



Dynamic Routing and Capsule Networks

(Copyright © 2018 Piero Scaruffi - Silicon Valley Artificial Intelligence Research Institute)

In 2017 Geoffrey Hinton spoke of the shortcomings of convolutional neural networks in recognizing images (despite the media's euphoria for it) and came up with a new idea, the "capsule network" ("Dynamic Routing Between Capsules", 2017). He first proposed it eight years earlier when he wrote: "This paper argues that convolutional neural networks are misguided in what they are trying to achieve" ("Transforming Auto-encoders", 2011). One problem of "convnets" is that they are "translation invariant": they detect the co-existence of some features and ignore their relative position. If an image has two eyes, a nose and a mouth, it gets classified as a face, even when the eyes are placed below the mouth. But a face is not just an aggregate of some features but an ordered aggregate of such features: it does make a difference whether the mouth is above or below the nose! To overcome this limitation, Hinton designed layers that consist not of individual networks, but rather of capsules, groups of functional networks. Each capsule is programmed to detect a particular attribute of the object being classified. (Of course, a traditional software engineer could just write a little bit of code to specify where the eyes are supposed to be in relation to the nose, but that would be truly old-fashioned).

The other problem of convolutional nets is that they are susceptible to "white-box adversarial attacks": one can easily embed a secret pattern into an image to make it look like something else to the neural network (but not to the human eye). Capsnets should be biologically more plausible. After a quarter of a century, Hinton rediscovered something that he had researched when he was still at UC San Diego surrounded by neuroscientists: a way to generate shape descriptions ("A Parallel Computation that Assigns Canonical Object-based Frames of Reference", 1981). That method was later improved by Bruno Olshausen, Charles Anderson, and David Van Essen at Caltech ("A Neurobiological Model of Visual Attention and Invariant

Pattern Recognition based on Dynamic Routing of Information", 1993), and the result was a better way to represent an object in space.

Anderson and Van Essen had been researching a computational model of visual attention ("Shifter Circuits", 1987), in particular the mechanism that regulates the flow of data within and between cortical areas. Olshausen worked on the hypothesis that the brain contains a population of control neurons whose only job is to route the data flow in the cortex. These neurons implement a process called "dynamic routing", the brain's equivalent of a computer's routing circuits. This provides a more plausible model of object recognition than the collections of loosely related "invariant features" proposed by Fukushima and LeCun with their convolutional networks. Furthermore, Olshausen's model seems to be a better approximation of the ventral stream in the visual cortex.

(Copyright © 2018 Piero Scaruffi - Silicon Valley Artificial Intelligence Research Institute)