

基于深度学习的计算机视觉图像分割算法研究

张利平 廖亚凤

邯郸科技职业学院 河北邯郸 056046

摘要：图像分割是计算机视觉领域的核心任务之一，旨在将图像像素划分为具有语义意义的区域，为场景理解、目标识别等高层任务提供关键支撑。随着深度学习技术的快速发展，基于深度神经网络的图像分割算法已逐步取代传统方法，成为研究热点。本文系统梳理了深度学习驱动图像分割技术演进历程，重点分析了全卷积网络（FCN）、U-Net、Mask R-CNN 等经典算法的网络结构与创新机制，并探讨了 Transformer、注意力机制、多尺度特征融合等前沿技术在分割任务中的应用。通过对比不同算法在医学影像、自动驾驶、遥感图像等领域的性能表现，本文总结了当前研究面临的小目标分割精度不足、实时性与模型复杂度矛盾、标注数据依赖等挑战，并展望了弱监督学习、多模态融合、轻量化模型设计等未来发展方向。研究表明，深度学习模型凭借端到端学习能力和特征表达优势，已显著提升图像分割的精度与泛化性，但其在复杂场景适应性和实际部署效率上仍需进一步突破。

关键词：深度学习；图像分割；计算机视觉；卷积神经网络；语义分割

引言

图像分割是将图像分解成多个具特定语义信息像素集合的技术，在连接底层视觉特征和高层语义理解方面起关键桥梁作用。传统图像分割方法，如阈值分割、边缘检测和区域生长等，依赖人工设计特征，面对光照变化、纹理多样性和目标形态差异等复杂场景时表现不佳。2012 年 AlexNet 在图像分类任务的突破，推动了深度学习技术在计算机视觉领域的广泛应用。基于深度神经网络的图像分割算法，凭借自动学习特征的能力，实现了从“像素级分类”到“端到端分割”的范式转变。近年来，随着深度学习发展，图像分割领域出现语义分割、实例分割和全景分割等细分任务，在医学影像分析、自动驾驶、机器人导航等领域有重要应用价值。本文将重点探讨深度学习在图像分割中的核心算法及其研究进展，为相关领域研究与应用提供参考，通过深入分析可更好理解其作用及解决实际问题的方法。

1 深度学习图像分割核心算法

（1）基于卷积神经网络（CNN）的经典算法：FCN：首次将 CNN 的应用领域从图像分类扩展到像素级的图像分割任务。这一创新的核心在于将传统的全连接层替换为卷积层，这一变革使得 FCN 能够处理任意尺寸的图像输入，极大地提高了算法的灵活性和通用性。更重要的是，这种替换使得网络能够在特征图上进行像素级的分类，从而实现精细

的图像分割。为了进一步提升分割效果，FCN 引入了上采样技术，通过反卷积或双线性插值等方法将特征图恢复到原始图像大小，同时结合跳跃连接，有效地融合了高低层特征，提高了分割的准确性和边界的清晰度。U-Net：提出了“编码器-解码器”对称结构的创新性算法，专门针对医学影像分割任务设计。编码器部分通过卷积层和池化层逐步提取图像的上下文特征，形成特征金字塔。解码器部分则通过上采样和卷积层逐步恢复图像的空间细节，同时引入跳跃连接，将编码器中相应层级的特征图与解码器中的特征图进行拼接，解决了梯度消失问题，实现了高低层特征的有效融合。这种结构在处理小样本数据集时表现出色，能够捕捉到更多的细节信息，因此成为了医学图像分割领域的基准模型。Mask R-CNN：结合了 Faster R-CNN 目标检测框架与 FCN 分割分支，提出了端到端的实例分割训练方法。Mask R-CNN 在 Faster R-CNN 的基础上增加了一个用于分割的分支，通过 RoIAlign 技术解决了特征对齐问题，使得网络能够在目标检测框内进行精确的像素级分割。这种方法不仅能够兼顾目标检测的精度和速度，还能够保证分割细节的准确性，进一步推动了图像分割技术的发展。同时，Mask R-CNN 的端到端训练方式简化了训练过程，提高了算法的效率和稳定性。

（2）新兴技术与模型改进：在新兴技术中，

Transformer 与 CNN 的结合方式呈现出多样化的趋势。以 SegFormer 为例,其采用了“混合结构”,即结合了 Transformer 的自注意力机制和 CNN 的局部特征提取能力。具体而言,SegFormer 通过 Transformer 模块捕捉图像中的长距离依赖关系,而 CNN 模块则负责提取图像的局部特征。这种结合方式使得 SegFormer 在复杂的场景中能够更好地理解图像中的细节和上下文信息,从而提升图像分割的精度。在注意力机制方面,CBAM 和 SE 等模块在多尺度特征中的应用也愈发广泛。以 CBAM 为例,它通过结合通道注意力和空间注意力,优化特征图的通道和空间权重,从而增强关键区域特征的表达能力。在多尺度特征中,CBAM 能够针对不同尺度的特征图进行自适应的调整,使得模型在处理不同尺度的目标时更加高效和准确。SE 模块则主要通过挤压和激励操作,重新校准特征图的通道权重,使得模型能够更加关注于重要的特征通道。一个简要的案例说明是,在自动驾驶场景中,车辆和行人的检测与分割是至关重要的。通过引入 Transformer 架构和 CBAM 等注意力模块,模型能够更好地捕捉车辆和行人的长距离依赖关系和关键特征,从而在复杂的交通场景中实现更加准确的检测与分割。这不仅提高了自动驾驶系统的安全性,还为其在实际应用中的推广提供了有力的技术支持。此外,DeepLab 系列模型也通过引入空洞卷积技术和 ASP 模块,有效地融合了多尺度的上下文信息,提高了图像分割的精度。而 HRNet 则通过保持高分辨率特征图贯穿整个网络,缓解了小目标分割精度损失的问题。这些技术的引入和发展,进一步推动了图像分割领域的进步和发展。针对标注成本高昂的问题,弱监督分割和半监督分割方法也成为了研究的热点。这些方法通过利用少量的标注数据和大量的无标注数据,结合自训练策略和伪标签生成等技术,降低了标注成本,提高了模型的泛化能力和分割精度。这不仅为图像分割领域的研究提供了新的思路和方法,也为其在实际应用中的推广提供了更加广阔的空间。

(3) 应用场景与性能评估:在深度学习领域,分割算法的应用场景非常广泛,并且在这些领域中已经得到了充分的验证和认可。特别是在医学影像领域,U-Net 及其各种变体算法被广泛应用于 CT 和 MRI 图像的器官和肿瘤分割任务中,极大地提高了临床诊断的准确性和手术规划的效率。例如,U-Net 算法在医学影像分割任务中表现出色,能够准确地识别出图像中的关键结构,从而为医生提供更为精确的

诊断依据。在自动驾驶领域,深度学习分割算法同样发挥着重要作用。通过在 Cityscapes 和 KITTI 等权威数据集上的训练和测试,DeepLab、ENet 等模型能够实时地对道路、车辆和行人进行精确分割,从而为自动驾驶汽车的环境感知系统提供了强有力的支持。这些模型能够快速识别和理解周围环境,确保自动驾驶汽车在复杂交通场景中的安全行驶。值得注意的是,Cityscapes 的复杂街景和 KITTI 等数据集的特点对算法的泛化性能提出了考验。在遥感图像处理方面,深度学习分割算法同样展现出了巨大的潜力。高分辨率卫星影像的土地覆盖分类任务中,SegNet 等算法能够有效地提取出不同类型的地理信息,如建筑物、植被、水域等,为城市规划和资源管理提供了宝贵的数据支持。PSPNet 等模型在建筑物提取任务中表现出色,能够准确地识别出城市中的建筑物分布,为城市规划和建设提供了重要的参考依据。为了评估这些深度学习分割算法的性能,研究人员通常会使用一些常用的指标,如平均交并比(mIoU)、像素准确率(PA)和帧率(FPS)等。例如,在 COCO 数据集的实例分割任务中,Mask R-CNN 算法的 mIoU 达到了 37.1%,显示出其在分割精度上的优势。而 COCO 数据集的多样物体对模型的泛化性同样提出了挑战。在 Cityscapes 数据集的语义分割任务中,SegFormer 算法的 mIoU 更是突破了 84%,展现了其在高精度分割任务中的卓越性能。Cityscapes 的复杂街景对模型的分割精度和泛化性提出了更高要求。然而,在实际应用中,实时性模型如 YOLOv8-seg 则需要在分割精度和处理速度之间进行权衡,以满足不同应用场景的需求。针对实时性与模型复杂度的矛盾,轻量化技术如模型剪枝和知识蒸馏等方法被广泛应用以优化模型的性能和效率。模型剪枝是一种通过移除神经网络中不重要的权重和神经元来减少模型复杂度的技术。通过剪枝,我们可以显著减少模型的参数数量和计算量,从而提高模型的推理速度并降低内存占用。同时,剪枝后的模型仍然能够保持较高的分割精度,从而满足实际应用的需求。知识蒸馏则是一种通过将一个复杂的大模型(教师模型)的知识转移到一个简单的小模型(学生模型)中的技术。通过知识蒸馏,我们可以利用大模型的强大性能来指导小模型的训练,从而使小模型在保持较低复杂度的同时获得较高的分割精度。这种方法不仅可以提高模型的实时性能,还可以降低模型的部署成本。

2 挑战与未来展望

(1) 当前的研究领域面的挑战: 这些挑战在很大程度上影响了研究的进展和应用的广泛性。首先, 当涉及到小目标检测和处理遮挡问题时, 研究者们发现, 在一些密集的场景中, 例如人群密集的场所或细胞显微图像中, 小目标的特征很容易被忽略, 这使得目标检测变得异常困难。此外, 遮挡现象的存在进一步加剧了这一问题, 因为遮挡会导致目标的边界变得模糊不清, 从而使得目标检测和识别变得更加复杂。其次, 在实时性和轻量化需求方面, 由于移动设备和嵌入式设备对模型的参数量和计算量有着严格的限制, 研究者们面临着巨大的挑战。他们需要在保证模型性能的同时, 尽可能地减少模型的复杂度, 以满足实时处理的需求。这意味着研究者们必须在模型的精度和效率之间找到一个平衡点, 这是一个非常具有挑战性的任务。最后, 在跨域泛化能力方面, 现有的模型在特定数据集上经过训练后, 往往难以适应具有不同光照条件、天气变化或传感器差异的新场景。这限制了模型在实际应用中的广泛应用, 因为现实世界中的场景千变万化, 模型需要具备足够的灵活性和适应性才能应对各种不同的环境。这些问题的存在, 使得研究者们需要进一步探索和改进, 以提高模型的鲁棒性和适用性, 从而使其能够在各种复杂多变的环境中都能表现出色。

(2) 未来的发展方向: 未来将会集中在几个关键的研究领域, 其中自监督学习和无监督学习将成为其中的重要组成部分。这些学习方法的优势在于它们能够减少我们对大量标注数据的依赖, 从而在一定程度上解决数据标注成本高和标注不准确的问题。通过图像重构、对比学习等技术手段, 这些方法能够有效地学习到图像分割的特征, 从而提高模型的性能和准确性。此外, 多模态融合技术也将得到进一步的发展和应用。这种技术通过结合 RGB 图像、红外图像、深度信息等多种数据源, 能够显著提升模型在复杂环境下的鲁棒性和适应性。例如, 在自动驾驶领域, 多模态融合技术可以结合视觉、雷达和激光雷达等多种传感器数据, 从而提高系统的安全性和可靠性。最后, 可解释性研究将成为一个重要课题。通过可视化技术, 我们可以深入解析模型的决策过程, 从而增强模型在医疗、安防等关键领域的可靠性和可信度。例如, 在医疗领域, 通过可解释的 AI 模型, 医生可以更好地理解模型的诊断依据, 从而提高对模型的信任度。这些研究方向不仅将为未来的技术进步提供坚实的基础, 还将

推动 AI 技术在各个领域的广泛应用和拓展。

3 结论

近年来, 深度学习技术推动图像分割算法发展, 从全卷积网络 (FCN) 到 Transformer 模型, 图像分割从单一语义分割扩展到全景分割, 模型能力不断突破。深度学习使图像分割算法能更精准识别分类图像区域, 实现精细分割。FCN 改造传统卷积神经网络, 能处理任意大小输入图像, 提高了灵活性和实用性。Transformer 模型自注意力机制可捕捉长距离依赖, 用于图像分割提升了精度和鲁棒性。随着技术进步, 图像分割任务复杂度增加, 全景分割要求识别分割独立对象, 为计算机视觉应用提供更多可能。然而, 实际应用中研究者面临挑战, 高精度伴随高计算成本, 需在精度、计算效率和泛化能力间找平衡, 如自动驾驶领域对实时性要求高。未来研究应结合理论创新与实际需求, 不仅要在算法层面创新, 还要优化硬件和系统, 推动图像分割技术向智能化、可靠化发展, 为计算机视觉产业化提供支撑, 促进行业快速发展。

参考文献:

- [1] 蒋弘凡. 基于深度学习的计算机视觉图像分割的研究 [D]. 江南大学, 2024. DOI:10.27169/d.cnki.gwqgu.2024.000652.
- [2] 徐从安, 李健伟, 董云龙, 等. 深度学习时代的计算机视觉算法 [M]. 人民邮电出版社: 202201:201.
- [3] 牟韵洁. 基于深度学习图像分割算法研究 [D]. 南京林业大学, 2024. DOI:10.27242/d.cnki.gnjlu.2024.000187.
- [4] 何雪东. 基于深度学习的生活场景图像的语义分割算法研究 [D]. 广西民族大学, 2023. DOI:10.27035/d.cnki.ggxmc.2023.000102.
- [5] 刘建鑫. 基于深度学习的图像语义分割算法及应用研究 [D]. 齐鲁工业大学, 2021. DOI:10.27278/d.cnki.gsdqc.2021.000477.

作者简介: 张利平 (1968–03), 男, 汉族, 河北省邯郸市临漳县, 职称: 高级讲师, 学历: 研究生, 研究方向: 计算机应用。

基金项目: 邯郸市科学技术研究与发展计划项目 (23422901135); 邯郸科技职业学院科学技术研究与发展计划项目 (KJ202201);