Convolutional Neural Networks Do Not Rely On Object Features Which Drive Human Overt Attention

MohammadHossein NikiMaleki[1], Hamid Karimi-Rouzbahani[2,3]

[1]Faculty of Computer Science and Engineering, Shahid Beheshti University, Iran

[2]Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, UK

[3]Department of Computing, Macquarie University, Australia

Deep Convolutional Neural Networks (DCNNs) are among the most accurate models of human object recognition. It has been shown that humans rely on specific segments of objects (called MIRCs) for recognition (Ullman et al., 2016). However, DCNNs did not show such sensitivity to identical MIRCs. Therefore, it remains unclear if humans and DCNNs use different mechanisms for object recognition. One potential difference can be that human recognition may involve bottom-up attentional mechanisms which make "different" parts of an image more salient for down-stream mechanisms. Such bottom-up attentional mechanisms are not explicitly incorporated in DCNNs. Computaional models of such bottom-up mechanisms use local low-level image statistics (e.g. color, orientation, contrast) to accurately predict the location of human gaze on the image (Kimura et al., 2013). To test if MIRCs are predicted by salient parts of images, we obtained MIRCs for one of the most brain-like DCNNs (VGG16) using the well-established Bubbles method (Gosselin and Schyns, 2001). We extracted MIRCs from 12 object categories (each including 16 examplars) of the ImageNet dataset (Deng et al., 2009). Results showed that the MIRCs obtained from the DCNN and the salient regions obtained from computational models of human attention were quantitatively and qualitatively different. This suggests that, rather than relying on salient low-level image statistics which may guide human attention, DCNNs may rely on object segments which contain semantic category information critical for object categorization. We are collecting human data to quantitatively compare to the results from our DCNN and the computational models of attention.