CognitiveCNN: Mimicking Human Cognitive Models to resolve Texture-Shape Bias

Satyam Mohla *, Anshul Nasery*, Biplab Banerjee, Subhasis Chaudhari
Indian Institute of Technology Bombay, Mumbai, India
satyammohla@gmail.com

April 26, 2020

Abstract

Recent works demonstrate the texture bias in Convolutional Neural Networks (CNNs), conflicting with early works claiming that networks identify objects using shape. It is commonly believed that the cost function forces the network to take a greedy route to increase accuracy using texture, failing to explore any global statistics. We propose a novel intuitive architecture, namely CognitiveCNN, inspired from feature integration theory in psychology to utilise human-interpretable feature like shape, texture, edges etc. to reconstruct, and classify the image. We define two metrics, namely *TIC* and *RIC* to quantify the importance of each stream using attention maps. We introduce a regulariser which ensures that the contribution of each feature is same for any task, as it is for reconstruction; and perform experiments to show the resulting boost in accuracy and robustness besides imparting explainability. Lastly, we adapt these ideas to conventional CNNs and propose *Augmented Cognitive CNN* to achieve superior performance in object recognition.

*Equal Contribution