A. Evaluation Protocol Details

A.1. Blender parameters

We render images using *Blender 3.3.21* with the *Cycles* rendering engine. Eval camera positions are taken from a sphere with a radius of 1.4, where the spherical coordinates φ (azimuth angle) and θ (polar angle) are derived from the setup described in Text2Tex. The HDRI light map from polyhaven used in our experiments is visualized in Figure 7(a); we set the environment map strength to 0.7.

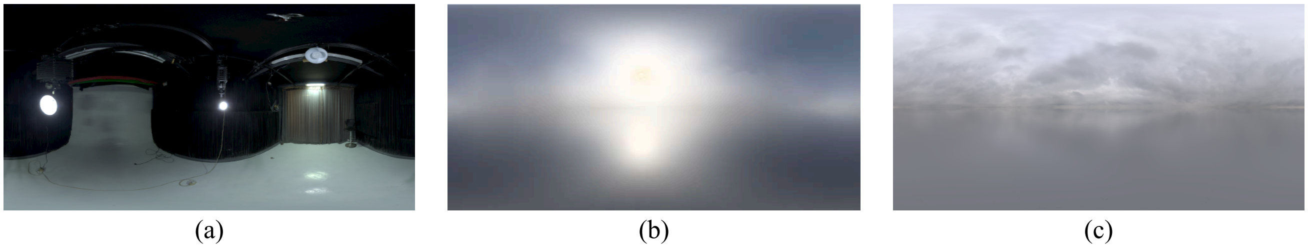
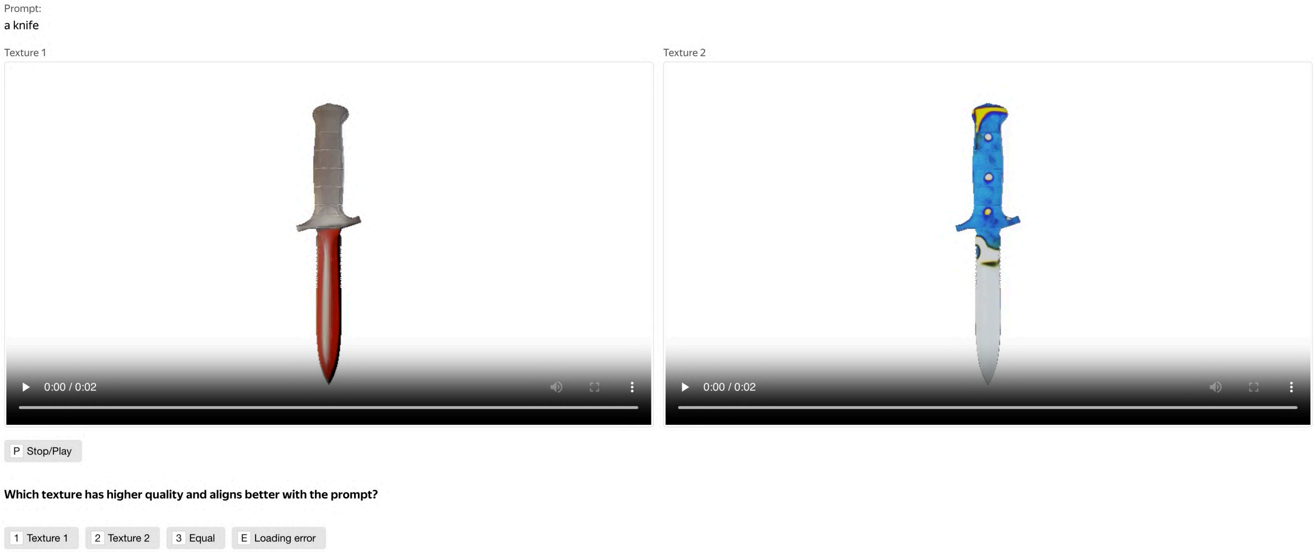

(a) (b) (c)

Figure 7. Used HDRI light maps: *(a)* for evaluation; *(b)* for our pipeline; *(c)* for Paint-it.

A.2. Details of User Study

Every assessor was asked to evaluate the setup shown in Figure 8.


Prompt:
a knife

Which texture has higher quality and aligns better with the prompt?

Figure 8. Assessor's evaluation setup. Professional assessors were tasked with evaluating the quality of the generated textures and selecting the texture that best aligned with the suggested prompt. To ensure unbiased evaluations, the order of the methods was randomly shuffled for each setup.

B. Environment lights

For our texture generation setup, we utilize the HDRI light map visualized in Figure 7(b). In contrast, the Paint-it HDRI light map, shown in Figure 7(c), represents an alternative lighting configuration. All the HDRI light maps used in this work are publicly available for download from polyhaven.

C. Texture Synthesis Results in Varying Setups

In this section, we present textures obtained with different diffusion model combinations to illustrate the impact of model size and the role of the super-resolution module. The results are shown in Figure 9. We also provide additional examples comparing our method with the baselines in Figure 10. Finally, Figure 11 shows a textured model after the first and the second stage of using our cascaded pixel diffusion, compared to the noisy output of latent diffusion.

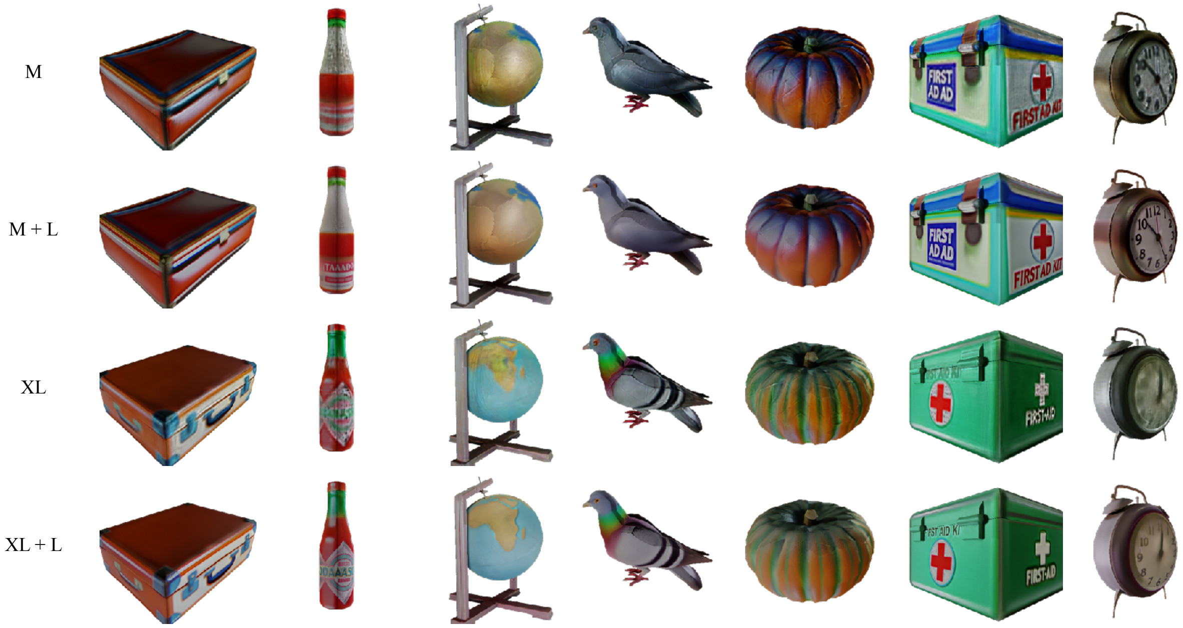


Figure 9. Textures obtained with middle (M) and extra-large (XL) diffusion models along with their versions improved with large (L) super-resolution model on the second stage.

D. High frequency study in DIP

Our investigation into integrating Deep Image Prior (DIP) with our framework reveals certain limitations that affect its practical utility in texture synthesis. Despite being computationally more expensive, our experiments suggest that DIP may not provide sufficient benefits to justify its implementation costs.

As shown in Figure 12, spectral analysis conducted on the Objaverse dataset [11] demonstrates why the parameterization of DIP performs poorly compared to our standard explicit approach. The spectral plot (Figure 12(a)) reveals DIP's significant deficiency in high-frequency components, visually translating to outputs with noticeably less textural detail and surface variation. To further illustrate this effect, we provide additional examples in Figure 12(b) using two cow prompts, which clearly demonstrate the degradation of the detail in the generation based on DIP.

This limitation in creating detailed textures further validates our choice of explicit parameterization, which achieves better visual quality and quantitative metrics with less computational cost.

E. Visualization of PBR Texture Maps

Even though physically based texture improve the overall results, our method occasionally struggles to fully disentangle the lighting effects. We visualize a few examples in Figure 14. For instance, consider partially baked highlights on the coffee maker or the suitcase.

Figure 13 further evaluates prompt-conditioned control on a fixed mesh: we change only the material description and generate the corresponding PBR maps. The generator consistently shifts the distributions of metalness, roughness, and normal amplitude in a physically meaningful way, indicating that it learns material semantics rather than merely recoloring the surface. Metals remain the most sensitive case, where map statistics can still correlate with lighting; we attribute this to limited light map diversity and leave stronger lighting disentanglement for future work.

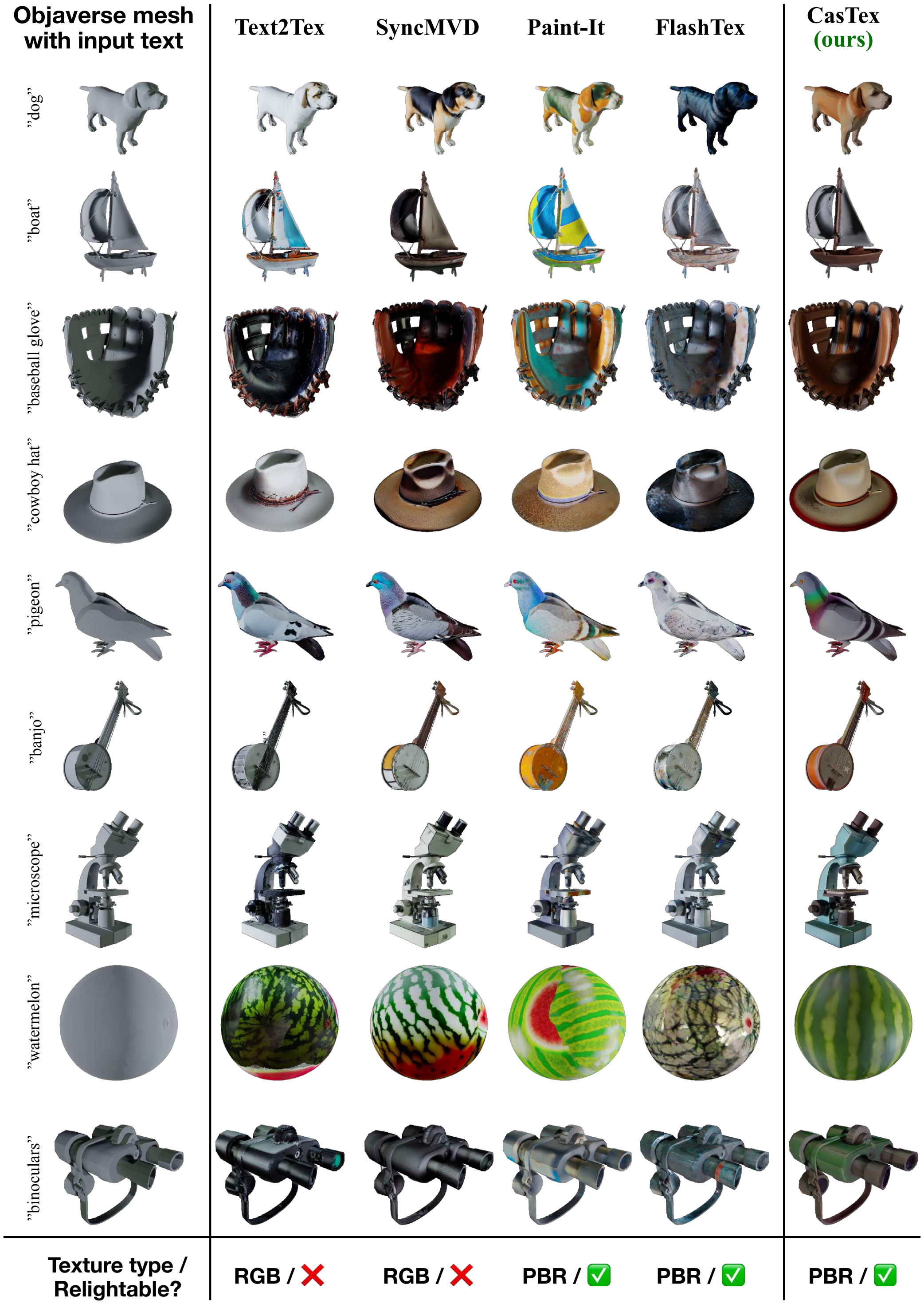


Figure 10. Additional samples from the Objaverse dataset.

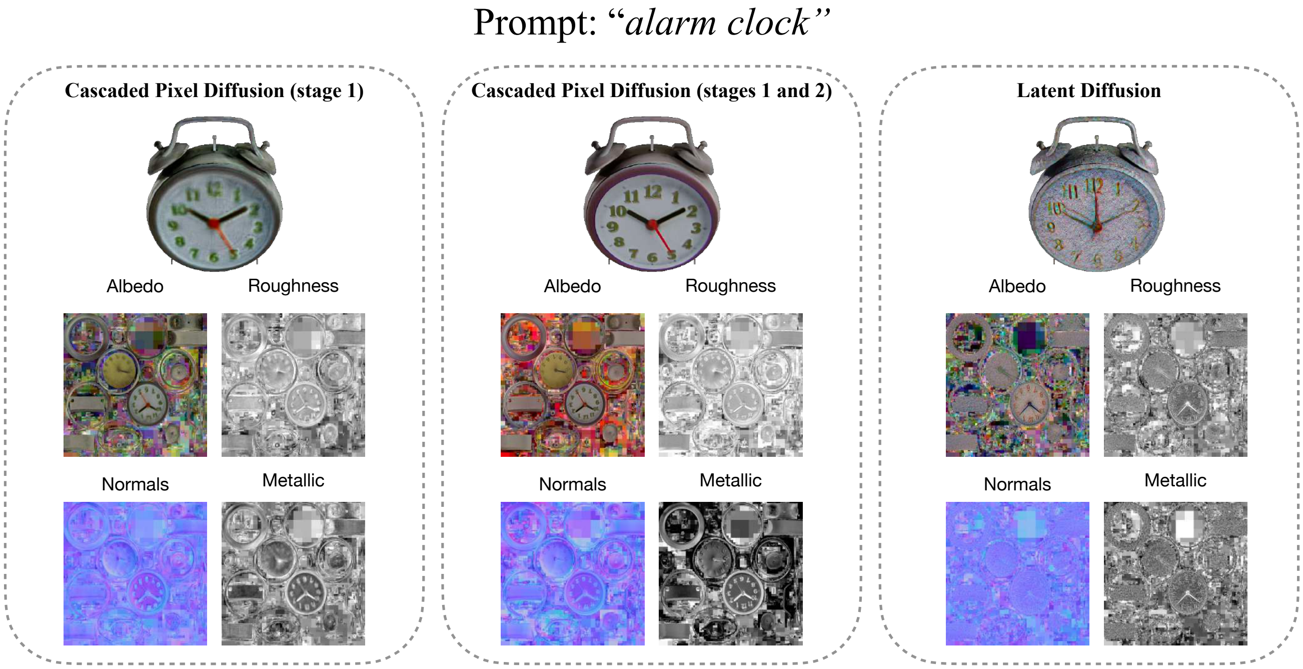

Prompt: "*alarm clock*"
Cascaded Pixel Diffusion (stage 1)
Cascaded Pixel Diffusion (stages 1 and 2)
Latent Diffusion
Albedo
Roughness
Normals
Metallic

Figure 11. Textured model after the first and the second stage of our cascaded pixel diffusion, compared to the noisy result from latent diffusion.

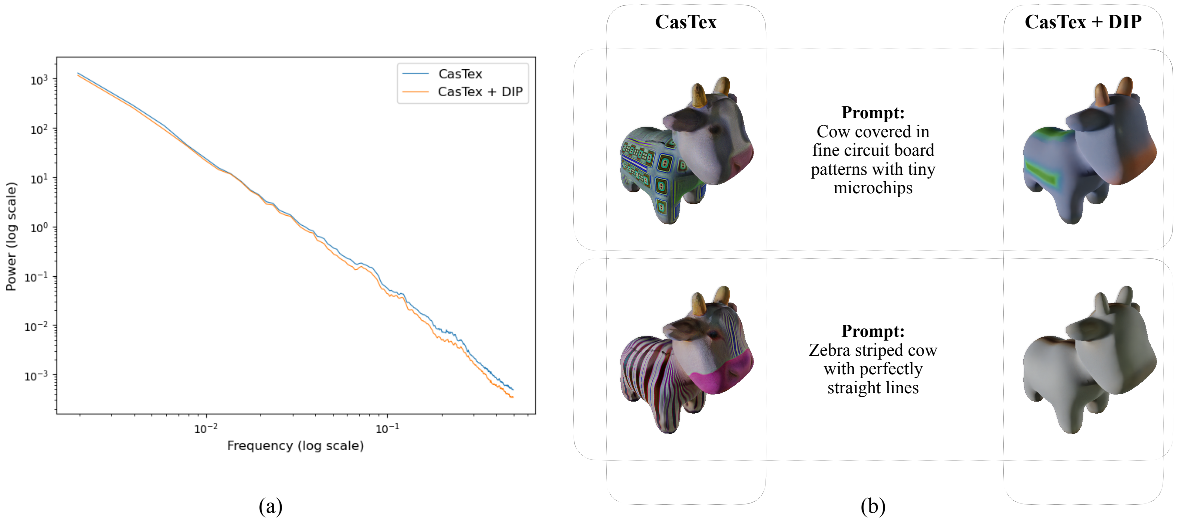


Figure 12. Spectral and visual comparison of texture generation approaches. (a) Power spectrum analysis (log scale) computed on the Objaverse subset, showing CasTex + DIP's reduced power in high-frequency components compared to standard CasTex . (b) Additional examples showing texture generation results for two test prompts. CasTex produces rich, detailed textures while CasTex + DIP generates simplified outputs lacking fine-grained details, directly corresponding to its high-frequency limitations observed in the spectral analysis.

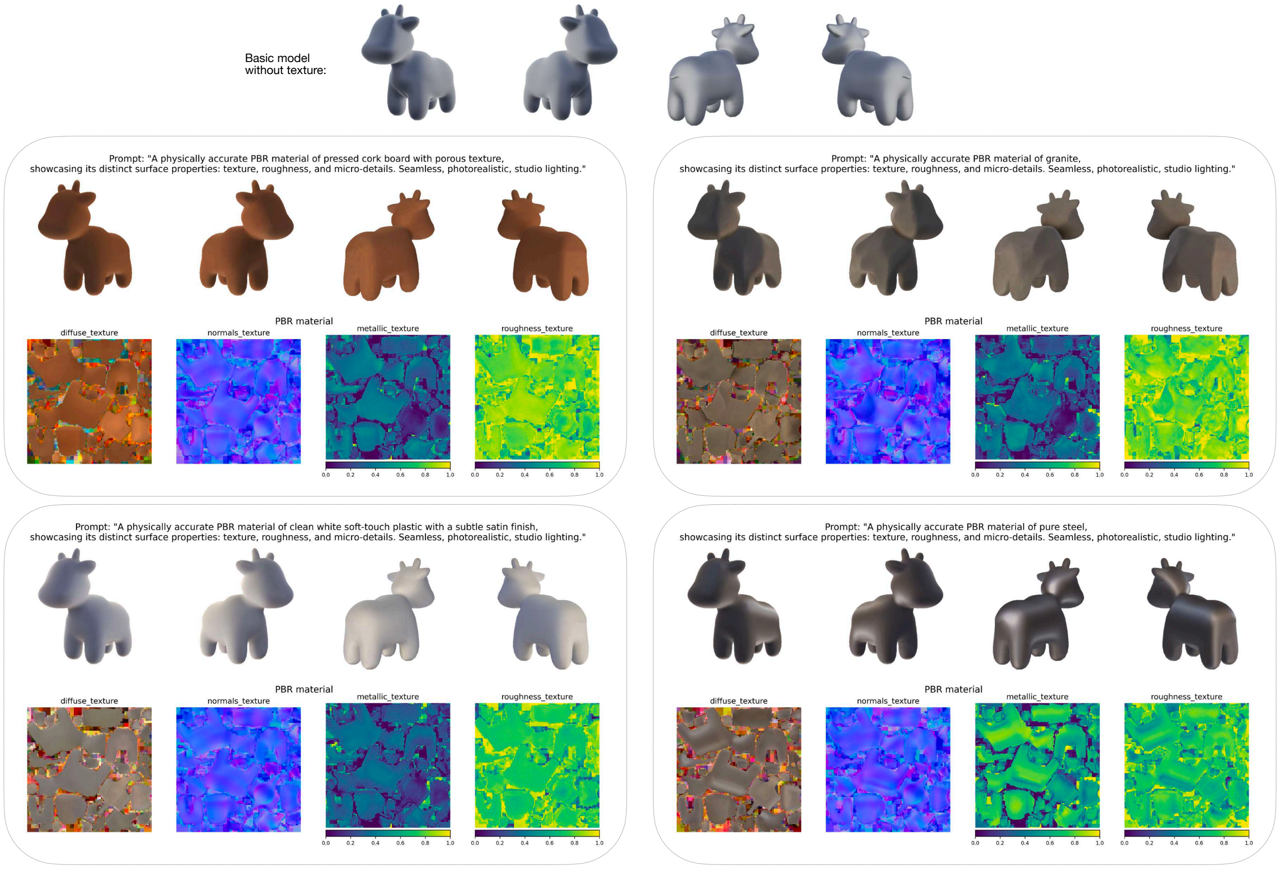

Basic model
without texture:
Prompt: "A physically accurate PBR material of pressed cork board with porous texture,
showcasing its distinct surface properties: texture, roughness, and micro-details. Seamless, photorealistic, studio lighting."
PBR material
diffuse_texture
normals_texture
metallic_texture
roughness_texture
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
Prompt: "A physically accurate PBR material of granite,
showcasing its distinct surface properties: texture, roughness, and micro-details. Seamless, photorealistic, studio lighting."
PBR material
diffuse_texture
normals_texture
metallic_texture
roughness_texture
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
Prompt: "A physically accurate PBR material of clean white soft-touch plastic with a subtle satin finish,
showcasing its distinct surface properties: texture, roughness, and micro-details. Seamless, photorealistic, studio lighting."
PBR material
diffuse_texture
normals_texture
metallic_texture
roughness_texture
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
Prompt: "A physically accurate PBR material of pure steel,
showcasing its distinct surface properties: texture, roughness, and micro-details. Seamless, photorealistic, studio lighting."
PBR material
diffuse_texture
normals_texture
metallic_texture
roughness_texture
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0

Figure 13. Prompt-conditioned PBR textures demonstrating prompt-level control over material properties. *Top-left (cork):* high roughness variation, porous normal detail, near-zero metalness. *Top-right (granite):* stone-like appearance from speckled albedo and fine normals; non-metallic, mid–high roughness. *Bottom-left (soft-touch plastic):* uniform albedo, near-zero metalness, satin roughness, low-amplitude normals. *Bottom-right (steel):* high metalness and lower roughness; residual noise and partially baked highlights remain—likely due to limited environment-lighting variation; improving metal disentanglement is future work.

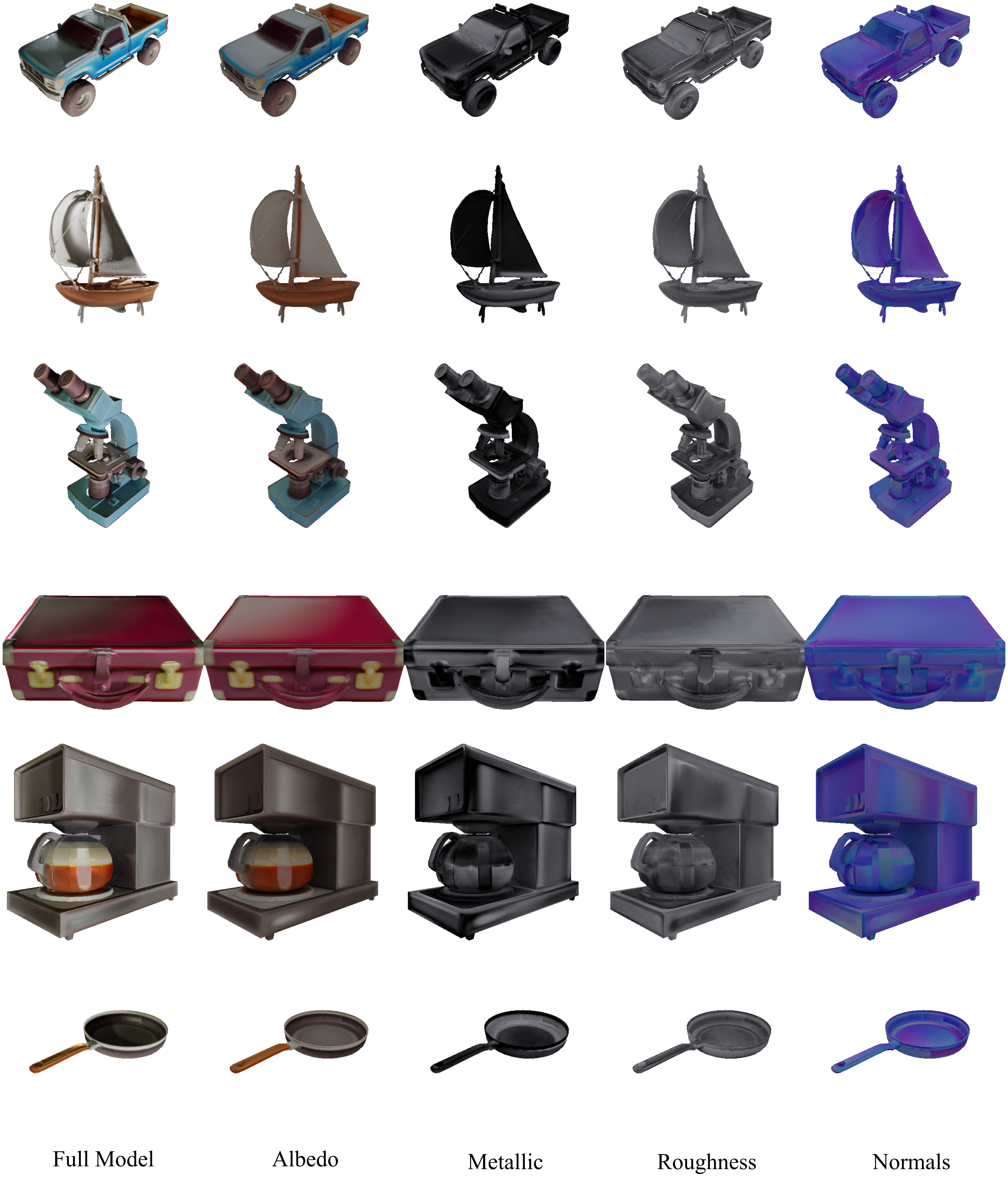


Figure 14. Visualization of the PBR textures. Lighter colors indicate higher metallic and roughness values for the corresponding texutre components.