多尺度特征融合在图像语义分割中的应用

陈金双 曹会念
重庆工商职业学院 重庆 401520

摘 要 图像语义分割作为计算机视觉的核心任务，需同时实现像素级分类精度与全局语义一致性。多尺度特征融合技术通过整合不同层级的视觉特征，有效解决了单一尺度特征难以兼顾细节信息与语义理解的瓶颈。本文系统梳理多尺度特征融合的核心原理与技术演进，分析金字塔池化、空洞卷积、注意力机制等主流融合策略的设计逻辑，探讨典型网络架构的特征融合机制，并结合实际应用场景总结技术优势。研究表明，多尺度特征融合通过互补局部细节与全局上下文信息，显著提升了复杂场景下语义分割的鲁棒性与准确性，为后续技术优化提供了理论参考与实践方向。

关键词 图像语义分割；多尺度特征融合；卷积神经网络；注意力机制

Multi-Scale Feature Fusion for Image Semantic Segmentation

Chen Jin-shuang, Cao Hui-nian
Chongqing Technology and Business Institute, Chongqing 401520, China

Abstract Image semantic segmentation, a core task in computer vision, demands both pixel-level classification accuracy and global semantic consistency. Multi-scale feature fusion integrates visual cues from different layers, overcoming the bottleneck of single-scale representations that struggle to balance fine details with high-level semantics. This paper systematically reviews the underlying principles and technical evolution of multi-scale feature fusion, analyzes the design rationale of mainstream strategies such as pyramid pooling, atrous convolution, and attention mechanisms, examines the fusion schemes of representative network architectures, and summarizes their advantages in real-world applications. Evidence shows that complementary local details and global context provided by multi-scale fusion significantly improve the robustness and accuracy of semantic segmentation in complex scenes, offering theoretical guidance and practical directions for future optimization.

Key words Image Semantic Segmentation; Multi-Scale Feature Fusion; Convolutional Neural Networks; Attention Mechanism

图像语义分割是自动驾驶、遥感监测、医疗影像分析等领域的关键技术，核心是将图像像素划分为特定语义类别区域。现实场景中目标尺度差异、语义模糊等问题，使传统分割方法难以兼顾细节恢复与语义理解。多尺度特征融合技术通过提取并整合不同层级特征，弥补了单一尺度特征的局限，成为提升分割性能的核心手段。近年来，基于卷积神经网络的多尺度融合架构持续创新，从编解码结构到注意力增强模型，不断提升融合的高效性与精准性。本文剖析该技术的原理、架构及应用，为相关研究提供系统性参考。

1 多尺度特征融合的核心原理与技术基础

1.1 多尺度问题的本质与上下文信息的作用

语义分割的多尺度问题源于目标物体尺寸差异与空间分布复杂性，如城市场景中近距车辆、行人与远端路标、信号灯的尺度反差，使单一尺度特征提取网络难以精准识别不同尺度目标。该问题核心是低、高层特征信息不对称：低层特征空间分辨率高，含边缘、纹理等细节但语义表达弱；高层特征经卷积、池化后语义信息丰富，却丢失大量空间细节。

上下文信息是多尺度特征融合的关键，其作用类似自然语言处理的语境分析，通过像素周边语义背景约束减少分类歧义，如孤立像素块结合水面、天空等场景特征可实现准确分类。有效利用上下文信息，既提升像素分类的准确性，又保障分割结果全局语义一致性，是解决多尺度问题的核心思路[1]。

1.2 多尺度特征融合的技术框架

多尺度特征融合的核心框架可概括为“特征提取—特征转换—特征融合—特征解码”4个阶段。特征提取阶段通过卷积神经网络（如ResNet、VGG等骨干网络）获取不同层级的特征

图，形成多尺度特征集合；特征转换阶段通过1×1卷积、归一化等操作调整特征通道数与维度，降低计算复杂度并增强特征表达；特征融合阶段采用特定策略整合多尺度特征，实现细节信息与语义信息的互补；特征解码阶段通过上采样、反卷积等操作恢复特征图分辨率，输出最终分割结果。

这一框架的关键在于特征融合阶段的策略设计，需解决不同尺度特征的语义间隙、空间对齐等问题。优秀的融合策略应能自适应权衡各尺度特征的权重，突出关键信息，抑制冗余噪声，为后续分割任务提供更具判别力的特征表示[2]。

2 多尺度特征融合的典型策略与架构

2.1 基于池化操作的融合策略

池化操作可在不同尺度聚合特征图，实现全局与局部上下文信息融合，金字塔池化技术是典型代表。PSPNet的金字塔池化模块通过1×1、2×2、3×3、6×6这四种尺度池化处理高层特征图，将多尺度特征向量与原始特征图拼接融合，既能借助全局池化获取图像级语义特征，又保留局部细节，提升了模型对多尺度目标的适应性。ParseNet则通过全局平均池化提取全局特征，经L2规范化后与局部特征融合，解决了传统FCN有效感受野不足的问题，以简洁高效的操作让局部特征受全局语义指导，降低分类歧义。这类池化融合策略结构简洁、计算开销低，适配实时性要求高的场景。

2.2 基于空洞卷积的融合策略

空洞卷积（Atrous Convolution）通过在卷积核中引入空洞率（扩张率），在不增加参数与计算量的前提下扩大感受野，成为多尺度特征提取的核心技术之一。Deeplab系列模型（如Deeplabv3、Deeplabv3+）将空洞卷积与空间金字塔池化（ASPP）结合，通过不同扩张率的空洞卷积并行提取多尺度特征，再将特征图融合输出。不同扩张率的空洞卷积对应不同的感受野，能够自适应捕获大、中、小尺度目标的特征信息，同时避免了下采样导致的细节丢失[3]。

空洞卷积的核心优势在于实现了感受野的灵活调整，无须额外的池化操作即可获取多尺度特征，保持了特征图的空间分辨率。在道路坑洞检测、遥感图像分割等场景中，该策略能够有效区分不同尺度的目标区域，平衡分割精度与计算效率。

2.3 基于注意力机制的融合策略

注意力机制通过对多尺度特征赋予自适应权重，突出关键特征、抑制无关信息，成为近年来多尺度融合的研究热点。UAVSeg提出的跨尺度轻量级集成（CSLI）模块，基于轻量级点积注意力机制，对双编码器提取的多尺度特征进行动态融合，在减少计算参数的同时提升了特征融合的精准度。该模块通过计算不同尺度特征间的相关性，自适应调整各特征的权重分配，有效缩小了不同层级特征的语义差距。

通道注意力模块（CAM）与空间注意力模块的结合，进一步提升了融合效果。SE-Net通过“挤压—激励”操作对通道特征重赋权，增强有效特征通道的表达；ENCNet则通过上下文编码模块学习全局共享的权重向量，优化特征融合过程。注意力机制的引入，使模型能够根据具体场景自适应地聚焦重要尺度的特征，显著提升了分割结果的细节精度与语义一致性。

2.4 基于高分辨率特征维持的融合策略

传统编解码结构在下采样过程中易丢失高分辨率细节，HRNet（高分辨率网络）提出了全新的融合思路：在网络整个训练过程中保持高分辨率特征图，并与低分辨率特征图进行并行融合。该架构通过多分支结构生成不同分辨率的特征图，各分支间通过反复的特征交换与融合，实现了高分辨率细节与低分辨率语义信息的持续互补。

与传统编解码结构相比，HRNet无须复杂的上采样操作即可保留丰富的细节特征，同时通过跨分辨率融合获得充足的语义信息。这种融合策略在人体姿态估计、遥感图像分割等对细节要求较高的任务中表现突出，为多尺度特征融合提供了新的技术路径[4]。

3 多尺度特征融合的应用场景与优势

3.1 典型应用场景

多尺度特征融合技术凭借其强大的多尺度目标适应能力，已广泛应用于多个领域，具体场景及技术适配性如表1所示。

表1 多尺度特征融合的典型应用场景与技术适配

应用领域	核心需求	适配融合策略	典型架构
自动驾驶	兼顾车辆、行人、路标等多尺度目标	空洞卷积+注意力融合	Deeplabv3+
遥感图像分割	地物尺度差异大、全局一致性要求高	金字塔池化+高分辨率融合	PSPNet、HRNet
医疗影像分析	病灶细节精准提取、语义无歧义	注意力机制+编解码融合	U-Net、UAVSeg
道路坑洞检测	坑洞尺度不一、边缘模糊	空洞卷积+通道注意力	MSFFM
城市场景解析	建筑、行人、车辆多尺度共存	金字塔池化+全局特征融合	ParseNet、PSPNet

在自动驾驶场景中，车辆需同时识别近距离的行人和远距离的交通标志，Deeplabv3+的空洞卷积与ASPP模块能够有效捕获不同尺度目标的特征，保障分割精度；遥感图像中，建筑物、道路、植被等目标尺度差异极大，PSPNet的金字塔池化技术可实现全局与局部特征的有效融合，提升地物分类的准确性[5]；医疗影像分析中，病灶区域的精准分割依赖细节信息与语义约束，注意力机制与编解码融合的结合能够减少误诊率。

3.2 技术优势分析

多尺度特征融合技术相较于传统单一尺度分割方法，具有三大核心优势：一是提升多尺度目标适应性，通过整合不同层级特征，使模型能够同时精准识别大尺度物体与小尺度细节，适应复杂场景的尺度差异；二是增强语义一致性，上下文信息的有效利用减少了像素分类的歧义性，保障了分割结果的全局语义连贯；三是平衡精度与效率，通过特征转换与注意力机制的优化，在提升分割精度的同时控制计算复杂度，以满足实际应用的实时性需求。

这些优势使得多尺度特征融合成为语义分割领域的核心技术，推动了分割模型在各行业的落地应用，解决了传统方法难以应对的复杂场景分割问题[6]。

4 多尺度特征融合的挑战与展望

多尺度特征融合技术虽已取得显著进展，但仍面临三大核心挑战：不同尺度特征的语义表达差异大，直接融合易引发特征冲突，语义间隙的缩小仍是关键；特征提取与融合需处理大量数据，导致模型参数多、推理速度慢，难以适配嵌入式设备等实时场景；现有固定融合策略缺乏动态尺度适应能力，对极端尺度目标的分割效果欠佳。

未来该技术将围绕“高效、自适应、轻量化”发展：结合Transformer全局建模能力设计动态权重分配机制，实现特征融合比例的自适应调整；通过深度可分离卷积等技术轻量化架构，减少参数与计算量；拓展至RGB与红外、雷达等跨模态数据融合，提升复杂环境下分割的鲁棒性[7]。

5 结束语

多尺度特征融合技术通过整合不同层级的视觉特征与上下文信息，有效解决了图像语义分割中的尺度差异与语义歧义问题，成为提升分割性能的核心支撑。本文系统分析了多尺度特征融合的原理、典型策略与应用场景，总结了金字塔池化、空洞卷积、注意力机制等融合方法的技术特点与优势。尽管当前技术仍面临语义间隙、计算复杂度等挑战，但随着轻量化架构、自适应策略等方向的研究推进，多尺度特征融合将在自动驾驶、遥感监测、医疗影像等领域发挥更大价值。未来，通过跨学科技术的融合创新，有望实现更高效、精准、通用的多尺度特征融合方案，推动图像语义分割技术迈向新高度。

参考文献

[1]郭琪周,袁春.基于空间语义信息特征融合的目标检测与分割[J].软件学报,2023,34(6):2776-2788.

[2]殷梓.多尺度特征融合技术在弱信息图像分割的运用[J].价值工程,2023,42(28):138-140.

[3]李敏,李学俊,廖竞.融合多层次卷积神经网络的知识图谱嵌入模型[J].计算机工程与应用,2025,61(6):192-198.

[4]张海鹏,段勇.基于注意力与多尺度特征融合的三维点云语义分割[J].传感技术学报,2025(7):76-82.

[5]易清明,张文婷,石敏,等.多尺度特征融合的道路场景语义分割[J].激光与光电子学进展,2023,60(12):82-90.

[6]孙国豪.基于多尺度信息融合的遥感图像小目标分割[D].北京:北京邮电大学,2023.

[7]郭丹青,符颖,朱烨,等.自注意力多尺度特征融合的遥感图像语义分割算法[J].计算机辅助设计与图形学学报,2023,35(8):1259-1268.

（本文编辑　曹　阳）

