Chapter 1

Introduction

This chapter provides the introduction of the thesis. The background of salient object detection is discussed in Section 1.1. The motivation behind choosing the topic of research is explained in Section 1.2. Section 1.3 defines the problem statement and lists the thesis objectives. The contributions of the thesis are mentioned in Section 1.4.

1.1 Background

Salient object detection is a critical task in the field of computer vision [1]. Humans can easily identify the most significant object in an image. Machines need to simulate human attention to extract the main region in an image [2,3]. For salient object detection, we need to localize and segment the salient region of the image [4].

FIGURE 1.1: Examples of salient objects in images.

The salient region has well-defined boundaries and has high contrast due to which it grabs the attention of the human visual system [5, 6]. A few examples of salient object in images is given in FIGURE 1.1.

The models created for salient object detection must extract the salient region rapidly and precisely [7]. The models need to find the most notable object in the foreground and ignore the background [8, 9].

Eye-fixation prediction is a related field in which an attempt is made to predict where humans look at the images [2]. The results of eye fixation data have been used by various researchers to mark the location of the salient object. Some researchers generally use a two-step process to find the salient object. First, they locate the area where a salient object is likely to be present. Next, they apply some processing to the specific area to extract the salient object [3]. Most of the visual saliency models depend on various priors to find the salient object like, location, semantic, color, background, boundary connectivity, focus [4], foreground, contrast, rarity, objectness, surroundedness [5], edge [10], shape [11], texture [12], intensity and orientation [13].

There are two methods of developing models for salient object detection: bottom-up methods and top-down methods. The bottom-up methods use low-level features, whereas top-down methods use high-level knowledge for salient object detection [4]. Bottom-up methods are data-driven and do not require any training. They are fast and task-independent. Top-down methods are task-driven and need supervised training. They are slower compared to bottom-up methods [5, 13]. There are two theories for the replication of biological mechanism of visual attention: Feature Integration Theory and Guided Search Theory. Feature Integration Theory explains the use of multiple feature dimensions for finding the salient object. Guided Search Theory models look for the region of interest by using guidance from scene [14].

A host of applications utilize the result of salient object detection as a previous step or to increase the performance of existing methods. Salient object detection can be used for the implementation of object of interest image segmentation and object recognition [1]. It can be used as a pre-processing step for content-based image retrieval [2]. Other uses are the image and video compression, video summarization, media re-targeting, photo collage, image quality assessment, image collection browsing, image manipulating, visual tracking, object discovery, and human-robot interaction [3]. It can also be employed for content-aware image editing [5]. The results of salient object detection are also applied in image resizing and scene classification [15]. Salient object detection is also relevant in the field of medical imaging [6]. It has a significant contribution in reducing the complexity of image analysis and image understanding [10]. It also has importance in the field of weakly supervised semantic segmentation, non-photo-realist rendering, photo blending, information discovery, and action recognition [16].

1.2 Motivation

There are several limitations in the existing methods. Most of the saliency maps provided by existing algorithms need to be thresholded by binarization [2]. They do not uniformly highlight the salient region [15]. Existing methods also treat locating salient objects and segmentation as a unified task, which increases the complexity of the procedure and does not provide good results [2]. Most of the datasets are biased by having single salient objects placed in the center of the image with minimum details in the background. The term salient object regarding an image is also subjective in nature. Therefore, the datasets must be prepared by having a widely agreed definition of saliency to provide objectivity [3]. Heterogeneous objects present a major challenge to the saliency algorithms. Algorithms that use hand-crafted features result in the biasing of features towards a specific dataset [16]. The presence of multiple objects is also a major hurdle for the existing algorithms [9]. In images where features of objects and background are similar, it becomes difficult to separate the salient object from the background [17]. Objects with complex shape area are also difficult to extract from image [18].

Salient object detection is applicable in various fields. Finding a salient object can be a performance improvement in terms of speed and accuracy of image annotation. For tagging salient objects in an image, previously, scanning of the whole image was required. Several tags are then generated irrespective of the label belonging to a non-salient object. This leads to several cluttered tags. To overcome this issue, we can first target on finding the salient object in the image. In the next step, we can provide a label for the salient object. In this way, we can reduce tagging time and prevent unnecessary tagging.

There is a clear boundary between salient object detection and scene analysis. For salient object detection, the most prominent objects are selected from the image. For scene analysis, the context behind an object is also important. A man standing in a desert and a man standing near a waterfall has the salient object as man, but considering scene analysis, the desert and waterfall are also significant. This makes both the fields different. There is no method to blur this boundary by providing techniques that help us to move from the salient object to scene analysis. A single algorithm to increase our context from the salient object to the nearby scene is very much required.

Moreover, the transition from the salient object detection to scene analysis must be smooth. At every step, only the next salient object must be included instead of including the whole context at once.

Several deep learning architectures have been proposed for salient object detection. But, the deep architectures are very static in terms that they provide no space for user interaction. The system is of no use if providing the best result does not satisfy the end-user. Providing space for user interaction has two advantages. First, the end-user who will employ the results of salient object detection gets to have some space for his/her opinion. Secondly, this dramatically improves the performance of the overall system. Thus, an attempt is made to make the salient object detection user interactive.

The above mentioned existing gaps in this field motivated us to design various models for salient object detection, which can overcome them and provide effective results.

1.3 Problem Statement and Thesis Objectives

Motivated by the existing research gaps, in this thesis, we aim to propose some methods and models for effective salient object detection and look for its application in image annotation and colon tumor detection.

To provide salient object detection models, three types of techniques have been explored:

1. Statistical models for salient object detection include the use of statistical features of the image for locating the salient object.

2. Machine learning models use Bayesian classifier to discover salient objects in an image.

3. Deep learning models use deep networks for detecting the salient object.

Attempts to smoothly translate from salient object detection to scene analysis is also done using machine learning methods. Also, we have tried to provide a user-interactive approach for salient object detection to make the salient object detection process dynamic.

The thesis also studies the application of salient object detection in the field of colon tumor detection and image annotation.The objectives of the thesis are:

1. To survey the state-of-the-art methods for salient object detection and image annotation.

2. To understand the issues and challenges of the field and limitations of existing models.

3. To propose a statistical, conventional, and deep learning model for salient object detection.

4. To use the proposed model of salient object detection in the medical field.

5. To use the proposed model of salient object detection for image annotation.

1.4 Contribution to the Thesis

The major contributions to the thesis are as follows:

1. A literature survey of salient object detection and image annotation is provided. It covers the methods and techniques of salient object detection and image annotation, together with issues and challenges. The benchmark datasets used in this field are described in detail. The performance metrics used to evaluate the algorithms are also studied.

2. Design and implementation of a salient object detection model using active contours and gradient vector flow. It provides statistical models using statistical cues like intensity, edges, directional contrast in combination with active contours and gradient vector flow.

3. Salient object detection model based on machine learning method is provided. It uses Bayesian classifier for detecting object proposals and other low-level features for locating the salient object.

4. Deep learning model is designed for salient object detection using the YOLOv2 network and boundary correction algorithm. The method is user interactive. A case study on the use of the technique for colon tumor detection is done.

5. Image annotation model is developed using the results from salient object detection. Texture features are used. Feature selection algorithms are comprehensively studied, and a new feature selection algorithm is proposed.

1.5 Outline of the Thesis

The outline of the thesis is as follows:

Chapter 1 introduces the topic of the thesis. It explains the motivation and problem statement of the thesis. Finally, the thesis objectives and contributions to the thesis are mentioned.

Chapter 2 lays the theoretical foundation of the salient object detection and automatic image annotation. It provides a literature survey of salient object detection, image annotation, and colon tumor localization. Issues and challenges existing in this field are mentioned. It also includes a list of benchmark databases and performance metrics on which the proposed models are evaluated.

Chapter 3 introduces the statistical models for salient object detection. In this chapter, three statistical models are designed and implemented using threshold and region-based segmentation, active contours, gradient flow, and directional contrast. A comparison with state-of-the-art methods is also shown.

Chapter 4 presents a Bayesian classifier-based model of salient object detection. The model is designed using background subtraction, directional background contrast, texture segmentation, object proposals, and graph-based segmentation. The model is compared with state-of-the-art algorithms, and the result is provided.

Chapter 5 describes a salient object detection model using deep learning. It also shows the application of a model for colon tumor localization. YOLOv2 with Resnet and Faster R-CNN is used for generating the results.

Chapter 6 shows the application of salient object detection for automatic image annotation. In this, a comprehensive analysis of feature selection is done, and a new method of feature selection is proposed. Multi-label classification is performed for generating tags for images.

Chapter 7 concludes the thesis and explains possible future work in the area of salient object detection.