

# New method helps AI navigate 3D space using 2D images

September 25 2023, by Matt Shipman

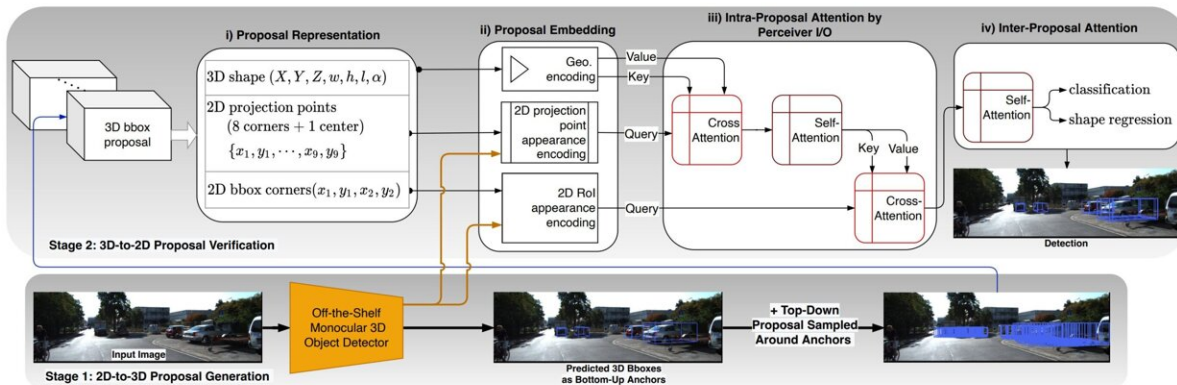


Illustration of the proposed MonoXiver method. MonoXiver is built for any off-the-shelf backbone monocular 3D object detectors. It consists of a 2D-to-3D proposal generation phase using a bottom-up anchoring and top-down sampling strategy, and a 3D-to-2D proposal verification (or denoising) phase using the Perceiver I/O [21] with a carefully designed 3D/2D input space to address the unique challenges. See text for details. Credit: *arXiv* (2023). DOI: [10.48550/arxiv.2304.01289](https://doi.org/10.48550/arxiv.2304.01289)

Photos are two-dimensional (2D), but autonomous vehicles and other technologies have to navigate the three-dimensional (3D) world. Researchers have developed a new method to help artificial intelligence (AI) extract 3D information from 2D images, making cameras more useful tools for these emerging technologies.

"Existing techniques for extracting 3D information from 2D images are good, but not good enough," says Tianfu Wu, co-author of a paper on the work and an associate professor of electrical and computer engineering at North Carolina State University. "Our new method, called MonoXiver, can be used in conjunction with existing techniques—and makes them significantly more accurate."

The work is particularly useful for applications such as autonomous vehicles. That's because cameras are less expensive than other tools used to navigate 3D spaces, such as LIDAR, which relies on lasers to measure distance. Because cameras are more affordable than these other technologies, designers of autonomous vehicles can install multiple cameras, building redundancy into the system.

But that's only useful if the AI in the autonomous vehicle can extract 3D navigational information from the 2D images taken by a camera. This is where MonoXiver comes in.

Existing techniques that extract 3D data from 2D images—such as [the MonoCon technique](#) developed by Wu and his collaborators—make use of "bounding boxes." Specifically, these techniques train AI to scan a 2D image and place 3D bounding boxes around objects in the 2D image, such as each car on a street.

These boxes are cuboids, which have eight points—think of the corners on a shoebox. The bounding boxes help the AI estimate the dimensions of the objects in an image, and where each object is in relation to other objects. In other words, the bounding boxes can help the AI determine how big a car is, and where it is in relation to the other cars on the road.

However, the bounding boxes of existing programs are imperfect, and often fail to include parts of a vehicle or other object that appears in a 2D image.

The new MonoXiver method uses each bounding box as a starting point, or anchor, and has the AI perform a second analysis of the area surrounding each bounding box. This second analysis results in the program producing many additional bounding boxes surrounding the anchor.

To determine which of these secondary boxes has best captured any "missing" parts of the object, the AI does two comparisons. One comparison looks at the "geometry" of each secondary box to see if it contains shapes that are consistent with the shapes in the anchor box. The other comparison looks at the "appearance" of each secondary box to see if it contains colors or other visual characteristics that are similar to the visual characteristics of what is within the anchor box.

"One significant advance here is that MonoXiver allows us to run this top-down sampling technique—creating and analyzing the secondary bounding boxes—very efficiently," Wu says.

To measure the accuracy of the MonoXiver method, the researchers tested it using two datasets of 2D images: the well-established KITTI dataset and the more challenging, large-scale Waymo dataset.

"We used the MonoXiver method in conjunction with MonoCon and two other existing programs that are designed to extract 3D data from 2D images, and MonoXiver significantly improved the performance of all three programs," Wu says. "We got the best performance when using MonoXiver in conjunction with MonoCon."

"It's also important to note that this improvement comes with relatively minor computational overhead," Wu says. "For example, MonoCon, by itself, can run at 55 frames per second. That slows down to 40 frames per second when you incorporate the MonoXiver method—which is still fast enough for practical utility."

"We are excited about this work, and will continue to evaluate and fine-tune it for use in [autonomous vehicles](#) and other applications," Wu says.

The paper, "[Monocular 3D Object Detection with Bounding Box Denoising in 3D by Perceiver](#)" is published on the *arXiv* preprint server.

**More information:** Xianpeng Liu et al, Monocular 3D Object Detection with Bounding Box Denoising in 3D by Perceiver, *arXiv* (2023). [DOI: 10.48550/arxiv.2304.01289](https://doi.org/10.48550/arxiv.2304.01289)

Provided by North Carolina State University

Citation: New method helps AI navigate 3D space using 2D images (2023, September 25)  
retrieved 28 April 2024 from <https://techxplore.com/news/2023-09-method-ai-3d-space-2d.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.