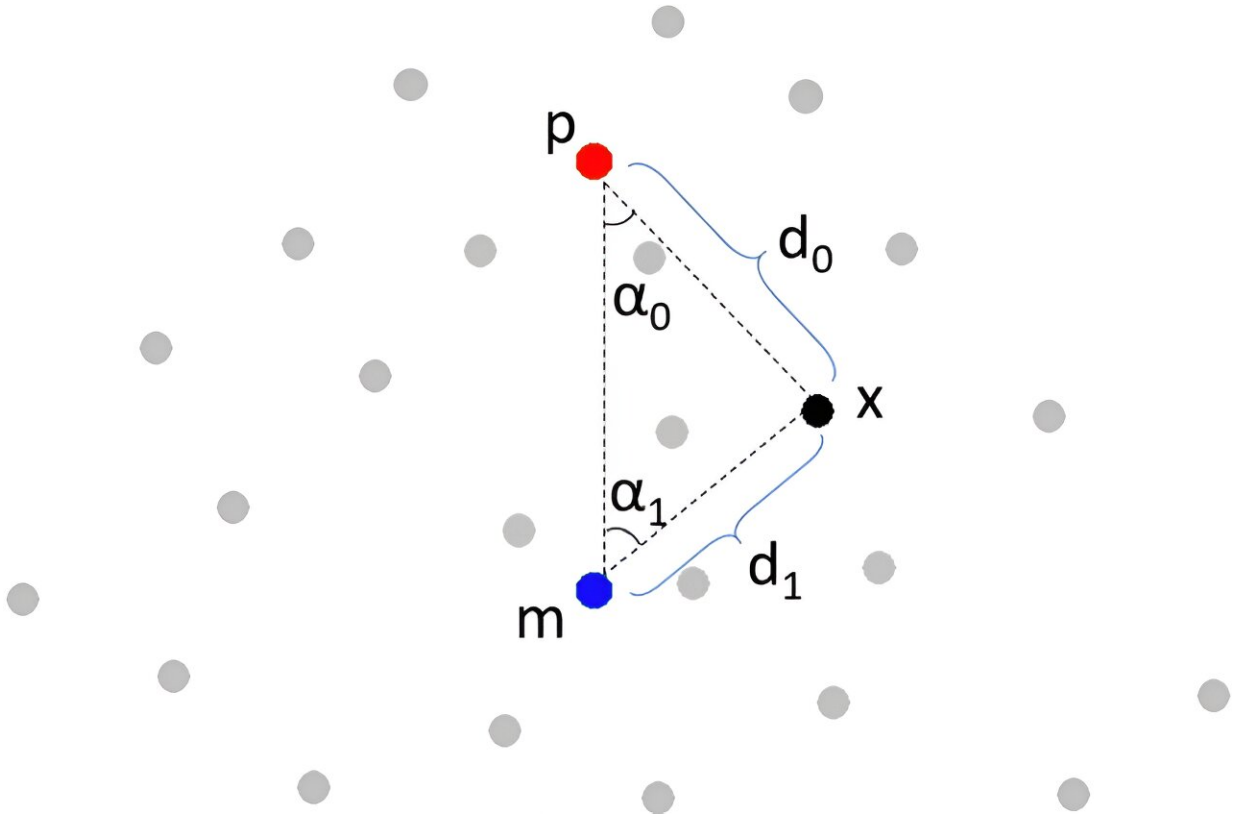


# Seeing 3D images through the eyes of AI

November 27 2023, by Alvin Lee

---



Rotation invariant features at a point  $x$  used by RConv (Zhang et al. 2019a). The distances and angles are constructed based on the reference vector  $pm$ , where  $p$  is the representative point and  $m$  the centroid of the local point set. Such features give good performance in the classification task, but are not as distinctive as features learned from translation-invariant convolutions. Credit: *International Journal of Computer Vision* (2022). DOI: 10.1007/s11263-022-01601-z

Image recognition technology has come a long way since 2012 when a group of computer scientists at the University of Toronto created a convolutional neural network (CNN)—dubbed "AlexNet" after its creator Alex Krizhevsky—that correctly identified images much better than others. Its findings have propelled successful use of CNNs in related fields such as video analysis and pattern recognition, and now researchers are now focusing on 3D deep learning networks.

Unlike the 2D images that AlexNet competently identified, 3D data represent a different challenge. Whereas the basic element of grid-based 2D images is the pixel, and the top left pixel is always the first pixel with all other pixels ordered in relation to it, that is not the case for 3D data. With three dimensions to consider—usually denoted by the axes  $x$ ,  $y$ , and  $z$ —3D CNNs deal with 3D "points" instead of pixels. The collection of these points makes up a 3D point cloud which is the direct output of numerous 3D scanners and the most popular 3D data representation.

"3D point clouds are not well-structured and the 3D points are scattered, sparse and orderless," explains Zhang Zhiyuan, Assistant Professor of Computer Science at the SMU School of Computing and Information Systems. "Thus, traditional [neural networks](#) like CNNs, which can perform well on the well-structured 2D images, cannot apply itself on orderless 3D point clouds directly, and we need to design new convolution operators for 3D point clouds."

## Castle on a (3D point) cloud

This lack of a first 3D point, unlike the first pixel in a 2D image, is called "point order ambiguity." Even though 3D data has been introduced to CNNs with some success and identification accuracy, complex network architecture is necessary and training speed is low. They are also not rotation invariant, i.e., able to identify two objects as rotated versions of each other.

"Existing neural networks can only recognize a 3D object that is in a similar pose contained in the training data," Professor Zhang explains.

"Take 3D human recognition, for example. During the training stage, all the 3D human models are in a standing pose, while in the real application (during the testing stage) the [network](#) is given the same human model but in a lying down pose. The well-trained neural networks cannot recognize it because they have not seen such a pose.

"Thus, rotation invariant networks are important such that they can be more generalized and effective in real applications. We cannot guarantee that the given 3D object is in the same pose as in the training stage. This is especially true for 3D data since it has more freedom of rotation than 2D data." This issue is resolved by Professor Zhang's paper, "RConv++: Effective Rotation Invariant Convolutions for 3D Point Clouds."

RConv++ achieves rotation invariance by designing informative rotation invariant features encoding angles and lengths between the 3D points.

The operation of convolution, which essentially is processing digital data into images comprehensible to humans, and the resolving of point order ambiguity make up the bulk of Professor Zhang's paper, "ShellNet: Efficient Point Cloud Convolutional Neural Networks using Concentric Shell Statistics," published in the *International Journal of Computer Vision*. ShellNet, the CNN that Professor Zhang details in his paper, includes ShellConv, a program that converts 3D data into shell structures and performs 1D (single-dimension) convolution.

"This is very efficient as only 1D convolution is needed," says Professor Zhang, pointing to how it speeds up training. "ShellConv not only achieves efficiency, it also solves the orderless problem in a very elegant way. It converts the orderless point set into shell structures and creates the order from inner to outer shell."

Professor Zhang's research interests encompass Object Oriented Programming as well as Artificial Intelligence, but this specific paper will have significant implications for the fields of autonomous driving, robot navigation, and unmanned aerial vehicles (UAVs) which require accurate and efficient 3D environment perception including 3D object recognition and 3D scene understanding.

He adds, "Nowadays, researchers mostly focus on how to improve the accuracy by designing complicated networks which are hard to run on intelligent moving devices like robots and UAVs. For such devices, lightweight networks are preferred. Thus, ShellConv and ShellNet change the research focus towards the real applications and effective 3D deep learning techniques."

**More information:** Zhiyuan Zhang et al, RConv++: Effective Rotation Invariant Convolutions for 3D Point Clouds Deep Learning, *International Journal of Computer Vision* (2022). [DOI: 10.1007/s11263-022-01601-z](https://doi.org/10.1007/s11263-022-01601-z)

Provided by Singapore Management University

Citation: Seeing 3D images through the eyes of AI (2023, November 27) retrieved 29 April 2024 from <https://techxplore.com/news/2023-11-3d-images-eyes-ai.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.