VideoNeuMat: Neural Material Extraction from Generative Video Models
Supplementary Material

Fig. 1. **Tilable texture results**. We include a few materials examples obtained from our tilable variant. The first image shows the location of the seams. In many cases, the patterns seem to match at the edges, however, the reflectance does seem to differ, causing visible seams. Further architectural changes can improve tilability such as circular convolutions in the LRM, etc.

1 Checkpoint Selection and Prior Preservation

We use the RPC metric to identify the optimal stopping point that balances trajectory compliance with prior preservation. Here we qualitatively validate this selection using out-of-distribution prompts.

We test "cat molded silver" and "dragon molded copper"—concepts absent from MatSynth—to verify that our selected checkpoint indeed preserves semantic knowledge from the pretrained prior. Supplementary Fig. 3 confirms our selection: at iteration 2000, the model generates recognizable cat sculptures and intricate dragon reliefs while following the learned trajectory.

The figure also illustrates what would happen with suboptimal selection. Training beyond our chosen checkpoint leads to progressive semantic collapse: the cat dissolves into featureless planes by 5k, while dragon details degrade to generic MatSynth-like textures by 9k.

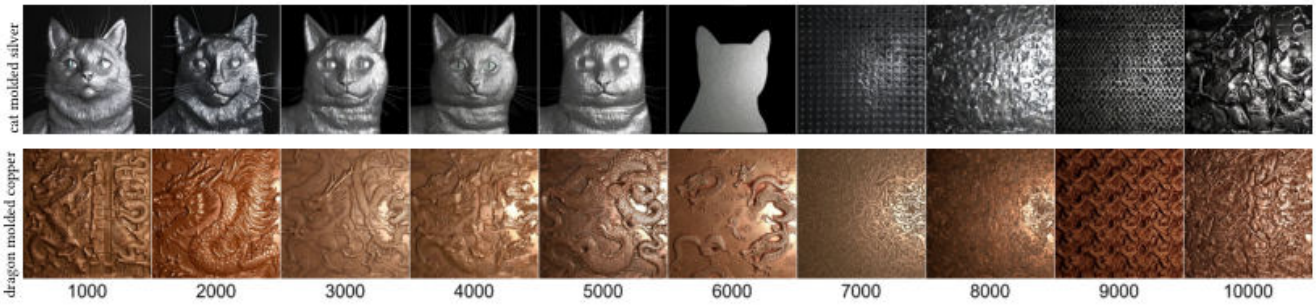

Fig. 2. **Effect of training iterations on out-of-distribution generalization.** Top: recognizable cat sculpture (1k) collapses to featureless planes (5k) then unrelated patterns (6k+). Bottom: detailed dragon reliefs (1k–2k) degrade to generic textures (9k–10k).

Fig. 3. Qualitative results at 512×512 using our model. For each material, the left image shows the inferred texture rendered on a cloth drape, while the right image is a zoomed-in crop highlighting high-quality displacement and parallax.

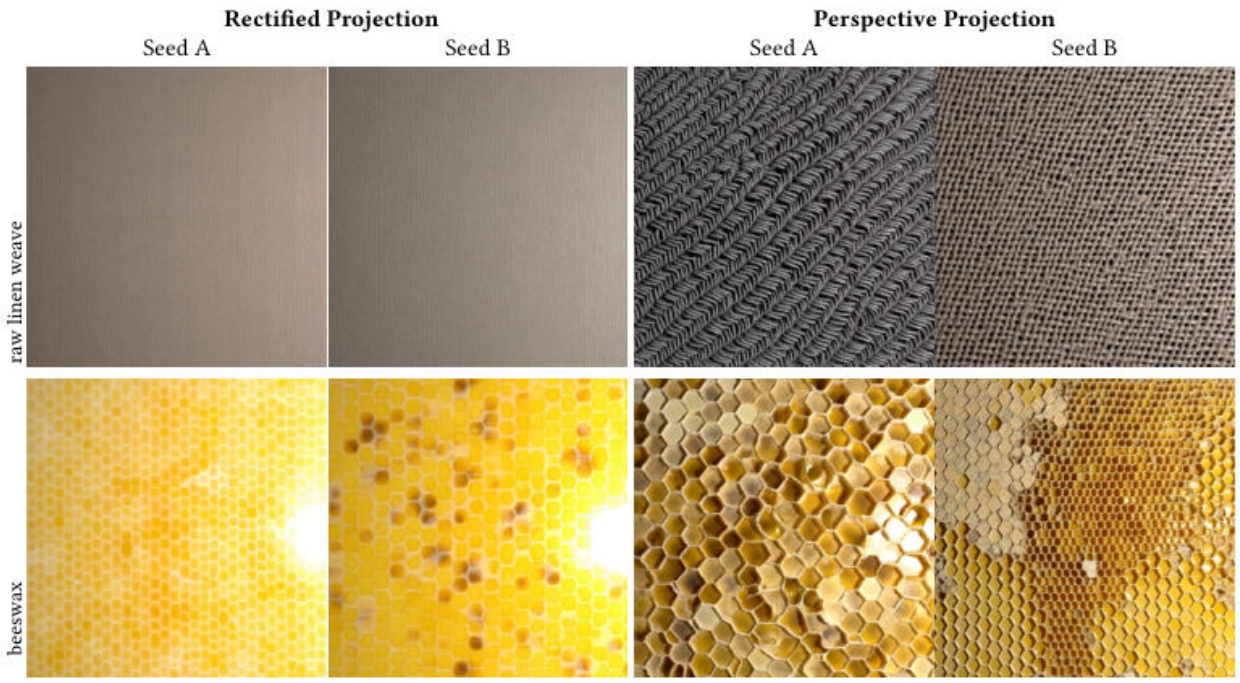

Fig. 4. Ablation on projection mode. For each text prompt, we generate material videos using two different random seeds. **Rectified projection** (left): The model suffers from mode collapse—different seeds produce nearly identical outputs with flat, low-detail textures that lack the diversity of the pretrained prior. **Perspective projection** (right, ours): Different seeds yield visually distinct materials with rich geometric detail and varied appearances, demonstrating that perspective projection preserves the generative diversity of the pretrained video model.

Fig. 5. **Reconstruction results.** A selection of materials generated by our pipeline from text prompts, shown on a flat and curved surface under different illuminations. Note the realism of the results and the ability to handle non-trivial geometry (leaves, fur, fabric) that cannot be represented by heightfields. Please see the extensive supplementary materials for animated results, showing parallax effects.

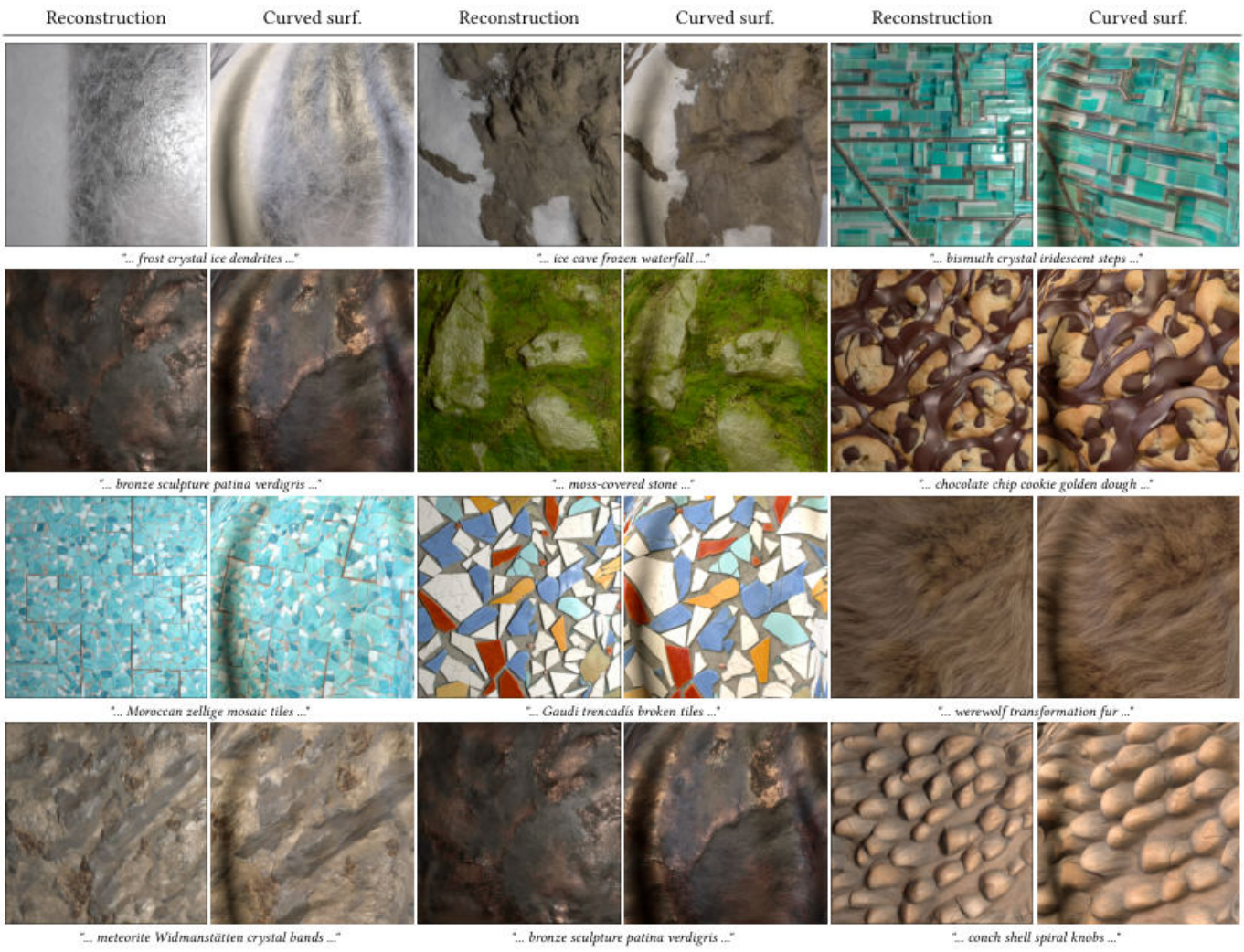

Fig. 6. **Reconstruction results.** A selection of materials generated by our pipeline from text prompts, shown on a flat and curved surface under different illuminations. Note the realism of the results and the ability to handle non-trivial geometry (leaves, fur, fabric) that cannot be represented by heightfields. Please see the extensive supplementary materials for animated results, showing parallax effects.

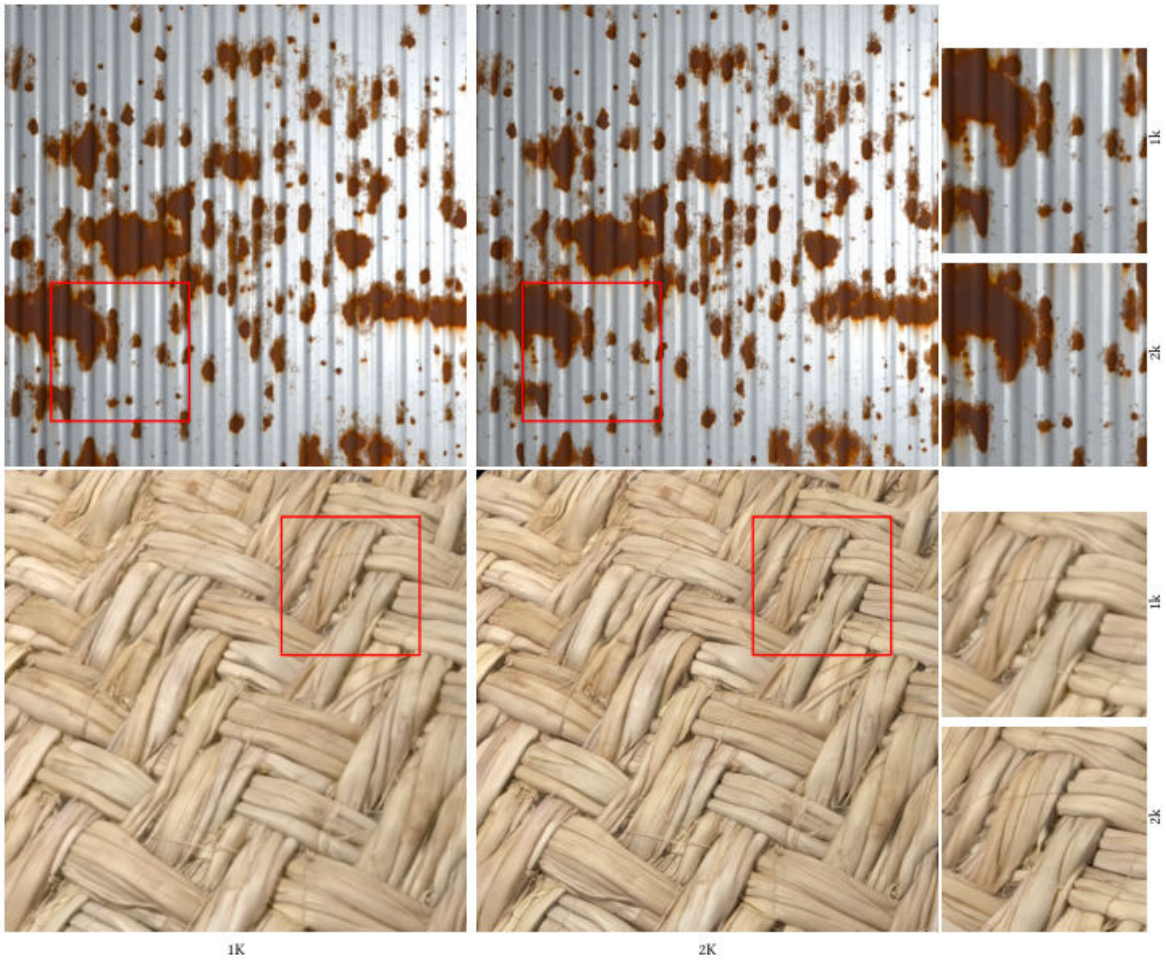

Fig. 7. Comparison showing full images and cropped regions. Red boxes indicate the zoomed area.