

A Simple Computational Model of Saliency Map Formation in the Brain

Abe Leite

Indiana University, Bloomington, Indiana, United States

Abstract

Many convolutional neural network (CNN)-based approaches are excellent functional models of visual attention, but lack cognitive and biological interpretations. In this work, I offer novel, cross-disciplinary justification for the Deep Gaze 1 model, which calculates saliency as a weighted average of feature maps from a pre-trained CNN. In the cognitive realm, experiments demonstrate that visual attention depends on multiple levels of real-world features (edges, text, faces). This is well-modeled using features from a naturalistically-trained CNN. Furthermore, neuroscience research strongly suggests that visual attention is computed in the superior colliculus, using information from multiple levels of the ventral visual stream; all information flow in Deep Gaze follows analogous pathways. To encourage broader adoption of this model, whose source code remains unpublished, I offer a readable implementation with minor changes for biological plausibility. It is validated on the MIT1003 dataset using features from MobileNetV2, with results comparable to the original Deep Gaze.