

# Supplementary

## MARBLE: Material Recomposition and Blending in CLIP-Space

Ta Ying Cheng  
University of Oxford

Prafull Sharma  
MIT CSAIL

Mark Boss  
Stability AI

Varun Jampani  
Stability AI

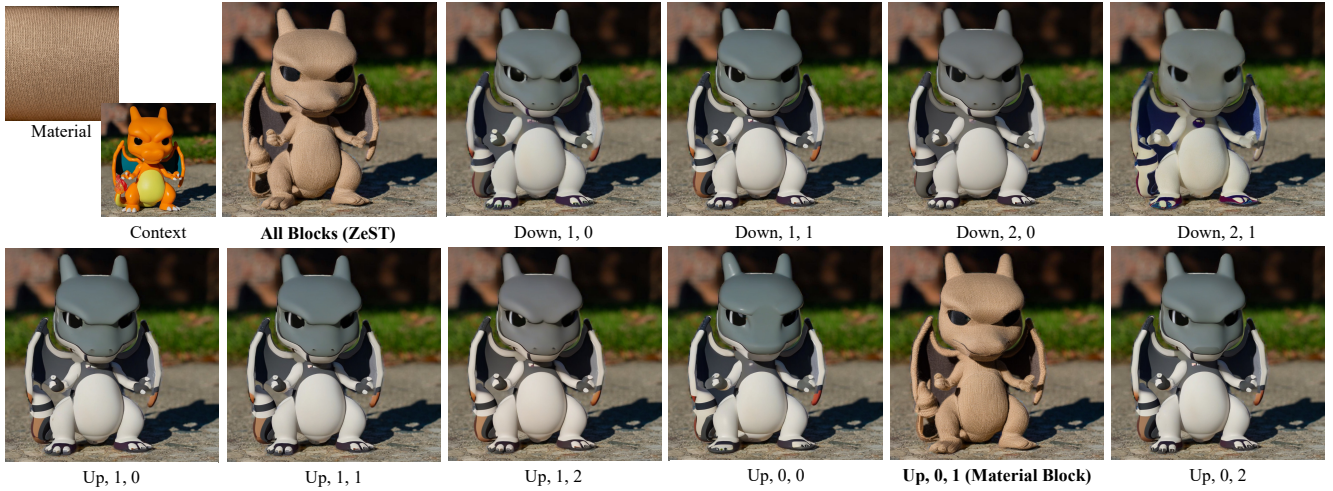


Figure 1. **Material Block Analysis.** We present the material transfer results when injecting only into a certain block. Each block is labelled as `section_blocknumber_attnnumber`. For example, “Up, 0, 1” represents `up_blocks.0.attentions.1`.

### 1. Table of contents

In this supplementary, we provide the following:

- Qualitative study on material block exploration
- Qualitative results on training with very few objects
- Survey example and user study details

### 2. Finding the Material Block

Inspired by InstantStyle [1], we exhaustively visualize the results of the generated images from injecting into each of the individual blocks. Figure 1 provides an example of the visualizations across all the blocks. Similar to the findings of `down_blocks.2.attentions.1` changes the layout (in this case the geometry) of the object and `up_blocks.0.attentions.1` changes the overlay textures. Since we are only interested in transferring only the material information while keeping all other information in tact, we only inject into `up_blocks.0.attentions.1`.

Compared to the results of injecting into all blocks, we see that only material injection preserves the geometry

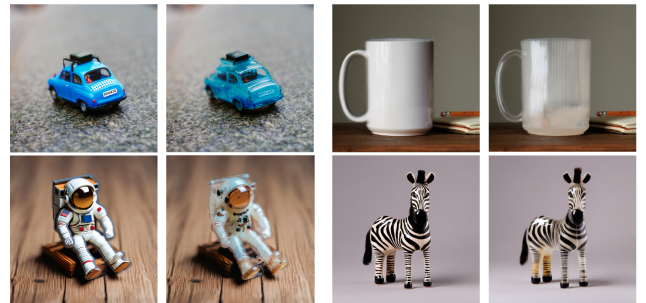


Figure 2. **Transparent results trained with just 6 objects.** We present four examples of the original reconstructed image (left) and increasing their transparency (right).

much better, as presented by the change in the facial structure of the toy figure.

### 3. Real-World Results on Small Datasets.

While we show in the main paper that quantitative metrics for parametric control remains similar even when we train

This is a user study on comparing generative AI model’s ability on controlling materials.

You are given one context image, one instruction, and two result images. Please indicate your preference, i.e., which image is generated better.

Question 1:

Context Image:



Instruction: Make the object surface more rough

Option (A)



Option (B)



Figure 3. **Example User Study.** We present an example user study question where we present an image and a guiding instruction and ask which generated example is preferred.

with a very small dataset, it is still important to analyse the results on real-world datasets.

Figures 2 presents fours examples trained on just 6 objects in the transparency dataset. We can see that even though the extent of transparency is limited, the model already learns the effect of the transparencies fairly well.

## 4. Details on User Study

We provide an example survey template used for our user study, as shown in Figure 3. As DDIM inversion already causes defects in the reconstructed image, most of the participants prefer MARBLE’s generated results even when Concept-Slider sometimes performs parametric control accurately.

## References

- [1] Haofan Wang, Matteo Spinelli, Qixun Wang, Xu Bai, Zekui Qin, and Anthony Chen. Instantstyle: Free lunch towards style-preserving in text-to-image generation. *arXiv preprint arXiv:2404.02733*, 2024. 1