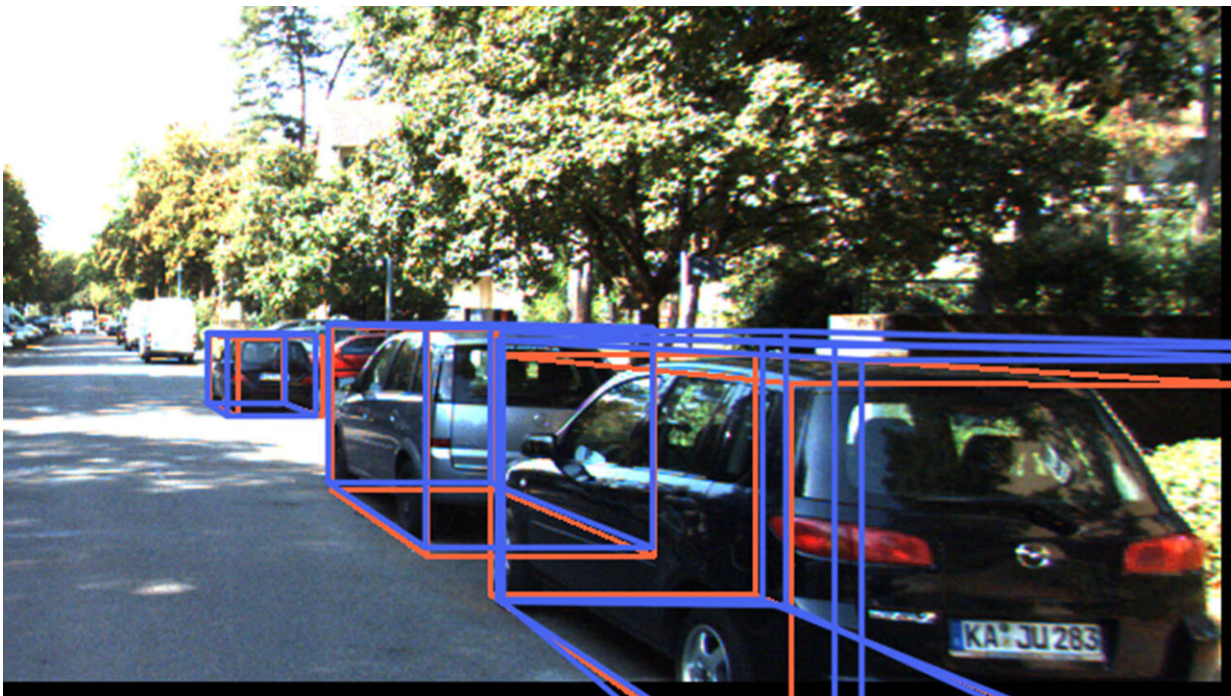


# Technique improves AI's ability to understand 3D space using 2D images

January 26 2022, by Matt Shipman

---



Researchers have developed a new technique, called MonoCon, that improves the ability of artificial intelligence (AI) programs to identify three-dimensional (3D) objects, and how those objects relate to each other in space, using two-dimensional (2D) images. This image shows how MonoCon places objects in "bounding boxes" for use in navigating the street. Credit: Tianfu Wu, NC State University

Researchers have developed a new technique, called MonoCon, that

improves the ability of artificial intelligence (AI) programs to identify three-dimensional (3D) objects, and how those objects relate to each other in space, using two-dimensional (2D) images. For example, the work would help the AI used in autonomous vehicles navigate in relation to other vehicles using the 2D images it receives from an onboard camera.

"We live in a 3D world, but when you take a picture, it records that world in a 2D image," says Tianfu Wu, corresponding author of a paper on the [work](#) and an assistant professor of electrical and computer engineering at North Carolina State University.

"AI programs receive visual input from cameras. So if we want AI to interact with the world, we need to ensure that it is able to interpret what 2D images can tell it about 3D space. In this research, we are focused on one part of that challenge: how we can get AI to accurately recognize 3D objects—such as people or cars—in 2D images, and place those objects in space."

While the work may be important for autonomous vehicles, it also has applications for manufacturing and robotics.

In the context of [autonomous vehicles](#), most existing systems rely on lidar—which uses lasers to measure distance—to navigate 3D space. However, lidar technology is expensive. And because lidar is expensive, autonomous systems don't include much redundancy. For example, it would be too expensive to put dozens of lidar sensors on a mass-produced driverless car.

"But if an autonomous vehicle could use visual inputs to navigate through space, you could build in redundancy," Wu says. "Because cameras are significantly less expensive than lidar, it would be economically feasible to include additional cameras—building

redundancy into the system and making it both safer and more robust.

"That's one practical application. However, we're also excited about the fundamental advance of this work: that it is possible to get 3D data from 2D objects."

Specifically, MonoCon is capable of identifying 3D objects in 2D images and placing them in a "bounding box," which effectively tells the AI the outermost edges of the relevant [object](#).

MonoCon builds on a substantial amount of existing work aimed at helping AI programs extract 3D data from 2D images. Many of these efforts train the AI by "showing" it 2D images and placing 3D bounding boxes around objects in the image. These boxes are cuboids, which have eight points—think of the corners on a shoebox. During training, the AI is given 3D coordinates for each of the box's eight corners, so that the AI "understands" the height, width and length of the "bounding box," as well as the distance between each of those corners and the camera. The training technique uses this to teach the AI how to estimate the dimensions of each bounding box and instructs the AI to predict the distance between the camera and the car. After each prediction, the trainers "correct" the AI, giving it the correct answers. Over time, this allows the AI to get better and better at identifying objects, placing them in a bounding box, and estimating the dimensions of the objects.

"What sets our work apart is how we train the AI, which builds on previous training techniques," Wu says. "Like the previous efforts, we place objects in 3D bounding boxes while training the AI. However, in addition to asking the AI to predict the camera-to-object distance and the dimensions of the bounding boxes, we also ask the AI to predict the locations of each of the box's eight points and its distance from the center of the bounding box in two dimensions. We call this 'auxiliary context,'" and we found that it helps the AI more accurately identify and

predict 3D objects based on 2D images.

"The proposed method is motivated by a well-known theorem in measure theory, the Cramér–Wold theorem. It is also potentially applicable to other structured-output prediction tasks in computer vision."

The researchers tested MonoCon using a widely used benchmark data set called KITTI.

"At the time we submitted this paper, MonoCon performed better than any of the dozens of other AI programs aimed at extracting 3D data on automobiles from 2D images," Wu says. MonoCon performed well at identifying pedestrians and bicycles, but was not the best AI program at those identification tasks.

"Moving forward, we are scaling this up and working with larger datasets to evaluate and fine-tune MonoCon for use in autonomous driving," Wu says. "We also want to explore applications in manufacturing, to see if we can improve the performance of tasks such as the use of robotic arms."

The paper, "[Learning Auxiliary Monocular Contexts Helps Monocular 3D Object Detection](#)," will be presented at the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, being held virtually from Feb. 22 to March 1. First author of the paper is Xienpeng Lu, a Ph.D. student at NC State. The paper was co-authored by Nan Xue of Wuhan University.

**More information:** Xianpeng Liu, Nan Xue, Tianfu Wu, Learning Auxiliary Monocular Contexts Helps Monocular 3D Object Detection. arXiv:2112.04628v1 [cs.CV], [arxiv.org/pdf/2112.04628.pdf](https://arxiv.org/pdf/2112.04628.pdf)

Provided by North Carolina State University

Citation: Technique improves AI's ability to understand 3D space using 2D images (2022, January 26) retrieved 1 June 2023 from <https://techxplore.com/news/2022-01-technique-ai-ability-3d-space.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.