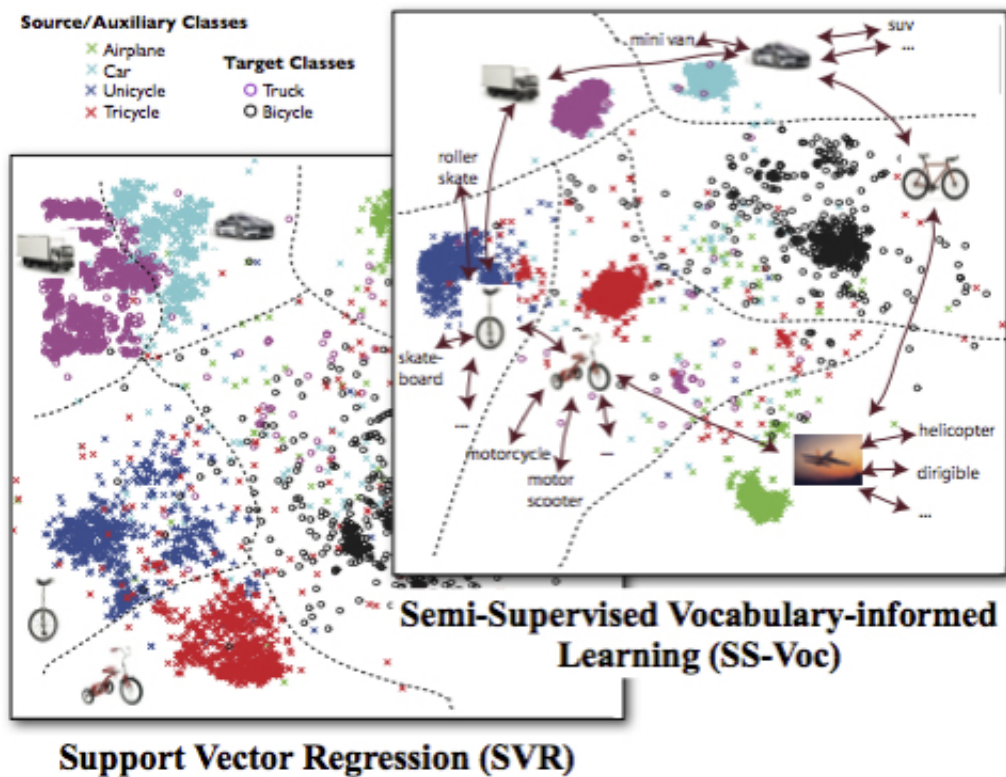


Computer vision system studies word use to recognize objects it has never seen before

June 23 2016



Credit: Disney Research

Computer vision systems typically learn how to recognize an object by analyzing images of thousands of examples. But scientists at Disney Research have shown that computers also can learn to recognize objects they have never seen before, based in part on studying vocabulary.

People, after all, can get an idea of what things might look like based on reading a book. Similarly, a computer that already has been taught to recognize certain objects - apples, for instance - can analyze word use to get hints about the existence of fruits such as pears and peaches, and how they might differ from apples, said Leonid Sigal, senior research scientist at Disney Research.

The knowledge that other fruits exist also is helpful in teaching the computer about important characteristics of apples themselves, he added.

"This opens the door to a new learning paradigm," Sigal said. By reducing the need to train [vision systems](#) with thousands of labeled images, it could help reduce the time necessary for computers to learn new objects and expand the number of object categories that computers can recognize.

Sigal and Yanwei Fu, a post-doctoral researcher at Disney Research, will present this new learning model, called semi-supervised vocabulary-informed learning, at the IEEE Conference on Computer Vision Pattern Recognition, CVPR 2016, June 26 in Las Vegas.

"We've seen unprecedented advances in object recognition and object categorization in recent years, thanks to the development of convolutional neural networks," said Jessica Hodgins, vice president at Disney Research. "But the need to train vision software with thousands of labeled examples for each object has created a bottleneck and limited the number of object classes that can be recognized. Vocabulary-informed learning promises to break that bottleneck and make computer vision more useful and reliable. "

For this study, the computer learned its vocabulary by being trained against all of the articles in Wikipedia and UMBC WebBase, a dataset

with three billion English words. From those articles, it gleaned more than 300,000 [object](#) categories and discovered statistical associations between them. For instance, the computer may have been trained to recognize cars and buses, but from the word analysis it could surmise that there are other categories of vehicles, such as vans, mini-vans and SUVs, and get hints about how each differs from a car or a bus based on its linguistic use.

Simply knowing that these categories exist helps the system as it is trained with images to recognize objects, Sigal said, resulting in the creation of better models for seen objects. Information it gets from the vocabulary analysis can then also suggest how it might recognize other, as-yet unseen objects. If it knows what an apple looks like, for instance, the vocabulary may suggest that a pear, which it has never seen, might be of similar size, but elongated.

"I've never been to Africa, but I read books so I know what to expect," Fu said. "We use our brains to organize information and contextualize how unknown things might look. Compared with previous semi-supervised learning, our vocabulary-informed paradigm is perhaps more similar to how humans reason.

In their testing, Sigal and Fu found that semi-supervised, vocabulary-informed learning worked better and required fewer training examples than other learning techniques, including zero-shot learning, a widely studied approach that introduces new objects during testing, rather than during training.

According to Sigal, [computer vision](#) systems now can recognize thousands of objects, but with this new method they can learn to recognize 300,000 categories based on the vocabulary it developed.

"We didn't try to mimic humans exactly, but making the learning

approach more human-like was a motivating factor," Sigal said. "It is a different form of [learning](#) and so will motivate researchers to develop different types of algorithms."

More information: "Semi-supervised Vocabulary-informed Learning-Paper" [[PDF, 3.49 MB](#)]

Provided by Disney Research

Citation: Computer vision system studies word use to recognize objects it has never seen before (2016, June 23) retrieved 19 April 2024 from <https://phys.org/news/2016-06-vision-word.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.