

基于 CSG 的二维 CAD 形状智能解析方法*

冯卓航, 申立勇[†]

(中国科学院大学数学科学学院, 北京 100049)

(2024 年 12 月 9 日收稿; 2025 年 3 月 3 日收修改稿)

Feng Z H, Shen L Y. An intelligent analysis method for 2D CAD shapes based on CSG [J]. Journal of University of Chinese Academy of Sciences, 2026, 43(5): 603-613. DOI: 10. 7523/j.ucas. 2025. 006.

摘要 构造实体几何 CSG 技术通过几何基元集与基元间的布尔运算来定义复杂形状, 但从已有的几何形状逆向建模恢复 CSG 构造序列一直是富有挑战性的研究问题。提出一种基于 Transformer 的深度网络架构, 解析输入的几何形状, 并输出该形状的 CSG 建模序列。采取 CSGNet 算法作为基线, 首先扩充训练所用的合成数据集, 使得模型能够学习并解析更广泛的形状。此外, 采取 VGG 编码器-Transformer 解码器的自回归学习架构, 基于注意力机制更好地关联生成序列, 将二维形状输入解析为固定格式的 CSG 序列输出, 得到更好的重建质量。还引入预测有效性矫正模块, 确保模型不会输出无效的重建序列。结果表明, 本文提出的模型在合成数据集和真实 CAD 数据集上都能够取得更好的质量。

关键词 构造实体几何(CSG); 注意力机制; 深度学习; 有效性矫正

中图分类号: TP2 文献标志码: A DOI: 10. 7523/j.ucas. 2025. 006

An intelligent analysis method for 2D CAD shapes based on CSG

FENG Zhuohang, SHEN Liyong

(School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract Constructive solid geometry (CSG) is a geometric modeling technique that defines complex shapes through Boolean operations between basic geometric primitives. However, reverse modeling to recover the CSG construction sequence from existing geometric shapes has been a challenging research problem. In this paper, we propose a Transformer-based deep network architecture that can parse input geometric shapes and output their corresponding CSG modeling sequences. Using the CSGNet algorithm as a baseline, we first expanded the synthetic dataset used for training, enabling the model to learn and parse a broader range of shapes. Additionally, we adopted an autoregressive learning architecture with a VGG encoder and a Transformer decoder, leveraging the attention mechanism to better correlate the generated sequences. This approach translates 2D shape inputs into CSG sequence outputs in a fixed format, resulting in improved reconstruction quality. We also introduced a validity correction module to ensure that the model

* 国家自然科学基金(12371384)和中央高校基本科研业务费专项资助

[†] 通信作者, E-mail: lyshen@ucas.ac.cn

does not output invalid reconstruction sequences. Experimental results show that the proposed model achieves better quality on both synthetic and real CAD datasets.

Keywords constructive solid geometry (CSG) ; attention mechanism; deep learning; validity correction

在现代设计中,计算机辅助设计(computer-aided design, CAD)技术被广泛应用于工业设计、建筑、制造等多个领域。而从已有形状中逆向重建CAD模型是几何建模设计者长期追求的目标之一。完整的CAD模型包含了构造的细节,除几何信息和简单的拓扑信息之外,还涉及组件的综合设计知识及其相互关系,这使得设计者能够更好地了解CAD形状的结构,并可以根据需要更方便地创建新的模型。构造实体几何(constructive solid geometry, CSG)是CAD设计中一种经典的几何表示方法,它通过在简单的几何图元(二维图元如圆形、三角形、矩形等,三维图元如球体、立方体、圆柱体等)上依序应用布尔运算(交、并和差)来生成更复杂的CAD形状,在自动化、模块化和精确控制方面具有显著优势。图1展示了二维形状的CSG形式构造步骤。

传统设计场景下,CAD形状的生成主要依赖于手动设计,设计师根据专业知识与经验逐步构建模型。这种手工设计方式虽然具备灵活性,但随着模型复杂度的增加,其操作的难度和时间成本也迅速攀升,且在设计过程中难免会因人为因素导致精度偏差或重复劳动,影响设计质量的稳定性和一致性。此外,在复杂设计项目中,设计师通常需要对已有的形状不断进行调整和复用,如重新构建基本形状模块、调整布尔操作顺序、重新设置参数等。这种模式下,设计师若希望对复杂

形状的局部区域进行修改,通常需要手动回溯并调整多个操作步骤,甚至可能要重新设计相应的构造模块。这不仅增加了设计难度,还限制了设计的可扩展性,使得设计的复用和形状的管理变得繁琐,严重影响了设计效率。

CSG解析方法通过将复杂CAD形状分解为基础几何模块和布尔操作序列,为设计提供了系统化的构造路径。这种解析方式不仅保留了每一步操作的顺序与参数,使得设计过程更易追溯和修改,且能快速对形状进行局部调整和复用,从而提高设计效率,确保模型构建的准确性和一致性,有效解决了传统手动设计中的复杂度、调整难度和复用性等问题。因此,如何将CAD形状解析为CSG序列的表达形式成为值得研究的热点问题。

当前,CAD形状的CSG解析任务面临很多挑战。首先,CAD形状的CSG表示并不具有唯一性,如图1所示,同一个CAD形状可能对应多种CSG构造方式,这会给学习带来不确定性。其次,在解析一个结构未知的CAD形状时,模型无法事先预知其所对应的基元数量及所需要的操作的数量,这导致目标输出的维度和CSG序列长度均无法固定。此外,CSG序列的微小变化(包括基元与操作的选择及其排列顺序)可能导致其对应生成的图形截然不同,且序列越长越容易导致微小误差的逐渐累积,最终与目标图形产生巨大差异。

目前已有很多对二维与三维CAD形状的解析研究,过去主流方法多侧重于搜索与优化技术,尽管这些方法能够生成具有一定结构的形状,但通常运算速度较慢,且对噪声不具有鲁棒性。近年来,基于深度学习的解析方法逐渐成为主流,为CAD形状生成提供了新的路径。其中,基于自回归的学习框架能够更好地利用已生成序列的信息做下一步解析,在近几年越发受到关注。基于序列生成的深度学习模型能够自动识别并生成复杂形状的构建过程,使得设计人员可以在无需手动操作的前提下获得所需的几何结构。相比于传统的静态形状表示方法,此类模型不仅能够逐步生成形状,而且便于设计师在设计过程中对形状进行修改和调优,这为CAD形状设计提供了智能

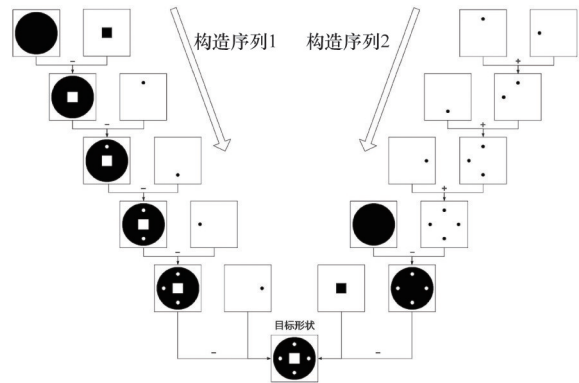


图1 同一二维形状的2种CSG构造序列示意图
Fig. 1 Two different CSG construction sequences of the same 2D shape

化、易编辑化的可能性。

本文提出的模型为基于Transformer^[1]的自回归框架,这主要是因为CSG表示在结构上可以视为一种CAD语言序列,因此可以利用Transformer架构强大的序列学习能力来做解析。整体来看,本文的模型基于编码器-解码器结构,采用卷积神经网络编码器学习输入的二维形状特征,然后利用Transformer解码器递归地进行解码以输出生成序列。本文以CSGNet^[2]作为基线算法,提出一种智能几何基元分析方法,主要贡献为:

1)提出一种基于视觉几何组(visual geometry group, VGG)编码器^[3]-Transformer解码器的端到端二维图形解析模型,有效结合多种训练方法,以自回归的方式学习CAD形状的CSG构造序列。

2)扩充了CSGNet中的合成数据集,引入更丰富的基元字典,使得模型可以学习到更广泛的基元类型,具备对更复杂形状的解析能力。

3)加入序列有效性校正模块,使得模型在预测时,不会生成语法无效的CSG序列,显著提高了重建的质量。

1 相关工作

在CAD领域,获取几何形状的CSG构造序列有利于设计者更好地复用和编辑形状。但如何直接从原始几何形状逆向恢复CSG建模信息是一个较复杂的问题。

近十几年来,CAD形状的解析研究已有丰富的积累,研究者们提出了多种方法,为复杂形状的分解与重构奠定了重要基础。这些方法大致可以分为2类:早期工作中基于搜索及优化的方法和近年来与机器学习相结合的解析方法。前者包括对象的识别与建模、将形状的边界表示(boundary representation, B-rep)转换为CSG程序、基元拟合等传统技术;后者则结合多种先进的神经网络直接学习目标形状的分解方式,特别是在深度学习和生成模型的推动下,研究者能够在较少人工干预的情况下,通过模型训练实现形状的自动解析与生成。本节简要总结了现有的代表性工作。

1.1 二维形状解析

在传统形状解析的工作中,研究者们首先关注如何从图像识别局部特征。一个早期例子是Fischler和Elschlager^[4]的图像结构表示和匹配方法,采用混合表示方案,使用树结构语法表示铰接

物体,能够在大量计算中找到描述对象。该工作随后被进一步拓展到基于部件的建模和对象识别^[5],应用于人脸和人体等在图像中的定位、静态图像中的人体姿态估计^[6]等方面。此外,研究者们也开始关注CSG构造信息的提取与利用。Weiss^[7]在早期提出基于几何的CAD规范树结构优化的方法,基于高度约束的CAD几何体,实现较全面的自动化结构变化和优化。Foley等^[8]则提出一个从B-rep表示到CSG表示的转换系统,通过构造额外的分离半空间来转换形状表示。Buchele和Crawford^[9]基于上述方法,进一步补充半空间简化B-rep表示到CSG转换过程的相关工作。

使用深度学习的形状解析中,二维领域的相关文献并不多,但已有研究者利用神经网络来推断一些通过特定领域语言表达的程序,例如Balog等^[10]提出的DeepCoder,是一种使用深度学习自动编程的方法,可以根据输入输出示例生成程序。Johnson等^[11]则提出一种视觉推理模型,根据程序生成器和执行引擎来显式获取推理预测过程。此后,Sharma等^[2]提出了在二维图形解析领域中具有代表性的工作CSGNet,该方法使用神经网络生成一组由形状建模指令组成的程序,继而通过形状渲染器组合建模形状来匹配目标图像。Zhou等^[12]则提出一种无监督学习模型,该模型无需预训练,结合语法编码树LSTM及路径采样来解析形状。Jones等^[13]通过引入伪标签和近似分布(pseudo-labels and approximate distributions, PLAD)构造数据对,提出PLAD训练框架,将一系列相关的自监督学习技术分组,用于形状程序推理。受上述研究工作的启发,本文延续图形编码与序列解码相结合的解析框架,设计并提出一种CAD语言序列生成模型。该模型有效整合多个模块组件,并结合监督学习、强化学习以及PLAD训练方法,实验结果表明其性能优于现有主流的二维形状解析模型。

1.2 三维形状解析

传统3D重建思路主要包括基元拟合、以3D网格等形式表示几何形状等方法,如SPFN^[14]与BRepNet^[15]等。但此类几何表示难以进行进一步编辑和修改。近年来,研究者的目光更多地集中在将三维形状解析为语言序列,且研究路径逐渐从开始的CSG分解框架转变到草图-拉伸序列学习框架。

在 CSG 分解框架下,模型被表示为三维基元间的布尔运算。CSGNet 作为此领域的早期工作,提供了有效的学习模式,但由于其使用的基元相对基础,难以适应复杂的三维情况。类似地,后续的一系列相关工作,如 UCSG-Net^[16]、CSG-Stump^[17]、BSP-Net^[18]、CAPRI-Net^[19]及 D²CSG^[20]等分别在学习方式和基元构造方面做出改进,取得了一定的成果。

在草图-拉伸序列框架下,第一次真正将 CAD 建模为语言序列的工作是 Wu 等^[21]提出的 DeepCAD,它构造了用于序列学习的数据集,并提出一个编码-解码的学习框架。这也引发了人们对于草图-拉伸语言序列的研究热情。SkexGen^[22]方法提出一种适合自回归学习的词元化方式,但其主要用于三维 CAD 形状的随机生成而非逆向建模。最新的研究如 CAD-SIGNet^[23]和 PS-CAD^[24]等,均从点云中提取形状信息,并以自回归的方式学习完整建模序列,取得了较好的重建结果。尽管一些前沿方法尚未开源,但它们都体现出 Transformer 架构在序列生成领域的强大功能,故有望应用至二维领域以取得更优的效果。

本文主要聚焦于二维图形,尽管当前三维形状学习已成为研究的主流,但从其发展趋势来看,草图-拉伸学习框架在学习能力上更为突出,且能更好地满足灵活编辑形状的需求。而草图本身是二维形状轮廓,故二维图形的解析对草图生成学习具有重要意义,能够为后续工作提供指导,进而推动相关成果向三维形状解析领域拓展。

2 网络模型训练

本文提出基于 VGG 编码器-Transformer 解码器的网络架构,给定二维图形,解析输入的几何形状,并以自回归的方式生成 CSG 表示序列。

2.1 编码-解码学习框架

设输入图像为 I ,编码器将图像 I 作为输入,经过多层卷积结构生成编码 $E(I)$ 。解码器 D 将 $E(I)$ 作为输入,根据注意力机制,学习生成表示为指令序列的程序 P 的概率分布。因此整个网络即可以写为 $\Theta(I) = D \circ E(I)$ 。

VGG 编码器结构遵循了简洁的设计理念:网络包含多个相同结构的卷积层、池化层和全连接层组合块,在每一层中使用相同大小(3×3)的卷积核,并通过堆叠这些卷积层来逐渐增加网络的

深度。VGG 网络相对于传统的卷积神经网络编码器,能够在不增加太多计算复杂度的情况下,捕捉到更复杂和深层次的特征。

从现有的 CAD 加工和几何形状逆向解析工作中可以看出,Transformer 不仅能有效处理三维点云的特征提取和压缩^[25-26],而且在 CAD 语言序列生成领域具有强大的潜力,与传统用于时间序列预测的循环神经网络(recurrent neural network, RNN)架构相比,Transformer 架构拥有显著的优势,主要体现在:1)长距离依赖建模能力。尽管如长短期记忆网络和门控循环单元网络等引入了门控机制,但依然在捕捉长程依赖时存在一定的局限性。Transformer 通过自注意力机制能够直接捕捉序列中任意 2 个位置之间的依赖关系,使得每个词在生成时都能基于整个序列,从而更好地捕捉全局信息。2)变长序列的处理能力。RNN 具有固定的时间步依赖关系,不能灵活处理输入序列的变长和不同的时间步。而 Transformer 的自注意力机制具有高度的灵活性,可以处理不同长度的输入序列,并且能够根据需要调整各个位置的权重。3)更高的训练速度和效率。RNN 的结构依赖于时间步的顺序执行,即当前时刻的输出依赖于前一时刻的隐藏状态,这导致了计算必须按顺序进行。Transformer 则通过自注意力机制摆脱了序列的顺序依赖,使得输入序列中的每个元素都可以在同一时间步进行处理,加速了训练过程。

本文的解码器采用基于 Transformer 的结构,以二维图形经过编码器得到的几何特征作为“记忆”输入,同时将基元、操作以及初始/终止符等转换为词元格式,从初始符开始,以自回归的形式基于当前构建序列预测下一步的操作概率。整体网络框架如图 2(a)所示。

2.2 两阶段训练框架

由于目前已有的二维图形或图像数据均以像素形式表示,不包含或本身并没有 CSG 构造序列的信息,因此在初始训练模型时,需要在人工合成的数据集上进行。与 CSGNet 中的形式相同,人工数据集表示为 CAD 语言序列的形式,二维像素表示的图像是在训练过程中通过图形渲染器得到。其中 CAD 语言序列需要遵循一定的语法规则才能够被读取并渲染为图像。基于人工构建的数据集,网络模型能够通过有监督学习的方式进行训

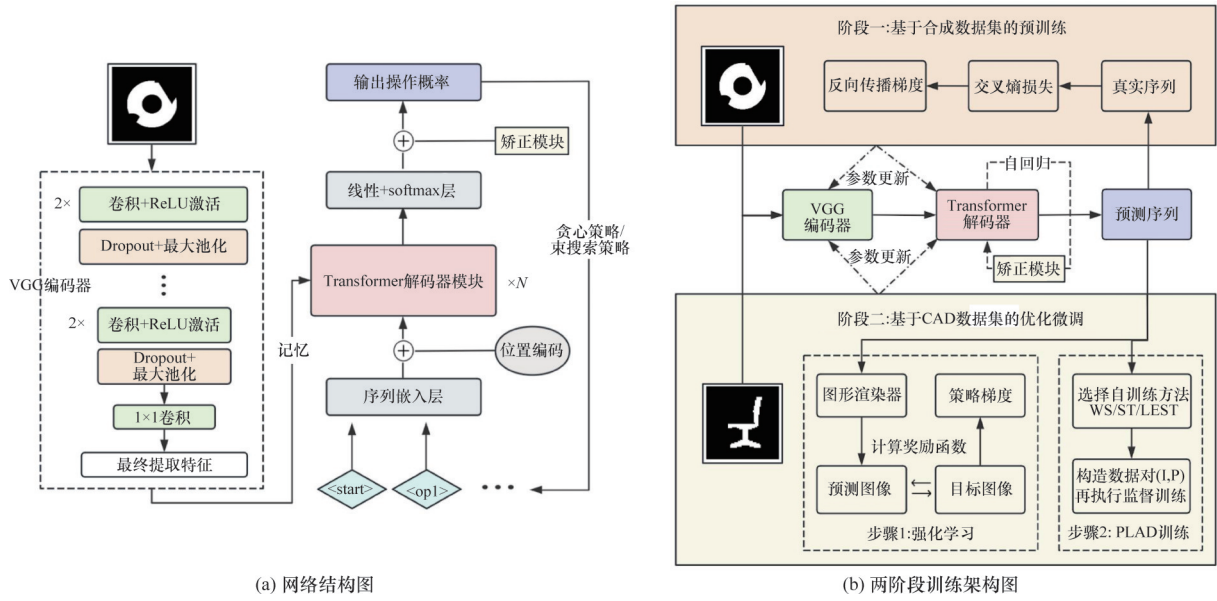


图2 网络结构与训练框架图

Fig. 2 Network structure and training framework

练,并根据交叉熵损失函数执行梯度反向传播策略以更新模型参数。由于生成模型在初期的预测效果很差,本文采用教师强制策略辅助训练,即在训练时,采用真值序列代替上一步的预测值作为Transformer解码器的每阶段输入词元,防止产生误差累积并提高模型的收敛速度。在测试和推理时,则正常按照模型预测计算出的结果作为下一步预测的输入。

由于合成数据集形状和教师强制策略本身的局限性,所获得的初步训练模型的能力和泛化性仍需提高,因此本文选择在CAD数据集上执行进一步的微调训练。真实二维CAD图形数据集由64×64大小的黑白图像构成,不包含CSG构造序列信息,故采用强化学习与PLAD框架^[13]进行训练。

与CSGNet类似,本文首先基于在合成数据集上训练所得的模型,在CAD数据上根据强化学习策略采样生成序列。其中,模型在每一生成步中输出字典各词元的概率向量,将这一概率分布视为强化学习中的状态-动作选择策略,依据对应概率随机采样生成当前步的词元。通过最大化预期奖励(等价于最小化执行程序获得的形状与目标形状之间的重建误差)来执行训练过程,使用策略梯度优化模型参数。选用蒙特卡罗方法采样 K 条轨迹来估计期望奖励,具体表示为

$$\hat{J}(\theta) = \frac{1}{K} \sum_{i=1}^K R^i \rightarrow J(\theta) = \mathbb{E}_{\pi_{\theta}}[R] = \mathbb{E}_{\pi_{\theta}} \left[\sum_{t=0}^T R_t \right]. \quad (1)$$

由于 $\nabla_{\theta} R_t = R_t \cdot \nabla_{\theta} \log \pi_{\theta}(a_t | s_t)$,则有

$$\begin{aligned} \nabla_{\theta} J(\theta) &= \mathbb{E}_{\pi_{\theta}} \left[\sum_{t=0}^T \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) R_t \right] \\ &\approx \frac{1}{K} \sum_{i=1}^K \sum_{t=0}^T \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) R_t^i. \end{aligned} \quad (2)$$

其中, $\pi_{\theta}(a_t | s_t)$ 为策略概率,即基于当前状态 s_t 下选择动作 a_t 的概率; R_t 为随机变量,表示序列生成第 t 步获得的奖励; ∇_{θ} 是表示对模型参数 θ 求梯度的符号。

强化学习初步增强了模型的能力,随后再基于PLAD思想做进一步微调训练。PLAD训练方法在推理时利用束搜索为CAD数据集中的每一幅图形都维护一个对应的最佳解析序列,并在每个批量样本的学习中随机以“觉醒-睡眠”、“自训练”或“自训练变体”构造(图形,生成序列)数据对来执行有监督训练,直至目标函数在验证集上收敛,得到最终模型。上述两阶段训练框架如图2(b)所示。

2.3 预测矫正模块

由于图形的CSG构造序列具有一定的语法约束,并非所有程序语言都是有效、可被恢复为二维图像的,因此所训练的网络还必须学习这一生成语法规则。然而在测试过程中发现,由于生成

的操作序列在顺序上具有较大的自由度,模型仍会在一些情况下生成无效的构造序列,且根据序列生成的特性,如果在序列的前段已经生成了不符合语法结构的词元,那么无论后续生成结果如何,该序列都无法被解析为有效的二维图像,这显然不符合任务的目标。

针对此问题,本文在 Transformer 解码器后加入一个矫正模块。此模块为每一条生成序列维护一组当前序列的状态,包括时间步、生成形状基元数、操作基元数、终止符数等。在模型推理时,矫正模块根据当前状态和 CSG 序列的语法规则,计算不合理的词元位置,并以掩码形式屏蔽其概率。此模块的输入为当前步的初始预测概率,输出经过修正后的词元概率分布,继续正常用于后续的策略选择和损失计算步骤。值得一提的是,由于

训练时采用教师强制策略,则矫正模块也应按照真值序列而非已生成序列在每一预测步骤生成矫正掩码。具体算法流程如图 3 所示。

在实际训练中,过早地引入预测矫正模块会导致模型的训练效果变差。这是因为该模块在屏蔽不合理的操作或结果的同时也会限制模型的探索空间。在训练过程中,模型需要通过多次尝试来逐步改进和调整预测,最终学习如何生成正确的操作或结果。如果掩码过早地限制了不合理操作的范围,可能会让模型错过某些潜在的学习路径,或者强迫模型从一开始就只探索较少的操作空间,导致学习的多样性减少,掩盖了模型自我修正能力,从而影响学习效果。

因此本文采用初期自由探索,后期引入矫正的策略。即在训练初期模型可以自由探索,更好

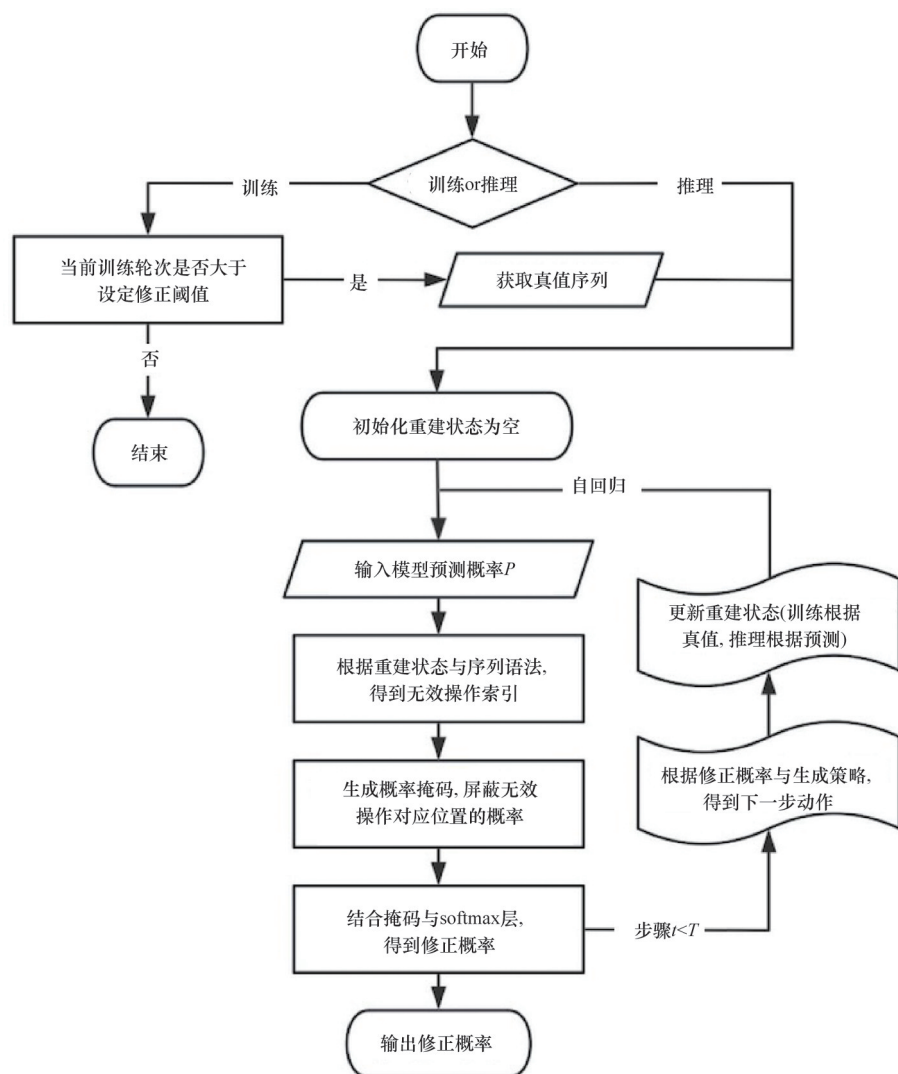


图 3 预测矫正模块流程图

Fig. 3 Prediction correction module flowchart

地了解整个任务的结构和数据分布,此时模型能够在相对较大的搜索空间中找到合适的解决方案,并学习到更加有效的特征表示。在模型训练至一定轮次后,模型已经掌握了任务基本规律,学习到了有效的特征表示和策略。这时,加入掩码可以帮助模型集中精力学习合理的操作,进一步提高训练效率,并减少不合理操作对模型学习的干扰。

3 实验与结果

本文对所提出的网络模型在合成数据集和真实CAD数据集上进行了实验验证。选择的对比基线算法有CSGNet^[2]、CSGNet-RL^[13]、SWRNet^[12]和PLAD^[13]。其中由于SWRNet作者并未提供预训练模型,故从初始执行其代码至目标函数收敛;其余算法均直接采用官方提供的预训练模型参数。本节展示了数值结果、视觉评估结果以及各模块的消融实验。

3.1 数据集与损失函数

3.1.1 数据集构造

本文首先扩充了CSGNet所提供的合成数据集。原数据集中,二维基元只有正方形、圆形和向上正三角形3类,每一类形状的位置和大小可以取多个离散值。数据以CSG语言序列的形式存储,符合规定语法规则的语言序列能够被图形渲染器解析为二维图形。

但上述3类基元对一些曲面形状的表达有限,而更多元化的基元则能够组合生成更多样的形状。因此本文选择加入长短轴之比为2:1的水平椭圆和竖直椭圆以及向下正三角形这3类基元。仍然采用CSGNet合成数据的语法规则,但微调了CSG构造序列的随机生成策略,据此生成了更广泛的词元字典(不同类型、不同大小或不同位置均被看作不同的词元),具体词元的符号表示与生成构造序列的语法规则如表1所示。

根据统计,CSGNet方法的字典共包含396个词元,而本文构造数据集的字典中共包含826个词元,更丰富的词元为提升模型的表达能力提供了良好的基础。

在二维CAD数据集方面,沿用CSGNet采用的数据集,其中的数据来源于Trimble 3D Warehouse数据集中收录的3类CAD形状:椅子、桌子和灯具,并从正面和侧面视图将这些CAD形状渲染为

表1 数据生成规则

Table 1 Data generation rules

类别	符号表示及说明
形状(S)	圆形(c)、三角形(t, T)、正方形(s)、椭圆形(e, E)
操作(O)	并集(+),交集(*),差集(-),终止(\$)
基元(P)	S(X, Y, R),由形状、坐标和半径组合表示
语法规则	任意步需满足: $\text{Max_len} // 2 + 1 \geq \text{num}(P) > \text{num}(O)$ 序列结束时需满足: $\text{num}(P) = \text{num}(O) + 1$
生成	开始 $\rightarrow P \rightarrow P \rightarrow P \parallel O \rightarrow \dots$
示例 (图1)	构造序列1(左): $c(32,32,24)s(32,32,8)-c(32,16,2)-c(32,48,2)-c(16,32,2)-c(48,32,2)-\$$ 构造序列2(右): $c(32,32,24)c(32,16,2)c(32,48,2)c(16,32,2)c(48,32,2)+++-s(32,32,8)-\$$

64×64大小的二进制图像共16 000幅,进一步将其分为训练集(10 000幅)、验证集(3 000幅)及测试集(3 000幅)。本文基于此数据集进行第二阶段的微调工作。

3.1.2 损失函数与度量

在合成数据集上,对模型进行有监督的训练,由于解码器每一步 t 的输出都为各词元的生成概率 $p_{t,i}^{(n)}$,因此将真值序列转换为独热编码形式 $y_{t,i}^{(n)}$,采用交叉熵损失(cross-entropy loss, L_{ce})函数作为目标函数,即

$$L_{ce} = -\frac{1}{N} \sum_{n=1}^N \sum_{t=1}^{T_n} \sum_{i=1}^C y_{t,i}^{(n)} \log \left(\frac{y_{t,i}^{(n)}}{p_{t,i}^{(n)}} \right). \quad (3)$$

其中, N 是训练中的批量大小, T_n 是当前数据真值序列的长度, C 是字典中词元的数量。

此外,对于图像的度量,选取交并比(intersection over union, IOU)和倒角距离(chamfer distance, CD)作为视觉定量衡量标准,对于2幅二值图像 A 和 B ,它们分别定义为

$$\text{IOU}(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (4)$$

$$\text{CD}(A, B) = \frac{1}{2|A|} \sum_{p \in A} \min_{q \in B} \|p - q\|_2^2 + \frac{1}{2|B|} \sum_{q \in B} \min_{p \in A} \|q - p\|_2^2. \quad (5)$$

对于强化学习,将奖励函数定义为

$$R = w \cdot \left(\text{clip}_{[0,1]} \left(1 - \frac{\text{CD}(I_{\text{target}}, I_{\text{pred}})}{S \cdot \sqrt{2}} \right) \right)^p + (1-w) \cdot \text{IOU}^p. \quad (6)$$

其与倒角距离呈负相关,与交并比呈正相关。其中 $\text{clip}_{[0,1]}$ 表示将输入值限制在区间 $[0,1]$ 内,若值小于0则取0,大于1则取1,否则保持原值。 S

为正方形画布的边长, $w \in [0, 1]$ 为权重系数, p 为放缩因子。

损失函数的选择上,使用滚动平均基准来减少方差,提升训练的稳定性。定义策略梯度损失函数为对数概率与优势函数(当前步奖励与基准的差异)相乘

$$L(\theta) = -\frac{1}{N \cdot K} \sum_{i=1}^N \sum_{j=1}^K \sum_{t=0}^{T_n} \log \pi_{\theta}(a_t^{i,j} | s_t^{i,j}) (R_t^{i,j} - b_t^{i,j}).$$

(7)

其中, K 为蒙特卡洛采样轨迹数, $b_t^{i,j}$ 为 t 步之前的平均奖励。模型基于此损失函数执行强化学习的策略梯度更新。

统一将各对比模型在每一生成步中输出的词元概率分布视为强化学习中的策略函数,计算生成序列的平均奖励值,将其也作为模型性能的衡量标准之一。

3.2 实验结果对比

在具体训练中,取 VGG 的模块数为 3,每个模块内卷积层数为 2;Transformer 解码器的模块数为 4;训练启用矫正阈值为 30,强化学习权重系数为 1,放缩因子为 20。字典中的可选操作共 830 项,包含词元、布尔运算操作以及终止符。本小节展示所得的实验结果。

3.2.1 合成数据集结果

目前主流的图形解析方法大多依赖于 CSGNet 预训练模型,然而由于本文的合成数据集中包含了 CSGNet 数据集未涉及的基元,因此在新

数据集上评估 CSGNet 算法并不具备代表性意义。分别在各自的数据集上训练并测试,得到结果如图 4 所示。从视觉效果定性来看,本文的方法与 CSGNet 都能还原原始图像的基本轮廓形状,但在细节方面本文的方法与原始图像更接近。为检验模型对于不同长度构造序列的解析鲁棒性,在表 2 中分别展示模型在真值构造序列长度 T 为短-中-长的 3 个子数据集上的衡量指标数值。从结果可以看出,本文提出的模型表现稳定且大幅优于 CSGNet,涵盖了其模型的功能,为第二阶段的微调学习提供了更高质量的初始预训练模型。

3.2.2 CAD 数据集结果

CSGNet-RL、SWRNet 和 PLAD 方法均采用 CSGNet 预训练模型进行后续的微调学习,最终与本文提出的模型在统一的测试框架下应用于 CAD 测试集,得到视觉结果如图 5 所示。从定性角度来看,本文的方法能够更准确地还原未知的 CAD 形状,新增的基元类型显著扩展了基元组合的可能性,同时使得生成结构更加合理。这表明本文提出的模型能够更高效地提取目标形状的几何特征,并利用注意力机制在生成序列中学习到的更广泛的关联性,从而具备对复杂形状更强的表示和解析能力。此外,表 3 展示了各方法在 CAD 数据集上的定量指标,本文的方法在各项指标上均优于当前其它主流方法。这表明模型各模块能够在解析过程中充分利用图形形状信息与序列上下文信息,合理地组装基元,并且具备避免无效序

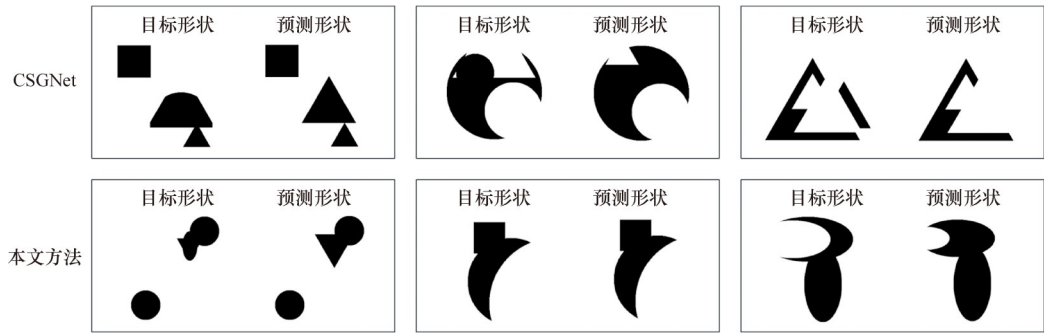


图 4 合成数据集测试结果对比

Fig. 4 Comparison of test results on synthetic dataset

表 2 不同序列长度合成数据测试结果对比

Table 2 Comparison of test results on synthetic data with different sequence lengths

	$T \leq 5$			$5 < T \leq 9$			$9 < T \leq 13$		
	$L_{ce} \downarrow$	IOU $\uparrow$	CD $\downarrow$	$L_{ce} \downarrow$	IOU $\uparrow$	CD $\downarrow$	$L_{ce} \downarrow$	IOU $\uparrow$	CD $\downarrow$
CSGNet	22.865	0.289	7.385	33.082	0.266	7.238	45.356	0.308	5.320
本文方法	2.706	0.631	1.967	3.152	0.603	2.103	3.270	0.580	2.218

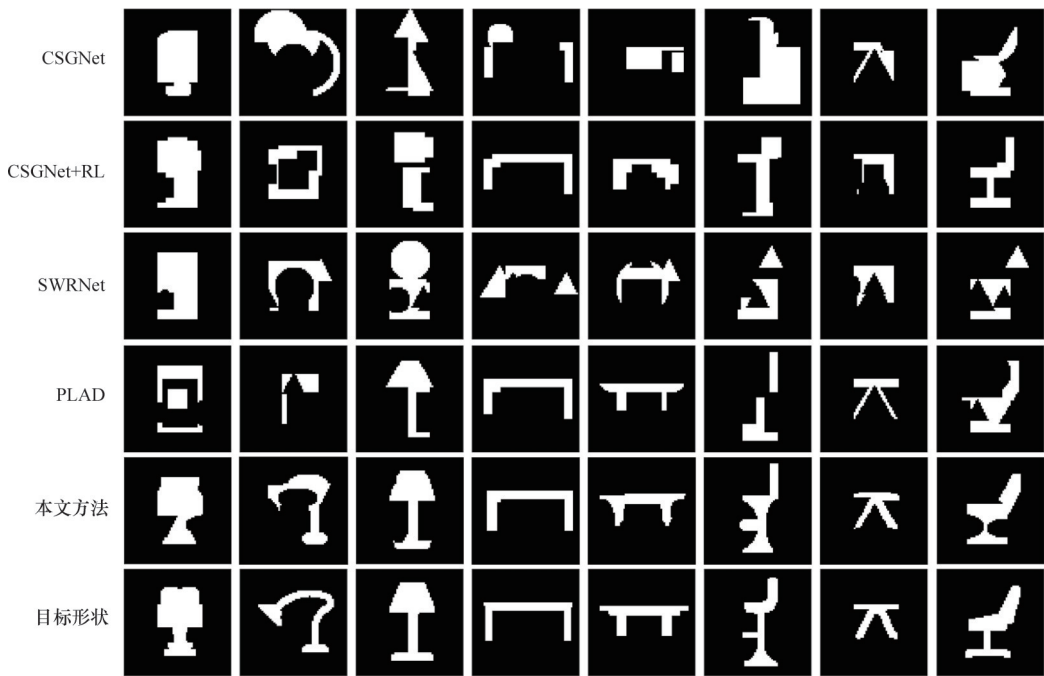


图 5 CAD数据集测试结果对比

Fig. 5 Comparison of test results on CAD dataset

表 3 CAD测试集上各项指标数值结果对比			
Table 3 Comparison of numerical results for different metrics on CAD dataset			
	IOU ↑	CD ↓	R ↑
CSGNet	0.447	2.304	0.614
CSGNet-RL	0.626	1.120	0.768
SWRNet	0.551	1.402	0.748
PLAD	0.699	0.903	0.821
本文方法	0.733	0.765	0.846

列生成的能力。对于单个形状而言,即不会出现因一步预测错误而导致整个序列变为无效的问题。

综上所述,本文提出的算法在合成和CAD数据集上都取得了更好的效果,进一步验证了其在实际应用中的潜力和优势,为复杂几何形状的解析和表示提供了新的思路。

3.3 消融实验

为验证本文所提出模型各部分结构的合理性和有效性,对各模块进行消融实验研究。表 4 展示了模型去除不同模块后的测试数值结果。实验结果表明,各模块在模型性能中均发挥了重要作用,去除任意一个模块都会导致测试结果的显著下降。具体来说,基元扩充增加了模型的表达能力,使其能够充分利用多样化的基元组合;VGG-Transformer 解析结构帮助模型更好地理解目标图

表 4 消融实验数值结果					
Table 4 Ablation experiment results					
数据 扩充	VGG-Transformer 结构	预测矫正 模块	IOU ↑	CD ↓	R ↑
×	×	×	0.699	0.903	0.821
√	×	×	0.721	0.796	0.840
×	√	×	0.712	0.850	0.830
√	√	×	0.732	0.770	0.845
√	√	√	0.733	0.765	0.846

像特征,并有效捕捉序列间的长程依赖;预测矫正模块则能在生成过程中及时纠正错误,避免无效序列的产生。这些模块的合理组合和协同作用是提升模型性能的关键因素。

此外,预测矫正模块的另一个优势是在第一阶段有监督训练中得到更好的预训练模型,本文对比了原始模型和加入预测矫正模块后,模型于训练过程中在训练集和验证集上的表现。训练集上交叉熵损失变化曲线和验证集上预测概率与真值的余弦相似度值变化曲线如图 6 所示。可以看到,在加入矫正模块后,有效提升了训练过程中的指标,其中训练损失下降约 30%,余弦相似度稳定值提升约 15%。以上结果表明,预测矫正模块给预训练模型带来了有效改进。

从实验中可以发现,预测矫正模块在性能较弱的模型上对指标的提升更明显,这是因为较弱

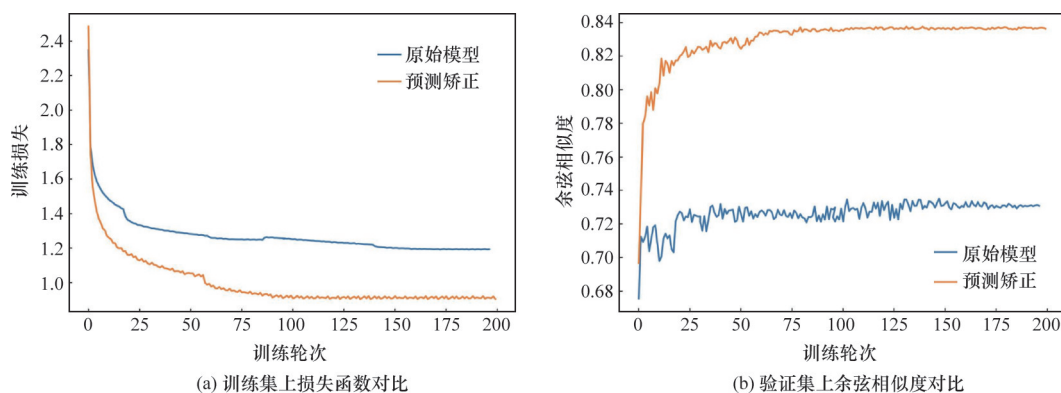


图6 原始模型与采用预测矫正结果对比

Fig. 6 Comparison of the original model with the prediction correction model

模型通常未学习到有效的语法规则,导致生成大量无效序列,因此引入预测矫正至少保证了模型所生成的程序能够通过图形渲染器得到形状。相对应地,对于性能较好,已经学习到语法规则的模型来说,该模块修正了少量的错误预测,指标提升幅度会相对小一些。

4 总结

本文提出一种基于Transformer的自回归深度学习架构,用于解析二维CAD图形的CSG序列。通过扩展训练数据集、采用VGG编码器和Transformer解码器的结合架构、并引入有效性矫正模块,在二维CSG建模任务中取得了显著的进展。实验结果表明,所提出的方法在合成数据集和真实CAD数据集上的表现均优于主流方法。

自动生成CSG序列能够降低复杂形状建模过程中对专业人员手动设计的依赖,能够提高设计的效率和灵活性。然而,当前的工作也面临一些挑战与局限性。例如,由于CSG表示的非唯一性,同一几何形状可能对应多种不同的CSG构造方式,这给模型的训练和应用带来了一定的复杂性。同时,如何进一步提升模型在更复杂形状解析中的鲁棒性,减少模型误差的累积,依然是未来需要深入研究的方向。

未来的工作将从以下几个方面开展:首先,进一步扩展训练数据集,增加不同领域和复杂度的CAD形状,以增强模型的泛化能力;其次,当前模型的序列生成逻辑与人类设计师存在较大差异,因此需要研究如何在解析过程中引入设计知识和语义信息,使得模型不仅能够高效解析几何形状,还能在生成过程中考虑设计意图和功能要求,为智能设计提供更有力的支持。此外,期望进一步

将二维形状解析方法推广至三维领域,例如先解析二维投影草图,并根据解析结果指导后续平面选择或三维拉伸等步骤。

参考文献

- [1] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. December 4 - 9, 2017, Long Beach, California, USA. New York: ACM, 2017: 6000-6010. DOI: 10.5555/3295222.3295349.
- [2] Sharma G, Goyal R, Liu D F, et al. CSGNet: neural shape parser for constructive solid geometry [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 5515-5523. DOI:10.1109/CVPR.2018.00578.
- [3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. arXiv 2014: 1409.1556. (2014-09-04) [2024-12-06]. <https://arxiv.org/abs/1409.1556v6>.
- [4] Fischler M A, Elschlager R A. The representation and matching of pictorial structures [J]. IEEE Transactions on Computers, 1973, C-22 (1): 67-92. DOI: 10.1109/T-C.1973.223602.
- [5] Felzenszwalb P F, Huttenlocher D P. Pictorial structures for object recognition [J]. International Journal of Computer Vision, 2005, 61 (1): 55-79. DOI: 10.1023/B:VISI.0000042934.15159.49.
- [6] Yang Y, Ramanan D. Articulated pose estimation with flexible mixtures-of-parts [C]//CVPR 2011. June 20-25, 2011, Colorado Springs, CO, USA. IEEE, 2011: 1385-1392. DOI:10.1109/CVPR.2011.5995741.
- [7] Weiss D. Geometry-based structural optimization on CAD specification trees [D/OL]. Zurich: ETH Zurich, 2009 [2024-12-06]. <https://www.research-collection.ethz.ch/handle/20.500.11850/45733>.
- [8] Foley J, Shapiro V, Vossler D L. Separation for boundary to CSG conversion [J]. ACM Transactions on Graphics, 1993, 12(1): 35-55. DOI: 10.1145/169728.169723.

- [9] Buchele S F, Crawford R H. Three-dimensional halfspace constructive solid geometry tree construction from implicit boundary representations [C]//Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications. June 16 - 20, 2003, Seattle, Washington, USA. ACM, 2003: 135-144. DOI:10.1145/781606.781629.
- [10] Balog M, Gaunt A, Brockschmidt M, et al. DeepCoder: learning to write programs [EB/OL]. arXiv 2016: 1611.01989. (2016-11-07) [2024-12-06]. <https://arxiv.org/abs/1611.01989>.
- [11] Johnson J, Hariharan B, Van Der Maaten L, et al. Inferring and executing programs for visual reasoning [C]//2017 IEEE International Conference on Computer Vision (ICCV). October 22-29, 2017, Venice, Italy. IEEE, 2017: 3008-3017. DOI:10.1109/ICCV.2017.325.
- [12] Zhou C H, Li C L, Póczos B. Unsupervised program synthesis for images by sampling without replacement [EB/OL]. arXiv 2020: 2001.10119. (2020-01-21) [2024-12-06]. <https://arxiv.org/abs/2001.10119>.
- [13] Jones R K, Walke H, Ritchie D. PLAD: learning to infer shape programs with pseudo-labels and approximate distributions [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 9861-9870. DOI:10.1109/CVPR52688.2022.00964.
- [14] Li L X, Sung M, Dubrovina A, et al. Supervised fitting of geometric primitives to 3D point clouds [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15-20, 2019, Long Beach, CA, USA. IEEE, 2019: 2647-2655. DOI: 10.1109/CVPR.2019.00276.
- [15] Lambourne J G, Willis K D D, Jayaraman P K, et al. BRepNet: a topological message passing system for solid models [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 12768-12777. DOI:10.1109/CVPR46437.2021.01258.
- [16] Kania K, Zięba M, Kajdanowicz T. UCSG-net: unsupervised discovering of constructive solid geometry tree [EB/OL]. arXiv 2020: 2006.09102. (2020-06-16) [2024-12-06]. <https://arxiv.org/abs/2006.09102v3>.
- [17] Ren D X, Zheng J M, Cai J F, et al. CSG-stump: a learning friendly CSG-like representation for interpretable shape parsing [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 12458-12467. DOI: 10.1109/ICCV48922.2021.01225.
- [18] Chen Z Q, Tagliasacchi A, Zhang H. BSP-net: generating compact meshes via binary space partitioning [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 42-51. DOI:10.1109/cvpr42600.2020.00012.
- [19] Yu F G, Chen Z Q, Li M Y, et al. CAPRI-net: learning compact CAD shapes with adaptive primitive assembly [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 11758-11768. DOI: 10.1109/CVPR52688.2022.01147.
- [20] Yu F G, Chen Z Q, Tanveer M, et al. D²CSG: unsupervised learning of compact CSG trees with dual complements and dropouts [EB/OL]. arXiv 2023: 2301.11497. (2023-01-27) [2024-12-06]. <https://arxiv.org/abs/2301.11497>.
- [21] Wu R D, Xiao C, Zheng C X. DeepCAD: a deep generative network for computer-aided design models [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 6752-6762. DOI:10.1109/ICCV48922.2021.00670.
- [22] Xu X, Willis K D D, Lambourne J G, et al. SkexGen: autoregressive generation of CAD construction sequences with disentangled codebooks [EB/OL]. arXiv 2022: 2207.04632. (2022-07-11) [2024-12-06]. <https://arxiv.org/abs/2207.04632v1>.
- [23] Khan M S, Dupont E, Ali S A, et al. CAD-SIGNet: CAD language inference from point clouds using layer-wise sketch instance guided attention [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 4713-4722. DOI:10.1109/CVPR52733.2024.00451.
- [24] Yang B C, Jiang H Y, Pan H, et al. PS-CAD: local geometry guidance via prompting and selection for CAD reconstruction [EB/OL]. arXiv 2024: 2405.15188. (2024-05-24) [2024-12-06]. <https://arxiv.org/abs/2405.15188v1>.
- [25] 何皓辰,方正,卢政达,等. PT-MFR: 一种基于 Point Transformer的CAD模型加工特征识别方法[J]. 中国科学院大学学报, 2026, 43(1): 115-124. DOI: 10.7523/j.ucas.2024.041.
- [26] 黄玉林,梁磊,李卫军,等. 基于多尺度特征和注意力机制的深度学习点云压缩[J]. 中国科学院大学学报, 2024, 41(5): 687-694. DOI:10.7523/j.ucas.2023.077.