Modelling Attention for object detection : a computational model

Mentor – Prof. Amitabha Mukherjee

Shubham Tulsiani
(Y9574)

Approach Used

- Saliency Map
- Context Based Cues
- Feature Based Cues
- Combined the above to get a Computational Model for visual attention

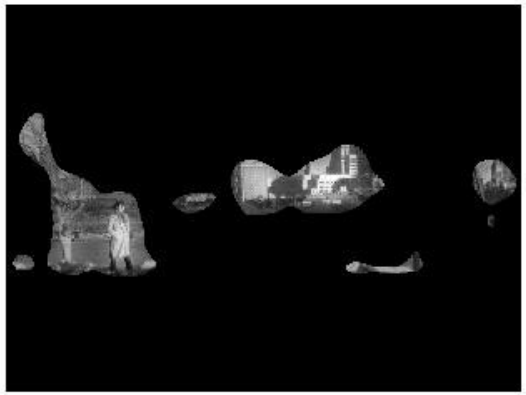
Saliency Maps

- Gives a measure of how much something stands out from its surroundings.
- Used the Itti-Koch model

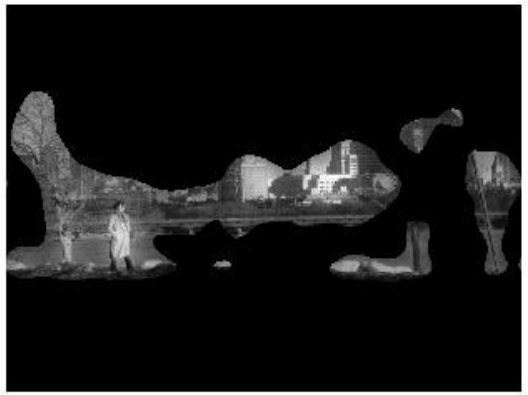
Saliency Maps

Original Stimulus: U0182__1LTE.jpg

SALIENCY Defined Region: Thresh 10%



SALIENCY Defined Region: Thresh 20%



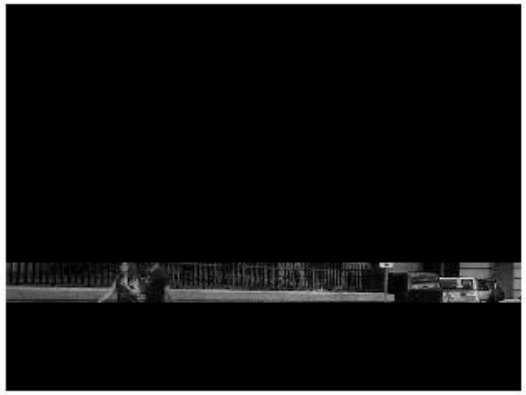
Contextual Maps

- Give an indication of where the object may be present taking into account the surroundings
- Used the LabelMe database to train a model for each object to be detected.
- Trained for Cars, Pedestrians and Trees on around 600 images containing these objects.

Contextual Map (Pedestrian Search)

Original Stimulus: U0146__1LBE.jpg

Context Defined Region: Thresh 10%

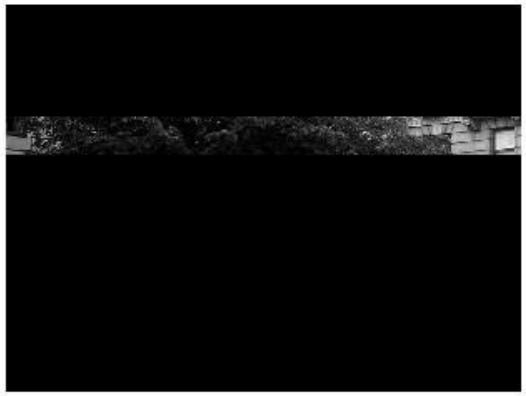


Context Defined Region: Thresh 20%

Contextual Map (Tree Search)

Original Stimulus: U0146__1LBE.jpg

Context Defined Region: Thresh 10%



Context Defined Region: Thresh 20%

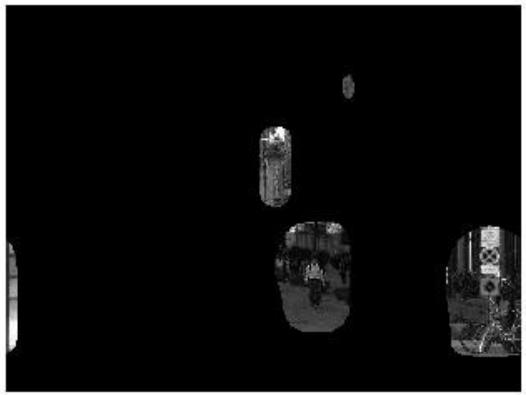
Feature Based Maps

- Indicate the resemblance to the object being searched for.
- Trained using LabelMe database on 100 images (1 positive and 10 random negative examples per image)
- Also used some images available on the internet with pre-extracted features for pedestrian search.

Feature Based Maps (Pedestrians)

Original Stimulus: U0161__1RBE.jpg

Target Features Region: Thresh 10%

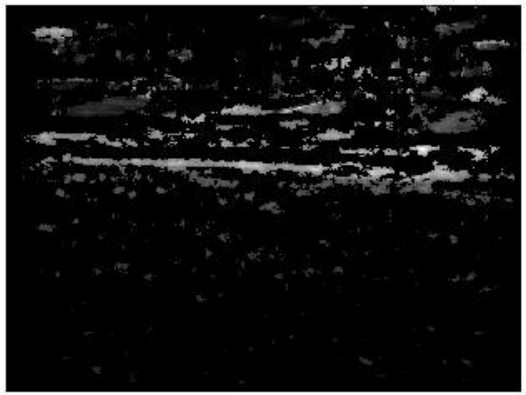


Target Features Region: Thresh 20%

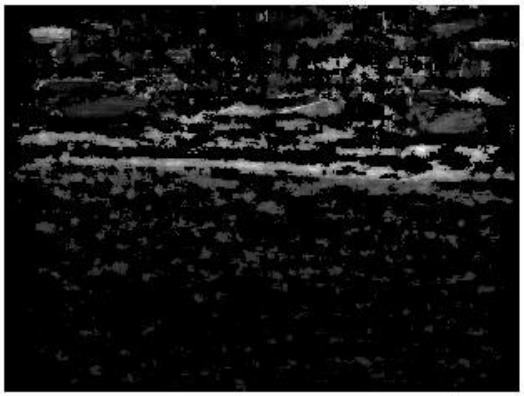
Feature Based Maps (Cars)

Original Stimulus: U0131__1RTE.jpg

Target Features Region: Thresh 10%



Target Features Region: Thresh 20%



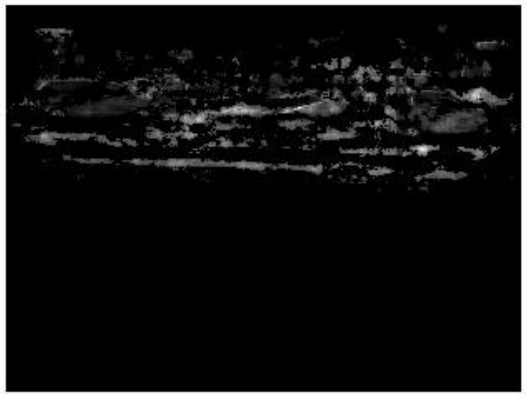
Combining the Maps for Object Search

- Combined these three maps in different ways by manipulating the weights assigned to each.
- Also tested the possible combinations by using weighted multiplication instead of addition while combining models.
- Output in form of selected 10% and 20% of the image which according to our model is most likely to have the target object.

Car Search

Original Stimulus: U0131__1RTE.jpg

Target Features Region: Thresh 10%

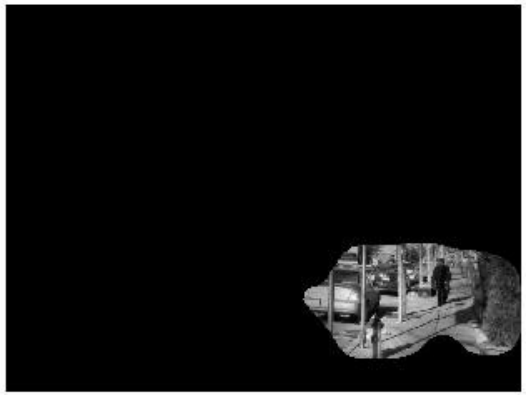


Target Features Region: Thresh 20%

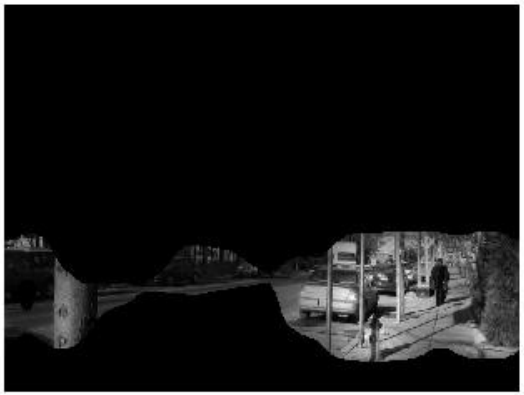
Pedestrian Search

Original Stimulus: U0293__1RBH.jpg

Target Features Region: Thresh 10%



Target Features Region: Thresh 20%



Criteria for Success

- Percentage of pedestrians in image who are present in the threshold region (top 10%, 20%) [Measure of success of object detection]

- Percentage of human eye fixations which lie in the threshold region.

  [Measure of success as model of human attention]

Number of Pedestrians

- The model impemented successfully included the pedestrian in the threshold region in all the cases
- We thus obtained an accuracy of nearly 100%
- Thus, this model can be effectively used as a preprocessing algorithm for object detection.

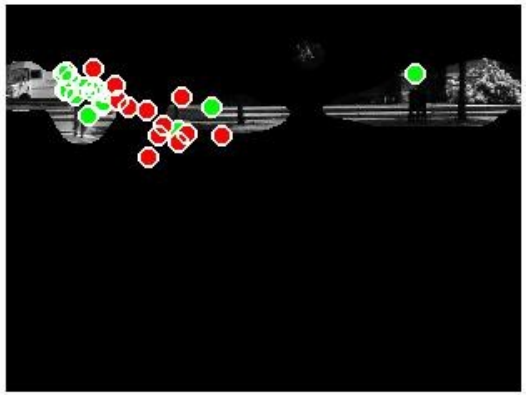
Success as a Model of Visual Attention

- Used a dataset by Antonio Torralba which gave the human eye fixations on image stimuli.
- Measured the success of the model by testing it on a set of 20 images.
- Total 411 fixations in 20 images

Success as a Model of Visual Attention

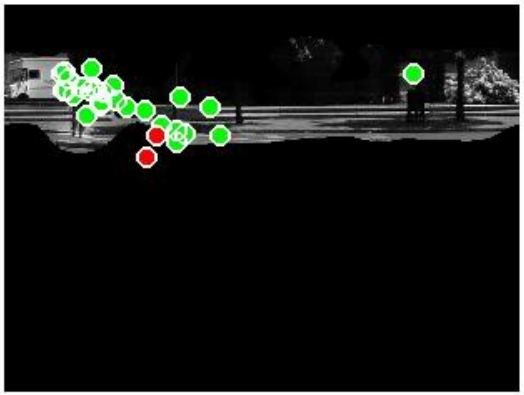
Original Stimulus: U0125__1LTE.jpg

Model Defined Region: Thresh 10%



17 out of 29 fixations IN region

Model Defined Region: Thresh 20%

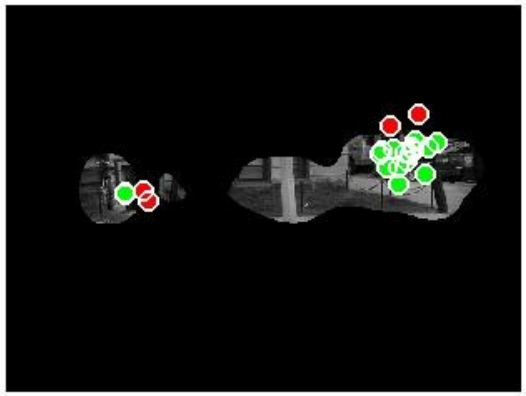


27 out of 29 fixations IN region

Success as a Model of Visual Attention

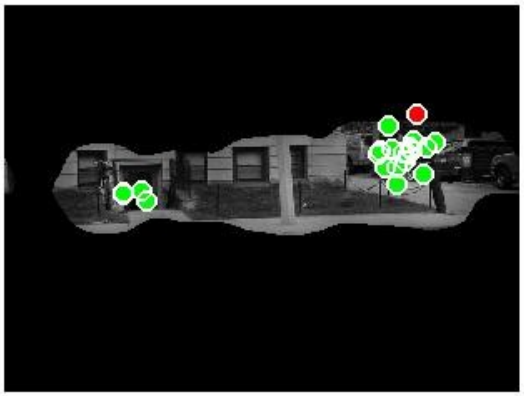
Original Stimulus: U0328__1RTH.jpg

Model Defined Region: Thresh 10%



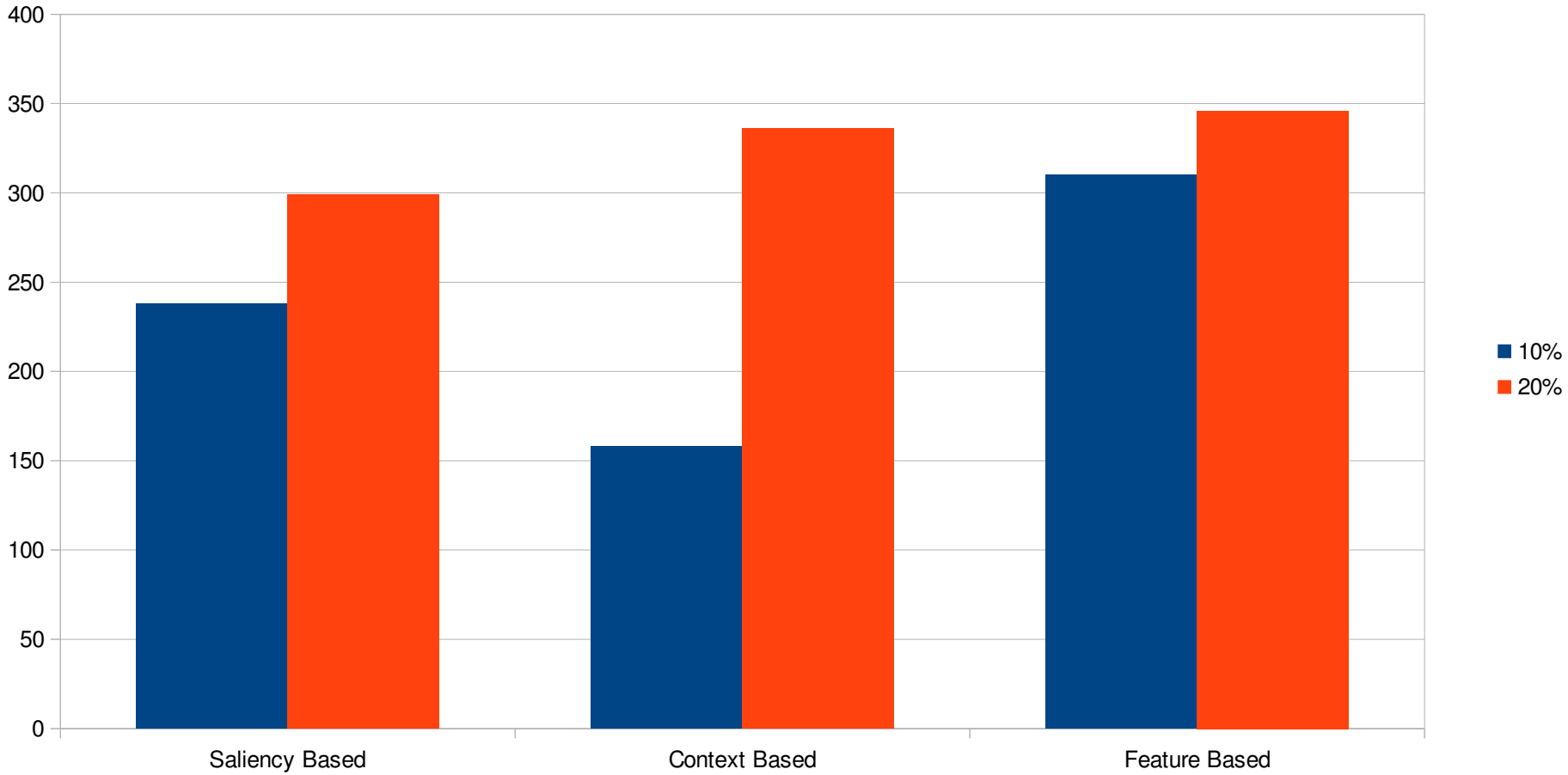
18 out of 22 fixations IN region

Model Defined Region: Thresh 20%

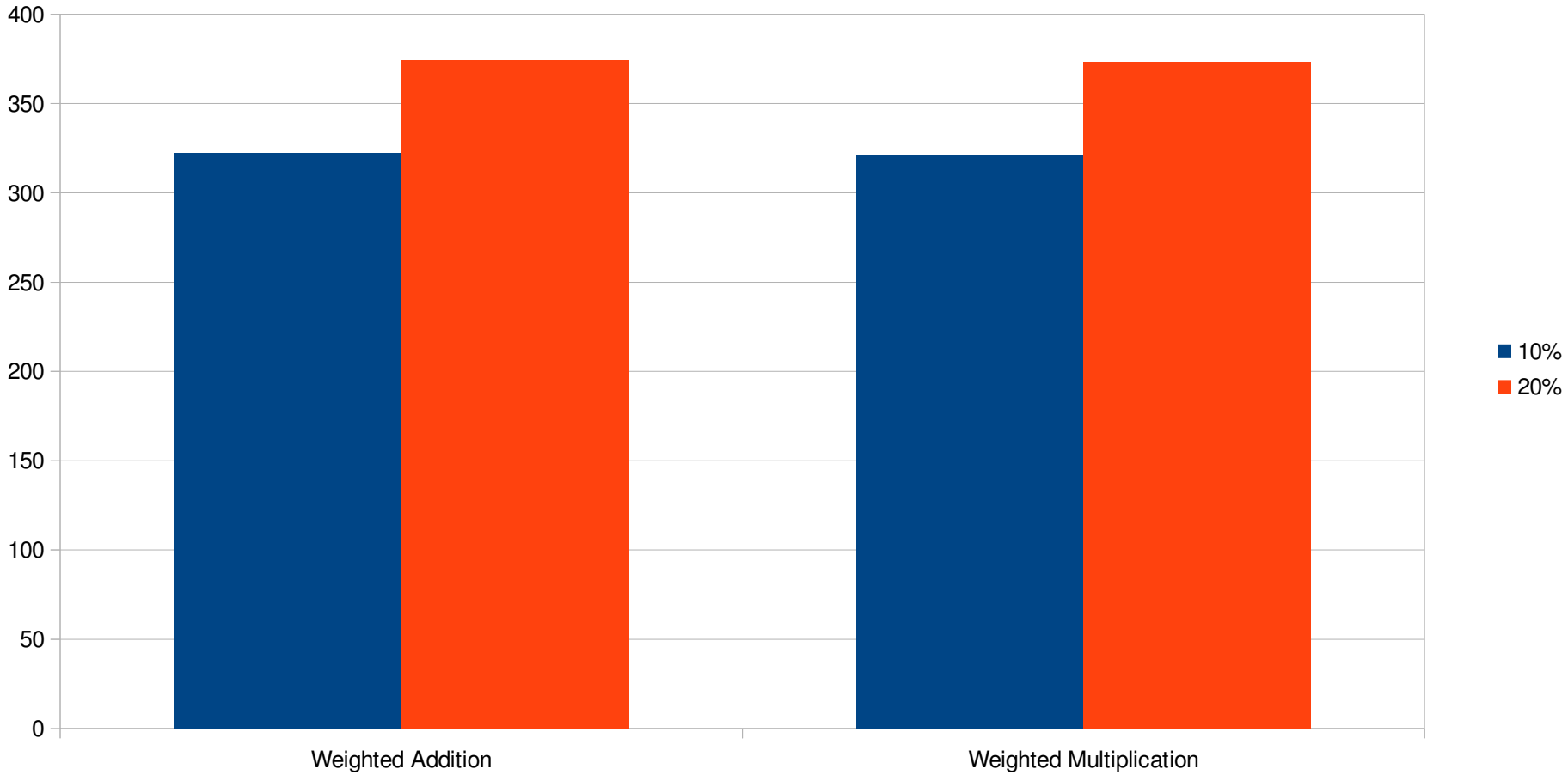


21 out of 22 fixations IN region

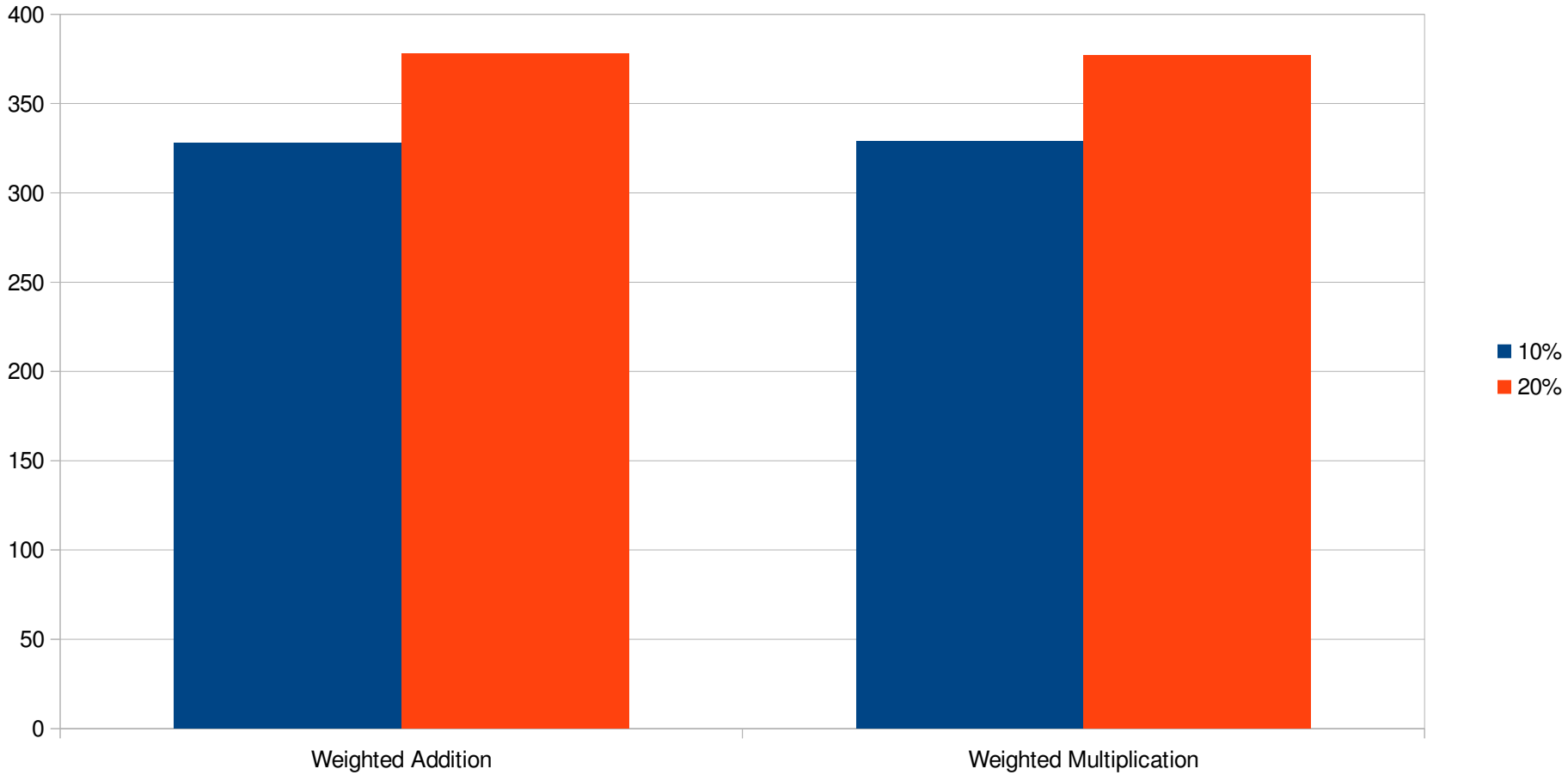
Single Source Models



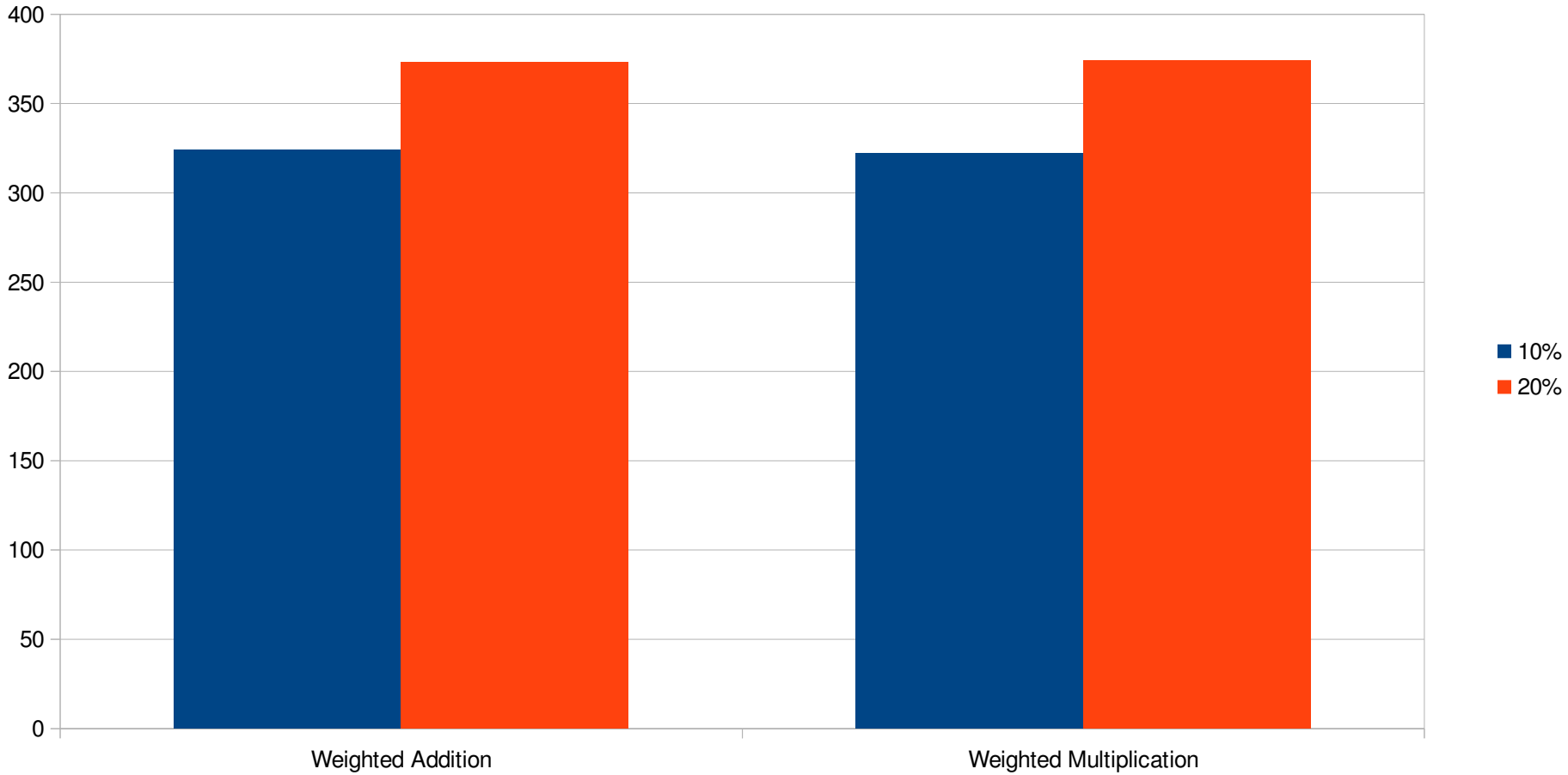
Combined Models : 80-15-5



Combined Models : 60-25-15



Combined Models : 40-35-25



Test on Negative Images – Target Absent

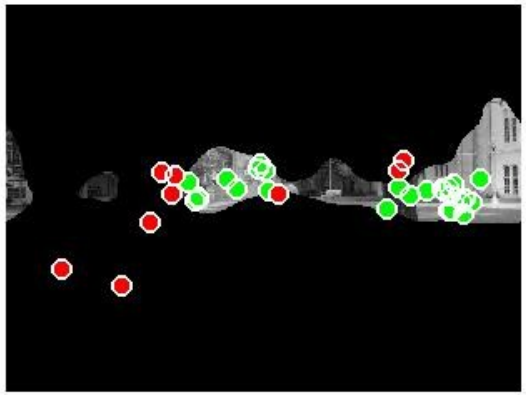
- Will answer question of whether hmans still follow the mentioned aspects to guide their attention

- 10 negative image stimuli to test the accuracy of the model – total 228 recorded fixations

Test on Negative Images – Target Absent

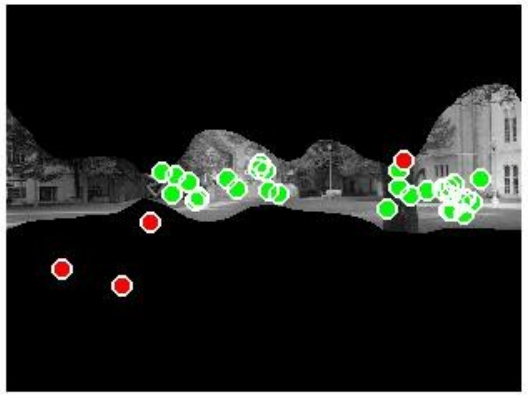
Original Stimulus: U0273__1RTH.jpg

Model Defined Region: Thresh 10%



26 out of 35 fixations IN region

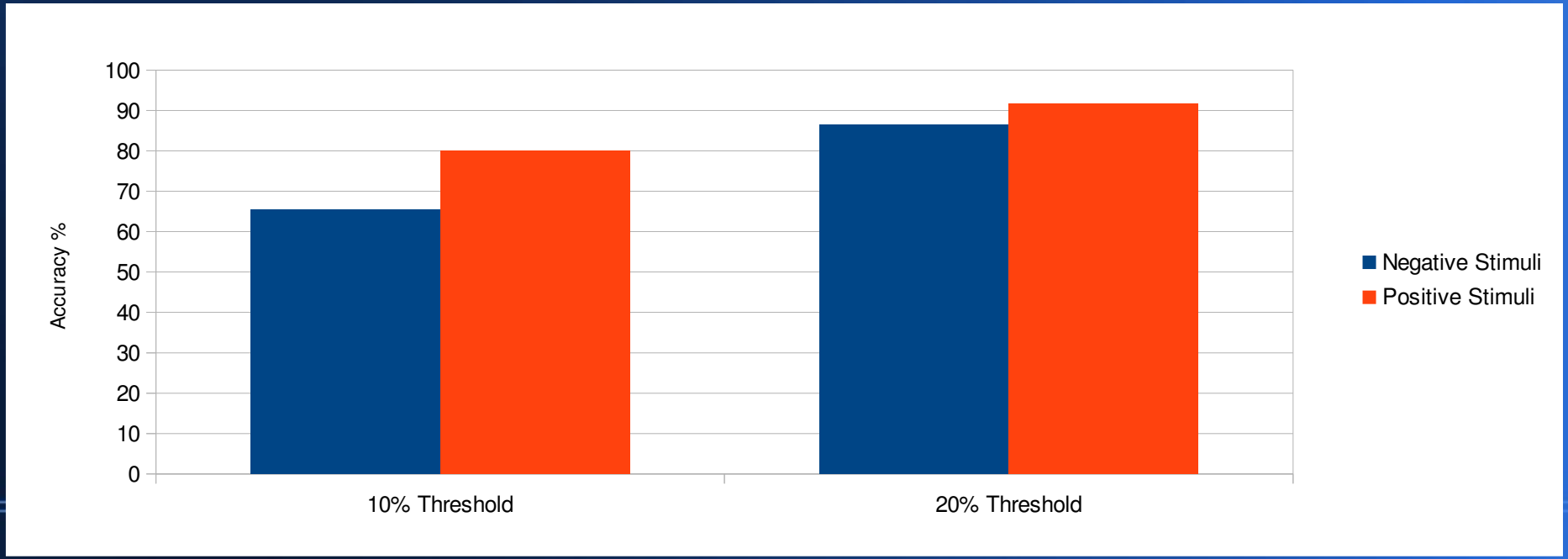
Model Defined Region: Thresh 20%



31 out of 35 fixations IN region

Test on Negative Images – Target Absent

- 149/228 fixations in 10% threshold region
- 197/228 fixations in 20% threshold region



Conclusions about the Combined Model

- It is a very good preprocessing step for object detection as the accuracy is nearly 100%
- For modelling human visual attention, it gives better results than any of the single source models.
- It gives an accuracy of more than 90% which is remarkable

Conclusions about the Combined Model

- As expected, for negative stimuli, it gives a slightly lower accuracy than for positive stimuli as other factors get involved.
- It is a reasonably good model of human visual attention as the regions of interest selected by it match with the ones where humans look with a high accuracy.