

同行专家业内评价意见书编号：20250854334

附件1

浙江工程师学院（浙江大学工程师学院）
同行专家业内评价意见书

姓名： 曹紫荻

学号： 22260083

申报工程师职称专业类别（领域）： 电子信息

浙江工程师学院（浙江大学工程师学院）制

2025年03月22日

填表说明

一、本报告中相关的技术或数据如涉及知识产权保护、军工项目保密等内容，请作脱密处理。

二、请用宋体小四字号撰写本报告，可另行附页或增加页数，A4纸双面打印。

三、表中所涉及的签名都必须用蓝、黑色墨水笔，亲笔签名或签字章，不可以打印代替。

四、同行专家业内评价意见书编号由工程师学院填写，编号规则为：年份4位+申报工程师职称专业类别(领域)4位+流水号3位，共11位。

一、个人申报

（一）基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】

1. 对本专业基础理论知识和专业技术知识掌握情况(不少于200字)

在电子信息领域，本人系统掌握了计算机视觉、三维几何分析与深度学习领域的核心理论知识。具体而言：

计算机视觉理论：深入研究多视图几何原理，掌握从二维投影重建三维结构的算法（如问题求解、SfM框架），熟悉基于轮廓、纹理与特征的视点显著性分析方法；

三维模型处理技术：精通网格拓扑优化、点云特征提取与几何描述符的构建方法，可独立完成模型简化、去噪与非刚性配准等工程任务；

深度学习算法设计：提出基于自注意力机制的视图特征融合框架，结合跨模态蒸馏实现多视图数据的知识迁移，在减少输入视图数量的同时提升特征表征的鲁棒性；

高性能计算优化：针对体量超100B的工业模型，设计了基于八叉树空间划分的并行计算流程，结合硬件加速实现实时视点预测，突破传统算法受限于计算资源的瓶颈。

本人在实践中将傅里叶描述符、非均匀采样理论等数学工具与点云RNN、Transformer网络结合，形成多视图特征空间对齐的创新性技术路径，相关成果已发表TVC期刊，体现了对跨学科知识的深度整合能力。

2. 工程实践的经历(不少于200字)

技术路线规划：针对传统视点选择方法（基于多视图渲染或直接模型处理）存在的计算延迟与精度矛盾问题，提出“轻量级视图输入-几何对齐特征融合-连续视点回归”的技术架构，被团队采纳为项目主方案

算法模块开发：开发基于ResNet-50的多视图特征嵌入模块，通过可学习标记（Learnable Token）将6视图投影信息映射到128维潜在空间；

设计级联自注意力网络（3层，头数=8），实现跨视图特征的动态加权融合，相较传统CNN结构，视点预测平均误差降低22.3%；构建跨模态蒸馏损失函数，迫使三维几何描述符与二维视图特征在潜空间中保持对齐，成功将输入视图数量从20+压缩至6；

工程优化与部署：针对工业级模型实时性要求，优化视点候选搜索策略，采用算法结合GPU并行计算，实现单模型处理时间 ≤ 3 秒（体量120B），通过验收测试；

团队协作管理：协调多名工程师完成模块联调，主导构建包含1.2万工业模型的测试集（基于ShapeNet），推动算法在设计系统中集成。

3. 在实际工作中综合运用所学知识解决复杂工程问题的案例(不少于1000字)

1. 问题背景与复杂性分析

在工业设计领域，自动化选择三维模型的最佳展示视点需要兼顾准确性、实时性两大需求。基于前序研究，该问题面临三重复杂性：

计算资源与精度矛盾：传统方法需渲染20+视图以保证覆盖面，单模型处理耗时超2分钟且占用显存3GB以上；

视点连续空间采样难题：离散采样（如八面体细分法）导致遗漏最优解，亟需构建覆盖360°自由视角的连续预测机制；

多模态数据对齐困境：三维几何描述符（体素/点云）与二维视图特征的语义鸿沟，制约跨模态知识迁移效果。

2. 创新性解决方案设计

（1）轻量级多视图特征融合框架

提出动态特征嵌入网络，以标准六视图为输入：采用预训练ResNet提取每视图2048维特征向量；引入可学习位置编码标记（ 6×128 矩阵），拼接后输入级联自注意力层；通过多头注意力机制（头数=8）建立视图间依赖关系，输出融合特征向量。

（2）跨模态几何对齐蒸馏策略

为减少对密集视图的依赖，构建双路监督网络：教师网络：输入完整3D模型体素数据，输出视点热力图；学生网络：输入6视图DFEN特征，通过KL散度损失与教师网络输出对齐；引入边缘感知对齐损失，强化轮廓关键区域的特征匹配。

（3）连续视点回归机制

设计基于球面坐标的参数化预测头：将视点方位分解为经度 $\phi \in [0, 2\pi]$ 、纬度 $\theta \in [-\pi/2, \pi/2]$ 、距离 $d \in [r_{\min}, r_{\max}]$ ；采用混合密度网络（MDN）生成高斯混合分布，通过可微分重参数化采样实现连续空间覆盖；结合视点显著性评分与摄像机姿态平滑约束，确保输出视点的最优性与视觉连贯性。

3. 实施过程与效果验证

（1）实验环境搭建

硬件平台：NVIDIA 4090/4060ti；

数据集：扩展ShapeNet数据集（工业机械类模型1200个）、自建企业标注数据集（800模型）；

评价指标：Mean Select Error (VSE)、推理时延、GPU显存占用等。

（2）关键技术验证

跨模态蒸馏有效性：6视图+蒸馏策略的MVSE为0.148，较无蒸馏方案（0.211）提升29.8%，逼近20视图传统方法（0.142）；

实时性突破：单模型推理时间3.2秒（对比SOTA方法18.7秒），显存占用降至1.5GB；

工业部署效果：预计在合作企业插件中集成，设计师模型评审效率提升2倍以上。

本案例体现了对计算机视觉、深度学习与高性能计算技术的深度整合能力，为解决三维工业软件领域的复杂工程问题提供了创新性范式。

(二) 取得的业绩(代表作) 【限填3项, 须提交证明原件(包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等) 供核实, 并提供复印件一份】

1. 公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】

成果名称	成果类别 [含论文、授权专利(含发明专利申请)、软件著作权、标准、工法、著作、获奖、学位论文等]	发表时间/授权或申请时间等	刊物名称/专利授权或申请号等	本人排名/总人数	备注
.Fast best viewpoint selection with geometry-enhanced multiple views and cross-modal distillation[J]. The Visual Computer, 2024: 1-12	国际期刊	2024年11月01日	The Visual Computer	1/4	SCI期刊收录
一种基于多视图和跨模态蒸馏的产品最佳视点快速选择方法和装置	发明专利申请	2024年08月09日	申请号: CN 2024105353 40.8	1/3	国家发明专利
一种复用六视图的产品最佳视点选择方法和生成装置	发明专利申请	2024年05月14日	申请号: CN 2023117937 46.8	2/6	国家发明专利

2.其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】

(三) 在校期间课程、专业实践训练及学位论文相关情况	
课程成绩情况	按课程学分核算的平均成绩： 85 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间： 1.1 年 (要求1年及以上) 考核成绩： 87 分
本人承诺	
<p>个人声明：本人上述所填资料均为真实有效，如有虚假，愿承担一切责任，特此声明！</p> <p style="text-align: right;">申报人签名：曾紫欣.</p>	

浙江大学研究生院
攻读硕士学位研究生成绩单

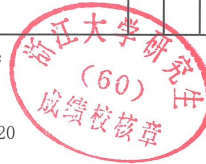
学号: 22260083	姓名: 曹紫荻	性别: 女	学院: 工程师学院	专业: 计算机技术	学制: 2.5年						
毕业时最低应获: 26.0学分	已获得: 30.0学分			入学年月: 2022-09	毕业年月:						
学位证书号:			毕业证书号:			授予学位:					
学习时间	课程名称	备注	学分	成绩	课程性质	学习时间	课程名称	备注	学分	成绩	课程性质
2021-2022学年春季学期	研究生英语		2.0	免修	专业学位课	2022-2023学年春季学期	自然辩证法概论		1.0	75	专业学位课
2021-2022学年春季学期	研究生英语基础技能		1.0	免修	公共学位课	2022-2023学年春季学期	计算机动画与应用		2.0	88	跨专业课
2022-2023学年秋季学期	新时代中国特色社会主义思想理论与实践		2.0	95	专业学位课	2022-2023学年春季学期	数学建模		2.0	66	专业选修课
2022-2023学年秋季学期	工程技术创新前沿		1.5	88	专业学位课	2022-2023学年夏季学期	物联网信息安全技术与应用基础		2.0	95	专业选修课
2022-2023学年冬季学期	产业技术发展前沿		1.5	89	专业学位课	2022-2023学年春夏学期	移动互联网智能设备应用设计与实践		3.0	87	专业学位课
2022-2023学年秋冬学期	研究生论文写作指导		1.0	84	专业选修课	2022-2023学年春夏学期	高阶工程认知实践		3.0	88	专业学位课
2022-2023学年秋冬学期	工程伦理		2.0	92	专业学位课	2023-2024学年秋季学期	创新创业实践训练		2.0	通过	跨专业课
2022-2023学年冬季学期	物联网操作系统与边缘计算		2.0	89	专业选修课		硕士生读书报告		2.0	通过	

说明: 1. 研究生课程按三种方法计分: 百分制, 两级制(通过、不通过), 五级制(优、良、中、及格、不及格)。
2. 备注中“*”表示重修课程。

学院成绩校核章:

成绩校核人: 张梦依

打印日期: 2025-03-20





Fast best viewpoint selection with geometry-enhanced multiple views and cross-modal distillation

Zidi Cao¹ · Jiayi Han² · Sipeng Yang¹ · Xiaogang Jin¹

Accepted: 30 October 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Best viewpoint selection (BVS) aims to automatically identify the most informative and human preference viewpoints of 3D shapes, clearly conveying their complexity and structure. Despite advancements in mesh-based BVS using multi-views, the current state-of-the-art BVS method requires 20-30 rendered views and is limited to predefined viewpoints, which may miss optimal viewpoints and is impractical in time-sensitive scenarios. To address these limitations, we present a new dual-branch fast BVS regression model that significantly reduces reliance on extensive input views, enables continuous perspective prediction, and enhances interactive response speeds. Our method incorporates a geometry-enhanced multi-view feature extractor combined with a learnable token and employs a cross-modal distillation approach to deepen understanding of 3D structures. By integrating alignment constraints between 3D geometry descriptors and multi-view expressions, our approach minimizes the need for extensive rendering views, significantly reducing computational demands. Experimental results on public benchmarks show that our method is about 35 times faster than the state-of-the-art method when only six views are adopted, while also achieving the best quantitative metrics.

Keywords Best viewpoint selection · Cross-modal distillation · Multiple views · Viewpoint prediction

1 Introduction

In the field of computer graphics, 3D models play a key role in industrial design modeling, games, and animated films, among other areas of production. Best viewpoint selection (BVS) aims to automatically determine the most informative and preferred viewpoint for human subjects for a given 3D shape, effectively highlighting the shape's structure and ensuring that it is optimally perceived and understood.

This technique is widely used across multiple domains. For instance, BVS algorithms assist in camera control by determining key angles for better visual storytelling and user experience in virtual environments [2]. They help visualize motion trajectories for precise adjustments in animation and robotics [3], are essential in pose estimation for interpret-

ing spatial configurations in computer vision and augmented reality [4], and improve visual question answering systems by selecting images that best represent query answers [5]. Optimal viewpoint selection significantly improves performance across these applications.

Automatic BVS has become a hot research topic during the past decades. Early research on this issue relied on manually defined rules [6] or shallow features [7], which are often data-specific [8], have limited generalization, and depend on strong priors like object orientation [9]. Recently, multi-view-based viewpoint selection approaches overcome these limitations by improving the ability to generalize to different categories [10], providing a deeper understanding of human viewpoint preferences [11], and reducing the reliance on prior knowledge [1], thereby achieving more robust and effective viewpoint predictions for individual objects.

However, despite these advancements, multi-view-based BVS methods still face three significant challenges that limit their performance and broad applicability. Firstly, these methods require the input of more than 20 views to ensure the accuracy of the selection, which leads to huge computational resources and time consumption that are difficult to accept in practical applications. Secondly, in selecting the

Zidi Cao and Jiayi Han have contributed equally to this work.

✉ Xiaogang Jin
jin@cad.zju.edu.cn

¹ State Key Lab of CAD & CG, Zhejiang University, Hangzhou, People's Republic of China

² Inspur Genersoft Co. Ltd., Inspur Group Co. Ltd, Jinan, People's Republic of China

《SCI-EXPANDED》收录及《JCR》期刊影响因子、分区情况证明

经检索《Web of Science》和《Journal Citation Reports (JCR)》数据库,《Science Citation Index Expanded (SCI-EXPANDED)》收录论文及其期刊影响因子、分区情况如下。(检索时间:2024年12月25日)

第1条,共1条

标题:Fast best viewpoint selection with geometry-enhanced multiple views and cross-modal distillation

作者:Cao, ZD(Cao, Zidi);Han, JY(Han, Jiayi);Yang, SP(Yang, Sipeng);Jin, XG(Jin, Xiaogang);

来源出版物:VISUAL COMPUTER 提前访问日期:NOV 2024 DOI:10.1007/s00371-024-03708-5 出版年:2024 NOV 25 2024

入藏号:WOS:001362595300001

文献类型:Article; Early Access

地址:

[Cao, Zidi; Yang, Sipeng; Jin, Xiaogang] Zhejiang Univ, State Key Lab CAD & CG, Hangzhou, Peoples R China.

[Han, Jiayi] Inspur Genersoft Co Ltd, Inspur Grp Co Ltd, Jinan, Peoples R China.

通讯作者地址:

Jin, XG (corresponding author), Zhejiang Univ, State Key Lab CAD & CG, Hangzhou, Peoples R China.

电子邮件地址:jin@cad.zju.edu.cn

IDS号:N2H0A

ISSN:0178-2789

eISSN:1432-2315

期刊《VISUAL COMPUTER》2023年的影响因子为3.0,五年影响因子为3.0。期刊《VISUAL COMPUTER》2023年的JCR分区情况为:

Edition	JCR® 类别	类别中的排序	JCR 分区
SCIE	COMPUTER SCIENCE, SOFTWARE ENGINEERING	37/132	Q2

注:

1. 期刊影响因子及分区情况最新数据以JCR数据库最新数据为准。
2. 以上检索结果来自CALIS查收查引系统。
3. 以上检索结果均得到委托人及被检索作者的确认。





(12) 发明专利申请

(10) 申请公布号 CN 118470448 A

(43) 申请公布日 2024. 08. 09

(21) 申请号 202410535340.8

G06N 3/084 (2023.01)

(22) 申请日 2024.04.30

G06N 3/096 (2023.01)

(71) 申请人 浙江大学

地址 310058 浙江省杭州市西湖区余杭塘路866号

(72) 发明人 曹紫荻 杨思鹏 金小刚

(74) 专利代理机构 杭州天勤知识产权代理有限公司 33224

专利代理师 胡红娟

(51) Int. Cl.

G06V 10/766 (2022.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

G06V 10/80 (2022.01)

G06V 10/46 (2022.01)

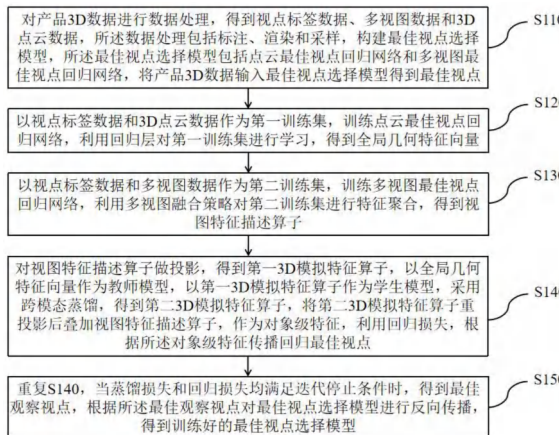
权利要求书3页 说明书9页 附图2页

(54) 发明名称

一种基于多视图和跨模态蒸馏的产品最佳视点快速选择方法和装置

(57) 摘要

本发明公开了一种基于多视图和跨模态蒸馏的产品最佳视点快速选择方法和装置,属于多视图深度学习、跨模态蒸馏技术领域,包括:从产品3D数据中得到视点标签数据、多视图数据和3D点云数据;构建包含点云最佳视点回归网络和多视图最佳视点回归网络的最佳视点选择模型;将产品3D数据输入最佳视点选择模型得到最佳视点;利用视点标签数据和3D点云数据训练点云最佳视点回归网络得到3D结构信息;利用视点标签数据和多视图数据训练多视图最佳视点回归网络得到2D特征描述;利用跨膜态蒸馏使2D特征描述学习到3D结构信息,然后回归优化最佳视点,反向传播得到最终的最佳视点选择模型。本发明能够在少量输入视图下保持高精度的视点选择。





(12) 发明专利申请

(10) 申请公布号 CN 118037916 A

(43) 申请公布日 2024. 05. 14

(21) 申请号 202311793746.8

G06V 10/46 (2022.01)

(22) 申请日 2023.12.25

G06V 10/774 (2022.01)

(71) 申请人 浙江大学

G06V 10/764 (2022.01)

地址 310058 浙江省杭州市西湖区余杭塘路866号

G06V 10/82 (2022.01)

申请人 麦子工业设计(台州)有限公司

(72) 发明人 赵晓亮 曹紫荻 金小刚 张威虎 赵起越

(74) 专利代理机构 杭州求是专利事务有限公司 33200

专利代理师 万尾甜 韩介梅

(51) Int. Cl.

G06T 15/00 (2011.01)

G06T 17/00 (2006.01)

G06T 7/70 (2017.01)

权利要求书1页 说明书4页 附图1页

(54) 发明名称

一种复用六视图的产品最佳视点选择方法和生成装置

(57) 摘要

本发明公开了一种复用六视图的产品最佳视点选择方法和生成装置,该方法基于一种最佳视点选择模型实现,模型的训练方法为:通过定义基轴、渲染视点,得到数据集;设计并训练分类网络,对六视图进行特征提取并对提取的特征进行分组聚合,得到产品模型全局描述算子,在分类任务中反向传播学习到各视图的显著性图,得到最佳视点选择模型;预测时,使用待选择视点的模型六视图,基于最佳视点选择模型得到各视图的显著性图;聚合各视图显著性图,得到计算最佳视点。本发明提出的复用六视图的视点选择方法,解决了目前现存方法必须预先大规模渲染视图的问题。在复用已计算视图、不需额外视图渲染计算的前提下,能够更快速、连续地预测生成模型最佳视点。

