

# A Result Paper on Salient Object Detection using Edge Preservation and Multi-Scale Contextual Neural Network for identification

Pooja P. Ingole, S.C. Nandedkar

<sup>1</sup>Department of Computer Science and Engineering Marathwada Shikshan Prasark Mandal's Deogiri Institute of Engineering & Management Studies, Aurangabad Maharashtra state, India 2018-2019

<sup>2</sup>Assistant Professor, Department of Computer Science and Engineering Marathwada Shikshan Prasark Mandal's Deogiri Institute of Engineering & Management Studies, Aurangabad Maharashtra state, India 2018-2019

**Abstract:** Everywhere we go we take photographs, images, selfies etc for any social or private lifestyle of human being. That photos must be cleared, sharpened, filtered for this we have different techniques, application etc the one method is Salient Object Detection using Edge Preservation and Multi-Scale Contextual Neural Network. In this paper, we propose a novelty on edge preserving and multi-scale contextual neural networking for saliency. The full proposed framework architecture is mainly aims towards different address two ceiling of the existing CNN based methods. In recent years, salient object detection, which aims to detect object that most attracts people's attention throughout an image scenes along with whole foreground and background, has been widely exploited. The center of attraction and focus is to get maximum optimized result using the techniques like edge preserving, contextual networking, multi-scaling, detecting the most particular important object with feature extraction using SIFT algorithm. It has also been widely utilized for many computer vision tasks and digital image processing digital marketing, such as semantic segmentation, object tracking and image classification. These proposed framework achieves and optimizes the goals, targets both clear detection boundary of an input image and multi-scale contextual neural networks robustness simultaneously time being, and thus achieves an optimized performance results. Therefore, we get different and many more experimental results that's sets higher benchmarks, result and applications using various datasets express that this designed proposed methods are to achieves through best and high leveled state-of-the-art performance optimally therefore our proposed system produce the best experimental result that demonstrates recognition, identification of rated object is higher than other method and also reach to high performance with optimal solution in an image .

**Keywords:** Salient object detection, edge preservation, multi-scale context neural networks, RGB-D saliency detection, object mask pooling, SVM, SIFT.

## 1. Introduction

Edge Preserving and Multi-Scale Contextual Neural Network for Salient Object Detection with many applications are used to various different approaches in different process now a days. Its application include computer vision , medical image analysis, financial security, law Image processing on high level Object detection, Data driven animation, Pattern recognition, Artificial intelligence ,Computer vision, Medical image analysis for engineering, Brain mapping and Finding tumors in mammograms Automatic target detection (e.g., finding traffic signs along the road or military vehicles in a savanna or dense forest areas), Robotics (using salient objects in the environment as navigation landmarks) Image and video compression (e.g., giving higher quality to salient objects at the expense of degrading background clutter) identification enforcement etc. It is most powerful way that people to get more cleared and high definition of any an image we have to work on the phenomenon like object detection, object recognition, enhancement, restoration, representation, object identification etc. Our topic targets the preserving and detecting and enhancements of techniques for the best result from the aspects that we are using for classification of an image into different format. This type of analysis is aims to analyze a very new framework for tackling the image processing techniques in advanced manner. Here we are proposing a novelty edge preserving and multi-scale contextual neural network for salient object detection to get

the maximum optimized output from the data that given to satisfy the all requirement form the client or customer. When observing such real-world objects with a camera or an eye, there is an addition scale problem due to perspective effects. A nearby object will appear larger in the image space than a distant object, although the two objects may have the same size in the world. The aim of this whole process is to get best result for this we have to implement our images, data informational product and many more. Being get Motivated by these, In a this paper presents [1], we first propose an end-to-end edge-preserved neural network based on Fast R-CNN framework that called as or named as *RegionNet* to efficiently generate saliency map with sharp object boundaries. Furthermore, our method can be generally applied to RGB-D saliency detection [2] by depth refinement as we already mention above. The proposed framework achieves both clear detection boundary and multiscale contextual robustness simultaneously for the first time, and thus achieves an optimized performance. Experiments on various benchmarks are the paradigm of versatile of datasets demonstrates that the proposed method achieves state-of-the-art performance.

## 2. Literature survey

**Preprocessing – Noise Removal-**Digital image processing is always an interesting field as it gives improved pictorial information for human interpretation and processing of image data for storage, transmission, and representation for machine perception. Image Processing is a technique to enhance raw images received from cameras/sensors placed on satellites,

space probes and aircrafts or pictures taken in normal day-to-day life for various applications. This field of image processing significantly improved in recent times and extended to various fields of science and technology. The image processing mainly deals with image acquisition, Image enhancement, image segmentation, feature extraction, image classification etc. The basic definition of image processing refers to processing of digital image, i.e. removing the noise and any kind of irregularities present in an image using the digital computer. The noise or irregularity may creep into the image either during its formation or during transformation etc. The elements are called picture elements, image elements, peels, and pixels. Pixel is the most widely used term to denote the elements of a digital image.

**Edge Preservation using Gradient Magnitude and Non-Linear Diffusion**-Traditionally and comparatively Saliency methods main purpose or aim is to generate a Heat map, which gives us each pixel a relative value of its level of saliency. Here 3 are used to indicate the corresponding values of R, G and B. Here the gradient operator will apply on RGB images to compute the gradient of each components of color image to determine the magnitude and the direction of any intensity value. Gradient ascent methods iteratively refine the super-pixel clusters to optimize the segmentation until convergence criteria are reached. These methods use a tree graph structure to associate pixels together according to some criteria to getting the preservation non linear diffusion is required.

**Combining the saliency map**-In computer vision and image processing salient region along with the object identification is testing issue and critical point. It has a broad variety of uses, for example, object classification and recognition. The entire image is subset of salient region. The bottom up method is use for salient region object identification. There are different visual cue are available for salient region detection. Two visual cues are used for salient object. The compactness and local contrast visual cues are use. These two visual cues are reciprocal to each other. The saliency map is addition of compactness and local contrast map. After getting saliency map of object, then approach towards recognition of that object. The geometrical features of salient object are calculated. The goal of salient region detection is in terms of saliency map, from an image where the detected regions would draw the attentions of people at the principal sight of an image. Saliency identification is utilized to choose consequently the sensory data that is notable to a human vision framework.

### CNN Based Methods

We are going to Review on Existing Papers currently available, in a previous paper [3], the author proposed a new way to tackle the recognition scheme, preserving and finding detecting most desirable object from input that have already given. The entire system concentrates on group-wise deformable image registration and imposing on problem which that image or data that already unclear or damaged. In this section of this review paper we introduce some traditional and fairly yet new different salient detection methods and the recent CNN based methods. There are some following many methods utilizing bottom-up priors are proposed, readers and students are encouraged to finding more about the details in a recent survey paper by Borji *et al.* [4]. In addition, to this work we also introduce some related works that integrate multi-scale context information and some topics related to salient object detection. This method achieves highest recognition rate of features. There are some traditional

methods present in this we have to elaborate in the computer vision society and in the forum of digital image processing Salient object detection was first exploited by Itti *et al.*, and later on this technique grabs the attention and attracted wide amount by the most followers over world. Salient object detection becomes the 'Epic method' for encapsulating lost object hidden probably. Hence this are some Traditional methods mostly rely on prior assumptions and most are unsupervised by many things which we are future sees in this paper.

The proposed system work is planned for the use in image. Here, we propose a saliency map which is more productive and this model uses highlights like compactness and local contrast of an image for producing saliency map. There are different visual cue are used for saliency detection. The uniqueness, background and compactness are visual cues. Salient objects is only the genuine items, henceforth they are assembled [5]. Saliency detection approaches may roughly classified into different handcrafted feature based methods and deep learning based methods. There are having an overview of these categories, we explore methods for multi-scale feature fusion in this section. Handcrafted Features for Saliency Detection [6] – Priority wise the deep learning revolution, conventional saliency methods were mainly relied on handcrafted features, We referred here by giving an over-segmented image, color contrast has been exploited that formulated saliency detection as an image segmentation problem. This paper [7] mainly focuses on a new method for improving region segmentation in sequences of images when temporal and spatial prior context is available. RGB-D Salient Object Detection and Multi-Scale Context the paper [8] consist Later on Multi-scale context has been proved to be useful for image segmentation tasks given by Hariharan *et al.* [9] Furthermore, In this paper, we aim to propose a unified framework which can preserve object boundaries and take multi-scale spatial context into consideration.

**Support Vector Machines (SVMs)**-Classification and identification of Images Using Support Vector Machines (SVMs) are a relatively new supervised classification technique to the land cover mapping community. They have their roots in Statistical Learning Theory and have gained prominence because they are robust, accurate and are effective even when using a small training sample. By their nature SVMs are essentially binary classifiers, however, they can be adopted to handle the multiple classification tasks common in remote sensing studies.

**Scale Invariant Feature Transform (SIFT)**-It is a technique for detecting salient, stable feature points in an image. For every such point, it also provides a set of "features" that "characterize/describe" a small image region around the point. These features are invariant to rotation and scale. Motivation for SIFT is Image matching where Estimation of affine transformation / homograph between images along with Estimation of fundamental matrix in stereo, Structure from motion, tracking, motion segmentation All these applications need to detect salient, stable points in two or more images, and determine correspondences between them. To determine correspondences correctly, we need some features characterizing a salient point. These features must not change with: Object position/pose, Scale, Illumination, Minor image artifacts/noise/blur One could try matching patches around the salient feature points – but these patches will themselves change if there is change in object pose or illumination. So these patches will lead to several false

matches/correspondences. SIFT provides features characterizing a salient point that remain invariant to changes in scale or rotation. Following are the steps evaluated and illustrated of the proposed system that used in procedure of Saliency detection over the all scenes in an image ,dotting or highlighting the important features of the Image along with white dots i.e. image with Key points are dots onto it.

### 3. System Development

System architecture 3.1 shows overall structure of proposed method. There are two main Stages atlas construction and Recognition stage. This architecture shows all sub process that is Feature extraction, testing, classification etc. Here, given a query image, estimate the correct expression type, such as image features. image sequence or dataset contains not only image appearance information in the spatial domain, but also evolution details in the temporal domain.

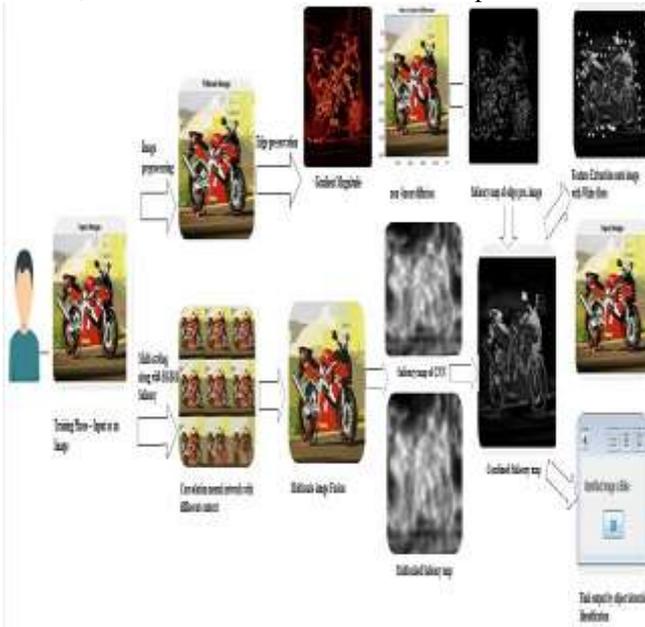


Figure 3.1 – System Architecture

Formulating salient object detection as a binary region classification task as the flow are heavy but steps are understandable, The image is first given as input gets segmented into regions, this regions are segmented by edge-preserving methods, saliency map generated by our network is naturally with sharp boundaries while using Edge preserving techniques like Non-linear diffusion gradient magnitude. We achieve this by taking advantages of dense image prediction. We also perform the process simultaneously like image fusion and feature extraction via identification.

### 3.2 Methodology

The idea of edge-preserving saliency detection of an image picture’s object based on a CNN network previously appeared and extended this idea with consideration of multi-scale spatial context. We target to propose a unified framework proposed with which we can preserve object boundaries and take multi-scale spatial context into consideration. 1) **Network Architecture:** In our structure is very efficient and general framework art in which the convolutional layers are shared on the entire image and the feature of each region is extracted by the RoI pooling layer. But Different from previous region-based methods which deal with each region of an image independently, our proposed structure processes all regions end-to-end per regions globally or locally over per pixels and with the entire image considered

as follows. 2) **Detection Pipeline:** As image, we segment it into regions using super pixel and edges of same as that we already used in framework similar with object detection tasks via semantic segmentation fragments of image scenes are formed. To generate the image region mask of scene in an image each pixel and then down sampling and downsizing it by 16 times and put it into the RoI pooling layer. Then, at the RoI pooling stage, features inside each RoI ( $h \times w$ ) matrix are pooled into a fixed scale  $H \times W$  ( $7 \times 7$  in our work), that's why each sub-window we having with scale  $h/H \times w/W$  is get converted to one(1) value with max-pooling. This all steps are utilized in extracting the features of irregular pixel-wise RoI region, we are going to pool that features inside its region mask while leaving others as 0 values only. The whole process of the proposed art mask-based of RoI pooling and masking is formulated as follows as

$$P_j = \begin{cases} \{(k|k \in SW_j^{max}, M_k = i)\}_k^F & i \in M(SW_j) \\ 0 & i \notin M(SW_j) \end{cases}$$

Where, region of that masked image having values with index  $i$ , and a certain sub-window as  $SW_j$ ,  $F$  for features before pooling, region mask as  $M$ , the pooled feature at sub-window  $SW_j$  as  $P_j$ , are the value that we denote vicesly in the equation. The superpixel is segmented using SLIC algorithm may be belongs here. In the context the first three layers of each branch are with  $3 \times 3$  convolutional filters and 64, 64, 128 channels, and the dilated convolution applied to increase the receptive field. The last two layers are fully convolutional layers with 128 and 1 channels to generate saliency map with one eighth scale of the original input images branches. The outputs of all branches are then fused into fully convolutional layers which learn the combination weights to generate saliency map  $SC$ . The final saliency map  $S$  is then got by fusing  $SS$ ,  $SE$ , and  $SC$  via a fully convolutional layer is in form of neural networks.

$S = Fusion(S_S, S_E, S_C)$  Results of previous region-based methods of that method and our  $SS$  and  $SE$ . We can see that misclassification or disposed of regions has a great impact on the final performance and most regions are assigned to near either 0 or 1, with few intermediate values. These will limit the precision at high recall when thresholding over image region global value or locally. **C. Loss-**During the proceed operations we may have to consider the loss over the full Image because of multitasking, frequent operation going on that image. So damage will be on while processing each step to signifies that we have to formulate that loss. We assume that the training data,  $D = \{(X_i, T_i)\}_{Ni=1}$ , consists of  $N$  training images and ground truth. Our goal is to train a convolutional network  $f(X; \theta)$  to predict saliency map for an given image. We have to define two kinds of loss for *ContextNet* to generate saliency map with high accuracy and clear object boundary. The first *Loss* is common used Cross Entropy Loss  $LC$ , which aims to make the output saliency map  $f(X; \theta)$  consistent with the groundtruth  $T$  may calculated or formulate by equation follows -

$$L_c = -\frac{1}{N} \sum_{i=1}^N [T_i \log(f(X_i; \theta)) + (1 - T_i) \log(1 - f(X_i; \theta))]$$

The second *Loss* is Edge Loss  $LE$  which aims to preserve edge of image and make the saliency map more uniform and precise. Since we have segmented image into regions with edge-preserving methods, that saliency map in the same region should share similar value, so that the final saliency map can also preserve edge and be more uniform we have to assume this method loss can bigger. We average saliency map  $f(X; \theta)$  in each region and marked the averaged map as  $f^-(X; \theta)$ . The Edge Loss is defined as the  $L2$  norm between saliency map  $f$

$(X; \theta)$  and the averaged map  $f^-(X; \theta).L_\epsilon = \frac{1}{2N} \sum_{i=1}^N \|f(X_i; \theta) - \bar{f}(X_i; \theta)\|_2^2$ . Saliency map with features at different scale, which accelerates convergence of the network and makes the final saliency map more precise and accurate similar to data originally.

**Steps of SIFT algorithm** •Determine approximate location and scale of salient feature points (also called key points)  
 •Refine their location and scale •Determine orientation(s) for each key point. •Determine descriptors for each key point.

**Step 1:** Approximate key point location, Look for intensity changes using the difference of Gaussians at two nearby scales:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$

$$= L(x, y, k\sigma) - L(x, y, \sigma)$$

Convolution operator: refers to the application of a filter (in this case Gaussian filter) to

Difference of Gaussians = “DoG” Scale refers to the  $\sigma$  of the Gaussian. **Step 2: Refining key point location**

•To find an extremum of the DoG values in this neighborhood, set the derivative of  $D(\cdot)$  to :

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{\sigma} \end{pmatrix} = - \left( \frac{\partial^2 D(x, y, \sigma)}{\partial(x, y, \sigma)^2} \right)^{-1}_{x=x_i, y=y_i, \sigma=\sigma_i} \begin{pmatrix} \frac{\partial D(x, y, \sigma)}{\partial(x, y, \sigma)} \end{pmatrix}_{x=x_i, y=y_i, \sigma=\sigma_i}$$

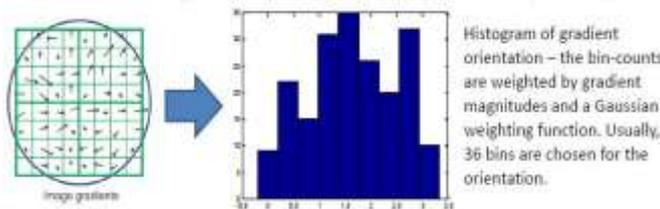
$$\therefore D_{\text{extremal}} = D(x_i, y_i, \sigma_i) + \frac{1}{2} \begin{pmatrix} \frac{\partial D(x, y, \sigma)}{\partial(x, y, \sigma)} \end{pmatrix}^T_{x=x_i, y=y_i, \sigma=\sigma_i} \begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{\sigma} \end{pmatrix}$$

The keypoint location is updated. All extrema with  $|D_{\text{extremal}}| < 0.03$ , are discarded as “weak extrema” or “low contrast points”. An edge will have high maximal curvature, but very low minimal curvature. A keypoint which is a corner (not an edge) will have high maximal and minimal curvature. **Step 3: Assigning orientations** -Compute the gradient magnitudes and orientations in a small window around the keypoint – at the appropriate scale.

$$I(x, y, \sigma) \quad G(x, y, \sigma) * I(x, y)$$

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)))$$



**Step 4: Descriptors for each keypoint** For scale-invariance, the size of the window should be adjusted as per scale of the keypoint. Larger scale = larger window. The SIFT descriptor (so far) is not illumination invariant –the histogram entries are weighted by gradient magnitude. •Hence the descriptor vector is normalized to unit magnitude. This will normalize scalar multiplicative intensity changes. •Scalar additive changes don’t matter –gradients are invariant to constant offsets anyway. •Not insensitive to non-linear illumination changes

Resistant to affine transformations of limited extent (works better for planar objects than full3Dobjects).

**4. PERFORMANCE ANALYSIS**

To verify the performance of proposed method, face images are collected as training data. We perform all process on test image. We implement and investigate our performance method using Caffe framework of MIT Database. The training process consists of several stages.

**Steps Followed next** - The confusion matrix for this test set is Selecting feature point locations using the Grid method. Extracting SURF features from the selected feature point locations. Finished creating Bag-Of-Features Training an image category classifier for 10 categories. BlackWhite, Indoor, Jumbled, lineDrawing, lowResolution, Noisy, Random, Satellite, Sketch, Social. Encoding features for 1000 images...done. Finished training the category classifier. Use evaluates to test the classifier on a test set.



Figure 1: Input image



Figure 2: Edge Preserving via Gradient Magnitude, Non-linear diffusion, Saliency edge map

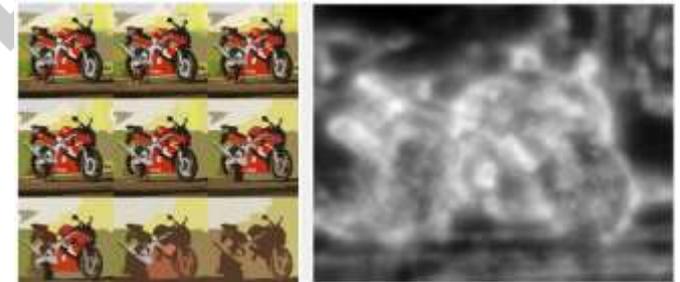


Figure 3: Convolutional Neural network for deep learn and with saliency map

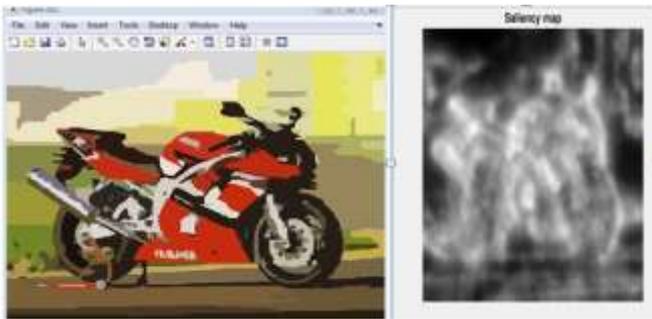


Figure 4: Multiscale contexting along with Saliency map

At the stage, we fine-tune the *RegionNet* using weights pre-trained on Image. We evaluated that, detection rate has been increased as well as improve performance accuracy and efficiency as compared to other methods.



Figure 5: Final output showing the exact category of which image is belong i.e. Identification image is Bike

As show in figure. Although experimental protocols of compared methods are not exactly the same Recognition rate is higher than other methods.state-of-the-art methods, including traditional methods: Black White Indoor Jumbled Line Drawing Low Resolution Noisy Random Satellite Sketch Social and CNN based methods

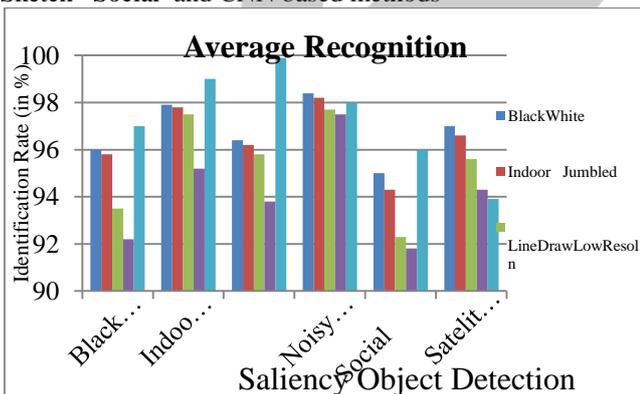


Figure 5 : Average Accuracy rate of diff Categories of Images

Methods	Identification accuracies (%)
BlackWhite	94.13
Indoor, Jumbled	95.40
LineDrawing ,LowResolution	96.48
Noisy, Random	96.78
Satellite, Sketch	99
Social	93

Table I : Average Identification rate of all Image Categories we use comparable training data. Fig. shows PR-curves, F-measure and MAE on six benchmark datasets. We

can see that our method outperforms other methods and our preliminary conference method by a large margin. For the state-of-the art

Method %	B/W	Ind, Jum	LiD/LRes	N/R	Sat/Sketh	Social
Black White	96	0	0	0	2.0	0
IndoorJ umbl	0	97.9	0	0.6	1.5	0
LineD LRes	0	0	100	0	0.5	1.5
Noisy Randm	0	0.5	0	98.4	0	0
Satelite , Sketch	2.4	0.8	0.5	0	95	0.8
Social	0	0	2.5	0.4	0	97

Table: II Confusion matrix of proposed method

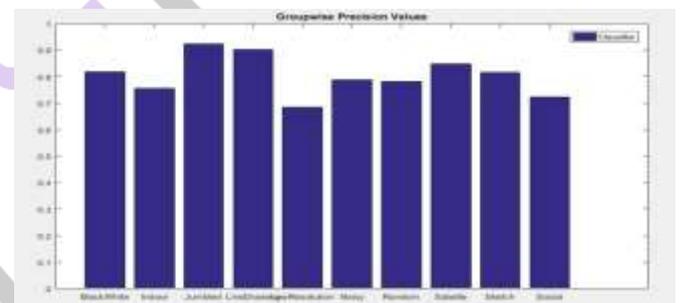


Figure 6-Average accuracies obtained by the proposed method with different values of precision GroupWise classifierβ.

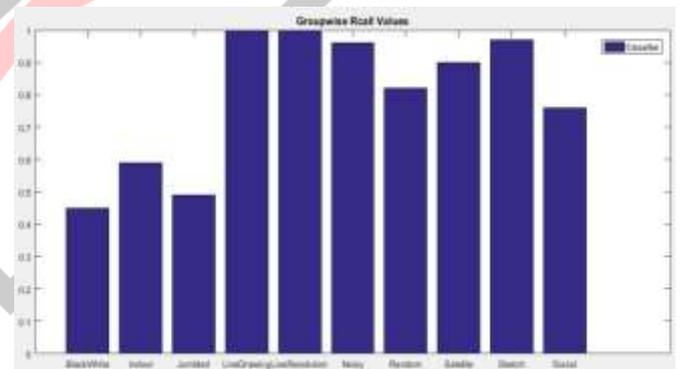


Figure 6: Recognition rate with different methods.

To study the effects of different values of N, average recognition accuracies with respect to different values of N are shown in Figure. It can be seen that when N is small (e.g., N = 4), inferior recognition accuracies are obtained because the Classifier describe the expression evolution process sufficiently. As N increases, the representation power of becomes stronger and higher recognition accuracies are obtained, For instance, when N = 12, satisfactory recognition accuracies are obtained (i.e., 97.2%). However, when N further increases, the recognition accuracy begins to saturate because the atlas sequence has almost reached its maximum description capacity and the gain in recognition accuracies becomes marginal.  $Precision = \frac{tp}{tp+fp}$ ,  $Recall = \frac{tp}{tp+fn}$   
Here,  $tp$  is the number of true positives,  $fp$

is the number of false positives, and  $fn$  is the number of false negatives, all of which are illustrated in Precision can be considered as background-suppression accuracy, whereas recall is related to salient-object detection accuracy. Precision and recall value pairs can be estimated by comparing two binary masks. Recall = 0.7940 Precision = 0.8041 Specificity = 0.9771, also We then calculate the precision-recall pair of the salient-object map that is binarized via the adaptive threshold. Finally, the F-measure can be computed as follows:

$$F - \text{Measure} = \frac{(\beta^2 + 1) \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}$$

Here,  $\beta^2 + 1$  is the coefficient of trade-off between precision and recall. The average F-Measure values of the salient object detectors on the well-known datasets are shown. The performance more or less follows the same trend, where it increases from uniqueness to spatial connectivity-based methods. The only exceptions are the uniqueness-based methods LR and RC, both of which perform only slightly better than the spatial variance-based methods. F\_score = 0.7811, Saliency --- Processing channel : 1 / 3 Processing channel : 2 / 3 Processing channel : 3 / 3 Time taken for finding the key points is : 1.233627 Time taken for finding key point descriptors is : 1.588664

## 5. Conclusion

Based on different technique in this, we propose *RexNet* which generates saliency map end-to-end and with sharp object boundaries framework, image is first segmented into two scales of complementary regions: super pixel regions and edge regions. The network then generates saliency score of regions end-to end and context in multiple layers are considered to fuse with region saliency scores. The proposed *RexNet* achieves both clear detection boundary and multi-scale contextual robustness simultaneously for the first time, thus achieves an optimized performance on datasets demonstrate that the proposed method achieves state-of-the-art performance detection of salient object. Due to the efficiency of the super-pixel wise mechanism along with region based, the proposed SuperCNN with Saliency object detection can be applied to other CNN applications, such as image segmentation, image classification and image parsing, image enhancement, image identification via methods in detection accuracy consistency and speed.

## References

- [1] Xiang Wang, Huimin Ma, Member IEEE, Xiaozhi Chen, and Shaodi You, "Edge Preserving and Multi-Scale Contextual Neural Network for Salient Object Detection" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 27, NO. 1, JANUARY 2018
- [2] R. Girshick, "Fast R-CNN," in *Proc. ICCV*, 2015, pp. 1440–1448.
- [3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [4] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. (2014). "Salient object detection: A survey." [Online]. Available: <https://arxiv.org/abs/1411.5878>
- [5] "DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection" by Xi Li, Liming Zhao, Lina Wei, Ming-Hsuan Yang, Fei Wu, Yueting Zhuang, Haibin Ling, and Jingdong Wang in 2017
- [6] Deep Edge-Aware Saliency Detection Jing Zhang, Student Member, IEEE, Yuchao Dai, Member, IEEE, Fatih Porikli, Fellow, IEEE and Mingyi He, Senior Member, IEEE arXiv:1708.04366v1 [cs.CV] 15 Aug 2017.
- [7] Contextual smoothing of image segmentation by Jonathan Letham & Neil M. Robertson Heriot-Watt University Edinburgh, UK..
- [8] SALIENT OBJECT DETECTION FOR RGB-D IMAGE VIA SALIENCY EVOLUTION Jingfan Guo<sup>1,2</sup>, Tongwei Ren<sup>1,2</sup>, Jia Bei<sup>1,2</sup> State Key Laboratory for Novel Software Technology, Nanjing University, China Software Institute, Nanjing University, China
- [9] Multi-Scale Salient Object Detection with Pyramid Spatial Pooling Jing Zhang, Yuchao Dai, Fatih Porikli and Mingyi He School of Electronics and Information, Northwestern Polytechnical University, China. E-mail: zjnwpu@gmail.com; myhe@nwpu.edu.cn Research School of Engineering, Australian National University,
- [10] What is a Salient Object? A Dataset and a Baseline Model for Salient Object Detection Ali Borji, Member, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 2, FEBRUARY 2015
- [11] Review of Visual Saliency Detection with Comprehensive Information Runmin Cong, Jianjun Lei, Senior Member, IEEE, Huazhu Fu, Ming-Ming Cheng, Weisi Lin, Fellow, IEEE, and Qingming Huang, Fellow, IEEE 2018.
- [12] B. Lei, E.-L. Tan, S. Chen, D. Ni, and T. Wang, "Saliency-driven image classification method based on histogram mining and image score," *Pattern Recognit.*, vol. 48, no. 8, pp. 2567–2580, 2015.
- [13] A SUPPORT VECTOR MACHINE APPROACH FOR OBJECT BASED IMAGE ANALYSIS Angelos Tzotsos\* Laboratory of Remote Sensing, Department of Surveying, School of Rural and Surveying Engineering, National Technical University of Athens, Greece – [tzotsos@gmail.com](mailto:tzotsos@gmail.com)
- [14] Fast SIFT scene matching algorithm based on saliency detection and frequency segmentation for downward-viewing images ZHAO Dan-pei Image processing center School of Astronautics Beihang University Beijing, WANG Jia-jia, China [zhodanpei@buaa.edu.cn](mailto:zhodanpei@buaa.edu.cn)
- [15] Convolutional Neural Networks for Image Processing: An Application in Robot Vision Matthew Browne and Saeed Shiry Ghidary GMD-Japan Research Laboratory, Collaboration Center 2-1 Hibikino, Wakamatsu-ku, Kitakyushu-city
- [16] Learning RGB-D Salient Object Detection using background enclosure, depth contrast, and top-down features Riku Shigematsu David Feng Australian National University [riku.research@gmail.com](mailto:riku.research@gmail.com) Australian National University