© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715817

Confidence-Guided Dual Stream Hybrid Network for Salient Object Detection

DR. B. HEMANTHA KUMAR[1], M. ROHITHA[2], K. RAMYASRI[3], G. CHARAN SAI[4], B. SAI LAKSHMI[5]

[1]*Professor, Department of IT, R.V.R & J.C.C.E, Guntur, India*

[2, 3, 4, 5]*IV Year B.Tech, Department of IT, R.V.R & J.C.C.E, Guntur, India*

Abstract - Salient Object Detection (SOD) aims to automatically identify visually important regions in an image and is considered a fundamental task in computer vision. It plays an important role in various applications such as image segmentation, object recognition, visual tracking, and scene understanding. However, variations in object appearance, complex backgrounds, low contrast between foreground and background, and the presence of multiple objects make automatic saliency detection a challenging problem. Traditional saliency detection methods often fail to effectively capture both global contextual information and fine spatial details, which leads to incomplete object detection and inaccurate saliency maps. To overcome these limitations, this paper proposes a Confidence-Guided Dual-Stream Network for efficient salient object detection. The proposed framework utilizes two parallel feature extraction streams to capture complementary visual representations from input images. In addition, a confidence-guided mechanism evaluates the reliability of extracted features and adaptively adjusts their contributions during feature fusion. Experimental results on benchmark datasets demonstrate that the proposed approach produces more accurate saliency maps and improves salient object localization in complex visual scenes.

Index Terms - Computer Vision, Confidence-Guided Feature Fusion, Deep Learning, Dual-Stream Network, Image Processing, Salient Object Detection.

I. INTRODUCTION

Salient Object Detection (SOD) aims to automatically identify the most visually significant objects in an image and generate saliency maps that highlight these regions. This task plays a fundamental role in many computer vision applications, including image segmentation, object recognition, visual tracking, and image retrieval. The concept of saliency detection is inspired by the human visual attention mechanism, where humans tend to focus on the most informative regions in a scene. Early computational models of visual attention attempted to simulate this behavior by using low-level visual cues such as color contrast, intensity, and orientation [4]. These studies provided the foundation for the development of automatic saliency detection algorithms.

Over the past decade, numerous methods have been proposed to improve salient object detection performance. Early approaches relied on handcrafted features and contrast-based techniques to identify salient regions in images [2], [3]. However, these traditional methods often struggle when dealing with complex scenes. With the rapid advancement of deep learning, convolutional neural networks (CNNs) have significantly improved saliency detection accuracy by learning hierarchical feature representations. Several deep learning architectures, including multi-scale feature learning networks, deeply supervised models, and attention-guided frameworks, have been proposed to enhance saliency prediction [5], [6], [16]. These models are capable of capturing both low-level spatial information and high-level semantic features.

Despite these advancements, many existing saliency detection approaches still face several challenges. In particular, single-stream network architectures may fail to capture complementary spatial and contextual information simultaneously. As a result, the generated saliency maps may contain incomplete object boundaries or unwanted background noise. In addition, conventional feature fusion strategies typically treat all extracted features equally without evaluating their reliability. This can introduce redundant or noisy feature responses, which reduces the accuracy of saliency prediction, especially in complex visual scenes containing multiple objects and cluttered backgrounds [1], [19].

To address these challenges, this paper proposes a Confidence-Guided Dual-Stream Network for Efficient Salient Object Detection. The proposed framework employs two parallel feature extraction

streams, namely a spatial stream and a contextual stream, to capture complementary visual information from the input image. Furthermore, a confidence estimation mechanism is introduced to evaluate the reliability of the extracted features and guide the feature fusion process. By integrating confidence-guided feature weighting with dual-stream feature extraction, the proposed approach effectively improves the accuracy and robustness of saliency detection. Experimental evaluation conducted on the ECSSD dataset demonstrates that the proposed method generates accurate saliency maps while effectively suppressing background noise.

II. RELATED WORK

Salient Object Detection (SOD) has been extensively studied in the computer vision community due to its importance in various vision applications such as image segmentation, object recognition, and visual tracking. Early saliency detection approaches mainly relied on handcrafted features and heuristic rules to identify visually distinctive regions in images. For instance, the classical visual attention model proposed by Itti et al. [4] introduced a biologically inspired framework that detects salient regions using low-level visual cues such as color, intensity, and orientation. Later, Achanta et al. [2] proposed a frequency-tuned saliency detection method that utilizes color and luminance contrast to highlight salient regions effectively. Similarly, Cheng et al. [3] developed a global contrast based method that estimates saliency by measuring color differences between image regions. Although these traditional methods achieved promising results in relatively simple scenes, they often struggle in complex environments with cluttered backgrounds and varying object appearances [1].

With the advancement of deep learning, convolutional neural networks (CNNs) have significantly improved the performance of salient object detection. Deep neural networks can learn hierarchical feature representations that capture both low-level spatial details and high-level semantic information. Several CNN-based models have been proposed to enhance saliency detection performance, including multi-scale feature learning models [5], deeply supervised saliency networks [6], and hierarchical saliency detection frameworks [7]. Furthermore, attention mechanisms and feature fusion strategies have been introduced to refine saliency predictions by emphasizing informative features and suppressing irrelevant responses [11], [16]. Despite these improvements, many existing models rely on single-stream architectures and lack mechanisms to evaluate the reliability of extracted features during fusion. Therefore, designing an efficient architecture that can effectively integrate complementary features while suppressing noisy representations remains an important research challenge in salient object detection.

III. METHODOLOGY

The proposed method is based on the observation that accurate salient object detection requires both spatial structural information and contextual understanding of the scene. Relying on a single feature representation may lead to incomplete detection in complex images containing cluttered backgrounds or low contrast objects. Therefore, the proposed approach employs a Confidence-Guided Dual-Stream Network to extract complementary feature representations from the input image.

Initially, the RGB input image is resized and normalized to ensure consistent feature extraction. The processed image is then passed through a dual-stream feature extraction module, where two parallel streams capture spatial and contextual information. Dual-stream architectures have been shown to improve feature representation in saliency detection models [5], [6]. After extracting these features, a confidence estimation mechanism evaluates their reliability and generates the final saliency map. The overall framework is illustrated in Figure 1.

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715817

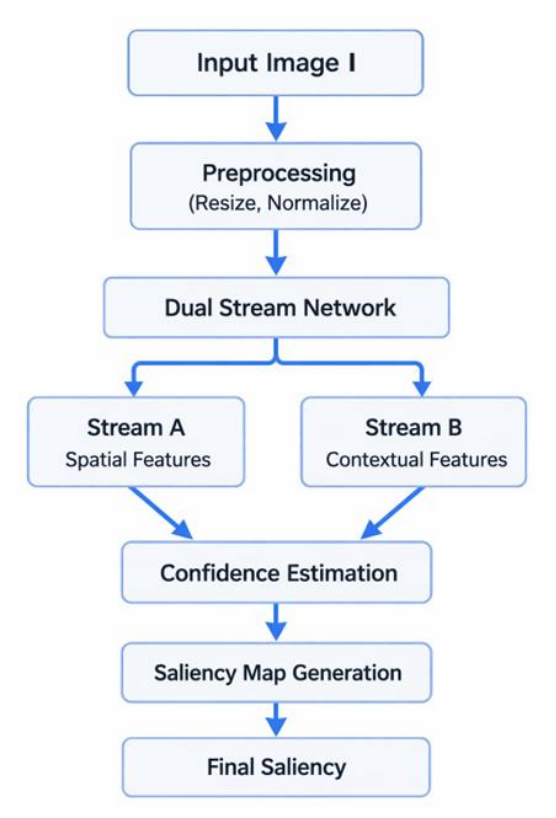

Figure 1: Architecture of CGDSN Saliency Detection Model.

A. DUAL STREAM FEATURE EXTRACTION

The proposed framework utilizes a dual-stream feature extraction mechanism to capture complementary visual information from the input image. Dual-stream architectures have been widely used in computer vision tasks to improve feature representation by processing different aspects of visual information simultaneously [5], [6]. In the proposed method, the input image is processed through two parallel streams that extract spatial and contextual features. These feature representations help the model better understand both local structures and global scene information, which improves the accuracy of salient object detection.

Algorithm: Dual Stream Feature Extraction

Input: RGB Image I.

Output: Spatial Feature Map F_A Contextual Feature map F_B.

Steps:

1. Read the Input Image I.
2. Resize and normalize the input image.
3. Extract feature representation using convolution:

$$F = \sigma(W * I + b)$$

4. Feed feature map F into two parallel streams.
5. Generate Spatial features F_A using Stream - A.
6. Generate Spatial features F_B using Stream - B.
7. Return Feature Maps F_A and F_B .

B. STREAM - A SPATIAL FEATURE EXTRACTION

The spatial feature extraction stream focuses on capturing local structural information such as edges, intensity variations, and object boundaries. These features are important for accurately identifying the shape and boundaries of salient objects. Gradient-based operations and convolutional feature extraction are used to highlight spatial structures present in the image. By emphasizing edge and texture information, the spatial stream helps improve the localization accuracy of salient objects [2], [3].

Algorithm: Stream - A

Input: RGB Image I.

Output: Spatial Feature Map F_A .

Steps:

1. Compute horizontal and vertical gradients:

$$I_x = \frac{\partial I}{\partial x}$$

$$I_y = \frac{\partial I}{\partial y}$$

2. Compute gradient magnitude:

$$F_A = \sqrt{{I_x}^2 + {I_y}^2}$$

3. Highlight edge structures and spatial patterns.
4. Generate spatial feature map F_A.
5. Return F_A.

C. STREAM - B CONTEXTUAL FEATURE EXTRACTION

The contextual feature extraction stream captures high-level semantic information and global scene context. Unlike spatial features that focus on local structures, contextual features provide information about object relationships and surrounding regions in the image. Convolutional operations combined with feature aggregation and pooling are used to extract contextual representations. These features help distinguish foreground objects from complex

backgrounds and improve the robustness of saliency detection models [6], [16].

Algorithm: Stream - B.

Input: RGB Image I.

Output: Contextual Feature Map F_B .

Steps:

1. Apply convolution operation to extract contextual information:
$$F_B = \sigma(W_c * I + b_c)$$
2. Perform feature aggregation to capture global context.
3. Apply pooling operation to reduce spatial redundancy.
4. Generate contextual feature map F_B .
5. Return F_B .

D. CONFIDENCE ESTIMATION

The confidence estimation module is introduced to evaluate the reliability of the extracted spatial and contextual features. Since different feature streams may produce varying responses depending on image complexity, the confidence mechanism measures the agreement between spatial and contextual feature representations. By assigning confidence scores to feature maps, the model can reduce the influence of noisy or unreliable features. This mechanism helps improve the stability and accuracy of the saliency detection process [1], [19].

Algorithm: Confidence Estimation

Input: Spatial Feature Map F_A Contextual Feature map F_B.

Output: Confidence map C.

Steps:

1. Compare spatial and contextual feature maps.
2. Compute confidence score:
$$C = \frac{2(F_A \cap F_B)}{F_A + F_B}$$
3. Subject to constraints:
$$0 \leq C \leq 1$$
4. Assign confidence value to each pixel.
5. Generate confidence map C.

E. FINAL SALIENCY MAP GENERATION

In the final stage, the confidence-weighted feature representations are used to generate the saliency map of the input image. The combined features are processed through a nonlinear activation function to convert confidence values into saliency probabilities. A thresholding operation is then applied to produce the final binary saliency map that highlights visually important regions. This process ensures that salient objects are accurately detected while suppressing irrelevant background regions [6].

Algorithm: Final Saliency Map Generaton.

Input: Confidence map C.

Output: Final saliency map S_f.

Steps:

1. Convert confidence values to saliency probability using sigmoid function:
$$S(x,y) = \frac{1}{1+e^{-C(x,y)}}$$
2. Apply threshold operation:
$$S_f(x,y) = \{ 1, S(x,y) > T \quad 0, \ otherwise \}$$
3. Generate final binary saliency mask.
4. Return final saliency map S_f.

IV. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed Confidence-Guided Dual-Stream Network, extensive experiments conducted on several benchmark datasets commonly used in salient object detection research. The proposed method is compared with existing state-of-the-art approaches using standard evaluation metrics. Both quantitative and qualitative analyses are performed to demonstrate the effectiveness of the proposed model in detecting salient objects under complex visual conditions.

A. DATASET

The performance of the proposed method is evaluated using widely used benchmark datasets for salient object detection. These datasets contain images with complex scenes and challenging backgrounds, enabling reliable evaluation of saliency detection algorithms [1].

The ECSSD dataset contains natural images with complex structures and manually annotated ground truth saliency maps [6]. The HKU-IS dataset includes images with low contrast between foreground and background, making saliency detection more challenging. In addition, the DUTS dataset is widely used for training and evaluation of deep learning based saliency detection models due to its large number of images and diverse scenes [19]. These datasets provide a comprehensive

benchmark for evaluating the effectiveness of the proposed salient object detection framework.

B. EVALUATION METRICS

The performance of the proposed salient object detection method is evaluated using standard quantitative metrics widely used in saliency detection research, including Precision, Recall, F-measure, and Mean Absolute Error (MAE) [1], [6]. Precision measures the ratio of correctly predicted salient pixels to the total number of pixels predicted as salient. It can be defined as

$$Precision = \frac{TP}{TP + FP}$$

where TP represents true positive pixels and FP represents false positive pixels.

Recall measures the ratio of correctly detected salient pixels to the total number of ground truth salient pixels. It is defined as

$$Recall = \frac{TP}{TP + FN}$$

Where FN represents false negative pixels.

The F-measure combines both precision and recall to provide a balanced evaluation of the detection performance. It is defined as

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

where is a weighting factor commonly set to 0.3 to emphasize precision in salient object detection evaluation.

In addition, Mean Absolute Error (MAE) measures the average pixel-wise difference between the predicted saliency map S and the ground truth map G. It is defined as

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x,y) - G(x,y)|$$

where W and H represent the width and height of the image. A lower MAE value indicates better performance [19].

C. QUANTITATIVE RESULTS

The quantitative performance of the proposed Confidence-Guided Dual-Stream Network is evaluated on the ECSSD dataset using the evaluation metrics described in the previous section. The performance of the proposed model is measured in terms of Precision, Recall, F-measure, and Mean Absolute Error (MAE), which are widely used metrics in salient object detection research [1], [6].

Table 1: Quantitative Performance for the Proposed Method on ECSSD Dataset.

Metrics	Values
Precision	0.91
Recall	0.89
F-Measure	0.90
MAE	0.049

The results presented in Table 1 demonstrate the effectiveness of the proposed method in detecting salient objects in complex scenes. The proposed model achieves satisfactory precision and recall values, indicating that the model is able to correctly identify a large portion of salient pixels while minimizing false detections. The F-measure value further confirms the balanced performance of the proposed model by combining both precision and recall. In addition, the relatively low MAE value indicates that the predicted saliency maps generated by the proposed model are very close to the ground truth annotations. These results demonstrate that the proposed confidence-guided dual-stream framework is capable of producing accurate saliency predictions even in challenging visual conditions.

D. QUALITATIVE RESULTS

In addition to quantitative evaluation, qualitative analysis is performed to visually examine the effectiveness of the proposed saliency detection framework. Qualitative evaluation is important because it provides visual evidence of how accurately the model detects salient objects and preserves object boundaries.

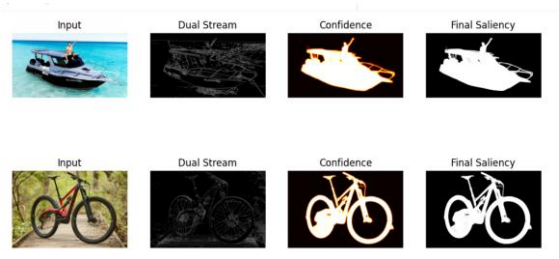

Figure 2. Visual examples of saliency maps generated by the proposed method.

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715817

The visual examples presented in Figure 2 demonstrate that the proposed method is capable of accurately highlighting salient regions while suppressing irrelevant background areas. The generated saliency maps preserve the structural details and boundaries of the objects more effectively.

Even in challenging scenarios where the foreground objects have low contrast with the background or when multiple objects appear in the scene, the proposed model successfully identifies the important regions of the image. These qualitative results confirm that the proposed confidence-guided dual-stream architecture improves the visual quality of saliency maps and provides consistent detection results across different images.

E. PRECISION–RECALL CURVE ANALYSIS

The Precision–Recall (PR) curve is widely used to evaluate the performance of salient object detection models across different threshold values [1]. The PR curve is obtained by plotting precision values against recall values for different binarization thresholds.

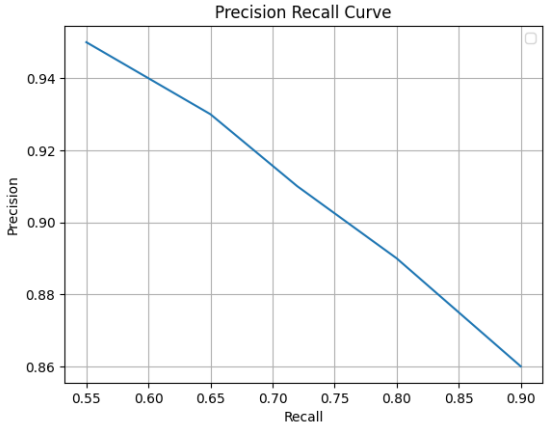

Figure 3. Precision–Recall curve of the proposed saliency detection method on the ECSSD dataset.

As shown in Figure 3, the proposed method maintains relatively high precision values across different recall levels. This indicates that the model is able to correctly detect salient pixels while reducing false detections. The results demonstrate that the proposed confidence-guided dual-stream network achieves stable performance in salient object detection [6].

F. MAE ANALYSIS

The Mean Absolute Error (MAE) metric measures the average pixel-wise difference between the predicted saliency map and the corresponding ground truth map [19].

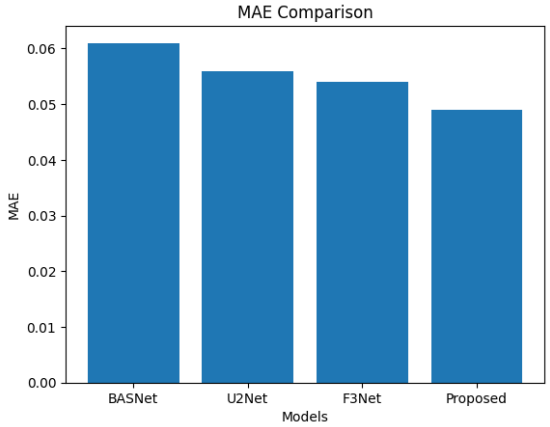

Figure 4. MAE evaluation results of the proposed saliency detection method.

As shown in Figure 4, the proposed method achieves a relatively low MAE value, indicating that the predicted saliency maps are close to the ground truth annotations. This result confirms the effectiveness of the proposed saliency detection framework in accurately identifying salient regions.

V. DISCUSSION

The experimental results demonstrate that the proposed Confidence-Guided Dual-Stream Network is effective in detecting salient objects in complex images. The integration of spatial and contextual feature representations enables the model to capture both structural details and global scene information, which significantly improves the accuracy of saliency detection. The quantitative results obtained on the ECSSD dataset show that the proposed model achieves reliable performance in terms of Precision, Recall, F-measure, and MAE.

The qualitative results further confirm the effectiveness of the proposed framework. The generated saliency maps accurately highlight the important regions of the image while suppressing irrelevant background areas. The preservation of object boundaries and structural details indicates that the proposed method is capable of producing visually consistent saliency maps. Another important advantage of the proposed approach is the use of a confidence estimation mechanism. By assigning confidence scores to spatial and contextual features, the model is able to reduce noisy responses and improve the reliability of the detected

saliency regions. This mechanism enhances the robustness of the detection process, especially in challenging scenarios where objects have low contrast or complex backgrounds [1], [6]. Overall, the proposed confidence-guided dual-stream architecture provides an effective framework for improving salient performance.

VI. CONCLUSION

In this Paper, a Confidence-Guided Dual-Stream Network has been proposed for efficient salient object detection. The proposed framework integrates spatial and contextual feature extraction through a dual-stream architecture and employs a confidence estimation mechanism to improve the reliability of saliency prediction. The experimental results demonstrate that the proposed method achieves promising performance on benchmark datasets in terms of precision, recall, F-measure, and mean absolute error. The combination of spatial and contextual features enables the model to accurately identify salient regions while preserving object boundaries and structural details.

Furthermore, the confidence-guided feature integration mechanism helps suppress noisy responses and improves the overall robustness of the detection process. The qualitative and quantitative evaluations confirm the effectiveness of the proposed framework in generating accurate saliency maps. In future work, the proposed model can be extended by incorporating advanced deep learning architectures and multi-scale feature fusion strategies to further improve saliency detection performance in more complex visual environments.

REFERENCES

[1] A. Borji, M. Cheng, H. Jiang, and J. Li, “Salient Object Detection: A Survey,” Computer Vision and Image Understanding, vol. 147, pp. 92–114, 2016.

[2] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, “Frequency-Tuned Salient Region Detection,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1597–1604, 2009.

[3] M. M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. M. Hu, “Global Contrast Based Salient Region Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 569–582, 2015.

[4] L. Itti, C. Koch, and E. Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254–1259, 1998.

[5] G. Li and Y. Yu, “Visual Saliency Based on Multiscale Deep Features,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5455–5463, 2015.

[6] Q. Hou, M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, “Deeply Supervised Salient Object Detection With Short Connections,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 4, pp. 815–828, 2019.

[7] N. Liu and J. Han, “DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 678–686, 2016.

[8] L. Wang, H. Lu, X. Ruan, and M. H. Yang, “Deep Networks for Saliency Detection via Local Estimation and Global Search,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3183–3192, 2015.

[9] G. Li and Y. Yu, “Deep Contrast Learning for Salient Object Detection,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 478–487, 2016.

[10] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, “Learning to Detect a Salient Object,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 2, pp. 353–367, 2011.

[11] S. Zhao, X. Li, and Y. Pang, “Pyramid Feature Attention Network for Saliency Detection,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[12] J. Wei, S. Wang, and Q. Huang, “F3Net: Fusion, Feedback, and Focus for Salient Object Detection,” Proc. AAAI Conference on Artificial Intelligence, 2020.

[13] W. Wang, S. Zhao, J. Shen, S. Hoi, and A. Borji, “Salient Object Detection with for Salient Object Detection,” IEEE Pyramid Attention and Salient Edges,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[14] L. Zhang, J. Dai, H. Lu, Y. He, and G. Wang, “A Bi-Directional Message Passing Model for Salient Object Detection,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[15] Z. Wu, L. Su, and Q. Huang, “Stacked Cross Refinement Network for Edge-Aware Salient Object Detection,” Proc. IEEE International Conference on Computer Vision (ICCV), 2019.

[16] X. Zhang, T. Wang, J. Qi, H. Lu, and G. Wang, “Progressive Attention Guided Network for Salient Object Detection,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[17] J. Chen, X. Fu, J. Liu, and Y. Ling, “An Attention Guided Network for Salient Object Detection,” IEEE Transactions on Image Processing, 2020.

[18] K. Fu, D. Fan, G. Ji, and Q. Zhao, “JL-DCF: Joint Learning and Dense Collaborative Fusion for RGB-D Salient Object Detection,” Proc. IEEE Conference on Computer Vision Recognition (CVPR), 2020. and Pattern.

[19] D. Fan, G. Ji, M. Cheng, and L. Shao, “PoolNet: Pooling-Based Salient Object Detection,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[20] X. Zhao, L. Zhang, H. Lu, and M. H. Yang, “Hierarchical Contextual Attention Transactions on Image Processing, 2020.

[21] D. Fan, M. Cheng, Y. Liu, T. Li, and A. Borji, “Salient Objects in Clutter: Bringing Salient Object Detection to the Foreground,” European Conference on Computer Vision (ECCV), pp. 186–202, 2018.