

## PREFACE

---

The phenomenal use of digital media in our day-to-day life and society brings us to the corner of another world. This transformation appeals and inspires me to innovate and add something to this revolution. Images play an essential role in online shopping and purchasing to online education. It empowers us in various ways, like photo and information sharing, medical imaging, simulation, and military handling of the pandemic. Together with the empowerment, it is necessary to organize these images and use them for various computer vision purposes. One of the most critical computer vision tasks is salient object detection. Salient object detection is to identify the most prominent and relevant object in images. Humans can do this task automatically. The computer can learn to do this by understanding the human visual attention mechanism. It has diverse applications in image captioning, image segmentation, object recognition, content-based image retrieval, image and video compression, and video summarization. This thesis aims to develop computer algorithms, methods, and models for salient object detection, which can detect salient objects similar to human beings.

The main focus of the thesis is to generate probabilistic models for salient object detection to simulate structure, shape, size, and background uncertainty. While designing these models, an attempt is made to discover the current research gaps in existing algorithms, methods, and models. For this, a comprehensive literature

survey is performed. Research gaps of these fields are identified. After that, statistical, probabilistic, and deep learning-based models are developed to address the identified research gaps.

The five models have been developed for salient object detection in 2D and 3D modalities. Among these models, three are based on probabilistic approaches and rest are based on deep learning based approaches. The first two models propose probabilistic contrast and edge enhanced Global Topographical Surface. This surface encloses the prominent object with all its structural and spatial information. Then it is used as a reference plane for regional depth, color, and spatial saliency integration. Most saliency methods generate prominent objects from 2D information, while human attention systems are 3D perception mechanisms. Inspired by this perception, the following models are proposed based on 3D saliency.

The third proposed models is utilized additional depth information from RGBD to robustly and correctly detect the salient object in a complex and cluttered background. To distinguish the salient object in complex and cluttered background, the Poisson probabilistic contrast space is proposed. This process produces a global concave reference surface. The various regional saliencies are integrated into this global concave reference surface to detect the salient object correctly. Background estimation and central saliency integration will thoroughly remove the background. This algorithm generates a robust conspicuous object.

The thesis is presented in a way so that the conventional, as well as newer models like

deep learning, are covered. The exiting models incur inconsistency or distribution loss of salient points and regions. These drawbacks are targets to design two deep network models. The fourth model is CSA-Net, which produces essential features: non-complementary, cross-complementary, intra-complementary, and deep localized improved high-level features. The designed  $2 \times 3$  encoder and decoder streams produce these essential features and assure modality-specific saliency preservation. The cross and intra- complementary fusion are deeply guided by proposed novel, cross-complementary self-attention to produce fused saliency. The attention map is computed by two-stage additive fusion based on a Non-Local network.

Finally, in fifth model, a composite backbone network-based deep CNN framework is designed to produce more acute saliency in a challenging and complex scenario. This composite backbone produces enhanced features that are fused with a self-learning-based dense decoder. A comparative analysis of the proposed models are summarized and discussed. The models are compared against state-of-the-art algorithms and evaluated on benchmark databases. Possible future works and research directions are also included in the last chapter.