Group 31 VISIONARIES

Ujjwal Tiwari(2018900100) Harshit Kumar Sankhla(2018900050), Udit Singh Parihar(2018701024)

Computer Vision

Saliency Filters:

Contrast Based Saliency Estimation

Figure 1: From left to right: input images, image abstraction into perceptually homogeneous elements, results of our saliency computation, ground truth labeling.

Project Overview

Saliency estimation is a challenging computer vision task of identifying image elements that are likely to catch the attention of a human observer. These computational methods are challenging due to the complex cross disciplinary nature of the problem. Methods that can find salient image regions find wide application base in image processing tasks such as image resizing, object detection and image segmentation.

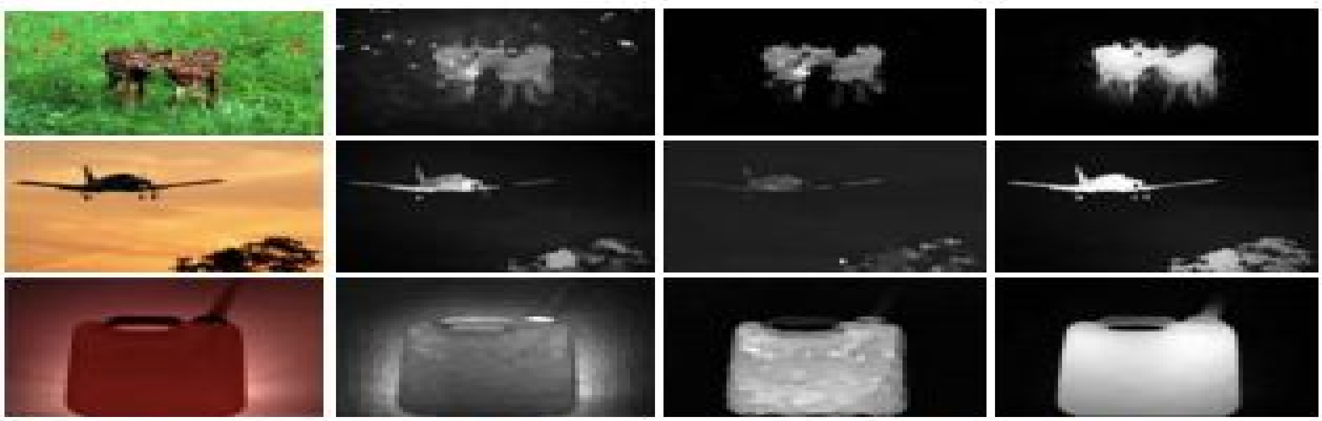
High- level methods take into account insights from a variety of interdisciplinary fields such as neuroscience, computer vision, biology etc. However, low level feature based computational methods based on contrast are quite successful.

The definition of contrast in related works are often related to color variation of individual pixels, edges, gradients, , spatial frequencies, structure and distribution of image patches, histograms, multi-scale descriptors, or combinations thereof. This ambiguous nature of contrast measure fails to attribute improvements in results to design considerations of specific algorithms.

As a part of this project we explore a intuitive yet effective algorithm proposed in CVPR'12 for contrast based saliency estimation. The proposed work improves upon previous work by formulating the problem of saliency estimation in four different steps:

1. **Finding Super Pixels:** Decomposition of Image into perceptually homogeneous clusters using a variant of SLIC superpixels. In this varint the superpixels are computed using K-Means algorithms in CIElab color space rather than in RGBXY space, this accounts for better compactness other than just being local and edge-aware superpixels. Geodesic image distance guarantees connectivity, while retaining the locality, compactness and edge awareness of SLIC superpixels.
2. **Uniqueness (first measure of contrast):** The notion of pixel wise saliency is derived from two different measures of contrast. The first of the these measures is element uniqueness. The uniqueness of every element found in the first step is calculated. We therefore evaluate how different each respective element is from all other elements constituting an image, essentially measuring the "rarity" of each element.
3. **Element Distribution (second measure of contrast):** Element uniqueness scores are conceptually weak directives to pixel saliency estimation as background superpixels can be just as unique as the ones that constitute the actual salient object. Therefore, the intuition behind this second measure of contrast is that salient features in an image will have low spatial variance, which renders unique elements more salient when they are grouped in a particular image region rather than evenly distributed over the whole image.
4. **Saliency Estimation:** In the final step, the two above contrast measures that are defined on a per-element level are clubbed together to assign the actual saliency values to the input image to get a pixel-accurate saliency map

Central to the contrast and saliency computation is their second main contribution; where they show that all involved operators can be formulated within a single high-dimensional Gaussian filtering framework. This formulation leads of highly efficient implementation with linear complexity.



(a) Source image. (b) Uniqueness. (c) Distribution. (d) Saliency.

Figure 3: Uniqueness, spatial distribution, and the combined saliency map. The uniqueness prefers rare colors, whereas the distribution favors compact objects. Combined together those measures provide better perfomance.

Implementation:

We have implemented this saliency estimation algorithm and produced comparable qualitative results. This is a python implementation which will be open sourced upon final submission and evaluation of this project.

For the implementation of this algorithm we have used the following python packages:

- OpenCV2.0
- Scipy
- Scikit-image
- Numpy

Final requirements will be made available in a requirements.txt file.

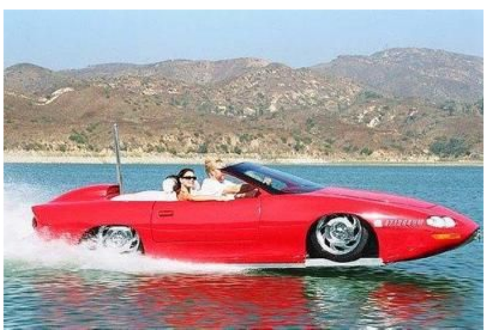
1. **ABSTRACTION :**

   In this step we abstract away the texture in the image by using sklearn based SLIC module implementation of K-means clustering algorithm. Our results are close to that provided by the original paper.

   K-means is a unsupervised learning algorithm that clusters together similar data entries based on a distance measure on the feature space. Number of image segments of this experiment is taken to be 300.

   Refer below to the results section for generated Superpixels on test image.

Results:



2. **ELEMENT UNIQUENESS :**

Element uniqueness is calculated by the mathematical formulation provided in the paper:

$$U_i = \sum_{j=1}^{N} \|\mathbf{c}_i - \mathbf{c}_j\|^2 \cdot \underbrace{w(\mathbf{p}_i, \mathbf{p}_j)}_{w_{ij}^{(p)}} .$$

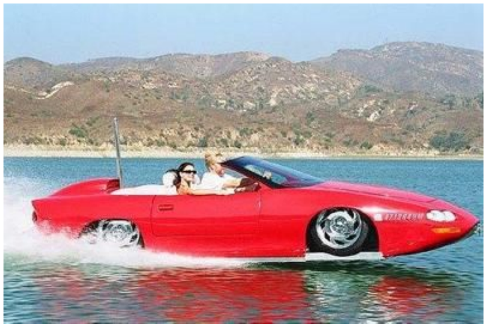
- The weights provide control over the global and local contrast estimation with radius of influence of the uniqueness operator.
- A local function yields a local contrast term, which tends to overemphasize object boundaries in the saliency estimation. whereas w (p) ij ≈ 1 yields a global uniqueness operator, which cannot represent sensitivity to local contrast variation.

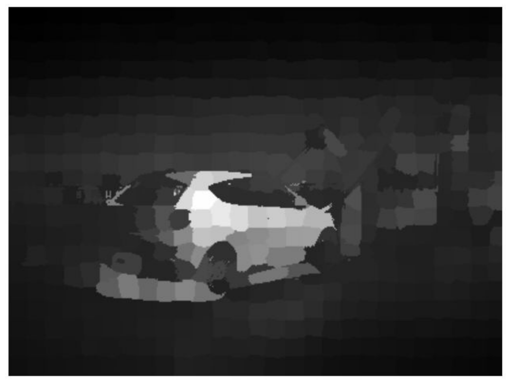
$$\begin{aligned} U_i &= \sum_{j=1}^{N} \|\mathbf{c}_i - \mathbf{c}_j\|^2 w_{ij}^{(p)} \\ &= \mathbf{c}_i^2 \underbrace{\sum_{j=1}^{N} w_{ij}^{(p)}}_{1} - 2\mathbf{c}_i \underbrace{\sum_{j=1}^{N} \mathbf{c}_j w_{ij}^{(p)}}_{\text{blur } \mathbf{c}_j} + \underbrace{\sum_{j=1}^{N} \mathbf{c}_j^2 w_{ij}^{(p)}}_{\text{blur } \mathbf{c}_j^2} .(\end{aligned}$$

- Using gaussian Kernel reduces time complexity from O(N^2) to O(N). It is advantageous over other methods that subsample the image to a lower resolution which leads to blurring of images.

Refer to the below results section for the Image filters and Uniqueness estimation on test image.

Results post abstraction and element uniqueness :





3. ELEMENT DISTRIBUTION :

Element Distribution measures the spatial distribution of colour in for every segment abstracted in the SLIC superpixel abstraction step. This measures it's occurrence elsewhere in the image.

- Low variance indicates a spatially compact object which is considered more salient as compared to spatially widely distributed object.
- Mathematical formulation for the computation of element distribution:

$$D_i = \sum_{j=1}^{N} \|\mathbf{p}_j - \mu_i\|^2 \underbrace{w(\mathbf{c}_i, \mathbf{c}_j)}_{w_{ij}^{(c)}},$$

- The sigma parameter of the gaussian controls the color sensitivity of the element distribution. Sigma = 20 is used in all the experiments.
- This parameter controls the local vs global trade off of spatial distribution based contrast measure.
- Choosing color similarity to be Gaussian of the form $w_{ij}^{(c)} = \frac{1}{Z_i} \exp(-\frac{1}{2\sigma_c^2}\|\mathbf{c}_i - \mathbf{c}_j\|^2)$, we can effectively decompose and evaluate this in linear time.

$$
\begin{aligned}
D_i &= \sum_{j=1}^{N} \|\mathbf{p}_j - \mu_i\|^2 w_{ij}^{(c)} \\
&= \sum_{j=1}^{N} \mathbf{p}_j^2 w_{ij}^{(c)} - 2\mu_i \underbrace{\sum_{j=1}^{N} \mathbf{p}_j w_{ij}^{(c)}}_{\mu_i} + \mu_i^2 \underbrace{\sum_{j=1}^{N} w_{ij}^{(c)}}_{1} \\
&= \underbrace{\sum_{j=1}^{N} \mathbf{p}_j^2 w_{ij}^{(c)}}_{\text{blur } \mathbf{p}_j^2} - \underbrace{\mu_i}_{\text{blur } \mathbf{p}_j}{}^2.
\end{aligned}
$$

Now that we have the element distribution and element uniqueness properties calculated on top of the superpixels. We need an accurate pixel-wise labelling method that combines these sparse labels to estimate pixel level saliency estimate.

4. Saliency Assignment :

- This is the final step can be seen as two distinct steps, first is to combine the two contrast based metrics into a single saliency estimate for each element:

$$S_i = U_i \cdot \exp(-k \cdot D_i),$$

- In practice, spatial distribution is of more significance and hence it's exponentiated before factoring into the saliency estimate.

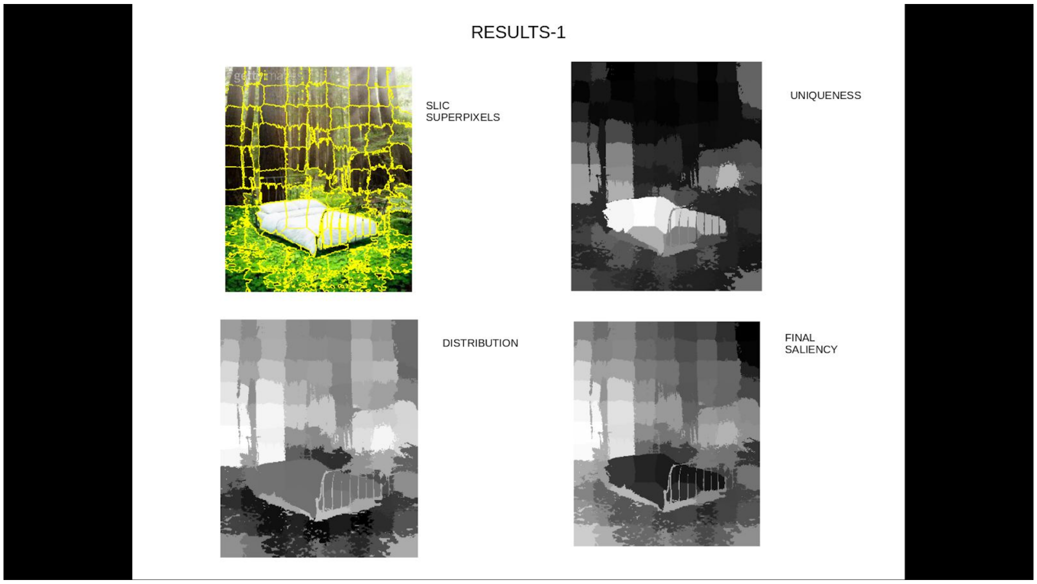
Note: Both spatial and element distribution are normalized in [0,1] range before this computation.

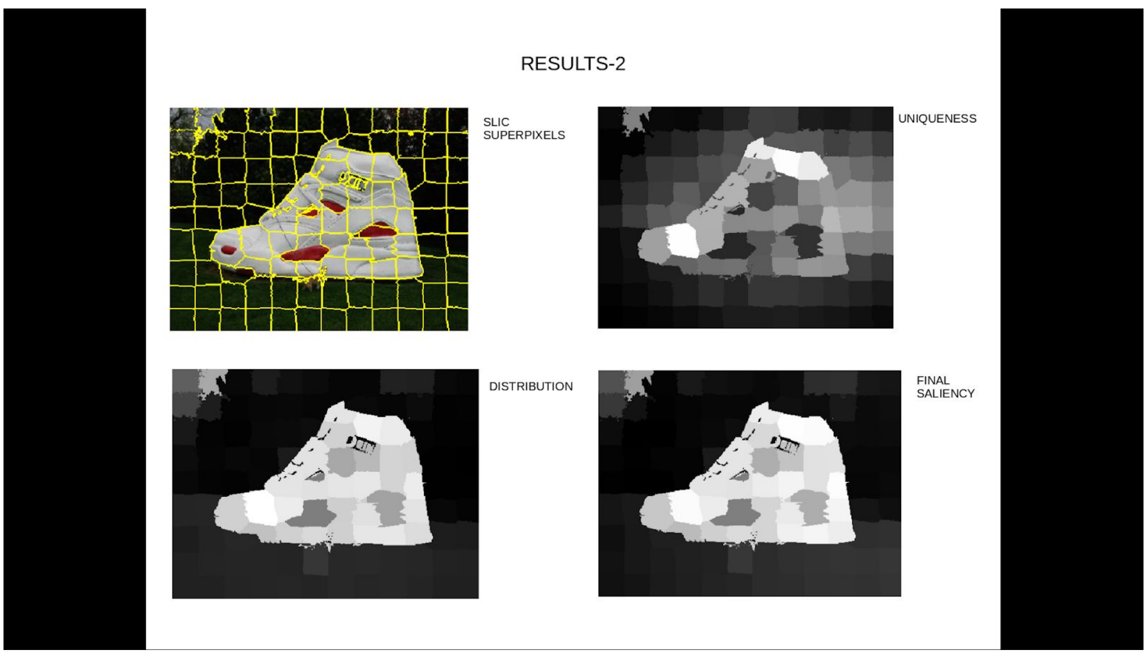
- As the final step we need to get finer estimate of saliency at pixel level from coarse saliency estimated.
- Saliency of a pixel is weighted linear combination of the saliency of its surrounding image elements.

$$\tilde{S}_i = \sum_{j=1}^{N} w_{ij} S_j.$$

- The choosing a Gaussian weights as $w_{ij} = \frac{1}{Z_i}\exp(-\frac{1}{2}(\alpha\|\mathbf{c}_i - \mathbf{c}_j\|^2 + \beta\|\mathbf{p}_i - \mathbf{p}_j\|^2)$, this ensures that the upsampling process is both local and color sensitive.

Results:


RESULTS-1
SLIC
SUPERPIXELS
UNIQUENESS
DISTRIBUTION
FINAL
SALIENCY


RESULTS-2
SLIC
SUPERPIXELS
UNIQUENESS
DISTRIBUTION
FINAL
SALIENCY

Limitations:

- Sensitive to lightning variations as it's color contrast based saliency estimate.
- When background and foreground is of similar color this methods produces inaccurate results.
- Thresholding based evaluations can result in noisy segmentations, to significantly reduce this effect is to perform a single min-cut as a post process using our saliency maps as a prior for the min-cut data term and color difference between neighbouring pixels for the smoothness term.

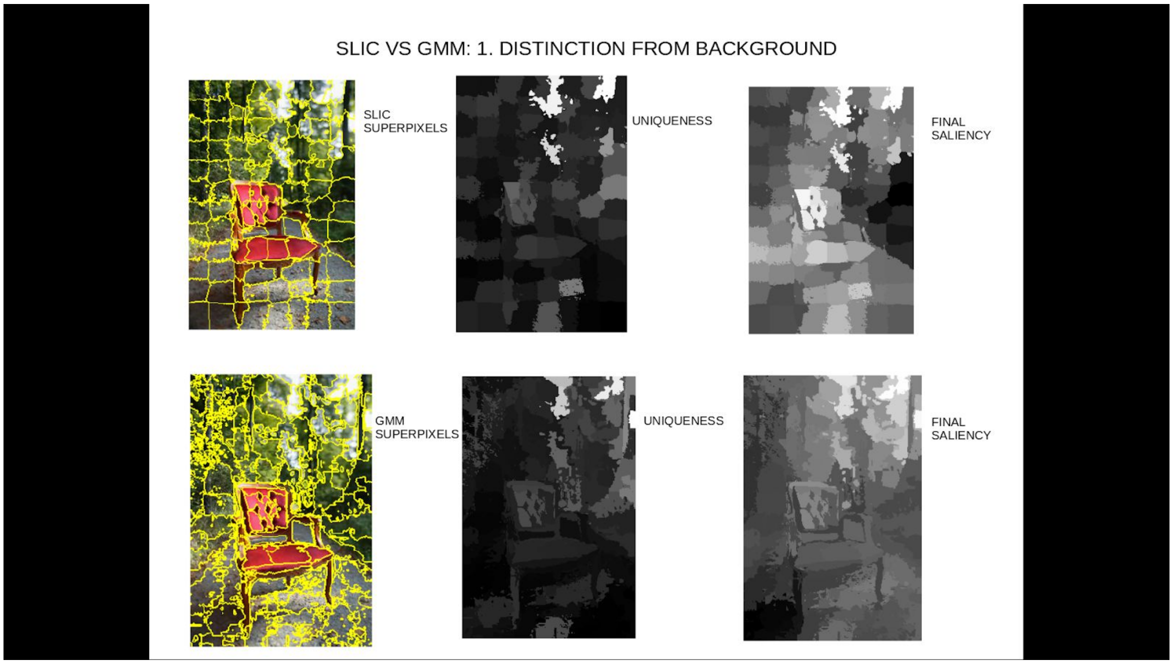
We further extend this project to find better saliency estimate using GMM based Image abstraction rather than using K-means clustering to find super pixels. K-means is a hard labelling whereas GMM gives soft abstraction.

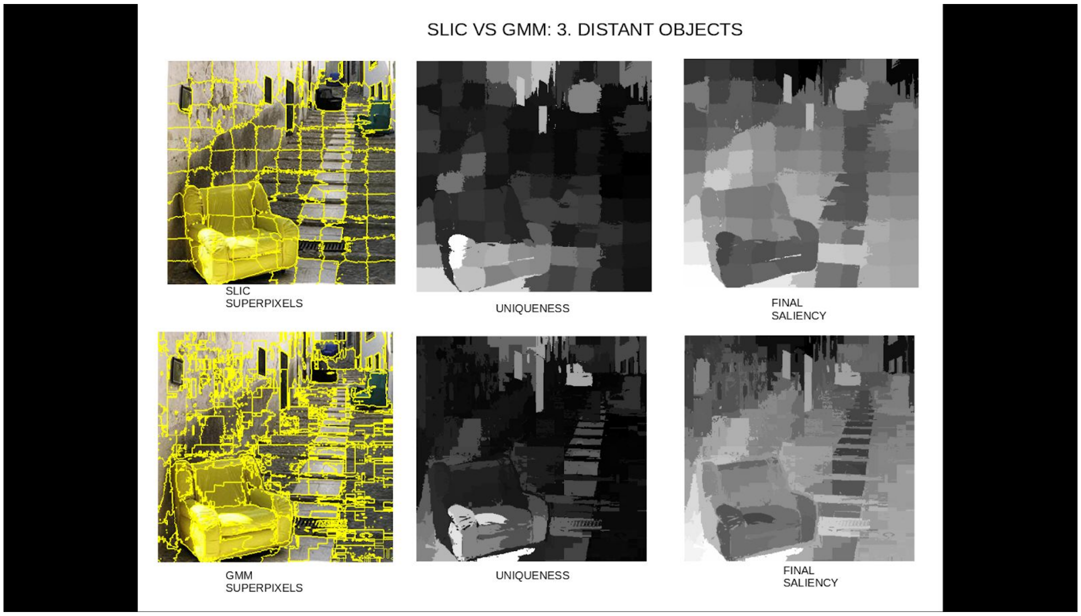
We make a comparative study between the two pipelines keeping other things like quantitative estimate of contrast keeping constant.

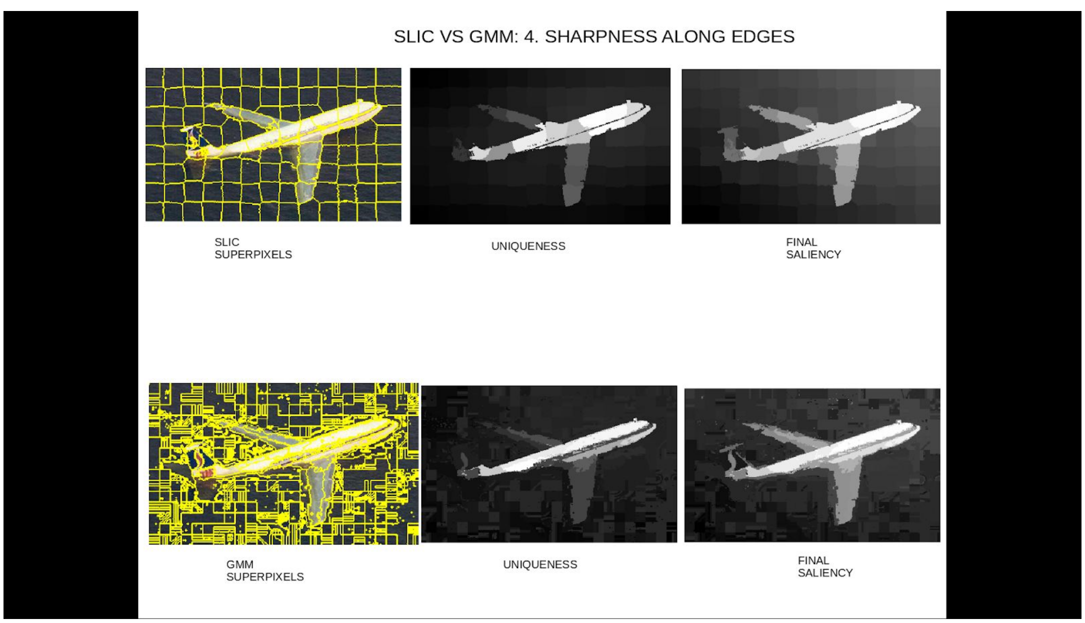
KEY IDEOLOGICAL DIFFERENCE:

This soft abstraction avoids the hard decision boundaries of super pixels, allowing abstraction components with very large spatial support. This allows the subsequent global saliency cues to uniformly highlight entire salient object regions. . These can be seen in our comparative results. Finally, we integrate the two global saliency cues, element uniqueness (EU) and Color Spatial Distribution (CSD), by automatically identifying which one is more likely to provide the correct identification of the salient region.

Comparative Results:




SLIC VS GMM: 3. DISTANT OBJECTS
SLIC
SUPERPIXELS
UNIQUENESS
FINAL
SALIENCY
GMM
SUPERPIXELS
UNIQUENESS
FINAL
SALIENCY


SLIC VS GMM: 4. SHARPNESS ALONG EDGES
SLIC
SUPERPIXELS
UNIQUENESS
FINAL
SALIENCY
GMM
SUPERPIXELS
UNIQUENESS
FINAL
SALIENCY

These are some ideas we had but could not extend our project due to time constraints.

Other key ideas for extension of this work:

1. Hierarchical clustering, Spectral clustering for image abstraction.
2. Saliency estimation post semantic segmentation proves to produce better results in literature.

REFERENCES :

THE FIRST TWO REFERENCES ARE DIRECTLY RELATED TO OUR IMPLEMENTATION.

1. Saliency Filters : contrast based saliency estimation. [LINK]
2. Efficient Salient Region Detection with Soft Image Abstraction.[LINK]
3. A model of saliency-based visual attention for rapid scene analysis. [LINK]
4. SLIC Superpixels. Technical report. [LINK]
5. Learning to detect a salient object. [LINK]
6. Improved saliency detection based on superpixel clustering and saliency propagation. [LINK]
7. Global contrast based salient region detection. [LINK]

Find the git repository to all project code.

[git_link]