

Deep Learning and Attentional Bias in Human Category Learning

Leyla Roksan Caglar
Rutgers University, Newark

Stephen Jose Hanson
Rutgers University, Newark

Abstract: Human category learning is known to be a function of both the complexity of the category rule and attentional bias. A classic and critically diagnostic human category problem involves learning integral stimuli (correlated features) using a condensation rule, or separable stimuli (independent features) using a filtration rule. Human category learning shows differential learning based on category rules that either require attentional binding or ignoring features. It has been shown that neural networks learning with backpropagation cannot differentially learn or distribute attention without built in perceptual bias. In effect neural networks fail to integrate the complexity of learning with the representational bias of the stimuli. In this paper we show that Deep Learning networks, through successive re-encoding and the development of more sensitive feature detectors, learn both the category rules while modeling the attentional bias consistent with the human performance in a task of categorizing realistic 3D-modeled faces.