A Review Paper on Salient Object Detection using Edge Preservation and Multi-Scale Contextual Neural Network

[1]Pooja P. Ingole, [2]S.C. Nandedkar

[1]Student, [2]Assistant Professor
[1]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. 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 achieve through best and high leveled state-of-the-art performance optimally.**

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

1. Introduction

Edge Preserving and Multi-Scale Contextual Neural Network for Salient Object Detection with many applications. 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 etc. Our topic targets the preserving and detecting and enhancements of techniques for the best result from the aspects that we are using. 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. The proposed framework in this is aiming to address two different but yet dependent limits of the existing CNN based methods. Furthermore, our method can be generally applied to RGB-D saliency detection by depth refinement. The proposed framework architecture achieves both clear detection boundary and multi-scale contextual robustness simultaneously for the first time, and thus achieves an optimized performance. It mainly depends on many stages: feature extraction ,these features are often represented in different form such as novel approaches like motion based method and orthogonal, ,geometric features like active appearance model , video metrics, classifier design, support vector machine is mostly used. 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. For this we targeting to address two limits of the existing CNN based methods.1) Region-based CNN methods lack of sufficient enough over the context to accurately locate the saliency of object since they deal with each region independently,wisely.2) Pixel-based CNN methods suffer or difficult from blurry boundaries due to the presence of convolutional and pooling layers. 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. Later, to further improve it, multi-scale spatial context is attached to *RegionNet* to consider the relationship between regions and the global scenes. Where R-CNN stands for R- Regions within an image, CNN- Convolutionary neural nets which having context as data. 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

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 un-supervised by many things which we are future sees in this paper.

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. Deep Saliency as we go inner side Multi-Tasking done deeply Neural Network Model for Salient Object Detection go deeper and deeper [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. First method 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.

Contextual neural network- 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. The proposed technique uses elementary classifiers on the basis of data to obtain a coarse segmentation per-pixel by pixel region-based methods. Contextual information is exploited in a smooth the segmentation between frames. In general the framework significantly, enhances segmentation from the classifiers alone. Context is being used increasingly in computer vision techniques to help perform scene recognition, region categorization and object detection along with the image processing two important methods such as Region-based methods, pixel-based methods.

As the new things comes on existence new methods invents so in paper next methods are RGB-D Salient Object Detection and Multi-Scale Context the paper [8] consist RGB-D saliency as an emerging topic and mostly their methods are based on fusing depth priors. Propose RGB-D saliency method invented by Ju *et al.*, in which saliency is measured as how much it outstands from surroundings. 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. In below table the comparison of key factors on Salient object detection and their methods are shown below---

Table 1: Comparision Table

Name of method	*Performance*	*Disadvantage*
Region-based CNN methods	Each region independency	Lack sufficient context to accurately locate salient object
Pixel-based CNN methods	End to end pooling layers	suffer from blurry boundaries
Local method for Edge recognition-preservation	Better recognition than filter by filter along preserved ones	Color information is not included.
Multiscling-Principal Component Analysis	Multiple scale Recognized in rate is higher	Storage requirement is higher
Salinecy over region	Recognition rate is low	Only single factor can be varied.

Contextual neural networks units	Gives multi-scale level higher and deep recognition rate	Doesn't recognize the full range of scenes regions of image too complex
Local Binary labeling pattern	Recognition and detection of the object rate is with static images	Dynamic images are not included.
Group-wise Registration and coming the saliency map	Performance is good	information lost durng operation

3. System Development

System architecture 3.1 shows overall structure of proposed method. There are main Stages construction and Recognition stage. This architecture shows all sub process that is Feature extraction, testing, classification, detection, updating datasets. Etc. Here, we propose a novel edge preserving and multi-scale contextual network for salient object detection. The proposed framework System architecture shown in below figure 3.1, achieves both clear boundary and multi-scale contextual robustness simultaneously for the first time as manner of illustrated the proposed structure, named *RexNet*, proposed for object detection to achieve superior performance optimal result because of the convolution features of entire image are shared .This features of each patch (or RoI) are extracted via the RoI pooling layer [10]. The architecture is shown stepwise manner and pictorial overall view of the system preservation, detection, so on.

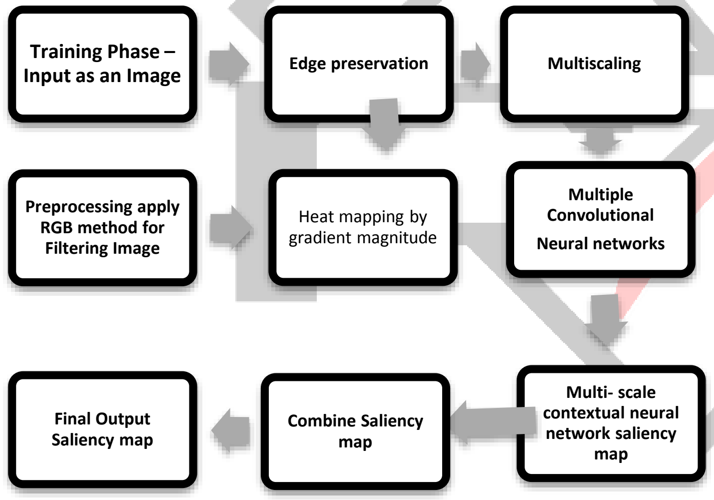

Figure 3.1: System Architecture Design

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-preserved methods, saliency map generated by our network is naturally with sharp boundaries. Second, the *ContextNet* aims to provide strongly reliable on multi-scale contextual information. This is based on the observation that different layers of CNN represent different levels of semantically gather all the considering context of different levels may be more sufficient. We achieve this by taking advantages of dense image prediction. For all max-pooling layers of *RegionNet*, we attach multiple convolutional layers to predict saliency map of different levels. Then all levels of saliency map are fused with *RegionNet* to generate the final saliency map. Our method generates saliency map with accurate location while keeping fine and absolute object boundaries.

3.1 Salient Object Detection by Edge Preserving and Multi-Scale Contextual Neural Network techniques

System architecture 3.1 shows overall structure of proposed method. There are few main Stages of construction, detection and Recognition stage. This architecture shows all sub process that is Feature extraction, testing, classification etc. Here, given a query image as input to the process, estimate the correct expression type, such as image features firstly Edge preservation. That l expression of image sequences or dataset contains not only image appearance information in the spatial domain, but also evolutes the details in the temporal domain. The image appearance information gather in the form of “Scene” .this scene together with evaluates information can be in form of noisy, impurity, blurred, unclear so we have to enhance future to recognition performance. But it will so tough and challegened that the dynamic information provided is useful or not that from image scene, there are challenges regarding how to capture this information reliably and robustly of the “Foreground object”. For instance, an image feature sequence normally constitutes of one or more object having blurry partial substance it’s difficult to handle. In order to capture temporal information and make temporal information of training and query image sequences comparable, correspondences between different phases as segmentation, multi-scaling, contexting neural network we need to be establishment. Finally in recognition stage, to fuse all the dispersed data together evenly by the type determined by comparing through corresponding query image input sequence with each sequence originally formed to get maximum output as detection of saliency.

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. For that we propose an effective, essential network, named *RegionNet*, which generates saliency score of each region end-to-end

1) Network Architecture: In Fast R-CNN 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 However, Fast R-CNN is only used for object detection and classification but not for saliency. The result of Fast R-CNN is bounding box but not pixel-wise. In this paper, we make the modification to enable edge preserving saliency by introducing mask-based RoI pooling. Different from previous region-based methods which deal with each region of an image independently, our proposed Fast R-CNN 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 Fast R-CNN 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 $H \times W$ (7×7 in our work).thats 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

$$P_j = \begin{cases} \{(k | k \in SW_j^{max}, M_k = i)_k^F & i \in M(SW_j), \\ 0 & i \pounds M(SW_j) \end{cases}, \quad \text{……eq (I)}$$

Where, region of that masked image having values with index i , and a certain sub-window as SWj, F *for* features before pooling, region mask as M, , the pooled feature at sub-window SWj as Pj, are the value that we denote vicesly in the equation. The superpixel is segmented using SLIC algorithm may be belongs here.

Salient object detection is a class-agnostic task, where in which image stated as whether a region is salient or not is largely depend on its surroundings, *i.e.*, context. In Our proposed network is shown in Based on the *RegionNet*, we propose to use multi-scale dense image prediction method is very complicated complexity is higher, to get model the relationship between regions and the global scenes at multiple levels to predict saliency maps of different levels.

In the context the first three layers of each branch are with 3×3 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.

$$S = Fusion(S_{S,} S_{E.} S_C) \text{ …..eq (II)}$$

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, $D = \{(Xi , Ti)\}Ni=1$,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 -

$$L_c = -\frac{1}{N}\sum_{i=1}^{N}\left[T_i \log\left(f(X_i;\theta)\right) + (1-T_i)\log\left(1 - f\left(X_{i;}\theta\right)\right)\right] \quad \ldots \quad \text{eq (III)}$$

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-preserved 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 $\bar{f}(X;\theta)$. The Edge Loss is defined as the *L*2 norm between saliency map $f(X;\theta)$ and the averaged map $\bar{f}(X;\theta)$.

$$L_\epsilon = \frac{1}{2N}\sum_{i=1}^{N}\left\|f(X_i;\theta) - \bar{f}(X_i;\theta)\right\|_2^2 \quad \ldots\ldots \textit{eq (IV)}$$

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 given.

4. Conclusion

Our framework of Salient object detection technique is obtained. Here we have compared many existing system and methods to get optimal result which is more efficient and very useful. The survey exhibits higher recognition rate gives higher performance. This technique gives high quality performance than other compared method.

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] 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.

[9] 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.

[10] Zhen Zuo1, Bing Shuai1, Gang Wang1,2, Xiao Liu1, Xingxin'g Wang1, Bing Wang1, Yushi Chen3 , "Convolutional Recurrent Neural Networks: Learning Spatial Dependencies for Image Representation "1Nanyang Technological University, Singapore 2Advanced Digital Sciences Center, Singapore 3Harbin Institute of Technology, China.

[11] Focusness Guided Salient Object Detection Xiaolin Xiao, Yicong Zhou Department of Computer and Information Science University of Macau, Macau 999078, China 2017