IEEE/CFV Winter Conference on Applications of Computer Vision (WACV) 2026

End-to-End Fine-Tuning of 3D Texture Generation using Differentiable Rewards

Amirhossein Zamani Tianhao Xie Amir Aghdam Tiberiu Popa Eugene Belilovsky

WACV 2026 TUCSON, AZ 3/6 - 3/10

Objective

Integrate task-specific preferences, encoded as differentiable rewards, into an end-to-end learning framework to generate texture images aligned with the geometry of a 3D mesh.

Advantages Our Method Offers

- **Efficient optimization:** Removes costly sampling-based reinforcement learning, reducing computation time.
- **End-to-end geometry awareness:** Backpropagates through the full 3D generative process, aligning textures naturally with mesh geometry.

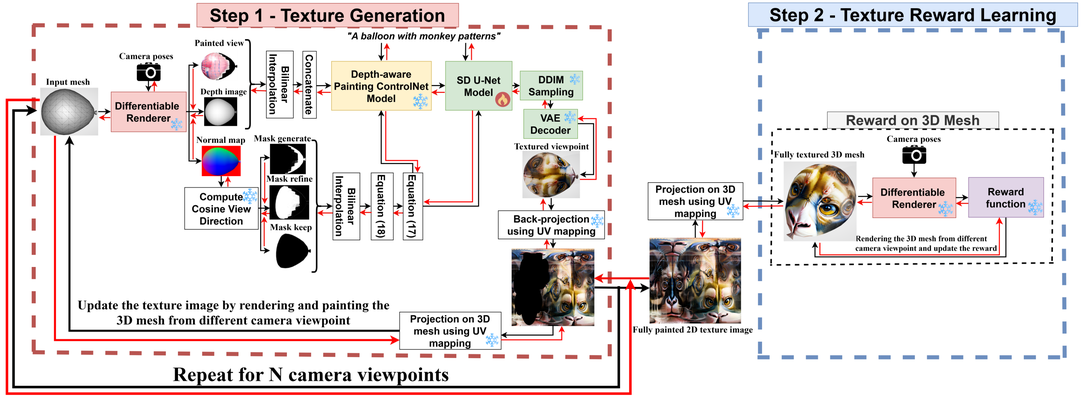
Core Idea: Align Texture with Geometry

- **Texture Generation (TexGen):** Render a 3D mesh from multiple viewpoints ($\boldsymbol{V_\theta}$) and apply a 2D text-to-image diffusion model ($\boldsymbol{\theta}$) to generate its texture from a text ($\boldsymbol{c}$).
- **Reward Design:** Compute alignment between texture features and 3D surface geometry through a differentiable reward ($\boldsymbol{r}$).
- **Reward Learning:** Given the TexGen texture, maximize the following objective:

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{c}\sim p_{\mathbf{c}}}\left[r\left(\mathrm{TexGen}\left(\boldsymbol{\theta}, \mathbf{c}, \mathbf{v_{gen}}\right), \mathbf{c}\right)\right]$$

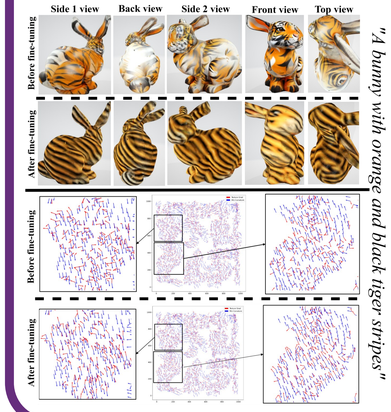
Systematic Approach to The Texture Preference Learning Problem

Texture Rewards Learning Pipeline

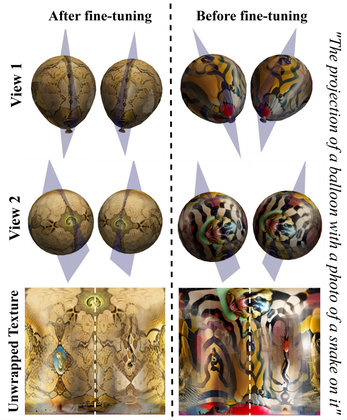


Geometry-Aware Reward Design

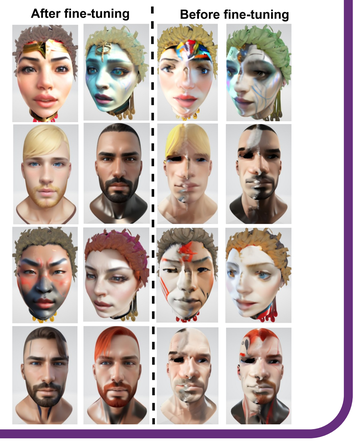
① Geometry-Texture Alignment



② Symmetry-Aware Texture Generation



③ Camera-Pose-Aware Texture Generation



Comparative Texture Generation Results

Qualitative Comparison

Quantitative Comparison

Method	Aesthetic ↑	ImageReward ↑	HPSv2 ↑	PickScore ↑	CLIPScore ↑	Inference Time (sec) ↓
TEXTure [1]	4.3910 ± 0.0328	-1.2245 ± 0.1199	0.1728 ± 0.0031	20.0719 ± 0.1083	0.3142 ± 0.0051	150
Text2Tex [2]	4.4454 ± 0.0387	-1.5088 ± 0.0823	0.1769 ± 0.0052	19.6886 ± 0.1862	0.2900 ± 0.0054	450
InTeX [3]	4.7467 ± 0.0381	-1.0859 ± 0.1469	0.1879 ± 0.0033	19.8832 ± 0.2368	0.3077 ± 0.0055	15
Paint3D [4]	4.7192 ± 0.0388	-1.6621 ± 0.0977	0.1549 ± 0.0089	18.8966 ± 0.2798	0.2798 ± 0.0065	30
Ours - Cam-Pose-Aware Reward	4.9328 ± 0.0271	-0.0479 ± 0.1748	0.2095 ± 0.0058	19.7609 ± 0.1573	0.3003 ± 0.0057	15
Ours - Geo-Tex-Align Reward	4.7722 ± 0.0324	-0.1758 ± 0.1136	0.2479 ± 0.0034	**21.5022** ± 0.1293	**0.3367** ± 0.0041	15
Ours - Sym-Aware Reward	4.9473 ± 0.0309	**0.0063** ± 0.0901	0.2118 ± 0.0031	20.8361 ± 0.0444	0.3076 ± 0.0032	15
Ours - Tex-Emphasis Reward	**5.0308** ± 0.0376	-0.5960 ± 0.1694	**0.2534** ± 0.0027	20.8822 ± 0.1125	0.3158 ± 0.0028	**15**

References

[1] Richardson, et al. "Texture: Text-guided texturing of 3d shapes," ACM SIGGRAPH 2023.

[2] Chen, et al. "Text2Tex: Text-Driven Texture Synthesis via Diffusion Models," ICCV 2023.

[3] Tang, et al."InTex: Interactive Text-to-Texture Synthesis via Unified Depth-Aware Inpainting," 2024.

[4] Zeng, et al. "Paint3D: Paint Anything 3D with Lighting-Less Texture Diffusion Models," CVPR 2024.