





香港科技大学(广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

三维大模型 年度进展报告

 陈颖聪 助理教授

 人工智能学域

 香港科技大学 (广州)

大模型时代

 OpenAI



文字大模型
(ChatGPT)



图像大模型
(SD/MJ)

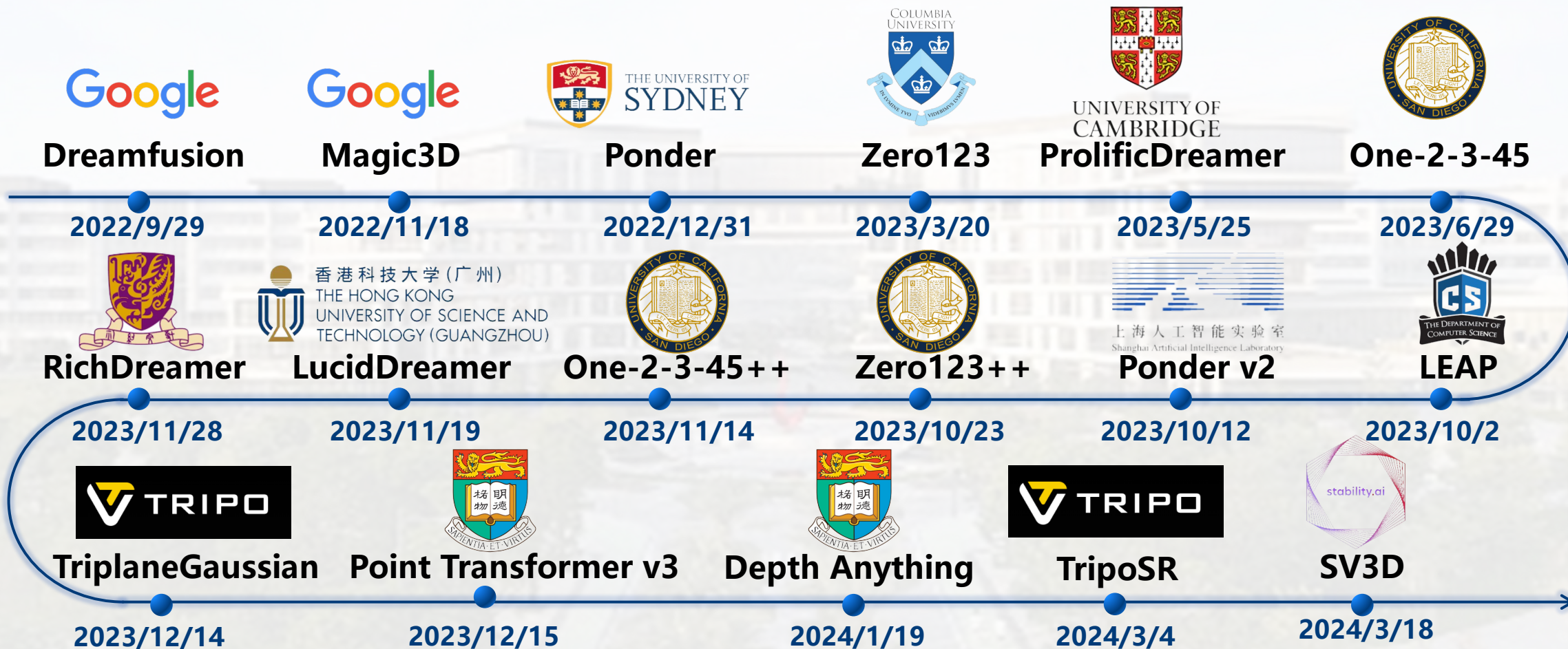


视频大模型
(Sora)

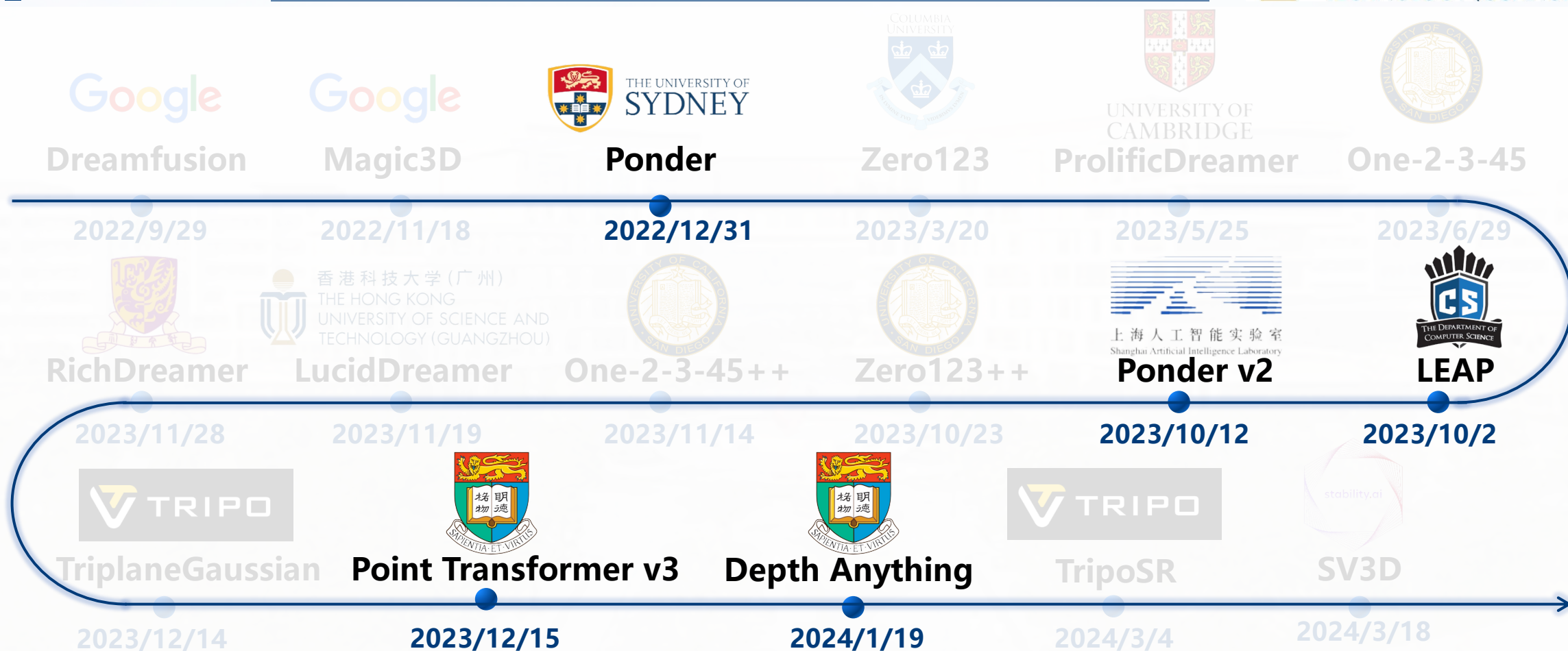


三维大模型

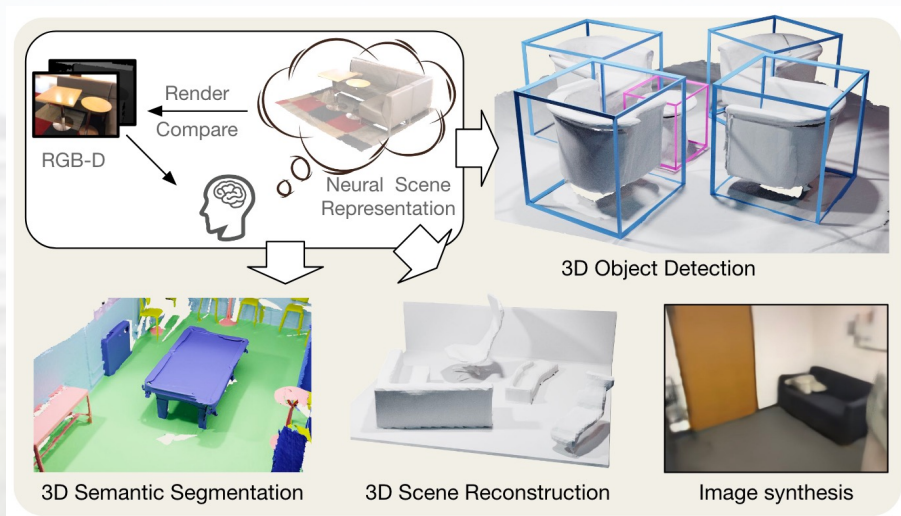
关键技术进展



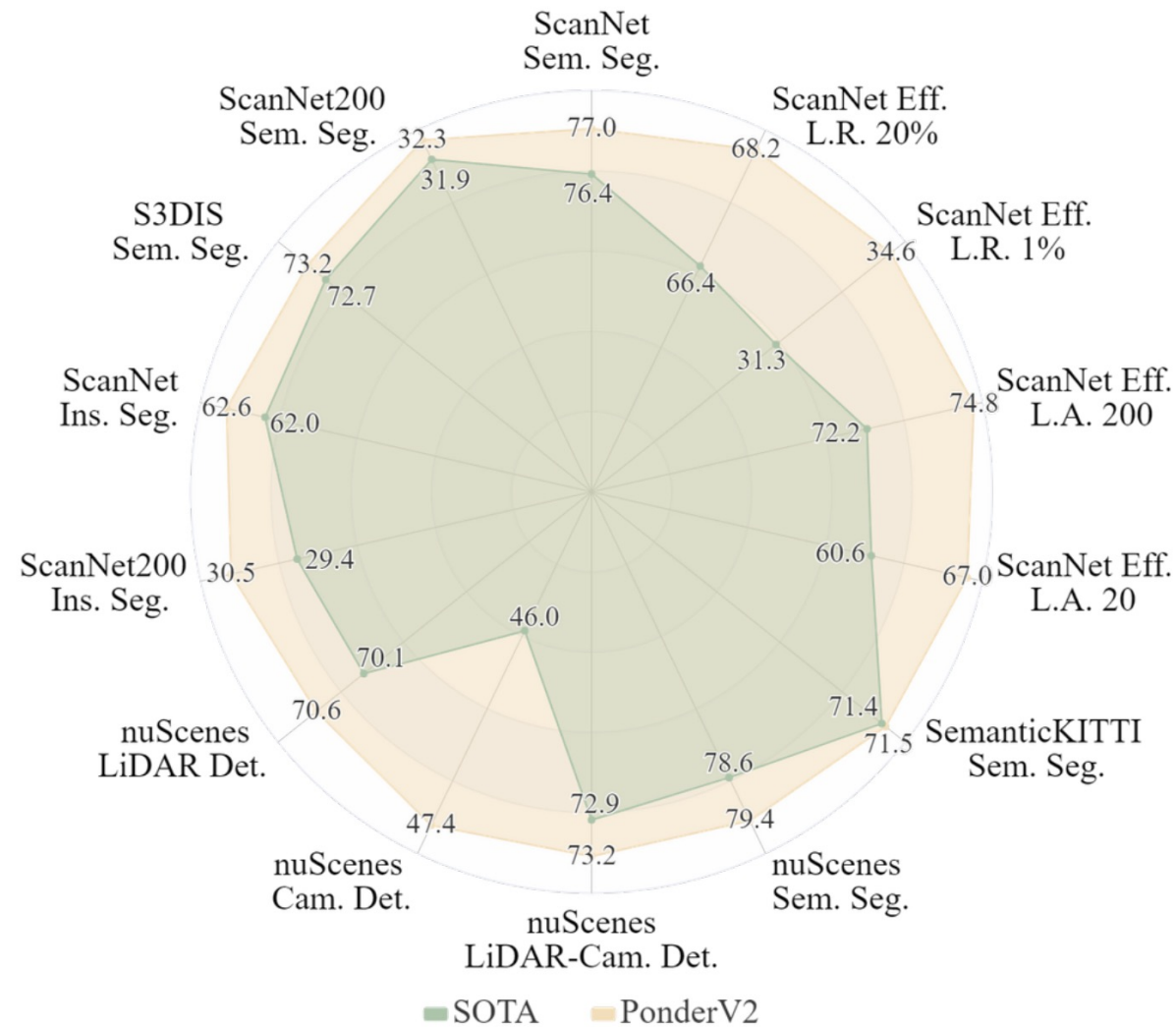
关键技术进展



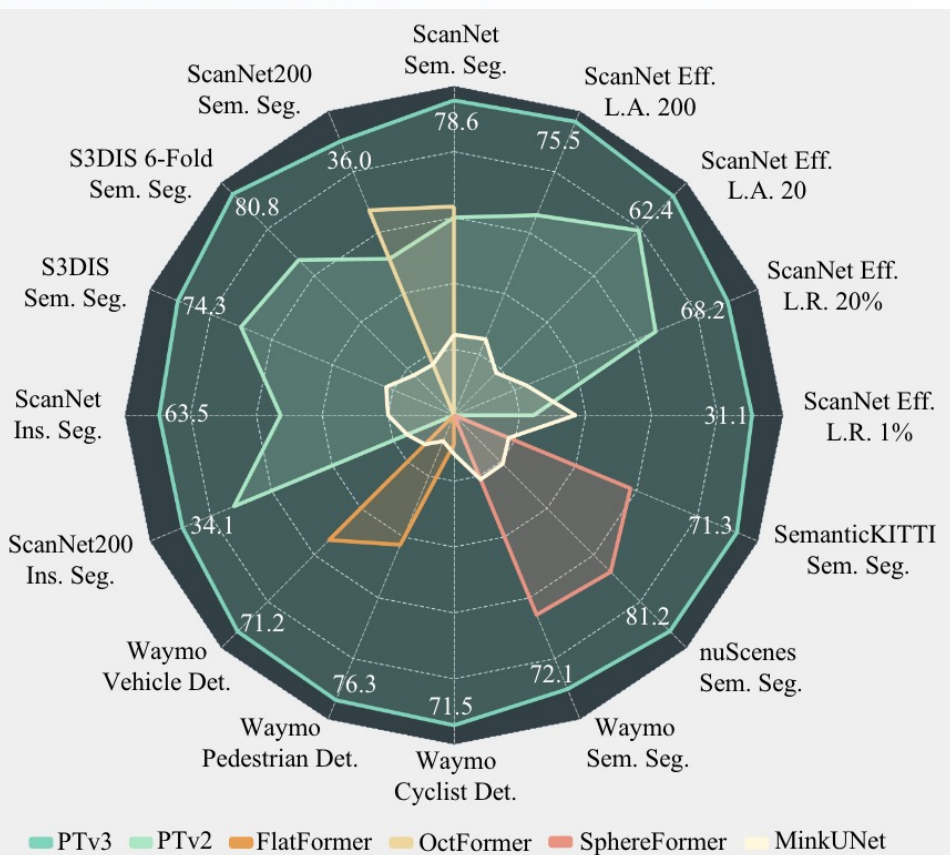
Ponder v2



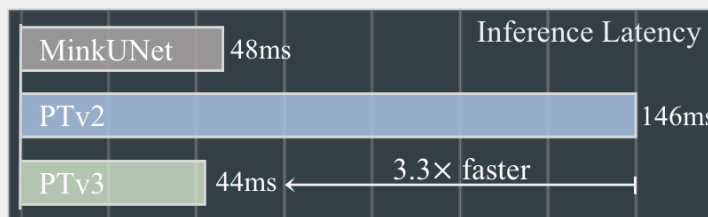
- 三维大模型预训练: Neural Rendering
- 其三维表示适配多种下游任务, 包括3D检测、3D分割, 3D重建、3D生成等
- 在11个涵盖室内外的基准测试中取得最佳效果



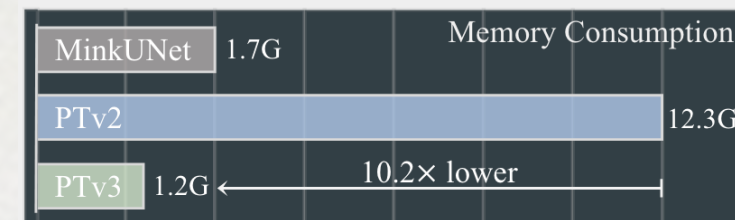
Point Transformer V3: Simpler, Faster, Stronger



- 在大模型时代，三维骨干网络的精度受规模化程度而非模型结构主导
- 其设计应以简单与效率为先
- 相比上一版，速度提升3.3倍，内存节省10.2倍
- 在超过20个下游任务中取得最先进的性能

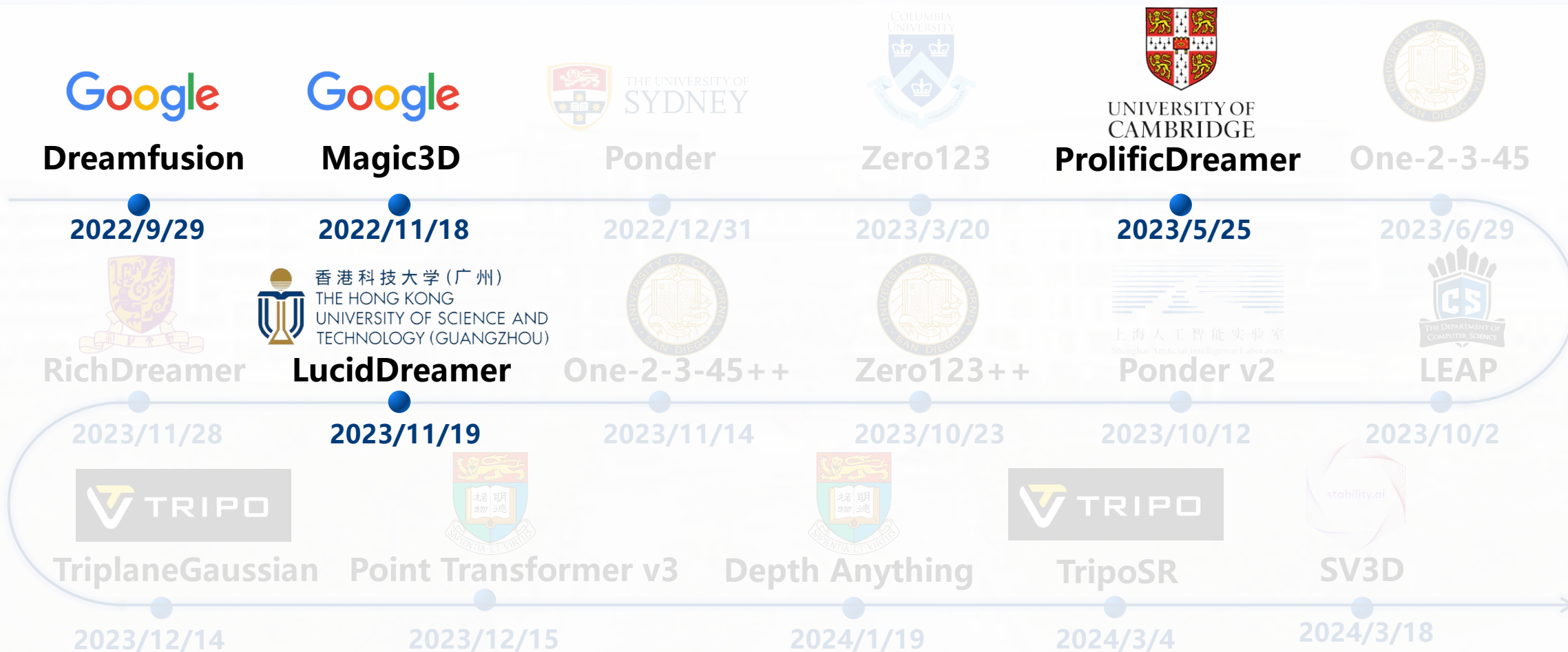


Faster Speed



Lower Memory Consumption

关键技术进展

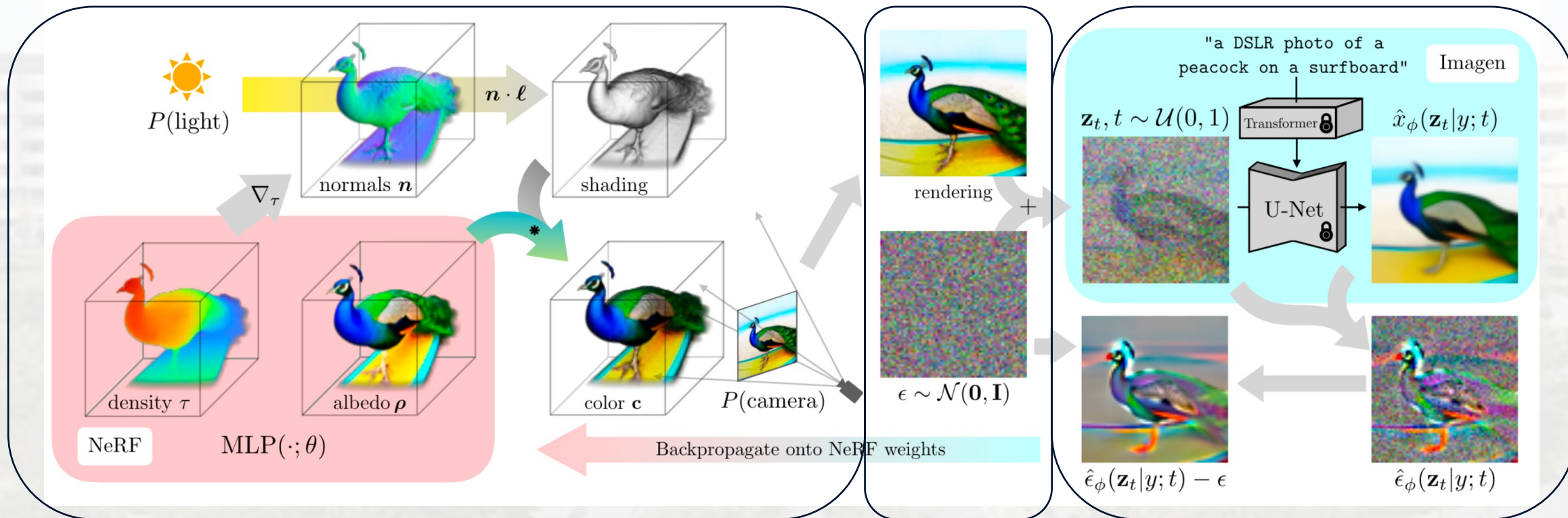


DreamFusion: Text-to-3D using 2D Diffusion

三维表示

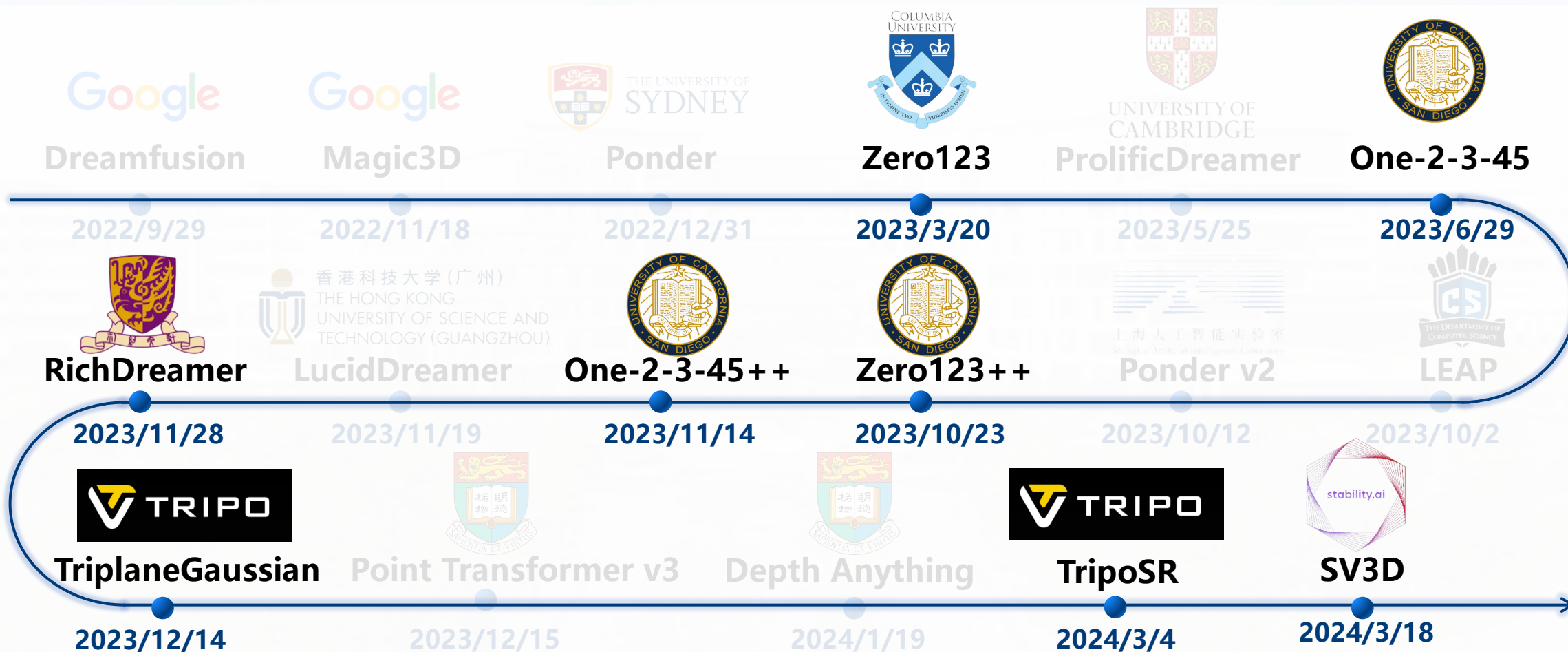
蒸馏

2D Diffusion



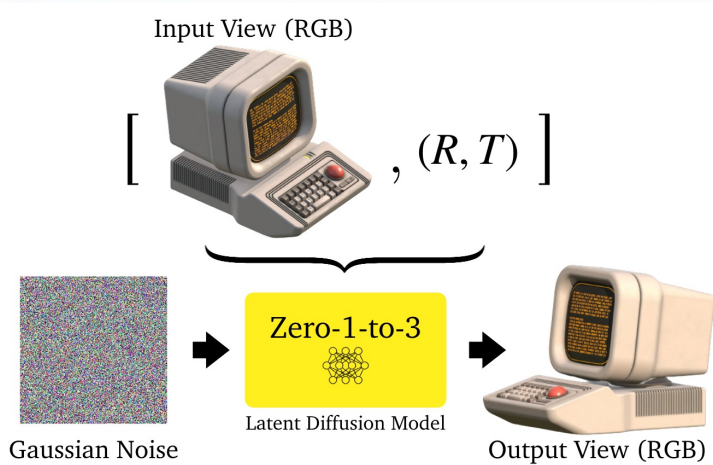
将二维大模型的知识蒸馏到三维表示中，生成三维模型，点燃了社区对三维模型生成的兴趣。

关键技术进展

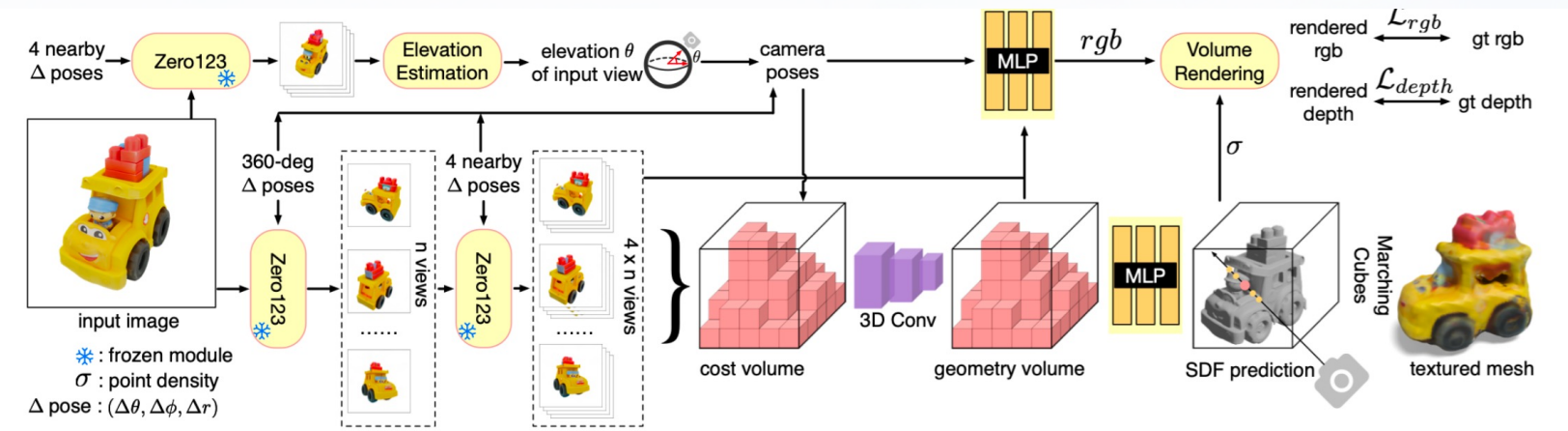


关键技术进展

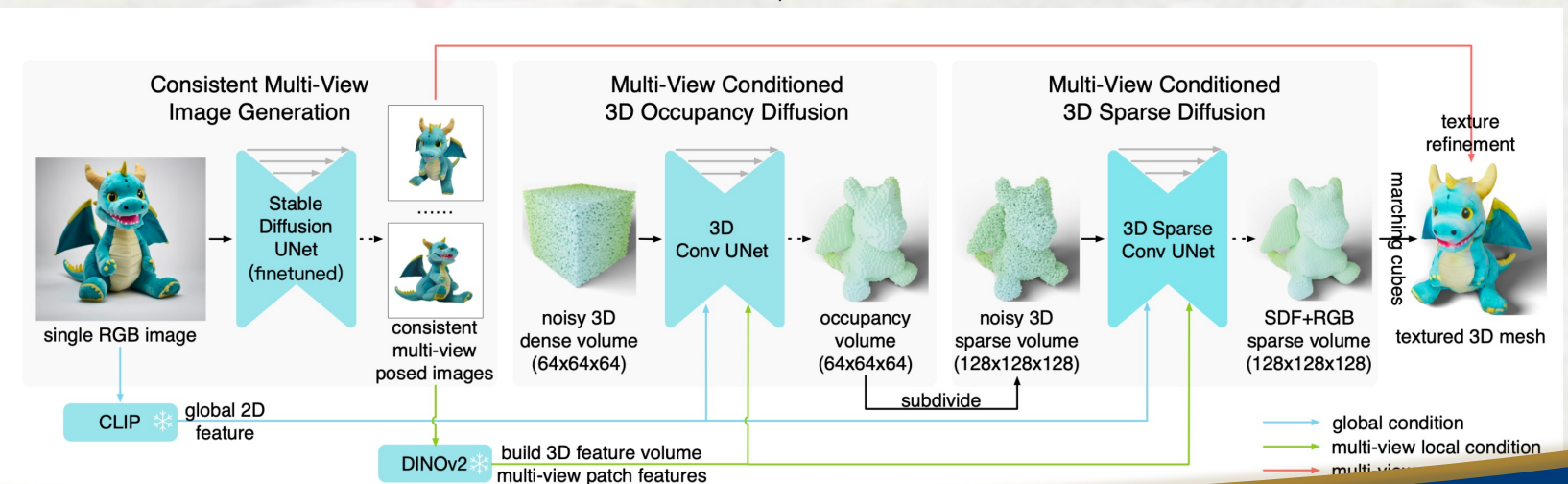
Zero-1-to-3



One-2-3-45



One-2-3-45++

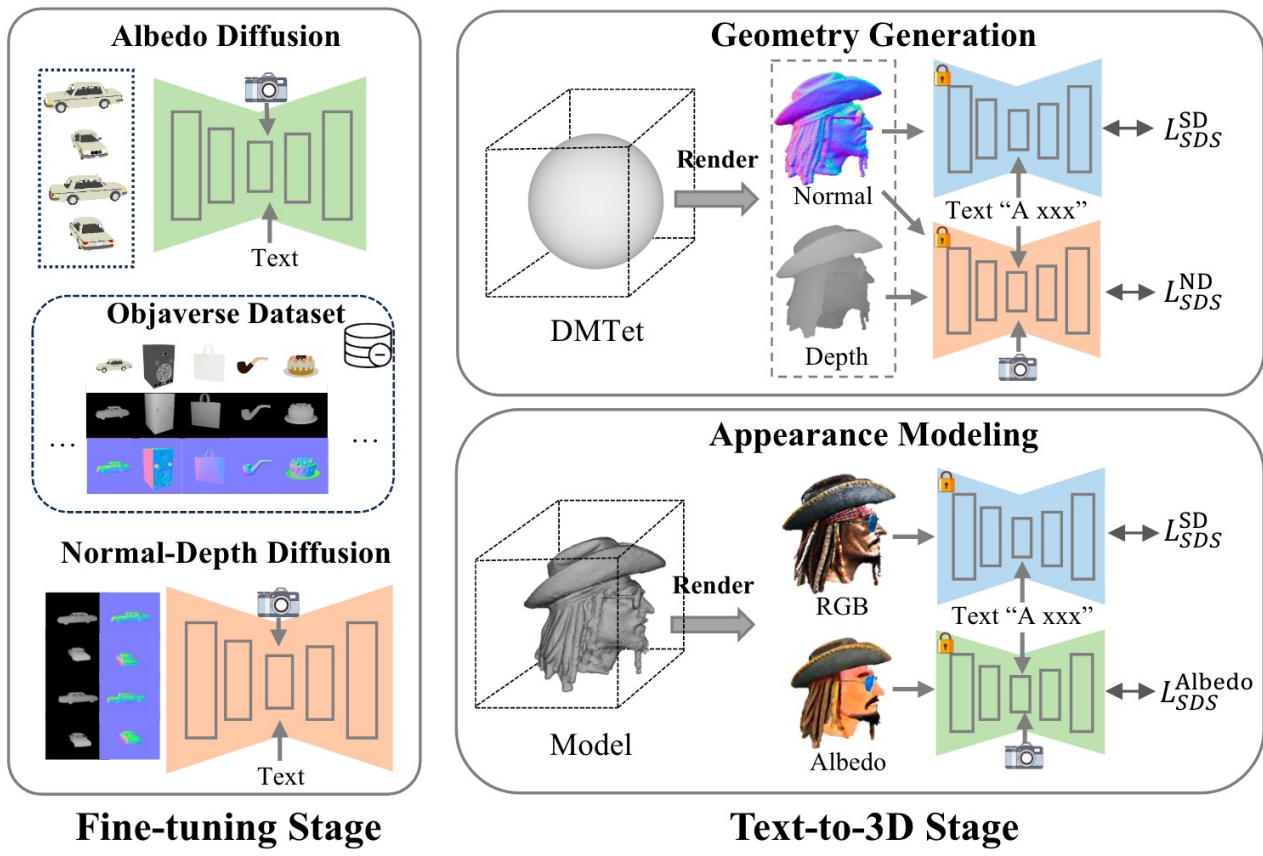


• Zero-1-to-3: 扩散模型可直接输出不同视图图像

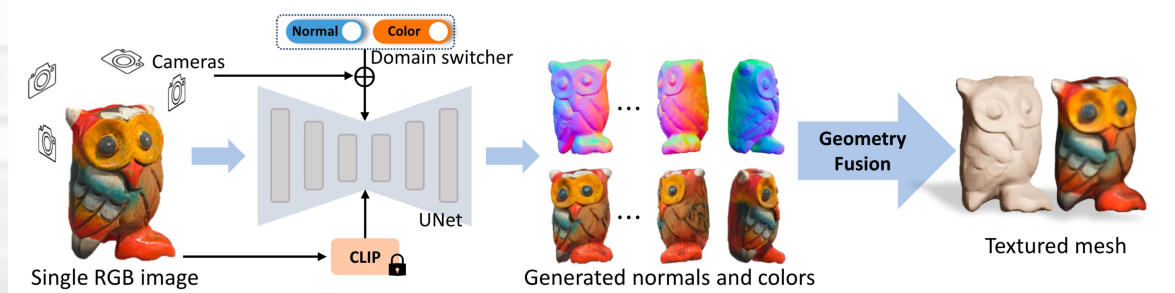
• One-2-3-45 (++)

- 提升多视角一致性 (性能)
- 专用可泛化重建模型 (速度)

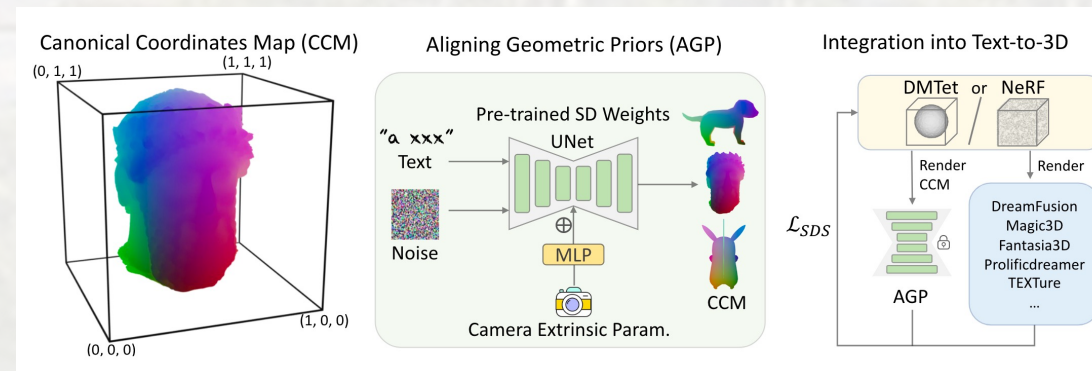
RichDreamer



Wonder3D

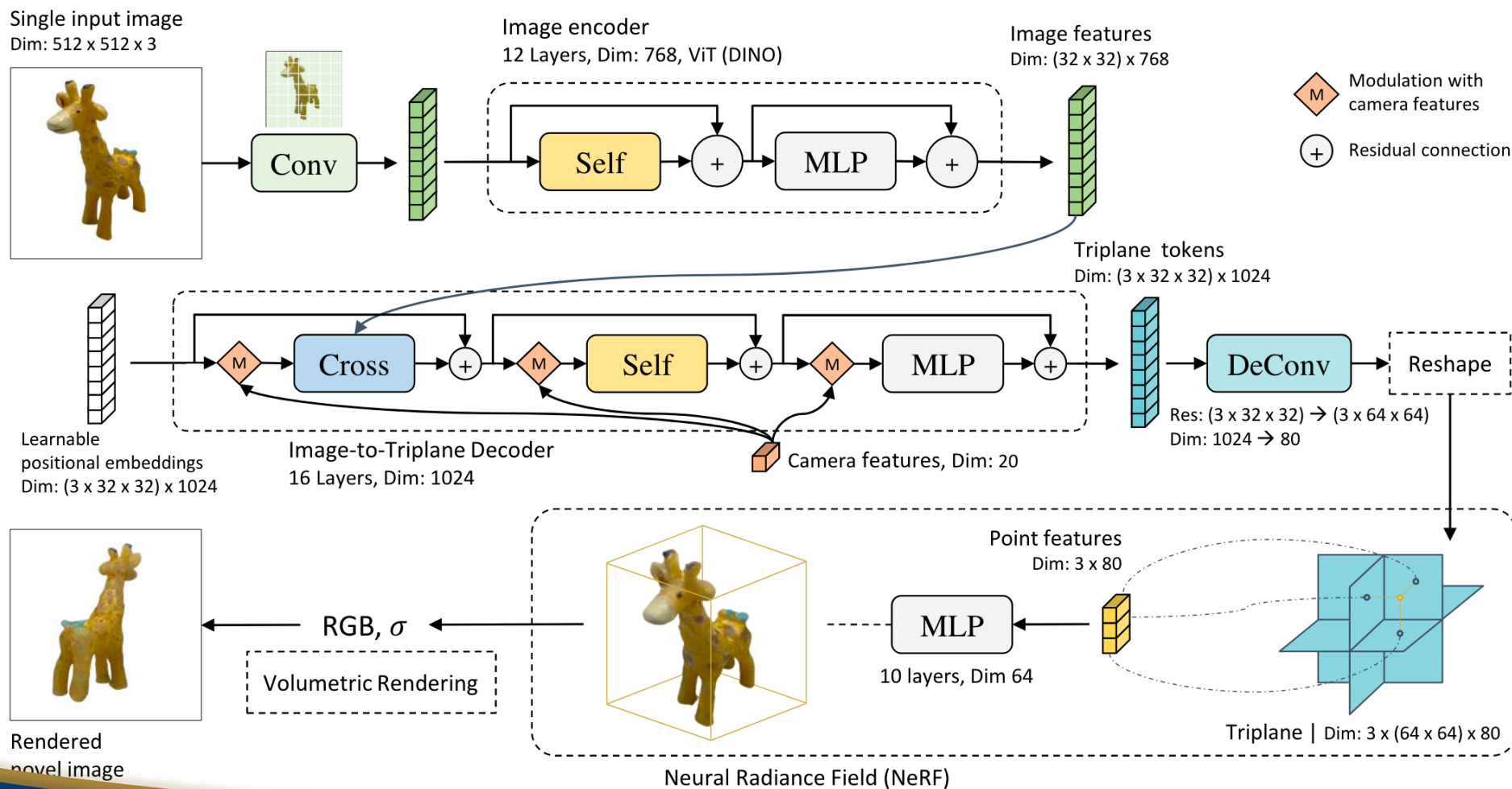


SweetDreamer



在扩散模型中嵌入三维先验信息（如normal, depth, CCM等）有助于提升几何一致性。

LRM: Large Reconstruction Model for Single Image to 3D

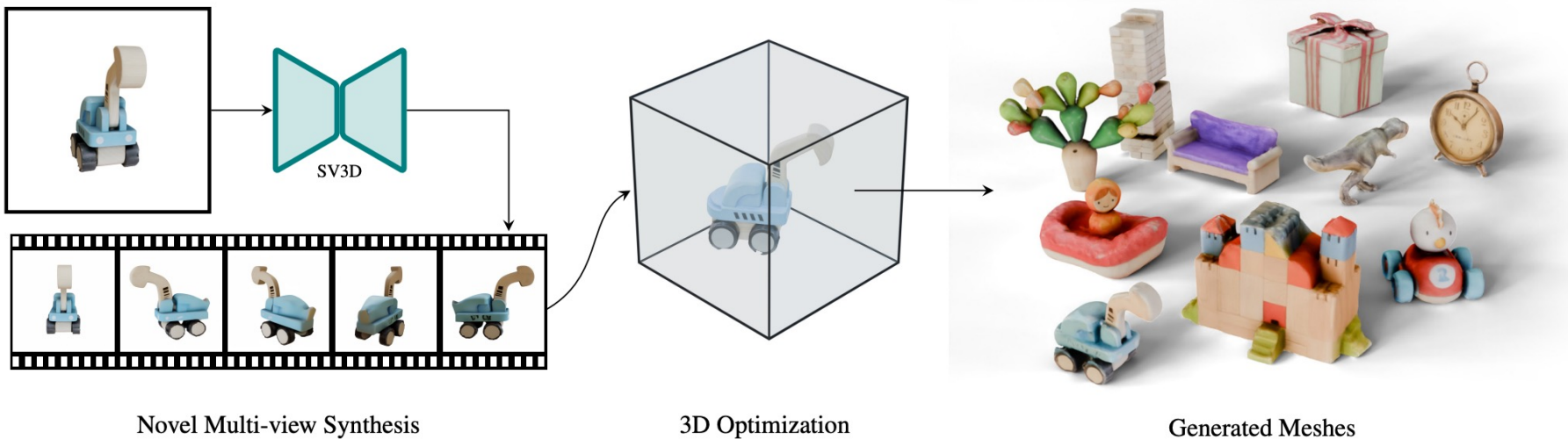


文字大模型的成功
取决于三大要素:

- Transformer
- 大数据集
- 自监督

这三大要素能否复制
到三维大模型中?

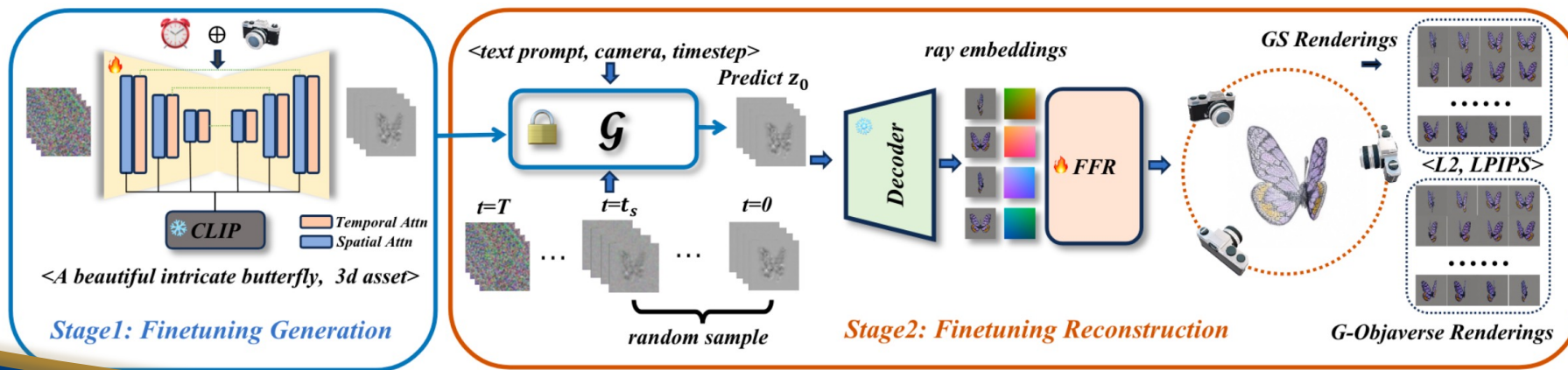
关键技术进展



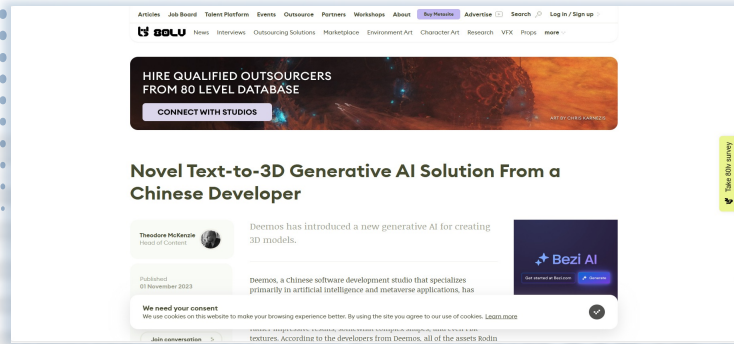
SV3D: Novel Multi-view Synthesis and 3D Generation from a Single Image using Latent Video Diffusion

基于视频大模型的3D生成

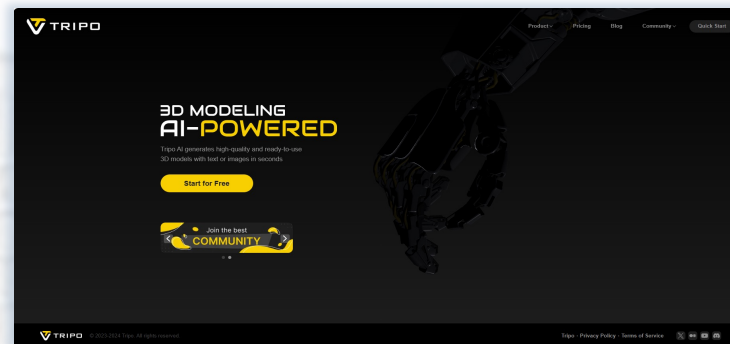
帧间一致性 \approx 多视图一致性



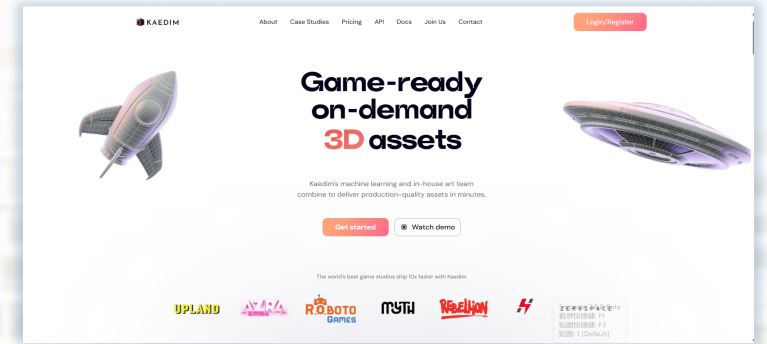
VideoMV: Consistent Multi-View Generation Based on Large Video Generative Model



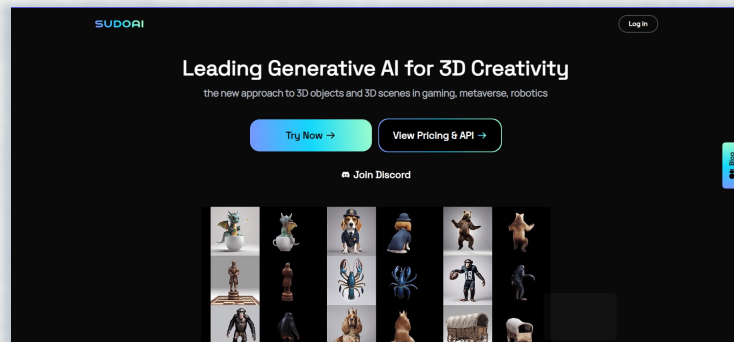
影眸科技



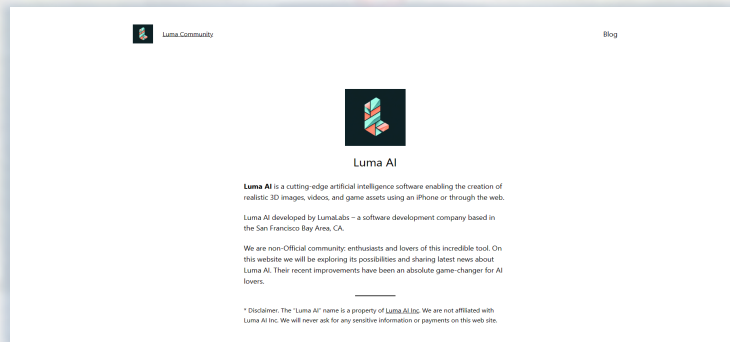
Vast AI



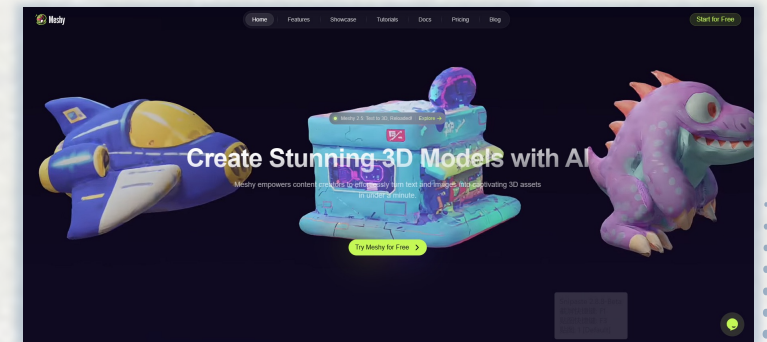
Kaedim



sudoAI



Luma AI



Meshy