Object Detection with Computer Vision

Rakesh Kumar Mahey[1], Tanima Arora[2]

[1]MCA Student, Department of Computer Application, D.A.V Institute of Engineering and Technology, Jalandhar,144008, Punjab, India
[2]MCA Student, Department of Computer Application, D.A.V Institute of Engineering and Technology, Jalandhar,144008, Punjab, India

---***---

Abstract

The goal of this research was to use [1]human hands to run computers, mainly to support and improve technology used in the field of education, especially for supporting lecturers during presentations. The programme created for this goal makes use of additional libraries including [2]FLTK, [3]OpenGL, and [4]OpenCV, as well as [5]computer vision techniques. Presenters need a projector and a webcam in order to use the programme. The programme sends appropriate signals to the computer based on the[6] recognised patterns by using the webcam to detect and analyse the shape and pattern of the presenter's hands. The study's output is a programme that successfully improves [7]teaching and presentation techniques, giving teachers a more seamless experience. In the subject of [8]computer vision, there is now a lot of work being done on the identification and localization of visually appealing regions inside images. Applications in [8]computer vision, [9]computer graphics, and multimedia can all benefit from the ability to automatically [10]recognise and partition such salient regions. Many salient object detection [11](SOD) techniques have been developed to imitate the capacity of the human visual system to identify salient areas in images. Based on how they engineer features, these techniques can be generally divided into two groups: deep learning-based methods and conventional methods.

This survey covers both traditional and deep learning-based methods, including the most significant developments in image-based SOD. The survey offers in-depth insights on saliency modelling trends, addressing important concerns, outlining fundamental methods, and exploring potential future directions.

***Keywords**:* **human hand detection FLTK, OpenGL, OpenCV, computer vision techniques, recognized patterns, teaching presentation techniques, computer vision, computer graphics, recognize, (SOD) techniques**

1.Introduction

The key task of salient object detection (SOD), which is based on the properties of the human visual system (HVS), is to precisely recognize and separate visually different sections inside images. The pre-attentive phase of the HVS, which focuses human attention on the most intriguing parts of a scene, is emulated by SOD models. In order to maximize efficiency and make the most use of available resources, following high-level vision tasks may benefit from the identification of salient regions in images.

SOD has shown useful in a variety of computer vision tasks as a preprocessing step, including visual tracking, picture captioning, and image/video segmentation. But because there are so many different kinds of scenes that are recorded in free-viewing situations, SOD runs into problems and difficulties.

A key component of deep learning-based salient object detection (SOD) is the use of fully convolutional neural networks (FCN) [12]. In order to do coarse saliency prediction and modify it for boundary-accurate saliency maps in a data-driven manner, FCNs make use of their potent hierarchical multi-scale feature representation. While conventional models for SOD have the advantage of real-time performance and use in real-world scenarios, deep learning-based models for SOD have substantial advantages, such as better accuracy.

Saliency priors have been added into various deep models in recent research to improve the representational capacity of multi-layer features and speed up the training process. Saliency estimations from various traditional methods were integrated by Wang et al. [13] to provide as background information for the identification of salient locations. The saliency identification procedure is guided by this prior knowledge.

Goal Of SOD

The purpose of this survey is to give readers a thorough and useful understanding of how Salient Object Detection (SOD) approaches have developed throughout time. By providing thorough insights into their essential components, the study seeks to [14] capture the essence of the motivational concepts behind both traditional and deep learning-based approaches. Although numerous methodologies have been addressed by current studies, they frequently lack in-depth technical information. These surveys were unable to offer adequate technical details for each method due to the emphasis on covering a wide number of ways.

The goal of this survey, in contrast, is to find a balance between the technical details provided for each approach and the coverage of pertinent methods. Thus, it hopes to close the gap and recent a more thorough rundown of SOD methods. The most recent and effective techniques from both traditional and deep learning-based approaches are covered in this survey. It also compares a number of cutting-edge techniques using four regularly used metrics from the SOD literature.

This review seeks to give readers a comprehensive overview of the status of the field today by combining a broad covering of methods with technical specifics, giving them an understanding of the various approaches and how well they perform in salient object recognition.

2. overview of Salient Object Detection

Cognitive psychologists and neuroscientists who have looked into the cognitive and psychological theories of the[15] Human Visual System (HVS) attention have made contributions to the interdisciplinary topic of saliency detection . Early saliency models have been built on the foundation of these notions. Itti et al.'s full implementation of the computational attention architecture is a substantial advancement in visual saliency.

In order to create a low-resolution saliency map, the model put forward in Reference established a feed-forward method that computes and combines multi-scale color contrast, intensity contrast, and orientation contrast. The most prominent areas of the image are highlighted by this computational technique. Additionally, a[16] winner-take-all (WTA) neural network is used repeatedly.

2.1 Global Contrast Based SOD

These steps can be used to determine a region's saliency value using a colour histogram: For the whole image I, calculate the colour histogram. The distribution of colours in an image is depicted by a colour histogram by counting how often each colour appears.

Segment the image I into appropriate parts. Consider that rk stands for one of these subdivided regions.

The segmented region rk's colour histogram should be calculated.

By contrasting region rk's colour histogram with the colour histograms of every other segmented region in the image, you can determine the region's overall contrast.

You can use a variety of methods, including histograms, to determine the global contrast.

2.2 . Diffusion Based SOD

Diffusion-based SOD models use a diffusion matrix and a graph structure on the picture to distribute the seed saliency values over the entire region of interest. The creation of foreground/background seeds, the regulation of the diffusion process, and the graph building of these models can all differ.

These approaches build a graph on the image, where each node corresponds to a pixel or patch, and edges link adjacent nodes. The spatial relationships between pixels or patches in the image are represented by the graph.To start the diffusion process, seed values are generated for the foreground and background. The initial saliency values allocated to particular areas of the image are shown by these seeds. Depending on the particular SOD model, the seed selection may change.Using a diffusion matrix, the seed saliency values are spread throughout the entire region of interest during the diffusion process. The saliency values' distribution across the graph is governed by the formula in this matrix. Normally, it takes into account both the saliency values of nearby nodes and the connectivity between nodes.

3.Deep Learning-Based Salient Object Detection

Convolutional Neural Networks[17] (CNNs) have a deep design that allows them to learn differentiable and hierarchical features at many levels. In order to create representations that are essential for saliency detection, deep learning-based Saliency Object Detection (SOD) models take use of this hierarchical feature hierarchy and integrate innovative architectural features into the network. Deep learning-based SOD models may automatically identify image regions with high saliency values at a coarse scale by utilising these sophisticated multi-layer features. Furthermore, the shallow layers of the hierarchy offer thorough information that is helpful for identifying the boundaries and fine structures of the salient object(s). Researchers can create novel models for the SOD problem using CNNs because of their wide range of capabilities.

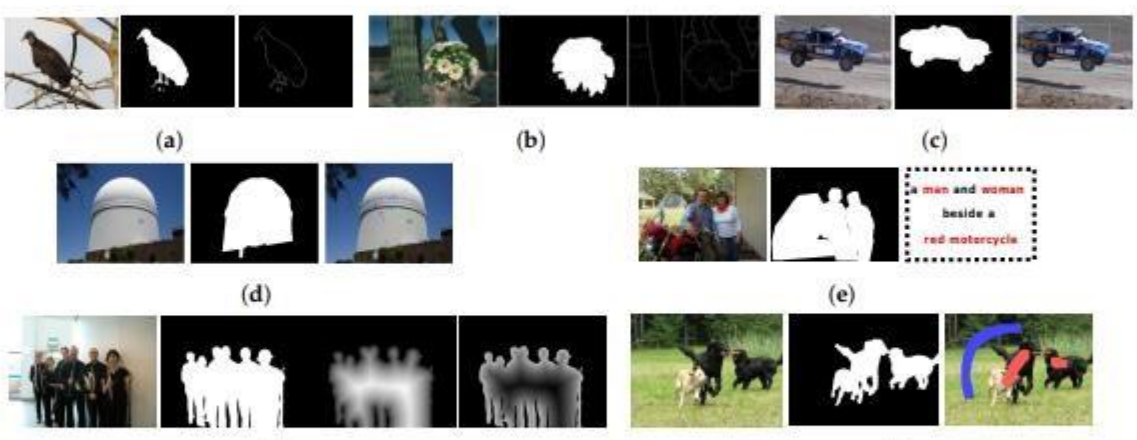

FIG3.1

4. Comparison and Analysis

Due to differences in their number, shape, size, location, and lighting conditions, the existence of many items with the same attributes presents a challenge in saliency object recognition. Edgebased algorithms have the potential to miss actual positives and emphasise true negatives by inaccurately detecting significant edges and suppressing background edges.

In order to highlight many items, the RAS model [18] improves object boundaries, as seen in Figure 4.1, . However, precisely capturing all of the semantic data is essential to the RAS model's effectiveness. Additional difficulties arise when attempting to detect small objects in scenes (Figure 4.2), including the need for suitable detection at a coarse level and efficient feature aggregation techniques to prevent interruptions during progressive fusion.

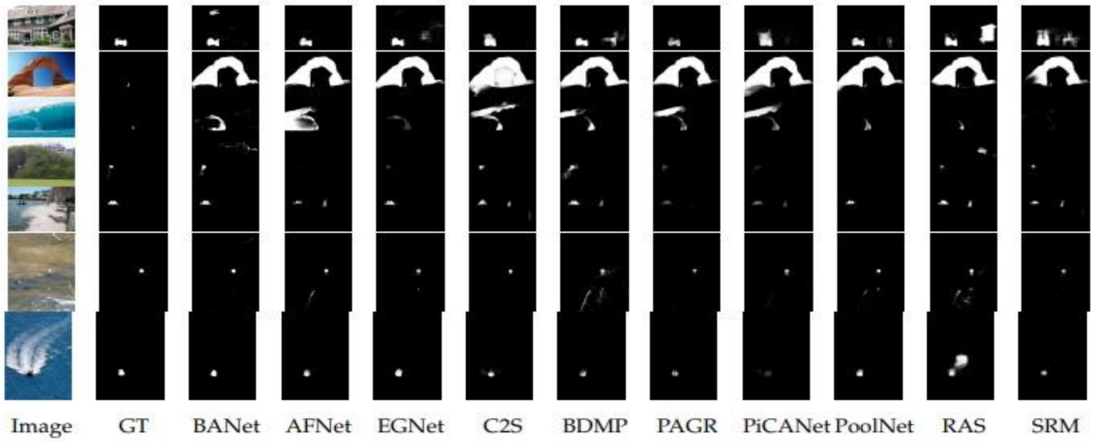

FIG 4.1

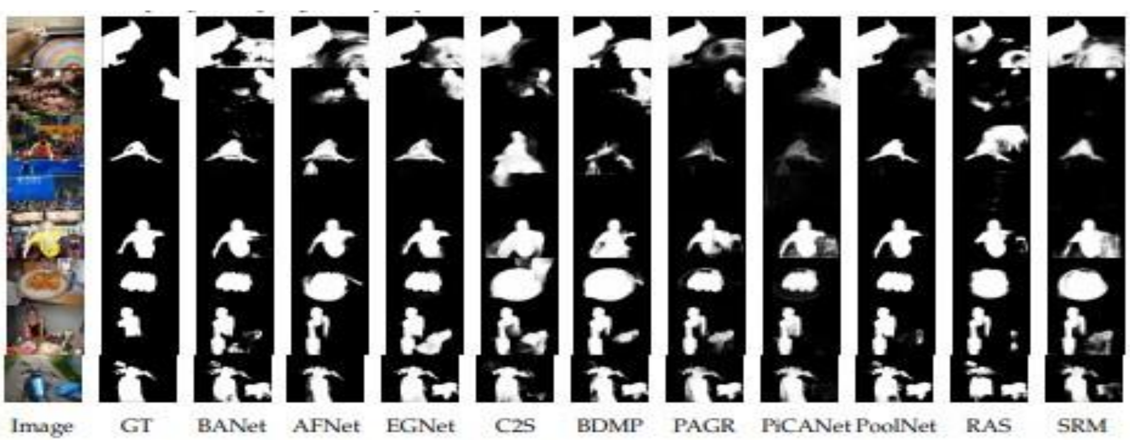

FIG 4.2

5. Future Recommendations

Deep neural network (DNN) models have recently emerged to meet the needs of embedded and mobile applications. A good example of one of these models is the residual learning model suggested in Reference , which focuses on learning the residual in each side-output of the Fully Convolutional Network (FCN) to gradually improve the global prediction. A convolution layer with fewer channels is used to apply this strategy, resulting in a compact and effective model. It is challenging to maintain high detection accuracy while reducing the model's capacity in such constrained environments. Therefore, these recent developments in DNN models for SOD aim to achieve real-time performance by optimizing model efficiency and addressing memory constraints.

Issues relating to datasets: For the development of SOD models, access to big, less biassed datasets is essential. Bias in the training dataset makes it more difficult for the model to attend salient objects in challenging settings. To swiftly check for the presence of centre bias and data selection bias, existing SOD datasets can be quickly examined. When gathering photos for the datasets, photographs that are too generic for salient object detectors are typically eliminated. Examples include photos with no prominent parts, a background that is overly busy, and salient objects that are outside the image's borders. It is crucial to create datasets with more biased-free, realistic circumstances while maintaining a big scale.

6. Conclusions

In order to give readers a broad understanding of the most important models and current developments in the field, this survey concentrates on salient object detection (SOD) in images. There have been many models presented over the past 20 years, including both traditional SOD strategies and deep learning-based techniques.

Conventional models frequently rely on heuristic priors or low-level hand-crafted features. In scenarios with a single object and a plain background, these models show efficiency and efficacy. They are unable to effectively extract semantic information due to their reliance on hand-crafted features and priors, which results in disappointing predictions in challenging cases. On the other hand, deep learning-based models have taken the lead in SOD. The performance of these models, which make use of deep learning techniques, is astounding, especially in difficult situations involving many objects, scale fluctuations, reflections, and background clutter.

Qualitative analyses show that even the most successful SOD models display performance variability across many difficult circumstances. Recent SOD models have begun including edge information and contextual information to boost performance, which further improves their capacity to accurately capture salient items.

References

1. Ma, C.; Miao, Z.; Zhang, X.P.; Li, M.. IEEE Trans. Multimed. 2017, 19, 2415–2424. [CrossRef]
2. Qin, C.; Zhang, G.; Zhou, Y.; Tao, W.; Cao, Z. Neurocomputing 2014, 129, 378–391. [CrossRef]'
3. Fu, H.; Xu, D.; Lin, S. Object-based multiple foreground segmentation in RGBD video. IEEE Trans. Image Process. 2017, 26, 1418–1427. [CrossRef] [PubMed]
4. Itti, L.; Koch, C.; Niebur, E. IEEE Trans. Pattern Anal. Mach. Intell. 1998, 20, 1254–1259. [CrossRef]
5. Long, J.; Shelhamer, E.; Darrell, T.; pp. 3431–3440.
6. Chen, S.; Tan, X.; Wang, B.; Lu, H.; Hu, X.; Fu, Y.. IEEE Trans. Image Process. 2020, 29, 3763–3776. [CrossRef]
7. Feng, G.; Bo, H.; Sun, J.; Zhang, L.; Lu, H. Neurocomputing 2020, 403, 33–44. [CrossRef]
8. Cao, X.; Zhang, C.; Fu, H.; Guo, X.; Tian, Q. IEEE Trans. Neural Netw. Learn. Syst. 2015, 27, 1253–1265. [CrossRef]
9. Chia, A.Y.S.; Zhuo, S.; Gupta, R.K.; Tai, Y.W.; Cho, S.Y.; Tan, P.; Lin, S. Semantic colorization with internet images. ACM Trans. Graph. (TOG) 2011, 30, 6. [CrossRef]
10. Margolin, R.; Zelnik-Manor, L.; Tal, A. Saliency for image manipulation. Vis. Comput. 2013, 29, 381–392.[CrossRef]
11. Goferman, S.; Tal, A.; Zelnik-Manor, L. Puzzle-like collage. In Computer Graphics Forum.
12. Fang, Y.; Wang, J.; Yuan, Y.; Lei, J.; Lin, W.; Le Callet, P. Saliency-based stereoscopic image retargeting. Inf. Sci. 2016, 372, 347–358. [CrossRef]
13. Borji, A.; Itti, L. State-of-the-art in visual attention modeling. IEEE Trans. Pattern Anal. Mach. Intell. 2012, 35, 185–207. [CrossRef]
14. Nguyen, T.V.; Zhao, Q.; Yan, S. Attentive systems: A survey. Int. J. Comput. Vis. 2018, 126, 86–110. [CrossRef]

15. Treisman, A.M.; Gelade, G. A feature-integration theory of attention. Cogn. Psychol. 1980, 12, 97–136. [CrossRef]

16. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 640–651. [CrossRef] [PubMed]

17. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation.

18. Chen, B.; Li, P.; Sun, C.; Wang, D.; Yang, G.; Lu, H. Multi attention module for visual tracking. Pattern Recognit. 2019, 87, 80–93. [CrossRef]

19. Shi, J.; Malik, J. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2000, 22, 888–905.

20. Alexe, B.; Deselaers, T.; Ferrari, V. Measuring the objectness of image windows. IEEE Trans. Pattern Anal. Mach. Intell. 2012, 34, 2189–2202. [CrossRef