Saliency Selector

Shoumik Bhattacharya MT21144
Sayan Mitra MT21142
Ramyanee Kashyap MT21139

1. Abstract

There are a number of approaches for finding saliency of images. Some of these methods use heuristics based on image matrix like distance between pixels with respect to various color spaces as well as spatial distances, while some of the methods are deep learning based methods that use a neural network to predict saliency of images. While heuristic based methods are often quite fast, they tend to fall short when the image has relatively homogenous color distribution. DL based methods on the other hand are computationally expensive, but deliver excellent results. Our aim is to suggest a suitably saliency method to process a given dataset depending on the time and quality of saliency computation by various methods. We aim to find a suitable tradeoff between time and quality of various saliency methods for a particular type of image data at hand.

2. Introduction

2.1. Problem Statement

Input: A small randomly sampled subset of an image dataset.
Output: The best saliency extraction method for the input set of images.
Pre-processing: Resize each input set image so that all images are of the same shape.
Output Evaluation: The output can be evaluated based on the quality of saliency map obtained.

2.2. Challenges

One of the major challenges in saliency extraction is to identify which saliency extraction method is best suited for the input image set since there are many saliency extraction methods available, each having their own advantages and disadvantages. Simple heuristic based methods work well on some images with distinct background and foreground separation but may fail in cases where the foreground and background are not well separated. Deep Learning based methods work on a large number of image types but are computationally expensive.
One way is to implement and test each method separately and observe the results and based on the quality of results obtained, choose the best method. But this is a very time consuming and cumbersome task, since there are a lot of methods available. Therefore, it requires expertise and thorough background knowledge of various saliency methods to select the correct saliency method for the type of image at hand.

2.3. Motivation

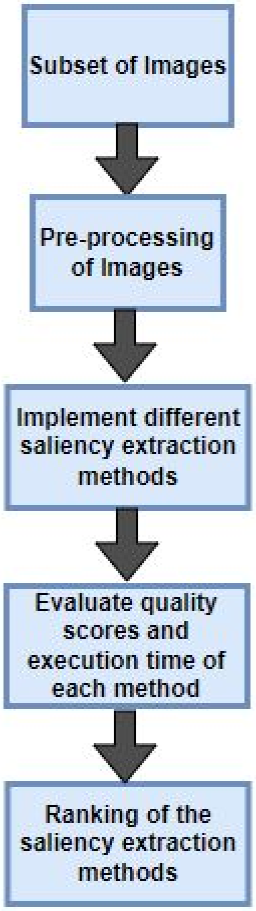
The objective of this project is to automate the process of selecting a suitable saliency detection method for a set of similar images,to find the best saliency method in an efficient manner and to get the best saliency map for the input image set and to take into consideration both computational efficiency and quality measure while selecting saliency method.
Since our proposed idea will internally find the best possible saliency extraction method for the input image and suggest the user with the best extraction method, it will eliminate the cumbersome task of testing each method separately by the user for determining the best method.

2.4. Contribution and Novelty

Current literature in saliency involves quality assessment of saliency methods based on their ability to find the region of the image which is most likely to grab the attention of the viewer. However little attention has been given to the need to select an efficient saliency model for a specific task at hand.
In this project we aim to find a robust method to select the **optimal saliency detection technique** given an image dataset while taking into consideration both quality measure and computational time. The goal is to arrive at a suitable tradeoff between saliency detection quality and the time required in the process to finally suggest the suitable method of saliency detection given an image dataset.

2.5. Summary

Our proposed solution is to take a small randomly sampled subset from an image dataset as input. Resizing of the images to a standard size would be performed as part of the pre-preprocessing. Once the images are resized, the algorithm will implement various saliency extraction methods on this subset of images and assess the quality of saliency maps obtained by each individual method using a suitable quality score, as well as take into account computation time required in each of the methods. Based on both the quality scores and execution time, a suitable weighted tradeoff mechanism will be formulated to rank the different saliency extraction techniques for the given dataset, which will allow the user to make an informed decision about which technique to be used for their task at hand.


Subset of Images
Pre-processing of Images
Implement different saliency extraction methods
Evaluate quality scores and execution time of each method
Ranking of the saliency extraction methods

3. Literature Survey

Borji et al [1] analyzed the effects of critical factors in Saliency models performance like center-bias and map smoothing. The study was conducted by comparing 32 state-of-the-art models to detect saliency and used shuffled AUC scores to discount for center bias. Evaluation metrics used in this study are the Correlation coefficient between a model saliency map and a human saliency map, Normalised Scanpath Saliency, and area under the ROC curve (AUC-ROC). The human saliency maps serve as a distribution of positive and negative samples, the salient region constituting the positive points and the negative points being uniformly selected. The saliency map obtained by particular models is then treated as a binary classifier and evaluated against its ability to separate the positive and negative points. Four well-known datasets were used for comparison: Toronto, NUSEF, MIT, and Kootstra in order to account for any dataset biases one particular dataset may have. How well a model is at predicting where the human tendency to look at is reported using shuffled AUC. Next, models are evaluated on their ability to predict the saccade sequence. Models are also compared based on their ability to decode task stimulus categories or sections of the human population the subject belongs to e.g. a normal person or an ADHD patient etc. Features are augmented from the statistics of fixations, saccades, and saliency at fixation to decode the stimulus categories. A multiclass SVM classifier with an RBF kernel was used to perform the classification. The study found that within their capacity of metrics AWS, LG, AIM, and HouNIPS consistently showed better performance than other models over all the four datasets considered.

Bylinskii et al [2] studied 8 major evaluation metrics of saliency maps and explored their property and suitability for different use cases. Some common evaluation metrics used in Saliency map evaluation are AUC-ROC, sAUC, Normalized Scanpath Saliency, Pearson's Correlation Coefficient, Earth Mover's Distance, Similarity or histogram intersection, Kullback-Leibler divergence, Information Gain. Experiments were carried out in the MIT300 dataset. Metrics are studied as functions that take the ground truth and the prediction as input and process them in order to output a numeric score assessing the closeness between them. Under the assumptions that all systematic dataset biases like center bias, blur, and scale, the study found NSS and CC to be the fairest metric to compare saliency maps. If saliency is probabilistically modeled, KL-divergence and information gain were found to be more appropriate metrics. Choice of a metric can also be based on specific application the saliency is being used for. For detection of salient regions, KL, AUC, and IG were found to be the most suitable whereas for applications like image-retargeting, compression, and progressive transmission, NSS or SIM are more appropriate.

The paper [3] proposes a co-saliency framework that explores the inter-image information via co-saliency and then performs co-saliency-based segmentation or localization on individual images. It has two key components, i.e Quality Measurement, which will measure and compare the quality of each saliency map with that of its corresponding co-saliency map, and Fusion Based Co-Saliency, which performs joint processing without introducing unnecessary parameters for tuning.

The proposed method has five points:-

• **Objective and Proposed Solution:** The total quality of any saliency map 'Si' is calculated by the product of ϕ(Si), i.e., the separation measure, and ψ(Si), i.e., the concentration measure. Thus, the objective is to achieve a saliency map set such that the total quality of comprising saliency maps is maximum.

• **Quality Measurement System:** Two measures used to determine the quality of the saliency map are Separation Measure (ϕ), which measures the separation between foreground and background, and Concentration Measure (ψ), which measures how concentrated the foreground pixels are.

• **Interaction:** It consists of 3 steps, i.e., Grouping, where images with similar appearance are grouped together in one group, using K-means clustering, and Saliency Warping, that involves alignment of one image w.r.t. another by establishing dense correspondence, and Saliency Fusion, where candidate saliency maps are fused together using ways like average, geometric mean, etc.

• **Improving Efficiency:** The original method required computing the costly dense correspondences for x(x 1) times, thus requiring quadratic time, whereas the modified method required just 2(x - 1) computations, thus requiring linear time, which is a much more efficient approach.

• **Applications:** The obtained high-quality saliency maps are further utilized for object level segmentation and localization. The datasets used for the co-segmentation evaluation are MSRC (14 categories with 419 images), iCoseg (38 categories with 643 images), Coseg-Rep (23 categories with 572 images), and Internet images dataset (3 categories: Airplane, Car, and Horse, with 4347, 6381 and 4542 images, respectively).

The dataset used for large-scale localization evaluation is the ImageNet dataset.

For evaluating the segmentation, Jaccard Similarity and Accuracy metrics are used, and for evaluating the localization, the CorLoc metric score is used. The results of both original and efficient methods were compared with the state-of-the-art co-segmentation and localization methods on different datasets, and it is seen that both the original and efficient methods performed better in terms of accuracy, Jaccard, CorLoc, and also in terms of speed.

To obtain good inferences from the data, their quality should be guaranteed; thus, prior to their analysis, the quality of the data should be evaluated. This paper [6] focuses on data quality metrics based on the notion of saliency score that is developed in terms of outliers. The proposed data quality metrics use visualization techniques to provide a holistic view of data quality in terms of outliers.

Pipino et al. addressed 16 subjective and objective data quality dimensions: accessibility, an appropriate amount of data, believability, completeness, concise representation, consistent representation, ease of manipulation, free-of-error, interpretability, objectivity, relevancy, reputation, security, timeliness, understandability, and value-added and also proposed three functional forms for developing objective data quality metrics: simple ratio, min or max operation, and weighted average.

Heinrich et al. presented six requirements that must be met in order for the data quality measures to hold: normalization, interval scale, interpretability, aggregation, adaptability, and feasibility, and also proposed metrics that meet these requirements such as correctness metric and timeliness metric.

In order to access the data quality, the saliency of each data item in associated attribute sets is evaluated, where the data space in associated attribute sets is partitioned into subspaces, and the data distribution in these subspaces suggests the data set quality. If a subspace contains a very small amount of data, then the data falling into that subspace may contain outliers. A new index called Saliency Score is proposed, which is sensitive to outliers. Using this saliency score, a ranking method is used to rank the subspaces according to their possibility of having outliers. The attribute sets with top-k ranking scores are determined as candidates of suspicious attribute subspaces. When applying the saliency score-based metrics, the associated attributes are first identified. Techniques such as 2 -measure association rule mining have been developed to select the associated attributes where 2 -measure evaluates the difference between real frequencies and expected frequencies and Association rule mining (like Apriori algorithm) is used to identify frequent attribute value combinations in a data set.

Experiments were conducted on the proposed method, in which the bank marketing data set from the UCI ML Repository was used, which consists of 45,211 data records with 17 attributes. It used the Apriori algorithm for selecting the associated attribute sets of size 3 or 4 and then computed saliency scores for them and visualized the data quality with and without noise. 43 attribute sets were selected as associate ones and the top 4 as suspicious ones. Thus, the experimental results showed that the proposed saliency score effectively brought out the existence of outliers in the data set

Human vision can solve many application related problems in Computer Vision effectively. The vision of human beings has a great capacity to detect distinct objects and regions of interest visually. That is the objects that look completely unique from others. Hence much focus is given to that object. This is referred to as saliency detection. In saliency detection, automatic identification of significant objects is done without any prior knowledge. It has attracted great attention in the field of computer vision technology. More than 100 saliency detection models have been proposed. In this paper 25 of them have been chosen and their performances are measured using four image datasets.

Ning Li et al [4] adopted the ECSSD, SED2, JuddDB, and DUTOMRON datasets. The ECSSD comprises 1,000 semantically meaningful but structurally complex images. SED2 contains 100 images. JuddDB contains multiple objects. The DUTOMRON dataset contains 5000 images. There are several categories of saliency detection that are done namely - Attention point prediction, Salient region detection(Global Contrast(Foreground Priori, Background Priori), Local Contrast(Foreground Priori, Background Priori) and Learning Algorithm) and Salient object detection and all the state of the art models are distributed among the categories.The Evaluation Measure that is being focussed conducted to effectively evaluate the models is Quantitative Evaluation as it is simple and more accurate. Measurements are applied to the overlapping area between the prediction map and the Ground Truth. The Evaluation Metrics of the Quantitative Evaluation are - Precision-Recall, Mean Absolute Error, and S-Measure. The study found that in Precision-Recall, after AM and UCF, DRF models have performed best in each of the 4 datasets. The SR performed the weakest among all models in three datasets except JuddDB. For MAE metric, AM and UCF models achieved good results compared to other models in each of the four datasets. In S-Measure also, AM and UCF models maintained good results. This shows that the convolution neural network models like AM and UCF can perform the best saliency detection whatever be the image and evaluation metric for measurement.

The paper written by Long Mai and Feng Liu [5] developed a method that will produce the best salient object detection result from many results produced by different methods for each and every input image. But here different salient object detection results were compared without any ground truth. A range of features was designed to measure the quality of salient object detection results. These features were passed in various machine learning algorithms to rank different salient object detection results. That is they used a learning to rank mapping method to rank salient object detection results of the same input image. That is they trained a binary classifier to compare the quality between every two saliency detection results and then combine these results to rank all the saliency detection results.There are many ways to design features like Saliency Coverage, Saliency Map Compactness, Saliency Histogram, Color Separation, Segmentation Quality and Boundary Quality.

Saliency Coverage- When the pixels of the Saliency Map covers a very large or very small area abnormally, the chances of getting a good map is less. Saliency Coverage Feature keeps track of it. It is denoted by fC(M) where M is the Saliency Map. A saliency map M(range[0,1]) is binarized using a threshold value t whose range is (0,1) and then the saliency coverage value is computed.

Saliency Map Compactness- A good saliency map should concentrate its salient pixels in a compact region in the image. Saliency Map Compactness value is denoted by fCP.. The more the fCP value, the better is the saliency map.

Saliency Histogram- If the saliency histogram contains concentrated peaks at both ends, then the saliency map is good as it well separates the foreground from the background. This feature is measured by fH.

Color Separation- The color separation feature fCS is measured as the intersection between the histograms for the salient and background region.

Segmentation Quality- The feature fNC was designed to measure the quality of a saliency map by assessing the segmentation result it induces. There is a threshold t that divides the image into Salient Region St and Background Region Bt. Then the map is binarized. A good segmentation result would maximize the intra-region similarity and minimize the inter-region similarity.

Boundary Quality- The saliency map boundary must match well with the edges of the input images. The boundary quality features fB takes care of it. Generally, better saliency maps have higher boundary quality feature values.

Now ground-truth based ranking is done by ranking all the saliency maps according to the AUC score. Now they had designed the pairwise preference model PMi,Mj which meant that probability that Mi has higher quality than Mj for which they trained a binary classifier(they have used SVM, Random Forest Classifier and MLP) that takes fMi,Mj(concatenation of feature vectors) as input and outputs the preference label 1 if Mi has higher quality than Mj and 0 otherwise2. After training the pairwise preference model, it was used for ranking salient object detection results on new images.The ranking is denoted by r(Mi). The overall ranking for every saliency map Mi can then be obtained by sorting to their scores.

Their methods were experimented on the public salient object detection benchmark THUS-10000. The dataset contains 10000 images. Each image is associated with a segmented salient object mask. For each of the experiments in this section, 2000 images were randomly selected for training their ranking model and the rest 8000 images for testing. The saliency maps were generated by 10 state-of-the-art salient object detection methods. Two methods were also used for baseline ranking the saliency images.They are Mean-AUC-Based Ranking(MAR) and Voting-Based Ranking(VBR).They were compared with the other methods that were done above. Now they compared their ranking results with ground-truth ranking using two rank correlation metrics.To evaluate how the ranking results agree with the ground-truth ranking, they computed their rank correlation. Two rank correlation metrics were used. In both the rank correlation metrics. The results showed that the saliency map ranking from their method has higher correlation with the ground-truth ranking than those from the two baseline methods.

Similarly Rank-n Accuracy metric also showed that the ranking predicted by their model is better than the baseline ranking in selecting the best salient object detection results.But it gives little details for the cases where the system fails to predict the true best saliency map.So finally the best saliency map will be considered correct if the AUC score of the saliency map differs from the true map by not more than tolerance value.It was found that their methods make less number of errors with small values of tolerance value compared to the baseline methods.Thus their method can be used to improve the overall salient object detection performance by select good saliency maps for each input image.

4. Methodology

1. At first, we computed the saliency masks of the images, randomly sampled from five different datasets, using various Deep learning and Non-Deep Learning methods. We used the following methods-

   NON–DL Based Methods:-

   (a) Saliency computation using SLIC super pixel segmentation.

   (b) Saliency computation using K-means clustering.

   (c) Saliency computation using OpenCV's static fine grained saliency detector.

   (d) Saliency computation using OpenCV's static saliency spectral residual detector.

   Deep Learning Based Methods:-

   (a) Saliency computation using PoolNet.

2. Using the OTSU's method, we extracted the foreground and background for saliency masks of all the five sampled images of the given dataset.

3. Using the foreground and background obtained, we retrieved their pixel distributions.

4. Then, we calculated the mean and standard deviations of both the foreground and background distributions.

5. Using the representative gaussian distributions of the foreground and background, we calculated the **KL Divergence between the foreground and background distributions**, which gives the separation measure between both the distributions. **The formula to compute KL Divergence between two normal distributions**,

$$KL(p,q) = -\int p(x)logq(x)dx + \int p(x)logp(x)dx$$

$$= \frac{1}{2}log(2\pi\sigma_2^2) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}(1 + log2\pi\sigma_1^2)$$

$$= log\frac{\sigma_2}{\sigma_1}\frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2} \quad (1)$$

   Where **p(x)** and **q(x)** are the Gaussian distributions,and
    are the standard deviations and means respectively.

6. We computed the **concentration measure** of the foreground as described in the paper [3].
   Concentration Measure: It measures, how much concentrated the foreground pixels are. A good quality saliency map should contain concentrated foreground pixels. The foreground often gets distributed into multiple components, but the ideal scenario demands that there should be one largest object component and the others (if exist) should dissipate from it. Higher the contribution of this largest component towards the foreground and lesser the dissipation of foreground into components, higher will be the foreground concentration.
   Let **O(S) = O1 (S), O2 (S),...,O—O(S)—(S)** denotes the set of object components, and **Cu(S)** is the fraction of the total foreground area covered by the object component **Ou(S)**, which is measured as,

$$C_u(S) = \frac{[O_u(S)]}{\sum_{u=1}^{[O(S)]}[O_u(S)]} \quad (2)$$

   Where [.] = area of Ou(S) , and —.— = cardinality.
   Finally, the concentration measure for saliency map 'S' is calculated using the formula,

$$CM(S) = C_{u^*}(S) + (1 - C_{u^*}(S))\frac{1}{|O(S)|} \quad (3)$$

   Where $C_{u^*}(S)$ is the contribution of the largest object component and $(1 - C_{u^*}(S))\frac{1}{|O(S)|}$ is the measure of amount of dissipation of foreground into object components.

7. Then the quality score is calculated as the sum of logarithm of separation measure(KL Divergence between foreground and background) and concentration

measure for all N samples divided by the number of samples.

The formula to compute Quality Score (Q) is given as,

$$Q = \frac{\sum_{i=1}^{N}(log_{10}(KLD(bg, fg))_i + CM_i)}{N} \quad (4)$$

Where **CM** denotes the **Concentration Measure** and **N** denotes the **Number of Samples**.

8. The time measure is calculated by taking into account the time required by the slowest method T_{max} and the time required by the fastest method T_{min}.The time rewards the score of a method by how much faster it is w.r.t.the slowest method and penalises it by how much slower it is w.r.t. the fastest method.The computation formula is shown below-

$$T = log_{10}(\frac{T_{max} - t_{method} + 0.5}{t_{method} - T_{min} + 0.5}) \quad (5)$$

Where T_{max} denotes the **maximum method time**, T_{min} denotes the **minimum method time**, and t_{method} denotes the **method time**.

9. Then we obtained the maximum quality score and finally the total score is computed for each saliency method by adding the quality score of the method under consideration with its quality score divided by the maximum quality score, and multiplied with the total time of that method. The total score formula is shown below-

$$Score = Q + \frac{Q}{Q_{max}}.T \quad (6)$$

Where Q_{max} denotes the **maximum quality score**, **Q** denotes the **quality score of the method under consideration**, and **T** denotes the **total time of that method**.

5. Experimental Results

The two tables below shows the Concentration Measure of the five images sampled from the dataset using the five different saliency methods.

Saliency Method	Image1	Image2	Image3	Image4	Image5
SLIC	0.78525391	0.8809082	0.92334595	0.81125276	0.51864299
Spectral	0.92325265	0.88987732	0.79123688	0.84976959	0.87145996
PoolNet	0.82992554	0.8182373	0.67837524	0.73944092	0.83001709
KMeans	0.72297974	0.86523438	0.70947266	0.80463664	0.59030151
FineGrained	0.77011719	0.53837077	0.68410492	0.5736618	0.73413086

In our experimentation, we have taken N=5.

The table below shows the Quality Score of each saliency method implemented for cars dataset.

Saliency Method	Quality Score
PoolNet	1.6461977043406797
Spectral	0.9755658860995446
FineGrained	0.6238775849766218
SLIC	0.5058538830657897
KMeans	0.329015814293206

The table below shows the Quality Score of each saliency method implemented for cats dataset.

Saliency Method	Quality Score
PoolNet	1.9798671249748536
Spectral	0.9874981646109104
FineGrained	0.7869570668381002
SLIC	0.5058538830657897
KMeans	0.329015814293206

The table below shows the Total Time of each saliency method implemented.

Saliency Method	Time
SLIC	6.617721425400001
Spectral	0.28571752589996324
PoolNet	2.38
KMeans	163.07433261199913
FineGrained	0.2885304813999937

The table below shows the Final Score (in sorted order) of each saliency method implemented.

Saliency Method	Final Score
PoolNet	3.4020093076960802
Spectral	2.4657671290707226
FineGrained	1.5729031968250538
SLIC	0.9488274866356937
Kmeans	-0.1735640548914374

The figures below shows results obtained from using saliency method along with their foreground and background distribution graphs.

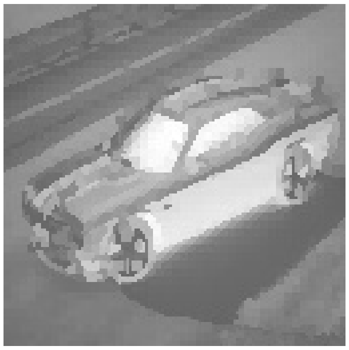


Figure 1. SLIC Result

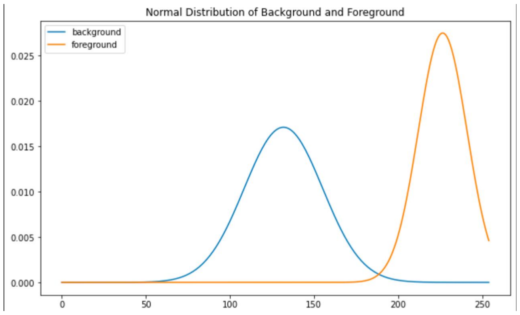


Figure 2. SLIC Distribution

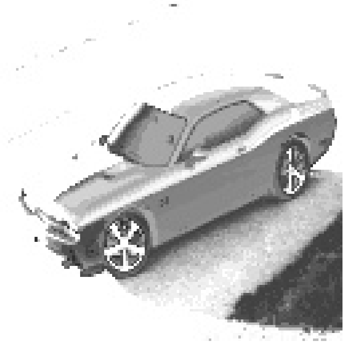


Figure 3. Kmeans Result

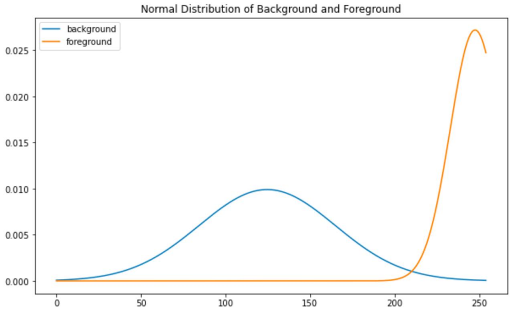


Figure 4. Kmeans Distribution

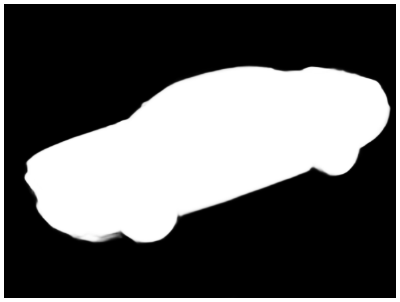


Figure 5. PoolNet Result

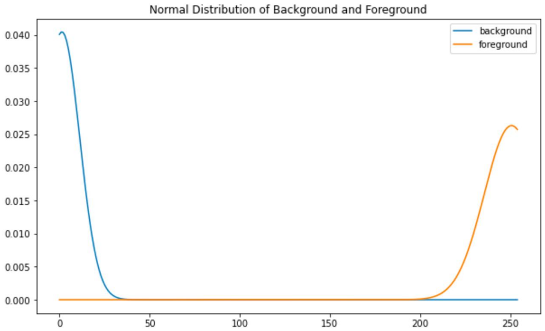


Figure 6. PoolNet Distribution

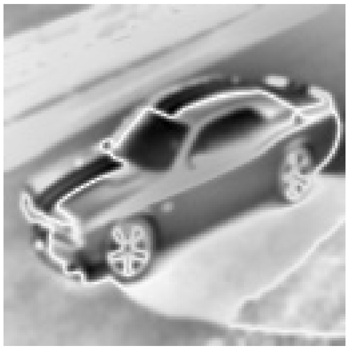


Figure 7. Fine Grained Result

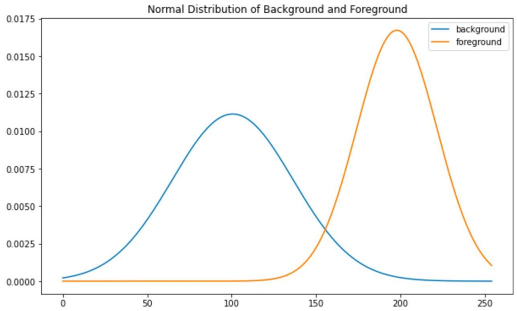


Figure 8. Fine Grained Distribution

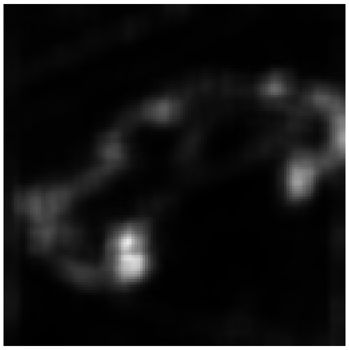


Figure 9. Spectral Result

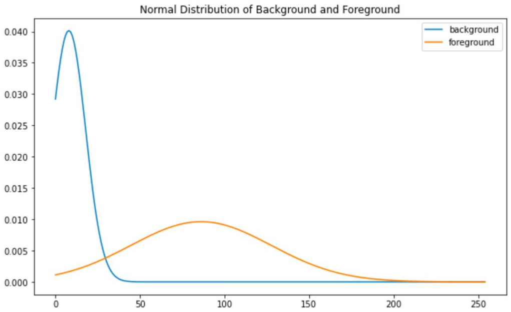


Figure 10. Spectral Distribution

6. Conclusion

Assuming distribution of foreground and background pixels follow normal distribution, KL Divergence between the two distributions can be computed using a closed form solution which makes the process very fast. KL Divergence gives us an idea about how well separated foreground and background pixels are in the saliency map. Concentration measure gives an idea about how well connected the foreground blobs are in the saliency maps. These two measures combined gives us the overall idea about the quality of the saliency map. This quality score has been rewarded or penalized on the basis of how fast the saliency extraction method has been.

The method proposed in this paper gives us the time sensitive ranking of various saliency extraction methods for a input dataset.The final rankings of various saliency extraction methods in the tables shown above highlight that Deep Learning based **PoolNet** gives the best quality saliency map with reasonably lesser time consumption.

References

[1] Ali Borji, Hamed R. Tavakoli, Dicky N. Sihite, and Laurent Itti. Analysis of scores, datasets, and models in visual saliency prediction. In *2013 IEEE International Conference on Computer Vision*, pages 921–928, 2013-12. 2

[2] Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. What do different evaluation metrics tell us about saliency models? 2017. 2

[3] Koteswar Rao Jerripothula, Jianfei Cai, and Junsong Yuan. Quality-guided fusion-based co-saliency estimation for image co-segmentation and colocalization. pages 2466–2477, 2018. 2, 5

[4] Ning Li, Hongbo Bi, Zheng Zhang, Xiaoxue Kong, and Di Lu. Performance comparison of saliency detection. In *Advances in Multimedia*, 2018. 4

[5] Mai Long and Liu Feng. Comparing salient object detection results without ground truth. In *Computer Vision – ECCV 2014*, pages 76–91, 2014. 4

[6] Keon Myung Lee Yong Ki Kim. Saliency score-based visualization for data quality evaluation. pages 289–294, 2015. 3