Google team rises to 2014 visual recognition challenge

September 8 2014, by Nancy Owano

Google's Christian Szegedy, software engineer, blogged Friday about GoogleNet's entry into a visual recognition challenge, the results of which indicate improvements in the state of machine vision technology. The Large Scale Visual Recognition Challenge is the largest academic challenge in computer vision, held annually to test state-of-the-art technology in image understanding—in recognizing objects in images and locating where they are. The competition tracks are classification, classification with localization, and detection. Szegedy detailed the tracks: "The classification track measures an algorithm's ability to assign

correct labels to an image. The classification with localization track is designed to assess how well an algorithm models both the labels of an image and the location of the underlying objects. Finally, the detection challenge is similar, but uses much stricter evaluation criteria."

How did GoogleNet do? GoogLeNet placed first in the classification and detection (with extra training data) tasks, he said.

The test is no piece of cake. Commenting on the significance of this challenge, John Markoff wrote in The New York Times how "a machine's ability to see the world around it is gaining the benefits of bigger computers and more accurate mathematical calculations, an advance that is visible in the contest results. Markoff quoted Fei-Fei Li, the director of the Stanford Artificial Intelligence Laboratory. He was one of the creators of a set of labeled digital images serving as the basis for the contest. Li said he considered this year as a historical year for the challenge. "What really excites us is that performance has taken a huge leap."

One of the features in this challenge is being presented with a lot of images with tiny objects that are hard to recognize. Superior performance in the detection challenge requires pushing beyond annotating an image with a "bag of labels," Szegedy said. "A model must be able to describe a complex scene by accurately locating and identifying many objects in it."

How they did it: Thank something called the DistBelief Infrastructure, This is a software framework that makes it possible to train neural networks in a distributed manner and rapidly iterate. It is "a radically redesigned convolutional network architecture," commented Szegedy.

The architecture involved leads to over 10 times reduction in the number of parameters if compared with most state of the art vision networks.

The advances enable better image understanding and the progress is directly transferable to Google products such as photo and image searches, YouTube, self-driving cars, "and any place where it is useful to understand what is in an image as well as where things are."

At Google, Szegedy works on machine learning, computer vision via deep learning and AI. The team included two PhD students who are in the intern program at Google.

  More information: research.google.com/pubs/pub40565.html googleresearch.blogspot.com/20 … nding-of-images.html

© 2014 Tech Xplore

provided for information purposes only.