Controlled diffusion model can change material properties of objects in images

May 28 2024, by Alex Shipps

MIT CSAIL researchers helped develop a diffusion model that can alter four material properties of objects in images: roughness, metallicity, albedo, and transparency. Credit: Alex Shipps/MIT CSAIL

Researchers from the MIT Computer Science and Artificial Intelligence

Laboratory (CSAIL) and Google Research may have just performed digital sorcery—in the form of a diffusion model that can change the material properties of objects in images.

Dubbed Alchemist, the system allows users to alter four attributes of both real and AI-generated pictures: roughness, metallicity, albedo (an object's initial base color), and transparency. As an image-to-image diffusion model, one can input any photo and then adjust each property within a continuous scale of -1 to 1 to create a new visual. These photo editing capabilities could potentially extend to improving the models in video games, expanding the capabilities of AI in visual effects, and enriching robotic training data.

The magic behind Alchemist starts with a denoising diffusion model: In practice, researchers used Stable Diffusion 1.5, which is a text-to-image model lauded for its photorealistic results and editing capabilities. Previous work had built on the popular model to enable users to make higher-level changes, like swapping objects or altering the depth of images. In contrast, CSAIL and Google Research's method applies this model to focus on low-level attributes, revising the finer details of an object's material properties with a unique, slider-based interface that outperforms its counterparts.

While prior diffusion systems could pull a proverbial rabbit out of a hat for an image, Alchemist could transform that same animal to look translucent. The system could also make a rubber duck appear metallic, remove the golden hue from a goldfish, and shine an old shoe. Programs like Photoshop have similar capabilities, but this model can change material properties in a more straightforward way. For instance, modifying the metallic look of a photo requires several steps in the widely used application.

"When you look at an image you've created, often the result is not exactly what you have in mind," says Prafull Sharma, MIT Ph.D. student in electrical engineering and computer science, CSAIL affiliate, and lead author on a new paper describing the work. "You want to control the picture while editing it, but the existing controls in image editors are not able to change the materials. With Alchemist, we capitalize on the photorealism of outputs from text-to-image models and tease out a slider control that allows us to modify a specific property after the initial picture is provided."

Precise control

"Text-to-image generative models have empowered everyday users to generate images as effortlessly as writing a sentence. However, controlling these models can be challenging," says Carnegie Mellon University Assistant Professor Jun-Yan Zhu, who was not involved in the paper.

"While generating a vase is simple, synthesizing a vase with specific material properties such as transparency and roughness requires users to spend hours trying different text prompts and random seeds. This can be frustrating, especially for professional users who require precision in their work. Alchemist presents a practical solution to this challenge by enabling precise control over the materials of an input image while harnessing the data-driven priors of large-scale diffusion models, inspiring future works to seamlessly incorporate generative models into the existing interfaces of commonly used content creation software," Zhu continued.

Alchemist's design capabilities could help tweak the appearance of different models in video games. Applying such a diffusion model in this domain could help creators speed up their design process, refining textures to fit the gameplay of a level. Moreover, the project by Sharma and his team could assist with altering graphic design elements, videos, and movie effects to enhance photorealism and achieve the desired material appearance with precision.

The method could also refine robotic training data for tasks like manipulation. By introducing the machines to more textures, they can better understand the diverse items they'll grasp in the real world. Alchemist can even potentially help with image classification, analyzing where a neural network fails to recognize the material changes of an image.

The team's work exceeded similar models at faithfully editing only the requested object of interest. For example, when a user prompted different models to tweak a dolphin to max transparency, only Alchemist achieved this feat while leaving the ocean backdrop unedited. When the researchers trained comparable diffusion model InstructPix2Pix on the same data as their method for comparison, they found that Alchemist achieved superior accuracy scores. Likewise, a user study revealed that the MIT model was preferred and seen as more photorealistic than its counterpart.

Keeping it real with synthetic data

According to the researchers, collecting real data was impractical. Instead, they trained their model on a synthetic dataset, randomly editing the material attributes of 1,200 materials applied to 100 publicly available, unique 3D objects in Blender, a popular computer graphics design tool.

"The control of generative AI image synthesis has so far been constrained by what text can describe," says Frédo Durand, the Amar Bose Professor of Computing in the MIT Department of Electrical Engineering and Computer Science (EECS) and CSAIL member, who is a senior author on the paper. "This work opens new and finer-grain control for visual attributes inherited from decades of computer-graphics research."

"Alchemist is the kind of technique that's needed to make machine learning and diffusion models practical and useful to the CGI community and graphic designers," adds Google Research senior software engineer and co-author Mark Matthews. "Without it, you're stuck with this kind of uncontrollable stochasticity. It's maybe fun for a while, but at some point, you need to get real work done and have it obey a creative vision."

Sharma's latest project comes a year after he led research on Materialistic, a machine-learning method that can identify similar materials in an image. This previous work demonstrated how AI models can refine their material understanding skills, and like Alchemist, was fine-tuned on a synthetic dataset of 3D models from Blender.

Still, Alchemist has a few limitations at the moment. The model struggles to correctly infer illumination, so it occasionally fails to follow a user's input. Sharma notes that this method sometimes generates physically implausible transparencies, too. Picture a hand partially inside a cereal box, for example—at Alchemist's maximum setting for this attribute, you'd see a clear container without the fingers reaching in.

The researchers would like to expand on how such a model could improve 3D assets for graphics at scene level. Also, Alchemist could help infer material properties from images. According to Sharma, this type of work could unlock links between objects' visual and mechanical traits in the future.

MIT EECS professor and CSAIL member William T. Freeman is also a senior author, joining Varun Jampani, and Google Research scientists Yuanzhen Li, Ph.D., Xuhui Jia, and Dmitry Lagun. The group's work will be highlighted at CVPR in June.

More information: Prafull Sharma et al, Alchemist: Parametric Control of Material Properties with Diffusion Models (2024)

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Controlled diffusion model can change material properties of objects in images (2024, May 28) retrieved 20 June 2024 from https://techxplore.com/news/2024-05-diffusion-material-properties-images.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.