Generation query network lets computer create multi-view 3-D model from 2-D photographs

June 15 2018, by Bob Yirka

An artist's interpretation of the paper by S.M. Ali Eslami et al., titled "Neural Scene Representation and Rendering." Credit: DeepMind

A team of researchers working with Google's DeepMind division in London has developed what they describe as a Generation Query Network (GQN)—it allows a computer to create a 3-D model of a scene

from 2-D photographs that can be viewed from different angles. In their paper published in the journal *Science*, the team describes the new type of neural network system and what it represents. They also offer <u>a more personal take</u> on their project in a post on their website. Matthias Zwicker, with the University of Maryland offers a <u>Perspective</u> on the work done by the team in the same journal issue.

In <u>computer</u> science, big jumps in systems engineering can seem small because of the seeming simplicity of results—it is not until someone applies the results that the big leap is truly recognized. This was the case, for example, when the first systems began to appear that were able to listen to a what a person says and extract meaning from it. In this new endeavor, the team at DeepMind might have made a similar leap.

In traditional computer applications, including deep learning networks, a computer must be spoon-fed data in order to behave as if it has learned something. That is not the case for the GQN, which learns purely from observation, like human infants. The system can observe a real-world scene, such as blocks sitting on a table, and then recreate a model of it able to show the scene from other angles. At first glance, as Zwicker notes, this might not seem all that groundbreaking. It is only when considering what the system must do to come up with those new angles that the real power of the system becomes clear. It has to look at the scene and infer characteristics of occluded objects that cannot be observed using only 2-D information provided by cameras. There is no radar or depth finder, or images of what blocks are supposed to look like stored in its data banks. All it has to work with are the few photographs it takes.

Accomplishing this, the team explains, involves using two neural networks, one to analyze the <u>scene</u>, the other to use the resulting data to create a 3-D model of it that can be viewed from angles not shown in the photographs. There is much more work to be done, of course, most

obviously, determining if it can be broadened to more complex objects—but in its primitive form, it clearly represents a new way to allow computers to learn.

Abstract

Scene representation—the process of converting visual sensory data into concise descriptions—is a requirement for intelligent behavior. Recent work has shown that neural networks excel at this task when provided with large, labeled datasets. However, removing the reliance on human labeling remains an important open problem. To this end, we introduce the Generative Query Network (GQN), a framework within which machines learn to represent scenes using only their own sensors. The GQN takes as input images of a scene taken from different viewpoints, constructs an internal representation, and uses this representation to predict the appearance of that scene from previously unobserved viewpoints. The GQN demonstrates representation learning without human labels or domain knowledge, paving the way toward machines that autonomously learn to understand the world around them.

Citation: Generation query network lets computer create multi-view 3-D model from 2-D photographs (2018, June 15) retrieved 18 April 2024 from https://techxplore.com/news/2018-06-query-network-multi-view-d.html