

# Nvidia works out speedy process to turn out 3-D models from 2-D images

December 11 2019, by Nancy Cohen

---



Credit: Nvidia

The goal: To change 2-D images into 3-D models using a special encoder-decoder architecture. The actors: Nvidia. The praise: A clever utilization of machine learning with beneficial real-world applications.

Paul Lilly in [Hot Hardware](#) was among the tech watchers who made note that the way they went from 2-D-to-3-D was news. It's no big surprise

when the path is the reverse—3-D into 2-D—but "to create a 3-D [model](#) without feeding a system 3-D data is far more challenging."

Lilly quoted Jun Gao, one of the research team who worked on the rendering approach. "This is essentially the first time ever that you can take just about any 2-D image and predict relevant 3-D properties."

Their magic sauce in producing a 3-D object from 2-D images is a "differentiable interpolation-based renderer," or DIB-R. The researchers at Nvidia trained their model on datasets that included bird images. After training, DIB-R had a capability of taking a bird image and delivering a 3-D portrayal, with the right shape and texture of a 3-D bird.

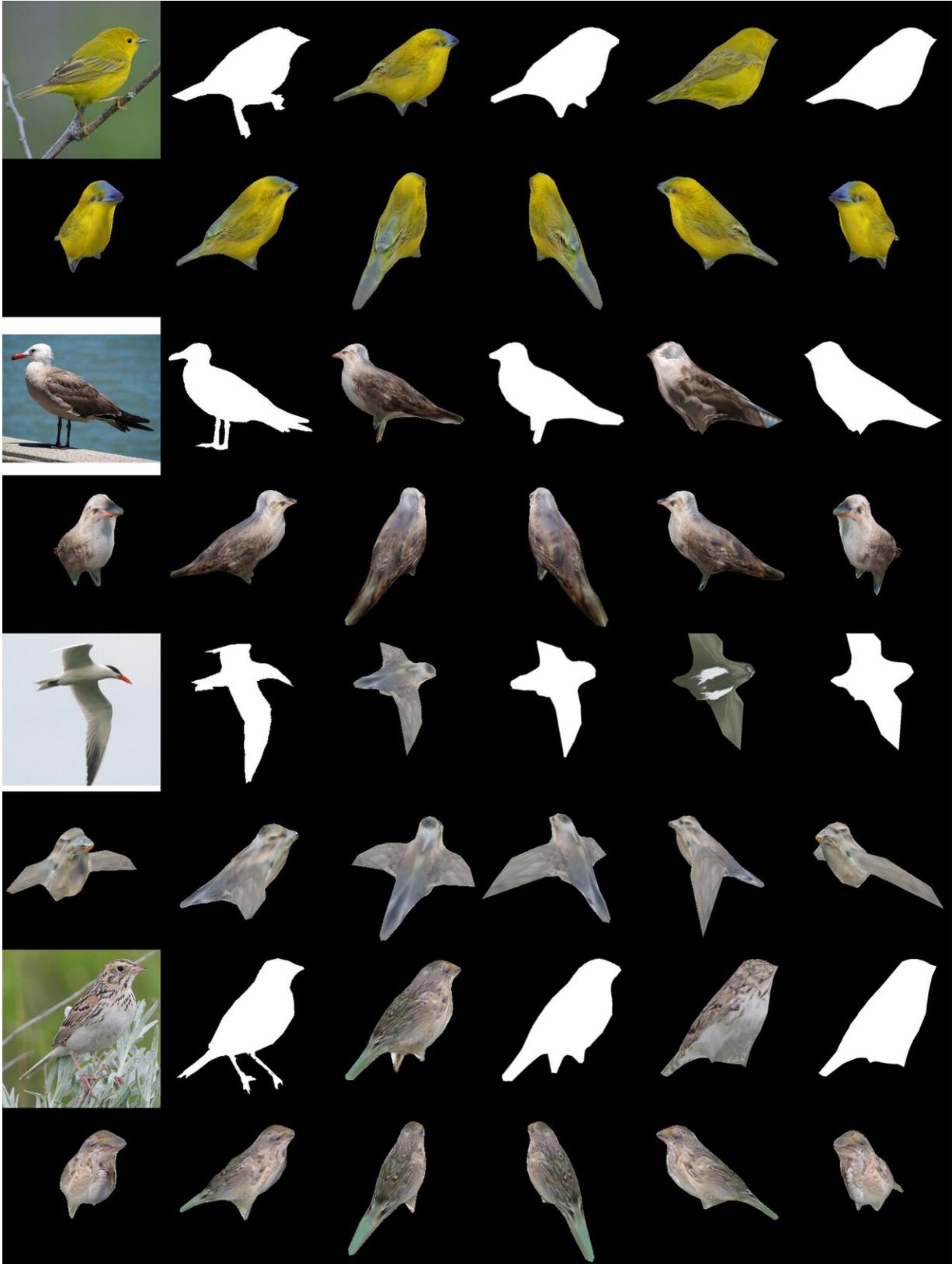
Nvidia further described input transformed into a feature map or vector that is used to predict specific information such as shape, color, texture and lighting of an image.

Why this matters: [Gizmodo](#)'s headline summed it up. "Nvidia Taught an AI to Instantly Generate Fully-Textured 3-D Models From Flat 2-D Images." That word "instantly" is important.

DIB-R can produce a 3-D object from a 2-D image in less than 100 milliseconds, [said](#) Nvidia's Lauren Finkle. "It does so by altering a polygon sphere—the traditional template that represents a 3-D shape. DIB-R alters it to match the real object shape portrayed in the 2-D images."

Andrew Liszewski in *Gizmodo* highlighted this 100-millisecond time element. "That impressive processing speed is what makes this tool particularly interesting because it has the potential to vastly improve how machines like robots, or [autonomous cars](#), see the world, and understand what lies before them."

Regarding autonomous cars, Liszewski said, "Still images pulled from a live video stream from a camera could be instantaneously converted to 3-D models allowing an autonomous car, for example, to accurately gauge the size of a large truck it needs to avoid."



The team tested DIB-R on four 2D images of birds (far left). The first experiment used a picture of a yellow warbler (top left) and produced a 3D object (top two rows). Credit: Nvidia

A model that could infer a 3-D object from a 2-D image would be able to perform better object tracking, and Lilly turned to thinking about its use in robotics. "By processing 2-D images into 3-D models, an autonomous robot would be in a better position to interact with its environment more safely and efficiently," he said.

Nvidia noted that autonomous robots, in order to do so, "must be able to sense and understand its surroundings. DIB-R could potentially improve those depth perception capabilities."

*Gizmodo's* Liszewski, meanwhile, mentioned what the Nvidia approach could do for security. "DIB-R could even improve the performance of security cameras tasked with identifying people and tracking them, as an instantly generated 3-D model would make it easier to perform image matches as a person moves through its field of view."

Nvidia researchers would be presenting their model this month at the annual Conference on Neural Information Processing Systems (NeurIPS), in Vancouver.

Those wanting to learn more about their research can check out their paper on [arXiv](#), "Learning to Predict 3-D Objects with an Interpolation-based Differentiable Renderer." The authors are Wenzheng Chen, Jun Gao, Huan Ling, Edward J. Smith, Jaakko Lehtinen, Alec Jacobson and Sanja Fidler.

They proposed "a complete rasterization-based differentiable renderer

for which gradients can be computed analytically." When wrapped around a neural network, their framework learned to predict shape, texture, and light from single images, they said, and they showcased their framework "to learn a generator of 3-D textured shapes."

In their abstract, the authors observed that "Many machine learning models operate on images, but ignore the fact that images are 2-D projections formed by 3-D geometry interacting with light, in a process called rendering. Enabling ML models to understand image formation might be key for generalization."

They presented DIB-R as a framework that allows gradients to be analytically computed for all pixels in an image.

They said that the key to their approach was "to view foreground rasterization as a weighted interpolation of local properties and background rasterization as an distance-based aggregation of global geometry. Our approach allows for accurate optimization over vertex positions, colors, normals, light directions and texture coordinates through a variety of lighting models."

**More information:** Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer arXiv:1908.01210 [cs.CV] [arxiv.org/abs/1908.01210](https://arxiv.org/abs/1908.01210)

© 2019 Science X Network

Citation: Nvidia works out speedy process to turn out 3-D models from 2-D images (2019, December 11) retrieved 17 April 2024 from <https://techxplore.com/news/2019-12-nvidia-speedy-d-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.