统一3D场景理解与生成模型

VGGT扩展实现新视角生成

许湘明

CONTENT

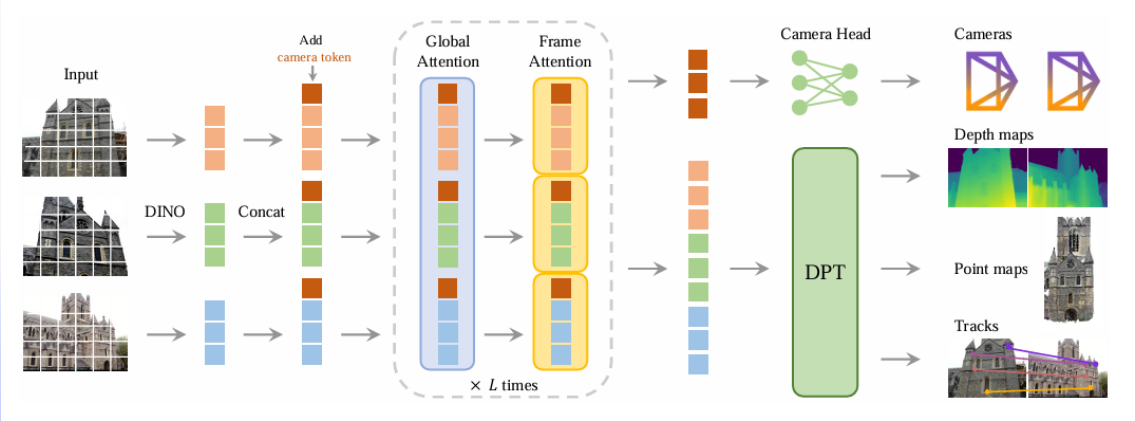
目录

01

相关工作进展

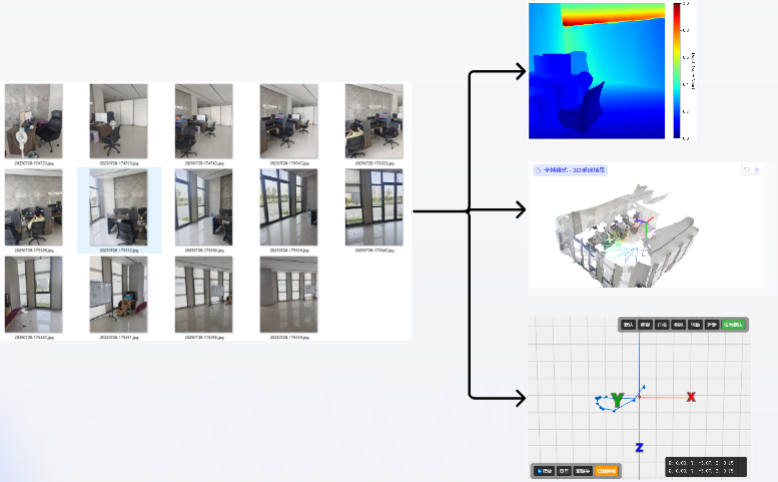
3D场景理解相关工作

调研了相关研究工作，主要以VGGT基础进行研究扩展
2025 CVPR BSET PAPER(facebook)
end-end feed-forward model



3D场景理解相关工作

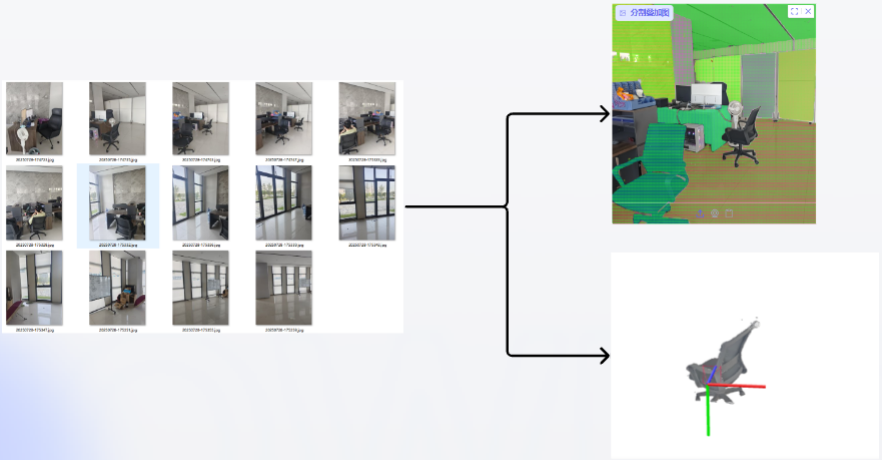
VGGT主要预测的深度任务:相机位姿，深度，点云，点轨迹跟踪



3D场景理解相关工作

VGGT主要预测的深度任务:相机位姿，深度，点云，点轨迹

扩展融合SAM，使其具有目标分割理解能力



3D场景理解小结

从传统的SFM特征匹配+几何计算方法，转到完全基于模型与数据驱动的通用范式已初见效果，其通用性，感知精度都已经达较高水准。

但是VGGT并不是3D场景理解的终点，基于它可以把3D场景理解扩展到更深更广的应用。

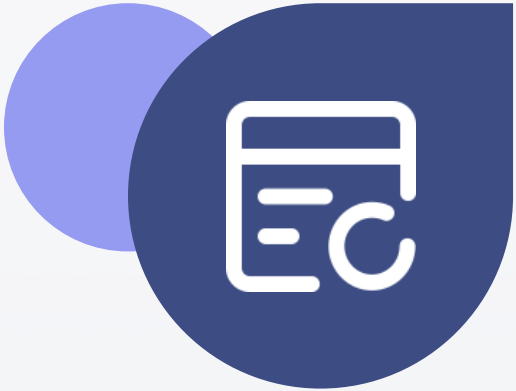
02

统一的3D场景理解与生成模型
--扩展VGGT至新视角生成

解决的问题

VGGT能做相机参数预测，深度，点云，但不能生成新视角图像

Neurf，GS Splat能做3d新视角渲染效果突出，但依赖离线训练和外部SfM点云（如COLMAP）

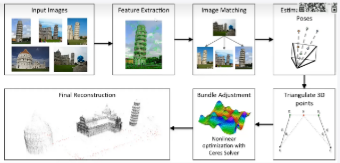


提出一个统一的3D场景理解与生成模型框架，扩展VGGT的模型支持新视角生成，使模型不仅支持多视角RGB输入，还可接收目标相机位姿作为条件，生成物理一致的新视角图像

研究背景

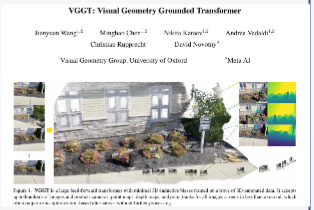
3D场景理解现状

3D场景理解任务
三阶段



特征匹配+几何计算
slamers:BA+全局优化

双视图frame pair wise
end-end model几何估计

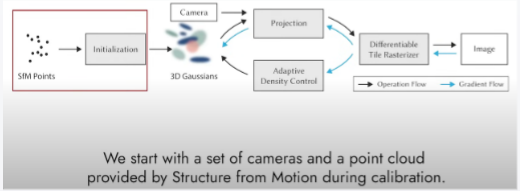


支持大规模RGB输入
完全端到端的3D任务输出

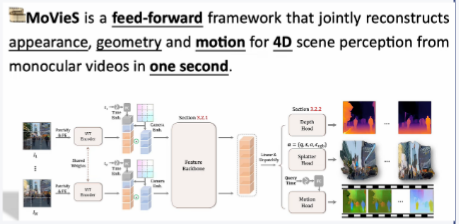
研究背景

3D场景生成

主流方案一



3D GS Splat
基于点云的场景渲染
开源

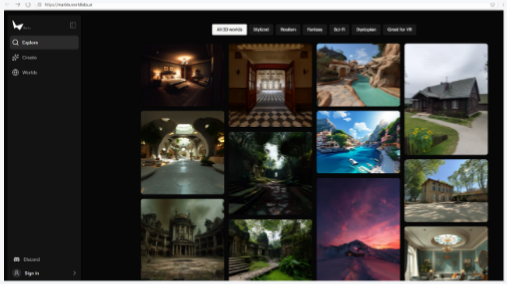


feed-forward 3d场景生成
基于时间序列的生成模型
开源

研究背景

3D场景
生成

主流方案二
一张图生成3D世界



marble
李飞飞空间智能
不开源

genie3
谷歌deepmind
genie3开始闭源

研究动机

统一的3D场景理解与生成模型，能理解3D场景并且能精准控制生成模拟3D场景，为机器人应用奠定感知模拟基础

03

技术方案与挑战

研究进展

目前方案设计与框架，完成了代码

正在验证，目前小数据集上已经能看到有生成效果，但是还未收敛

研究进展

目前实验效果

优化计划

1.增大数据集，增加训练时长

2.继续局部尝试更优模型设计

感谢观看！

THANKS!