

基于MFS的图像分割边界优化策略

王宇帅, 雷敏*

太原理工大学数学学院, 山西 太原

收稿日期: 2026年3月13日; 录用日期: 2026年4月6日; 发布日期: 2026年4月14日

摘要

基础视觉大模型(例如SAM)极大地推动了自然场景下的计算机视觉任务, 然而其分割掩码的锯齿状边缘缺陷, 影响了分割的质量。传统边界细化技术通常通过在原网络架构中附加参与训练的额外分支来解决此问题, 不可避免地带来了庞大的训练开销。针对这一痛点, 本文提出了一种专门适配SAM模型的低成本边界细化模块, 该模块基于基本解隐式方法构建。具体而言, 我们对SAM的初始分割边界点进行重采样, 并执行隐式重构, 从而在不改变目标原始拓扑形状的情况下显著平滑边界。基于COCO数据集的评估结果证实, 该优化策略在保留SAM分割精度的同时, 使曲率指标提升了接近十倍。该方法为自然图像的高级视觉处理提供了一种高效、稳健且无需昂贵训练成本的边界优化方案。

关键词

边界细化, 图像分割, 零训练成本, 隐式函数曲线重构

Boundary Optimization Strategy for Image Segmentation Based on MFS

Yushuai Wang, Min Lei*

Department of Mathematics, Taiyuan University of Technology, Taiyuan Shanxi

Received: March 13, 2026; accepted: April 6, 2026; published: April 14, 2026

Abstract

SAM has significantly advanced computer vision tasks in natural scenes. However, the jagged edges of their segmentation masks often compromise the overall quality. Traditional boundary refinement techniques typically address this by appending additional trainable branches to the original architecture, which inevitably incurs substantial training overhead. To address this pain point, this paper proposes a low-cost boundary refinement module specifically tailored for SAM,

*通讯作者。

built upon the implicit method of fundamental solutions. Specifically, we resample the initial boundary points from SAM's segmentation and perform global implicit reconstruction, significantly smoothing the boundaries while preserving the object's original topology. Evaluation results on the COCO dataset demonstrate that this optimization strategy improves curvature metrics nearly tenfold while maintaining SAM's full segmentation accuracy. This approach provides an efficient, robust, and training-free boundary optimization solution for advanced visual processing in natural imagery.

Keywords

Boundary Refinement, Image Segmentation, Training-Free, Implicit Function Curve Reconstruction

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在计算机视觉领域, 图像分割(Image Segmentation)作为实现像素级场景理解的核心任务, 在自动驾驶、医学影像诊断及工业自动化检测等高精度需求场景中发挥着至关重要的作用。自深度学习介入计算机视觉领域以来, 图像分割技术经历了从启发式架构到通用基础模型的跨越式发展。U-Net [1]通过提出对称的编码器-解码器结构与跳跃连接, 有效地解决了下采样导致的细节丢失问题, 成为了影像分割领域经典的基准架构。SegNet [2]则通过记录池化索引进行非线性上采样, 在保持较高分割精度的同时显著降低了显存开销, 开启了对轻量化分割架构的早期探索。

为了应对复杂场景中的尺度变化, DeepLab 系列[3]引入了空洞卷积与空间金字塔池化(ASPP)机制扩大感受野, 极大地提升了模型对大尺寸目标的捕获能力。在实例级别, Mask R-CNN [4]在检测框架中并行引入掩码分支并利用 RoIAlign 来改进 Faster RCNN [5]里 RoIPooling 中以降低量化误差, 进一步提高了对个体目标的精准提取, 提高了实例分割任务的精度。在 HRNet [6]提出之前, 绝大多数网络都采用先下采样获取强语义信息、再上采样恢复高分辨率位置信息的范式, 但这种反复上下采样的方式会造成大量有效信息丢失。HRNet 则通过并行构建多个不同分辨率的特征分支, 并在分支间持续进行信息交互与融合, 能够同时保留丰富的语义信息与精准的位置信息, 有效解决了传统结构在分辨率恢复过程中信息损失的问题。

步入 Transformer [7]时代, SegFormer [8]采用全新的分层 Transformer 编码器, 能够输出多尺度特征, 且不依赖位置编码, 有效避免了因测试与训练分辨率不一致、位置编码插值带来的性能下降问题; 舍弃传统复杂解码器结构, 设计了简洁高效的 MLP 解码器, 可融合不同层级特征, 同时兼顾局部与全局注意力, 从而得到表达能力更强的特征表示。Mask2Former [9]在 Transformer 解码器中采用以预测片段(根据分组语义可对应物体或区域)为中心的局部特征掩蔽注意力, 将注意力限制在局部范围内, 而非标准 Transformer 解码器交叉注意力那样关注图像所有位置, 从而实现更快收敛与更优性能; 同时引入多尺度高分辨率特征, 有效提升模型对小物体/小区域的分割效果; 此外, 还通过调整自注意力与交叉注意力的顺序、引入可学习查询特征、去除 dropout 等优化改进, 在不增加计算量的前提下进一步提升模型性能。2023 年, Segment Anything Model (SAM) [10]的问世标志着分割任务进入基础模型时代, 其通过在大规模数据集上的预训练实现了强大的零样本泛化能力, 允许用户通过交互式提示进行全场景分割。然而, 分割结果在微观几何特征上的表现却往往差强人意。具体而言, 受限于网络架构中频繁的下采样操作以及

特征图分辨率的丢失, 预测生成的分割掩膜(Mask)在物体边缘处常会出现明显的“锯齿现象”(Jagged Artifacts)。这种不规则的边缘不仅破坏了物理世界中物体的自然形态, 也直接导致了下游任务如精准抠图、机器人抓取控制及医学病灶精确测量在应用中的失效。导致这些边界不平滑的一个主要原因是, 在训练和预测过程中受限于图像分辨率。低分辨率的输入通常会导致输出结果较为粗糙, 且轮廓质量较差。虽然提高图像分辨率有可能提升边界的精度, 但这种方法往往会受到硬件条件的限制, 或者受制于图像放大(上采样)过程中所引入的插值误差。

为了攻克这一难题, 学术界开展了大量研究: Kirillov 等人提出的 PointRend [11]借鉴了图形学渲染思想, 通过对边界不确定点进行自适应采样来提升细节分辨率; SegFix [12]则利用坐标偏移预测将边界像素向内部高置信度区域修正; Gated-SCNN [13]和 Boundary-preserving Mask R-CNN [14]分别通过引入门控形状流和显式边界回归分支, 试图在特征提取阶段强化几何约束。RefineMask [15]旨在融合两阶段实例分割方法在实例区分上的固有优势, 以及细粒度特征在捕捉空间细节上的补偿作用。该框架的核心创新在于: 首先, 在特征金字塔(FPN)的高分辨率层级上构建语义头, 提取关键的细粒度特征; 其次, 在掩码头(Mask Head)中引入多阶段精炼策略, 通过在 RoI-Align 后逐步上采样并融入前述特征, 有效遏制了传统方法中的信息衰减。此外, 通过引入边界感知机制, RefineMask 实现了对复杂边缘的精准建模。CascadePSP [16]实现了从低分辨率到高分辨率的跨尺度掩码优化。CascadePSP 通过对初步生成的粗糙掩码进行二次加工, 显著提升物体边缘的预测精度与空间分辨率。其核心架构采用级联式设计, 遵循由粗到精(coarse-to-fine)的精炼逻辑: 早期层级侧重于捕获并稳定物体的全局结构, 进而为后续层级提供可靠的空间先验, 以实现边界细节的精准恢复与拓扑纠正。然而, 这些方法的共同局限性在于: 它们均需要在原有网络架构上额外引入一个辅助分支。由于该分支必须参与联合训练, 这不仅导致其在适配不同基线模型时面临繁琐的结构调整, 还显著提高了整体的计算开销与训练成本。

物体的曲面或曲线重构在逆向工程、计算机图形学和医学扫描等众多领域中具有广泛的应用基础。在处理由物理测量设备获取的不规则散乱点云数据时, 基于偏微分方程(PDE)的无网格数值方法在隐式曲面建模中展现出了独特的优势。其中, 基本解法(MFS)作为一种高效的计算技术, 常与势场(PF)法相结合, 通过纯解析的方式进行精确的三维几何重建。在此理论基础之上, Tankelevich 等人[17]使用了一种利用 PF/MFS 技术进行三维图像与曲面重构的方法, 将逼近真实物体的重构表面转化为一个由单一隐式方程 $f(P)=0$ 定义连续模型。为了进一步提升重构的整体表现, Chen 等人[18]随后提出了一种包含四阶偏微分方程的新型几何重构模型。该研究表明, 在求解这类新提出的高阶 PDE 时, MFS 作为一种简单且易于实现的无网格方法, 能够成功免除对传统虚拟边界(fictitious boundary)的依赖, 从而大幅简化了数值方法的实现流程, 并在有效重构复杂几何形状方面展现出了极大的优势。

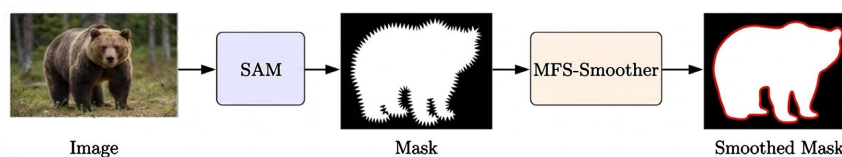


Figure 1. Overall flowchart

图 1. 本文整体流程

本文提出了一种全新的基于偏微分方程(PDE)基本解的图像光滑模块 MFS-Smoother。本文整体流程如图 1 所示, 经 SAM 生成初始掩码后, 由 MFS-Smoother 优化得到平滑掩码。不同于传统的离散优化策略, 该方法将掩膜边界视为受特定物理规律约束的连续演化, 利用偏微分方程基本解在数学上的解析优良性, 通过全局或局部的场演化实现对锯齿噪声的自然滤除。实验结果表明, 该方法在保持目标宏观拓

扑结构的同时, 能够赋予分割边界极高的几何平滑度, 为解决神经网络分割掩膜的质量缺陷提供了一种兼具数学严谨性与计算鲁棒性的新途径。

2. SAM 模型介绍

Segment Anything Model (SAM)作为计算机视觉领域的一项开创性工作。图 2 展示了 SAM 网络的整体架构。SAM 创造性地定义了“可提示分割”(Promptable Segmentation)的全新范式, 使得模型能够通过点、边界框、文本或粗略掩码等多样化的交互式提示, 灵活且精准地定位并分割图像中的任意目标。这一框架的巨大成功, 很大程度上得益于其背后规模空前的 SA-1B 数据集, 该数据集包含了超过 10 亿个高质量的掩码和 1100 万张高分辨率图像。其核心优势在于模型掌握了具备高度普适性的“物体”先验知识。打破了传统特定类别分割的限制, 该模型在零样本(zero-shot)学习框架下表现优异, 无需额外训练即可直接泛化并分割出未知的物体类型。此外, SAM 在系统设计上巧妙地解耦了繁重的图像全局特征提取与轻量级的提示交互过程, 从而在保证高精度特征表示的同时, 实现了极高的实时响应效率。这种兼具高度通用性、交互灵活性与计算高效性的特征, 为诸多复杂的下游应用提供了强大的底层视觉感知引擎, 极大地降低了密集预测任务中的像素级标注成本。

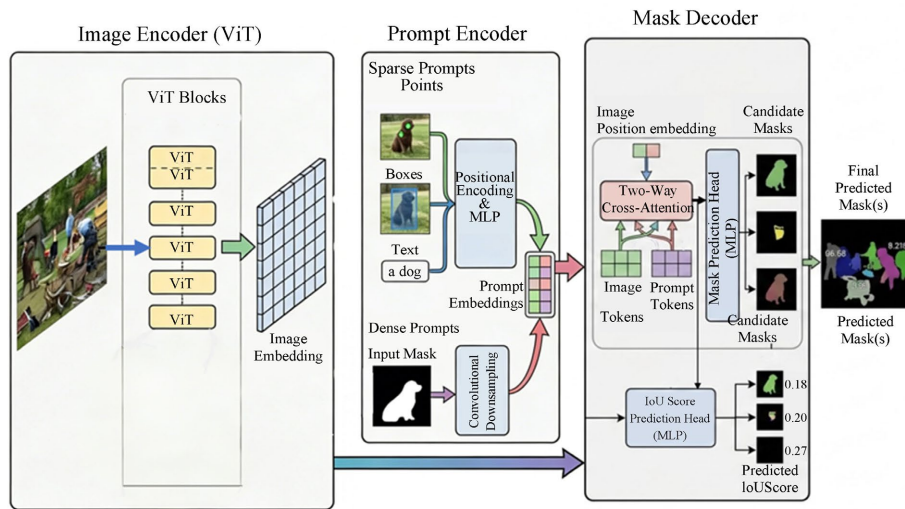


Figure 2. Network of SAM
图 2. SAM 结构图

2.1. Image Encoder

在 Segment Anything Model (SAM)的图像编码器中, 数据从原始像素空间到高维语义特征空间的演进经历了一个严密且高效的张量变换流程。首先, 经过预处理和标准化的高分辨率输入图像(尺寸通常固定为 $1024 \times 1024 \times 3$)被均匀划分为 16×16 大小的互不重叠的图像块(Patches)。随后, 这些图像块通过线性投影层展开并映射至高维隐藏空间(设通道维度为 D), 同时叠加可学习的绝对位置编码以保留原始的空间拓扑关系, 此时数据由三维图像张量转化为了空间尺寸为 64×64 、通道数为 D 的特征序列。在深度特征提取阶段, 该特征序列依次穿过多个视觉 Transformer (ViT)堆叠块, 网络通过交替应用局部窗口自注意力机制(Windowed MSA)与全局自注意力机制(Global MSA), 在始终维持 $64 \times 64 \times D$ 张量维度不变的前提下, 不断提纯语义特征并捕获长距离的上下文依赖。最终, 经过多层自注意力计算的深度特征进入轻量级的卷积颈部网络(Neck Module), 其高维通道 D 被显著压缩, 最终输出尺寸为 $64 \times 64 \times 256$ 的密

集图像嵌入(Dense Image Embedding)。从宏观数据流向来看，图像编码器完成了一次空间分辨率至原图 1/16 的大幅降采样与特征通道的扩张再压缩。

2.2. Prompt Encoder

在提示编码器(Prompt Encoder)中，不同模态的交互指令被分别转化为统一维度的特征向量(通道维度通常为 256)，以便与图像嵌入进行无缝融合。对于稀疏提示(Sparse Prompts)，例如用户点击的空间点坐标与边界框，网络首先利用位置编码(Positional Encodings)将其映射为高维特征，并与表征提示类型的可学习嵌入向量相加；若输入为文本指令，则直接引入预训练的 CLIP 模型提取跨模态特征序列。对于密集提示(Dense Prompts)，即用户提供的掩码(Mask)，其初始输入分辨率通常为 256×256 。该密集掩码会依次穿过多层卷积神经网络，执行空间维度的 4 倍下采样与通道升维，最终输出与图像特征尺寸完全一致的 $64 \times 64 \times 256$ 张量。在数据流的终点，这些经过卷积处理的密集提示特征会与图像编码器输出的全局图像特征进行逐元素的空间相加，而稀疏提示特征则作为独立的 Token 序列，直接送入下游的掩码解码器中。

2.3. Mask Decoder

掩码解码器(Mask Decoder)的数据处理流程是一个高效的多模态特征融合与上采样解码过程。首先，解码器接收来自前置模块的两个核心输入：一是空间尺寸为 $64 \times 64 \times 256$ 的图像嵌入特征，二是包含稀疏提示特征与可学习输出占位符(Output Tokens)的一维 Token 序列。在轻量级的 Transformer 解码块内部，数据流交替穿过自注意力机制与双向交叉注意力机制(Two-way Transformer)：一方面，提示 Token 序列作为查询向量(Query)去检索图像特征(Img-EmDp)，从而动态更新自身的语义状态(T2IAttention)；另一方面，图像位置特征(Img-EmDpPe)反向去检索提示 Token，以融入局部的引导信息(I2TAttention)。经过两次解码块的迭代融合后，更新后的图像特征会输入至空间上采样模块，通过两层转置卷积被放大 4 倍，恢复至 256×256 的空间分辨率。同时，更新后的输出 Token 则经过多层感知机(MLP)映射为动态分类器权重。最终，两者通过空间维度的矩阵乘法(点积操作)，并行输出多个(通常为 3 至 4 个)分辨率为 256×256 的预测掩码通道，以及对应的掩码置信度得分(IoU 分数)。图 3 展示了双向 Transformer 的网络结构，其核心在于双向交叉注意力机制。

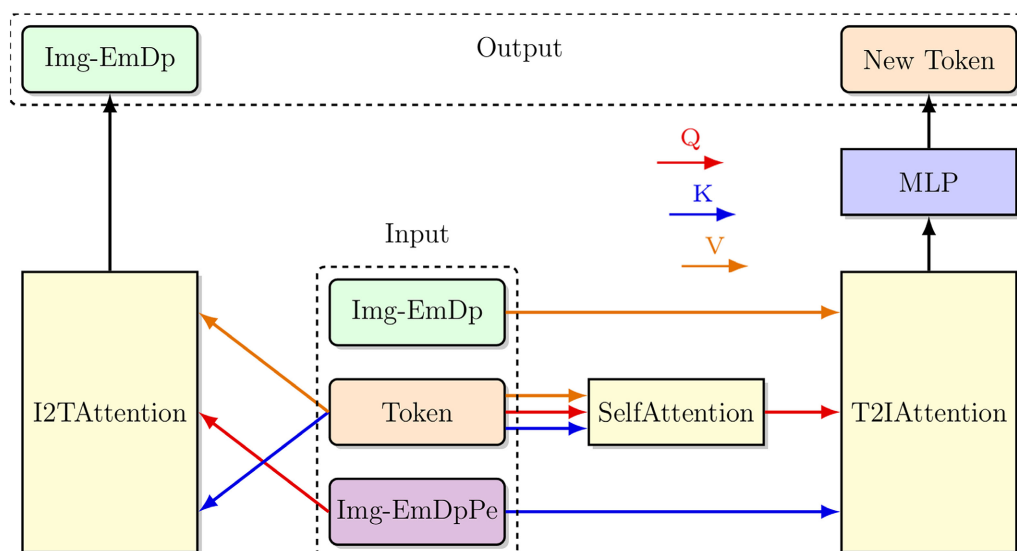


Figure 3. Structure of two-way transformer
图 3. 两步交叉注意力机制的结构

3. MFS 方法

3.1. 推导过程

在处理二维平面点云并进行隐式场重构时, 本文采用一种改进型双拉普拉斯 - 亥姆霍兹(modified bi-Helmholtz)算子作为控制方程的基础。以下为该二维算法的核心演化与降维推导全过程。

边值问题与控制方程

设定二维欧氏平面内的目标几何实体由封闭边界曲线 Γ 以及其内部平面区域 Ω 构成。对于平面内任意坐标 $\mathbf{p}=(x,y)$, 定义一个函数 $\phi(\mathbf{p})$, 其在整个求解域内需满足偏微分方程及其边界条件:

$$\begin{aligned} (\nabla^2 - \kappa^2) \phi(\mathbf{p}) &= 0, \quad \forall \mathbf{p} \in \Omega \\ \phi(\mathbf{p}) &= 1, \quad \mathbf{p} \in \Gamma \\ \frac{\partial \phi(\mathbf{p})}{\partial \mathbf{n}}(\mathbf{p}) &= h(\mathbf{p}), \quad \mathbf{p} \in \Gamma \end{aligned}$$

其中, ∇^2 代表二维空间下的拉普拉斯算子, κ 是待定参数。

上述偏微分方程组的解为

$$\begin{aligned} \tilde{\phi}(\mathbf{p}) &= \Phi_A(\mathbf{p}) + \Phi_B(\mathbf{p}) \\ \Phi_A(\mathbf{p}) &= \sum_{l=1}^N \alpha_l K_A(\rho_l, \kappa), \quad \Phi_B(\mathbf{p}) = \sum_{l=1}^N \beta_l K_B(\rho_l, \kappa) \end{aligned}$$

此处, $\rho_l = \|\mathbf{p} - \mathbf{q}_l\|_2$ 表示评估点 \mathbf{p} 与第 l 个离散源点 \mathbf{q}_l 之间的二维欧氏距离。 α_l 和 β_l 是由边界条件决定的待定权值系数。

其中,

$$\begin{aligned} K_A(\rho, \kappa) &= \frac{\exp(-\kappa\rho)}{4\pi\rho} \\ K_B(\rho, \kappa) &= \frac{\exp(-\kappa\rho)}{8\pi\kappa} \end{aligned}$$

选取 N 个位于二维边界曲线 Γ 上的点云数据作为配点。将数值解 $\tilde{\phi}$ 代入前述的边界条件中, 即可装配出一个稠密的分块线性代数方程组:

$$\begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{e} \\ \mathbf{H} \end{bmatrix}$$

在这里, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_N]^T$ 为未知的列向量; \mathbf{e} 是全为 1 的列向量; $\mathbf{H} = [h(\mathbf{p}_1), \dots, h(\mathbf{p}_N)]^T$ 。

$$\begin{bmatrix} [K_1(\rho_{ij}, \lambda)]_{1 \leq i, j \leq n} & [K_2(\rho_{ij}, \lambda)]_{1 \leq i, j \leq n} \\ [\frac{\partial}{\partial n} K_1(\rho_{ij}, \lambda)]_{1 \leq i, j \leq n} & [\frac{\partial}{\partial n} K_2(\rho_{ij}, \lambda)]_{1 \leq i, j \leq n} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{e} \\ \mathbf{H} \end{bmatrix}$$

直接求解上述大型分块矩阵计算代价高昂。为此, 引入一种代数降阶策略: 通过人为构造函数 $h(\mathbf{p})$ 的值, 使得方程组的第一个系数向量发生零化, 即:

$$\boldsymbol{\alpha} = \mathbf{0}$$

在该机制下, 原本庞大且繁杂的方程组变为仅依赖于第二组基的简化形式:

$$M_{12}\beta = e$$

伴随这一降维过程, 平面内任意一点的函数估值公式也得到极大简化:

$$\tilde{\phi}(\mathbf{p}) = \sum_{l=1}^N \beta_l K_B(\rho_l, \kappa) = \sum_{l=1}^N \beta_l \left(\frac{\exp(-\kappa\rho_l)}{8\pi\kappa} \right)$$

3.2. 计算过程: 基于点集的实际操作步骤

基于上述完成降维与非奇异化处理的理论模型, 当给定一个平面点云数据集后, 进行 2D 隐式曲线重建的具体操作过程可以归纳为以下五个核心步骤:

第一步: 数据准备(输入点集)

获取目标物体的 2D 点云数据(即平面中的一系列坐标点)。在上述简化的基本解方法模型中, 这些已知的数据点不仅作为边界上的配置点, 同时直接被用作源点。

第二步: 计算关键参数 κ

计算输入点集中所有点对之间的最近邻距离, 求出平均最小距离, 记为 d_{\min} 。

利用经验公式计算出模型所需的 κ 值。

$$\kappa = 5/d_{\min}$$

第三步: 构造线性方程组

利用前文推导出的双修正亥姆霍兹方程基本解建立数学关系。其公式为:

$$K_B(\rho, \kappa) = \frac{\exp(-\kappa\rho)}{8\pi\kappa}$$

根据所有点对之间的欧氏距离 ρ 和算好的参数 κ , 得到线性方程组:

$$M_{12}\beta = e$$

第四步: 求解权重系数

将上述构造好的矩阵输入到线性求解器中, 解出线性方程组 $M_{12}\beta = e$, 得到未知权重系数向量 β_k 。

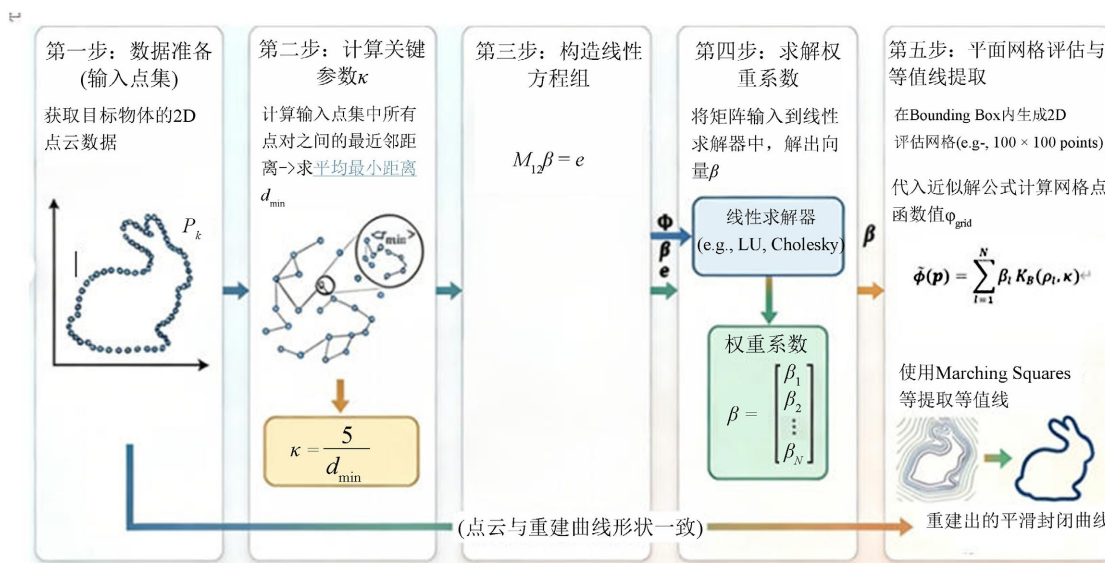


Figure 4. MFS calculation process

图 4. MFS 计算过程

第五步: 平面网格评估与等值线提取(渲染曲线)

网格评估: 在包含该点云的平面边界框(Bounding Box)内, 生成密集的二维网格点(例如 100×100 的评估点)。将求得的系数向量 β 代入近似解公式, 计算平面网格中每一个评估点的标量场函数值:

$$\tilde{\phi}(\mathbf{p}) = \sum_{l=1}^N \beta_l K_B(\rho_l, \kappa)$$

提取曲线: 在计算完整个平面的函数值后, 使用诸如 Marching Squares 等算法提取出函数值 $\tilde{\phi}(\mathbf{p}) = 1$ 所在的所有坐标位置。这个 $\tilde{\phi}(\mathbf{p}) = 1$ 的等值线(Isocontour)就是最终重建出来的平滑且封闭的 2D 曲线。图 4 展示了 MFS (边界平滑算法)的具体计算流程, 详细呈现了掩码边界优化的核心运算步骤。

3.3. 双拉普拉斯 - 亥姆霍兹方程的优势

在 MFS 的理论体系中, 控制方程的选择被视为算法的核心, 它不仅决定了算法的操作难度, 更直接影响到重构表面的效果和稳定性。例如拉普拉斯算子[17]的基本解形式为:

$$\phi_N(\mathbf{p}) = \sum_{l=1}^N \beta_l K(\mathbf{p}, \mathbf{q}_l), \quad K(\mathbf{p}, \mathbf{q}_l) = \frac{1}{4\pi|\mathbf{p} - \mathbf{q}_l|}$$

K 作为基本解具有奇异性, \mathbf{q}_l 需要人来手动设定它的分布位置。 \mathbf{q}_l 的最优选择仍是公开的挑战。对于亥姆霍兹方程[19], 它的基本解形如下:

$$K(\rho) = \left(\frac{\kappa}{2\pi\rho} \right)^{\frac{d}{2}-1} J_{\frac{d}{2}-1}(\kappa\rho)$$

其中, d 指维度, J 是第一类贝塞尔函数(Bessel function of the first kind)类似于上述的拉普拉斯算子, $K(\rho)$ 仍有奇异性, 面临同上述方程类似的挑战。因此, 本文选择的改进型双拉普拉斯 - 亥姆霍兹算子在此方面具有显著优势。

4. 评价指标介绍

4.1. 曲率

曲率能表征曲线或曲面的局部弯曲程度, 被广泛应用在曲线的光滑度评估中: 曲率值越小, 表明形状越平滑、越笔直; 曲率值越大, 则反映出弯曲越剧烈。三点共圆法[20]是一种直观且经典的数值计算方法, 常用于估算离散曲线(例如图像分割提取出的边界轮廓点集合)上某一点的局部曲率。其核心原理基于一个简单的几何理论: 不在同一直线上的三个点可以唯一确定一个外接圆。曲线上某一点的曲率 θ 等于该点密切圆半径 R 的倒数。

对于局部区域内的点, 点 Y 的曲率可以通过其相邻节点 X, Z 来进行评估, 公式如下:

$$\theta_y = \frac{4\sqrt{s(s-x)(s-y)(s-z)}}{xyz}, \quad s = \frac{1}{2}(x+y+z),$$

其中, 节点间的成对距离定义为 $x = \|X - Y\|, y = \|X - Z\|, z = \|Z - Y\|$ 。

4.2. 交并比(IoU)

IoU (Intersection over Union, 交并比)是计算机视觉领域中衡量目标检测与图像分割准确度的核心指标, 其本质是通过计算模型预测区域与真实标注(Ground Truth)区域的重叠比例来评估预测质量。在数学表达上, IoU 等于两个区域交集的面积除以它们并集的面积, 取值范围在 0 到 1 之间: 数值越接近 1, 代

表预测结果与真实目标越吻合；反之，若数值趋于 0，则说明模型完全偏离了目标。

假设 A 代表预测区域(Prediction)， B 代表真实标签区域(Ground Truth)，其公式如下：

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

5. 实验设计与实验结果

在实验环节，本研究基于 NVIDIA RTX 4090 GPU 硬件平台部署了 SAM (Segment Anything Model) 预训练模型。我们从 COCO 2017 数据集中随机抽取了 306 张不同分辨率的图像，进行推理与分割，并针对模型输出的分割边界进行了曲率评估测试，同时我们比较了 CascadePSP, PointRend 同 MFS-Smoother 的效果，实验结果表明，MFS-Smoother 优化后，分割精度 IoU 几乎不变，同时轮廓曲率值大幅降低，边缘平滑效果提升显著；与 CascadePSP、PointRend 等主流算法相比，该方法在保持高精度分割的同时，实现了更优的轮廓平滑性能，综合表现更具优势。为验证所提 MFS-Smoother 算法的边界优化效果，本文开展了两组对比实验，实验结果如表 1 与表 2 所示。表 1 展示了 MFS-Smoother 优化前后的核心指标变化：优化后模型的交并比(IoU)达到 0.9586，同时边界曲率值降至 0.025，表明算法在保证分割精度的同时，有效平滑了分割边界。表 2 进一步将 MFS-Smoother 与当前主流的边界优化方法 CascadePSP、PointRend 进行横向对比。实验数据显示，MFS-Smoother 在 IoU 指标上达到 0.9586，显著优于 PointRend 的 0.8224；同时在边界平滑度指标上，MFS-Smoother 的曲率值仅为 0.025，远低于 PointRend 的 0.2878，充分验证了所提算法在分割精度与边界质量上的双重优势，能够有效解决传统方法中边界锯齿、不连续的问题。

Table 1. Experimental data before and after MFS-Smoother optimization

表 1. MFS-Smoother 优化前后的实验数据

| | 优化后 | 优化前 |
|-----|--------|--------|
| IoU | 0.9586 | 0.9598 |
| 曲率值 | 0.025 | 0.300 |

Table 2. Experimental data of MFS-Smoother compared with CascadePSP and PointRend

表 2. MFS-Smoother 同 CascadePSP 和 PointRend 的实验数据

| | MFS_Smoother | PointRend | CascadePSP |
|-----|--------------|-----------|------------|
| IoU | 0.9586 | 0.8224 | 0.9599 |
| 曲率值 | 0.025 | 0.2878 | 0.303 |

通过图 5 的定性对比可以发现，MFS 有效解决了现有算法在边缘细化中的痛点。PointRend 和 CascadePSP 在处理复杂边缘时，往往会在局部生成离散的锯齿状轮廓，且前者因受限于 MaskR-CNN 的基础分割框架，随着新模型不断提高分割精度，导致 IoU 指标相比最新的 SAM 有一定下降。

图 6 中自左向右选取了四种具有不同边缘特征的测试对象：带有圆润边缘的红色消防栓、具有细长复杂肢体的长颈鹿、包含笔直锐角几何边缘的飞机，以及具有长直线条的黑白火车头。图像自上而下呈现了完整的对比过程：最上方两行首先展示了这四个物体的原始输入照片，以及覆盖了亮绿色蒙版的真实标签(Ground Truth)，以此标定理想的像素级分割区域；中间两行提取了目标物体的整体边界，并用矩形框圈出了待观察的特定局部，此时已经能初步观察到代表“优化后”的黄色整体线条比代表“优化前”

的绿色线条更加流畅; 最下方两行则是整张图的核心, 将矩形框内的局部边缘放大了 8 倍进行对比, 在微观像素级别下, 绿色线条暴露出极为明显的阶梯状的锯齿问题, 而采用该方法优化后的黄色线条即使在高度放大下依然保持着极高的连续性, 呈现出极度平滑的直线或自然曲线。该图通过从原始输入、宏观轮廓到微观像素细节的层层递进, 有力地证明了该算法能够有效消除锯齿伪影, 提取出高质量且边缘平滑的物体轮廓。图 7 展示了另一组边界优化效果的对比实验, 进一步验证了所提 MFS-Smoother 算法在不同场景下的边界平滑能力。

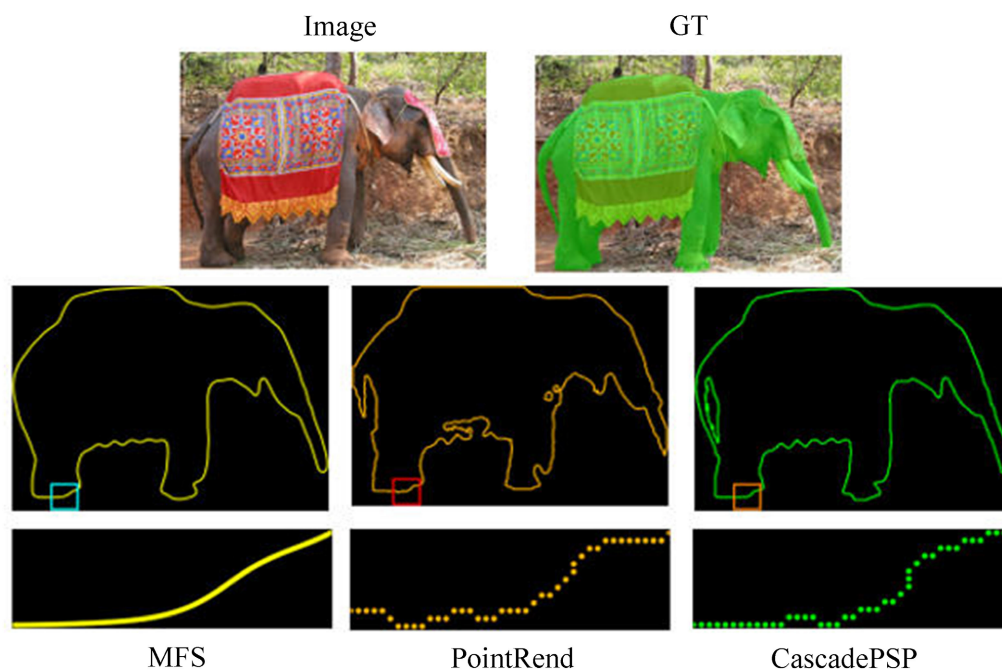
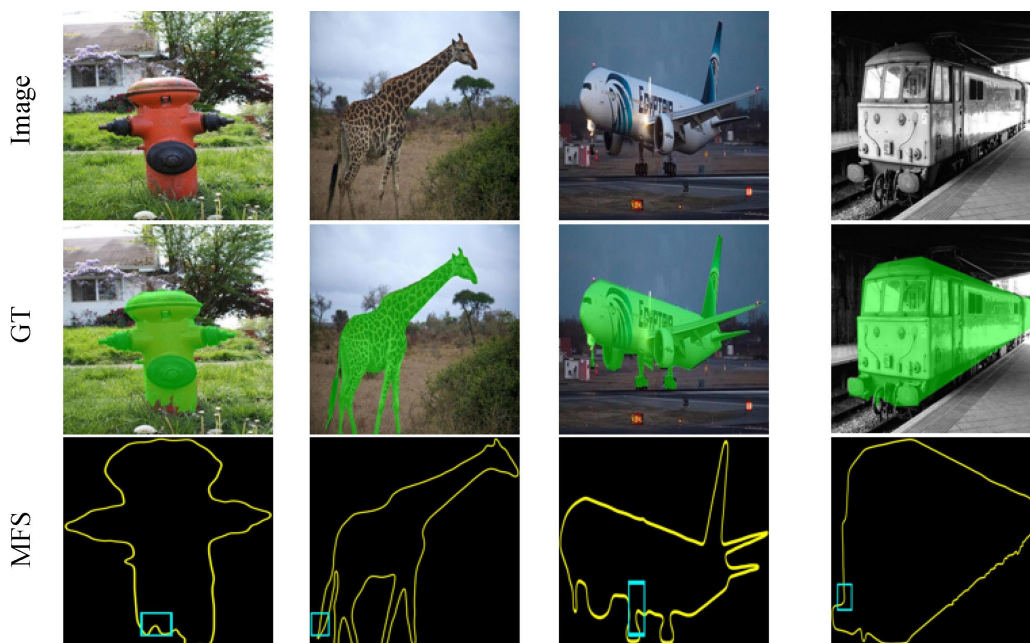


Figure 5. Comparison of different boundary refinement methods
图 5. 不同边界优化方法的比较



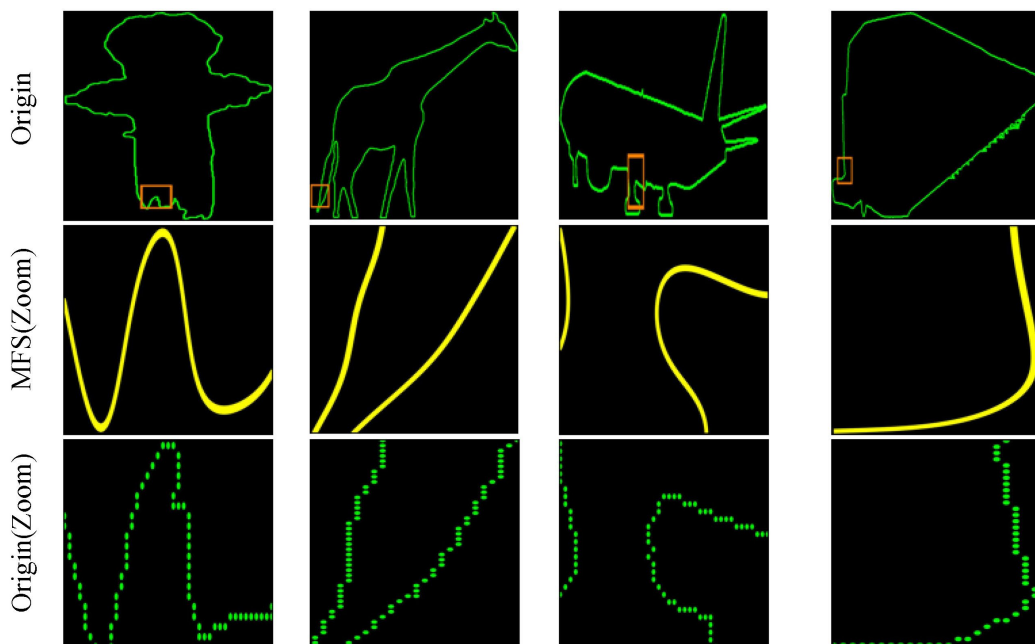


Figure 6. Boundary refinement comparison (1)

图 6. 边界优化效果对比(1)



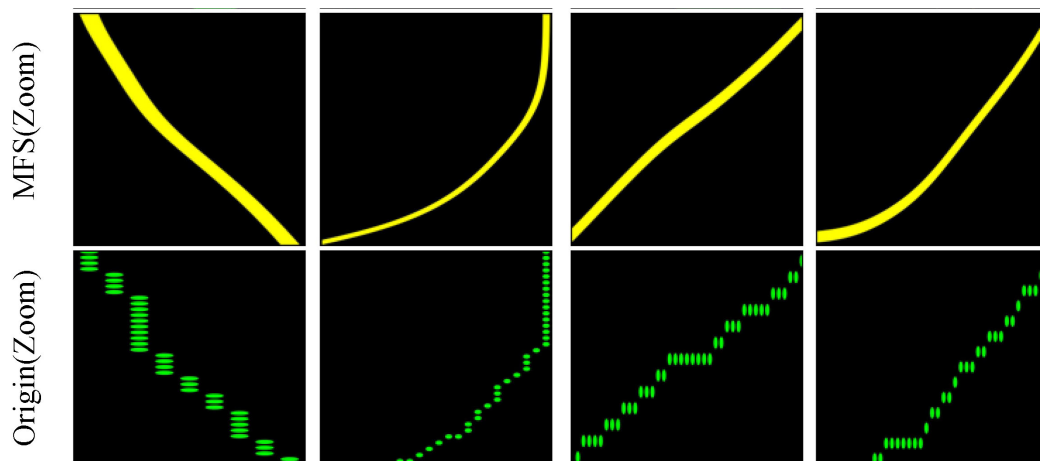


Figure 7. Boundary refinement comparison (2)
图 7. 边界优化效果对比(2)

6. 实验结论

本文提出了一种基于基本解法(Method of Fundamental Solutions, MFS)的图像分割边缘细化模块 MFS-Smoother, 旨在解决自然场景图像分割中长期存在的边界质量低下问题。

不同于现有的主流细化方案, 本研究分析了传统方法在优化边界时的局限性——即这类方法往往需要通过增加额外的训练分支或复杂的子网络结构来实现边缘修正, 这不仅显著提高了模型训练的计算成本, 也破坏了原有分割框架的简洁性。针对这一痛点, 本文引入的 MFS-Smoother 模块通过数学建模手段, 利用偏微分方程的解析特性, 在无需大规模修改网络架构的前提下, 实现了对物体边缘的精准建模。

在 COCO 2017 数据集上的综合实验结果表明, 该方法展现出了卓越的鲁棒性。通过对比分析, 引入 MFS-Smoother 后的分割结果在保持原有预测精度 IoU 值没有明显损失的同时, 边界的曲率值得到了显著下降。这一量化结果直观地体现在视觉效果上, 即大幅优化了分割掩码边缘常见的锯齿状走样, 生成了更加符合真实物体物理轮廓的平滑且连续的边界曲线。

从计算机视觉的实际应用价值来看, 这种对边界质量的本质提升具有深远的意义。高质量的形状表征能够直接赋能多个下游核心任务: 在实例分割与物体识别中, 更精准的轮廓有助于提升重叠物体的区分度; 在图像编辑与创意视觉创作中, 平滑的掩码边缘能大幅减少后期处理的手工参与成本; 在自动驾驶与机器人环境感知领域, 可靠的边界数据是实现精准距离感知、机器人避障以及目标抓取控制的关键基础。综上所述, 本文所提出的基于 MFS-Smoother 的边界细化方法不仅在技术层面解决了锯齿效应, 更在计算效率与优化效果之间取得了平衡, 为自然场景下的高性能图像分割研究提供了新的视角。

基金项目

本研究得到“基于超声造影-细胞角蛋白的精确诊断胆囊息肉样病变的人工智能模型构建研究”项目(2025L041)的资助。

参考文献

- [1] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [2] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2016) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495.

- <https://doi.org/10.1109/tpami.2016.2644615>
- [3] Chen, L., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49
- [4] He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2018) Mask R-CNN. <https://doi.org/10.48550/arXiv.1703.06870>
- [5] Ren, S., He, K., Girshick, R. and Sun, J. (2016) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. <https://doi.org/10.48550/arXiv.1506.01497>
- [6] Wang, J., Sun, K., Cheng, T., Jiang, B., *et al.* (2020) Deep High-Resolution Representation Learning for Visual Recognition. <https://doi.org/10.48550/arXiv.1908.07919>
- [7] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2023) Attention Is All You Need. <https://arxiv.org/abs/1706.03762>
- [8] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M. and Luo, P. (2021) SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. <https://arxiv.org/abs/2105.15203>
- [9] Cheng, B., Misra, I., Schwing, A.G., Kirillov, A. and Girdhar, R. (2022) Masked-Attention Mask Transformer for Universal Image Segmentation. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 1280-1289. <https://doi.org/10.1109/cvpr52688.2022.00135>
- [10] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., *et al.* (2023) Segment Anything. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 3992-4003. <https://doi.org/10.1109/iccv51070.2023.00371>
- [11] Kirillov, A., Wu, Y., He, K. and Girshick, R. (2020) PointRend: Image Segmentation as Rendering. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 9796-9805. <https://doi.org/10.1109/cvpr42600.2020.00982>
- [12] Yuan, Y., Xie, J., Chen, X. and Wang, J. (2020) SegFix: Model-Agnostic Boundary Refinement for Segmentation. In: *Lecture Notes in Computer Science*, Springer, 489-506. https://doi.org/10.1007/978-3-030-58610-2_29
- [13] Takikawa, T., Acuna, D., Jampani, V. and Fidler, S. (2019) Gated-SCNN: Gated Shape CNNs for Semantic Segmentation. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 5228-5237. <https://doi.org/10.1109/iccv.2019.00533>
- [14] Cheng, T.H., Wang, X.G., Huang, L.C., *et al.* (2020) Boundary-Preserving Mask R-CNN. In: Vedaldi, A., Bischof, H., Brox, T., *et al.*, Eds., *Computer Vision—ECCV 2020*. Springer International Publishing, Cham, 660-676.
- [15] Zhang, G., Lu, X., Tan, J., Li, J., Zhang, Z., Li, Q., *et al.* (2021) RefineMask: Towards High-Quality Instance Segmentation with Fine-Grained Features. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 6857-6865. <https://doi.org/10.1109/cvpr46437.2021.00679>
- [16] Cheng, H.K., Chung, J., Tai, Y. and Tang, C. (2020) CascadePSP: Toward Class-Agnostic and Very High-Resolution Segmentation via Global and Local Refinement. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 8887-8896. <https://doi.org/10.1109/cvpr42600.2020.00891>
- [17] Tankelevich, R., Fairweather, G. and Karageorghis, A. (2009) Three-Dimensional Image Reconstruction Using the PF/MFS Technique. *Engineering Analysis with Boundary Elements*, **33**, 1403-1410. <https://doi.org/10.1016/j.enganabound.2009.04.015>
- [18] Chen, C.S., Amuzu, L., Acheampong, K. and Zhu, H. (2021) Improved Geometric Modeling Using the Method of Fundamental Solutions. *Engineering Analysis with Boundary Elements*, **130**, 49-57. <https://doi.org/10.1016/j.enganabound.2021.04.025>
- [19] Lei, M., Liu, L., Chen, C.S. and Zhao, W. (2023) The Enhanced Boundary Knot Method with Fictitious Sources for Solving Helmholtz-Type Equations. *International Journal of Computer Mathematics*, **100**, 1500-1511. <https://doi.org/10.1080/00207160.2023.2191749>
- [20] Belyaev, A.G. (1999) A Note on Invariant Three-Point Curvature Approximations. <https://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/1111-16.pdf>