Paint-it: Text-to-Texture Synthesis via Deep Convolutional Texture Map Optimization and Physically-Based Rendering

POSTECH
AMI Algorithmic Machine Intelligence Lab.
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

https://kim-youwang.github.io/paint-it

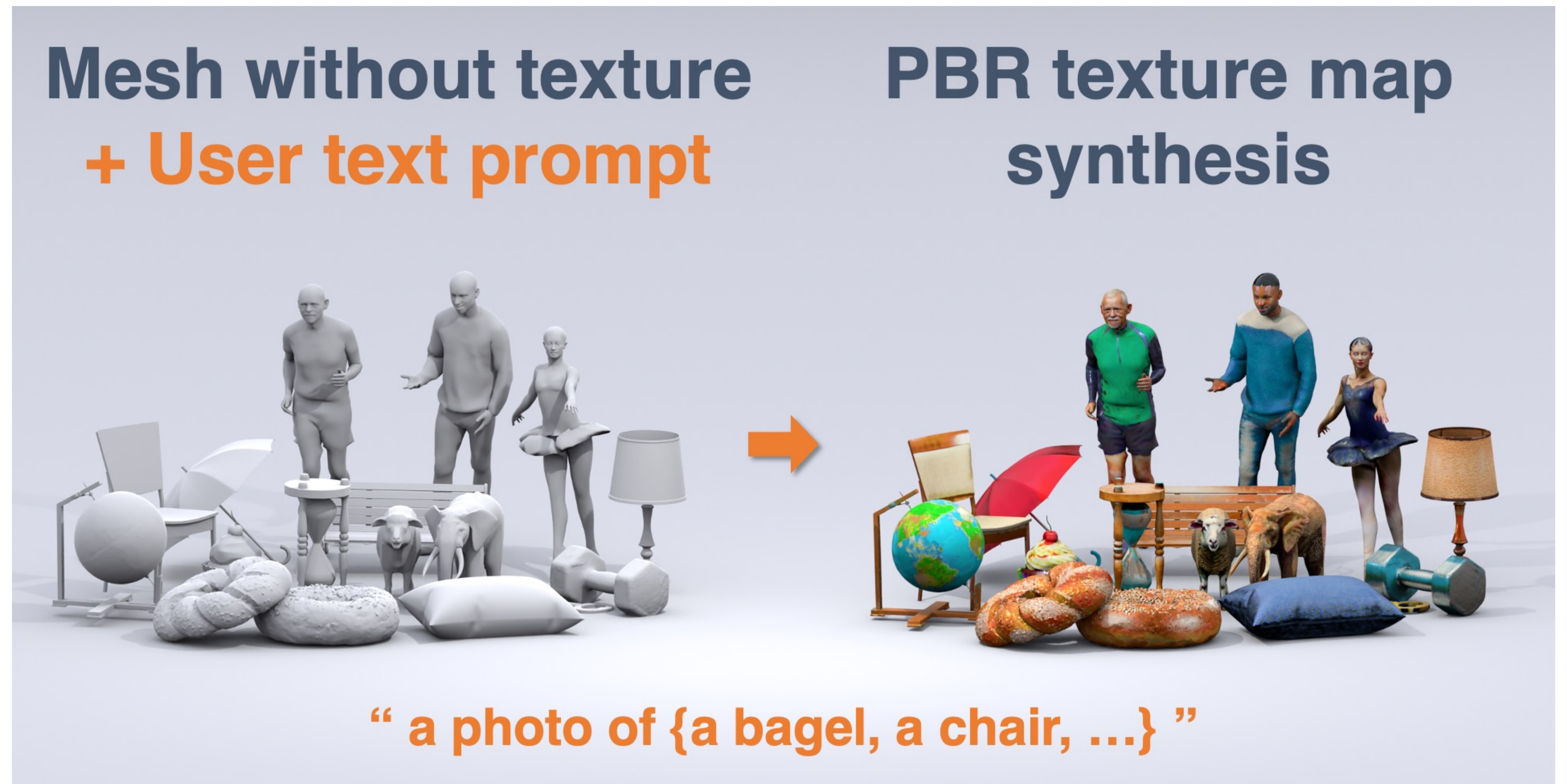
Kim Youwang[1] Tae-Hyun Oh[1] Gerard Pons-Moll[2]

[1]POSTECH [2]University of Tübingen

CVPR SEATTLE, WA JUNE 17-21, 2024

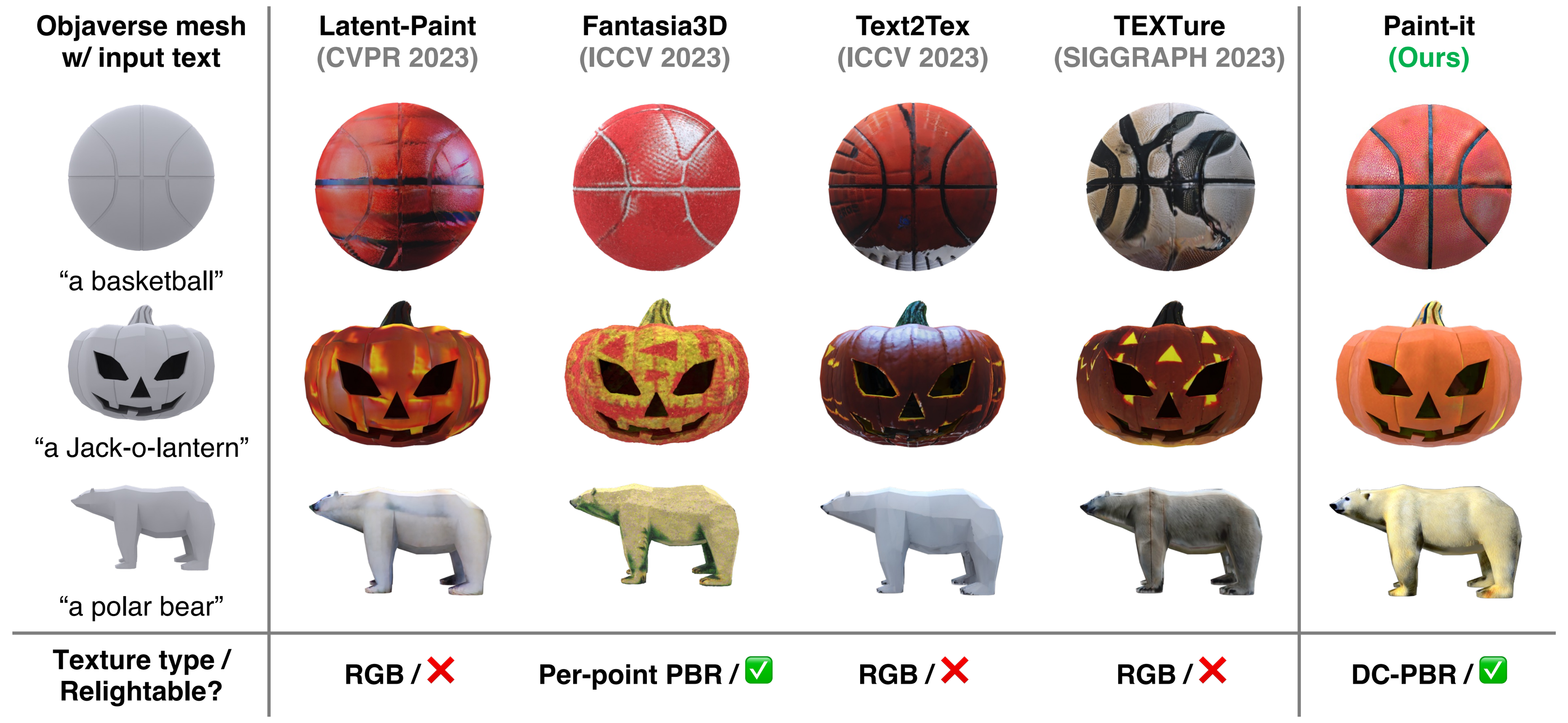
Overview

TL;DR Convolutional parameterization of PBR texture maps helps high-quality text-to-texture synthesis!

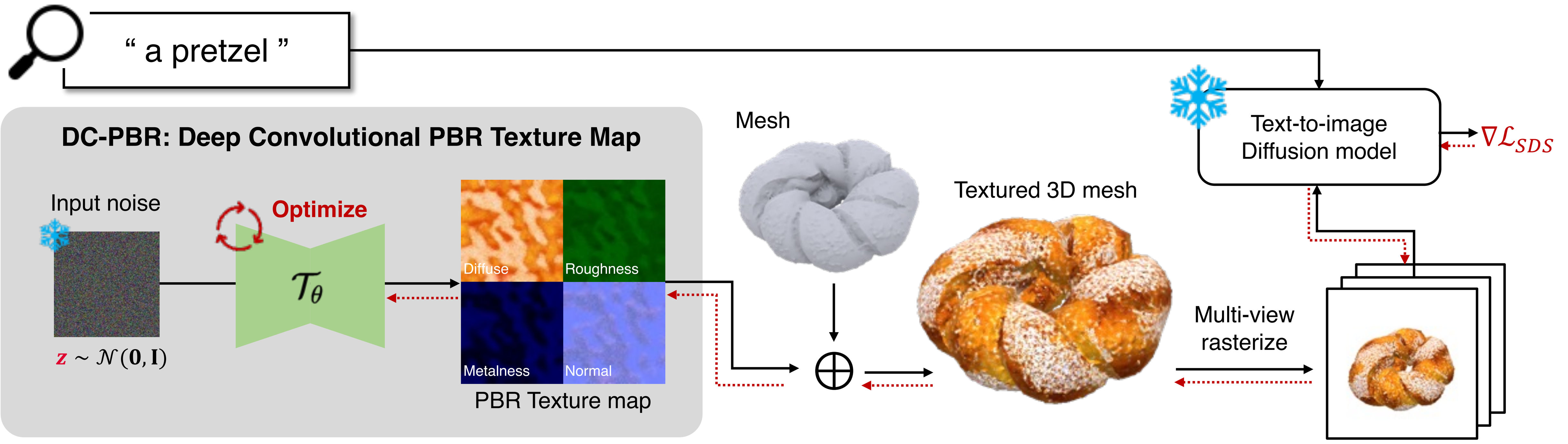


Contributions

- **Paint-it**: Text → 3D mesh texture synthesis method, generating PBR maps compatible with graphics engines
- **DC-PBR**: CNN re-parameterization of PBR texture maps guides frequency-scheduled texture synthesis



Paint-it: Text-Driven PBR Texture Map Synthesis



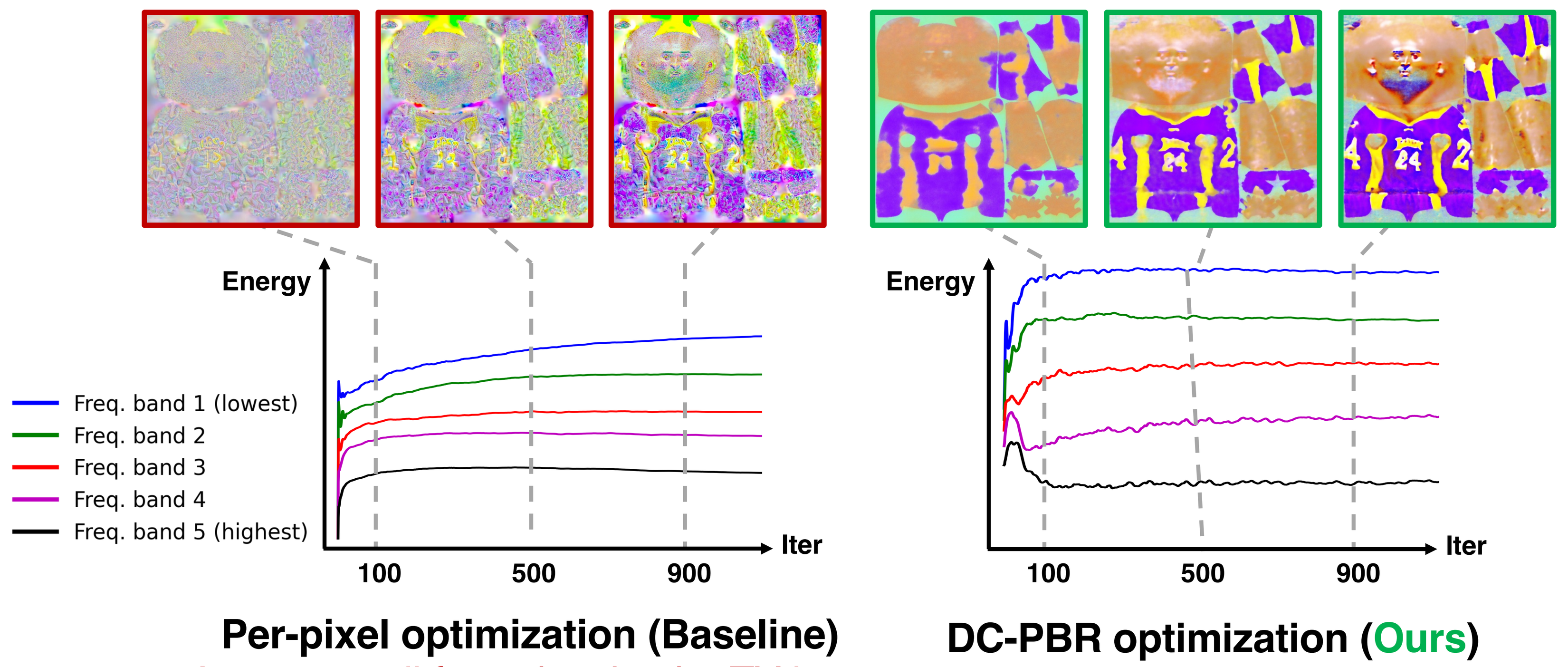
- Step 1: Randomly initialize CNN U-Net ($\mathcal{T}_\theta$) & Fixed input noise ($\mathbf{z}$)
- Step 2: Feed fixed $\mathbf{z}$ to U-Net $\mathcal{T}_\theta$ → Synthesize PBR texture maps
- Step 3: Render 3D mesh with the synthesized PBR texture maps
- Step 3: Update **CNN kernels θ**, using Score-Distillation Sampling (SDS) gradient

$$[\mathbf{K}^d_\theta, \mathbf{K}^{rm}_\theta, \mathbf{K}^n_\theta] = \mathcal{T}_\theta(\mathbf{z})$$

Explicit PBR texture maps — DC-PBR : Our PBR representation

Deep Convolutional PBR Texture Map Re-param. (DC-PBR)

Our observation: Re-parameterizing PBR texture maps with CNN kernels naturally induces **"coarse-to-fine frequency scheduling"** for texture synthesis & noise robustness



Per-pixel optimization (Baseline)
Increases all freq. despite the TV loss

DC-PBR optimization (Ours)
Learns low freq. to high freq.

(Baseline) Per-pixel PBR optimization

$$[\mathbf{K}^{d*}, \mathbf{K}^{rm*}, \mathbf{K}^{n*}] = \arg\min_{\mathbf{K}^d,\mathbf{K}^{rm},\mathbf{K}^n} \mathbb{E}_{t,\epsilon}\left[\|\hat{\epsilon}_\phi(\mathcal{R}^M_t(\mathbf{K}^d,\mathbf{K}^{rm},\mathbf{K}^n);y,t)-\epsilon\|_2^2\right] + \mathcal{L}_{TV}$$

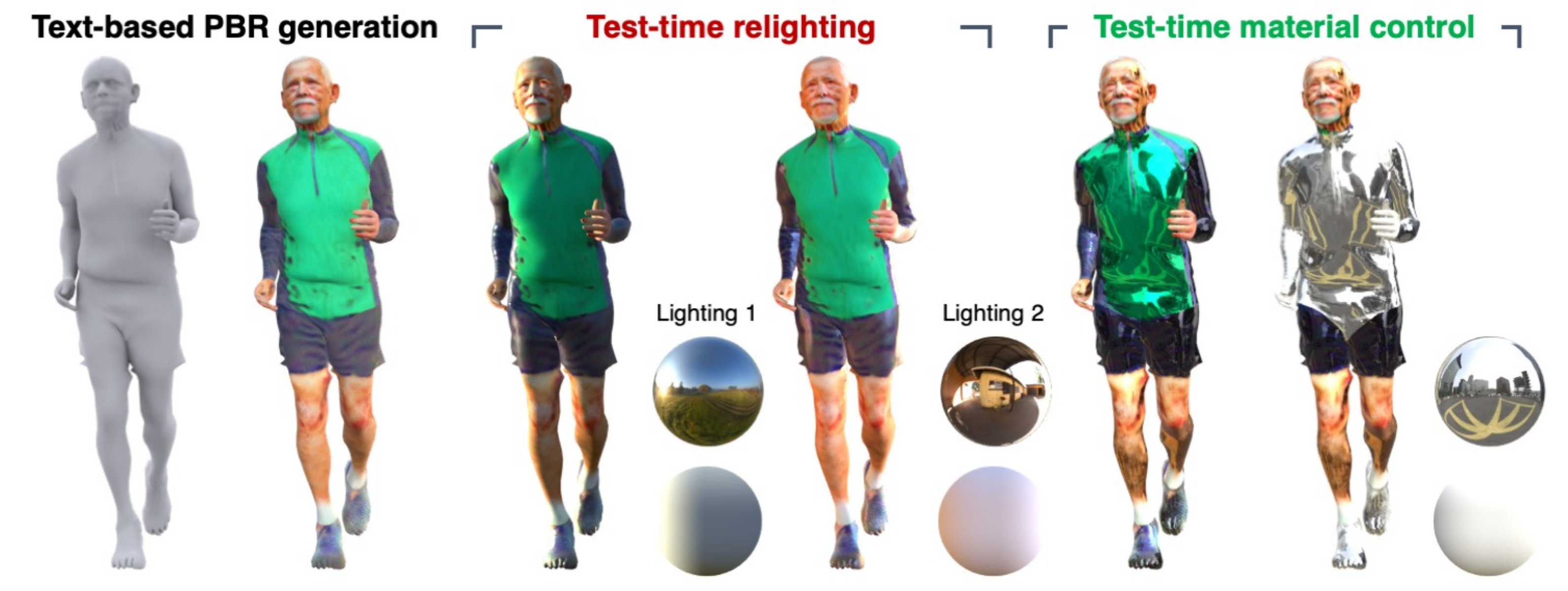
(Ours) DC-PBR re-parameterized optimization

$$\theta^* = \arg\min_\theta \mathbb{E}_{t,\epsilon}\left[\|\hat{\epsilon}_\phi(\mathcal{R}^M_t(\mathbf{K}^d_\theta,\mathbf{K}^{rm}_\theta,\mathbf{K}^n_\theta);y,t)-\epsilon\|_2^2\right],\quad [\mathbf{K}^d_\theta, \mathbf{K}^{rm}_\theta, \mathbf{K}^n_\theta] = \mathcal{T}_\theta(\mathbf{z})$$

PBR Texture Synthesis Results

Qualitative results - Objaverse, RenderPeople

Test-time applications with CG engines



Take-home message

The SDS loss is known to be noisy. We show a better PBR texture representation (parameterization) sufficiently stabilizes the SDS loss and its empirical mechanism.