

## Research

[Home](#) [Blog](#)

# Smoothly editing material properties of objects with text-to-image models and synthetic data



July 26, 2024 Mark Matthews and Yuanzhen Li, Software Engineers, Google Research

We present a method that augments an image generation model with parametric editing of the material properties, such as color, shininess or transparency, of objects in any photograph. The resulting parametric model leverages the real-world understanding of generative text-to-image models through fine-tuning with a synthetic dataset.

## QUICK LINKS

[Paper](#)[Share](#)

Many existing tools allow us to edit the photographs we take, from making an object in a photo pop to visualizing what a spare room might look like in the color mauve. Smoothly controllable (or parametric) edits are ideal as they provide precise control over how shiny an object appears (e.g., a coffee cup) or the exact shade of paint on a wall. However, making these kinds of edits while preserving [photorealism](#) typically requires expert-level skill using existing programs. Enabling users to make these kinds of edits while preserving photorealism has remained a difficult problem in computer vision.

Previous approaches like [intrinsic image decomposition](#) break down an image into layers representing “fundamental” visual components, such as [base color](#) (also known as “albedo”), [specularity](#), and lighting conditions. These decomposed layers can be individually edited and recombined to make a photo-realistic image. The challenge is that there is a great deal of ambiguity in determining these visual components: Does a ball look darker on one side because its color is darker or because it’s being shadowed? Is that a highlight due to a bright light, or is the surface white there? People are usually able to disambiguate these, yet even we are occasionally fooled, making this a hard problem for computers.

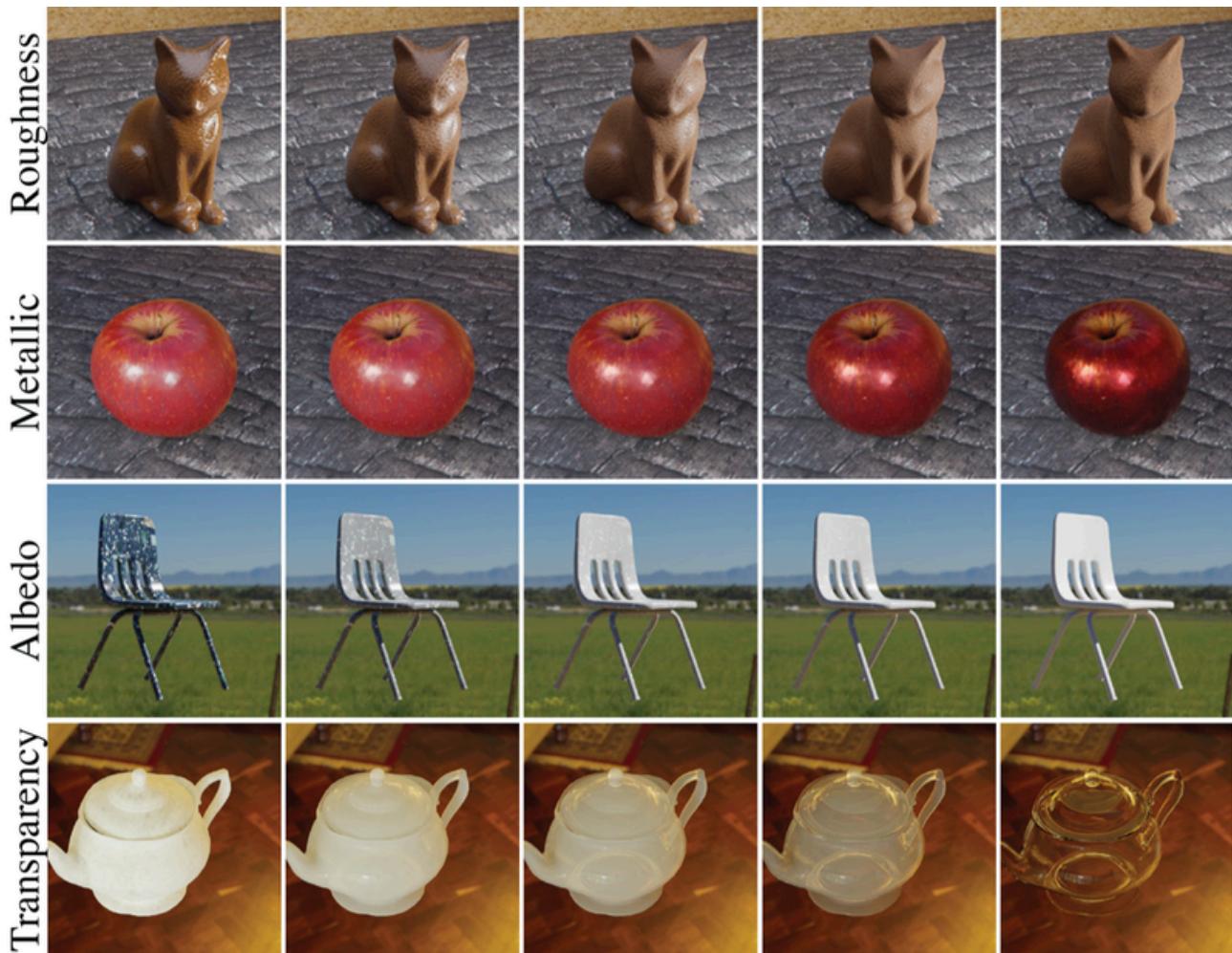
Other recent approaches leverage generative text-to-image (T2I) models, which excel at photorealistic image generation, to edit objects in images. However, [these approaches](#) struggle to disentangle material and shape information. For example, trying to change the color of a house from blue to yellow may also change its shape. We observe similar issues in [StyleDrop](#), which can generate different appearances but does not preserve object shape between styles. Could we find a way to edit the material appearance of an object while preserving its geometric shape?

In “[Alchemist: Parametric Control of Material Properties with Diffusion Models](#)”, published at [CVPR 2024](#), we introduce a technique that harnesses the photorealistic prior of T2I models to give users parametric editing control of specific material properties ([roughness](#), [metallic appearance](#), [base color](#))

## Research

# The Method

We use [traditional computer graphics](#) and [physically based rendering](#) techniques, which have enabled the realism of visual effects in film and TV for years, to render a synthetic dataset that gives us full control of material attributes. We begin with a collection of 100 3D models of household objects of varying geometric shapes. Creating an image of one of these models requires the selection of a material, a camera angle, and lighting conditions. We select these randomly, allowing us to create a large number of “base images” of each object. For each base image, we then change a single attribute of the material, say roughness or transparency, to produce multiple image versions with various *edit strengths* while keeping object shape, lighting and camera angle the same. We define the *edit strength* as a scalar value that changes a material attribute. Defining these is a heuristic design choice, but for simplicity we set 0 as “no change”, -1 as “minimum change”, and +1 “maximum change” depending on the attribute.



*Samples from our synthetic dataset illustrating appearance change for a linear attribute change.*

Next, we modify the architecture of [Stable Diffusion 1.5](#), a latent diffusion model for T2I generation, to accept the edit strength value, enabling the fine-grained control of material parameters we seek. To teach the model how to change *only* the material property we want, we fine-tune it on the dataset of synthetic images that illustrate edits to only the desired material property, simultaneously inputting the corresponding edit strength. The model learns how to edit material attributes given a context image, instruction, and a scalar value defining the desired relative attribute change.

To edit material properties of objects in real-world images, we simply provide the new real-world image to the trained model and input whatever edit strength the user wants. The model generalizes from the relatively small amount of synthetic data to real-world images, unlocking material editing of real-world images while keeping all other attributes the same. This relatively simple approach of fine tuning on a task-specific dataset, demonstrates the power of T2I models to generalize over a wide domain of input images.

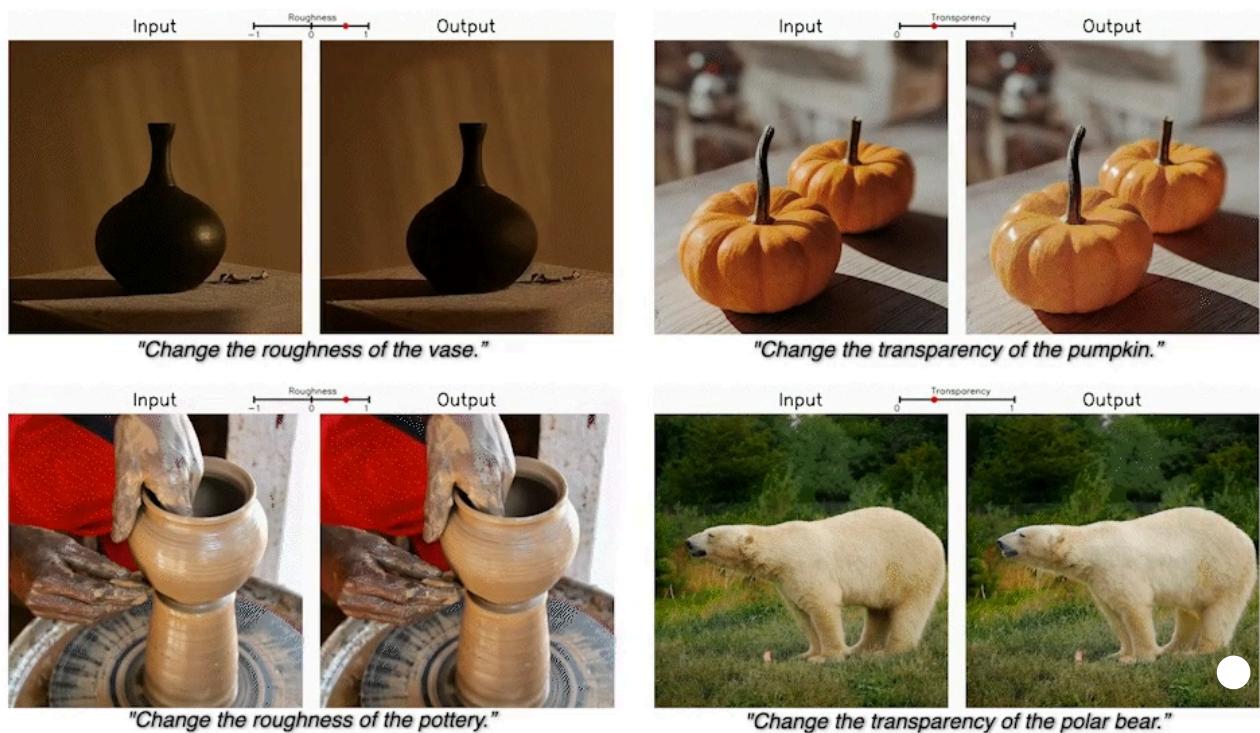
## Results

We were impressed at how well this method works. When asked to make objects metallic, our model effectively changes the appearance of the object while retaining the object’s shape and image lighting the same. When asked to make an object transparent, it realistically fills in the background behind the object, hidden interior structures, and [caustic effects](#) (refracted light moving through the object).

## Research



Examples of edits. **Input** shows a novel, held-out image the model has never seen before. **Output** shows the model output, which successfully edits material properties. Note the caustic lighting effects from the pumpkin, and invisible geometry inside the chair.



## Research



*Smooth editing of material attributes. **Input** shows a novel, held-out image the model has never seen before. **Output** shows model output. Observe how the output image varies material properties smoothly with changing edit strength.*

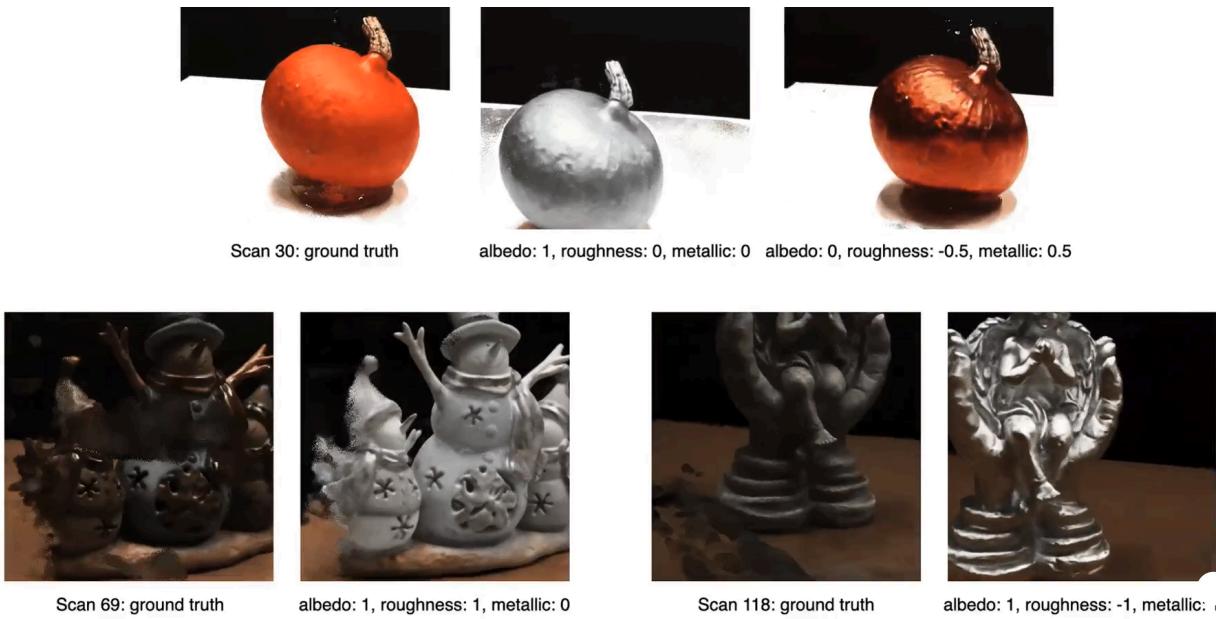
Moreover, in a user study we compared our method to a baseline method, [InstructPix2Pix](#), trained on the same synthetic dataset. Internal volunteers were asked to review 12 pairs of edited images and choose: (1) the most photorealistic image, and (2) the image they preferred. The study reported that our method had more photo-realistic edits (69.6% vs. 30.4%) and was strongly preferred overall (70.2% vs. 29.8%) compared to the baseline method.

## Applications

The potential use-cases of this technology are wide-ranging. Besides being able to more easily imagine what repainting your spare room might look like, architects, artists and designers could more easily mock-up new product designs. We also demonstrate that the edits our model performs are visually consistent, allowing them to be employed in downstream 3D tasks.

For example, given a number of images of a scene, [NeRF](#) reconstruction allows new views to be synthesized. We simply edit the same input images, changing the material appearance of those input images. We then use NeRF to synthesize new views of the scene. We observe 3D consistent renderings of our material property edits in the scene. We show below the result of this process.

## Research



*NeRF material editing. The **top left** shows a NeRF created with ground truth images. In the **top middle** and **top right** images, we edit the input images with our model, then train a new NeRF from scratch. The **bottom** row shows similar edits on other scenes.*

## Conclusion

We present a technique that leverages pre-trained text-to-image models and synthetic data to allow users to edit the material properties of objects in images, in a photorealistic and controllable fashion. Though the model struggles to make hidden detail in some instances, we are encouraged by the method's potential for controllable material edits. See the [paper](#) and [project site](#) to learn more.

## Acknowledgements

*This work would not be possible without the exemplary work of our lead author and Google intern Prafull Sharma. We also thank our paper co-authors Varun Jampani, Xuhui Jia, Dmitry Lagun, Fredo Durand, and William Freeman, as well as Florian Schroff and Hartwig Adam for continuous help in building this technology. We also thank Dilip Krishnan, Jason Baldridge and Sergey Ioffe for their helpful feedback on the paper draft. The authors of this post are now part of Google DeepMind.*

Labels:

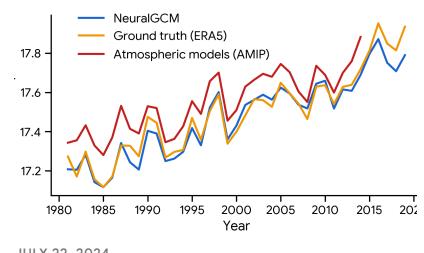
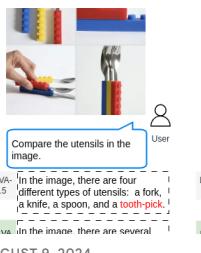
[Generative AI](#)

[Human-Computer Interaction and Visualization](#)

[Machine Intelligence](#)

[Machine Perception](#)

## Other posts of interest



**Google Research**

Generative AI · Machine Intelligence

PREDICTION

Health &amp; Bioscience · Machine Intelligence

Climate &amp; Sustainability · General Science ·

Machine Intelligence ·

Open Source Models &amp; Datasets

Follow us    X            

About Google

Google Products

Privacy

Terms



Help

Submit feedback