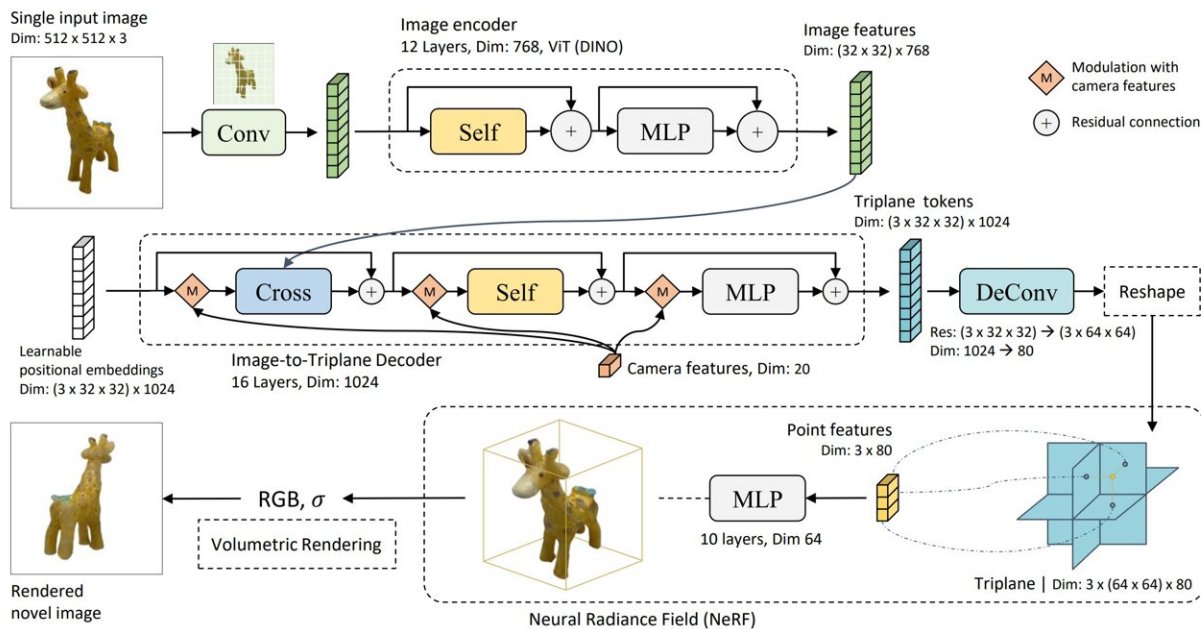# AI model instantly generates 3D image from 2D sample

November 13 2023, by Peter Grad

The overall architecture of LRM, a fully-differentiable transformer-based encoder-decoder framework for single-image to NeRF reconstruction. LRM applies a pre-trained vision model (DINO) to encode the input image (Sec. 3.1), where the image features are projected to a 3D triplane representation by a large transformer decoder via cross-attention (Sec. 3.2), followed by a multi-layer perceptron to predict the point color and density for volumetric rendering (Sec. 3.3). The entire network is trained end-to-end on around a million of 3D data (Sec. 4.1) with simple image reconstruction losses (Sec. 3.4). Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2311.04400

In the rapidly emerging world of large-scale computing, it was just a matter of time before a game-changing achievement was poised to shake up the field of 3D visualizations.

Adobe Research and the Australian National University (ANU) have announced the first artificial intelligence model capable of generating 3D images from a single 2D image.

In a development that will transform 3D model creation, researchers say their new algorithm, which trains on massive samplings of images, can generate such 3D images in a matter of seconds.

Yicong Hong, an Adobe intern and former graduate student of the College of Engineering, Computing, and Cybernetics at ANU, said their large reconstruction model (LRM) is based on a highly scalable neural network containing one million datasets with 500 million parameters. Such datasets include images, 3D shapes, and videos.

"This combination of a high-capacity model and large-scale training data empowers our model to be highly generalizable and produce high-quality 3D reconstructions from various testing inputs," Hong, the lead author of a report on the project, said.

"To the best of our knowledge, [our] LRM is the first large-scale 3D reconstruction model."

Augmented reality and virtual reality systems, gaming, cinematic animation, and industrial design can be expected to capitalize on the transformative technology.

Early 3D imaging software did well only in specific subject categories with pre-established shapes. Hong explained that later advances in image generation were achieved with programs such as DALL-E and Stable

Diffusion, which "leveraged the remarkable generalization capability of 2D diffusion models to enable multi-views." However, results with those programs were limited to pre-trained 2D generative models.

Other systems utilized per-shape optimization to achieve impressive results but they are "often slow and impractical," according to Hong.

The evolution of natural language models within massive transformer networks that utilized large-scale data to maximize next-word prediction tasks, Hong said, encouraged his team to ask the question, "Is it possible to learn a generic 3D prior for reconstructing an object from a single image?"

Their answer was "Yes."

"LRM can reconstruct high-fidelity 3D shapes from a wide range of images captured in the real world, as well as images created by generative models," Hong said. "LRM is also a highly practical solution for downstream applications since it can produce a 3D shape in just five seconds without post-optimization."

The program's success lies in its ability to draw upon its database of millions of image parameters and predict a neural radiance field (NeRF). That's the capacity to generate realistic-looking 3D imagery based solely on 2D pictures—even if those pictures are low-resolution. NeRF has image synthesis, object detection, and image segmentation capabilities.

It was 60 years ago that the first computer program that allowed users to generate and manipulate simple 3D shapes was created. Sketchpad, designed by Ivan Sutherland as part of his Ph.D. thesis at MIT, had a total of 64K of memory.

Over the decades, 3D programs grew by leaps and bounds with such

programs as AutoCAD, 3D Studio, SoftImage 3D, RenderMan and Maya.

Hong's paper, "LRM: Large Reconstruction Model for Single Image to 3D," was uploaded to the preprint server *arXiv* on Nov. 8.

  More information: Yicong Hong et al, LRM: Large Reconstruction Model for Single Image to 3D, *arXiv* (2023). DOI: 10.48550/arxiv.2311.04400

Project page: yiconghong.me/LRM/