行对比互学习，并进一步开发了一种软伪标签再学习（PReL）方案来弥补两种模态之间存在性能差距。

5 基于评价机制的改进

最近，由于大模型和预训练模型的流行趋势，深度学习需要的庞大计算成本所带来的问题显得比任何时候都更加重要，而预训练模型所提出的“预训练+精调”范式也引发了广泛的关注，学者们在众多领域都表现出了对通用模型的兴趣。基于研究工作的优化方向，本文提出了近两年研究中基于评价机制改进的分类：轻量化、泛化。在本文中所提到的评价机制与一般所认为的评估标准不同，这种评价机制是整体趋势性的，而非模型性的，这两种评价机制的建立基于大量最新文献贡献的总结。

5.1 轻量化

在模型训练成本和预测结果的质量之间取得平衡是图像分割任务的重要研究方向，如何在兼顾精度、实时性等性能的前提下尽量减少模型的运算量和参数也一直是深度学习的一大挑战。为

了在提升分割性能的同时不增加额外的计算量，或在同等的性能下减少成本，研究人员通常将模型的功能模块化，减少冗余计算量，如 Zhaozheng Chen 等人[77]介绍了简单但有效方法：使用 SoftMax 交叉熵损失（SCE）重新激活收敛的 CAM，称为 ReCAM。ReCAM 不仅可以生成高质量的掩膜，还可以在任何 CAM 变体中支持即插即用，减少了成本。

虽然以前的许多作品都探讨学术方法和工业需求之间的问题，但它们之间仍然存在差距：首先，现有的模型效率不够，无法使用低功耗设备；其次，当它们被用来改进现有的遮罩时，它们表现很差，因为它们不能避免破坏正确的部分。FocalClick[78]通过预测和更新本地化区域的掩码，同时解决了这两个问题。Yifan Zhang 等人[79]将图像解释为可学习区域的嵌套，每个区域都具有灵活的几何图形，并携带齐次语义，尽量避免破坏正确的图形。受基于对象中心学习的对象表示的启发，Yi Zhou 等人[80]提出了 Slot-VPS，这是第一个视频全光分割（VPS）的端到端框架，实现了对象的统一定位、分割、区分和关联。为了有效地对视频剪辑中的关键时间信息进行建模，Shusheng Yang 等人[81]提出了一种用于视频实例分割（VIS）的时间高效视觉转换器（TeViT），充分利用了帧级和实例级的时间上下文信息，获得了强大的时间建模能力，而额外的计算成本可以忽略不计。Lin Wang 等人[82]提出了一种简单而灵活的双迁移学习（DTL）框架，可以在不增加额外推理成本的情况下有效地提高终端任务的性能。Buddhadev Sasmal 等人[83]的研究展示了最近超像素技术在图像分割中起到平衡性能与成本的作用：最小化像素数量可以减少计算时间与计算复杂度。Bangze Zhang 等人[84]提出了一种用于医学图像分割的相关增强轻量级网络（CeLNet），网络采用连体结构进行权重共享和参数节省，提出了一种点深度卷积并行块（PDP 块），以降低模型参数和计算成本，同时提高编码器的特征提取能力。Yanjun Peng 等人[85]提出了与 h 网络相结合的多尺度特征（MShNet），构造了一种多尺度特征与 h-Net 算法相结合的算法 MShNet，它通过融合下采样模块和卷积金字塔模块实现多尺度特征的提取，以较少的参数有效地提高了分割性能。

5.2 泛化

图像分割将具有不同语义的像素进行分组，例如，类别或实例隶属度，每个语义的选择都定义了一个任务，虽然只有每个任务的语义不同，但目前的研究还是在为每个任务设计专门的体系结构，在图像分割领域中，泛化性始终是一项艰巨的挑战。本文同样关注研制统一、通用模型或框架的工作，并且认为这些工作的贡献将对未来提升图像分割方法的泛化性有所帮助。

Jie Qin 等人[86]提出了 FreeSeg 通用框架，用于实现统一、开放的图像分割，与单任务训练相比，FreeSeg 将训练成本成功地降低了约三分之二， 在各种设定的分割任务、训练数据集和零样本学习泛化方面都取得了最好的性能。Li Wang 等人[87]在 DTL 的基础上进一步提出了一种简单、灵活、通用的语义分割方法，称为跨数据集协作学习（CDCL），CDCL 兼具了 DTL 的轻量性，且通过利用来自所有数据集的信息来提高每个数据集的性能，从而训练出统一的模型。Fabian Isensee 等人[88]开发了一种基于深度学习的分割方法 nnU-Net，它可以自动配置自己适应不同的新任务，这些配置包括任何新任务的预处理、网络架构、训练和后处理。Bowen Cheng 等人提出了 Mask2Former，这是一种能够处理任何图像分割任务（全视、实例或语义）的新架构。Zhiqi Li 等人[42]提出了全光分割器，一个具有变压器的全光

分割的通用框架。Naiyu Gao 等人[89]提出了一个统一的深度感知全光分割（DPS）框架，目的是从一幅图像的实例级语义重建三维场景，并希望对 DPS 的统一解决方案能够在这一领域引领一种新的范式。Anirud Thyagharajan 等人[90]提出了 Segment-Fusion，一种新的基于注意力的语义和实例信息的分层融合方法，以解决部分的错误分类。Segment-Fusion 可以灵活地应用于任何网络架构，用于语义/实例分割。

6 面临的挑战与展望

从本文总结中可以看出，在近两年里图像分割研究利用深度学习解决了许多现有的挑战，也为未来的研究指引了方向，本文在此基础上进一步分析了基于深度学习的图像分割研究所面临的挑战。

6.1 对象边界

随着时间的推移，能否有效处理小物体形状和具有边界物体的分割已被证明是一个持续的挑战。具体来说，图像分割在准确检测小物体边界，以及处理分割目标与背景或其他物体共享边界的情况下，图像的分割会变得不够准确或有效，其中必须注意的是，分割目标的不同形状和姿势会导致预测错误率的增加。本文认为基于网络的对象优化方法是全局最优的，不能有效地解决复杂的对象的局部形状或边界问题，尽管研究已经为减少边界的错误预测做出了广泛的努力，但仍未能完全克服此类问题。

6.2 多域数据与泛化性

近年来，多域数据与通用网络模型的交叉问题引起了人们的关注。这些方法旨在开发能够处理各种任务的模型，例如全景模型、实例模型和语义分割模型，以及处理不同的域数据的模型，数据包括现实数据和动画数据。在多域任务中，可以将多域的数据整合到一个单一的域，比如空间、频率或特征域等，这允许研究将不同域特定的输入问题转换为单一域的输入问题，或许潜在地对整个模型有益。但多域数据整合本身也是基于数据改进的分割方法中一个不小的挑战。模型的泛化性差与域间差距大有紧密的联系，图像分割通用模型是否能有效处理多域和多模态任务，是目前继续增强泛化性需要克服的挑战。

6.3 多模态

在多模态任务中需要讨论的是，多模态的数据研究是否是一个有希望或有意义的研究方向。多模态研究本身是为探索不同模式的信息融合，最近像 ChatGPT 这样的大模型已经开始流行起来，它们对单模数据的处理非常擅长，不过 GPT3.0 或 GPT3.5 并不能够处理除文字外的多模态数据，尽管 GPT4.0 已经在多模态的道路上进行了探索，但这仍将是未来很长一段时间内的一个令人感兴趣且突出的研究方向。

6.4 可解释性

图像分割可解释性研究的目标是理解和解释深度学习模型对图像进行分割的过程和结果，这有助于提高模型的鲁棒性、可靠性和可信度，使模型能够适用于更广泛的应用，并增强用户对模型决策的信任，同时帮助我们更合理地验证和调试算法。但这也是人工智能目前面临的巨大挑战。Loan Dao 等人[91]在最新的工作中阐述了可解释人工智能（XAI）在医学图像分割中的重要性，并对可解释人工智能中的“Interpretability”（白盒）和“Explainability”（黑盒）进行了简要分析，提到图像分割可解释性的研究能够帮助专家深入获取模型中的信息，提升精准医学模型的性能。可解释性的本质是从环境或事物本身获取足够的信息去理解和掌握它，当一种现象发生时，如果我们能够解释它，那么我们通常会透过现象看到本质，并且意识到其中的一些潜在关联。在深度学习中这意味着，当我们知道模型在训练时学到了什么知识、受到了什么影响时，我们就能够更加信任模型给出的结果，并且具有对模型应

用时产生的潜在风险评估的能力和依据。尤其地，在医学领域可解释性的研究具有很高价值。对于图像分割的可解释性研究来说，可以从以下几个方面进行思考：可视化图像分割的过程、基于稀疏的方法、敏感性分析。

结束语 本文全面回顾了近两年来基于深度学习的图像分割研究的最新进展。在本文中，将基于深度学习的图像分割技术分为三大类：基于网络结构的改进、基于数据的改进和基于评价机制的改进，围绕方法的提升和解决的问题进行了综合分析。此外，本文还讨论了图像分割面临的一些挑战，并提出了一些未来的研究方向。由于深度学习技术仍处于快速发展阶段，此后很有可能开发出准确性、精确度更高，计算、标注成本更低，通用性更高的新模型，本文对近期研究的总结和探讨也是为了向相关领域的研究人员提供参考和帮助。

参考文献

[1] TAGHANAKI S A, ABHISHEK K, COHEN J P, et al. Deep Semantic Segmentation of Natural and Medical Images: A Review[J]. 2019.

[2] ALJABRI, MANAR. A review on the use of deep learning for medical images segmentation[J]. NEUROCOMPUTING, 2022,506.

[3] MOORTHY, JAYASHREE. A Survey on Medical Image Segmentation Based on Deep Learning Techniques[J]. BIG DATA AND COGNITIVE COMPUTING, 2022,6(4).

[4] BENNAI, MOHAMED. Multi-agent medical image segmentation: A survey[J]. COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE, 2023,232.

[5] LIN T, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context[M]//Cham: Springer International Publishing, 2014:740-755.

[6] CORDTS M, OMRAN M, RAMOS S, et al. The Cityscapes Dataset for Semantic Urban Scene Understanding[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 3213-3223.

[7] ZHOU B, ZHAO H, PUIG X, et al. Semantic Understanding of Scenes Through the ADE20K Dataset[J]. International Journal of Computer Vision, 2016,127: 302-321.

[8] MOTTAGHI R, CHEN X, LIU X, et al. The Role of Context for Object Detection and Semantic Segmentation in the Wild[J]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 891-898.

[9] YU F, CHEN H, WANG X, et al. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning[J]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 2633-2642.

[10] BROSTOW G J, FAUQUEUR J, CIPOLLA R. Semantic object classes in video: A high-definition ground truth database[J]. Pattern Recognit. Lett., 2009,30: 88-97.

[11] EVERINGHAM M, ESLAMI S M A, Van GOOL L, et al. The Pascal Visual Object Classes Challenge: A Retrospective[J]. International Journal of Computer Vision, 2014,111: 98-136.

[12] DAI A, CHANG A X, SAVVA M, et al. ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2432-2443.

[13] XU N, YANG L, FAN Y, et al. YouTube-VOS: Sequence-to-Sequence Video Object

Segmentation[J]. 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[14] PERAZZI F, PONT-TUSET J, MCWILLIAMS B, et al. A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 724-732.

[15] WAH C, BRANSON S, WELINDER P, et al. The Caltech-UCSD Birds-200-2011 Dataset[J]. Computation & Neural Systems Technical Report, 2011.

[16] BEHLEY J, GARBADE M, MILIOTO A, et al. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences[J]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9296-9306.

[17] GUPTA A, R P D A, GIRSHICK R B. LVIS: A Dataset for Large Vocabulary Instance Segmentation[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 5351-5359.

[18] LIANG X, GONG K, SHEN X, et al. Look into Person: Joint Body Parsing \& Pose Estimation Network and a New Benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018,41: 871-885.

[19] ROTHER C, KOLMOGOROV V, BLAKE A. "GrabCut": interactive foreground extraction using iterated graph cuts[J]. ACM SIGGRAPH 2004 Papers, 2004.

[20] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2012,25(2).

[21] ALI, KHALIFA F, EMAN. Deep Learning for Image Segmentation: A Focus on Medical Imaging[J]. 计算机、材料和连续体（英文）, 2023,75(4).

[22] LIN, HUAIJIA. Video instance segmentation with a propose-reduce paradigm[J]. arXiv, 2021.

[23] HE H, HUANG Z, DING Y, et al. CDNet: Centripetal Direction Network for Nuclear Instance Segmentation[C]// International Conference on Computer Vision: IEEE, 2021: 4006-4015.

[24] HUANG S, MA Z, MU T, et al. Supervoxel Convolution for Online 3D Semantic Segmentation[J]. ACM Transactions on Graphics, 2021,40(3): 1-15.

[25] XU G, WU X, ZHANG X, et al. LeViT-UNet: Make Faster Encoders with Transformer for Medical Image Segmentation[Z]. 2021.

[26] CAO J, LENG H, LISCHINSKI D, et al. ShapeConv: Shape-aware Convolutional Layer for Indoor RGB-D Semantic Segmentation[J]. 2021.

[27] GOODFELLOW, IAN. Generative Adversarial Networks[J]. COMMUNICATIONS OF THE ACM, 2020,63(11).

[28] HE, XINGZHE. GANSeg: Learning to Segment by Unsupervised Hierarchical Image Generation[J]. arXiv, 2021.

[29] KIM D, HONG B. Unsupervised Segmentation incorporating Shape Prior via Generative Adversarial Networks[C]// International Conference on Computer Vision: IEEE, 2021: 7304-7314.

[30] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need[J]. arXiv, 2017.

[31] STRUDEL, ROBIN. Segmenter: Transformer for semantic segmentation[J]. arXiv, 2021.

[32] XIE, ENZE. SegFormer: Simple and efficient

design for semantic segmentation with transformers[J]. arXiv, 2021.

[33] ZHAO T, ZHANG N, NING X, et al. CodedVTR: Codebook-based Sparse Voxel Transformer with Geometric Guidance[J]. arXiv, 2022.

[34] LIU H, MIAO X, MERTZ C, et al. CrackFormer: Transformer Network for Fine-Grained Crack Detection[C]// International Conference on Computer Vision: IEEE, 2021: 3763-3772.

[35] PU M, HUANG Y, LIU Y, et al. EDTER: Edge Detection with Transformer[J]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 1392-1402.

[36] de BRUIJNE M, CATTIN P C, COTIN S, et al. Multi-compound Transformer for Accurate Biomedical Image Segmentation[M]//Switzerland: Springer International Publishing AG, 2021:326-336.

[37] H. L, C. L, X. L, et al. Neural Recognition of Dashed Curves with Gestalt Law of Continuity[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 1363-1372.

[38] DING Y, YU X, YANG Y. RFNet: Region-aware Fusion Network for Incomplete Multi-modal Brain Tumor Segmentation[C]// International Conference on Computer Vision: IEEE, 2021: 3955-3964.

[39] LUO X, HU M, SONG T, et al. Semi-Supervised Medical Image Segmentation via Cross Teaching between CNN and Transformer[J]. arXiv e-prints, 2021.

[40] CHENG B, MISRA I, SCHWING A G, et al. Masked-attention Mask Transformer for Universal Image Segmentation[J]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 1280-1289.

[41] LI Z, WANG W, XIE E, et al. Panoptic SegFormer: Delving Deeper into Panoptic Segmentation with Transformers[J]. arXiv, 2021.

[42] LI, ZHIQI. Panoptic SegFormer: Delving Deeper into Panoptic Segmentation with Transformers[J]. arXiv, 2021.

[43] HUANG X, DENG Z, LI D, et al. MISSFormer: An Effective Transformer for 2D Medical Image Segmentation[J]. IEEE transactions on medical imaging, 2023,42(5): 1484-1494.

[44] PAN J, BI Q, YANG Y, et al. Label-efficient Hybrid-supervised Learning for Medical Image Segmentation[J]. 2022.

[45] MAO B, ZHANG X, WANG L, et al. Learning from the Target: Dual Prototype Network for Few Shot Semantic Segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022,36(2): 1953-1961.

[46] SEIBOLD C M, REIß S, KLEESIEK J, et al. Reference-Guided Pseudo-Label Generation for Medical Semantic Segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022,36(2): 2171-2179.

[47] SONG Y, YU L, LEI B, et al. Data Discernment for Affordable Training in Medical Image Segmentation[J]. IEEE transactions on medical imaging, 2023,42(5): 1431-1445.

[48] MONKAM P, JIN S, LU W. Annotation Cost Minimization for Ultrasound Image Segmentation using Cross-domain Transfer Learning[J]. IEEE journal of biomedical and health informatics, 2023,PP(4): 1-11.

[49] CHAITANYA K, ERDIL E, KARANI N, et al.

Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation[J]. Medical image analysis, 2023,87: 102792.

[50] GARNOT V S F, LANDRIEU L. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks[J]. 2021.

[51] FENG Z, WANG Z, WANG X, et al. Mutual-Complementing Framework for Nuclei Detection and Segmentation in Pathology Image[C]// International Conference on Computer Vision: IEEE, 2021: 4016-4025.

[52] LINGTENG QIU Z X X W, GUANYING CHEN X H S C. ETHSeg: An Amodel Instance Segmentation Network and a Real-world Dataset for X-Ray Waste Inspection[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[53] CHENG, BOWEN. Pointly-Supervised Instance Segmentation[J]. arXiv, 2021.

[54] MARIN D, BOYKOV Y. Robust Trust Region for Weakly Supervised Segmentation[J]. 2021.

[55] WANG Y, WANG H, SHEN Y, et al. Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels[J]. arXiv e-prints, 2022.

[56] ZHUANG J, WANG Z, GAO Y. Semi-Supervised Video Semantic Segmentation With Inter-Frame Feature Reconstruction, 2022.June.

[57] LEI T, ZHANG D, DU X, et al. Semi-Supervised Medical Image Segmentation Using Adversarial Consistency Learning and Dynamic Convolution Network[J]. IEEE transactions on medical imaging, 2023,42(5): 1265-1277.

[58] ZHANG C, PAN T, LI Y, et al. MosaicOS: A Simple and Effective Use of Object-Centric Images for Long-Tailed Object Detection[J]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 407-417.

[59] XU Q, YAO L, JIANG Z, et al. DIRL: Domain-Invariant Representation Learning for Generalizable Semantic Segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022,36(3): 2884-2892.

[60] YAO H, HU X, LI X. Enhancing Pseudo Label Quality for Semi-SupervisedDomain-Generalized Medical Image Segmentation[J]. arXiv e-prints, 2022.

[61] WU X, WU Z, LU Y, et al. Style Mixing and Patchwise Prototypical Matching for One-Shot Unsupervised Domain Adaptive Semantic Segmentation[J]. 2021.

[62] WOOD, ERROLL. Fake it till you make it: Face analysis in the wild using synthetic data alone[J]. arXiv, 2021.

[63] TANG H, LIU X, SUN S, et al. Recurrent Mask Refinement for Few-Shot Medical Image Segmentation[J]. 2021.

[64] XU Q. A Fourier-based framework for domain generalization[J]. arXiv, 2021.

[65] ZHAO Z, ZHOU F, XU K, et al. LE-UDA: Label-Efficient Unsupervised Domain Adaptation for Medical Image Segmentation[J]. IEEE transactions on medical imaging, 2023,42(3): 633-646.

[66] WANG J, ZHONG C, FENG C, et al. Disentangled Representation for Cross-Domain Medical Image Segmentation[J]. IEEE transactions on instrumentation and measurement, 2023,72: 1.

[67] WANG H, CHU H, FU S, et al. Renovate Yourself: Calibrating Feature Representation of Misclassified Pixels for Semantic

Segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022,36(3): 2450-2458.

[68] SU Y, SUN R, LIN G, et al. Context Decoupling Augmentation for Weakly Supervised Semantic Segmentation[J]. 2021.

[69] RAJU A, MIAO S, CHENG C T, et al. Deep Implicit Statistical Shape Models for 3D Medical Image Delineation[J]. 2021.

[70] DING Y, YU X, YANG Y. RFNet: Region-aware Fusion Network for Incomplete Multi-modal Brain Tumor Segmentation[C]// International Conference on Computer Vision: IEEE, 2021: 3955-3964.

[71] TORR S S X Y, INTELLIGENCE U O O C. Aggregation with Feature Detection[J]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021.

[72] S. B, H. P, H. C, et al. Panoptic, Instance and Semantic Relations: A Relational Context Encoder to Enhance Panoptic Segmentation[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 1259-1269.

[73] WENWEN PAN H S Z Z, JUN YU F W Q T. Wnet: Audio-Guided Video Object Segmentation via Wavelet-Based Cross-Modal Denoising Networks[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[74] ZHENG S, TAN J, JIANG C, et al. Automated multi-modal Transformer network (AMTNet) for 3D medical images segmentation[J]. Physics in medicine & biology, 2023,68(2): 25014.

[75] ZHANG W, HU M, TAN Q, et al. Cross-modal attention guided visual reasoning for referring image segmentation[J]. Multimedia tools and applications, 2023,82(19): 28853-28872.

[76] ZHANG S, ZHANG J, TIAN B, et al. Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation[J]. Medical image analysis, 2023,83: 102656.

[77] CHEN, ZHAOZHENG. Class Re-Activation Maps for Weakly-Supervised Semantic Segmentation[J]. arXiv, 2022.

[78] CHEN, XI. FocalClick: Towards Practical Interactive Image Segmentation[J]. arXiv, 2022.

[79] ZHANG Y, PANG B, LU C. Semantic Segmentation by Early Region Proxy[J]. 2022.

[80] ZHOU Y, ZHANG H, LEE H, et al. Slot-VPS: Object-centric Representation Learning for Video Panoptic Segmentation[J]. arXiv e-prints, 2021.

[81] YANG, SHUSHENG. Temporally Efficient Vision Transformer for Video Instance Segmentation[J]. arXiv, 2022.

[82] WANG L, CHAE Y, YOON K J. Dual Transfer Learning for Event-based End-task Prediction via Pluggable Event to Image Translation, 2021.

[83] SASMAL B, DHAL K G. A survey on the utilization of Superpixel image for clustering based image segmentation[J]. Multimedia tools and applications, 2023: 1-63.

[84] ZHANG B, WANG X, LIU L, et al. CeLNet: a correlation-enhanced lightweight network for medical image segmentation[J]. Physics in medicine & biology, 2023,68(11).

[85] PENG Y, YU D, GUO Y. MShNet: Multi-scale feature combined with h-network for medical image segmentation[J]. Biomedical signal processing and control, 2023,79(Part2): 104167.

[86] QIN J, WU J, YAN P, et al. FreeSeg: Unified, Universal and Open-Vocabulary Image Segmentation[J]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 19446-19455.

[87] WANG L, LI D, ZHU Y, et al. Cross-Dataset Collaborative Learning for Semantic Segmentation[J]. 2021.

[88] ISENSEE F, JAEGER P F, KOHL S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature Methods, 2020.

[89] GAO, NAIYU. PanopticDepth: A Unified Framework for Depth-aware Panoptic Segmentation[J]. arXiv, 2022.

[90] ANIRUD THYAGHARAJAN B U P L. Segment-Fusion: Hierarchical Context Fusion for Robust 3D Semantic Segmentation[J]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[91] DAO L, LY N Q. A Comprehensive Study on Medical Image Segmentation using Deep Neural Networks[J]. International journal of advanced computer science & applications, 2023,14(3).

黄雯珂，出生于2000年，硕士研究生，计算机学会（CCF）学生会员，主要研究领域为联邦学习、图像分割；滕飞(通讯作者)，出生于1984年，博士，副教授，计算机学会(CCF)会员（提供会员号），主要研究领域为医学信息学、云计算、医疗大数据分析；王子丹，出生于1999年，硕士研究生，主要研究领域为知识推理；冯力，出生于1974年，博士，研究员，主要研究领域为军事智能计算。

HUANG Wenke, born in 2000, postgraduate, is a member of CCF. Her main research interests include , image segmentation and so on.

TENG Fei, born in 1984, professor, is a member of CCF. Her main research interests include Medical Informatics, cloud computing, medical big data analyse and so on.