Sharif University Of Technology

Advanced Topics in Neuroscience

Simulation 08 (Visual Attention)

Instructor: Dr. Ali Ghazizadeh

Zahra Kavian - 98102121

Contents

List of Figures

Section 1: Eye Tracking Database

The datasets get by an eye tracker that records the data point that a subject fixed on them the first time. I plot the fixation points of a subject for some images in figure 1:5.

Most attention is on the points which have contrast with their background. Also, in figure 5, we can see the fixation points are on the people's faces and the text (the nearest player is more attractive than the players who are on the darker background).

Section 2: Saliency Model

In this part, saliency map is found for each images. It reflects the importance of a pixel to human visual system base on a specific features, such as color, edges and etc. I show the saliency map for some images base on different features. I think saliency map base on "Subband",and "Torrallba", and whole features are more similar to eye-track map. The subject first look at the pixel which are brighter (more important in saliency map).

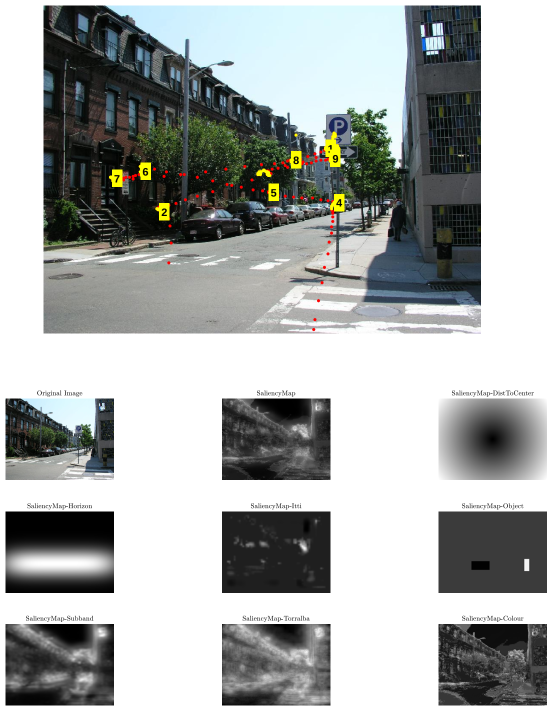

Figure 1: Eye-track and salinecy map (base on each seven features)

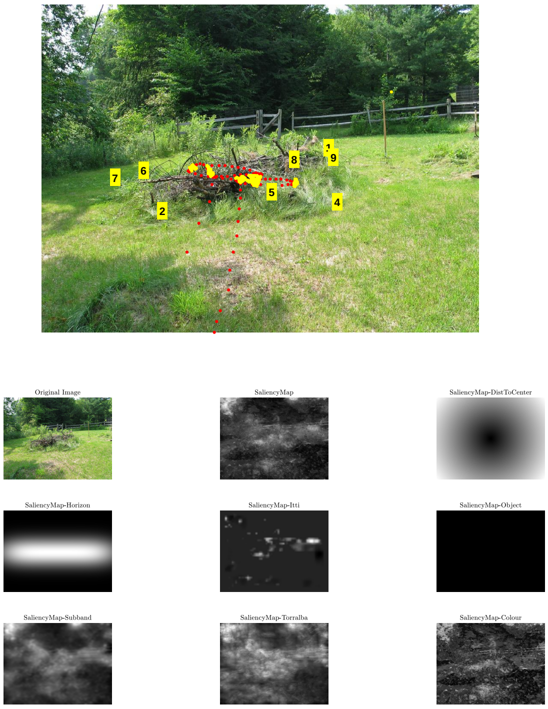

Figure 2: Eye-track and salinecy map (base on each seven features)

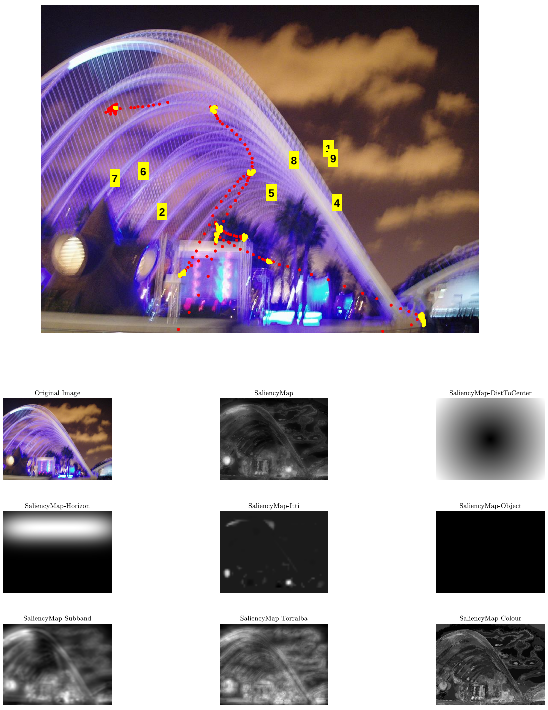

Figure 3: Eye-track and salinecy map (base on each seven features)

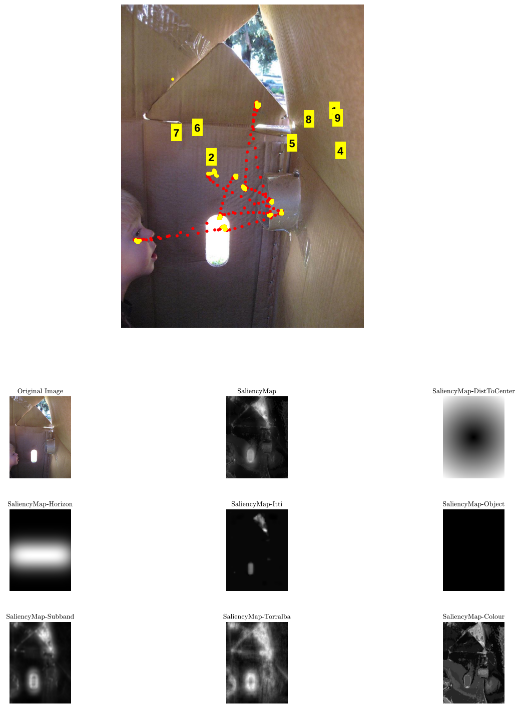

Figure 4: Eye-track and salinecy map (base on each seven features)

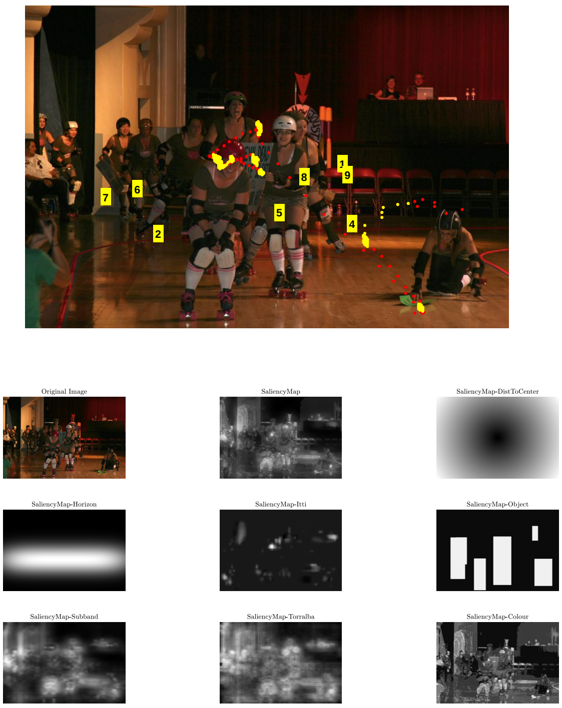

Figure 5: Eye-track and salinecy map (base on each seven features)

More information about the features:[1]

Low-level features:

1- The local energy of the steerable pyramid filters as features (subbands in four orientations and three scales).

2- Intensity, orientation and color contrast detection (Itti and Koch's method).

3- The values of the red, green and blue channels, the probabilities of each of these channels, and the probability of each color (computed from 3D color histograms of the image filtered with a median filter at 6 different scales, Torralba & Rosenholtz method).

Mid- level features:

1- Horizon line detection (where humans naturally look for salient objects).

High-level features:

1- Face detection (Viola Jones)

2- Person detection (Felzenszwalb)

3- The distance to the center (human naturally frame an object of interest near the center of the image).

Section 3: Comparison Saliency Maps to Fixations

Area under curve (ROC) is used to measure the performance of the saliency map and how much it could find the fixation point. At first, I run the model with on of features at a time and compare the performance of it with the model which contain all features (I only use 320 images to train the model). Also, we are interest to separate the top-down and bottom-up features, which I talk about it later. As you can see in figure 6, the model based on Horixation features have the best performance. After that is the model with whole features, Subband, and Torralba.

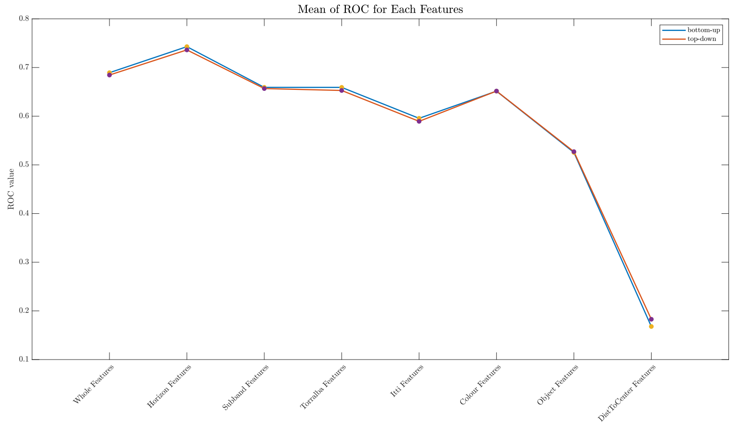

Figure 6: The average ROC over all images

Now, I add the features in each steps base on the article result. In other word, I run the model with only subband features, the add Torralba's features, color features, horizon detection, face and object detectors, and Itti channels. Figure 7 show the result.

[1] **judd2009learning**.

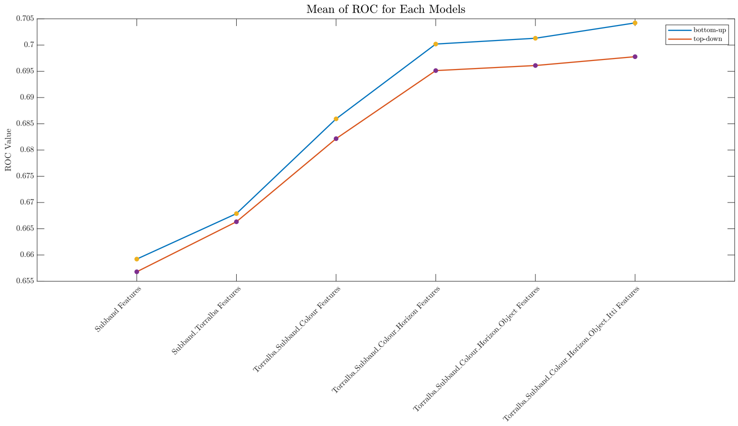

Figure 7: The average ROC over all images

Bottom-up features are the simple features that our visual system recognize from the image automatically, such as color, edges, etc. Top-down features guide attention on specific target volitionally.

To compare top-down and bottom-up effect, the fixation point is divided to two groups, first and second 1.5 s. In figure 8, you can compare the ROC distribution of each group for each models (histogram show pdf of the distribution). There is more dissimilarity between bottom-up and top-down time sample of low-level features (e.g. color, intensity, orientation and ...) than high-level (such as object or face features). Figure 9 and 10, compare each model result with the base model (used all features). For both botton-up and top-down samples Subband, Torralba's, color and horizon features have the most similarity respectively (as same as the article result).

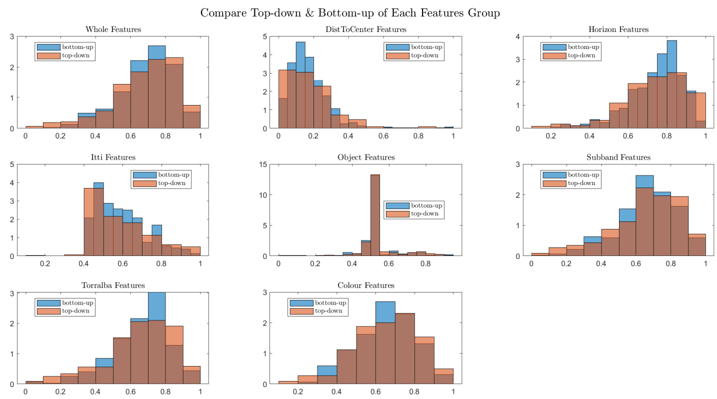

Figure 8: Different model ROC (top-down vs bottom-up)

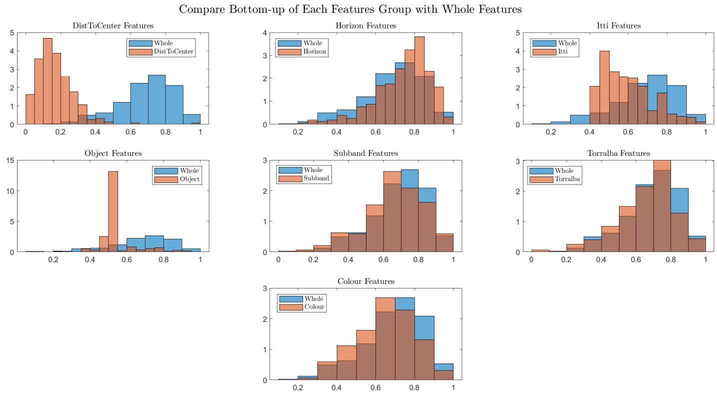

Figure 9: Different model ROC (bottom-up)

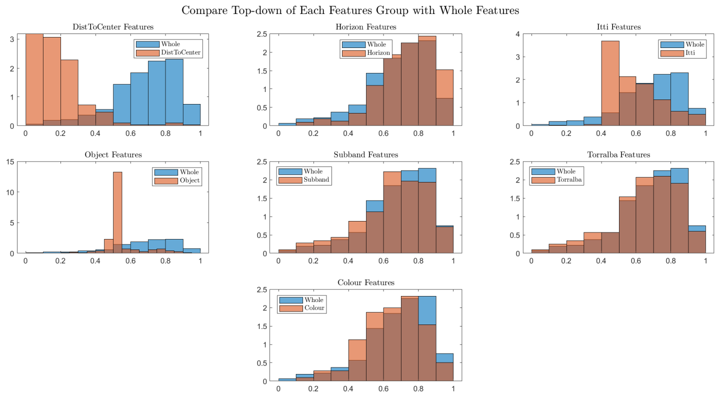

Figure 10: Different model ROC (top-down)

Section 4: Up to Down Images (Extra)

In this part, I flip all images vertically and same as in the previous part, add features to the salience map function at a time. I expect the model can't detect high-level features (such as person detection), unlike low-level features (intensity, orientation, ..., that are the same in both images).

In the figure 11, you can compare the saliency of the original (found in the previous part) and the inverted image for each group of features. For the subband, we have some differences but ultimately, both salience maps are similar. Maybe because most of the features are low-level. ('findHorizonFeatures' couldn't run for some inverted images!)

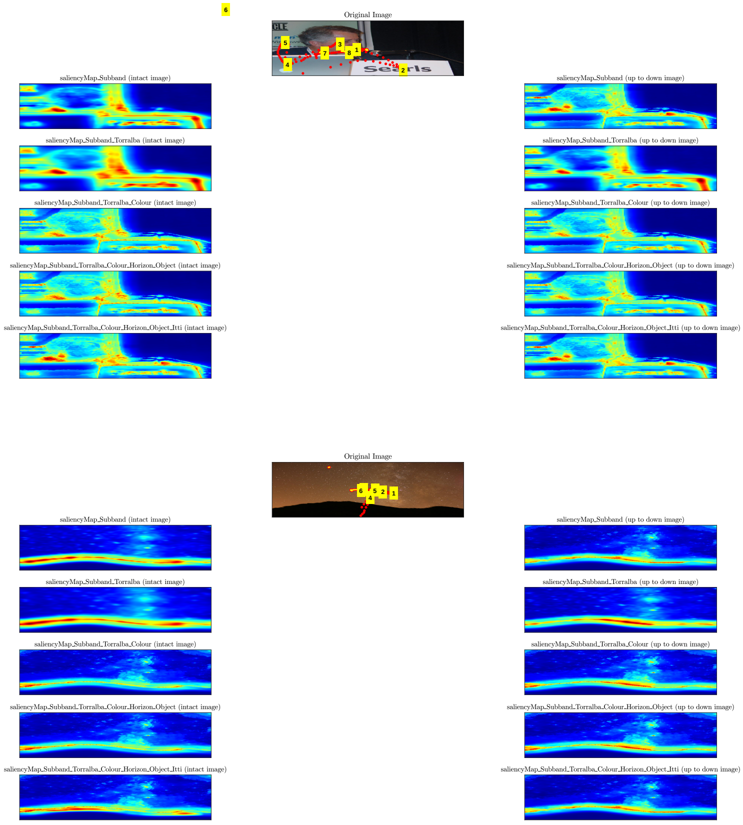

Figure 11: Salience map of intact and inverted image

In the figure 12, the model performance is measured with this flipped salience map, as same as figure 7. Only color and Itti features to improve model performance (color for both top-down and bottom-up and Itti only bottom-up).

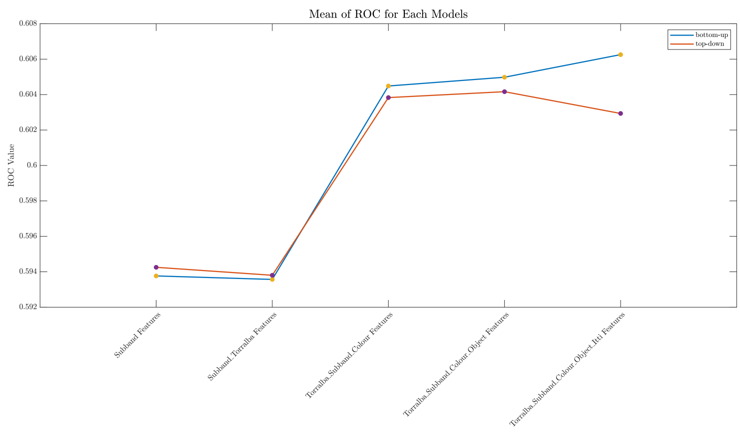

Figure 12: The average ROC over all images (top-down)