Project Proposal
Image Saliency Prediction

G Pradyumn
2018CSB1088
Divyanshu Mathpal
2018CSB1086
Indian Institute of Technology, Ropar
Punjab, India

Abstract

The task is to predict the spatial locations of interesting image regions. Saliency is what is interesting or distinctive in a photo or scenario, enabling your eyes to quickly focus on the most important regions. Given an image, the brain selects part of the scene for further detailed processing, whereas the rest is discarded.It also prioritizes the data in such a way, that the selected parts are processed first
This selection and ordering process is known as selective attention or visual saliency. We attempt to make a prediction at this 'saliency' property of humans. The input will be an image, the output of the model will be an image with the salient features highlighted. The dataset that we are going to use is the SALICON dataset made from MS COCO dataset.

Specifications

The importance of saliency prediction lies in their applications in the various fields that they serve, such as computer vision, robotics, and graphics. One of the advantages of saliency map prediction is that it narrows down the region of interest. Reduction in the amount of computation and time follows.
CNN's tend to perform better than the classical approaches due to their ability to extract low-level and high-level features. The high-level feature helps in the robustness of the saliency map with respect to localization of the object and the low-level feature helps in refining the saliency map.
The baseline that we are considering is the following paper
https://arxiv.org/pdf/1611.09571v4.pdf. Multiple approaches have been compared in the paper. In this paper, gaze maps are computed with a feed-forward network and present a novel model that can predict accurate saliency maps by incorporating neural attentive mechanisms. The star feature of the paper is the Convolutional LSTM that focuses on the most salient regions of the input image to iteratively refine the predicted saliency map.
LSTM, which has traditionally be used in NLP, for capturing long-distance relations in the text has been used along with the traditional convolutional neural networks of computer vision tasks. This helps us in encapturing relationships between high level and low-level features, making it better than other approaches, whereas CNN's outperforming of other classical techniques making the model even better.