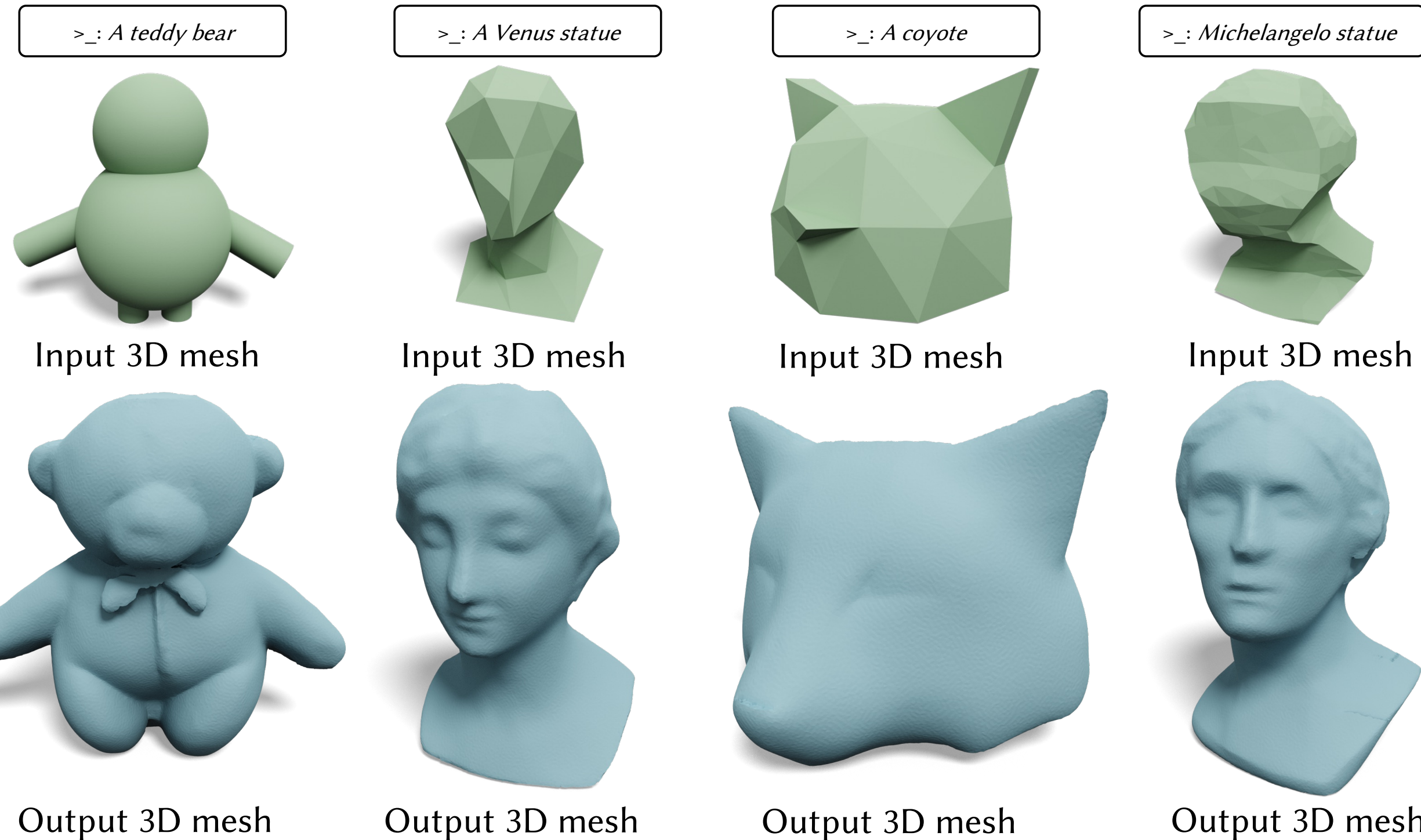




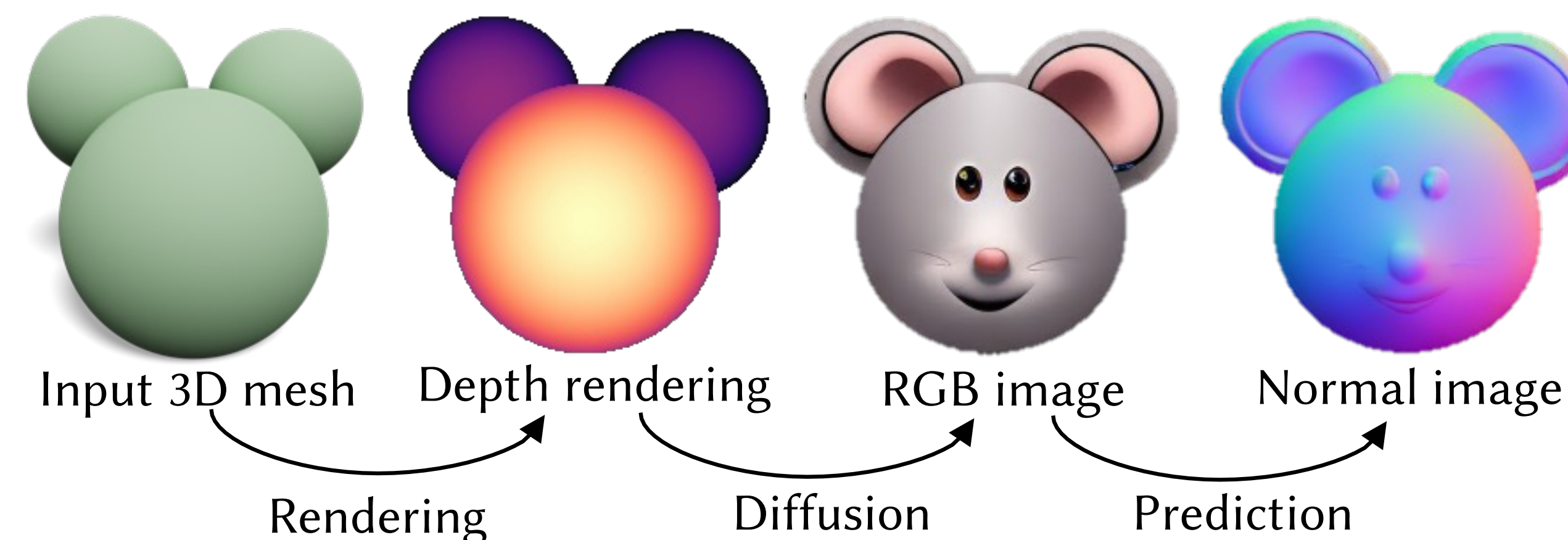
Text-guided Mesh Refinement

- Input: A coarse 3D mesh and a text prompt
- Goal: Create a 3D mesh by adding more geometric detail to the input coarse mesh guided by the input text prompt

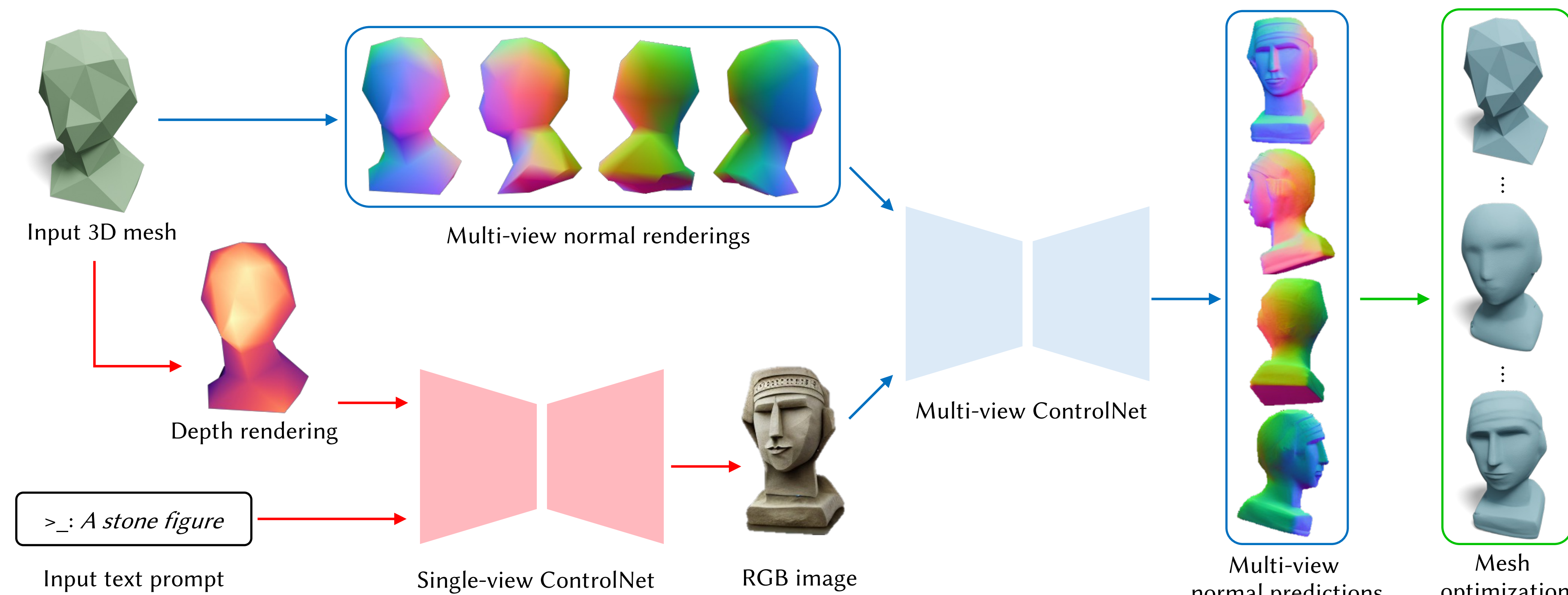


Key Insight

- Pre-trained text-to-image models generate images with rich geometric detail
- First stage: Single-view RGB generation for fast preview
- Second stage: Multi-view normal generation
- Third stage: Mesh optimization



Method Overview



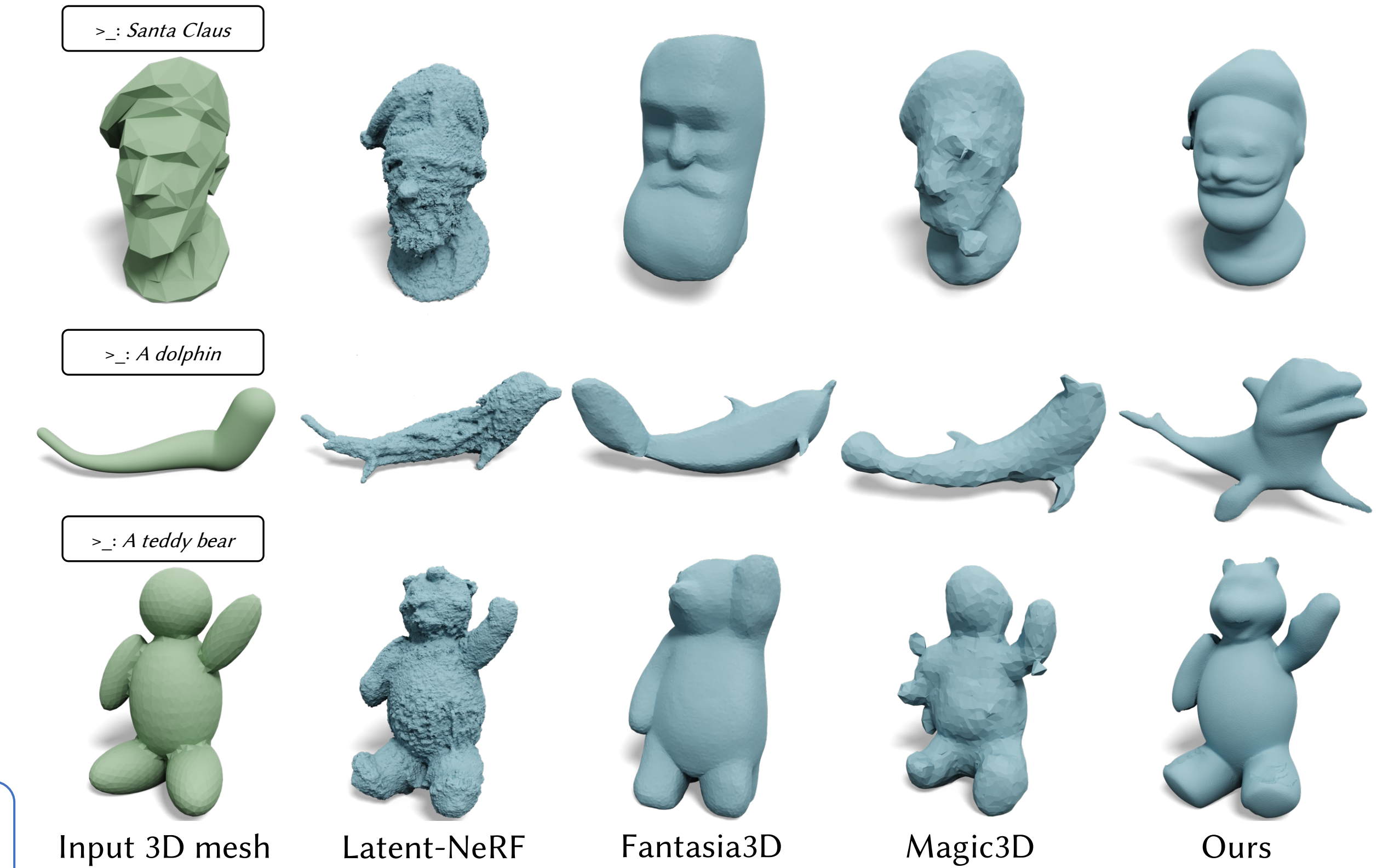
$$\mathcal{L}_n = \frac{1}{|\Theta|} \sum_{\theta \in \Theta} \|\mathcal{R}_n[M, \theta] - I_{n,\theta}\|_1$$

$$\mathcal{L}_{\nabla n} = \frac{1}{|\Theta|} \sum_{\theta \in \Theta} \left\| \frac{\partial \mathcal{R}_n[M, \theta]}{\partial x} - \frac{\partial I_{n,\theta}}{\partial x} \right\|_1 + \left\| \frac{\partial \mathcal{R}_n[M, \theta]}{\partial y} - \frac{\partial I_{n,\theta}}{\partial y} \right\|_1$$

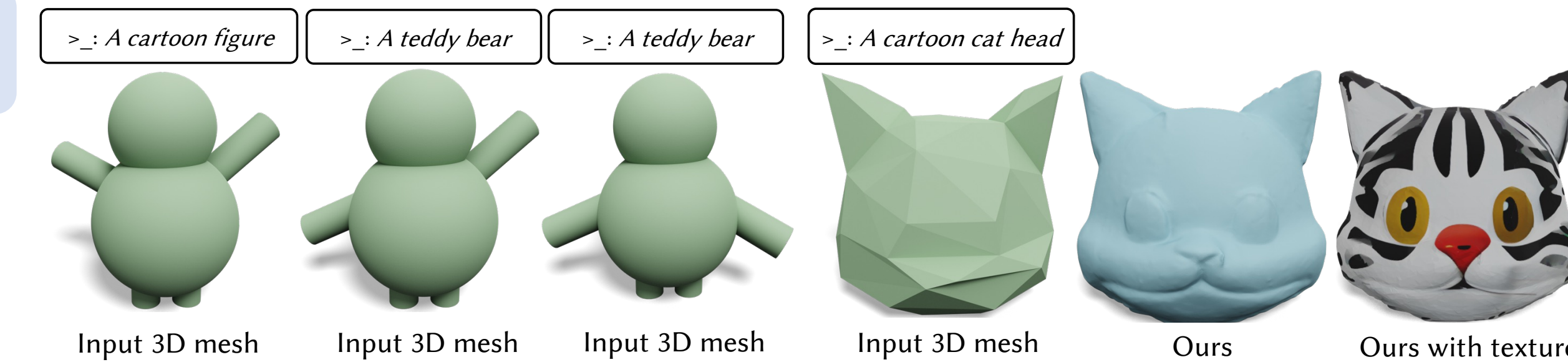
$$\mathcal{L}_{sil} = \frac{1}{|\Theta|} \sum_{\theta \in \Theta} \|\mathcal{R}_s[M, \theta] - \mathcal{S}[I_{rgb,\theta}]\|_1$$

Experimental Results

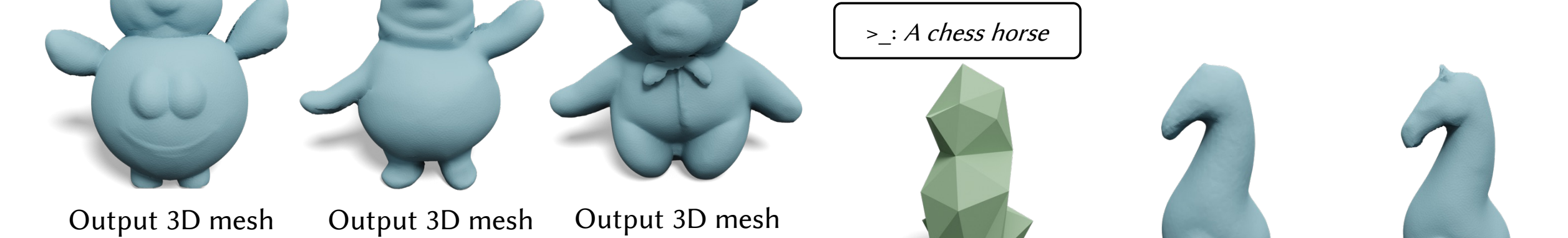
Comparison to state-of-the-art methods



Controlling pose with 3D input



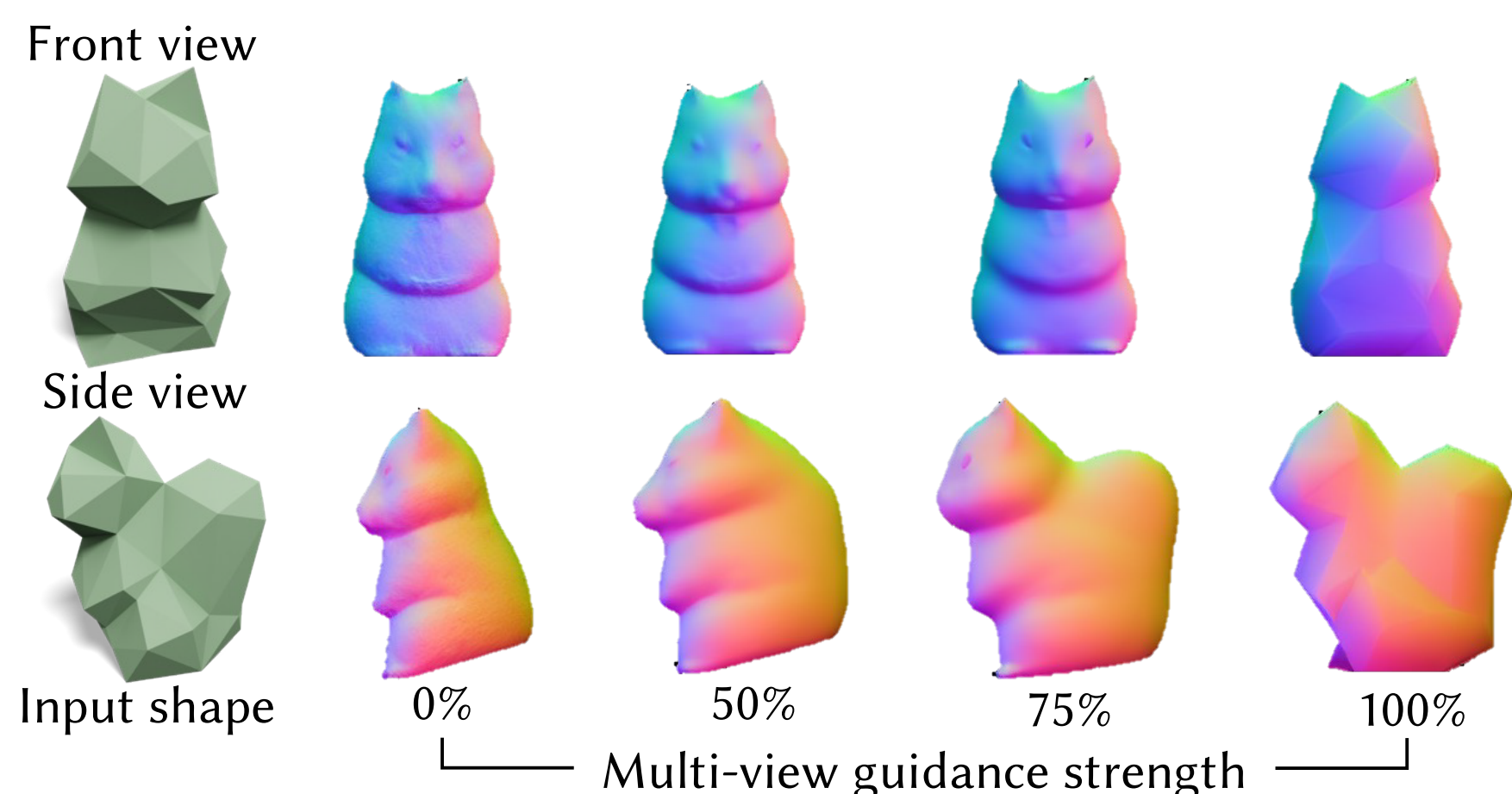
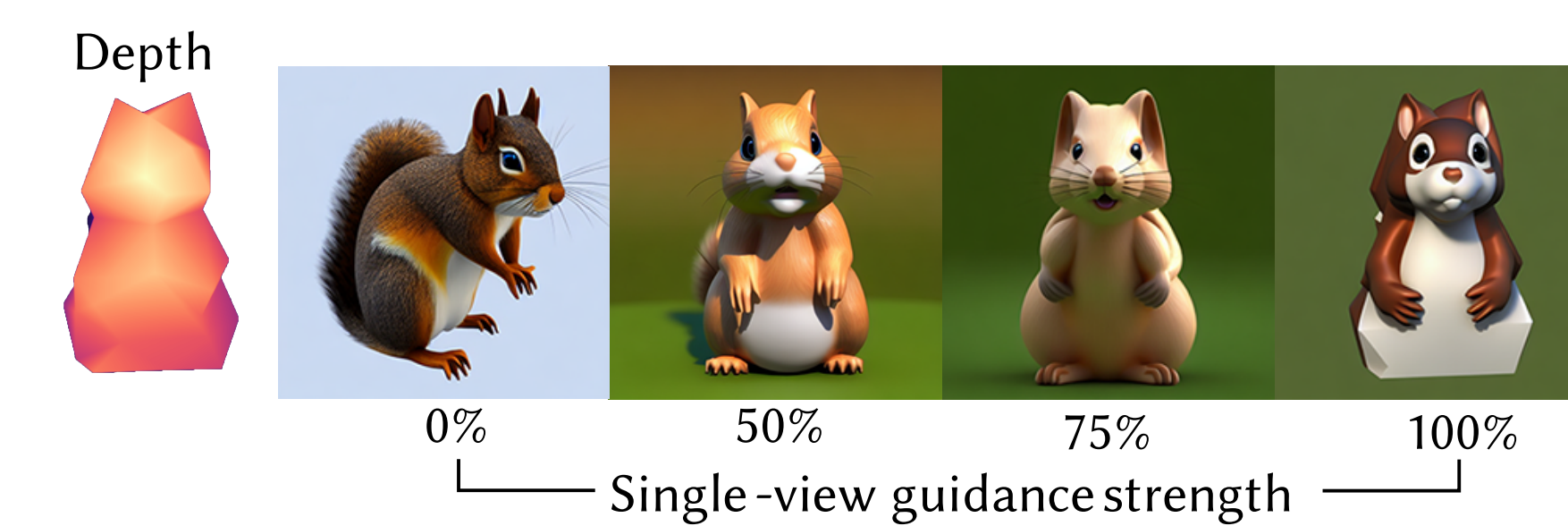
Mesh texturing



References

- Palfinger. *Continuous Remeshing for Inverse Rendering*.
- Lin et al. *Magic3D: High-Resolution Text-to-3D Content Creation*.
- Long et al. *Wonder3D: Single Image to 3D using Cross-Domain Diffusion*.
- Zhang et al. *Adding Conditional Control to Text-to-Image Diffusion Models*.
- Metzner et al. *Latent-NeRF for Shape-Guided Generation of 3D Shapes and Textures*.
- Chen et al. *Fantasia3D: Disentangling Geometry and Appearance for High-quality Text-to-3D Content Creation*.

Influence of Guidance Strength



Multi-view Control

