MonoPop3D: Single-Image Volumetric 3D Pop-up Generation

Yunyi Shi, Ruizhe (Rachel) Tan, Lynn Li
MScAC, University of Toronto

Motivation

Why 3D From a Single Image?

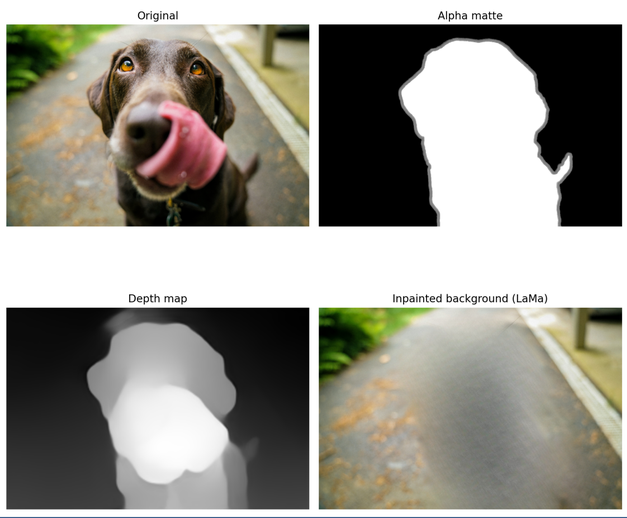
- 2D photos lack depth cues and feel flat.
- Lightweight 3D effects are increasingly desired for education, storytelling, etc.
- True 3D reconstruction pipelines (NeRF, MVS) are complex, heave, and require multiple views.

Problems with Existing Approaches

- Cardboard Cutout Effect: Objects are treated as rigid, flat planes, causing objects slide without rotating.
- Foreground extraction with hard masks fail on fuzzy edges (hair, fur, etc.).
- Inpainting methods can produce blurry fills or mismatched textures.

Our Approach

- Use Deep Learning Priors (SAM, MiDaS, LaMa) for segmentation, depth and pinpointing.
- Apply test-time mathematical optimization to enforce geometric consistency and volumetric shape.
- Animate layers using depth-aware parallax + volume warping for realistic 3D motion.



Related Work

- **Layered Depth Images (LDI):** Shade et al. [2] proposed storing multiple depth layers per pixel. We simplify this into a foreground/ background model for robustness.
- **3D Photography:** *Shih et al. [4]* introduced depth-based inpainting. We extend this by integrating LaMa for better semantic consistency than traditional inpainting.
- **Neural Rendering:** Recent NeRF-based methods require long training times per scene. Our approach uses pre-trained priors (SAM, MiDaS) for zero-shot inference.

Limitations of Previous Approaches

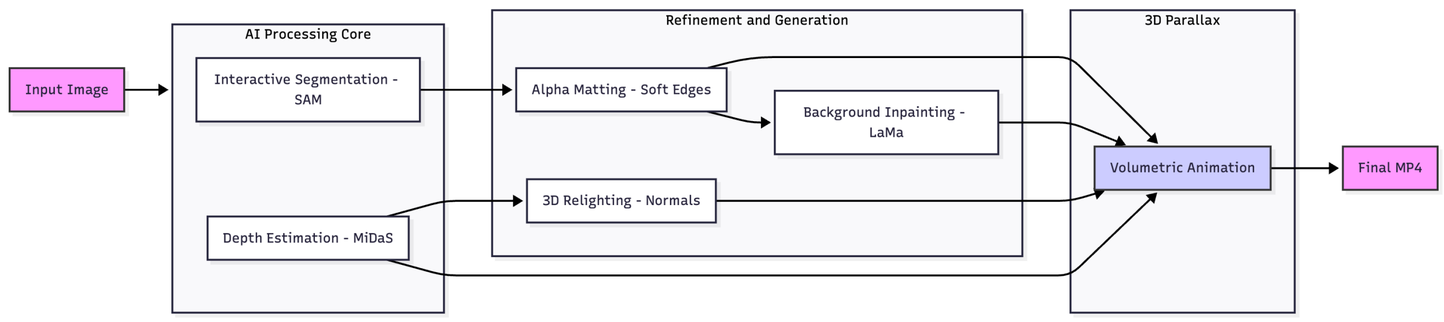
- **Jagged Edges:** Binary masks fail on furry objects (teddy bears, pets).
- **Flat Geometry:** Most "Parallax" apps treat objects as rigid planes, breaking the 3D illusion during rotation.

References

[1] A. Kirillov, E. Mintun, N. Ravi, et al. "Segment Anything." *ICCV*, 2023.
[2] J. Shade, S. Gortler, L.-W. He, and R. Szeliski. "Layered Depth Images." *SIGGRAPH*, 1998.
[3] R. Suvorov, E. Logacheva, et al. "Resolution-robust Large Mask Inpainting with Fourier Convolutions." *WACV*, 2022.
[4] M.-L. Shih, S.-Y. Su, J. Kopf, and J.-B. Huang. "3D Photography Using Context-Aware Layered Depth Inpainting." *CVPR*, 2020.
[5] K. He, J. Sun, and X. Tang. "Guided Image Filtering." *IEEE TPAMI*, 2013.

New Technique

We propose a hybrid pipeline combining **Deep Learning Priors** with **Test-Time Geometric Optimization**.



AI Processing Core

- **Interactive Segmentation (SAM):** Users click on the object parts they want, and we mathematically "stitch" these clicks together to form one complete shape.
- **Depth Estimation (MiDaS):** Predicts coarse depth and geometric cues necessary for parallax

Refinement and Generation

- **Alpha Matting:** Analyzes edge pixels and produces soft, natural boundaries (hair, fur, foliage) to avoid the Cardboard Cutout Effect.
- **Background Inpainting (LaMa):** "Hallucinates" a realistic background (like a continuing floor or wall) to fill the empty space seamlessly.
- **3D Relighting and Bokeh:** Uses the depth map to estimate surface curvature and add dynamic highlights (e.g., making the nose shine), while applying background blur to mimic the shallow depth-of-field of a professional camera lens.

3D Parallax

- **Volumetric Warping:** Warps pixels based on their depth—so the center moves faster than the edges—creating the illusion of a solid, rotating object.

Experimental Results

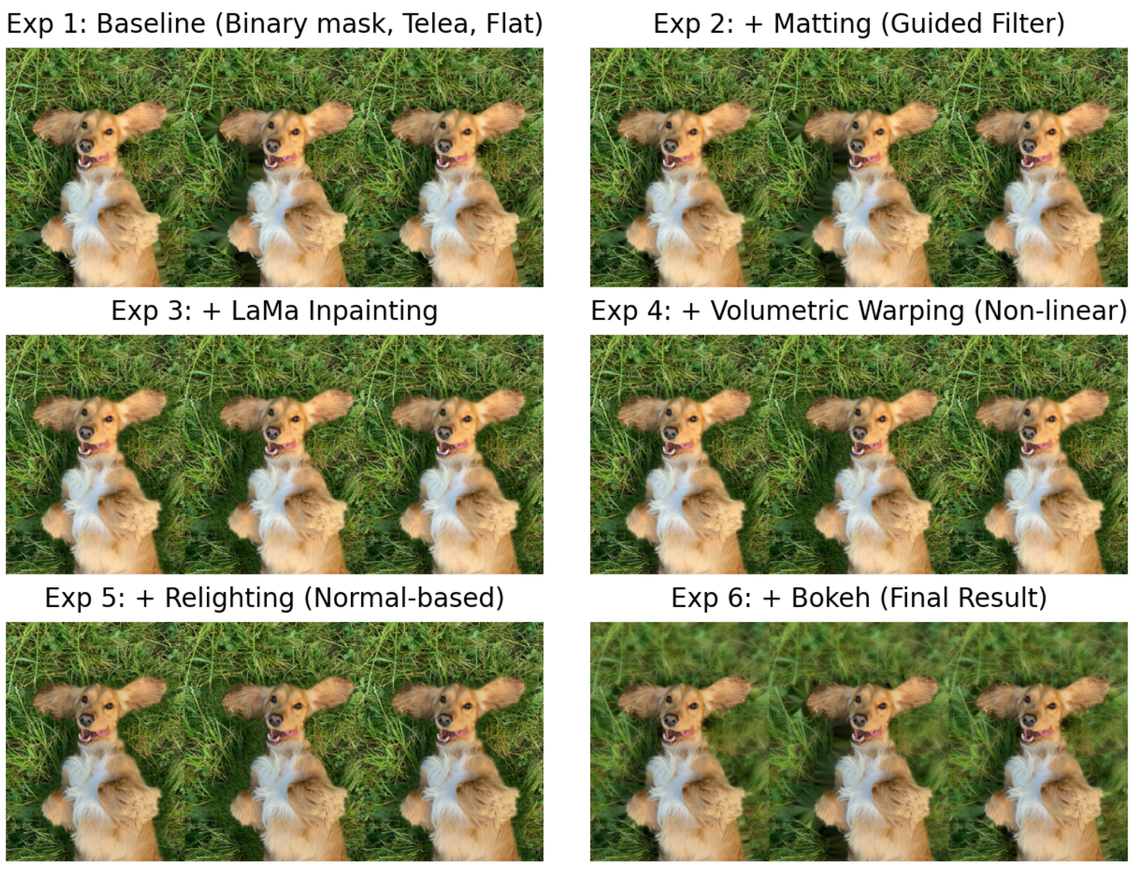
Quantitative Evaluation

- We measured the impact of our optimization modules using image statistics quantify "Realism" and “Focus.”

Metric	Baseline	Ours (Optimized)	Impact	Interpretation
Gradient Mean (Texture)	438.46	**626.69**	**+42.9%**	Relighting significantly enhances surface definition and 3D perceptibility.
High-Freq Energy (BG)	3.91e7	**2.56e7**	**-34.4%**	Bokeh successfully reduces background visual clutter, isolating the subject.
Inpainting PSNR	38.61 dB	**38.77 dB**	**+0.16 dB**	LaMa maintains higher structural fidelity than Telea (OpenCV)
Boundary Smoothness (Max Gradient)	4.47	**1.21**	**-73.0%**	Image matting brings significant reduction in edge hardness

Qualitative Ablation Study (ablation_runner)

- We build a gradio app to try different combinations and visualize the results



Scan to try our MonoPop3D ablation demo with your own image!

Scan to see our amazing results!