

TECHNICAL SOLUTION / V0.1

物体照片生成 近似 3D 展示

单图深度、多图纹理与形状先验的本地 PoC 技术方案

适用：方案验证、商品概念展示、视觉原型

不适用：工业测量、精确逆向工程、完整遮挡面恢复

编制：Joseph / Single View Lab 日期：2026-08-24

01 / EXECUTIVE SUMMARY

方案摘要与目标边界

通过普通照片快速生成带纹理网格、旋转预览和浏览器展示页面。系统不是单一算法，而是根据物体结构和材质选择不同重建路线。

业务目标 把“拍照后看近似 3D 效果”的验证周期压缩到分钟以内。	输入 1 至 4 张手机照片；主体尽量完整，视角可互补。
输出 GLB 网格、PNG 海报、GIF 旋转预览、HTML 展示页。	部署 当前全部本地处理；演示站使用 Cloudflare Pages 静态托管。
质量定位 优先保证正面可视区域可信，隐藏面允许使用可解释的先验补全。	非目标 不保证毫米级尺寸、真实内部结构、完整背面或工程拓扑。

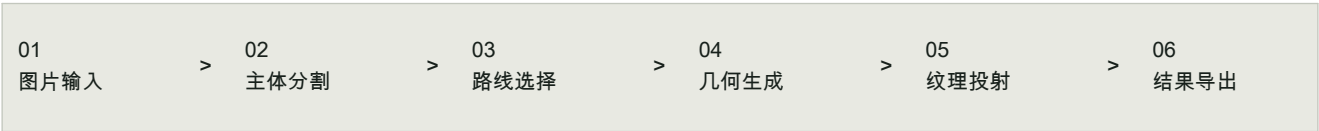
核心成功标准

- 输入照片不离开本地环境，单次 PoC 在 60 秒内完成。
- 输出文件可以被标准 GLB 查看器加载，并支持旋转和缩放。
- 原始照片纹理必须进入最终网格，禁止用生成式图片伪装重建结果。
- 对隐藏面、镜像纹理和形状先验进行明确标注，避免把视觉近似描述为真实扫描。

02 / ARCHITECTURE

总体架构与处理流程

统一入口负责图像接收和物体分类，路由层选择单图深度、结构先验或多图平面纹理方案，输出层统一生成 GLB 和静态预览。



逻辑分层

层级	职责	当前实现
输入层	读取图片、方向归一化、裁切、质量检查	Pillow / OpenCV
感知层	主体遮罩、轮廓、相对深度、平面角点	分割模型或 GrabCut + 规则
路由层	根据材质、形状、视角数量选择算法路线	当前为显式参数，后续自动分类
几何层	深度网格、参数化曲面、规则方盒网格	NumPy + trimesh
纹理层	UV、透视校正、多图图集、照片投射	OpenCV + Pillow
输出层	GLB、GIF、PNG、HTML、manifest	trimesh + 软件渲染器

统一输出目录

每个案例输出 result.glb、preview.gif、preview.png、viewer.html 和 manifest.json。manifest 记录方法、输入、耗时、指标及限制。

03 / ROUTE A

路线 A：单图分割 + 深度网格

适用于苹果、包装袋等轮廓完整、表面连续、正面纹理占主导的物体。该路线保留原照片像素，因此正面观察时真实感最高。

步骤	实现要点	关键输出
1. 主体分割	得到二值遮罩，清理小连通区域并羽化边缘	mask.png
2. 深度估计	生成相对深度；做范围裁剪、平滑和边缘保持	depth.png
3. 点云生成	把像素坐标与深度映射到局部三维坐标	vertices
4. 网格连接	按像素邻接关系生成三角面，跳过深度突变和背景	faces
5. 厚度补全	对边缘挤出或镜像，封闭背面，减少破面	watertight mesh
6. UV 投射	原图坐标直接作为 UV，照片成为正面纹理	texture.png

核心计算

对前景像素 (u, v) ，将深度值 d 归一化后映射为 $x = (u - cx) / f, y = -(v - cy) / f, z = scale \times d$ 。网格连接时，如果相邻顶点深度差超过阈值，则不生成对应三角面，以避免跨越物体边界。

```
for each foreground pixel (u, v):
    vertex = project(u, v, normalized_depth[u, v])
    for each 2x2 pixel cell:
        if all vertices valid and depth_jump < threshold:
            emit two triangles
```

优势与限制

- 优势：输入最少、速度快、原图纹理保真，适合正面小角度旋转。
- 限制：背面不可见；大角度旋转时会暴露拉伸、镜像或浅浮雕特征。
- 失败模式：透明、镜面、镂空、极薄结构会让深度和遮罩同时不稳定。

04 / ROUTE B

路线 B：分割结果 + 形状先验兜底

适用于风扇、瓶体、薄片和镂空结构。深度图只负责可见细节，整体体积由可解释的参数化先验控制。

风扇先验 圆形外框、浅圆顶深度、中心轴和扇叶区域分别建模，避免网罩被深度噪声撕裂。	瓶体先验 使用旋转曲面、超椭圆截面和堆叠圆柱描述瓶身、瓶盖、吸嘴与提手。
厚度策略 薄片结构使用最小厚度；实体结构按轮廓比例估计前后深度。	纹理策略 可见面继续使用实拍纹理；不可见区域采用同材质颜色或镜像纹理。

先验参数

参数组	示例参数	来源
轮廓	宽高比、圆角半径、肩部收缩曲线	分割遮罩与边界框
体积	深度比例、曲面指数、最小厚度	侧视图或类别默认值
部件	瓶盖半径、中心轴、提手路径	局部检测或人工配置
材质	基础色、粗糙度、透明度	前景像素统计与物体类别

为什么必须标注“先验”

形状先验提高稳定性，但它不是从照片中恢复出的事实。系统应在 manifest 和展示页中记录先验类型，并限制旋转范围，避免用户把补全区域误认为真实背面。

05 / ROUTE C

路线 C：多图平面校正 + 纹理方盒

适用于茶叶铁盒、包装盒、书本等规则不透明物体。顶面、底面和 45 度图分别提供可验证的平面纹理。

01 角点标注 > 02 透视变换 > 03 面纹理输出 > 04 方盒网格 > 05 UV 绑定 > 06 GLB 导出

平面透视校正

每个可见面使用四个角点计算 3×3 单应矩阵 H 。OpenCV

将原图中的倾斜四边形映射为规则矩形，输出顶面、底面、正面和侧面纹理。45 度照片中的两个可见面分别校正。

```
H = getPerspectiveTransform(source_quad, target_rect)
texture = warpPerspective(image, H, target_size)
mesh_face.uv = [(0,1), (1,1), (1,0), (0,0)]
```

当前补全策略

- 顶面和底面完全使用对应实拍照片，不进行生成式补画。
- 正面和右侧使用 45 度照片拆分后的纹理。
- 未拍摄的背面和左侧使用已拍摄面的镜像纹理，并在结果限制中显式记录。
- 盒体尺寸来自透视比例估计；若需要更准确比例，应提供正交正面、侧面和尺标。



图 1：三图平面纹理方盒的网格渲染海报

06 / CAPTURE GUIDE

输入拍摄规范

输入质量比模型复杂度更重要。标准化拍摄可以显著减少遮罩错误、视角配准失败和纹理接缝。

项目	推荐做法	应避免
背景	纯色、无纹理、与主体颜色有反差	杂乱桌面、屏幕、强阴影
光线	柔和漫射光，曝光一致	直射高光、局部过曝、频闪
物体状态	物体固定，只有相机绕物体移动	每张图都重新拿起或改变姿态
角度	每 15 至 30 度一张，覆盖完整一圈	只有正面和完全无重叠的背面
距离	焦距、距离和画面占比尽量一致	广角贴近拍摄、尺度变化过大
特殊材质	透明、反光物体增加柔光和背景标记	把玻璃、镜面当普通不透明物体

最小拍摄集合

单图连续曲面

正面 30 至 45 度，主体完整，边缘清晰。

规则方盒

正面 45 度 + 顶面 + 底面；最好补充背面 45 度。

多图通用重建

固定物体，连续环绕 20 至 40 张，视角间有足够纹理重叠。

视频替代

缓慢绕物体一圈，抽取稳定帧；避免运动模糊和自动曝光跳变。

07 / IMPLEMENTATION

代码结构、数据契约与运行方式

项目代码保持在 single-view-3d-poc，知识库只存方案与演示归档；静态站点从项目产物复制展示资源。

路径	职责
generate_3d.py	单图分割、深度网格、纹理和 GLB 输出
generate_multiview.py	多图轮廓融合、参数化瓶体、纹理图集
generate_box_multiview.py	平面透视校正、方盒网格、面纹理绑定
input/	原始输入图片；不覆盖、不二次压缩
output/	GLB、GIF、PNG、viewer、manifest 和诊断图
site/	纯静态演示站及可下载技术文档

manifest 数据契约

```
{
  "method": "three-view planar rectification + textured box mesh",
  "inputs": ["01-bottom.jpg", "02-top.jpg", "03-front-45.jpg"],
  "elapsed_seconds": 32.47,
  "outputs": {"mesh": "result.glb", "preview": "preview.gif"},
  "limitations": ["unseen faces use mirrored textures"]
}
```

本地技术栈

- Python 3.9+；NumPy 负责几何计算，OpenCV 负责分割与透视变换。
- Pillow 负责纹理、图集和 GIF；trimesh 负责网格组织与 GLB 导出。
- HTML 使用 model-viewer 加载 GLB；演示站本身只依赖静态文件。

08 / QUALITY & PERFORMANCE

实测指标与质量评估

当前评估以视觉可用性、输出完整性和处理耗时为主，不使用尺寸误差作为指标。

案例	路线	耗时	模型	主要结论
苹果	单图深度	6.57s	5.1MB / watertight	连续曲面效果较稳定
桌面风扇	深度 + 形状先验	6.86s	6.8MB / 非封闭	薄片和镂空仍有噪声
透明黑水杯	多图轮廓 + 瓶体先验	21.17s	带纹理 GLB	透明反光材质不适合当前路线
方形铁盒	三图平面纹理	32.47s	带纹理 GLB	规则不透明物体明显更适配

建议质量指标

- 轮廓一致性：渲染主视角轮廓与输入遮罩的 IoU。
- 纹理一致性：可见面重投影后的 SSIM 或 LPIPS。
- 完整性：GLB 可加载、材质存在、面数和文件大小不超预算。
- 交互体验：首屏海报先展示，GLB 按需加载；移动端不自动加载多个模型。
- 透明度：展示页必须明确标注真实纹理区域、镜像区域和先验补全区域。

验收分级

P0 - 可运行 流程成功、文件齐全、GLB 能加载。	P1 - 可展示 主视角轮廓稳定、纹理方向正确、无明显破面。
P2 - 可交付 多角度纹理连续、隐藏面标注、移动端体验稳定。	P3 - 产品化 自动拍摄质检、任务队列、监控、版本与成本治理。

09 / DELIVERY

前端展示、部署和安全边界

当前演示站不执行重建，只展示图片、GIF、案例指标和技术说明，因此可以完全静态托管。

模块	当前方案	约束
静态站	HTML + CSS , Cloudflare Pages	不上传用户图片，不运行模型
资源	图片与 GIF 随站点发布	控制单文件大小，使用懒加载
GLB 查看	独立 viewer.html + model-viewer	需要浏览器 WebGL
下载文档	PDF 位于 site/downloads/	版本号与页面说明保持一致
隐私	本地脚本读取照片	不得把照片自动发送到第三方生成服务

安全与合规

- 输入照片可能包含人物、屏幕、二维码和办公环境；生产化前必须增加裁切、脱敏和保留周期。
- 输出纹理会继承照片中的文字、商标和个人信息；展示前需要人工确认发布范围。
- 不把 API Key、模型凭据或本地绝对路径写入网页与 manifest。
- 公开站点只发布已筛选案例，不提供任意文件上传入口。

10 / ROADMAP

产品化路线与实施计划

先把拍摄质量和路线选择标准化，再引入后台任务系统。直接建设上传后台而没有输入规范，会放大失败率和算力成本。

阶段	目标	交付物	退出标准
阶段 1	稳定本地 PoC	拍摄规范、三条路线、案例集	规则物体 P1 通过率 ≥ 80%
阶段 2	自动路由与质检	材质分类、模糊检测、角度覆盖评分	失败可提前提示，不进入重建
阶段 3	本地 Web 工具	上传页面、任务状态、结果预览	单机可管理多任务和产物
阶段 4	异步后台	API、队列、GPU Worker、对象存储	任务可恢复、可追踪、可限流
阶段 5	高级多视图	相机配准、神经重建或 3DGS	标准环绕数据达到 P2

后台目标架构



首个产品化迭代建议

- 只支持规则不透明物体和连续曲面物体，暂不承诺透明、镜面和镂空结构。
- 上传前提供拍摄引导，服务端先做质量评分，不合格直接返回重拍建议。
- 输出默认限制旋转角度；只有多视图覆盖充分时才开放 360 度。
- 保留每次任务的输入哈希、模型版本、参数、耗时和限制说明，保证结果可追溯。

11 / ACCEPTANCE

验收清单与结论

该方案当前已经证明“普通照片可以在本地生成可展示的近似 3D 结果”，但产品边界必须按物体类别管理。

检查项	通过标准
输入完整性	照片可读取、方向正确、主体占比满足阈值
分割质量	主体边缘无大面积缺失，背景残留不进入主网格
几何质量	无严重翻面、爆点、悬空部件；主视角比例合理
纹理质量	方向正确、无整体镜像、文字与图案不跨面错位
文件完整性	GLB、GIF、PNG、viewer、manifest 均生成
浏览器兼容	桌面和移动端可显示预览；GLB 查看器可旋转缩放
限制披露	隐藏面、镜像面和形状先验在 manifest 与页面中一致
隐私发布	公开资源不包含未授权人物、屏幕内容和敏感路径

结论

当前最可行的推进方式不是追求一个算法覆盖所有物体，而是建立“输入质检 + 路线选择 + 明确限制”的组合系统。
短期优先支持规则不透明物体和连续曲面物体；透明、镜面、镂空与测量级需求单独立项。

当前状态 本地 PoC、三个公开案例、静态演示站、可下载技术方案。	下一决策 是否进入标准拍摄工具与本地 Web 上传界面阶段。
---	--

在线演示：<https://joseph-3d-demo.pages.dev/>