Thesis No: CSER-M-20-02

A Study on the Effect of Seam, Color and Boundary Priors on Salient Region Detection

by

Aminul Islam

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

December, 2019

A Study on the Effect of Seam, Color and Boundary Priors on Salient Region Detection

by

Aminul Islam

Roll No: 1507551

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

Khulna University of Engineering & Technology
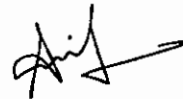
Khulna 9203, Bangladesh

December, 2019

Declaration

This is to certify that the thesis work entitled **A Study on the Effect of Seam, Color and Boundary Priors on Salient Region Detection** has been carried out by **Aminul Islam** in the Department of Computer Science and Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh. The above thesis work or any part of this work has not been submitted anywhere for the award of any degree or diploma.
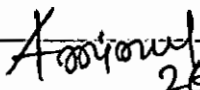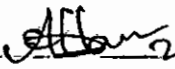
Signature of Supervisor

Signature of Candidate

Approval

This is to certify that the thesis work submitted by Aminul Islam, entitled "A Study on the Effect of Seam. Color and Boundary Priors on Salient Region Detection" has been approved by the board of examiners for the partial fulfilment of the requirements for the degree of "Master of Science" in the Department of Computer Science and Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh in December 2019.

BOARD OF EXAMINERS

1. _____   Chairman
   Dr. Sk. Md. Masudul Ahsan   (Supervisor)
   Professor
   Department of Computer Science and Engineering
   Khulna University of Engineering & Technology

2. _____   Member
   Head
   Department of Computer Science and Engineering
   Khulna University of Engineering & Technology

3. _____   Member
   Dr. K. M. Azharul Hasan
   Professor
   Department of Computer Science and Engineering
   Khulna University of Engineering & Technology

4. _____   Member
   Dr. Pintu Chandra Shill
   Professor
   Department of Computer Science and Engineering
   Khulna University of Engineering & Technology

5. _____   Member
   Dr. Kamrul Hasan Talukder   (External)
   Professor
   Computer Science and Engineering Discipline
   Khulna University

Acknowledgement

At first, I would like to thank almighty Allah for giving all his blessings on me to complete my task. It is my immense pleasure to express my deepest gratitude to my supervisor Dr. Sk. Md. Masudul Ahsan, Professor, Department of Computer Science and Engineering (CSE), Khulna University of Engineering & Technology (KUET) for his continuous encouragement, constant guidance and keen supervision throughout of this study. He helped me a lot at every stage of this work. I had a great fear of research works in the beginning. At times, I started thinking it may not be possible by me. But now, because of his inspiring guidelines, research has become easier for me and I am a much braver fighter to take any challenge in the field. His valuable comments and timely support have enriched this work. I will remember his inspiring guidelines in my future in shaa Allah.

I am also remembering with gratefulness Prof. Dr. Syed Ali Ashraf (pbuh), whom I had never met, but I am a direct beneficiary of his life long work. His quality and immortal inspiration always pushes me ahead in pursuing knowledge. I would like to convey my heartily ovation to all the faculty members, officials and staffs of the Department of Computer Science and Engineering as they have always extended their co-operation to complete this work. I am extremely indebted to the members of my examination committee for their constructive comments on this manuscript. Last but not least, I am grateful to my parents, brothers, sisters, all my teachers, friends and specially my wife and son for their patience, support and continuous encouragement during this period.

Abstract

Being the best creation of almighty Allah, human beings have the ability to analyse a visual scenario. They not only see an image, but can judge the importance of different parts of a visual area. They can easily differentiate a running car from its background, can tell the color of different objects in an image and can focus attention to some important parts of a visible scene. The value of this visual power of human is easily understood if one thinks about these. Achieving this wonderful human quality using machine is one of the most precious goals of today's scientific researches. With the continuous advancement of imaging technologies, more and more visual data are being collected all over the world. But, a major portion of these data are left unprocessed. Image processing is used to process these types of visual data. For all image processing techniques, the initial goal is to find some target region for extracting information from the image. Saliency detection is the technique of computationally finding important regions of an image. It is usually done using the contrast information present in an image. Seam map is the combination of cumulative summation of energy values from different directions. In this thesis, a combined method is proposed which uses seam importance map along with boundary aware color importance map. Color importance map is the weighted average of different color channels of Lab color space. Some intermediate combinations which are closer to the proposed optimized version but differ in the optimization technique are also presented in this thesis. Several standard benchmark datasets including the famous MSRA 10k and ECSSD datasets are used to evaluate performance of the suggested method. The proposed saliency model has been compared with several state of the art methods for each dataset. The qualitative and quantitative results from those comparisons make things easier to understand. Besides that, comparison with those state of the art techniques and precision recall curves and F-beta values found from the experiments on several datasets prove the superiority of the proposed method. This robust saliency model can improve almost all types of vision related applications including object detection, robotics, medical image analysis, content aware image resizing etc. In this research, the proposed architecture of saliency detection technique has been applied into a simple implementation of pedestrian detection. Its application has significantly improved the result of pedestrian detection in terms of performance and accuracy.

Contents

List of Figures

List of Tables

List of Abbreviations

BARC   Boundary Aware Regional Contrast Based Visual Saliency Detection

BARCS   Boundary Aware Regional Contrast Based Seam-map

BACS   Boundary Aware Color-map and Seam-map

HC   Histogram based Contrast Method

RC   Region based Contrast

MSS   Maximum Symmetric Surround

GB   Graph-based Visual Saliency

SWD   Spatially Weighted Dissimilarity

IT   A model of saliency-based visual attention for rapid scene analysis by Itti et al.

FT   Frequency-tuned Salient Region Detection

CA   Context-aware Saliency Detection,

CB   Context Based Saliency Detection

AIM   Attention by Information Maximization

SLIC   Simple Linear Iterative Clustering

CIE Lab   A Color space Defined by Commission Internationale de l'Eclairage

RGB   Red, Green, Blue based color space

ID   Identity

HOG   Histogram of Oriented Gradients

MSRA   Micrsoft Research Asia

CSSD   Complex Scene Saliency Dataset

ECSSD   Extended Complex Scene Saliency Dataset

DDR   Double Data Rate SDRAM

RAM   Random Access Memory

PR   Precision Recall

GT   Ground Truth

INRIA   National Research Institute for Digital Science and Technology of France

List of Symbols

d_c euclidian color distance

σ constant to control intensity strength

B_i dissimilarity with border regions

b_i i^{th} border regions

d_s Euclidean region distance

d_{lab} Lab distance

D_s Sum of distance

C_k k^{th} center

r_i i^{th} region

F_β F-beta measure

Dedicated to Prof. Dr. Syed Ali Ashraf (pbuh) whose work on "Islamization of Education" has great influence in my life

Chapter I

Introduction

1.1 Background

From the beginning, humans are gifted with the ability to judge the importance of visual scenario. They can focus attention on selected parts of a view area. Finding those salient image regions computationally is a desirable goal, since it allows efficient allocation of computational resources in many image analysis and synthesis techniques.

Saliency is usually detected by variations in image properties like gradient, color, edges and boundaries. It originates from visual uniqueness, unpredictability, rarity or surprise. Based on how people perceive and process visual stimuli, it is investigated by multiple disciplines including neurobiology [1] [2] [3] [4], cognitive psychology [5] [6] [7] and computer vision [8] [9] [10]. According to theories of human attention, the human vision system only processes parts of an image in detail, leaving others nearly unprocessed.

Two stages of visual saliency are proposed by early saliency detection studies [11], [12], [12] - one is fast, pre-attentive, bottom-up, data driven saliency extraction and another is slower, task dependent, top-down, goal driven saliency extraction. This study follows those two stages by extracting the saliency using high level cues and then by combining them using the goal based optimization technique. It results in a successful efficient method for saliency detection.

In saliency detection researches, most of the methods rely on different priors about objects and their background properties. Contrast and background priors are the most widely used cues among those. Contrast prior suggests that the appearance contrast between objects and their surrounding regions are high. This presumption is used almost everywhere [13], [14], [15], [16], [17], [18], [19].

Another prominent cue is boundary prior which assumes that regions at the boundary are mostly backgrounds [17]. Many recent papers [20], [21], [13], [17], [22], [14] proves its

effectiveness. Ahsan et. al. has used contrast map and spatial distance information of salient regions along with this distinctive cue in [13] to produce state of the art result.

Seam map is an interesting cue which is used to extract the salient region from its backgrounds. It is an optimal 8-connected path of pixels in an input image. Optimality in a seam map is defined by an energy function from top to bottom, left to right, bottom to top and right to left. Seam carving is used in [23] for content aware image resizing. The authors also have suggested that seam map can be used for image enhancement tasks. The work in [24] has successfully proved that seam map can be used to detect salient regions more accurately by foreground enhancement and background exclusion. In this study, these ideas are applied into [13] to examine the effect of seam map on it. Examination of the combination of seam map and BARC shows promising performance. Moreover, the region based seam importance map produces better detection results in terms of time and performance.

Color-map is another interesting prior which improves saliency detection performance. It is the weighted average of different color channels of an image. Joy and Hossan has presented boundary and color importance map to detect salient region from a digital image [25]. In this study, boundary aware weighted color-map is used in combination with weighted seam-map to produce the final saliency map.

1.2 Motivation

Saliency detection is a prominent field of research in the present world. It is the quality of an image region to capture human attention which makes an object or group of pixels stand out from the neighbourhood. The main target of these type of researches is to extract the salient region of a digital image. Contrast and boundary priors are used preferably to extract salient regions of an input image. It can enhance the performance of various computer vision applications including content aware image resizing [26] (Fig. 1.1(a)), image cropping [27] (Fig. 1.1(b)), object segmentation [28] (Fig. 1.1(c)), object detection [29] (Fig. 1.1(d)) etc. Very few researchers have tried this interesting methodology in application level. So, it is expected to find state of the art result by exploiting seam map with boundary prior in visual saliency detection. Moreover, applying this technique to some application level also presents some interesting output.

(a)   **(b)**

(c)   **(d)**
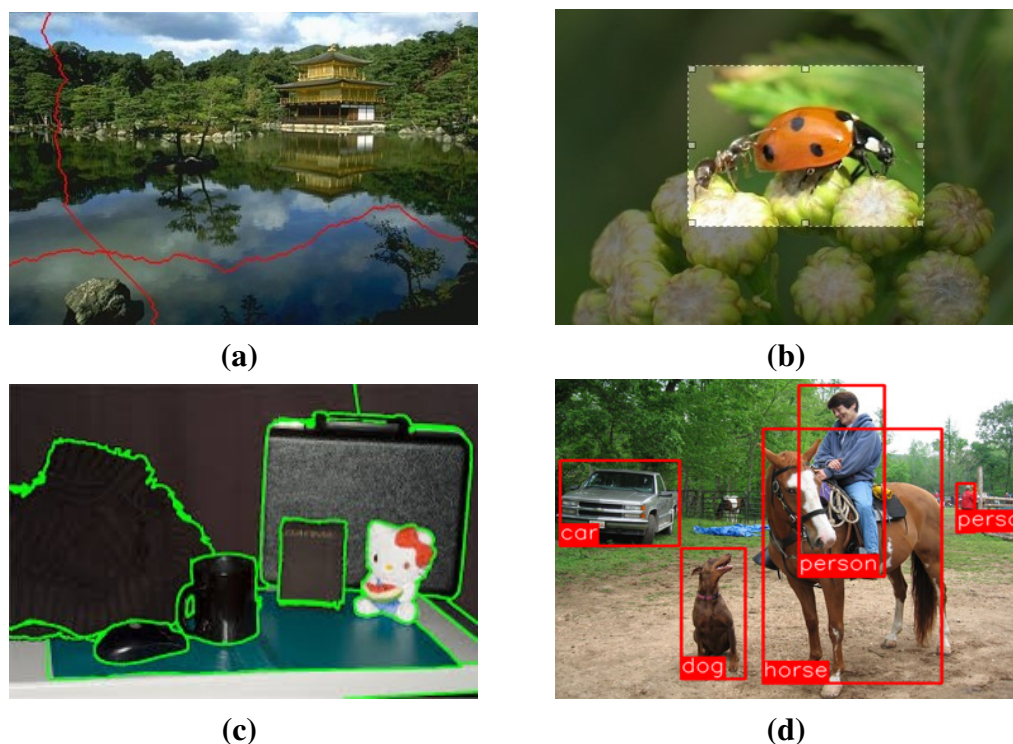
Figure 1.1: Possible applications of saliency detection.

1.3 Problem Statement

Computer vision techniques are usually complex and time consuming, since every pixel of the input image needs to be processed through them. Saliency detection is the technique of finding the salient or important region in an image. So, researches in this field have become popular to reduce this type of computational complexities and shrink the search area of an input image in various vision related applications including object detection, action recognition, image segmentation, video surveillance, medical data processing, robotics, car vision and many more.

Color contrast present between the salient region and the background is a popular and successful cue in saliency detection solutions. But, this cue is challenged when the salient region has less color dissimilarity with the background. It also fails to highlight some part of the salient region when part of the salient region shares similar properties of some background pixels.

Boundary prior which assumes that the boundary regions are mostly backgrounds is another prominent cue in saliency studies. But, this also fails when the salient regions are closer to

| (a) | (b) |

Figure 1.2: Objective of saliency detection technique.

the border or salient regions are part of the boundary regions.

There are saliency measurement methods which combines two or more high level cues that may improve detection results. But, most of them are not target oriented. The target is to match the result with the ground truth image which is a binary image that clearly separates foreground from background. It is shown in this study that better integration of multiple cues is possible by using some optimization technique.

In those complex conditions, relying only on contrast and boundary priors may not produce good result. So, a new method is proposed which does not rely only on contrast based dissimilarity but combines some other high level cues like seam and color priors. This novel method offers better detection and efficiency in terms of quantitative and qualitative analysis. Experimental results with popular standard datasets and comparison with some state of the art techniques proves the superiority of the proposed system.

1.4 Aims and Objectives

The main objective of this research is to devise a novel idea that can detect salient region efficiently as seen in Fig. 1.2 from digital scenery using computer vision technique. Starting from an input image like in Fig. 1.2(a), the target is to find the salient region shown in Fig. 1.2(b). It is useful in the preprocessing stage of automatic car vision, robotics, image analysis, traffic control, video surveillance and other related fields. The previous researches have shown outstanding performance in the field, but there exists space for improvements. The aims and specific objectives of this thesis which are discussed in the forthcoming chapters are to –

- Propose a new combination of low level saliency cues.

- Combine seam importance map with the contrast map of [13].

- Examine the effects of boundary and color map on the new model.

- Optimize the combination using the optimization procedure presented in [30].

- Show application of the proposed saliency detection model to detect salient objects more efficiently.

- Use multiple standard datasets to evaluate the method.

- Compare the results with other state of the art techniques.

- Explore the effectiveness of the proposed saliency model by implementing in a simple pedestrian detection application.

1.5 Organization of this Thesis

The thesis has been divided into five chapters.

Chapter-I contains introduction followed by the aims and objectives of the present work, motivation for choosing this research work, application of saliency detection and organization of the thesis.

Important previous works on this research topic and their historical backgrounds has been discussed in chapter-II.

The proposed method is discussed in details in Chapter-III. The superpixel generation procedure, seam importance map, border contrast map, color importance map and the optimization technique are described in details.

In Chapter IV, the detailed results regarding the proposed method are compared with different state of the art methods and discussed.

And Chapter V contains conclusion, achievement of works and further suggestion of this work.

Finally, a complete list of references has been given towards the end of the chapters.

Chapter II

Literature Review

2.1 Introduction

There is usually a visible contrast between the salient object and its background. So, many researchers mostly relied on contrast prior. Moreover, boundary prior offered promising results in most of the cases. In this chapter, different approaches of previous researches, their advantages and shortcomings are briefly discussed.

2.2 Basic Terms

Prior

It is "prior information" available in a set of images that can be used in image analysis problems to enhance results, ease the choice of processing parameters, resolve indeterminacies etc. For instance, it may be known before applying any procedure that the image is noisy, should contain only four colors or that pixels follow a specific distribution. These priors or their approximations can be put into math form and can be merged into the processing (filtering, deconvolution, segmentation) and reduce the set of feasible solutions. In this study, different priors are chosen like - contrast prior, boundary prior, seam importance map, color importance map etc. These are also called cues throughout this thesis.

The "prior" may have different meanings based on its use in the literature. For example, the "prior" is a measure over a set of classifiers in "prediction bounds & online learning" that expresses the degree to which it is hoped that the classifier will predict well.

Contrast Prior

"Contrast Prior" is the mostly used prior in saliency detection methods. It assumes that the appearance contrast between the objects and their surroundings are high. It is a very successful cue in saliency detection methods and combination of other cues can result in better saliency map. In this study, this prior is mainly used for smoothing the final output in optimization phase. The detailed process is discussed in section 3.4.

To compute the contrast prior, for a region r_i, the color intensity dissimilarity from r_i to another region r_j is calculated as in eq. 2.1. When the intensity difference of two regions r_i and r_j is too large, the Euclidean color distance d_c^2 becomes large. So, the exponential distribution of eq. 2.1 is closer to 1. On the other hand, when the difference is too small, value of d_c^2 reaches to 0 and the output is closer to 0. Similarly, this exponential function outputs value in the range [0,1] for all regions. This constitutes the color distribution matrix. Since, object usually covers relatively smaller region than the background. So, the contrast map is found by averaging the color distributions of each region to all other regions using eq. 2.2. Because of another common assumption - most of the image pixels constitutes the background, it highlights the foreground of the image.
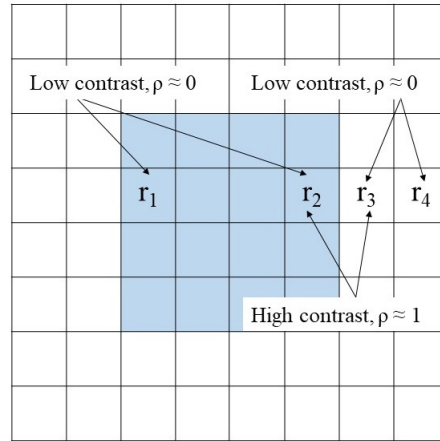
Figure 2.1: Example of contrast map generation process.

Figure 2.2: Example color distribution matrix and average calculation.
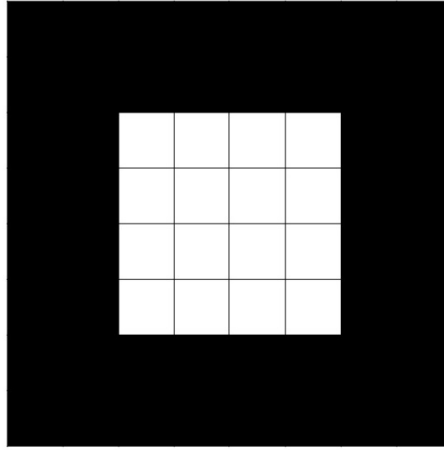
Figure 2.3: Example binary output for the input image of Fig. 2.1.

A 64 regions superpixel image is considered for example as in Fig. 2.1. There is the salient region in the middle with 16 regions. As seen in Fig. 2.1, value of $\rho(r_i, r_j) \approx 1$ for similar region and $\rho(r_i, r_j) \approx 0$ in contrast regions. For this simple example, the color distribution matrix and the average calculation are shown in Fig 2.2. So, as per the calculation in Fig. 2.2, foreground regions would have value that approximately equals to 0.75 and the background will have value approximately equals to 0.25. After applying normalization those values equal to 1 and 0 respectively. Thus, a binary image clearly separating the foreground and background of the example image is found as shown in Fig. 2.3.

$$\rho(r_i, r_j) = 1 - exp\left(\frac{-d_c^2(r_i, r_j)}{\sigma^2}\right)$$
$$d_c(r_i, r_j) = \sqrt{(c_i - c_j)^2}$$

$$c_i, c_j = \text{Mean color vector of } r_i, r_j$$

$$\sigma = \text{constant to control strength}$$

(2.1)

$$S_G(r_i) = \frac{1}{N}\sum_{k=1}^{N} \rho(r_i, r_j)$$

$$N = \text{Number of regions}$$

(2.2)

(a)   (b)
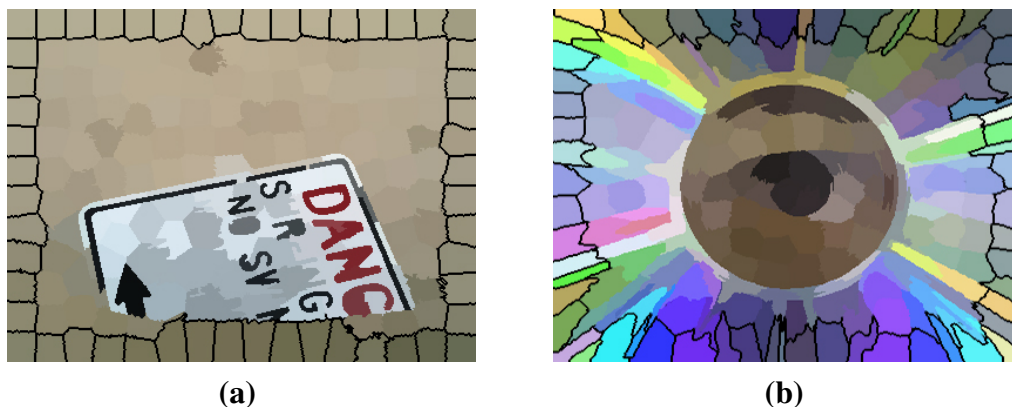
Figure 2.4: Examples of simple and complex border regions.

Boundary Prior

Boundary Prior is another famous and effective prior used in various saliency related applications. It assumes that the boundary regions are mostly backgrounds. It works on the idea of contrast prior. Instead of measuring color distance of all superpixel regions, it only considers measuring the dissimilarity with the border regions. Since, it assumes border regions to be part of the background. Although, it is very effective in practical application, it has some drawbacks i.e., treating all image boundary as backgrounds and lack of ability to work independently without being combined with another saliency measure. In addition to that, if some part of the object is touching the boundary, it will fail to detect the salient region. In this thesis, an intelligent way is used by choosing only the first 30% of the sorted values to reduce this problem. Moreover, it is combined with other low level saliency cues to minimize this behaviour.

This can be easily understood by examining the simple examples and their surroundings in Fig. 2.4. In that figure, the backgrounds of both 2.4(a) and 2.4(b) have similar colors like the border regions. Although, 2.4(a) is simple and have a background similar to the border regions and 2.4(b)'s boundary regions have several color intensities but all of them constructs the boundary and significantly differ from the salient object or region.

Seam Importance Map

Seam is an optimal 8-connected path of pixels. It is the cumulative sum of minimum energy values from all four directions of an image namely top to bottom, left to right, bottom to top and right to left. It is first introduced by Avidan et al. [23] for content-aware image retargeting. The main target of their proposed method is preserving important content by

(a)   (b)
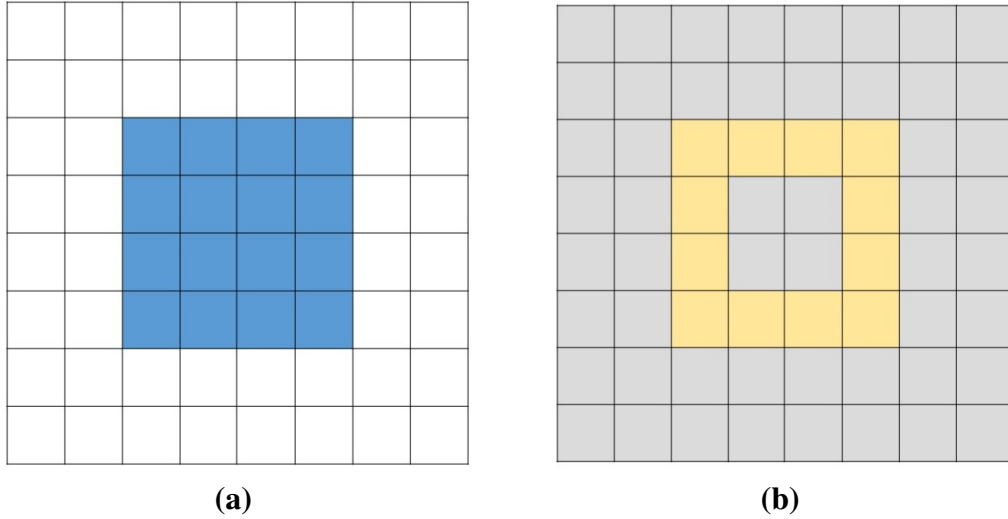
Figure 2.5: Energy map of of an example 8x8 simple input image.

minimizing the distortion during the retargeting process. In this method, seam carving it-
eratively reduces the input image size by removing seams with least horizontal or vertical
energy according to an energy map to obtain the target image size. Gradient information and
dynamic programming are used to compute the optimal seams. In this study, the seam from
all direction of an input image is calculated and combined in a way to enumerate the salient
region. The energy function is defined by Sobel operator to get the gradient image which
minimizes the cost. The procedure has been discussed in details in section 3.3.

$$\text{Energy Image}, e(I) = \left|\frac{\delta}{\delta x}I\right| + \left|\frac{\delta}{\delta y}I\right| \tag{2.3}$$

$$M_T(x,y) = \begin{cases} e(x,y); & \text{if } y = 0 \\ e(x,y) + min[M_T(x-1,y-1), \\ M_T(x-y,y), M_T(x-1,y+1)]; & \text{otherwise} \end{cases} \tag{2.4}$$

$$Imp_{seam}(x,y) = min[M_L(x,y), M_T(x,y), M_R(x,y), M_B(x,y)] \tag{2.5}$$
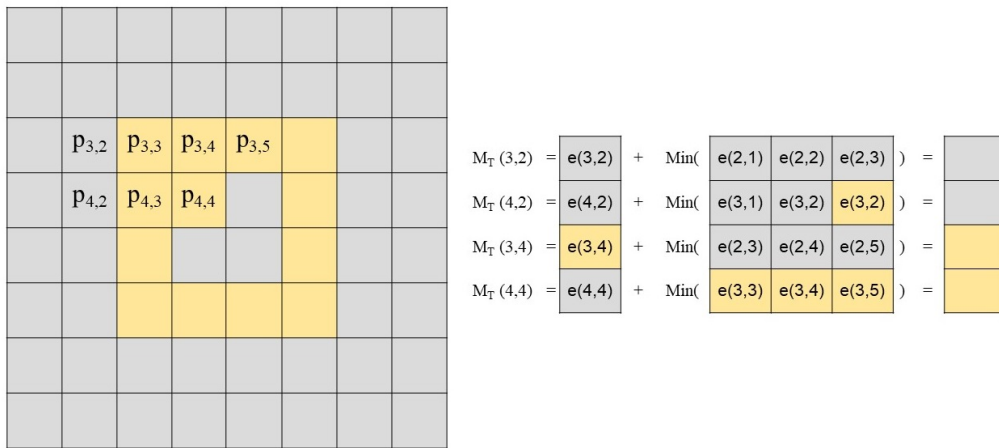
Figure 2.6: Example of top seam calculation (Best viewed in color).

Let us consider a simple 8×8 input image of Fig. 2.5(a) for understanding seam map generation procedure. Each small square box represents a pixel in that simple input image. Fig. 2.5(b) shows the energy map of this input image found using eq. 2.3 where the yellowish border represents the edge of the blue salient square object of Fig. 2.5(a). Fig. 2.6 shows pictorial representation of top seam calculation of some pixels according to eq. 2.4 for understanding the procedure. After completion of top seam calculation, the resulting output would look like Fig. 2.7(a). Similarly, left seam of Fig. 2.7(b), bottom seam of 2.7(c) and right seam of 2.7(d) are computed. Finally, the combined seam map of Fig. 2.7(e) is found by eq. 2.5.

Color Importance Map

Commonly, object and background contain different colors in a digital image. So, this is an important prior in case of saliency detection. In this study, it is shown that this color information can be used in a very simple way by calculating some weighted average of different color channels of Lab color space and combining them in an intelligent way. The resulting color map has significant effect in the final saliency output.

In the simple example image of Fig. 2.8 the butterfly is the only salient object in its bluish background. If the average intensity of the color channels of this simple image is considered with respect to the global space, it will be much greater than the average intensities of the salient butterfly. This prior information can be used intelligently as an important cue for saliency detection which is discussed elaborately in section 3.5.
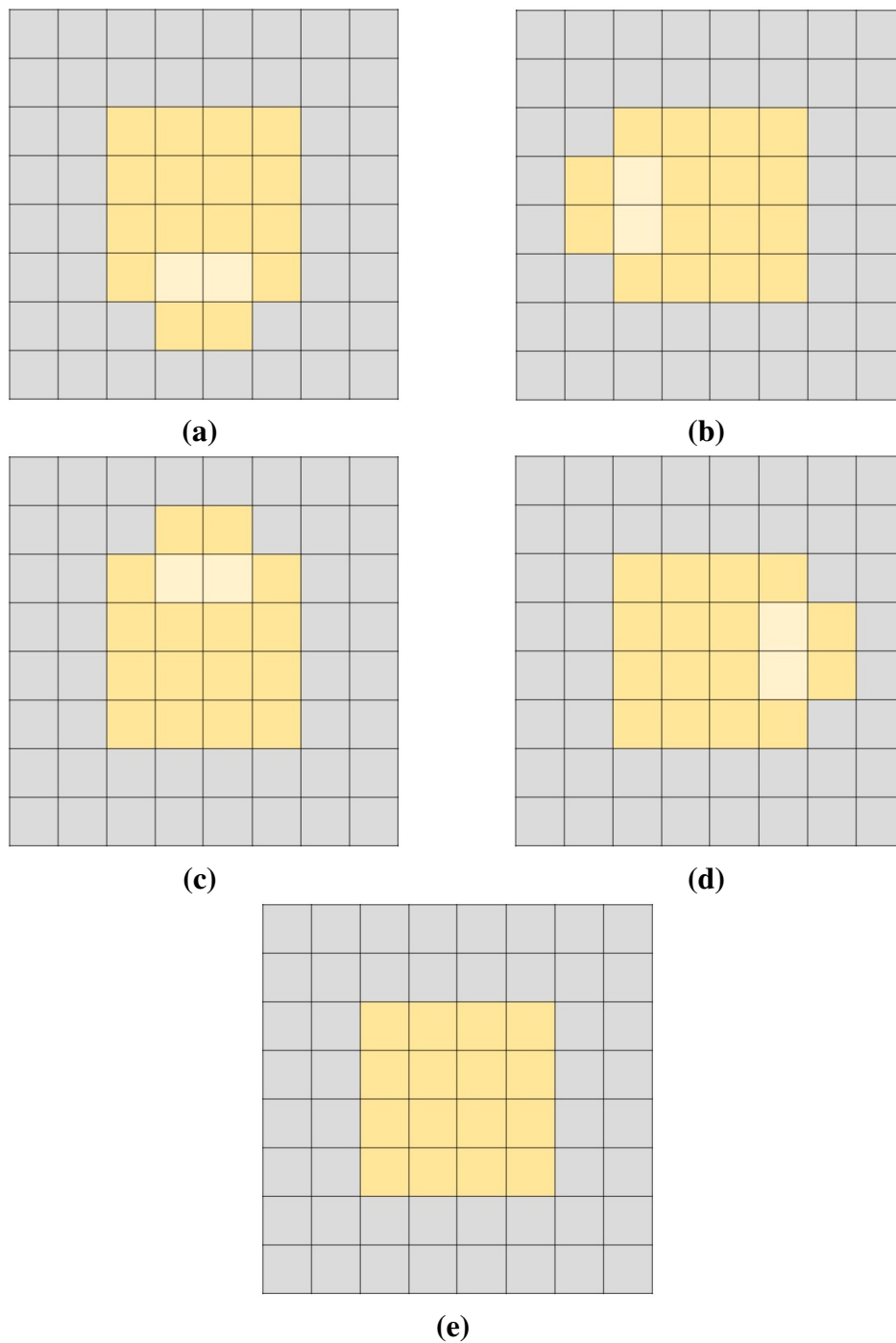
Figure 2.7: Example seam map from four directions and their combination.

Superpixel

In recent years, superpixels are becoming useful in many Computer Vision and Image processing applications like Image Segmentation, Object Detection and Tracking, Semantic Labelling etc. It is a group of pixels that share common characteristics i.e., pixel intensity, color, pattern, grey levels etc.

Figure 2.8: Simple example to understand color importance map.

In other words, it is finding groups of pixels that are most similar to each other and labelling those as being of the same type. It facilitates working with image regions instead of working with every pixel in an image and thus significantly reduces computational time and complexities.

Xiaofeng Ren and Jitendra Malik [31] introduced the concept of superpixel in 2003. They used superpixels instead of pixels to do image segmentation. They mainly proposed a novel approach of image segmentation by dividing an image into hundreds of non-overlapping superpixels. They have perceptual meaning because pixels belonging to a given superpixel share similar visual properties which carry more information than pixels. They produce a compact and convenient form of images that can significantly enhance the performance of computationally demanding problems.

There are two major advantages for using superpixels.

- More meaningful features can be computed on regions.

- Input entities can be reduced for many algorithms.

Superpixel segmentation have been successfully used in many computer vision technologies including image classification, semantic segmentation, visual tracking and so on.

SLIC Algorithm

SLIC algorithm [32] is used in this study to generate superpixelized version of the input image. It calculates superpixels by selecting pixels based on common characteristics i.e., pixel intensity, color, pattern, grey levels etc. in the image plane. It takes an approximate number of equally-sized superpixels, K as input. So, the approximate size of each superpixel

is therefore N/K pixels for an image with N pixels. So, there would be a superpixel center at every grid interval $S = \sqrt{N/K}$ for roughly equally sized superpixels. Based on the value of K superpixel size will differ for an input image as shown in Fig. 2.9.

The algorithm is performed in the five-dimensional *labxy* space, where Lab is the pixel color vector in CIELAB color space and xy is the pixel position. The spatial distances are normalized in order to use the Euclidean distance in this 5D space. It is done this way because the spatial distance in the xy plane depends on the image size and the maximum possible distance is limited between two colors in the CIELAB color space. Therefore, In order to separate pixels in this 5D space, a new distance measure was introduced that considers superpixel size. It is described below.

Distance Measure: At the beginning of the algorithm, K superpixel cluster centers are considered as $C_k = [l_k, a_k, b_k, x_k, y_k]$ with $k = [1, K]$ at regular grid intervals S. Since, the approximate area of any super-pixel is approximately S^2, it is safely assumed that pixels that are part of this cluster center lie within a $2S \times 2S$ area on the xy plane around the superpixel center. The normalized distance measure D_s to be used in the 5D space is defined as :

$$
\begin{aligned}
d_{lab} &= \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \\
d_{xy} &= \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \\
D_s &= d_{lab} + \frac{m}{S} d_{xy}
\end{aligned}
\tag{2.6}
$$

D_s is the sum of the lab distance d_{lab} and the xy plane distance d_{xy} which is normalized by grid interval S. Here, m is introduced as a variable in D_s which controls the compactness of a superpixel. Greater value of m produces more compact cluster. The variable m can be in the range [1,20].

The Algorithm: This algorithm begins by sampling K regularly spaced cluster centers. It moves them to seed locations corresponding to the lowest gradient position in a 3×3 neighborhood. The chances of choosing a noisy pixel and placing them at an edge is reduced by doing that. Image gradients are computed by using eq. 2.7.
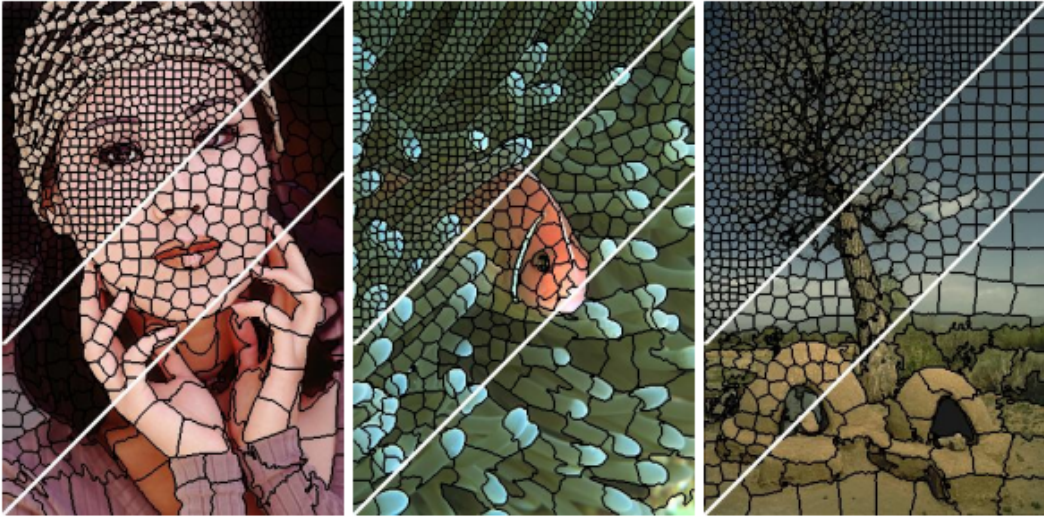
Figure 2.9: Superpixel generation examples from [33] with (approximate) size 64, 256 and 1024 pixels.

$$G(x,y) = \|I(x+1,y) - I(x-1,y)\|^2 + \|I(x,y+1) - I(x,y-1)\|^2 \qquad (2.7)$$

Algorithm 1: SLIC Superpixel generation algorithm.

Initialize cluster centers $C_k = [l_k, a_k, b_k, x_k, y_k]$ with $k = [1,K]$ at regular grid intervals S;

Perturb cluster centers in an $n \times n$ neighborhood, to the lowest gradient position;

repeat

> **for** *each cluster center C_k* **do**
>
>> Assign the best matching pixels from $2S \times 2S$ square neighborhood around the
>>
>> cluster center according to the distance measure
>
> **end**
>
> Compute new cluster centers and residual error E L_1 distance between previous
>
> centers and recomputed centers

until $E \leq threshold$;

Enforce connectivity.

Here, I(x,y) is the lab vector corresponding to the pixel at position (x,y), and $\|.\|$ is the L_2 norm. This considers both color and intensity information.

Each pixel is associated with the nearest cluster center whose search area overlaps the corresponding pixel in this image. After all the pixels are associated with the nearest center, the

average labxy vector of all the pixels belonging to the cluster is computed as a new center.

The process is repeated until convergence by associating pixels with the nearest cluster center and recomputing the cluster center. A few stray labels having the same label but not connected to it may remain at the end of this process in the vicinity of a larger segment. Connectivity can be enforced in the last step of the algorithm by relabelling disjoint segments with the labels of the largest neighbouring cluster.

Lab Color Space

The **Lab** color space was first defined by International Commission on Illumination (CIE) in 1976. L channel is used for the lightness from black (0) to white (100), a channel is from green $(-)$ to red $(+)$, and b channel is from blue $(-)$ to yellow $(+)$. It was designed to match the amount of numerical change in these values to visually perceived difference.

The Lab model is device independent with respect to a given white point. It defines colors independently of how they are created or displayed. It is designed to approximate human vision which is not the case of RGB and CMYK color models. RGB or CMYK spaces models the output of physical devices rather than human visual perception. Lab model aspires to perceptual uniformity as its L component closely matches human perception of lightness. It can be used to make accurate color balance corrections by modifying output curves in the a and b components, or by adjusting the lightness contrast using the L component.

Lab model is copyright and licence-free and fully mathematically defined. It is freely usable and integrable in any application and experiments. Because of the mentioned advantages, Lab color space is used in this research for different calculations.

2.3 Related Works

2.3.1 Contrast Based Approaches

A common belief is that human cortical cells preferentially respond to high contrast stimulus in their receptive fields [34]. According to the observations of [35] a global contrast based method is preferred over local contrast based methods. Because, global contrast method separates a large-scale object from its surroundings and the local contrast method only produces high saliency values at or near object edges. It also suggests that global considerations

can uniformly highlight entire objects by enabling assignment of comparable saliency values to similar image regions. Moreover, contrasts to distant regions are less significant and saliency of a region depends mainly on its contrast to the nearby regions. So, they proposed a histogram-based contrast method (HC) to produce full resolution saliency maps.

In [35], a regional contrast based saliency extraction method is presented. They show that, there are local and global contrast based saliency detection methods. Local contrast based methods generally detects higher saliency values near object edges and fails to highlight the entire salient region. On the other hand, they criticized some global contrast based methods for being insufficient to analyse common variations in natural images. Inspired from biological vision they propose a region based contrast method and a histogram-based contrast method including spatial information for saliency detection.

The main weakness of this approach becomes visible if a region has similar color pixel or superpixel like any background superpixel within it or if the salient region's color is closer to the background color. For instance, if some middle superpixel of the example image contains the same white color like the background, this method detects it as background. It will also fail if the salient region covers up more pixels than the number of background pixels in an input image.

2.3.2 Boundary Prior Based Methods

Early saliency based researches emphasized on center prior which assumes that the salient regions are usually at the center of an image. Those methods usually biased the center of the image with higher saliency values. It was not always true and there can be more examples where the saliency is not at the center of an image. So, it was a fragile assumption and researchers are no more interested in that cue.

On the other hand, boundary prior is an opposite assumption to center prior. Unlike its counterpart, it works better for objects which are not strictly at the center of an image and resulted in better detection in many researches. Although, this assumption has some drawbacks too. As it does not consider the condition where an object contains some of the boundary parts.

In [17] the shortest-path distance from an image region to its boundary is used as a cue for saliency detection. It shows that the boundary pixels can easily be connected to the images

background instead of the foreground. It also results in incorrect detection if an object is touching the boundary region.

Ahsan et al. [13] has used contrast based spatial features along with boundary prior to generate their state of the art saliency detection result. They measured region based color dissimilarity of a superpixelized image found by applying SLIC algorithm on an input image and suppressed the background using boundary prior as a cue to it. They also combined the contrast based output with boundary based output and fine-tuned the combined saliency map using Gestalt thresholding technique to generate the final saliency output.

Zhu et al. [30] have taken into consideration the problem of boundary prior and introduced weighted boundary connectivity to successfully solve such types of problem and presented a better boundary prior based implementation. They also have proposed a robust optimization technique to combine multiple saliency cues to achieve cleaner saliency maps. In this study, their optimization technology is adapted to efficiently combine the proposed saliency cues. The result is discussed in the upcoming chapters.

2.3.3   Use of Seam Importance Map

Seam is an optimal 8-connected path of pixels in an image. It is introduced by Avidan et al. in [23] for content aware image resizing and suggested for image enhancement too. Yijun et al. has successfully used seam information in their saliency detection technique [24]. For every pixel, they choose the minimum of its corresponding four seam cost as the final saliency. They mentioned the problem of seam map approach is- pixels faraway from the top boundary gets high seam cost values, since the long distance causes the cumulative effect and sometimes the 8-connected path unavoidably travels across some strong gradient point. They also gave the solution to this problem by exploiting the minimum of four cost values. They avoided selecting the direct multiplication of four maps which causes over-suppression of the salient object. This approach is robust to the image where the salient object is cropped on the boundary, which violates the boundary prior. This way, the problem of boundary prior can be solved by using seam importance map as a cue for saliency detection.

Seam map is also used along with other high level cues like contrast prior and boundary prior in [21] and [20] to successfully detect salient region. In this study, those combinations are

further exploited and used in the excellent target oriented optimization technique to produce the final saliency detector.

2.3.4   Application of Color Importance Map

Joy and Hossan [25] has presented a color map based saliency detection technique. It is the weighted average of different color channels. They combined border map to produce the final saliency output. This color map is used successfully in [20] to generate state of the art result. They merged the color importance cue with boundary prior to produce their saliency detection technique.

It is also used in [20] along with seam and boundary prior with Gestalt thresholding based optimization technique to produce state of the art result. In this study, those cues are combined and further optimized using cost based optimization procedure to generate better performing saliency detection technique.

2.3.5   Other Approaches

Achanta et al. [33] has proposed a saliency detection approach based on maximum symmetric surround. They used the low level features of luminance and color. They narrow down the search area for symmetric surround as they approach to the borders. The disadvantage of their method is that if the salient region touches the border or is cut by the border, it cannot detect it and is treated as the background.

Li et al. [36] proposed saliency detection based on reconstruction errors. They used superpixel based regional model and computed dense and sparse reconstruction errors. They applied Bayes formula to integrate saliency cues based on dense sparse reconstruction errors. The work in [37] considers salient objects as sparse noises and detects salient region by solving a low rank matrix recovery problem. Yang et al. [38] ranks the similarity of image patches via graph-based manifold ranking. Jiang et al. [39] treats salient region selection as a facility location problem and solves it by maximizing sub-modular objective function.

All these models are hand-crafted saliency methods. They are efficient and effective, but cannot handle complex scenarios. In recent years, learning based methods have received more attention from the community which can automatically learn to detect saliency by

training detectors (e.g., deep networks [40] [41] [42] [43], convolutional network [44] [45], random forests [46] [47] etc.) on image data with annotations. In those learning based detectors, deep network based saliency models have shown very competitive performance.

These methods adapt viewpoints and optimization techniques from other problems for saliency estimation. Unlike all the aforementioned methods, the proposed optimization directly integrates low level cues in an intuitive and effective manner.

2.4 Discussion

In the previous sections it is seen that there are biologically inspired methods and computationally optimized methods, local and global contrast based methods. There are also combination of biological and computational models. Harel et al. [48] extracted feature map using Itti's [8] biology based method and used graph based approach to normalize the output. Some of the methods computes entire feature map or some combines multiple maps to produce the final saliency output.

Many methods including [21], [49], [50], [51], [18] have combined multiple saliency maps to produce better optimized version successfully. Some of them may became complex and their implementation and performance in real world situations are still to be tested. In [14], saliency detection problem has been solved using graph based manifold ranking. In [52], a color and spatial contrast based saliency model is presented which uses neoteric background prior by selecting four corners of an image as background. The method was inspired by reverse-management methods and it also applied energy function to further optimize foreground and background.

The main challenges of saliency detection problem are low color dissimilarity of objects with its surrounding regions, size and position of a salient region in an image, computational complexity and usability of the method in real life applications.

Chapter III

Methodology

3.1 Introduction

In the previous chapter it is shown that there are different approaches of saliency detection - some use pixel based computation technique, other use region based approach; some may emphasize on local features, other may highlight some global characteristics. A combination of those approaches along with seam and color cues are proposed in this study with a target of highlighting the salient portion using better optimization procedure. Structure of the proposed method is shown in the flowchart of Fig. 3.1. The raw input image is transformed into superpixel image. Seam and color importance maps are generated from this superpixelized image. These two saliency maps are then combined and optimized using a suitable optimization procedure to generate the final saliency map. There are more internal steps seen in the flow chart. In this chapter the components of this flowchart are discussed in details.
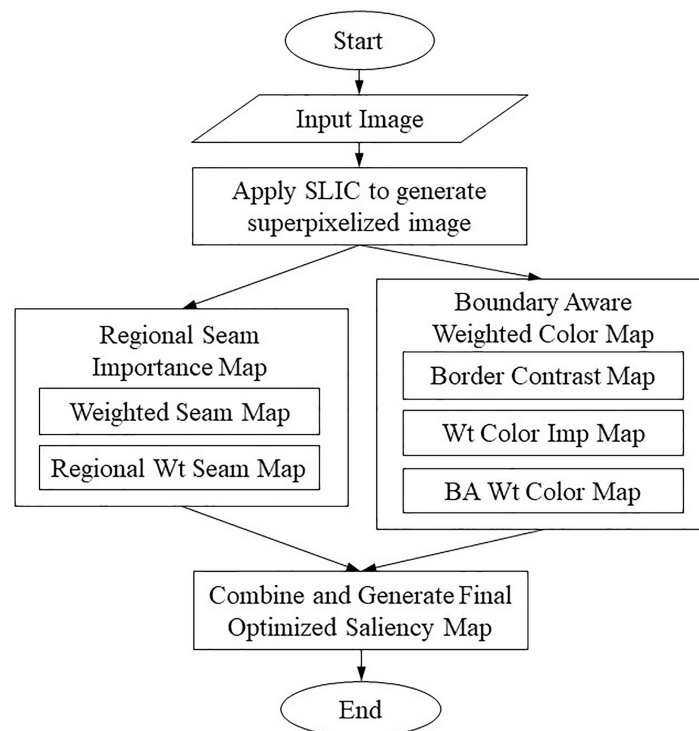
Figure 3.1: Flow chart of the proposed method.

(a) input　　　　　　　**(b)** SLIC Boundary　　　　　**(c)** Superpixelized
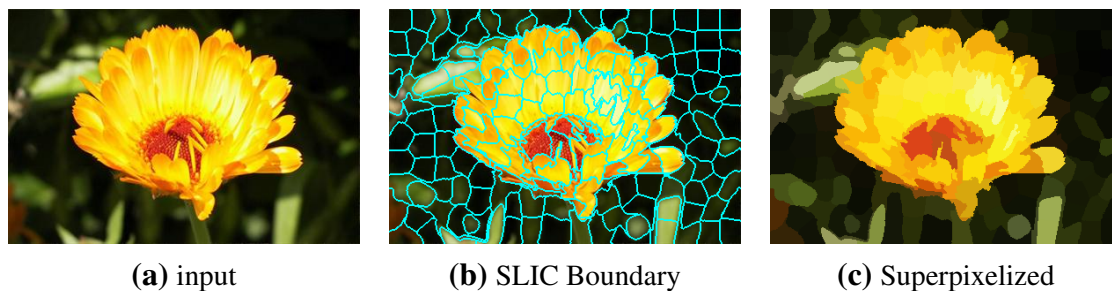
Figure 3.2: From input image to superpixel image.

3.2 Superpixel Image

At first, the input image is converted into superpixels using SLIC method [32]. It groups similar pixels together using some clustering method around randomly chosen pixel centers. Each region contains the averaged pixel intensity value in that region. It is done to speed up the calculation procedure. Since, working in pixel level will require more time to complete the calculation, it assigns a region ID to each pixel and computation can be continued using those regions only. By doing that the calculation becomes faster and quality is not much considered. It can be converted into the full size image anytime. Our main goal is to detect the salient region only. So, the pixel level perfection is not necessary. Fig. 3.2 depicts how an input image is transformed into its superpixel form where 3.2(a) is the input image, 3.2(b) resembles the superpixel boundaries and 3.2(c) shows the processed superpixel image..

3.3 Seam Importance Map

Seam map is calculated from the energy image of an input image. Unlike [23] which computed a seam for the whole image, we follow [24] to get the seam for each pixel of the superpixelized image. The superpixelized input image is taken as the input to the final seam map generation process following [20]. It produces better result as compared to [21] which uses the raw input image in the seam map generation procedure. The reason for this enhancement is probably the smoothness and stronger edges of the superpixelized image. The computation process is as follows: to compute the seam in the direction of top to bottom, each pixel follows eq. 2.4. The minimum cumulative sum of energy values of its previous row's 8-connected neighbours are summed up with the current pixel's energy value according to eq. 2.4 to calculate the current pixel's seam value. Other seam maps are generated in a similar fashion. Since, the background pixels take lower values in any of the four maps as

(a) Energy Image **(b)** Left Seam **(c)** Right Seam

(d) Top Seam **(e)** Bottom Seam **(f)** Combined Seam
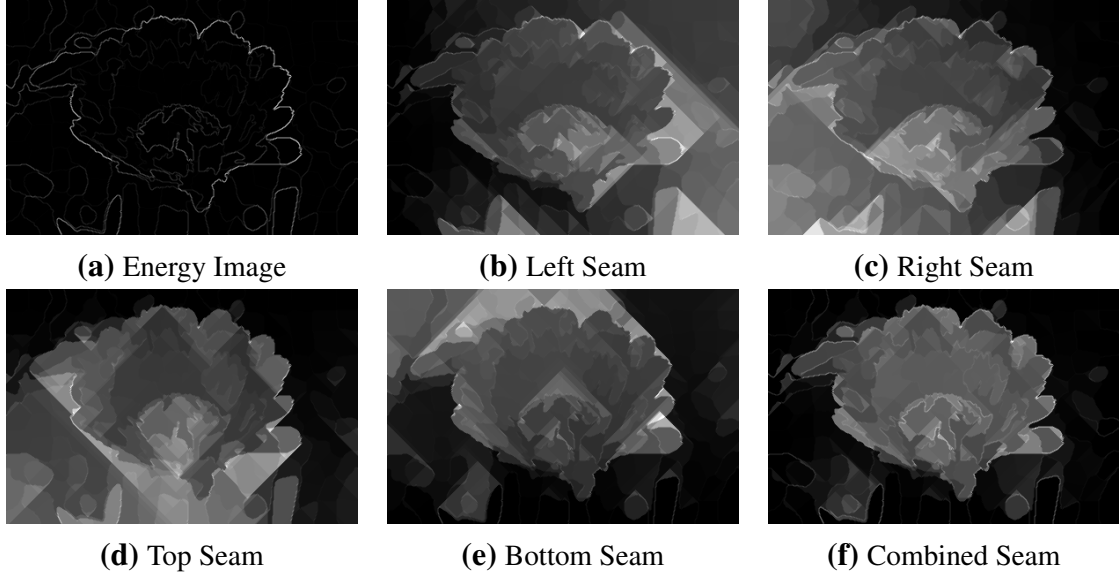
Figure 3.3: Different stages of seam map generation method.

seen in Fig. 3.3, the minimum seam value at each pixel is taken in eq. 2.5 which suppresses the background well. In Fig. 3.3 Figure 3.3(b), 3.3(c), 3.3(d), 3.3(e) are generated from the energy image 3.3(a) and 3.3(f) is generated by combining Figs. 3.3(b), 3.3(c), 3.3(d), 3.3(e) using eq. 2.5. This way the combined map is found by selecting the minimum of four values for each pixel by eq. 2.5. It removes the disadvantage of boundary prior and successfully enumerates the salient object even if it cuts the boundary of an image. The region level seam map is found by averaging seam values in each region as seen in eq. 3.1. The final seam map is generated by further down-weighting distant regions using average spatial distance as in eq. 3.2. Pictorial representation of the seam map generation process is visually presented in Fig 3.3.

$$Imp_{seam}(r_i) = \frac{1}{|r_i|} \sum_{\forall I_{cst}(x,y)=i} Imp_{seam}(x,y), where$$

$$|r_i| = \text{Number of pixels in region } r_i$$

(3.1)

$$S_{seam,s}(r_i) = \frac{Imp_{seam}(r_i)}{d_s^2(r_i)}$$

$$d_s = \text{Euclidian region distance}$$

(3.2)

(a)   **(b)**

Figure 3.4: Border contrast map of the input image.

3.4 Border Contrast Map

Boundary prior is a very successful cue in case of saliency detection. Many researchers have found this cue useful in generating state of the art results. Its main assumption is boundary regions are mostly backgrounds. To get advantage of this simple but significant characteristic of a digital image, every region is compared with the border regions to find the dissimilarity of that region with its boundary.

In eq. 3.3 these contrast dissimilarities are computed and the border contrast map is found by averaging the dissimilarities of each region with all boundary regions using eq. 3.4. The idea of boundary prior is extracted from the contrast map prior. Eq. 3.3 and 3.4 explains the required changes of eq. 2.1 and 2.2 to produce the border contrast map.

Fig. 3.4(b) shows the border map of input image 3.4(a) which is computed from its superpixel image by considering the dissimilarities with its border regions using eq. 3.3 and 3.4.

$$\rho(r_i, b_j) = 1 - exp\left(\frac{-d_c^2(r_i, b_j)}{\sigma^2}\right)$$

$$d_c = \text{Euclidian color distance}$$

$$\sigma = \text{constant to control strength}$$

$$b_j = j^{th} \text{ region of total M border regions}$$

(3.3)

(a) L channel   **(b)** a channel   **(c)** b channel

(d) diff. L   **(e)** diff. a   **(f)** diff. b

(g) G_{map} **(h)** N_{map} **(i)** Imp_{clr}
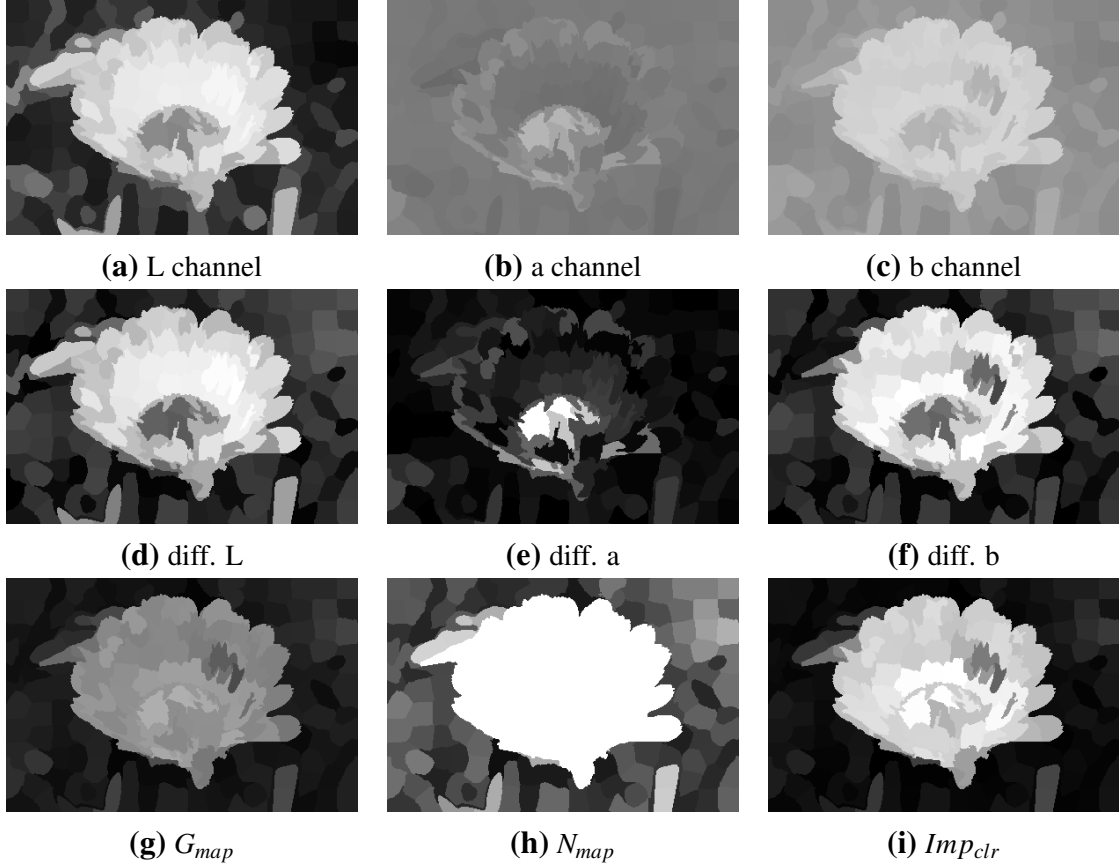
Figure 3.5: Different stages of color map generation method.

$$S_B(r_i) = \frac{1}{\alpha M} \sum_{k=1}^{\alpha M} B_i(k)$$

$$B_i(k) = \rho(r_i, b_j), \text{Dissimilarity vector for } r_i \tag{3.4}$$

$$\alpha = \text{constant } 0 < \alpha < 1$$

3.5 Color Importance Map

A color map consists of the weighted average of different color channels of CIE Lab color space. In a digital image most of the image pixels are backgrounds. So, the main idea of color importance map is to suppress those background pixels by applying some weighting techniques to make the salient regions prominent. The device independent Lab color space is used to calculate the weighted color map. Because, it provides us more information than

RGB color space where all three channels contain similar information. The visual appear-
ance of L, a and b channels of a test image are presented in Fig. 3.5(a), 3.5(b) and 3.5(c). The
average of L, a and b channels are calculated using eq. 3.5 and the color difference images
of Fig. 3.5(d), 3.5(e) and 3.5(f) are found by subtracting each channel's average from color
intensity values of each pixel as of eq. 3.6. The average calculation of the L channel is a
bit different. Because, in this channel the intensity values can differ in a range of 0 to 255.
Thus the mean and mode can have a large difference and because of frequency, the mode
may become insignificant. For this reason, the minimum of $(L_{mean}+L_{mode})/2$ and L_{mean}
is taken as the weighted average in eq. 3.5. The Grey map (G_{map}) in eq. 3.7 sums up all
three channels and Normalized map (N_{map}) in eq. 3.8 produces the magnitude value of the
color space. Grey map and normalized map are shown in Figs 3.5(g) and 3.5(h) respectively.
These two maps are combined into the single saliency map of Fig. 3.5(i) using eq. 3.9.

$$C_{avg} = [L_{avg}, a_{avg}, b_{avg}]$$
$$L_{avg} = min\left(\frac{L_{mean}+L_{mode}}{2}, L_{mean}\right)$$
$$a_{avg} = \frac{a_{mean}+a_{mode}}{2}$$
$$b_{avg} = \frac{b_{mean}+b_{mode}}{2}$$
(3.5)

$$dI_{Lab}(x,y) = \left|I_{Lab}(x,y) - C_{avg}\right|$$
$$I_{Lab}(x,y) = \text{Color intensity of pixel(x,y) Lab image}$$
(3.6)

$$G_{map}(x,y) = w_L d_L(x,y) + w_a d_a(x,y) + w_b d_b(x,y)$$
$$where, w_L = 0.1 \text{ and } w_a = w_b = 0.45$$
(3.7)

$$N_{map}(x,y) = \sqrt{d_L^2(x,y) + d_a^2(x,y) + d_b^2(x,y)}$$
(3.8)

$$Imp_{clr}(x,y) = G_{map}^{\gamma}(x,y) \times N_{map}(x,y)$$

$$where, \gamma = 1.5^{-1}$$

(3.9)

$$Imp_{clr}(r_i) = \frac{1}{|r_i|} \sum_{\forall I_{clr}(x,y)=i} Imp_{clr}(x,y), where$$

$$|r_i| = \text{Number of pixels in region } r_i$$

(3.10)

Though, gamma correction improves the result very slightly, it is performed on the grey map before the combination in eq. 3.9. The pixel level color map is transformed into region level map using the region average in eq. 3.10. The combined map highlights the salient region but it also fails at the boundary. So, the border contrast map is introduced with weighted color map in eq. 3.12 to produce the boundary aware color importance map. Fig. 3.6 shows the boundary aware version of the color importance map. This boundary aware color map of Fig. 3.6(c) is found by merging border contrast map of Fig. 3.6(a) and seam map of Fig. 3.6(b). As seen in this figure, the resulting combination is a better candidate for saliency detection.

$$S_{clr,s}(r_i) = \frac{Imp_{clr}(r_i)}{\frac{1}{N}\sum_{k=1}^{N} d_s(r_i, r_k)}$$

$$d_s = \text{Euclidian region distance}$$

$$N = \text{Number of regions}$$

(3.11)

$$S_{bClr}(r_i) = S_B(r_i) \times [1 + S_{clr,s}(r_i)]$$

$$S_B(r_i) = \text{Border contrast map from eq. 3.4}$$

(3.12)

(a)   **(b)**   **(c)**
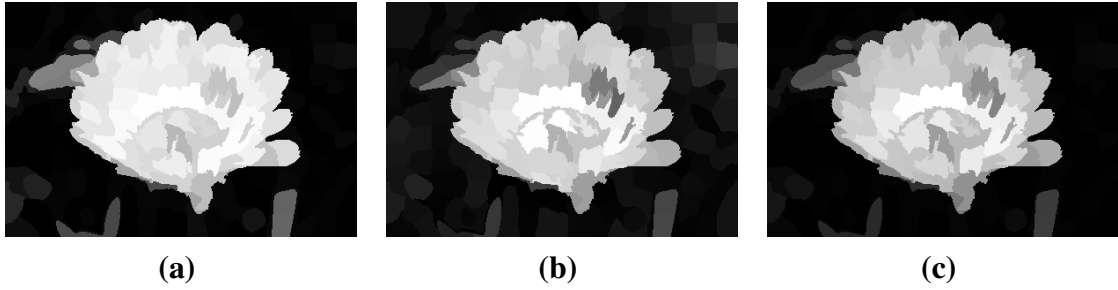
Figure 3.6: Generation of boundary aware color map by merging border contrast map and seam importance map.

(a)   **(b)**   **(c)**
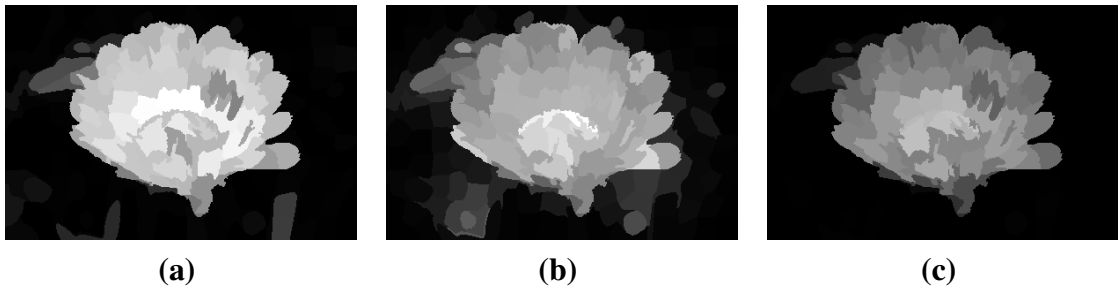
Figure 3.7: Merging boundary aware color map and seam importance map.

3.6 Combining Seam and Color Importance Maps

Following the traditional way of combining saliency maps, the boundary aware color map is combined with seam map by eq. 3.13 using multiplication as seen in Fig. 3.7. The combined saliency map of Fig. 3.7(c) is found by merging boundary aware color map of Fig. 3.7(a) and seam map of Fig. 3.7(b). The samples in Fig. 3.7(a) and Fig. 3.7(b) both contains highlighted portions of backgrounds. Though, the combined result successfully suppresses the background, it does not enumerate the salient region sufficiently. So, an optimization procedure is required to produce a more robust saliency map.

$$Sal'(r_i) = S_{bClr}(r_i) \times S_{seam,s}(r_i) \tag{3.13}$$

3.7 Color Distance and Contrast Dissimilarity

The existing color contrast between the regions in an image is one of the main attraction of saliency detection studies. The euclidean color distance in Lab color space of the superpixels

is used to measure the contrast difference between the superpixels as in eq. 3.14. The global contrast map is found by using eq. 3.15. This is used in the optimization equation as the smoothness weight.

$$\rho(r_i, r_j) = exp\left(\frac{-d_c^2(r_i, r_j)}{\sigma^2}\right)$$
$$d_c(r_i, r_j) = \sqrt{(c_i - c_j)^2}$$
$$c_i, c_j = \text{Mean color vector of } r_i, r_j$$
$$\sigma = \text{constant to control strength}$$

(3.14)

$$S_G(r_i) = \frac{1}{N}\sum_{k=1}^{N}\rho(r_i, r_j)$$
$$N = \text{Number of regions}$$

(3.15)

3.8 Optimization

The produced saliency map more often contains clutter or noise. To remove those clutter or noise some smoothing operations are usually performed. Gestalt thresholding is a successful technique in this regard. Moreover, combined saliency maps usually observe these type of clutters. So, optimization technique becomes a necessary step in saliency map generation procedure.

3.8.1 Gestalt Smoothing Technique

Gestalt Law is used to produce a smooth saliency map by removing clutters and noises. Here, we use the same sets of smoothing equations used in [20], [21] and [13] to explore regions that are closer to foci of attention in a combined saliency output of eq. 3.13. So, by using this law an optimized version of the seam map is found as seen in Fig. 3.8(a).

In [20] and [21] Gestalt Law [53] is used to smooth the combined saliency cues in order to produce a more efficient saliency map. According to that principle, visual systems are more likely to group similar regions together. So, equations 3.16 and 3.17 are used in an input

(a) Gestalt Smoothing **(b)** Cost Minimization
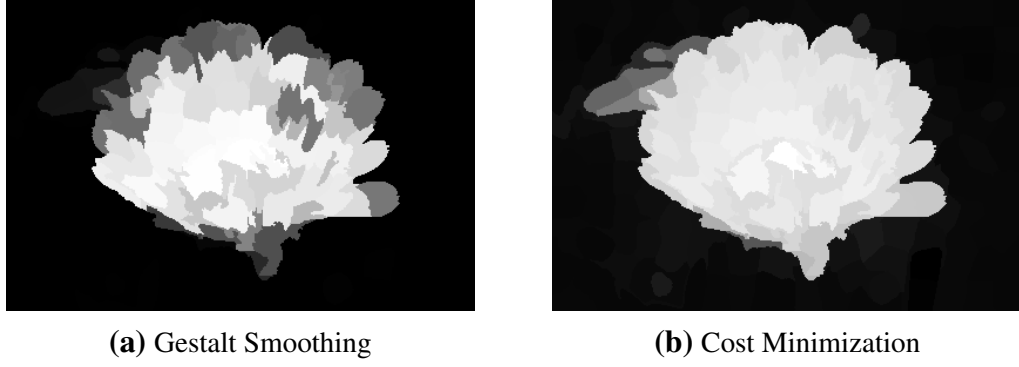
Figure 3.8: Optimized saliency maps.

image to explore regions that are closer to foci of attention. Finally, The optimized saliency map is found by using eq. 3.17. This approach is also applied in [13], [21] and [20].

$$Sal'(r_i) = \begin{cases} min(1.0, Sal'(r_i) \times (1 - d_{foci}(i))^{-1}; & \text{if } d_{foci}(i) \leq 0.2 \text{ and } Sal'(r_i) > \bar{m} + sd \\ Sal'(r_i) \times (1 - d_{foci}(i)); & \text{if } d_{foci}(i) \leq 0.5 \\ Sal'(r_i) \times (1 - d_{foci}(i))^2; & \text{otherwise} \end{cases}$$

(3.16)

$$Sal(r_i) = \frac{1}{Z_i} \sum_{k=1}^{N} Sal'(r_i) \times exp \frac{-|Sal'(r_i) - Sal'(r_k)|}{\delta}$$

(3.17)

3.8.2 Cost Minimization Function

Most of the previous researchers use multiplication or weighted summation to combine multiple saliency cues. The resulted combination does not do well in case of generalization. The main target of saliency detection techniques is to generate a map closer to the binary ground truth image. So, Zhu et. al. have developed an optimization technique towards this goal. In this research, optimization is done by minimizing their quadratic cost function 3.18 proposed in [30]. All three terms of this cost function are square errors and it can be solved by using least square method. The first term of the cost function represents the background which encourages a superpixel p_i with large background weight w_i^{bg} to take a small value s_i closer

(a)   **(b)**   **(c)**
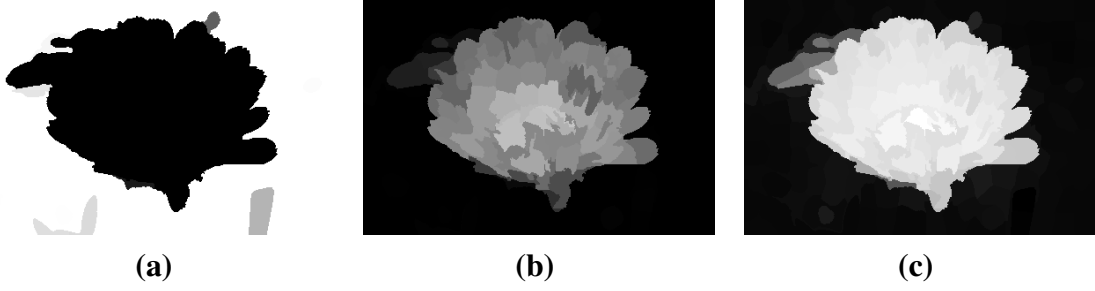
Figure 3.9: The optimal saliency map is found by solving the quadratic equation proposed in [30] using background map and foreground maps.

to 0. In this research, the background weight w_i^{bg} is $1 - Saliency$. The second term represents the foreground which encourages a superpixel p_i with large foreground weight w_i^{fg} to take a large value closer to 1. Here, the foreground weight is one of the saliency maps found in this study. The third term is the smoothness term which smooths both the background and foreground by removing small noises. The color dissimilarity based global contrast map from eq. 3.15 is used as the weight for the smoothness term. Since, the target of this weight is to smooth the combined saliency map, connecting two levels of nearest neighbours and the boundary regions before dissimilarity measure produces better result as shown in Fig. 3.9. The optimal saliency map of Fig. 3.9(c) is found by solving the quadratic equation proposed in [30] using 3.9(a) as background map and 3.9(b) as foreground map. In this study, different combinations of saliency measures are used as the foreground and background weight in eq. 3.18 as shown in table 4.1 and a balanced combination is proposed as the optimal combination.

It smooths the combined saliency map with a target to minimize the cost of foreground and background separation. It is the target oriented optimization technique that simultaneously draws a line between foreground and background pixels and results in a highlighted foreground with a suppressed background. It also smooths the resulted output in both foreground and background region.

$$\sum_{i=0}^{N} w_i^{bg} s_i^2 + \sum_{i=0}^{N} w_i^{fg} (s_i - 1)^2 + \sum_{i,j} w_{i,j} (s_i - s_j)^2 \tag{3.18}$$

3.9 Summary

In this chapter, the proposed architecture is discussed in details. It is shown how the input raw image is converted into the superpixelized version and from that superpixelized image foreground and background saliency maps are calculated using seam and color information. The contribution of border prior and its calculation procedures are also described. Besides that, two possible optimization procedure is discussed which can furnish the combined saliency map as the final output of the proposed methodology.

Chapter IV

Experiments, Results and Discussion

4.1 Introduction

Different types of experiment with the proposed method has been performed on several state of the art datasets. In this chapter, detailed outcomes of those experiments have been presented along with some historical background of the datasets. Moreover, The basis of evaluation and comparison of the proposed method with other state of the art techniques has been presented in this chapter.

4.2 Datasets

Standard datasets like MSRA 1k, MSRA 10k, CSSD and ECSSD are used to evaluate the performance of this study. Here, MSRA 1k and 10k are subsets of widely used MSRA original dataset which contains 20,000 images along with their hand labelled rectangular ground truth images which are manually annotated by 3-9 users. Wang and Li [54] and Achanta et. al. [15] has claimed that original MSRA dataset has limitation in fine grained evaluation because of being too coarse in some cases. In [15] the authors presented the MSRA 1k or ASD dataset by manually segmenting each salient region which overcome
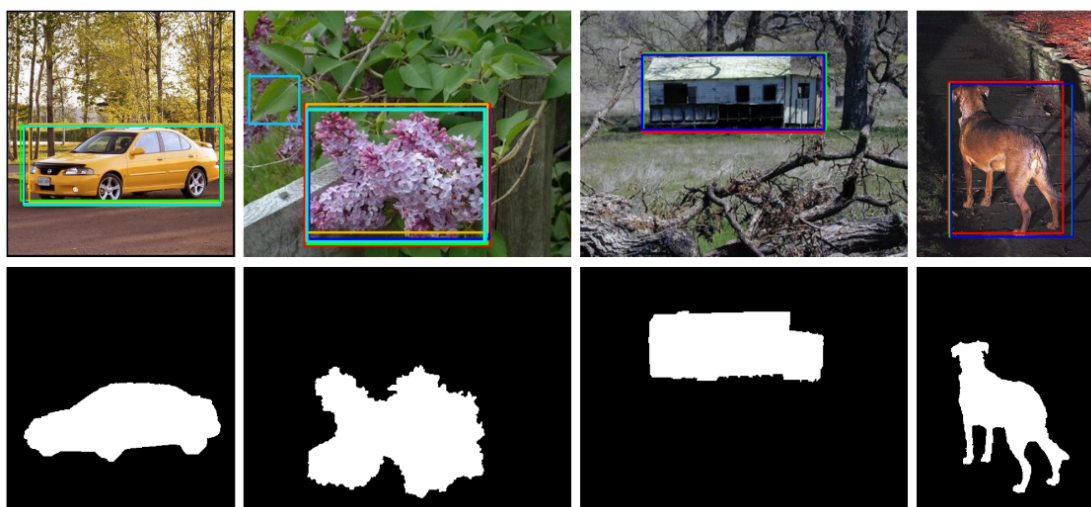
Figure 4.1: Annotated examples of MSRA and MSRA10k dataset.
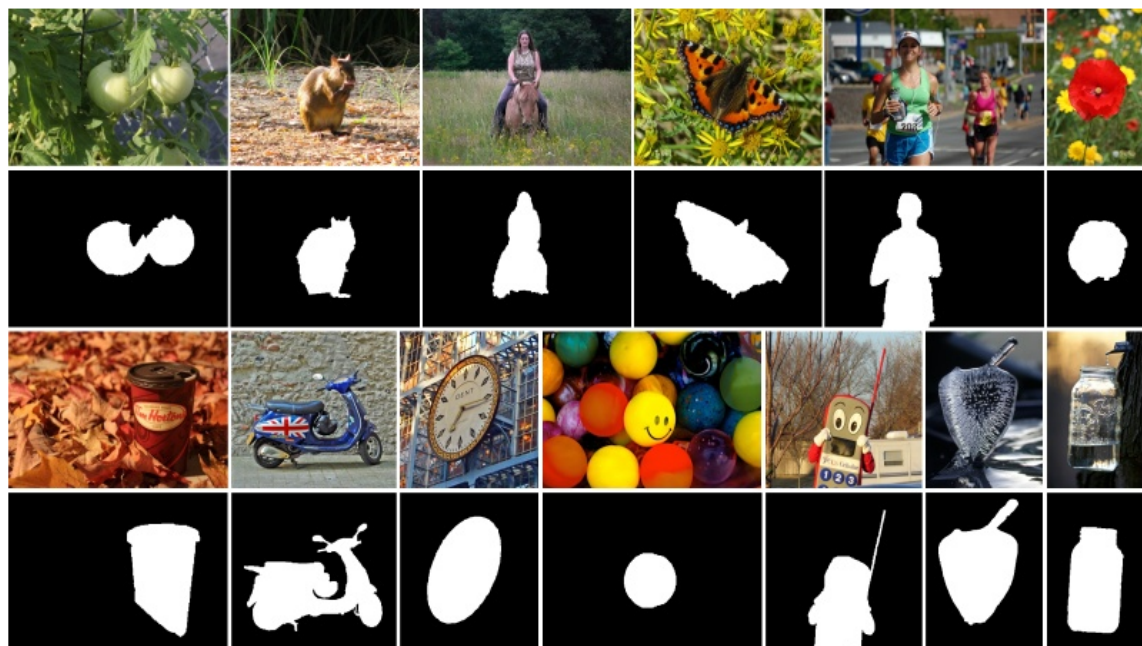
Figure 4.2: Some members of CSSD and ECSSD datasets with complex backgrounds and low contrast surroundings.

the limitations of original MSRA dataset. An extended version of MSRA 1k dataset which contains pixel level binary ground truth images of 10,000 images from the original MSRA dataset is presented in [35]. They named this dataset MSRA10k which was previously called THUS10000. Some examples of MSRA and MSRA10k datasets are presented in FIg. 4.1. The first row shows some bounding box annotation from original MSRA dataset and the second row shows the pixel level ground truth annotation from MSRA10k dataset. Almost all modern saliency detection techniques use this dataset as a benchmark for evaluation of saliency detection methods.

Microsoft Research Asia (MSRA) based datasets are good for saliency detection evaluation and contain a large variety in contents, but their backgrounds are relatively simple and smooth. To evaluate the complex situations in saliency detection, where objects and backgrounds have relatively low contrast and object surroundings are comparatively complex, Yan et. al. [18] have proposed CSSD dataset with pixel level binary ground truths in 2013 which is a more challenging dataset with 200 images containing complex situations and they extended the dataset by increasing the images to 1,000 in [55] and named it Extended Complex Scene Saliency Dataset (ECSSD).

4.3   Evaluation Metrics

There are usually two types of evaluation strategies for every computer vision technique - qualitative analysis and quantitative analysis. Results found in this study are compared with several state of the art methods. Qualitative analysis of this study is presented in Fig. 4.14 with some of the mentioned methods containing comparative samples from respective technologies. For quantitative analysis precision recall curves are shown for different datasets along with other techniques and clearly proves the superiority of the proposed method. Eq. 4.1 and 4.2 represents precision and recall respectively for a single image q with threshold T. To draw the precision recall curve simple binarization is performed by varying the threshold T in eq. 4.3 and 4.4 within the range from 0 to 255 and the curves are plotted with respect to the ground truth data available for each input image in the corresponding dataset. Since, precision-recall curves only consider whether the object saliency is higher than the background saliency, to evaluate the overall performance, F-beta measures in Fig. 4.9 are also presented in this study to better understand the performance of the compared methods.

$$precision(q,T) = \frac{tp}{tp+fp} \tag{4.1}$$

$$recall(q,T) = \frac{tp}{tp+fn} \tag{4.2}$$

$$Precision_T = \frac{1}{Q}\sum_{q=1}^{Q} precision(q,T) \tag{4.3}$$

$$Recall_T = \frac{1}{Q}\sum_{q=1}^{Q} recall(q,T) \tag{4.4}$$

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{4.5}$$

4.4 Experimental Setup

Standard datasets that are discussed in section 4.2 are used to evaluate the proposed method. The configuration of the windows machine for this study is second generation Intel core i5 processor with 8 Gb of DDR3 RAM. All the experiments of this thesis are performed in Matlab programming environment.

4.5 Quantitative Results and Analysis

4.5.1 Effects of Various Combinations

As observed in the previous chapter, optimization is a very useful step in enhancing the performance of the saliency detection procedure. It is done in different ways. In this section some optimization procedures are explored for saliency detection and the best of them for the final output is recommended.

Comparison of Optimization Techniques

Gestalt thresholding technique and cost minimizing method have been compared in Fig. 4.3(a) and 4.3(b). Both of the figures represent combination of seam, color and background cues on MSRA 1k dataset. In this illustration, the costMinimizing and gestaltSmoothing optimizations have F_β scores 0.892 and 0.874 respectively. So, the cost minimizing approach
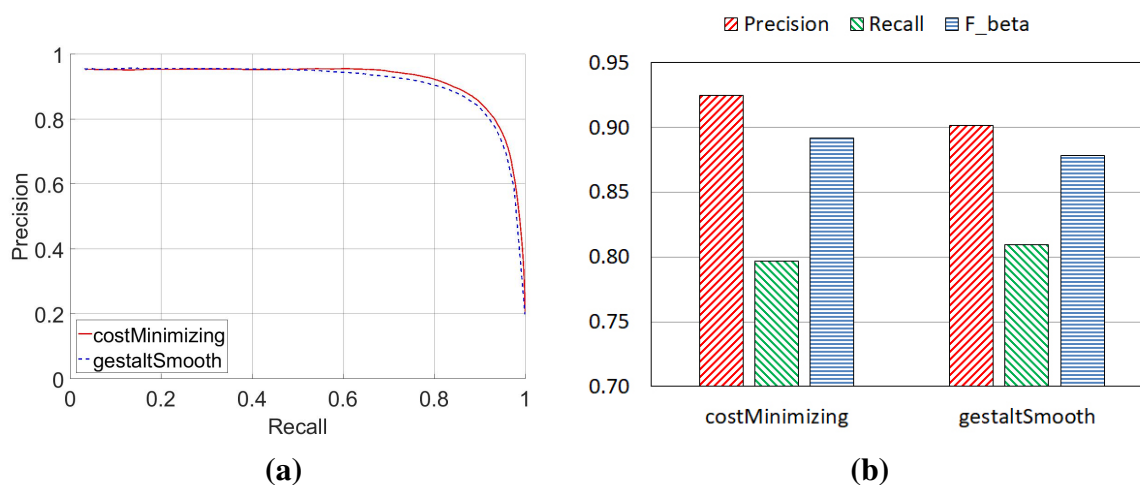

Figure 4.3: Comparison of optimization techniques.

outperforms the Gestalt Smoothing technique and this is considered as the optimization procedure of the proposed method.

Cost Minimizing Optimization Experiments

In this study, several promising combinations of saliency maps are explored by using cost minimization optimization technique. Some of them are briefly discussed in this section and mentioned in table 4.1. In those three combinations, the smoothness weight is kept the same which is the global contrast map between the superpixels from eq. 3.15. If the border contrast map is used as the background weight in the proposed cost minimization based optimization method and the seam importance map is used as the foreground weight, closer PR curves for comb2 and comb3 are found as seen in Fig. 4.4. The f-beta score is also closer for those combinations as depicted in that figure. Comb2 shows slightly degraded performance in terms of quantitative analysis. To better understand the differences, same combinations are analysed with comparatively complex ECSSD dataset containing same number of members like MSRA 1k. The results are shown in FIg. 4.4 and F_β scores from both the tests are given in table 4.2. Performance of comb2 is further degraded for ECSSD experiment and now comb3 clearly outperforms both comb1 and comb2 as seen in both Fig. 4.4 and table 4.2.

Table 4.1: Proposed optimization combinations.

Method	Foreground	Background
Comb1	Border Contrast Map from eq. 3.4	Seam Importance Map from eq. 3.1
Comb2	Border Contrast Map from eq. 3.4	Combined Seam and Boundary Aware Color Map from eq. 3.13
Comb3 (Proposed)	Boundary Aware Color Map from eq. 3.12	Weighted Seam Map from eq. 3.2

The average runtime for those combinations are given in table 4.3 for MSRA 1k experiment. As seen in that table, the runtime is not good enough for real time applications, since the program is written in Matlab for experiments and the experimental machine was not an ideal one for real time application. It would be much faster for real time use which will be possible by implementing an optimized version of this program in some application level programming language like C++ or Java on a suitable platform. However, the target of this thesis is to find a better method for saliency detection in terms of accuracy. So, application level programming is not implemented.

(a) MSRA1k

(b) MSRA1k

(c) ECSSD

(d) ECSSD
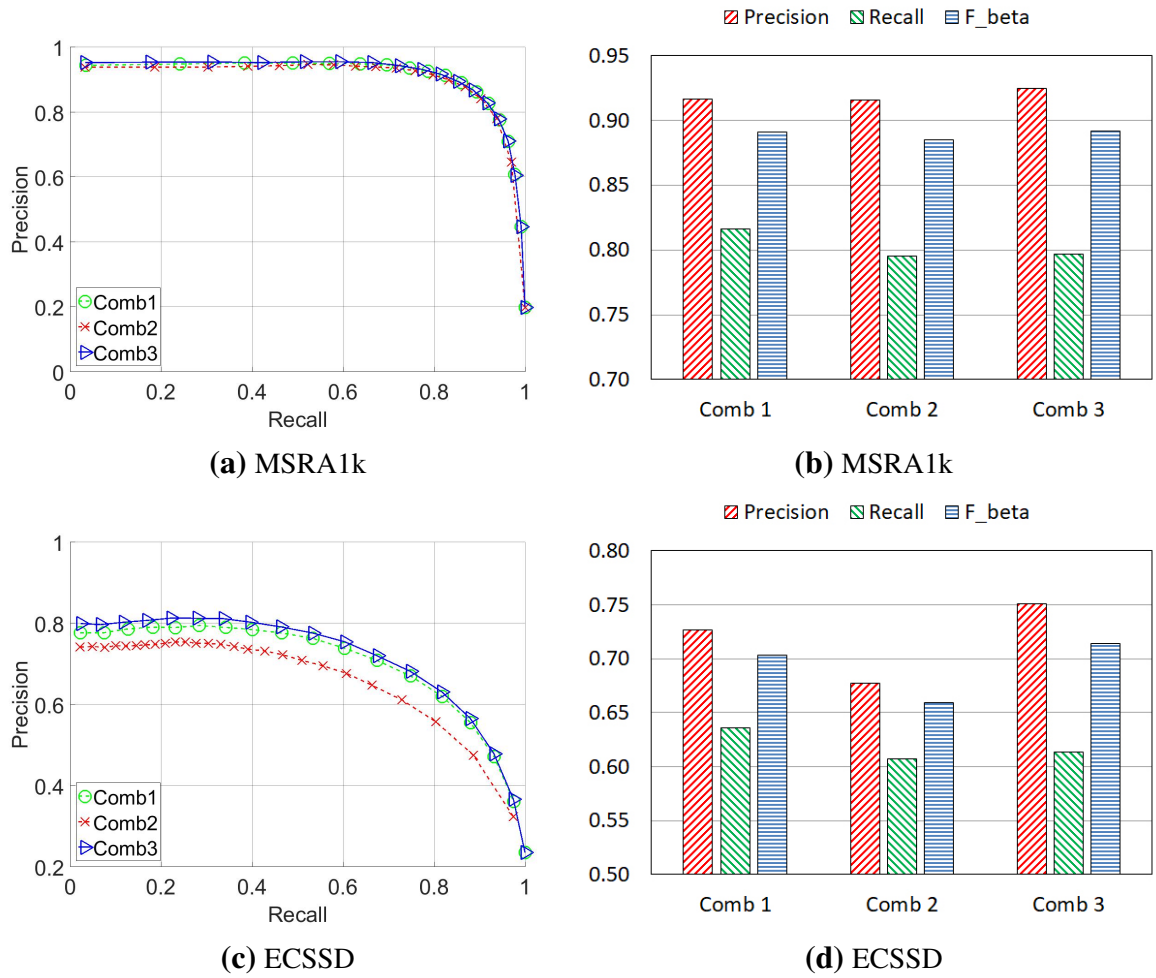
Figure 4.4: Precision recall curves for different combinations.

From the results in Fig. 4.4 and runtimes in table 4.3 it is seen that comb1 and comb3 are closer to each other. But the second combination is the slowest since it introduces more computational complexities and the first combination is the fastest in terms of runtime comparison. Qualitative results are also closer for comb1 and comb3 combinations (as seen in

Table 4.2: F_β score comparison of proposed combinations.

Dataset	Comb1	Comb2	Comb3 (Proposed)
MSRA 1k	0.891	0.885	0.892
ECSSD	0.703	0.659	0.714

Table 4.3: Runtime comparison of proposed combinations.

Method	Comb1	Comb2	Comb3 (Proposed)
Time (sec)	1.63	2.65	1.745
Code	Matlab	Matlab	Matlab

Fig. 4.10). On the other hand, for MSRA 1k dataset the second combination does well in case of PR curve and F-beta test. But, the third one does well in complex situations like the ECSSD dataset. Its performance is also competitive in MSRA 1k dataset and the runtime is quite acceptable in comparison to the fastest method and its excellent performance on the complex dataset. Thus the third combination is chosen as the proposed method in this study and it is be termed as *Our* method onwards.

4.5.2 Effects of Various Parameters

As mentioned before, superpixel size is considered as 600 pixels and $\alpha = 0.3$ is used for border contrast map generation. It is done based on some experiments done on the proposed method. Performance difference is also observed by using raw or superpixel version of the input image for seam and color map generation methods. In this section, all the experiments are performed on MSRA 1k dataset. The results of those experiments are presented below.

Performance of Raw and Superpixelized Seam and Color Importance Maps

The seam and color maps can be generated from raw input image or from superpixelized version of the input image. In [21], seam and color maps are directly calculated from the raw input image and in [20] the seam and color maps are generated from the superpixelized input image. In the precision-recall curve and F-beta chart of Fig. 4.5 the superiority of the
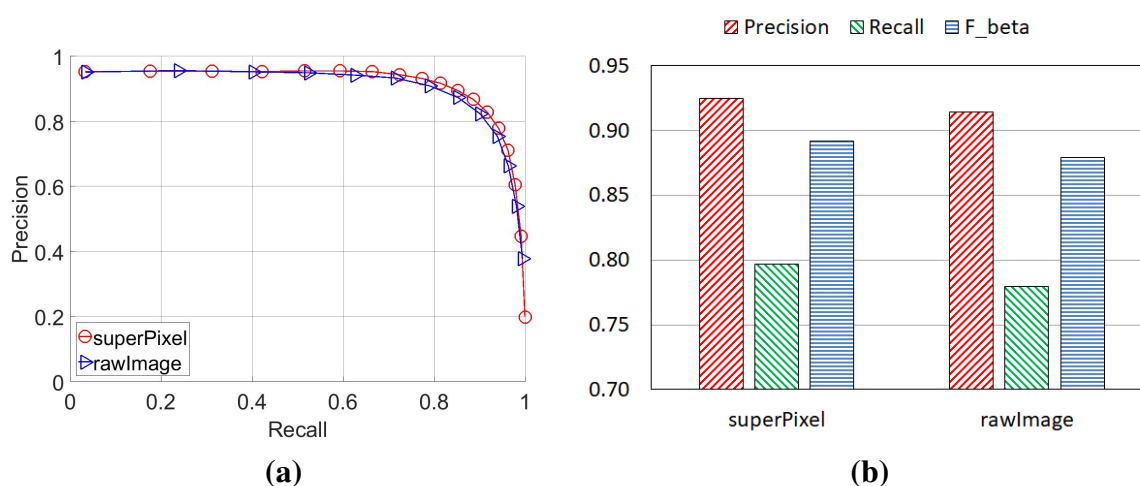

Figure 4.5: Comparison of raw and superpixelized input for seam and color map generation.

superpixel version of the seam and color maps are clearly demonstrated. Here, the F_β values for superPixel and rawImage are 0.892 and 0.879 respectively.

It is probably that the superpixelized image is more smoother and contains stronger edges. So, the superpixel image is choosen as the input to seam and color map generation procedure.

Effect of Different Superpixel Sizes

The SLIC superpixel generation procedure requires the approximate number of pixels for each superpixel as a parameter for an input image. It determines each superpixel size or the number of superpixels in the output superpixelized image. Experiments have been performed on the proposed method using MSRA 1k dataset with different superpixel sizes - 300 pixels, 600 pixels and 900 pixels. The PR curve and F-beta charts are given in Fig. 4.6. It is seen that performance of 600 pixel and 900 pixels are quite closer and better than 300 pixels and 1200 pixels. The F_β values for 300px, 600px, 900px and 1200px are 0.890, 0.892, 0.892 and 0.885 respectively.

As observed in the experiments, the bigger size of superpixels smooth the image and usually better for saliency detection. But, it starts loosing required edges at some point. For this reason performance degrades after increasing the superpixel size to 1200 pixels. Since, 600 pixels results in better precision value as seen in F-beta chart, it is chosen as the preferred size in the proposed saliency detection technique for superpixel generation.
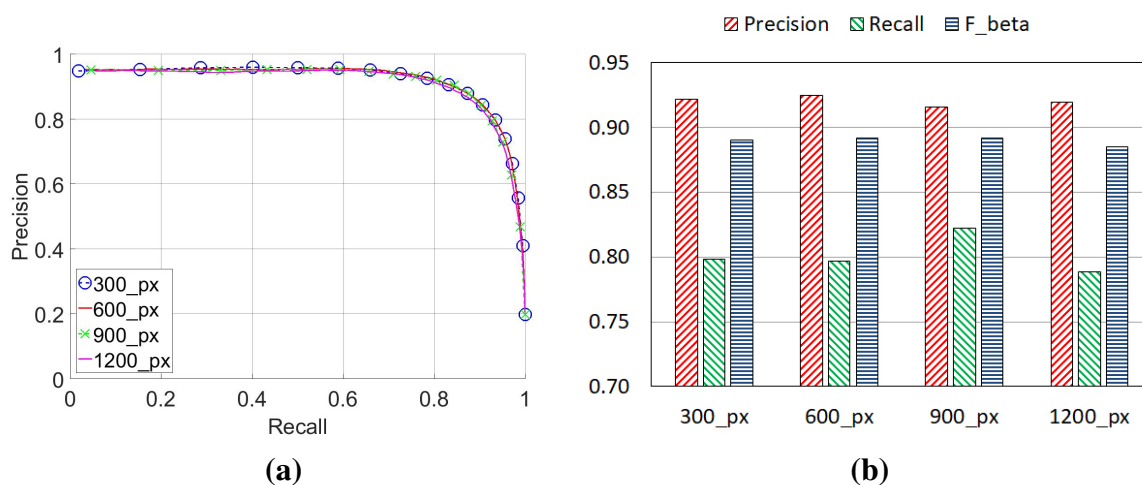

(a)   **(b)**

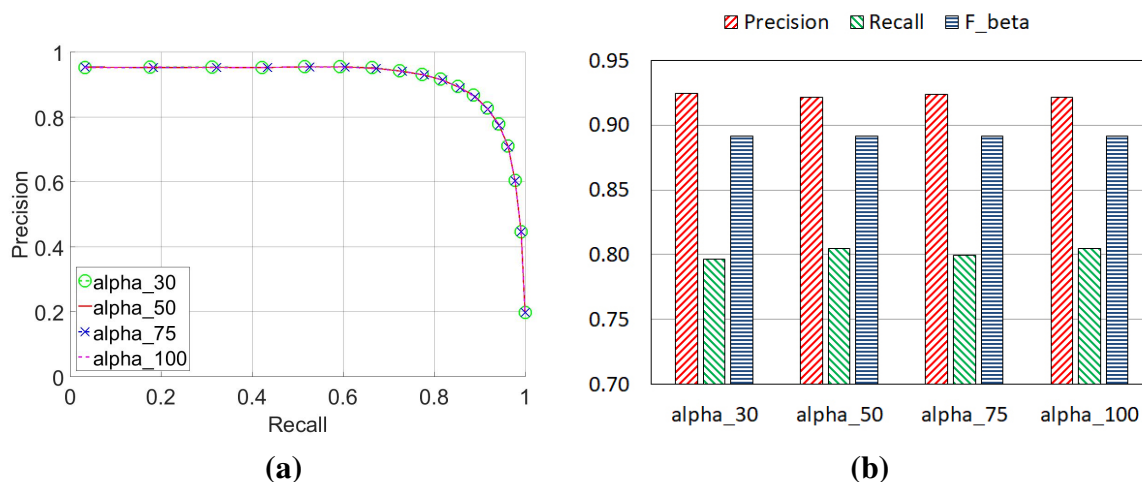**Figure 4.6:** Comparison of different superpixel sizes.

Figure 4.7: Comparison of different alpha values.

Effect of Different α Values

In eq. 3.4 α determines the percentage of background superpixels that is considered for comparison in border map generation procedure. In Fig. 4.7 α is varied from 30% to 100%. It is seen that performance remains almost same as the PR curves overlaps each other and F-beta value is constant at 0.891 for all four cases.

It is because in almost all cases, most of the image pixels are background and a very small portion constitutes the salient region. Taking smaller α value makes calculation less complex and faster. Besides that, the F-beta chart shows that precision is higher for $\alpha = 0.3$. So, this is chosen for as the α value in this thesis.

Based on the experiments in this section, the parameters in the following table 4.4 are chosen for the proposed saliency model.

Table 4.4: Chosen parameters for the proposed saliency model.

Superpixel Size	Alpha (α)	Compactness	Foreground	Background
600 px	0.30	20	Boundary Aware Color Map from eq. 3.12	Weighted Seam Map from eq. 3.2

4.5.3 Comparison with Other State of the Art Methods

Results found in this study are compared with several state of the art methods which are labelled as HC[35], MSS[33], GB[48], SWD[56], RC[57], AC[58], IT[8], GR[59], LC[60], FT[26], CA[16], AIM[61], GS[17], SF[50], SR[62], BARC[13], BARCS[21] and BACS[20].

(a) MSRA1k   **(b)** MSRA10k

(c) CSSD   **(d)** ECSSD
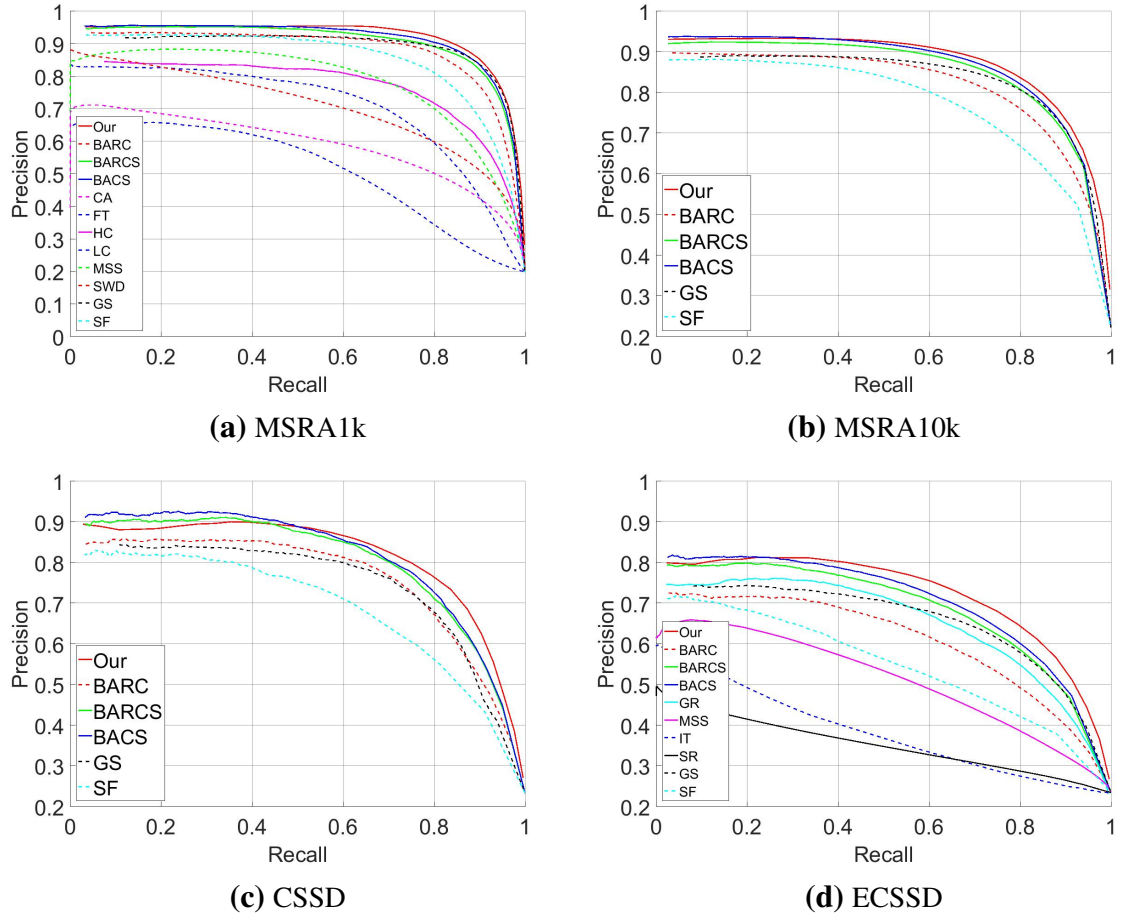
Figure 4.8: Precision recall curves for different datasets.

Most of these labels are used in previous researches like [15], [35], [63]. Other labels are constructed by taking some meaningful letters from the titles of the related publications. Based on publicly available results for specific dataset, the proposed method has been compared with other state of the art methods as shown in table 4.5.

Table 4.5: Datasets and compared methods.

Dataset	Compared Methods
MSRA 1k	CA[16], FT[26], HC[35], LC[60], MSS[33], SWD[56], GS[17], SF[50], BARC[13], BARCS[21], BACS[20]
MSRA 10k	GS[17], SF[50], BARC[13], BARCS[21], BACS[20]
CSSD	GS[17], SF[50], BARC[13], BARCS[21], BACS[20]
ECSSD	GR[59], IT[8], SR[62], MSS[33], GS[17], SF[50], BARC[13], BARCS[21], BACS[20]

The qualitative analysis in Fig. 4.14 visually presents the performance of the presented method compared to other state of the art techniques. The first row shows the input images and the last row contains the target pixel level ground truth for corresponding input image.

(a) MSRA1k

(b) MSRA10k

(c) CSSD

(d) ECSSD
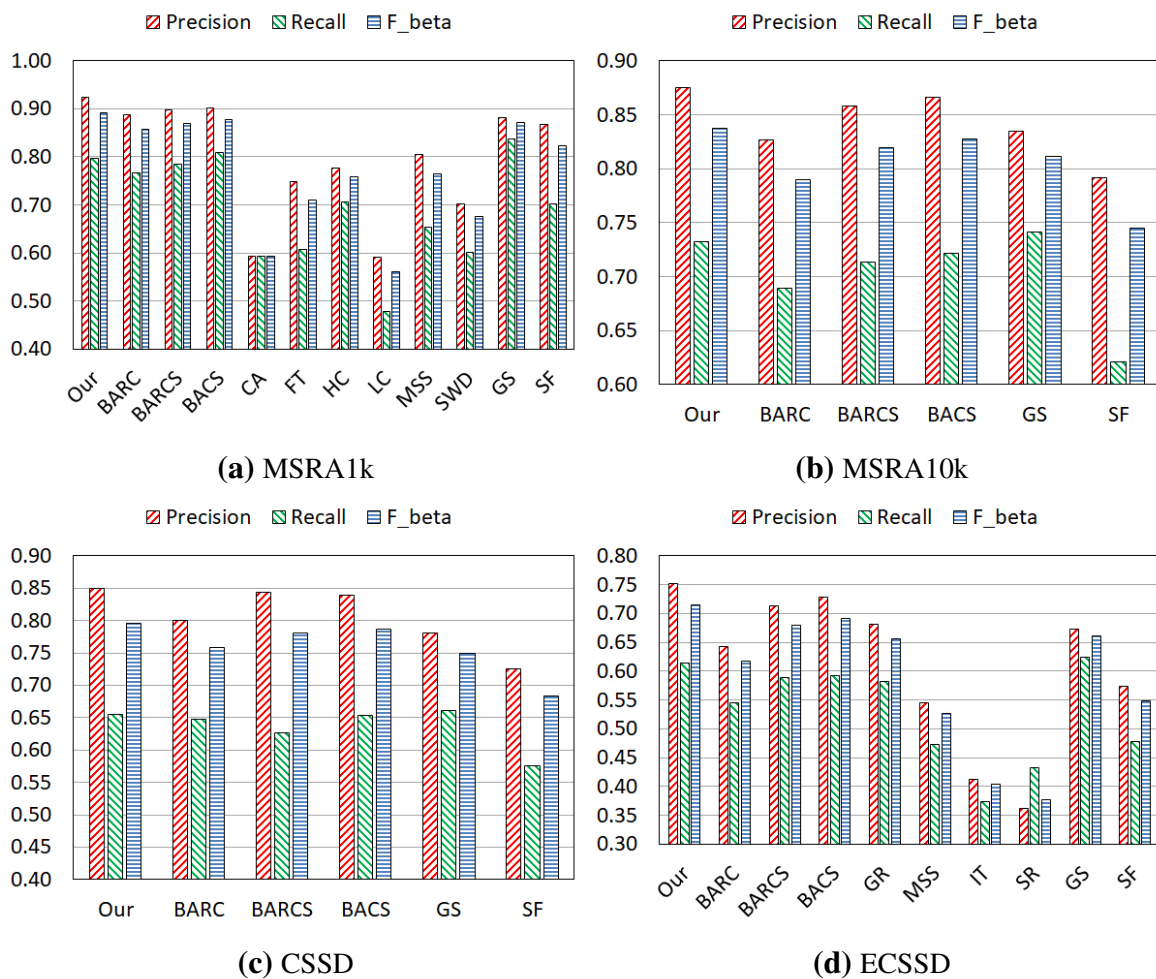
Figure 4.9: F-beta curves for different datasets.

From this figure it is seen that the proposed method done very well with other compared results.

In case of quantitative comparison, the precision-recall curve in Fig. 4.8 clearly demonstrates the superiority of the proposed method. As pointed in [35], all the methods have the same precision and recall values at maximum recall for threshold $= 0$ and all pixels are considered to be in the foreground. On the other hand, at minimum recall values of the proposed method are higher than the other compared methods as a result of smoother saliency containing more pixels in the salient region. $\beta = 0.3$ is considered in eq. 4.5 like [21] and [13] to draw F-beta measures in Fig. 4.9. The bars in this figure demonstrates that the presented method clearly outperforms other state of the art techniques.

The average runtime of different methods are given in table 4.6 as presented in [35]. The given times are depended on the programming platform and a different machine specification

(a) Comb1   (b) Comb2   (c) Comb3 (Proposed)

Figure 4.10: Optimization combinations.

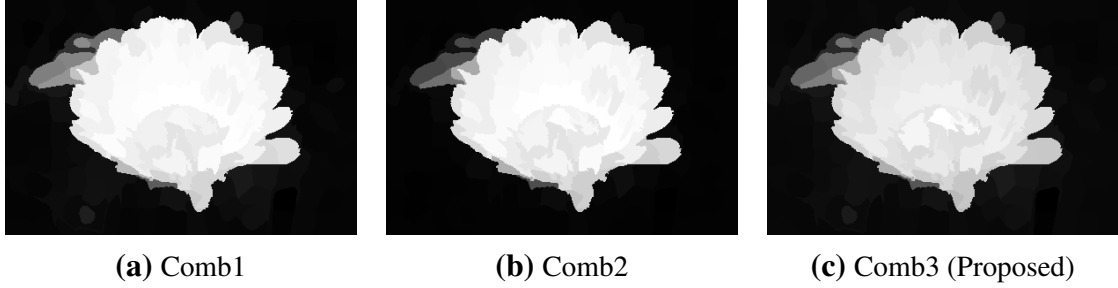and may vary on other computer or programming platform. So, this is mentioned here only for reference. The average runtime of the proposed method is also mentioned in that table. The configuration of the windows machine for this study is given in section 4.4.

Table 4.6: Time comparison of different methods.

Method	Our	AIM	CA	CB	FT	GB	HC	IT	LC	MSS	RC	SWD
Time (s)	1.745	4.288	53.1	5.568	0.102	1.614	0.019	0.611	0.018	0.106	0.254	0.100
Code	Matlab	Matlab	Matlab	M&C	C++	Matlab	C++	Matlab	C++	Matlab	C++	Matlab

4.6 Qualitative Results

4.6.1 Visual Effects of Various Combinations

Three different combinations of foreground and background saliency maps are experimented in this research. The quantitative analysis has been presented in section section 4.5.1 and the qualitative comparison is presented in Fig. 4.10. As mentioned before, there is not much visual difference among the combinations. From the discussion in section 4.5.1, it is clear that comb3 shows better performance for complex ECSSD dataset (see Fig. 4.4) and its runtime is better than comb1. So, comb3 is the proposed model of this thesis.

4.6.2 Visual Effects of Various Parameters

In the previous section, the quantitative effects of different parameters have been explored. In this section, the qualitative outputs of those experiments are presented. Though, the quantitative outputs clearly presents the differences of different parameter values, the figures of this section also supports their relating quantitative statements in the preceding section.
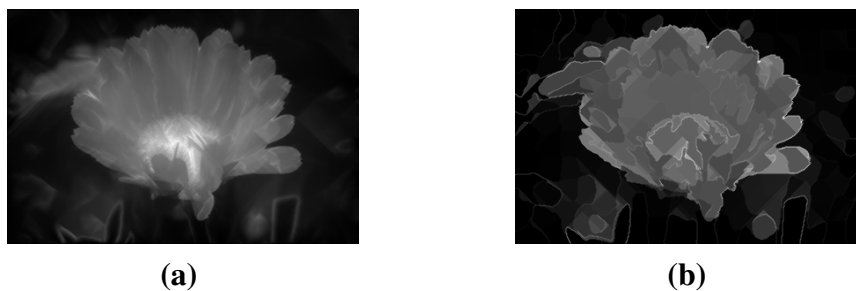
(a)   (b)

Figure 4.11: Seam map from raw input image and superpixelized input
image.

(a) 300 px   **(b)** 600 px   **(c)** 900 px   **(d)** 1200 px
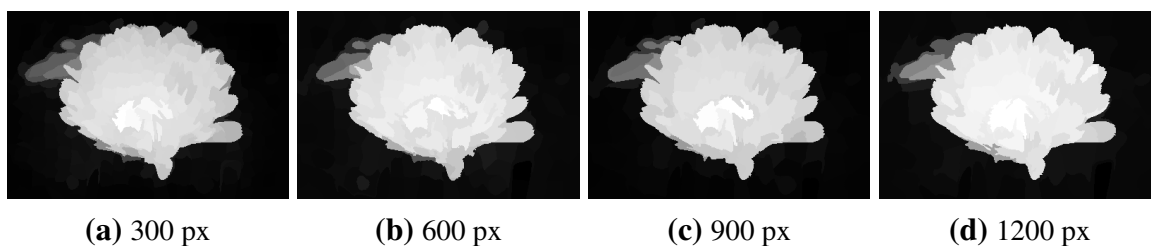
Figure 4.12: Visual examples of different superpixel sizes.

Raw vs Superpixel Input Image

The visual difference can be observed in Fig. 4.11. It is seen that the superpixelized version
results in better seam map because in the input superpixel image is more smoother and has
strong edges. This is also proved in quantitative results.

Visual Effect of Superpixel Sizes

The visual difference of different superpixel sizes is presented in Fig. 4.12. It is seen that the
superpixelized version results in better seam map because in the input superpixel image is
more smoother and has strong edges. This supports the statements in quantitative analysis.

Visual Effect of α variants

The visual difference of different superpixel sizes is presented in Fig. 4.13. It is seen that the
superpixelized version results in better seam map because in the input superpixel image is
more smoother and has strong edges. This supports the statements in quantitative analysis.

(a) 0.30	**(b)** 0.50	**(c)** 0.75	**(d)** 1.00
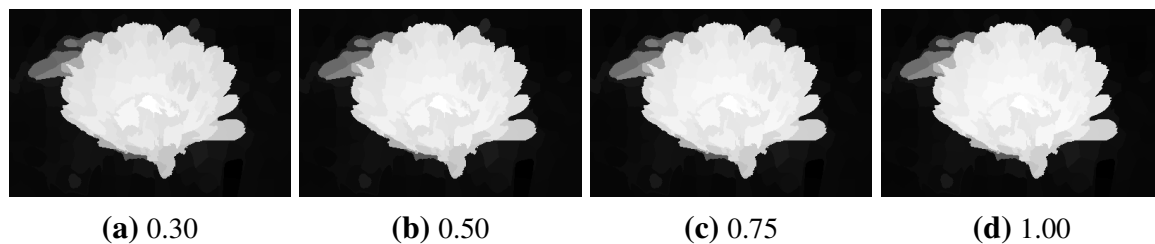
Figure 4.13: Visual examples of different alpha values.

4.6.3   Qualitative Comparison with Other Methods

The qualitative comparison of the proposed saliency model is presented in Fig. 4.14, 4.15, 4.16 and 4.17 on MSRA 1k, MSRA 10k, CSSD and ECSSD datasets respectively. The proposed saliency model is compared with publicly available results of other state of the art techniques for specific datasets mentioned in section 4.5.3. In all of the qualitative results it is seen that the proposed model does better than the compared techniques as it covers more area of the ground truth image in the last row of each figures which is the target output. The visual results also proves the superiority of the proposed model compared to other state of the art techniques.

Figure 4.14: Qualitative comparison of our saliency output with other state of the art methods on MSRA 1k dataset.

48

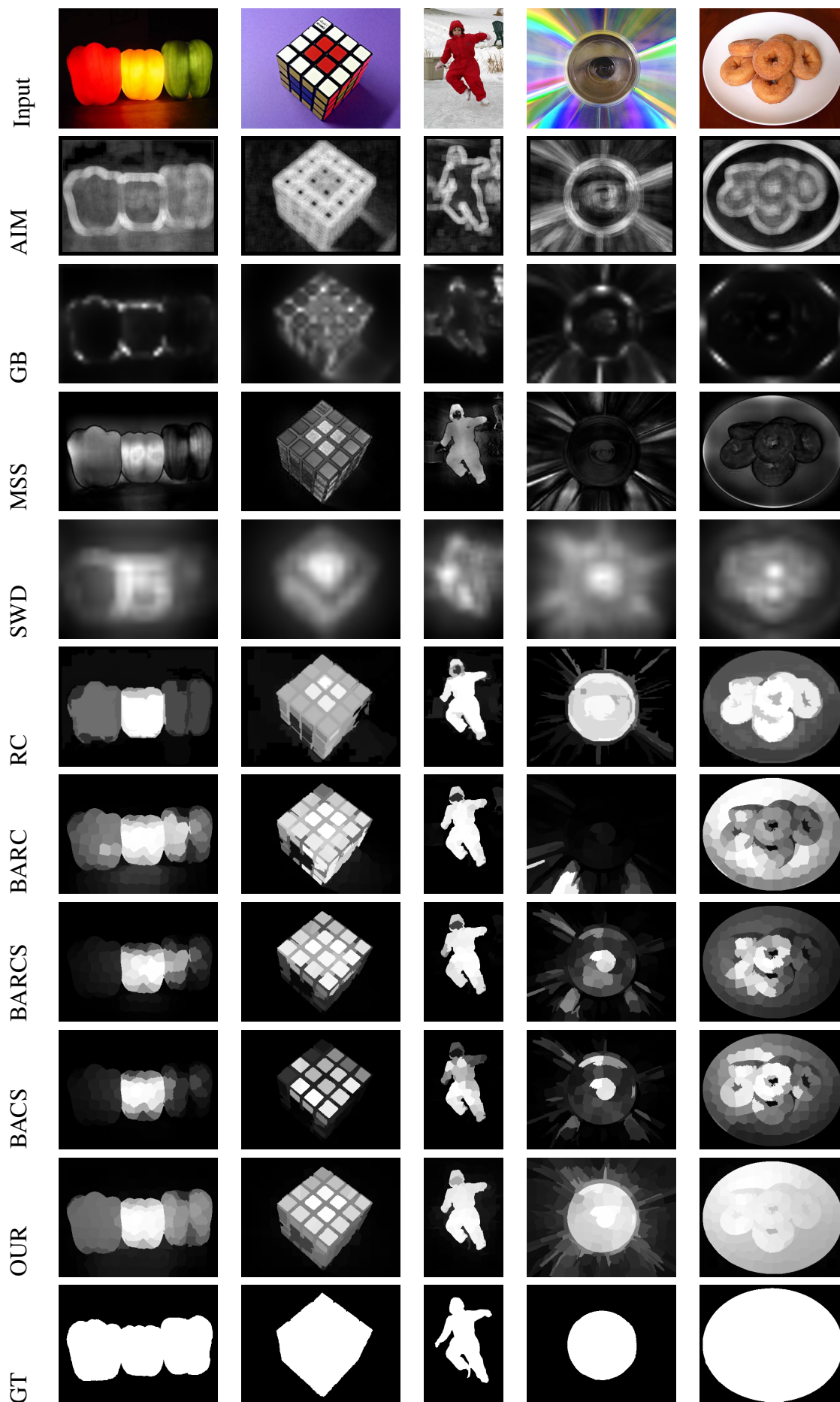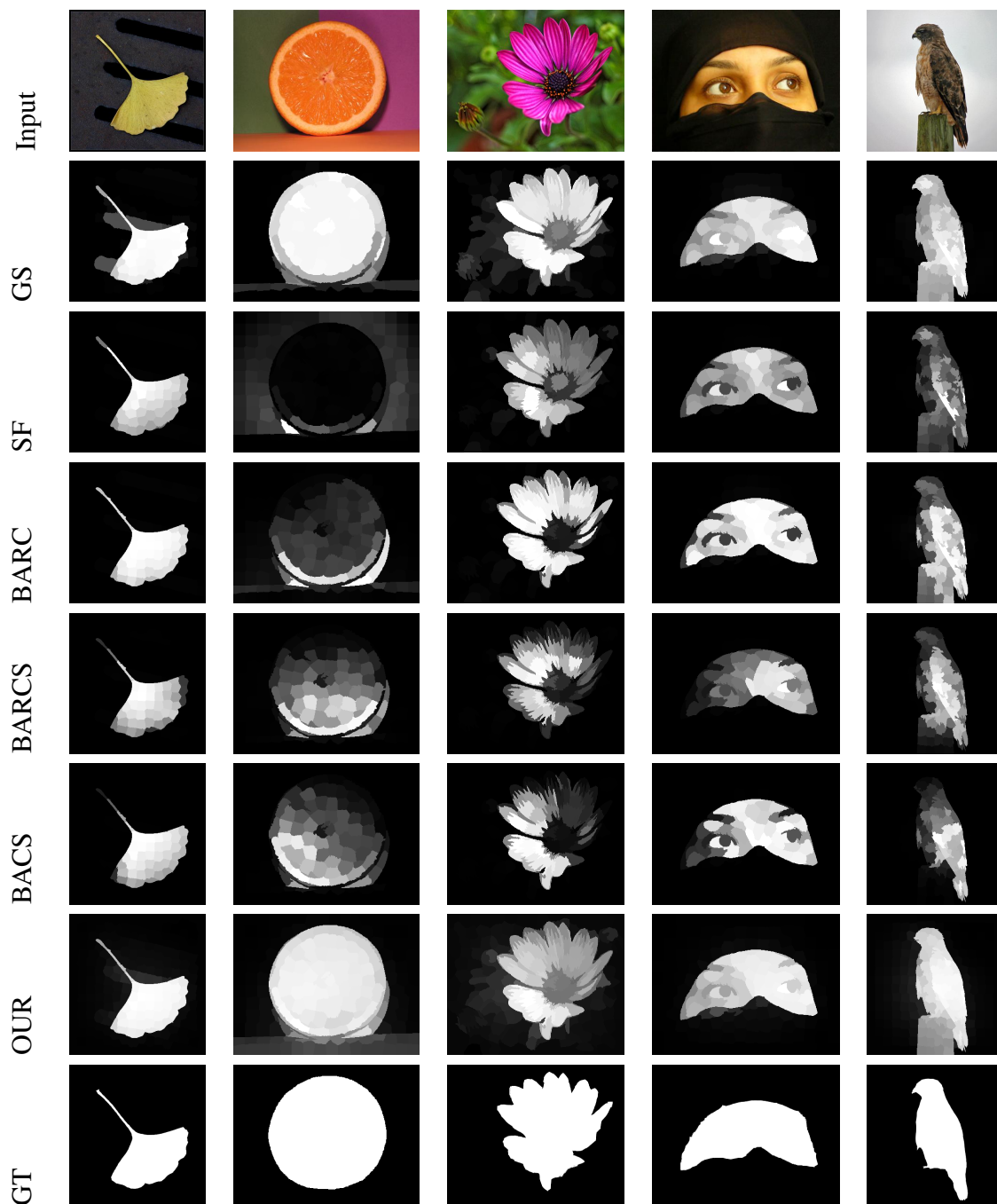**Figure 4.15:** Qualitative comparison of our saliency output with other state of the art methods on MSRA 10k dataset.
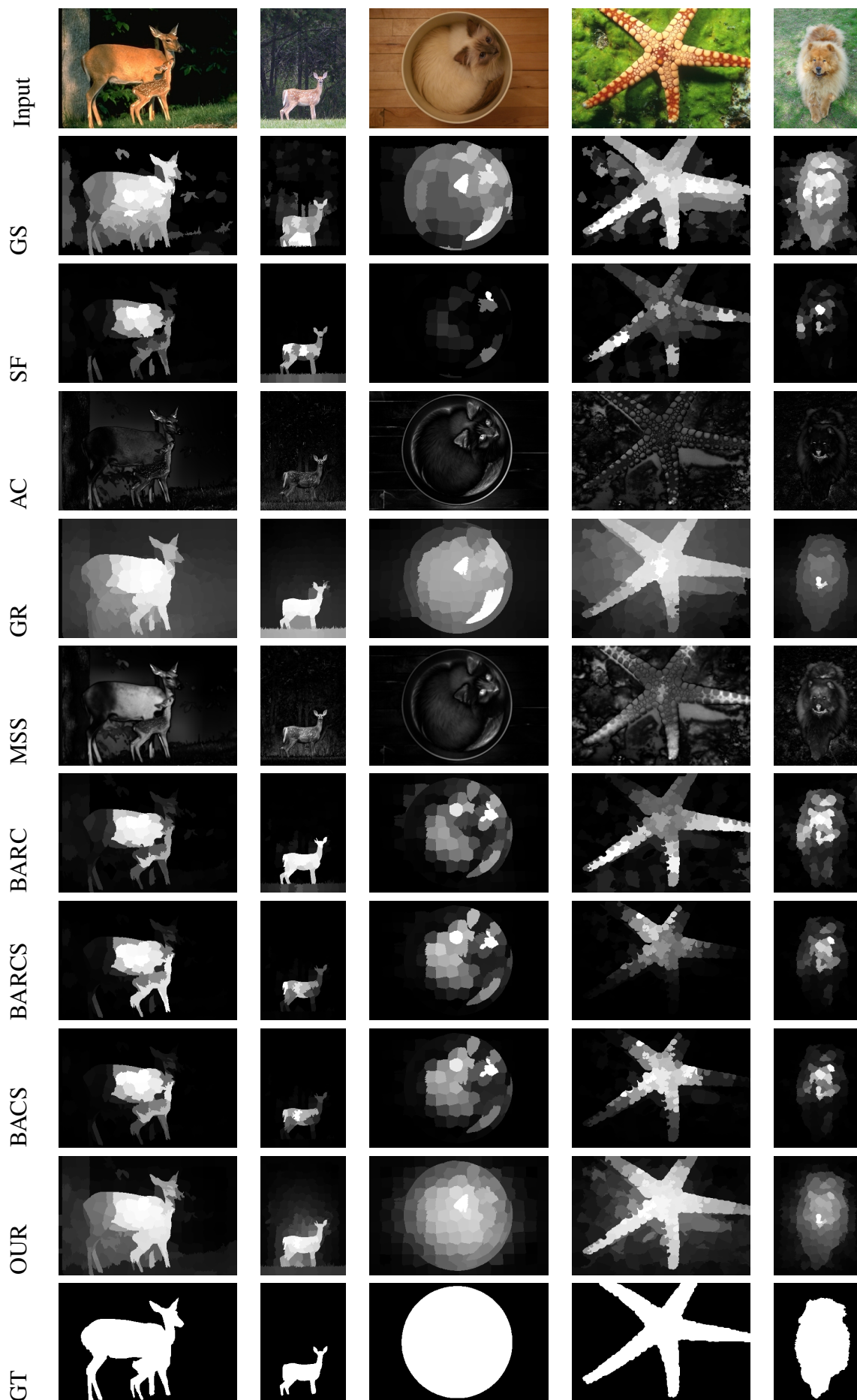
Figure 4.16: Qualitative comparison of our saliency output with other state of the art methods on CSSD dataset.
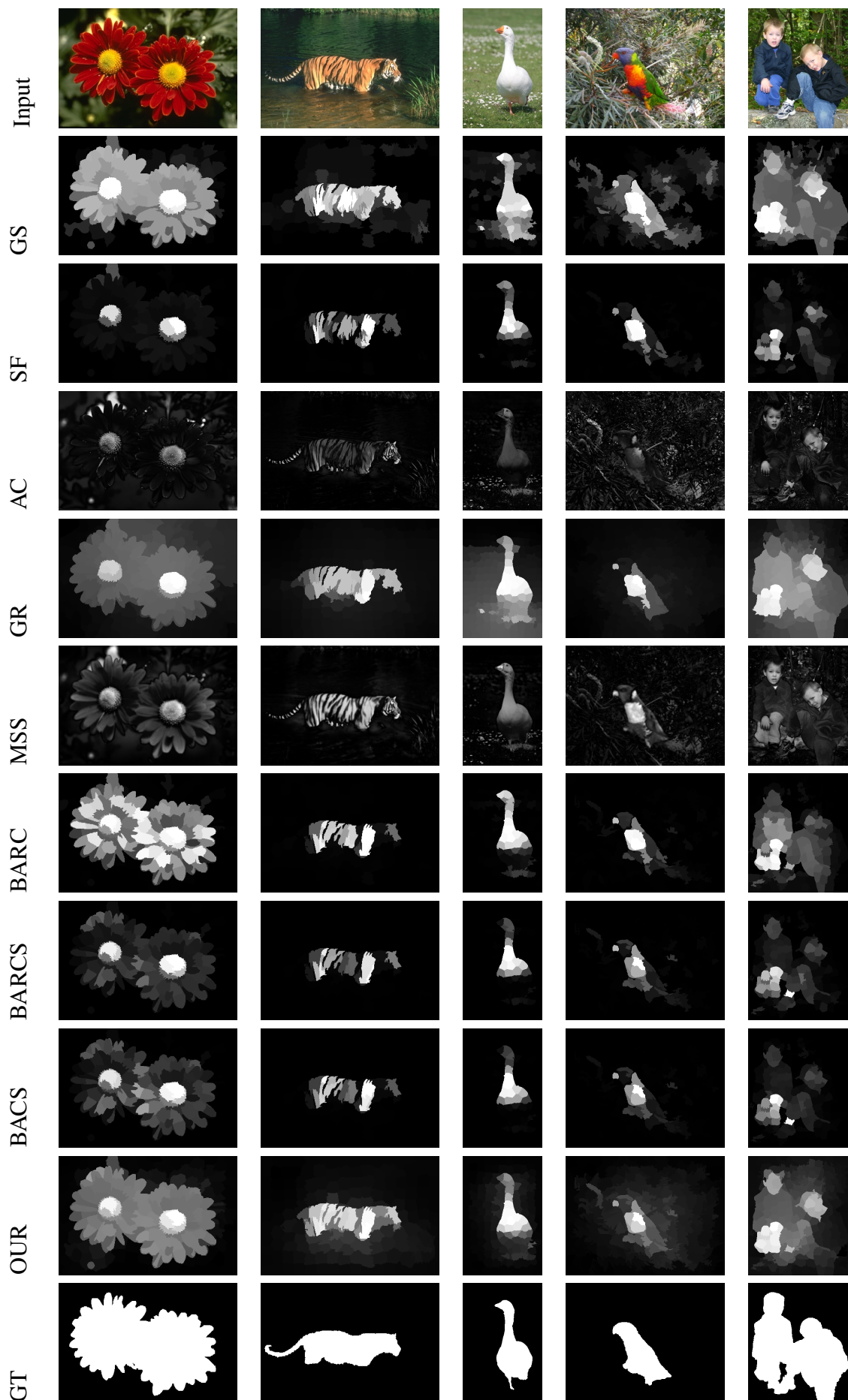
Figure 4.17: Qualitative comparison of our saliency output with other state of the art methods on ECSSD dataset.

4.7 Complex Cases and Miss Detections

In previous sections, it is seen that the proposed saliency model does better than other state of the art techniques in case of saliency detection. But, there are complex cases in each dataset where the proposed model may not detect well. Some of such cases are presented in Fig. 4.18. In this figure, input and ground truth images are presented along with BARC, BARCS and *Our* proposed saliency model. Each of the rows in this figure presents a comparatively complex case with a low contrast image. In spite of that, it is noticeable that the proposed model does well than other compared methods as seen in the fourth column of the figure. Specially, it performs outstanding in case of the first and third row. In other cases, the proposed model covers more area of the ground truth image than other compared approaches.

4.8 Application of the proposed method in pedestrian detection

In this study, the proposed saliency model is implemented in a simple HOG based pedestrian detection method to evaluate its performance in application level. Experiments reveal that the implementation of saliency detection makes the searching process faster as it can easily skip background regions from the search area. It also reduces false positives which are detected in the background.

A small annotated test dataset in combination of famous INRIA person dataset has been used to train the Hog based pedestrian model. The saliency detection technique is applied on some test images. First the salient region of the test image is calculated. Then the window based search technique is applied in the salient region. We only search the window for a possible pedestrian if the salient binary mask as seen in Fig. 4.20 of that window cover more than 40% of the foreground - i.e. white portion of the binary mask image. It reduces the search time as expected. Besides, it helps to reduce the false positives which resulted in the background region as seen in Fig. 4.19 and 4.21. The search time of the detection procedure is also reduced as is skips searching in pure backgrounds as detected by the saliency map. For the example in Fig. 4.21 search time is reduced by 24.30% by introducing saliency detection in the search procedure. This is applicable to most of the computer vision techniques. Computational time and performance of many application can be greatly improved by reducing the search area within the salient region.

Input	BARC	BARCS	OUR	GT
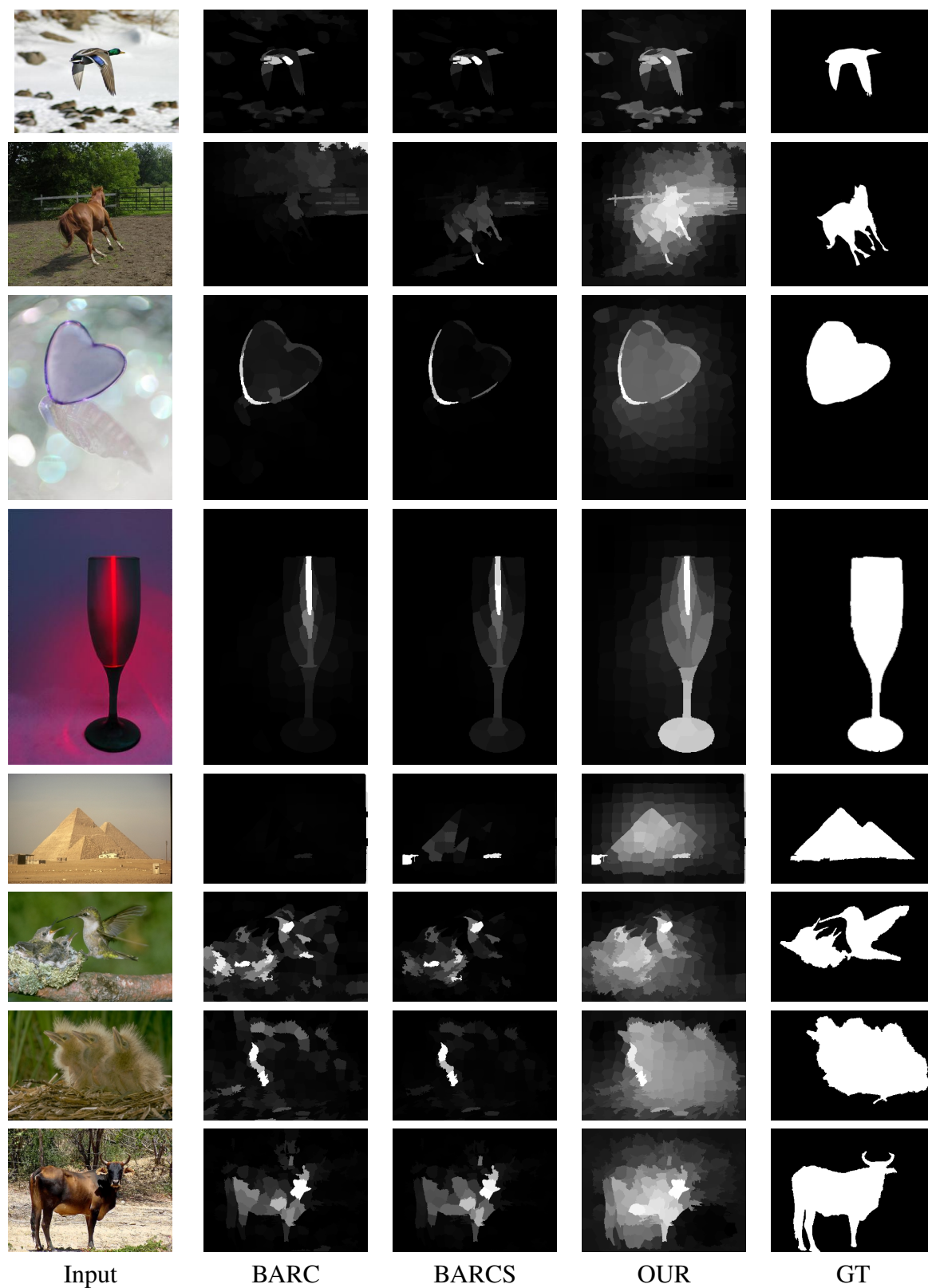
Figure 4.18: Qualitative comparison of proposed saliency output with a pair of complex members of each dataset namely - MSRA 1k, MSRA 10k, CSSD and ECSSD respectively.

Similar way, saliency detection can be applied to many image analysis system including but not limited to image classification, object detection, object tracking, semntic segmentation, person identification, image captioning, robot navigation, video surveillance, medical imaging etc. and significantly improve their performance in terms of time and accuracy.
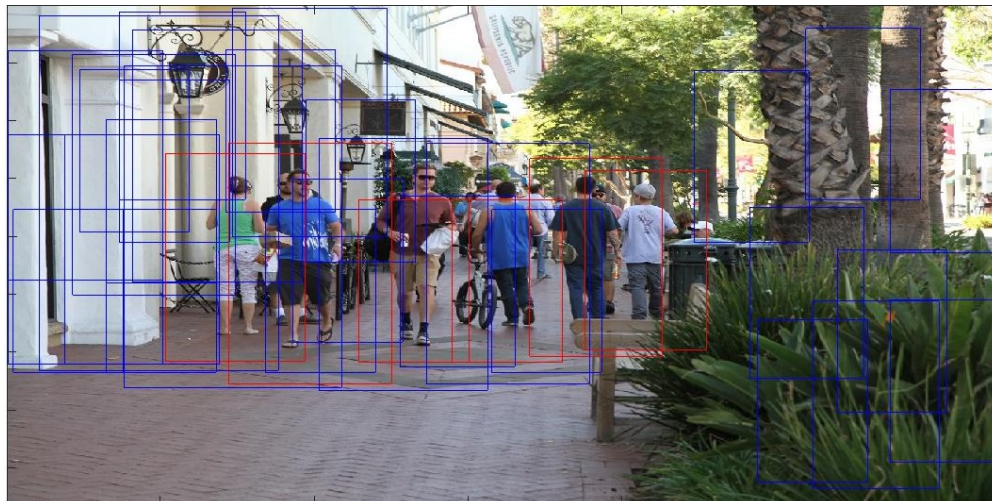
Figure 4.19: Pedestrian detection without saliency detection technique.

Figure 4.20: Binary mask of the salient region.

4.9 Summary

In this chapter, a brief history of the famous saliency datasets have been discussed. The performance of raw and superpixel input image has been explored with experimental results and explanation. In addition to that, two optimization techniques have been compared and the cost minimizing optimization has been suggested as the better one for this proposed architecture. In the result section, qualitative and quantitative analysis is done with several
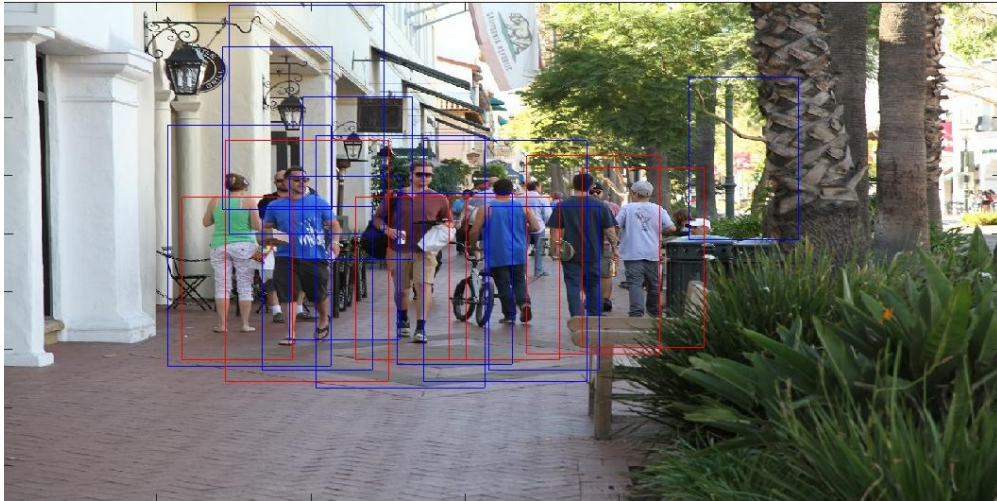
Figure 4.21: Pedestrian detection with saliency detection technique.

other state of the art saliency detection methods. Moreover, the use of the proposed method in pedestrian detection has been implemented and how the addition of this method significantly enhances the output of that simple implementation of pedestrian detection has been discussed. The results and discussion clearly demonstrates the superiority of the proposed architecture over other efficient technologies.

Chapter V

Conclusion

5.1 Introduction

The summary of this study is discussed in this chapter. Moreover, how the study can be useful in practical applications and the possible future research directions are suggested here.

5.2 Summary

A gradual enhancement of the saliency detection technique from the boundary aware regional contrast based method to the final proposed seam and color based boundary aware regional saliency detection procedure has been presented in this study. Introduction of the quadratic equation based optimization method has improved the result significantly. Some intermediate combinations are also presented with their detailed comparison. Though, one of the combinations is marked as the proposed solution balancing the runtime and performance, any of them can be used with respect to the application demand. Moreover, the presented results in the experimental section of this study clearly demonstrate the superiority of the proposed method over other state of the art techniques in terms of qualitative and quantitative analysis. The use of standard benchmark datasets strengthens the acceptability of this study and its publicly available results make an opportunity for other researches to compare their findings with the proposed method.

5.3 Recommendations

Besides the optimal final version of the proposed saliency model, different combinations relating the seam and color cues has been explored in this study. Some of them are computationally more efficient than their close competitors, others may show slightly more accurate result. So, the recommendation is to use the version as per requirement. Saliency detection techniques are usually useful in the preprocessing steps of different computer vision applications like - real time object detection, video surveillance etc. In those application level, detection time can be a crucial issue. In those cases, we suggest to experiment with less

computationally expensive computations from this study. On the other hand, there are applications like content aware image resizing, offline object search etc, where the time is not that much important but the exactness is the main concern. In these situations, we suggest to use the proposed final saliency model.

5.4 Future Scopes of Work

Computational devices are improving day by day. So, in future exactness of saliency detection technique will be the main concern. Moreover, fine tuned version of the proposed method is possible and its implementation in a application level programming language like C++ would certainly improve its runtime. Besides that, different saliency weight calculation procedures of this study like border contrast map, color map and contrast based dissimilarity can be parallelized and this way runtime can come to its minimum. The recent trend is moving towards learning based method like deep learning, random forest. So, in future the proposed saliency detection technique can be improved by using some learning architecture and can be applied into various computer vision related applications like pedestrian detection, action recognition, object detection and recognition etc.

References

[1]   T. Moore and M. Zirnsak. "Neural mechanisms of selective visual attention". In: *Annual review of psychology* 68 (2017), pp. 47–72.

[2]   S. K. Mannan, C. Kennard, and M. Husain. "The role of visual salience in directing eye movements in visual object agnosia". In: *Current biology* 19.6 (2009), R247–R248.

[3]   R. Desimone and J. Duncan. "Neural mechanisms of selective visual attention". In: *Annual review of neuroscience* 18.1 (1995), pp. 193–222.

[4]   S. Becker. "Which features guide visual attention, and how do they do it?" In: *Journal of Vision* 17 (2017), p. 6. ISSN: 1534-7362. DOI: `10.1167/17.10.6`.

[5]   H.-L. Teuber. "Physiological psychology". In: *Annual review of psychology* 6.1 (1955), pp. 267–296.

[6]   S. E. Taylor and S. T. Fiske. *Salience, Attention, and Attribution: Top of the Head Phenomena*. 1978. DOI: `10.1016/s0065-2601(08)60009-x`.

[7]   J. M. Wolfe and T. S. Horowitz. "What attributes guide the deployment of visual attention and how do they do it?" In: 5 (2004), pp. 495–501. ISSN: 1471-003X. DOI: `10.1038/nrn1411`.

[8]   L. Itti, C. Koch, and E. Niebur. "A model of saliency-based visual attention for rapid scene analysis". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.11 (Nov. 1998), pp. 1254–1259. ISSN: 0162-8828. DOI: `10.1109/34.730558`.

[9]   T. Lou and J. Xu. *Visual saliency based on natural scene statistics*. 2011. DOI: `10.1109/icnc.2011.6022196`.

[10] M. Xu, Y. Liu, R. Hu, and F. He. "Find Who to Look at: Turning From Action to Saliency". In: *IEEE Transactions on Image Processing* 27.9 (Sept. 2018), pp. 4529–4544. ISSN: 1941-0042. DOI: `10.1109/TIP.2018.2837106`.

[11] A. M. Treisman and G. Gelade. "A feature-integration theory of attention". In: *Cognitive psychology* 12.1 (1980), pp. 97–136. ISSN: 0010-0285. DOI: `10.1016/0010-0285(80)90005-5`.

[12] C. Koch and S. Ullman. *Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry*. 1987. DOI: `10.1007/978-94-009-3833-5_5`.

[13] S. M. M. Ahsan, J. K. Tan, H. Kim, and S. Ishikawa. "Boundary Aware Regional Contrast Based Visual Saliency Detection". In: *The Twenty-First Int. Symposium on Artificial Life and Robotics*. 2016, pp. 258–262.

[14] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. Yang. "Saliency Detection via Graph-Based Manifold Ranking". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 2013, pp. 3166–3173. DOI: 10.1109/CVPR.2013.407.

[15] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. "Frequency-tuned salient region detection". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 2009, pp. 1597–1604. DOI: 10.1109/CVPR.2009.5206596.

[16] S. Goferman, L. Zelnik-Manor, and A. Tal. "Context-Aware Saliency Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.10 (Oct. 2012), pp. 1915–1926. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.272.

[17] Y. Wei, F. Wen, W. Zhu, and J. Sun. "Geodesic Saliency Using Background Priors". In: *12th European Conf. on Computer Vision (ECCV), Proc., Part III*. Vol. 7574. Springer, Oct. 2012, pp. 29–42. DOI: 10.1007/978-3-642-33712-3_3.

[18] Q. Yan, L. Xu, J. Shi, and J. Jia. "Hierarchical Saliency Detection". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 2013, pp. 1155–1162. DOI: 10.1109/CVPR.2013.153.

[19] Z. Jiang and L. S. Davis. "Submodular Salient Region Detection". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 2013, pp. 2043–2050. DOI: 10.1109/CVPR.2013.266.

[20] A. Islam, S. M. M. Ahsan, and J. K. Tan. "Saliency Detection using the Combination of Boundary Aware Color-map and Seam-map". In: *International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*. 2019.

[21] A. Islam, S. M. M. Ahsan, and J. K. Tan. "Saliency Detection using Boundary Aware Regional Contrast Based Seam-map". In: *Proc. Int. Conf. Innovation in Engineering and Technology (ICIET)*. 2018. DOI: 10.1109/CIET.2018.8660825.

[22] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. "Salient Object Detection: A Discriminative Regional Feature Integration Approach". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 2013, pp. 2083–2090. DOI: 10.1109/CVPR.2013.271.

[23]   S. Avidan and A. Shamir. "Seam Carving for Content-aware Image Resizing". In: *ACM Trans. Graph.* 26.3 (July 2007). ISSN: 0730-0301. DOI: `10.1145/1276377.1276390`.

[24]   Y. Li, K. Fu, L. Zhou, Y. Qiao, and J. Yang. "Saliency detection via foreground rendering and background exclusion". In: *Proc. IEEE Int. Conf. Image Processing (ICIP)*. Oct. 2014, pp. 3263–3267. DOI: `10.1109/ICIP.2014.7025660`.

[25]   T. G. Joy and M. M. Hossan. "A Study on Salient Region Detection using Bounday and Color Cue". In: *Khulna University of Engineering & Thechnology* (2018). Thesis no. CSER-18-26.

[26]   R. Achanta and S. Süsstrunk. "Saliency detection for content-aware image resizing". In: *Proc. 16th IEEE Int. Conf. Image Processing (ICIP)*. Nov. 2009, pp. 1005–1008. DOI: `10.1109/ICIP.2009.5413815`.

[27]   Y. Chou, C. Fang, P. Su, and Y. Chien. "Content-Based cropping using visual saliency and blur detection". In: *Proc. 10th Int. Conf. Ubi-media Computing and Workshops (Ubi-Media)*. Aug. 2017. DOI: `10.1109/UMEDIA.2017.8074087`.

[28]   S. Chebbout and H. F. Merouani. "An object segmentation method based on saliency map and spectral clustering". In: *Proc. World Congress Information Technology and Computer Applications (WCITCA)*. 2015. DOI: `10.1109/WCITCA.2015.7367036`.

[29]   X. Wu, X. Ma, J. Zhang, A. Wang, and Z. Jin. "Salient Object Detection Via Deformed Smoothness Constraint". In: *Proc. 25th IEEE Int. Conf. Image Processing (ICIP)*. Oct. 2018, pp. 2815–2819. DOI: `10.1109/ICIP.2018.8451169`.

[30]   W. Zhu, S. Liang, Y. Wei, and J. Sun. "Saliency Optimization from Robust Background Detection". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 2014, pp. 2814–2821. DOI: `10.1109/CVPR.2014.360`.

[31]   Ren and Malik. "Learning a classification model for segmentation". In: *Proc. Ninth IEEE Int. Conf. Computer Vision*. Oct. 2003, 10–17 vol.1. DOI: `10.1109/ICCV.2003.1238308`.

[32]   R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.11 (Nov. 2012), pp. 2274–2282. ISSN: 0162-8828. DOI: `10.1109/TPAMI.2012.120`.

[33] R. Achanta and S. Süsstrunk. "Saliency detection using maximum symmetric sur-round". In: *Proc. IEEE Int. Conf. Image Processing*. Sept. 2010, pp. 2653–2656. DOI: `10.1109/ICIP.2010.5652636`.

[34] J. H. Reynolds and R. Desimone. "Interacting Roles of Attention and Visual Salience in V4". In: 37 (2003), pp. 853–863. ISSN: 0896-6273. DOI: `10.1016/s0896-6273(03)00097-7`.

[35] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu. "Global contrast based salient region detection". In: *Proc. CVPR 2011*. June 2011, pp. 409–416. DOI: `10.1109/CVPR.2011.5995344`.

[36] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang. *Saliency Detection via Dense and Sparse Reconstruction*. 2013. DOI: `10.1109/iccv.2013.370`.

[37] X. Shen and Y. Wu. *A unified approach to salient object detection via low rank matrix recovery*. 2012. DOI: `10.1109/cvpr.2012.6247758`.

[38] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. *Saliency Detection via Graph-Based Manifold Ranking*. 2013. DOI: `10.1109/cvpr.2013.407`.

[39] Z. Jiang and L. S. Davis. *Submodular Salient Region Detection*. 2013. DOI: `10.1109/cvpr.2013.266`.

[40] J. Zhang, Y. Dai, and F. Porikli. *Deep Salient Object Detection by Integrating Multi-level Cues*. 2017. DOI: `10.1109/wacv.2017.8`.

[41] L. Wang, H. Lu, X. Ruan, and M.-H. Yang. *Deep networks for saliency detection via local estimation and global search*. 2015. DOI: `10.1109/cvpr.2015.7298938`.

[42] R. Zhao, W. Ouyang, H. Li, and X. Wang. *Saliency detection by multi-context deep learning*. 2015. DOI: `10.1109/cvpr.2015.7298731`.

[43] and Guanbin Li and Y. Yu. "Visual saliency based on multiscale deep features". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 5455–5463. DOI: `10.1109/CVPR.2015.7299184`.

[44] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan. "Salient Object Detection with Recurrent Fully Convolutional Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.7 (July 2019), pp. 1734–1746. ISSN: 1939-3539. DOI: `10.1109/TPAMI.2018.2846598`.

[45] G. Li and Y. Yu. "Visual Saliency Detection Based on Multiscale Deep CNN Features". In: *IEEE Transactions on Image Processing* 25.11 (Nov. 2016), pp. 5012–5024. ISSN: 1941-0042. DOI: `10.1109/TIP.2016.2602079`.

[46] S. Nah and K. M. Lee. "Random forest with data ensemble for saliency detection". In: *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA)*. Dec. 2015, pp. 604–607. DOI: `10.1109/APSIPA.2015.7415340`.

[47] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. *The Secrets of Salient Object Segmentation*. 2014. DOI: `10.1109/cvpr.2014.43`.

[48] J. Harel, C. Koch, and P. Perona. "Graph-based visual saliency". In: *Advances in neural information processing systems*. 2007, pp. 545–552.

[49] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum. "Learning to Detect A Salient Object". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 2007, pp. 1–8. DOI: `10.1109/CVPR.2007.383047`.

[50] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung. "Saliency filters: Contrast based filtering for salient region detection". In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. June 2012, pp. 733–740. DOI: `10.1109/CVPR.2012.6247743`.

[51] A. Borji, M. Cheng, H. Jiang, and J. Li. "Salient Object Detection: A Benchmark". In: *IEEE Transactions on Image Processing* 24.12 (Dec. 2015), pp. 5706–5722. DOI: `10.1109/TIP.2015.2487833`.

[52] H. Zhang and C. Xia. "Saliency Detection combining Multi-layer Integration algorithm with background prior and energy function". In: *arXiv preprint arXiv:1603.01684* (2016).

[53] A. Richtsfeld, M. Zillich, and M. Vincze. "Implementation of Gestalt principles for object segmentation". In: *Proc. 21st Int. Conf. Pattern Recognition (ICPR2012)*. Nov. 2012, pp. 1330–1333.

[54] Z. Wang and B. Li. "A two-stage approach to saliency detection in images". In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2008, pp. 965–968.

[55]   J. Shi, Q. Yan, L. Xu, and J. Jia. "Hierarchical Image Saliency Detection on Extended CSSD". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.4 (Apr. 2016), pp. 717–729. DOI: `10.1109/TPAMI.2015.2465960`.

[56]   L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu. "Visual saliency detection by spatially weighted dissimilarity". In: *Proc. CVPR*. 2011, pp. 473–480. DOI: `10.1109/CVPR. 2011.5995676`.

[57]   M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu. "Global Contrast Based Salient Region Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.3 (Mar. 2015), pp. 569–582. ISSN: 1939-3539. DOI: `10.1109/ TPAMI.2014.2345401`.

[58]   R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk. "Salient region detection and segmentation". In: *International conference on computer vision systems*. Springer. 2008, pp. 66–75.

[59]   C. Yang, L. Zhang, and H. Lu. "Graph-regularized saliency detection with convex-hull-based center prior". In: *IEEE Signal Processing Letters* 20.7 (2013), pp. 637–640.

[60]   Y. Zhai and M. Shah. "Visual Attention Detection in Video Sequences Using Spatiotemporal Cues". In: *Proc. of the 14th ACM Int. Conf. on Multimedia*. MM '06. Santa Barbara, CA, USA: ACM, 2006, pp. 815–824. ISBN: 1-59593-447-2. DOI: `10. 1145/1180639.1180824`.

[61]   N. D. B. Bruce and J. K. Tsotsos. "Saliency, attention, and visual search: An information theoretic approach". In: *Journal of Vision* 9 (2009), pp. 5–5. ISSN: 1534-7362. DOI: `10.1167/9.3.5`.

[62]   X. Hou and L. Zhang. "Saliency detection: A spectral residual approach". In: *2007 IEEE Conference on computer vision and pattern recognition*. Ieee. 2007, pp. 1–8.

[63]   H. Chen, P. Wang, and M. Liu. "From co-saliency detection to object co-segmentation: A unified multi-stage low-rank matrix recovery approach". In: *Proc. IEEE Int. Conf. Robotics and Biomimetics (ROBIO)*. Dec. 2015, pp. 1602–1607. DOI: `10.1109/ ROBIO.2015.7419000`.

List of Publications

[1] A. Islam, S. M. M. Ahsan and J. K. Tan, "Saliency detection using the combination of boundary aware color-map and seam-map," in *5th IEEE Intl. Conf. on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Bangladesh*, 2019.

[2] A. Islam, S. M. M. Ahsan and J. K. Tan, "Saliency detection using boundary aware regional contrast based seam-map," in *IEEE International Conference on Innovation in Engineering and Technology (ICIET), Bangladesh*, 2018.

[3] S. M. M. Ahsan and A. Islam, "Saliency Detection using Seam and Color Cues," *Computer Science Journal (ISSN: 1508-2806; e-ISSN: 2300-7036).* [submitted]