Chapter 4

Salient Object Detection using background subtraction, Gabor filters, objectness and directional backgroundness

This chapter describes the Bayesian classifier based model developed for salient object detection. The model is introduced in Section 4.1. The theoretical foundations of the proposed model and brief introduction of state-of-the-art methods are described in Section 4.2. The description of the proposed model is given in Section 4.3. Section 4.4 gives the result generated by the proposed model and its analysis. Section 4.5 concludes the chapter.

4.1 Background

The main contribution of this chapter is to propose a method that solves the salient object detection problem using background subtraction, Gabor filters, minimum directional backgroundness, and objectness. The first step is to calculate a backgroundness score for each region by calculating the difference between the feature vector of image boundary and image regions. This backgroundness map is then used for calculating the minimum directional background difference. The image is segmented using Gabor filters, and then the objectness criterion is used to choose the segment containing the salient object. The normalized foreground saliency map is then used to refine the selected segment. Further enhancement of this intermediate output is done using morphological operations, and boundary correction are done using the method of lazy snapping. The algorithm is tested on eight publicly available datasets and is compared against five algorithms. The performance is evaluated by PR-curve, F-Measure curve, and Mean Absolute Error.

4.2 Literature of backgroundness, Gabor filters, objectness and lazy snapping

This section explains the related work that has been done in this field and is used to develop the model of this chapter.

In [188], the authors treat the salient object detection as a regression problem.

Authors perform multi-level segmentation of images and generate regional saliency maps for each of the segments. High-level integration of these regional saliency maps is performed in such a way that the discriminative features are preserved. Regional saliency maps use local contrast and backgroundness scores. Instead of concatenating various feature cues on a trial-and-error basis, the feature vector of a region is directly mapped to a saliency score using Random Forest regressor. In this work, the backgroundness score is used to generate the background saliency map.

In [189], the authors provide a texture segmentation procedure by multi-channel filtering of images using Gabor filter bank over the entire spatial-frequency domain. The filtered images are passed through a selection scheme for using them for reconstructing the original image. To calculate the energy for each pixel in a neighborhood window, the selected filtered images are non-linearly transformed. These feature images are then clustered for integration and segmentation using spatial adjacency. Psychophysiological experiments support the use of multi-channel texture filters to understand the human visual system. The proposed model uses the Gabor segmentation method, as explained by the authors.

In [190], the authors, try to measure the probability of an image window of containing an object. For this, the authors calculate multiple image cues like –

- Multi-scale saliency: In this, the spectral residue of Fast Fourier Transform (FFT) of an image is calculated at multiple scales.
- Colour Contrast: This is used to calculate how different the current image window is from the surrounding area.

- Edge Density: It helps in marking strong edges at image windows.
- Superpixel Straddling: It is used to check whether the image window completely covers the object.

These cues are then fused in a Bayesian framework. Authors propose the idea that combined saliency cues perform better than stand-alone saliency cues. The objectness score helps to find the location to find a salient object. The model is independent of class restriction, i.e., it can find objects belonging to various class categories. The proposed model uses this work to find areas of the image with high objectness to locate the salient object.

In [177], the authors, try to calculate foreground saliency based on minimum directional contrast. The authors utilize the spatial distribution of contrast, which was ignored in previous works. The previous works dealt only with the global distribution of contrast. For every pixel, directional contrast is calculated in four directions: Top left, top right, bottom right, and bottom left. The minimum of these four values is the minimum directional contrast of the pixel. Since foreground objects have to stand apart from the background, they have high contrast in all directions. Background pixels have to connect to the background; they have less contrast in at least one direction. Thus, the minimum directional contrast of a foreground pixel is higher than the minimum directional contrast of a background pixel. The minimum directional contrast value is then used to build a saliency map. From this work, the proposed model takes the cue of generating a foreground saliency map using backgroundness scores instead of intensity values.

The work in [191] separates the image background from the foreground despite the presence of weak edges and edges with low contrast. It combines graph-cut and over-segmentation of the image. The authors rely on marking the foreground object and editing the boundary of the object. The model claims that it outperforms Magnetic Lasso in terms of ease and efficiency. The proposed model uses this method to correct the boundary of the intermediate result to produce the final salient object.

In this chapter, the proposed algorithm is compared against five state-of-the-art algorithms: FBP [192], MDC [177], MST [193], EQCUT [194] and SGP [195].

In FBP[1] [192], authors generate foreground and background map separately and fuse them using geodesic distance information.

MDC [177] uses minimum directional contrast to find the salient object in the image.

MST [193] uses minimum spanning tree to find the object geometry in a scene.

EQCUT [194] forms a graph from superpixel segmentation, and it performs optimization related to image boundary, area, and local contrast.

In SGP [195], authors try to emulate human eye fixations by using Super-Gaussian Component response.

[1]No working code was available for the algorithm. Only outputs on ASD dataset were available.

4.3 Methods and Models

This section explains the methods used to find the salient object. Subsection 4.3.1 explains the background subtraction technique. Subsection 4.3.2 explains the process of calculating Minimum Directional Backgroundness. Subsection 4.3.3 explains the method of image segmentation using Gabor filters. Subsection 4.3.4 explains the objectness approach. Subsection 4.3.5 explains morphological methods that enhance the intermediate output. Subsection 4.3.6 describes the Lazy Snapping approach for boundary correction. Subsection 4.3.7 demonstrates the algorithm of the proposed model.

4.3.1 Background Subtraction

For salient object detection, one of the methods to focus on foreground objects is to subtract the background from the image. To find the background of an image, the proposed model uses the technique suggested by [188]. The authors extract a 15-pixel wide boundary from the image and calculate the following features: Average RGB values, Histogram of RGB boundary, Average LAB values, Histogram of LAB boundary, Average HSV values, Histogram of HSV boundary, LBP feature, the absolute response of LM filters, Histogram of max response of LM filters. From these features, the authors form a 29-dimensional feature vector. Then patches are taken from the image regions, and the feature vector of each patch is calculated. The distance between the background feature vector and the local image patches

TABLE 4.1: Features for calculating backgroundness score

Color and Texture Features		Difference of Features	
Features	Dimensions	Definition	Dimensions
Average RGB Values	3	$d\left(a_1^{R_i}, a_1^S\right)$	3
RGB Histogram	256	$\chi^2\left(h_1^{R_i}, h_1^S\right)$	1
Average HSV Values	3	$d\left(a_2^{R_i}, a_2^S\right)$	3
HSV Histogram	256	$\chi^2\left(h_2^{R_i}, h_2^S\right)$	1
Average LAB Values	3	$d\left(a_3^{R_i}, a_3^S\right)$	3
LAB Histogram	256	$\chi^2\left(h_3^{R_i}, h_3^S\right)$	1
Histogram	59	$\chi^2\left(h_4^{R_i}, h_4^S\right)$	1

feature vector is then used to calculate the backgroundness score. The higher the score obtained, the more the region is away from the background. Similarly, the smaller the score, the closer the image region is to the background.

Out of these features, the proposed model eliminates the use of LM filters and thus proceeded only with RGB, LAB, HSV, and LBP features of the image. LM filters are used for texture analysis. Since, in this work, Gabor filter segmentation is used, the response of LM filters was not proving useful. So to cut time cost, the proposed model eliminates LM filters. A 13-dimensional feature vector is obtained. The proposed model takes 15×15 patch for an image of size 300×300. Thus, the backgroundness score for 400 image regions is obtained. The TABLE 4.1 describes the color and texture features that are used to generate the backgroundness score. The distances are defined as follows:

$$d\left(x_1, x_2\right) = \left(\left|x_{11} - x_{21}\right|, \ldots, \left|x_{1c} - x_{2c}\right|\right) \tag{4.1}$$

where c is the number of channels.

$$\chi^2(h_1, h_2) = \sum_{i=1}^{b} \frac{2(h_{1i} - h_{2i})^2}{h_{1i} + h_{2i}} \tag{4.2}$$

where b is the number of histogram bins. The backgroundness score for a region is calculated as

$$y(R_i) = \sum_{i=1}^{N} D\left(v^{R_i}, v^B\right) \tag{4.3}$$

where R_i is a region of the image, v^{R_i} is the feature vector of the region and v^B is the feature vector of background. $D(v^{R_i}, v^B)$ represents the difference between the feature vectors. N is the number of patches.

FIGURE 4.1 shows the background saliency maps that are built from the backgroundness score.

4.3.2 Saliency Map using Minimum Directional Backgroundness (MDB)

For salient object detection, the proposed model also builds a foreground identification saliency map taking hints from a similar saliency map introduced by Huang and Zhang [177] for minimum directional contrast. In [177], values for each channel of CIE-lab color space is used to find minimum directional contrast for each pixel. In this work, instead of using values from a color space, the backgroundness score is used for generating the foreground saliency map. FIGURE 4.2 shows the four

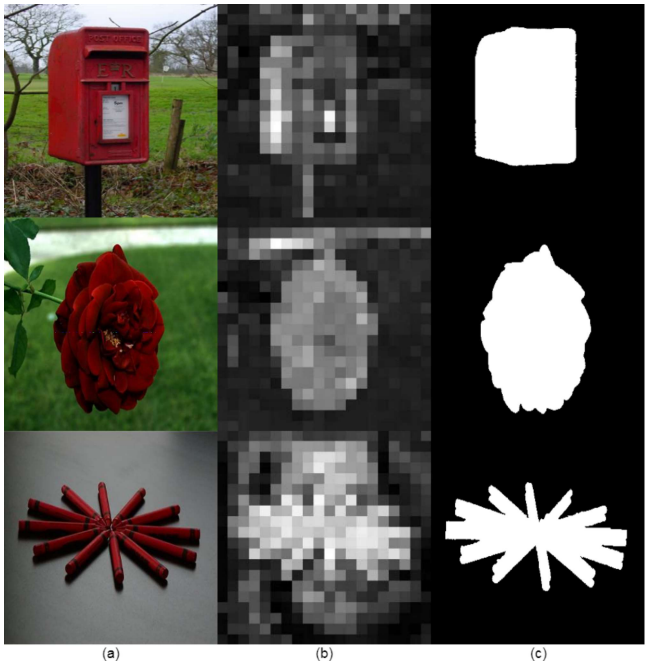


FIGURE 4.1: Background Saliency Map. (a) Original Image (b) Background Saliency Maps (c) Ground Truth.

directions concerning the pixel n. The regional backgroundness score gives an idea about foreground pixels. Foreground pixels will have a higher difference from all regions in all directions, whereas background pixels will have a higher difference from a maximum of three directions as they form the background in one direction. So, MDB for a foreground pixel will be higher than a background pixel.

By using backgroundness scores, it is ensured that the background saliency map and foreground saliency map correspond to each other. The saliency map thus formed helps in the clear demarcation of salient and non-salient pixels. The complexity is also reduced since backgroundness score has only one channel, whereas, in [177], it is calculated for three channels. This also reduces the time it takes for the computation.

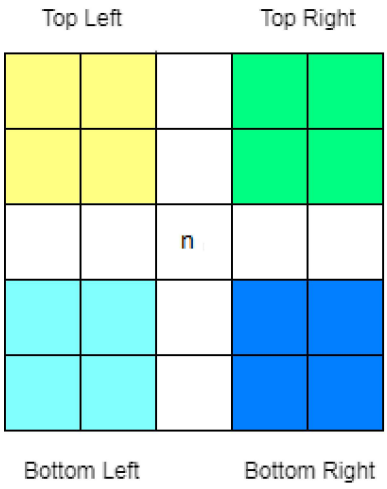


FIGURE 4.2: Dividing an image into four regions with respect to center pixel n.

The directional backgroundness for a pixel n is calculated as

$$DC_{n,\Gamma} = \sqrt{\sum_{o \in \Gamma} (I_n - I_o)^2} \tag{4.4}$$

where I represents the backgroundness score map . Γ represents the spatial domain of the image.

The minimum directional backgroundness for the pixel n is calculated as

$$V(n) = \sqrt{min\left(\sum_{o \in \Gamma} (I_n - I_o)^2\right)} \quad \Gamma \in TL, TR, BL, BR \tag{4.5}$$

where TL, TR, BL and BR represent Top-left, Top-Right, Bottom-Left and Bottom-Right of the image with respect to pixel n.

The foreground saliency map obtained is normalized. All the pixels having value less than 0.6 are set to 0. The saliency map thus obtained is called the truncated

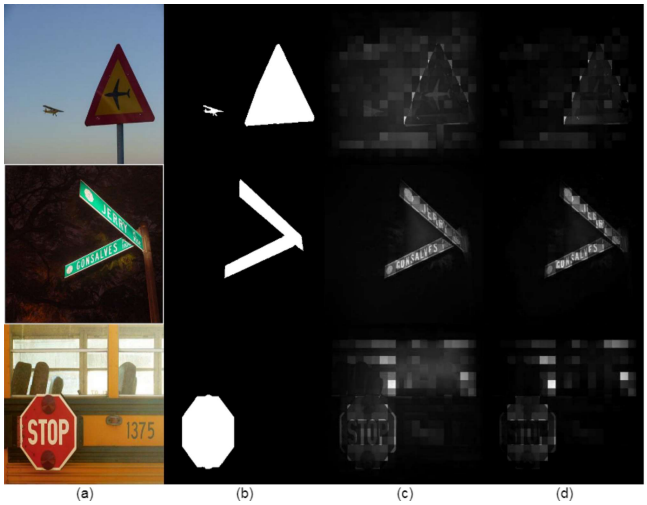


FIGURE 4.3: Foreground Saliency Map. (a) Original Image (b) Ground Truth (c) Saliency Map (d) Normalized Truncated Saliency Map.

normalized saliency map. FIGURE 4.3 shows some images with their saliency maps and normalized truncated saliency maps.

4.3.3 Texture segmentation using Gabor Filters

The vision system of mammals for distinguishing textures can be modeled using Gabor filters. So for salient object detection, the separation of the image into textures can help to distinguish between foreground and background. The separation using Gabor filters is inspired by the work of Jain and Farrokhnia [189], where they provide the values for orientation and wavelength used for generating Gabor filters. The following equation represents a generic Gabor filter in the spatial domain:

$$h\left(x,y\right) = exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right]\right\}\cos\left(2\pi u_0 x+\phi\right) \tag{4.6}$$

where u_0 and ϕ are the frequency and phase of the sinusoidal plane wave along 0° orientation. σ_x and σ_y are the Gaussian smoothing constants in x- and y- directions. For the proposed model, the image is segmented in three regions. The segmentation into three regions gives us a more appropriate location of finding the location of the object of interest. The segmentation of the image into two regions adds many background details into the region of interest, so the proposed model proceeds with three regions. The orientations used for generating Gabor filters are sampled between 0° and 135° in steps of 45 degrees. Wavelengths are sampled in increasing power of 2 starting from $\frac{4}{\sqrt{2}}$ till the length of the hypotenuse of the input image. In the case of the proposed model, the image size is 300×300. So wavelengths are sampled from $\frac{4}{\sqrt{2}}$,$\frac{8}{\sqrt{2}}$,$\frac{16}{\sqrt{2}}$,. . . ,$\frac{128}{\sqrt{2}}$. Thus, there are a total of 24 filters used: 4 orientations and six radial frequencies. The magnitude response of each filter is post-processed using Gaussian smoothing, the addition of spatial information (both in x and y-direction), and normalizing the features of different filters to have a common mean and variance. After the post-processing, the proposed model has, for each pixel, 26 features – 24 Gabor features from 24 Gabor filters and two features of spatial information. These features are then fed to the clustering algorithm for segmentation. FIGURE 4.4 shows some examples of texture segmentation using Gabor filters.

4.3.4 Objectness

It involves the process of finding a region that has the highest probability of containing an object. It acts as a controller for other methods in this work. The proposed

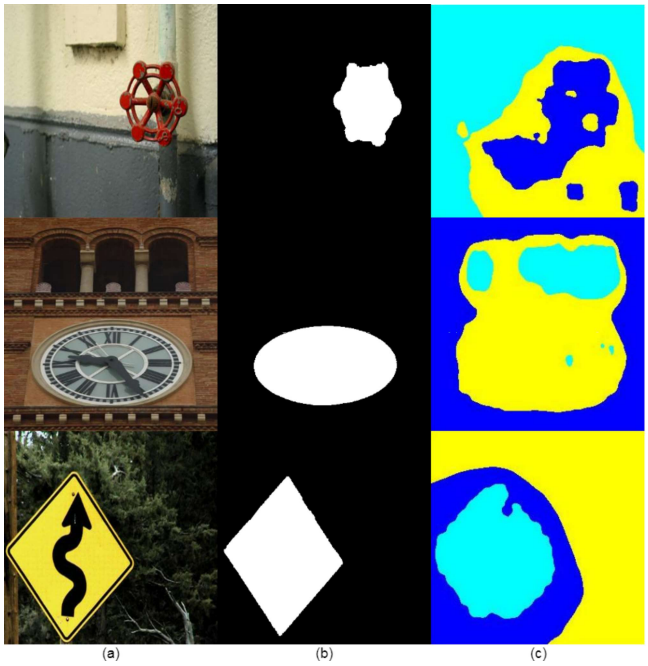


FIGURE 4.4: Image segmentation using Gabor filters. (a) Original Image (b) Ground Truth (c) Segmented Image.

model uses the Objectness method proposed by Alexe *et al* [190]. In this work, the authors propose ten bounding boxes. From all these bounding boxes, the proposed model finds the top three regions which are present in maximum bounding boxes using a Bayesian classifier. These top three regions are further used to identify the regions where the salient object might be present.

The images in FIGURE 4.5 illustrate the use of objectness in locating salient objects. From all the three segments that are produced by Gabor filters, the one that contains the region with maximum objectness is selected to proceed for further enhancements.

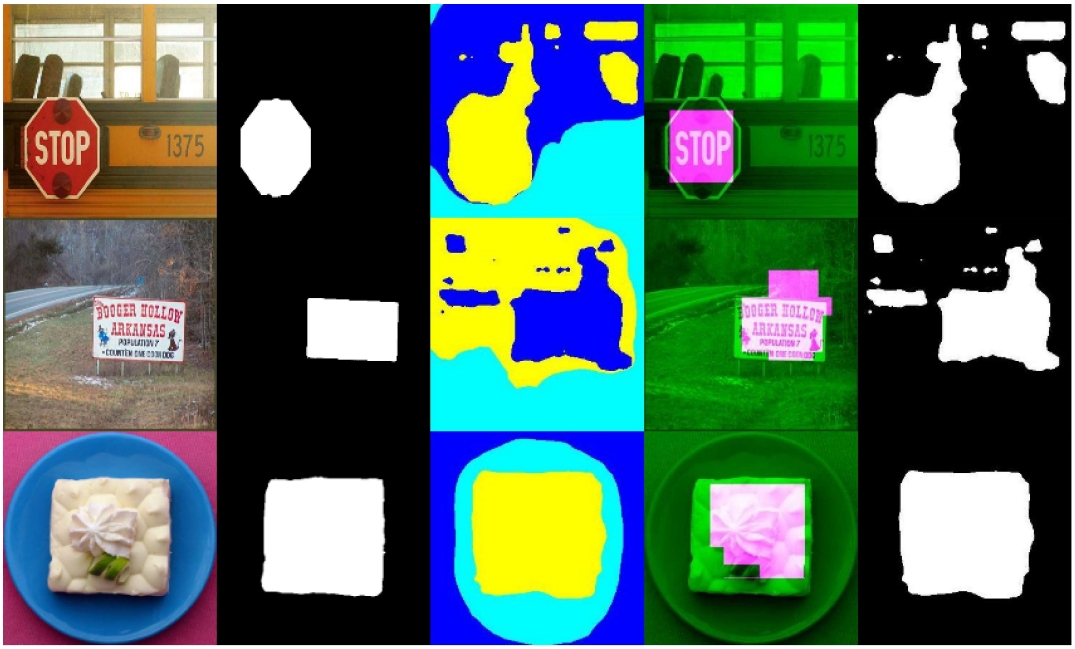


FIGURE 4.5: Using objectness to find segments that contain salient object. (a) Original Image (b) Ground Truth (c) Segmented Image (d) Region which has high objectness value, shown in pink shaded region (e) Segment of image which contains the salient object.

4.3.5 Morphological Operations

After the salient region is obtained from the above processes, the final region is obtained by utilizing the various morphological operation. The first process is to fill gaps between edges. Then the region is dilated, and the region's holes are filled. Following this is a process of erosion for smoothing out region boundaries. FIGURE 4.6 shows images with the intermediate and morphologically processed output.

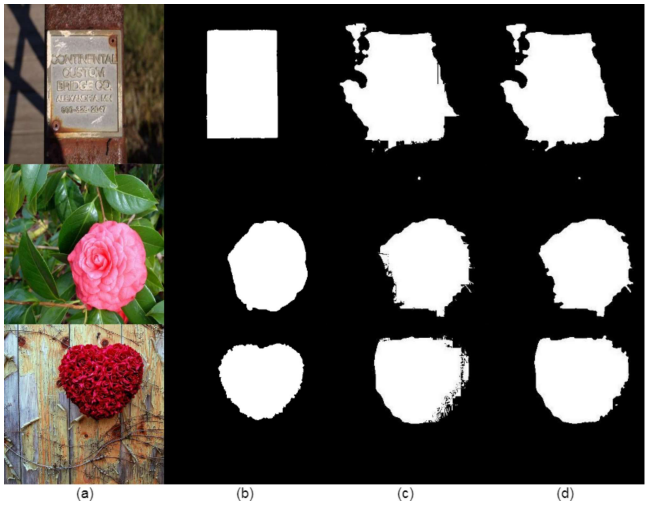


FIGURE 4.6: Morphological Operations. (a) Original Image (b) Ground Truth (c) Intermediate Output (d) Morphologically Processed Output.

4.3.6 Boundary correction using Lazy Snapping

For the correction of boundaries, the final output is obtained by passing the result through the graph-based segmentation of the image. Li *et al.* [191] in 2004 proposed the graph-based segmentation method known as Lazy Snapping. It takes as input the foreground and background masks. For these masks, the proposed model takes the outputs obtained from Gabor filter segmentation. The segment chosen (considered to contain the salient object) is considered to be a foreground mask, and the remaining two segments form the background mask. FIGURE 4.7 displays the image with boundary correction after morphological operations.

4.3.7 Algorithm

The algorithm for the proposed model is given in Algorithm 3.

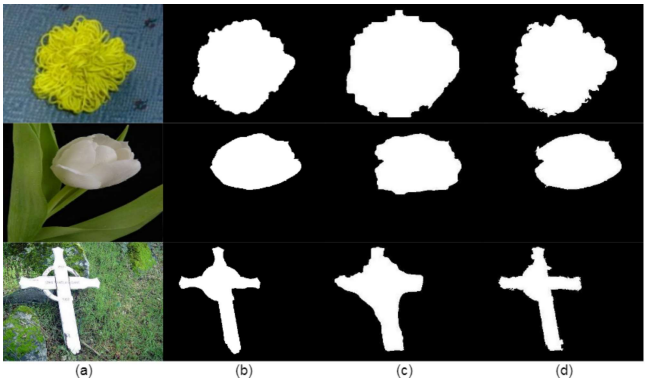


FIGURE 4.7: Boundary Correction. (a) Original Image (b) Ground Truth (c) Morphologically Processed Image (d) Boundary Corrected Image.

Algorithm 3: Proposed Algorithm

Result: Salient Object

1 Read an input image I;
2 Segment the input image I using Gabor filter bank;
3 Use objectness to select the segment $Gseg$ which contains the salient object;
4 Obtain background saliency map D ;
5 Obtain the foreground saliency map M;
6 Obtain a thresholded saliency map $truncM$;
7 Obtain an intermediate output C by intersection of $truncM$ and $Gseg$;
8 Apply morphological operations ;
9 Apply boundary correction to obtain the final salient object F;

The proposed algorithm can be represented as a single line equation as

$$\text{SalientObject} = \text{LS(MO(T(N(MDB(B(I))))\&(O(GS(I)))))} \tag{4.7}$$

where LS refers to Lazy Snapping, MO refers to Morphological Operations, T represents Truncation, N stands for normalization, MDB means Minimum Directional Backgroundness, B stands for Backgroundness, I is the input image, O represents Objectness and GS stands for Gabor Segmentation. A block diagram of the proposed algorithm is shown in FIGURE 4.8. FIGURE 4.9 shows the output obtained

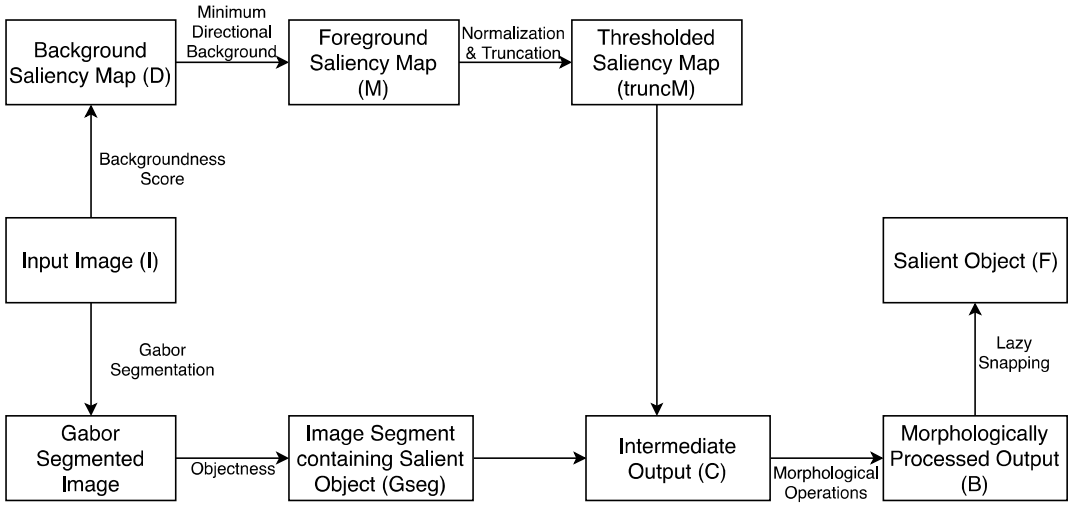


FIGURE 4.8: Block diagram of the proposed Bayesian classifier based algorithm.

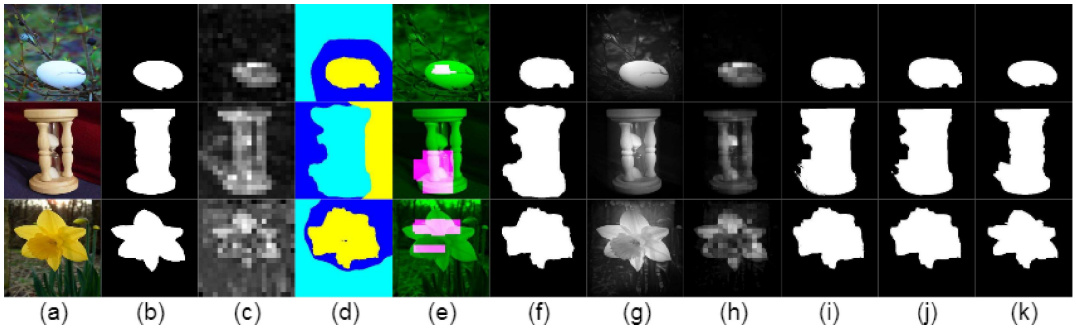


FIGURE 4.9: Outputs obtained at each step. (a) Original image (b) Ground Truth (c) Background Saliency Map (d) Gabor Segmented Image (e) Objectness patch (f) Image Segment containing Salient Object (g) Saliency Map (h) Thresholded Saliency Map (i) Intermediate Output (j) Morphologically Processed output (k) Final output with corrected boundary.

at each step.

4.4 Results

This section explains the results obtained by the proposed model. Subsection 4.4.1 shows the saliency results obtained by the proposed model. Subsection 4.4.2 justifies

TABLE 4.2: MAE of different algorithms on various dataset. Red: Least MAE; Green: Second Smallest MAE; Blue: Third Smallest MAE

Methods \ Datasets	ASD	BRUCE	DUT-OMRON	ECSSD	JUDD	MSRA	PASCAL-S	THUR
EQCUT [194]	0.0564	0.1088	0.1252	0.1733	0.1506	0.1136	0.2319	0.1075
FBP [192]	0.0436	-	-	-	-	-	-	-
MDC [177]	0.0454	0.2015	0.1431	0.1454	0.2132	0.0850	0.2067	0.1212
MST [193]	0.1412	0.1941	0.1647	0.1567	0.2179	0.1006	0.2106	0.1211
SGP [195]	0.3020	0.3253	0.3195	0.3672	0.3268	0.2930	0.3465	0.3151
Proposed Model	0.0420	0.1011	0.1131	0.1432	0.1132	0.0800	0.2015	0.1011

the choice of the value of parameters for the number of patches and the threshold value. Failure cases are discussed in subsection 4.4.3.

4.4.1 Saliency Results

The PR-curves and F-Measure curves for different datasets are shown in FIGURE 4.10 and 4.11. FIGURE 4.12 shows a visual comparison of different algorithms for various input images. TABLE 4.2 shows the mean absolute error obtained for all the algorithms.

According to the results obtained it can be said that, JUDD is the most difficult dataset for salient object detection. ASD is the simplest database to work. The results also confirm with the type of images the databases contain. ASD has clutter-free background whereas the background of JUDD images is very complex and cluttered.

Ablation studies are also conducted. The proposed algorithm is compared against the results of using only backgroundness map, saliency map and Gabor segments

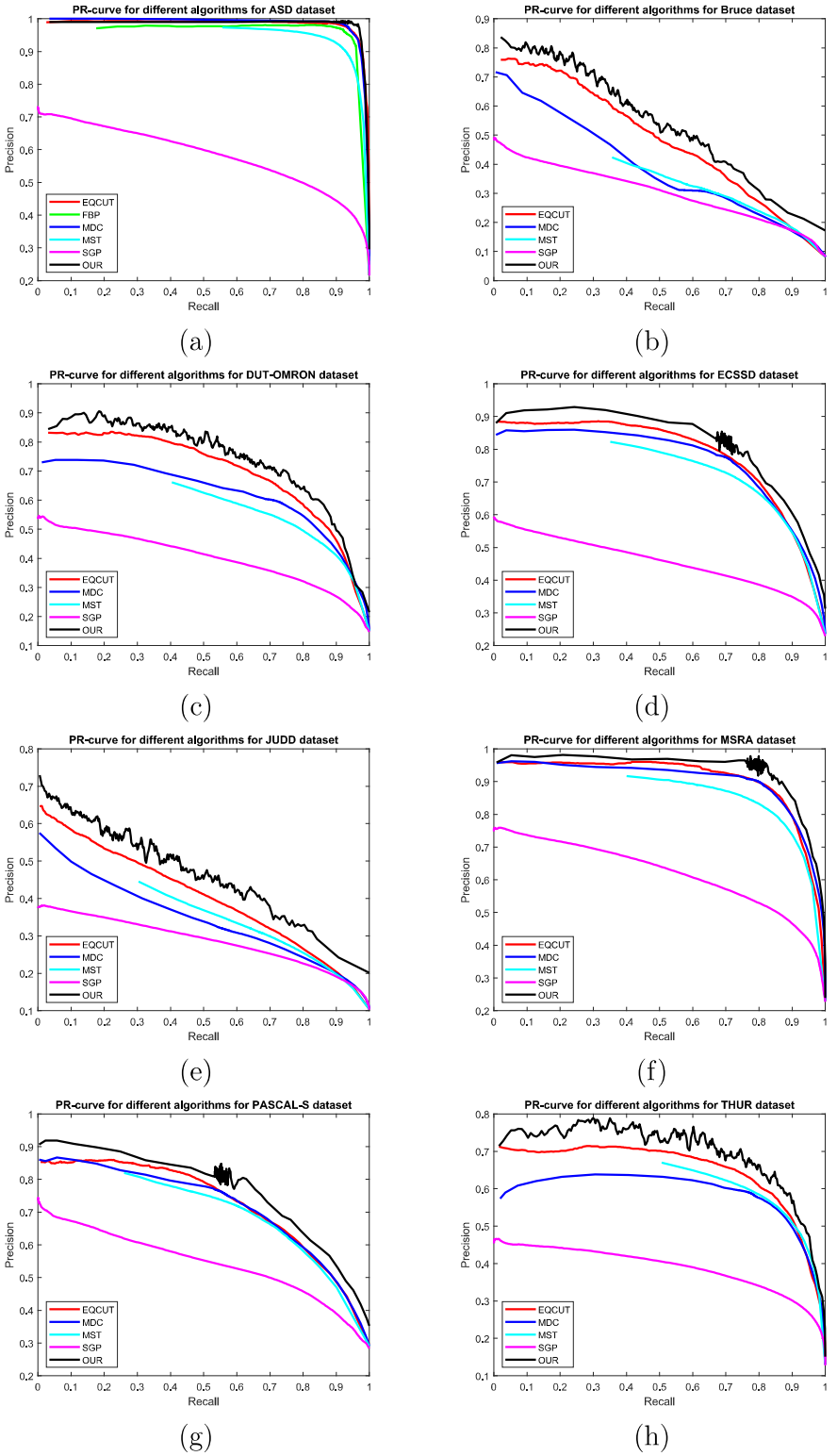


FIGURE 4.10: PR-Curves: (a) ASD (b) Bruce (c) DUT-OMRON (d) ECSSD (e) JUDD (f) MSRA (g) PASCAL-S (h) THUR.

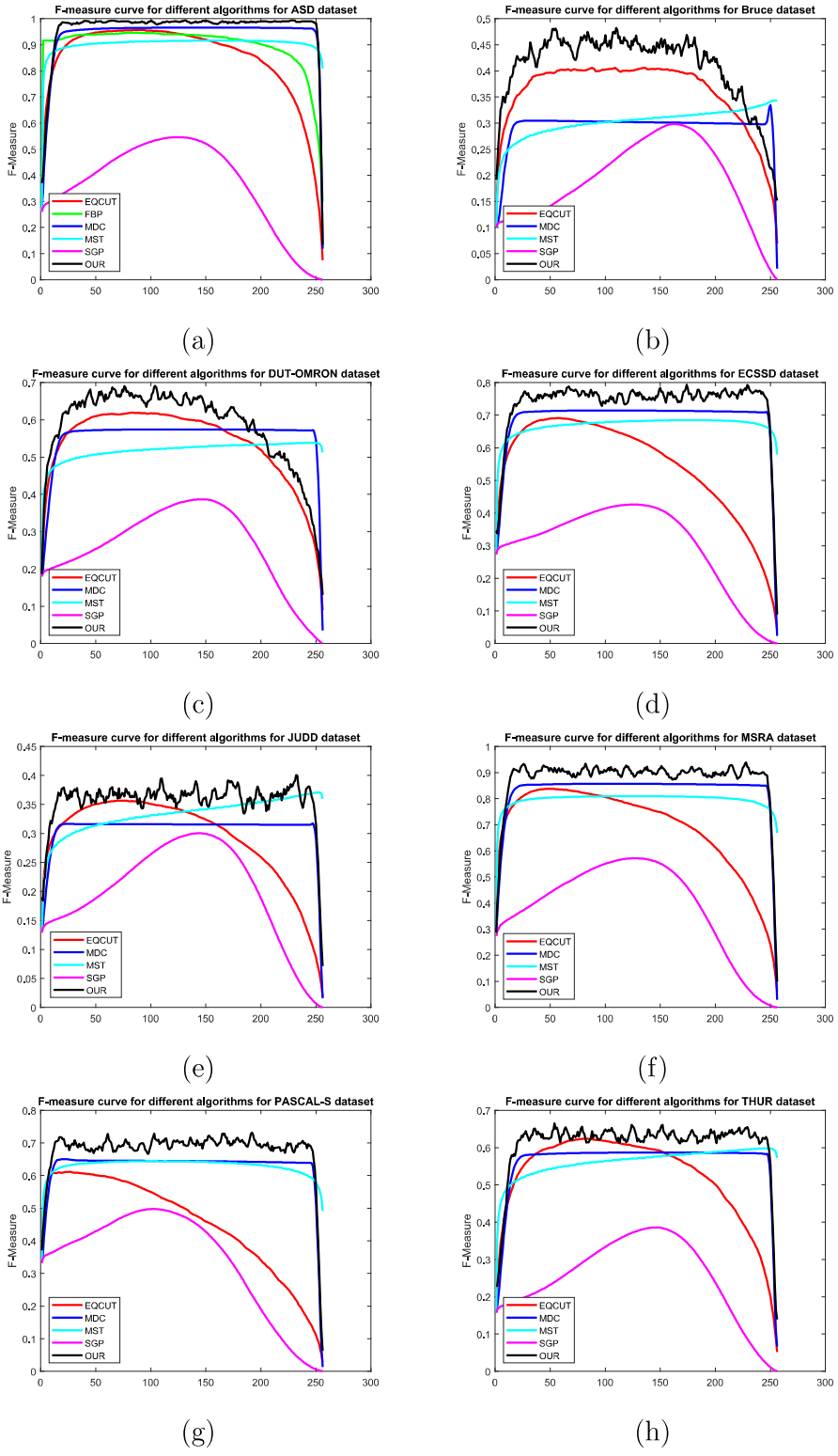


FIGURE 4.11: F-Measure Curves: (a) ASD (b) Bruce (c) DUT-OMRON (d) ECSSD (e) JUDD (f) MSRA (g) PASCAL-S (h) THUR.

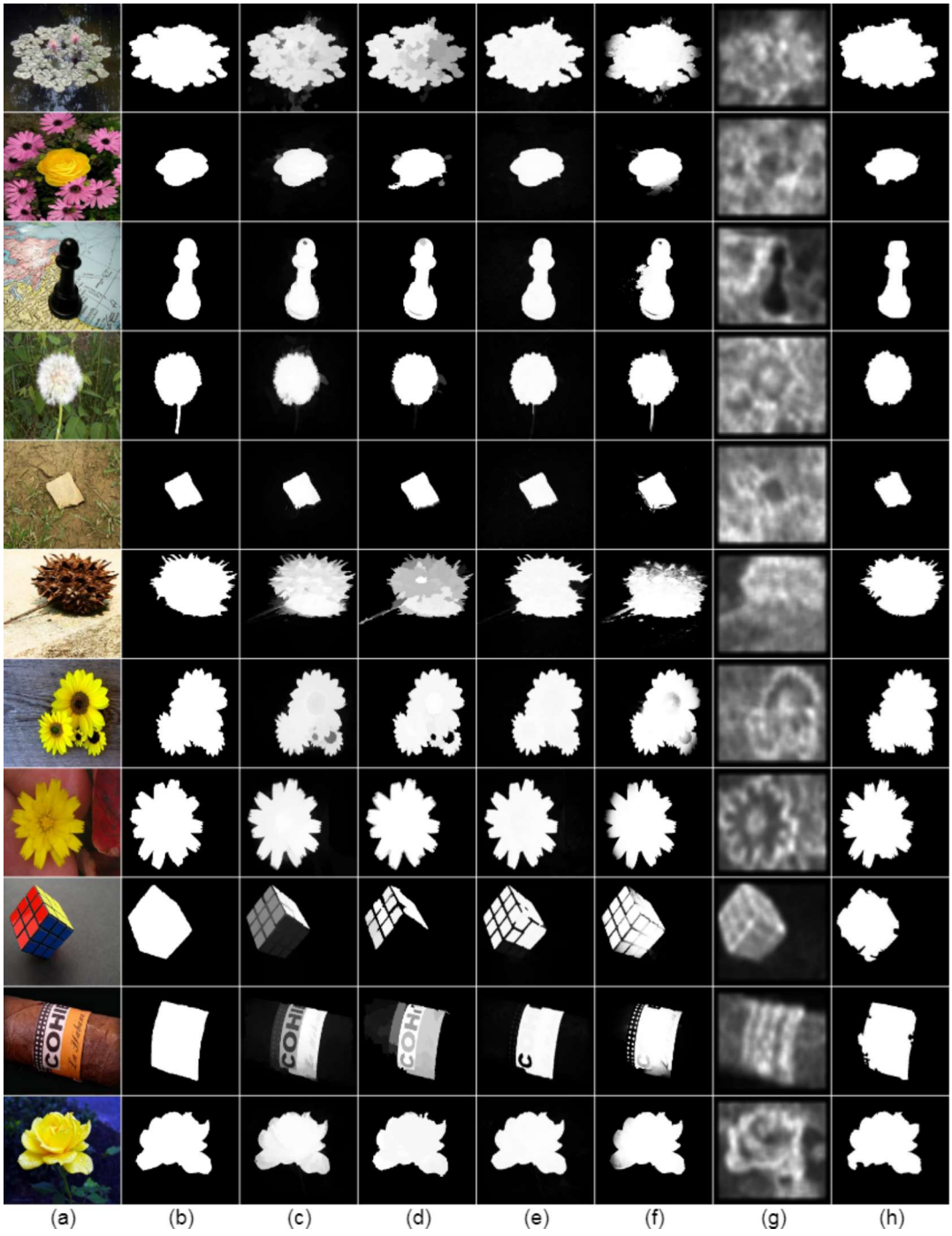
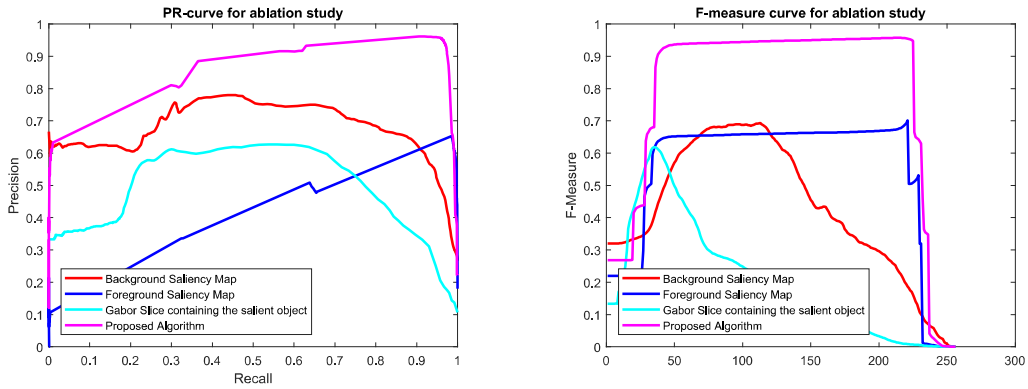


FIGURE 4.12: Visual comparison of the results with various algorithms. (a) Original Image; (b) Ground Truth; (c) EQCUT [194]; (d) FBP [192]; (e) MDC [177]; (f) MST [193]; (g) SGP [195]; (h) The proposed model.



(a) (b)

FIGURE 4.13: (a) PR-Curve for ablation study (b) F-Measure Curve for ablation study.

containing the object. From FIGURE 4.13 it can be clearly seen that none of them are individually able to perform better than the proposed algorithm.

4.4.2 Parameter Change Results

In the objectness part, the top three regions are selected, which appear in maximum bounding boxes. The choice of choosing three regions is based on observing the mean absolute error. By conducting experiments for the top two regions, the mean absolute error for the ASD dataset is 0.0648. By selecting the top four regions, the mean absolute error is 0.0649. On choosing the top three regions, the mean absolute error obtained was 0.0420. So, for implementation, the top 3 regions were included. The threshold for normalized foreground saliency map was experimented for values 0.5, 0.6, and 0.7. Mean Absolute Error obtained was 0.04209, 0.0420, and 0.0436, respectively. So, the threshold was set to 0.6.

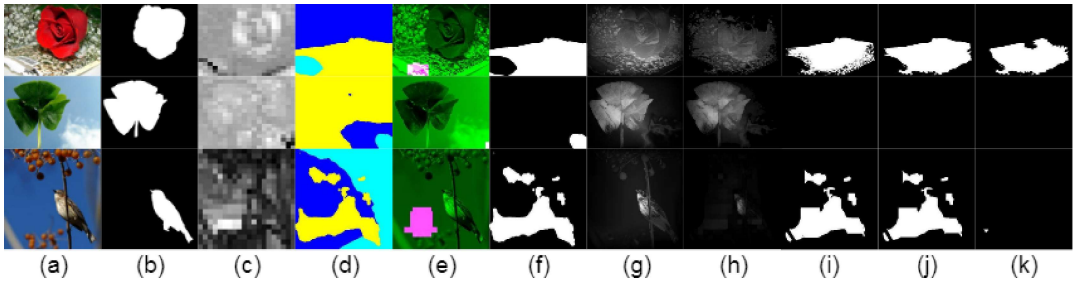


FIGURE 4.14: Result of failure cases. (a) Original image (b) Ground Truth (c) Background Saliency Map (d) Gabor Segmented Image (e) Objectness patch (f) Image Segment containing Salient Object (g) Saliency Map (h) Thresholded Saliency Map (i) Intermediate Output (j) Morphologically Processes output (k) Final output with corrected boundary.

4.4.3 Failure Results

This section deals with the failure cases of the proposed algorithm. FIGURE 4.14 displays some of the failure results. It can be observed that the final result fails when the objectness module fails. This points the weak link in the proposed algorithm. In the future, this can be explored. The strength of the algorithm will be distributed on various modules so that the failure of one of them does not result in complete failure to detect a salient object.

Also, some deep network features [196,197] can be explored to be incorporated in the objectness module to provide more accurate predictions. Feature maps produced by Convolutional Neural Networks can be tuned for use in one of the various modules employed in this chapter.

4.5 Conclusion

In this chapter, the proposed model attempted to solve the problem of salient object detection using background subtraction, Gabor filters, objectness, and minimum directional backgroundness. The novelty of the work lied in introducing a foreground saliency map built from background saliency map for a clear boundary between background and foreground. The work was compared against five state-of-the-art algorithms for eight different types of databases. The evaluation is done based on PR-curve, F-measure curve, and Mean Absolute Error. The proposed algorithm was found better than state-of-the-art algorithms. The failure situations of the proposed algorithm were also discussed, which led to exploring strengthening other modules of the proposed algorithm. Results on JUDD, Bruce and THUR datasets were still very far from human accuracy.

Further research will be focused on improving outcomes for these complex datasets. Deep learning algorithms have not been considered in this chapter. Results from deep learning are very appealing, and efforts can be put in generating hybrid algorithms for salient object detection.