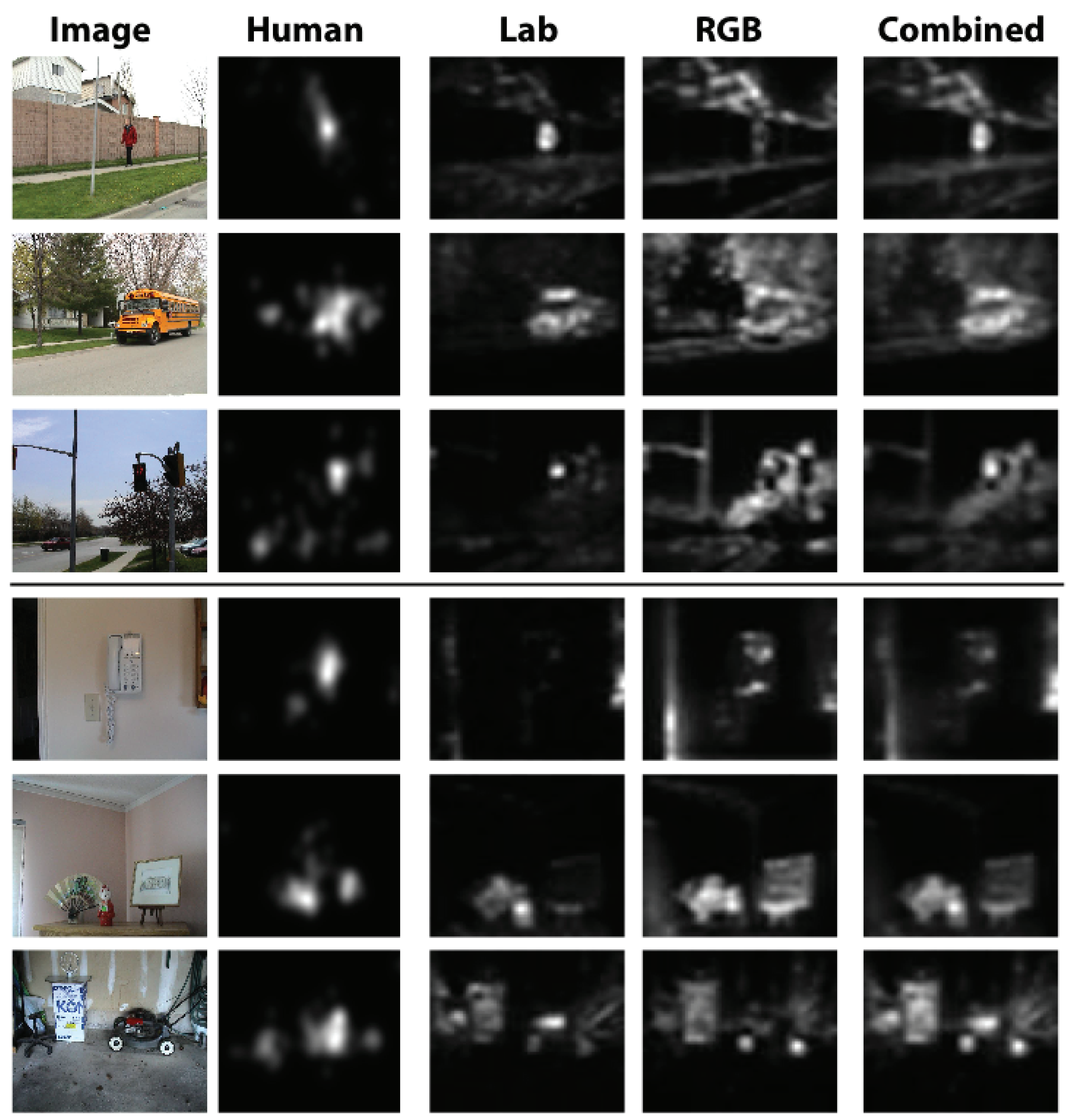
# Exploiting Local and Global Patch Rarities for Saliency Detection

Ali Borji and Laurent Itti
{borji, itti}@usc.edu

## Introduction

One color system does not work for all images.




Top (Bottom): Sample images where our model is able to detect the outliers in CIE Lab (RGB) color space.

## Contributions

+ Our **first contribution** is to propose a unified saliency model that benefits from the advantages of local and global approaches, which thus far have been treated independently. Note that the ideas of local and global context have been (separately) considered in the past by salient object detection/segmentation approaches, but those have not yet been tested with human fixation prediction, which is the goal of most models including ours.

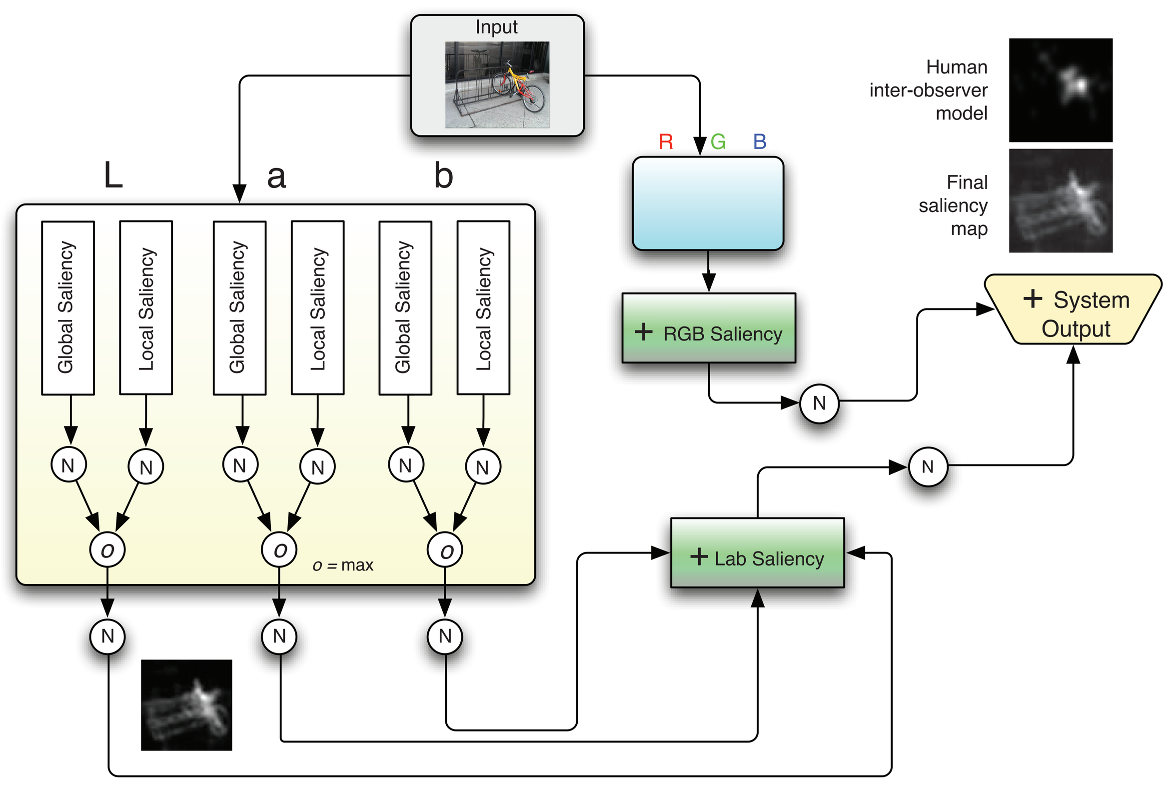
+ We argue that employing just one color system does not always lead to successful outlier detection. In figure above, we show that interesting objects in some images are more salient in Lab color space, while, for some others, saliency detection works better in RGB. Hence, a yet unexplored strategy, which is our **second contribution**, is combining saliency maps from both color spaces.

## Acknowledgements

*Supported by the National Science Foundation (grant number BCS- 0827764), and the Army Research Office (W911NF-08-1-0360 and W911NF-11-1-0046), and U.S. Army (W81XWH-10-2-0076).*

## Proposed Saliency Model

### Block Diagram




First, the input image is transformed into Lab and RGB formats. Then, in each channel of a color space (i.e., R, G, ...), a **global saliency map** based on rarity of an image patch in the entire scene, and a **local saliency map**, the dissimilarity between a patch and its surrounding window, are computed. These maps are then normalized and combined. Outputs of color channels (i.e., L, a, or b, similarly for RGB) are normalized and combined once more to form the output of a color system. The final map is the summation of the normalized maps in two color spaces.

### Image Representation

A dictionary of 200 basis functions learned from a large repository of natural images for the L channel of the Lab color space. Image size and patch size (w) were 512 × 512 and 8 × 8, respectively.

$$\alpha^*(x, D) = \arg\min_{\alpha \in R^n} \frac{1}{2}\|x - D\alpha\|_2^2 + \lambda_1\|\alpha\|_1$$

### Measuring Visual Saliency

**Step 1: Local Saliency**

$$S_l^c(p_i) = \frac{1}{L}\sum^{L} W_{ij}^{-1} D_{ij}^c$$

**Step 2: Global Saliency**

$$S_g^c(p_i) = P(p_i)^{-1} = \left(\Pi_{j=1}^{n} P(\alpha_{ij})\right)^{-1}$$

$$\log\left(S_g^c(p_i)\right) = -\log(P(p_i)) = -\sum_{j=1}^{n}\log(P(\alpha_{ij}))$$

$$S_g^c(p_i) \propto -\sum_{j=1}^{n}\log(P(\alpha_{ij}))$$

**Step 3: Combined Saliency**

$$S_{lg}^{Lab}(p_i) = \sum_{c \in L,a,b} N(S_{lg}^c(p_i))$$

$$S_{lg}^c(p_i) = N(S_l^c(p_i)) \circ N(S_g^c(p_i))$$

$$S_{lg}(p_i) = N(S_{lg}^{Lab}(p_i)) + N(S_{lg}^{RGB}(p_i))$$

**Extension to the scale space:**

$$S(x) = \frac{1}{M}\sum_{i=1}^{M} N(S_{lg}^i(x))$$

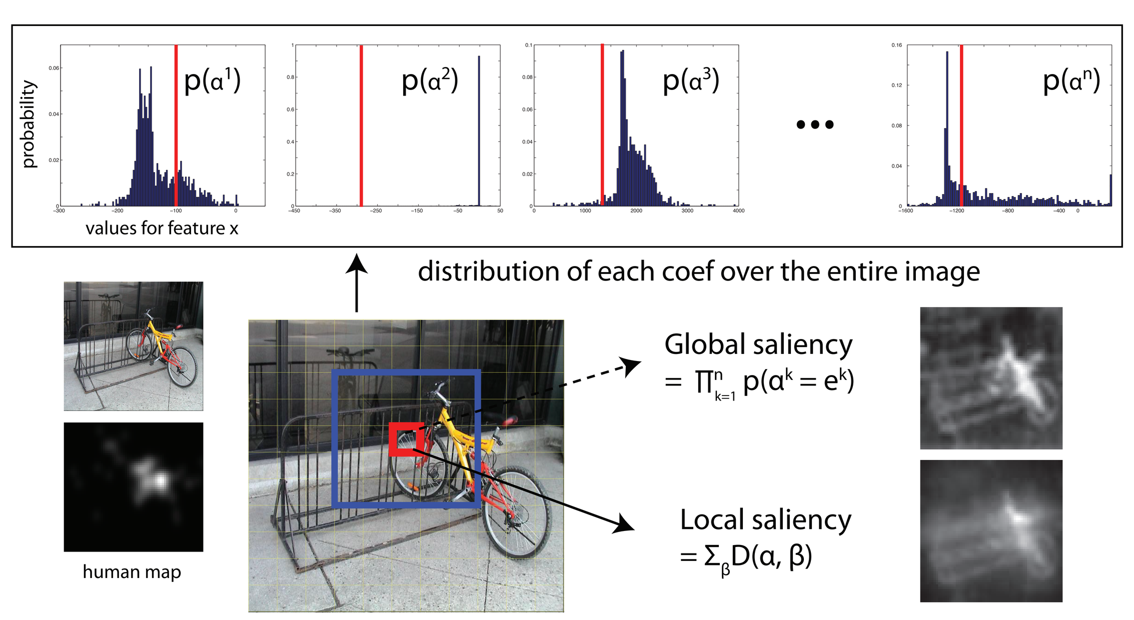
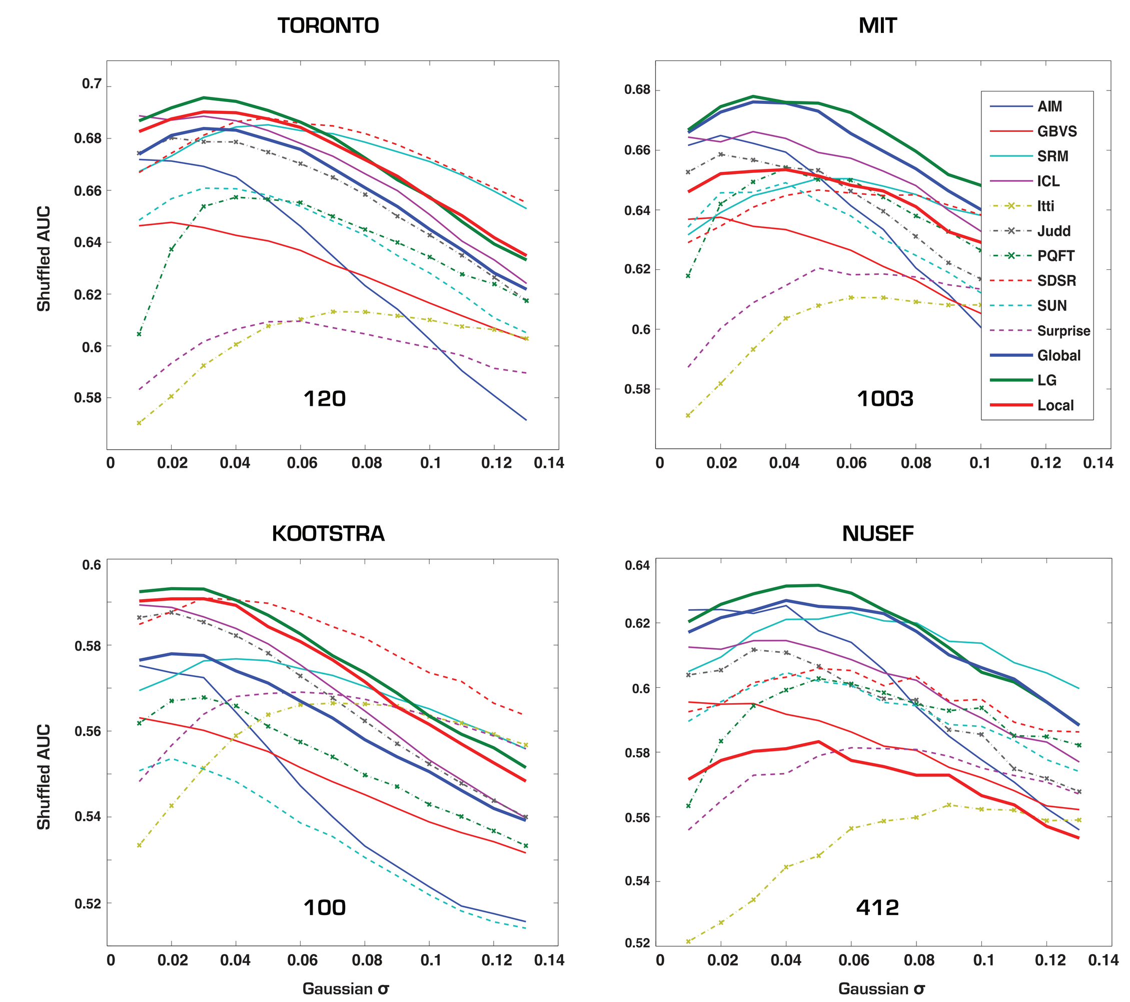



Illustration of global and local saliency for an image patch. Global saliency measures the rarity of a patch in the entire scene while local rarity measures the difference between a patch and its surrounding context.

## Performance Evaluation

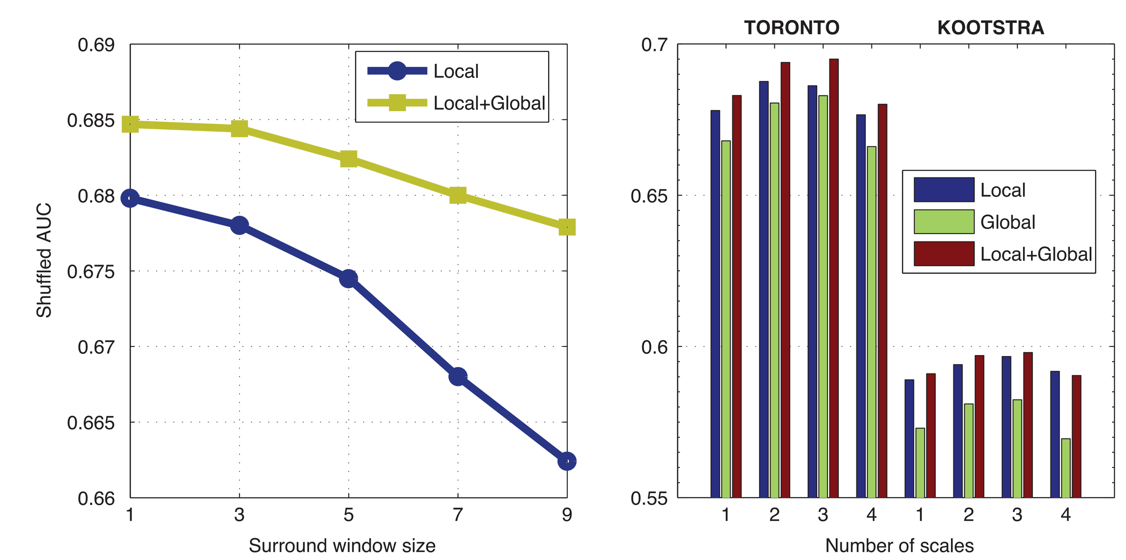



**Model comparison.** Fixation prediction accuracy of our saliency operations (Local, Global, LG (Local + Global)) along with 10 state-of-the-art models over 4 benchmark datasets. X-axis indicates the σ of the Gaussian kernel (in image width) by which maps are smoothed. Only 412 images of the NUSEF dataset are used here.

We used the Shuffled AUC score for discounting center-bias in model comparison.

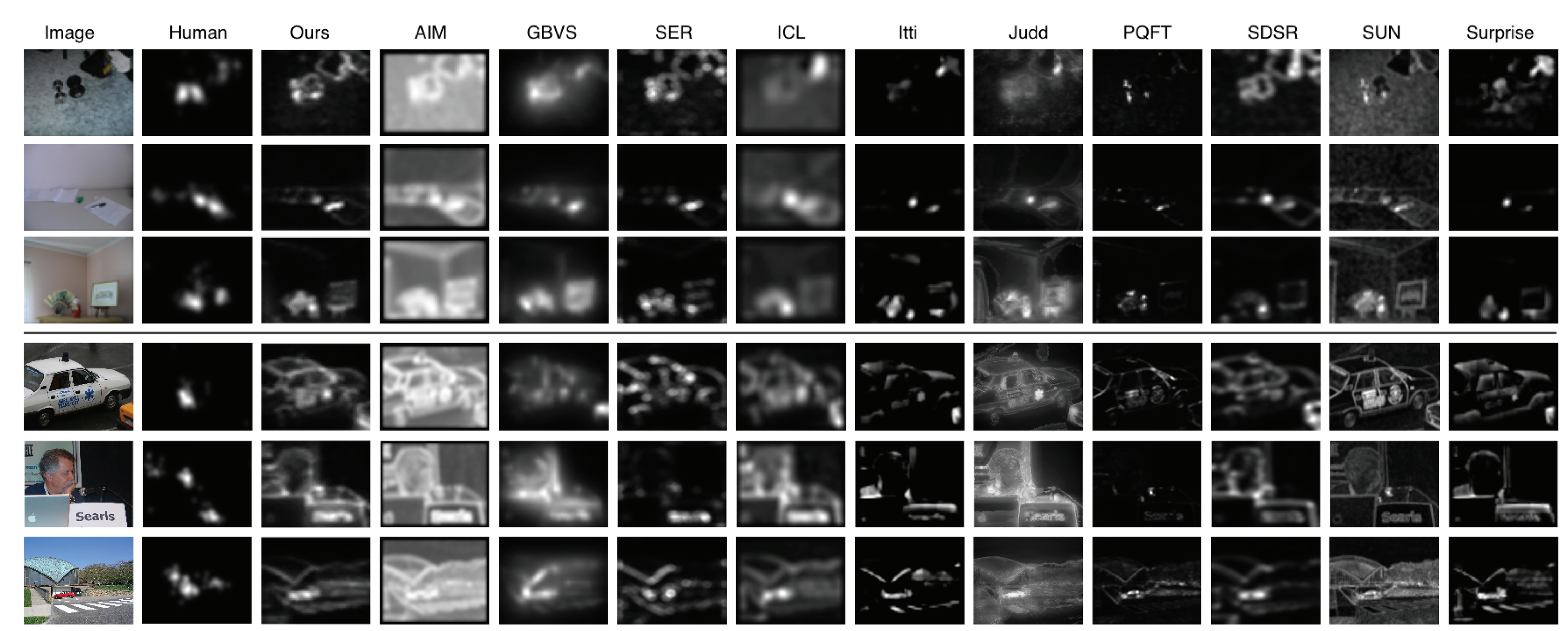
| Dataset | RGB | | | Lab | | | RGB + Lab | | |
|---|---|---|---|---|---|---|---|---|---|
| | $S_l$ | $S_g$ | $S_{lg}$ | $S_l$ | $S_g$ | $S_{lg}$ | $S_l$ | $S_g$ | $S_{lg}$ |
| TORONTO | 0.646 | 0.647 | 0.653 | 0.670 | 0.660 | 0.660 | 0.678 | 0.668 | 0.683 |
| MIT | 0.627 | 0.639 | 0.640 | 0.646 | 0.644 | 0.651 | 0.658 | 0.663 | 0.667 |
| KOOTSTRA | 0.574 | 0.572 | 0.578 | 0.572 | 0.555 | 0.570 | 0.589 | 0.573 | 0.591 |
| NUSEF | 0.599 | 0.610 | 0.610 | 0.556 | 0.596 | 0.592 | 0.569 | 0.614 | 0.616 |

**RGB vs. Lab for saliency detection.** Sl: Local; Sg: Global; Slg: Local + Global. Parameter settings: scales (M) =1 (256×256); Window size = 1. Results are over original saliency maps without smoothing.




**Parameter analysis.** Left: Effect of the surround window size on accuracy over TORONTO dataset using 256×256 images (M = 1). Right: Influence of scale on results over TORONTO and KOOTSTRA datasets (window size =1). First three bars are 256,128,64 and fourth one represents four scales 512,256,128,64.

**Visual comparison of our combined saliency model and 10 state-of-the-art models over samples from TORONTO (top) and MIT (bottom) datasets.**




**We conclude that integration of local and global saliency operators works better than just using either one, which encourages more research in this direction. Similarly, combining both color systems strongly benefits saliency detection and eye fixation prediction.**