Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries

ISSN: XXXX-XXXX

Volume 01 Issue 01 (November) 2024

MULTIMODAL FEATURE FUSION AND HYBRIDIZED CLASSIFIER FOR SALIENT OBJECT DETECTION

Noorayisahbe Binti Mohd Yaacob[1], Dr. Rajesh Kar[2]

[1] School of Computer Science & Engineering, Faculty of Innovation & Technology, Malaysia
[2]Central Highlands Institute of Technology, India

Abstract:

Salient object recognition is a critical technique for image analysis training, as it assists in the identification and emphasis of the most pertinent items or regions, thereby enhancing performance. Due to the intricate and diverse nature of training picture data, achieving high precision and dependability in this assignment is exceedingly challenging. The training image data's complexity and diversity make it difficult to achieve high levels of precision and dependability in this task. This research proposes a framework known as SOD-MFCNN to enhance salient object identification (SOD) in training photos. This framework combines multimodal feature (MF) merging with a hybrid classifier that employs Convolutional Neural Networks (CNNs) to improve detection accuracy and provide a more comprehensive understanding of critical features in training contexts. Machine learning techniques, including CNNs, can autonomously learn to extract intricate information from these images. This enables the system to understand and manage various visual inputs. Support vector machines (SVMs) and more conventional machine learning techniques, such as CNNs, collaborate to enhance detection efforts within the hybrid approach. A multimodal data that comprises training images is implemented to evaluate the system. The images in this dataset were captured during various training sessions and depict individuals and locations. The proposed method obtains detection accuracies of up to 95%, significantly higher than the 88% accuracy achieved with single-modality data. This system demonstrates the potential for improving image recognition tools by utilising MF fusion and hybrid classifiers. The use of these instruments can enhance performance results by fine-tuning training analysis.

Keywords: Salient Object Detection, Hybridized Classifier, Multimodal Features Fusion, Medical image diagnosis, Convolutional Neural Network, Support Vector Machine.

1. Introduction

One of the most critical tasks in computer vision is SOD, which attempts to recognize and separate visually different areas of images from the viewpoint of the human visual system (HVS). As a rule, SOD models should act similarly to how HVS's pre-attentive stage directs viewers' gaze to visually appealing parts of a scene [1], as shown in Figure 1.

Fig.1 Real image and Object detected image (Example)

Finding the most noticeable object or objects in a scene is the goal of salient object detection. Many real-world applications rely on it, including object tracking, person re-identification, action recognition, hidden object detection, video detection and segmentation, image retrieval, semantic segmentation, medical image segmentation, co-saliency detection, stereo matching, and image understanding. [2]. Improving the spatial

consistency of forecasts in this context and extracting more excellent value from scale variation data are two areas that still require attention despite the tremendous progress that has been accomplished [3]. It is common practice to employ top-down and bottom-up algorithms for saliency detection. Top-down algorithms are task-driven and often integrate neuroscience, biology, and computer vision insights. On the other hand, bottom-up algorithms identify crucial parts of an image by collecting low-level picture properties like color, edge, and shape from stimuli [4]. The enormous model sizes and calculations needed to apply multiscale CNNs for RGB-D saliency detection are obstacles. The primary obstacles in developing a multiscale convolutional neural network (CNN) for RGB-D SOD include (1) the size of the model. 2) the sharing of data at many levels [5].

Extracting beneficial characteristics from downsampled the RGB and thermal infrared pictures is the goal of this paper's multi-layer feature fusion technique. We employ two VGG16 networks, one for thermal infrared feature extraction and one for RGB feature extraction, that can keep their separate attributes before upsampling to accomplish this task [6]. DL has recently achieved impressive results in saliency detection, but it is still far from ready to completely supplant more conventional approaches. The primary obstacles are the complicated architectures and the massive computing power needed. They should also keep the object's edges and boundary intact. Applications with limited data and resources can rely on classical saliency detection because it often uses less computer power and memory [7]. Most algorithms used for saliency identification focus solely on local saliency features derived from the image's basic properties; nevertheless, the overall importance is often disregarded. The global saliency map shows information about the world based on basic traits. It reliably and precisely identifies the primary item [8]. Several computer vision tasks have extensively used the Fully Convolutional Neural Network (FCN) in recent years. Many FCN-based models have been developed to detect salient objects based on RGB. These FCN-based salient object detection models constructed multi-level and multi-scale feature representations, which resulted in impressive performance [9]. An enormous spike in the cost of running time is caused by the recurrent modules' efforts to eradicate the erroneous results produced by the preceding block. On the other hand, fine-grained saliency maps can be created by hierarchical feature aggregation [10] by utilizing enriched semantic features from higher levels and detailed border information from shallower layers.

The SOD-MFCNN system aims to improve salient object detection performance through a mixed classification approach and multimodal feature fusion. This system combines CNNs' sophisticated feature extraction capabilities with more traditional machine learning techniques, like SVMs, to enhance detection performance and give a better picture of the training settings. The key objectives of this technique are improving training analysis and performance in various domains and creating a trustworthy tool for object detection.

This study primarily aims to

- Improve SOD accuracy in various training images by integrating multimodal feature fusion with a hybrid CNN-based classifier.
- Improve training-condition feature understanding by combining convolutional neural networks (CNNs) with more conventional machine learning techniques.
- Construct a reliable system that can automatically extract complicated information from training images over a wide range of topics and environments, and that can interpret a variety of visual inputs.

Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries

ISSN: XXXX-XXXX

Volume 01 Issue 01 (November) 2024

- Deal with the difficulties of complicated and diverse training picture data that arise when achieving high accuracy and reliability in SOD.

The purpose of this study is to build a strong framework, SOD-MFCNN, that can identify important objects in training images using hybrid classification and multimodal features more effectively. The system aims to enhance detection accuracy and reliability by combining CNNs with more conventional machine learning methods, such SVMs. It integrates data from various imaging modalities (depth, infrared, and RGB) to provide additional information, ensuring consistent performance in many conditions, including complicated backdrops and varied illumination.

2. Literature Review

Authors	Work	Objective	Advantage	Result	Limitation
Kousik, N et al. [11]	The deep learning model for video salient object detection using CRNN	To address challenges in video saliency detection by combining CNN and RNN.	Improved accuracy with blurry targets, rapid movements, and background occlusions.	achieved higher precision and F-measure with the reduced computational load.	Limitations may occur in highly complex dynamic environments.
Ahmed, K. et al [12]	A comprehen sive study on SOD techniques	To review and analyze recent salient object detection (SOD) techniques and provide a thorough familiarity of challenges and accomplishment s.	Provided a broad perspective on SOD methods, including machine learning and deep learning.	Summarized various image segmentation techniques, classified learning methods, and reviewed datasets and model comparisons.	Lacks in-depth exploration of specific challenges and limitations in current SOD techniques.
Huo, L. et al. [13]	GMANet for SOD in Optical Remote Sensing (ORS) Images	to improve (SOD) in ORS images by addressing global context and large-scale variations.	Used a transformer-based backbone (Pyramid Vision Transformer) for global information and remote dependencies.	Outperformed 28 cutting-edge techniques on 6 metrics, including E-measure and F-measure, showing improved performance with a coarse-to-fine strategy.	It may require significant computational resources due to complex model architecture and extensive feature extraction.
Zheng, P. et al [14]	Co-Saliency of ImageNet (CoSINe) dataset and HICOME approach	to improve Co-Salient Object Detection (CoSOD) by introducing a new, comprehensive dataset and a novel detection approach.	Provided a large and diverse dataset (CoSINe) for CoSOD with better performance and fewer images.	CoSINe dataset outperformed existing datasets in terms of performance with fewer images.	Challenges include remaining inefficiencies in existing CoSOD methods and the need for further refinement

					and improvements .
Akram, T. et al. [15]	improved skin lesion detection and classificatio n method	To efficiently detect and classify skin lesions using enhanced segmentation and feature selection criteria.	utilized ternary colour spaces, a unique weighting criterion, and advanced feature extraction techniques for improved accuracy.	It outperformed existing techniques on PH2, ISBI 2016, and ISIC datasets with better performance metrics, including sensitivity, FPR, specificity, accuracy and FNR.	The method might still face challenges in real-world clinical settings or with highly variable skin lesion images.
Wang, W. et al. [16]	comprehen sive survey on deep SOD	To provide an in-depth review of deep SOD, covering algorithms, datasets, and open issues.	offered a thorough review of deep SOD algorithms, datasets, metrics, benchmark results, and unexplored attributes.	summarized various SOD algorithms benchmarked them and analysed their robustness, transferability, and performance under different attributes and perturbations.	It may not cover every possible SOD dataset; ongoing research may introduce new issues and solutions not included in this survey.
Zhuge, M. et al. [17]	Integrity Cognition Network (ICON) for SOD	To enhance the integrity of SOD by improving both micro and macro-level object identification.	Introduces diverse feature integrity channel enhancement, aggregation, and part-whole verification to improve detection integrity.	It outperforms baseline methods on seven benchmarks, achieving about 10% relative improvement in the average false negative ratio (FNR) across six datasets.	The model's effectiveness may vary with different types of images or real-world scenarios not covered in benchmarks.

3. Proposed Work

a. Dataset Explanation

DUTS is a massive dataset for saliency detection, with 5,019 test images and 10,553 training images [18]. We use the ImageNet DET training/value sets for training purposes and the ImageNet DET test set in conjunction with the SUN data set for testing purposes. Detecting saliency is no picnic; the test and training sets include complicated cases. Fifty people took the time to annotate precise ground realities down to the pixel level.

Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries

ISSN: XXXX-XXXX

Volume 01 Issue 01 (November) 2024

b. Implementation of the SOD-MFCNN Approach

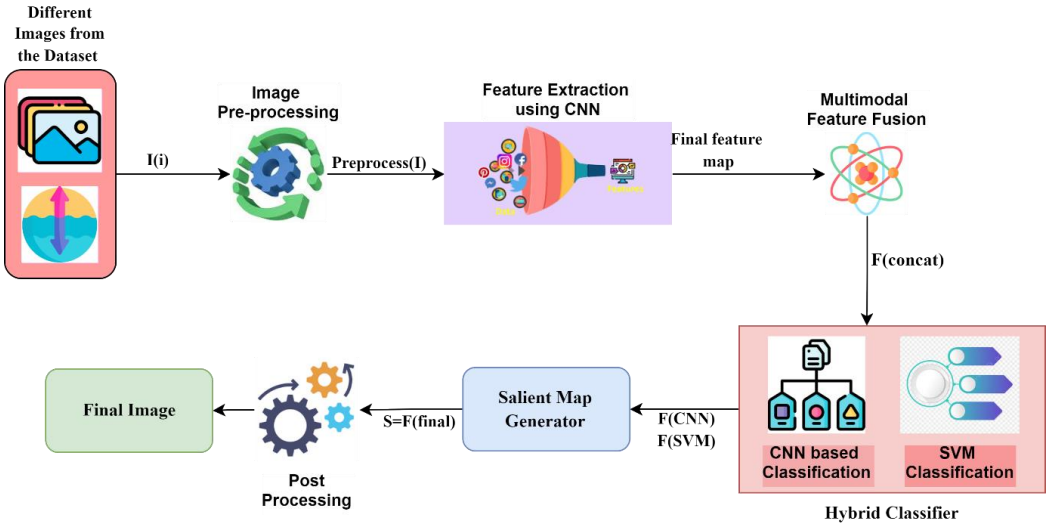

Fig.2 Application of the SOD-MFCNN Method

Figure 2 displays the SOD-MFCNN algorithm, a state-of-the-art technique for extracting crucial details from complex training images. It combines deep feature learning with hybrid classification algorithms to make the most of multimodal data. First, various input image formats are pre-processed, including RGB, depth, and thermal. Separate Convolutional Neural Networks (CNNs) process each modality to extract features. These collected features are combined using concatenation and weighted sum methods to construct a complete representation. The combined characteristics are subsequently fed into a hybrid classifier that employs both CNN-based and Support Vector Machine (SVM) methods for classification. A saliency map is created from the result and then refined using post-processing. The detailed implementation of the proposed work is discussed in the below sections.

i. Input and Preprocessing

The dataset consists of RGB, Depth, and Thermal images. Each image must be pre-processed to normalize size and colour ranges. The set of images can be represented by equation (1).

$$I_i = \{I_1, I_2, I_3 \dots I_n\} \quad \text{(Eq.1)}$$

The preprocessing step can be obtained from the equation (2).

$$I' = Preprocess(I) \quad \text{(Eq.2)}$$

The preprocessing pipeline standardizes the format of all images, regardless of their original attributes or modality, so that the next step is easy to process.

ii. Feature Extraction using CNNs

Feature extraction with CNNs involves passing the input images over a multi-layer processing pipeline to extract task-specific information. As shown in Figure 3, each modality can have its unique CNN architecture in a multimodal dataset to extract characteristics exclusive to that modality.

Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries

ISSN: XXXX-XXXX

Volume 01 Issue 01 (November) 2024

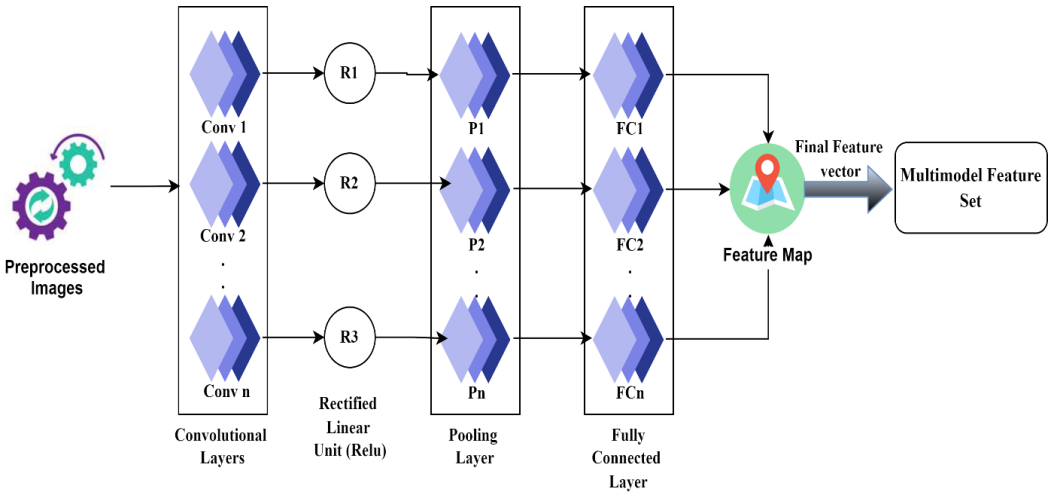

Fig.3 CNN Architecture

Convolutional Layers: These layers use convolutional procedures with learnt filters to isolate specific features in the source picture. It can be obtained by the equation (3).

$$Y_i(l) = f\big(A_i(l) * Y_i(l-1) + b_i(l)\big) \quad \text{(Eq.3)}$$

where $Y_i(l)$ refers to the convolutional layer output for the l-th layer, $Y_i(l-1)$ is the feature map input from the layer before it. $A_i(l)$ is the matrix of weights for the l-th layer in the i-th CNN. $*$ is the bias term, $f()$ is the activation function (Relu).

Pooling Layer: These layers keep the most essential information by reducing the feature maps' spatial dimensions. The feature map is down-sampled using the pooling procedure, commonly called max-pooling or average-pooling, which can be obtained by the equation (4).

$$Y_i(l+1) = Pooling(Y_i(l)) \quad \text{(Eq.4)}$$

where $Y_i(l+1)$ is the pooled feature map.

Fully Connected Layer: Finally, the retrieved features are combined and processed into a feature vector using fully connected layers, following the convolutional and pooling layers. Finally, a series of wholly connected layers flatten the final pooling layer's output into a vector. This can be obtained by the equation (5).

$$F_i = f\big(A_i^{fc} \cdot Y_i^{flat} + b_i^{fc}\big) \quad \text{(Eq.5)}$$

where A_i^{fc} and b_i^{fc} are the biases and weights of the fully linked layer and Y_i^{flat} is the flattened feature map from the final pooling layer.

Final Feature Vector (or Map): The final high-dimensional feature map or feature vector for the i-th modality is provided by the fully connected layer as output F_i, is calculated as shown in equation (6).

$$F_i = CNN_i(I_i) \quad \text{(Eq.6)}$$

The related convolutional neural network (CNN) analyzes each image I_i from the i-th modality through its layers, extracting features at each level until it produces the final feature vector F_i.

iii. *Multimodal Feature Fusion*

Feature extraction from many modalities is fused in multimodal feature fusion to generate a holistic representation that incorporates all modalities' information. This technique performs well for assignments when the different modalities offer

Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries
ISSN: XXXX-XXXX
Volume 01 Issue 01 (November) 2024

complementary but distinct insights. The two fusion methods used are concatenation and weighted sum.

Concatenation: Concatenation combines feature maps extracted from distinct modalities using convolutional neural networks (CNNs). The result is a unified feature representation incorporating all features from the different modalities. It is mentioned in the equation (7).

$$F_{concat} = |F_i| \quad \text{(Eq.7)}$$

where F_i is the feature map of the different images generated by CNN. A single feature representation is created by directly appending the feature maps of each modality. The hybrid classifier receives this combined feature map (F_{concat}) and processes it further.

Weighted Sum: In the weighted sum method, the feature maps of each modality are given a weight before being added together. Because these weights are learnable properties, the model can prioritize or downplay modalities according to their task-related importance. The weighted sum is calculated as in the equation (8).

$$F_{weighted} = A_1 \cdot F_1 + A_2 \cdot F_2 + \cdots + A_n \cdot F_n \quad \text{(Eq.8)}$$

where $A_1, A_2, \ldots, A_n$ are the weights for each modality and $F_1, F_2, \ldots, F_n$ are the feature maps from each CNN. Hybrid classifiers and weights are trained together so that the model can learn the relative importance of each modality in the fused feature map.

iv. Hybrid Classifier

The hybrid classifier takes advantage of the best features of traditional machine learning methods, such as SVMs, and recent developments in the field, such as CNNs.

CNN-based Classification: After multimodal feature fusion, the fused features F_{concat}As shown in equation (9), undergo additional refinement and classification using additional convolutional and fully linked layers.

$$F_{CNN} = FCN(F_{concat}) \quad \text{(Eq.9)}$$

where F_{concat} is the fused feature map from the multimodal feature fusion layer. $FCN()$ stands for the fused features' processing by the last set of convolutional and fully connected layers and F_{CNN} is the product of the CNN classifier, which is usually a probability distribution over classes or a high-level feature vector.

The extra CNN layers are utilized to better capture intricate patterns and interactions within the fused feature map. Prior to adding to the saliency map, these layers improve the feature representation to make it more applicable to the current job, such as classification or localization.

SVM-based Classification: The same fused characteristics are passed through an SVM classifier in parallel with the CNN-based technique. Particularly effective in high-dimensional spaces, support vector machines (SVMs) are robust conventional classifiers renowned for their capacity to construct appropriate decision boundaries, as shown in equation (10).

$$F_{SVM} = SVM(F_{concat}) \quad \text{(Eq.10)}$$

where SVM represents the SVM applied to the fused feature map and F_{SVM} The SVM classifier's results may be represented by an estimate of probability, a margin, or a decision score.

Decision Fusion: The final choice is based on the combined outputs of the CNN-based and SVM-based classifiers. A learnable parameter α balances both classifiers' contributions, which control this fusion. The final output is obtained from the equation (11).

Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries

ISSN: XXXX-XXXX

Volume 01 Issue 01 (November) 2024

$$F_{final} = \alpha \cdot F_{CNN} + (1 - \alpha) \cdot F_{SVM} \quad \text{(Eq.11)}$$

Depending on the task or data, the model can adaptively balance the impact of CNN and SVM using the learnable parameter (α). For example, if the CNN detects more significant patterns, the weight A can be changed to favour the CNN, and the reverse is true for the SVM.

v. Saliency Map Generation

As the last stage in the SOD-MFCNN system, Saliency Map Generation involves visualizing salient regions using the processed features. The function in equation (12) applies a sigmoid activation σ to each spatial position using the final feature output F_{final}.

$$S = \sigma(F_{final}) \quad \text{(Eq.12)}$$

This sigmoid function compresses the values to the interval [0, 1]; areas that are not extremely prominent are represented by 0, and those that are highly salient by 1. S, the final saliency map is a two-dimensional grayscale picture that mirrors the dimensions of the input image. Generally, the saliency map S is up-sampled to conform to the original picture dimensions if F_{final}'s resolution is smaller than the original image. Two methods that can accomplish this include up-sampling and bilinear interpolation. One common way to colourize a saliency map for visualization is to use a heatmap colour scheme, like a jet colour map. In this scheme, cool blues represent low saliency, and warm reds represent high saliency. To further enhance the saliency map, several post-processing procedures might be used:

- Noise reduction through smoothing.
- Saliency boundaries are aligned with the edges of the image via edge-aware filtering.
- Using connected component analysis to eliminate locally significant but tiny areas

4. Results and Discussion

a. Experimental Setup

The DUTS dataset, which includes 10,553 training images and 5,019 test images sourced from diverse sources such as ImageNet and SUN, was used to assess the suggested SOD-MFCNN framework. The photos depict a wide range of challenging situations involving salient object detection. Using different CNN architectures, the framework extracts features from RGB, depth, and thermal modalities. A hybrid classifier, which integrates CNN-based and SVM-based methods, then processes the fused features.

b. Performance Metrics

Intersection over Union (IoU), Mean Average Precision (mAP), and F1 score, were used to test the proposed SOD-MFCNN framework on the DUTS dataset. Comparisons with state-of-the-art approaches such as GMANet [13], CRNN [11], and ICON [17] demonstrated its superior accuracy and resilience across a variety of imaging settings.

i. Mean Average Precision (mAP)

mAP is the mean of Average Precision calculated across all images in a dataset and for all classes (in multiclass detection). For salient object detection, since there is typically only one class (salient vs. non-salient), mAP is the average of AP values computed for all test images. If there are N images, it can be obtained by the equation (13).

Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries

ISSN: XXXX-XXXX

Volume 01 Issue 01 (November) 2024

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad \text{(Eq.13)}$$

where AP is the average precision is deriveable from the given formula (14).

$$AP = \sum_{i=1}^{N} Precision\ of\ i \quad \text{(Eq.14)}$$

Where,

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \quad \text{(Eq.15)}$$

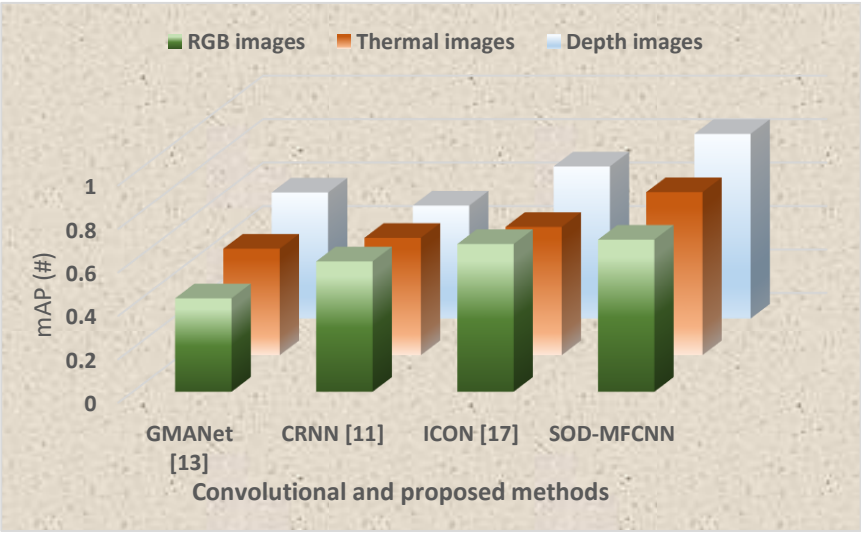

Fig.4 mAP Analysis

Figure 4 compares the suggested SOD-MFCNN framework to state-of-the-art approaches like GMANet, CRNN, and ICON regarding Mean Average Precision (mAP). SOD-MFCNN handles all three types of images—RGB, thermal, and depth—better than competing algorithms. The effectiveness of the framework's hybrid classification method in merging multimodal data to identify key objects is demonstrated here.

ii. F1 Score

The F1 Score considers real and false positives, balancing accuracy and recall. When class distribution is unbalanced or accuracy and recall must be balanced, it is particularly beneficial to employ the F1 Score (under the harmonic mean) to assign a greater weight to smaller numbers than the arithmetic mean. Equation (16) can be used to compute the F1 Score:

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{(Eq.16)}$$

The F1 Score is an essential metric for assessing the success of the suggested SOD-MFCNN approach. This statistic ensures that the model reliably identifies true positives by reducing the number of false positives and negatives. The SOD-MFCNN system can consistently and reliably locate significant objects by maximizing recall and accuracy, irrespective of the complexity of the information. The hybrid classifier enhances the F1 Score by enhancing the classification process of the SOD-MFCNN architecture. As depicted in Figure 5, the improved F1 Score indicates the system's improved recall and accuracy. This result demonstrates that the SOD-MFCNN system is capable of efficiently managing various and non-normalized raw data, a critical attribute for applications that necessitate salient object detection. By maintaining a delicate balance between memory

Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries
ISSN: XXXX-XXXX
Volume 01 Issue 01 (November) 2024

and accuracy, the methodology consistently produces highly precise saliency maps, even for the most complex images.

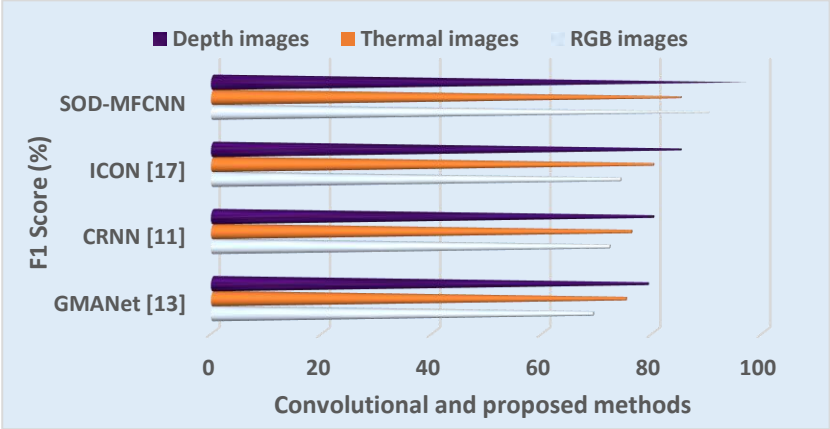

Fig. 5 F1-Score Analysis

iii. Intersection over Union (IoU)

The accuracy of a method for identifying objects can be assessed using the IoU metric. The overlap among the expected and genuine bounding frames is quantified. According to model terms, the predicted border area is the region in which the object is expected to be situated. To determine the region's IoU, divide the total distance of the actual and projected truth areas by the location of their union, as illustrated in equation (17). Applications that depend on accurate object localization include salient object detection.

$$IoU = \frac{A_{intersection}}{A_{union}} \quad \text{(Eq.17)}$$

where $A_{intersection}$ represents the overlap between the expected and actual limits and A_{union} the anticipated and ground truth models' associated bounding boxes cover the entire area.

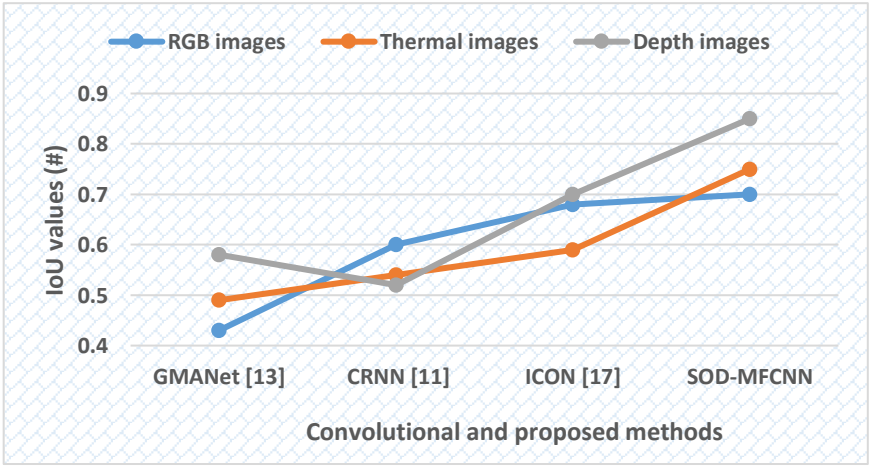

Fig. 6 IoU Analysis

Journal of Revolutionizing Neurological Biomedical Sensing Data Discoveries

ISSN: XXXX-XXXX

Volume 01 Issue 01 (November) 2024

The effectiveness of the SOD-MFCNN framework for SOD is critical, and Figure 6 investigates the intersection over union (IoU) statistic. The model's object localization accuracy is evaluated by determining how much the anticipated and true bounding boxes overlap. A higher IoU value in RGB, thermal, and depth images demonstrates the system's effectiveness and longevity. Consequently, it is possible to establish relevant similarities with established standards.

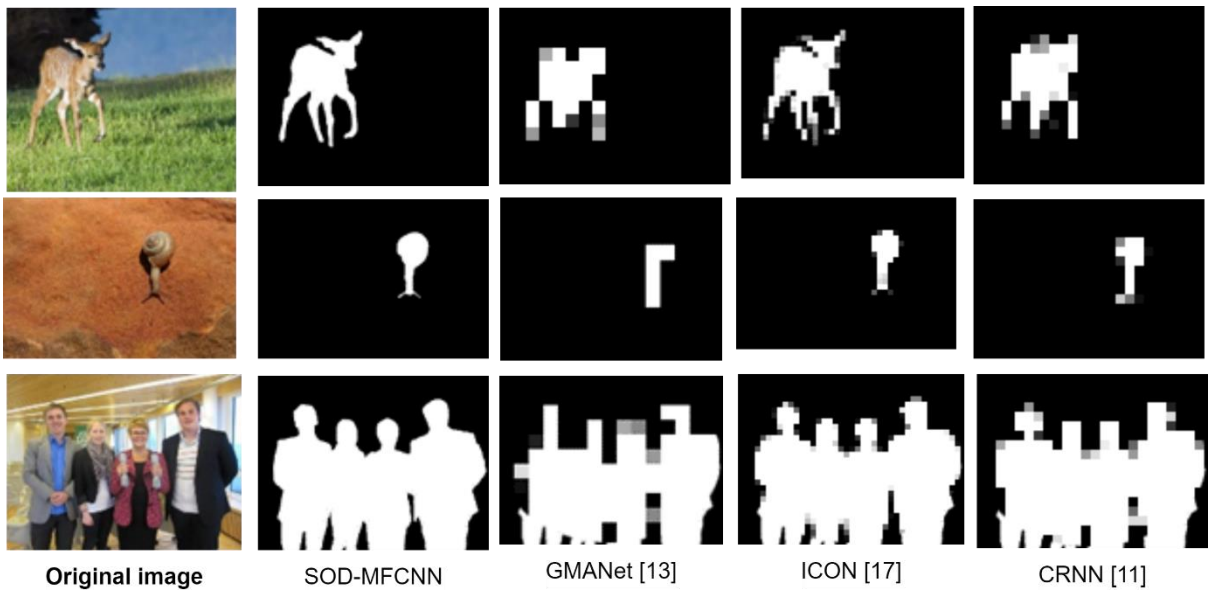

Fig.7 Result evaluation of the proposed method and the convolutional approaches

Figure 7 compares the results of the convolutional neural network methods and the proposed method. Regarding responsibilities that require the identification of significant objects, SOD-MFCNN is of superior quality. The proposed method utilizes a hybrid classification system incorporating multimodal data to improve precision and robustness. The system's capacity to accommodate a variety of visual modalities enhances its detecting precision in various contexts, as demonstrated by the results.

5. Conclusion

The SOD-MFCNN architecture improves salient object detection by integrating multimodal feature fusion, support vector machines (SVMs), and convolutional neural networks (CNNs). This method enhances detection accuracy by combining thermal, RGB, and depth imaging techniques. SOD-MFCNN's prospective value in fields that necessitate specific recognition of objects, such as medical care and sports analysis, is supported by experimental results demonstrating its superiority over state-of-the-art solutions. Effective incorporation of multimodal data is the architecture's greatest asset, as it facilitates robust detection and enhanced accuracy. Various industries could potentially be significantly impacted by this capability. SOD-MFCNN's applicability in time-sensitive or resource-constrained circumstances may be restricted by its computational complexity. Future research should aim to improve the framework's flexibility and extend compatibility with more intricate image modalities.

6. References

[1]. Gupta, A. K., Seal, A., Prasad, M., & Khanna, P. (2020). Salient object detection techniques in computer vision—A survey. Entropy, 22(10), 1174.

[2]. Zhou, T., Fan, D. P., Cheng, M. M., Shen, J., & Shao, L. (2021). RGB-D salient object detection: A survey. Computational Visual Media, 7, 37-69.

[3]. Pang, Y., Zhao, X., Zhang, L., & Lu, H. (2020). Multi-scale interactive network for salient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9413-9422).

[4]. Xiao, F., Li, B., Peng, Y., Cao, C., Hu, K., & Gao, X. (2020). Multi-modal weights sharing and hierarchical feature fusion for RGBD salient object detection. IEEE Access, 8, 26602-26611.

[5]. Huang, R., Zhao, Q., Xing, Y., Gao, S., Xu, W., Zhang, Y., & Fan, W. (2024, April). A saliency enhanced feature fusion based multiscale RGB-D salient object detection network. In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 9356-9360). IEEE.

[6]. Tu, Z., Ma, Y., Li, Z., Li, C., Xu, J., & Liu, Y. (2022). RGBT salient object detection: A large-scale dataset and benchmark. IEEE Transactions on Multimedia, 25, 4163-4176.

[7]. Makram, A. W., Salem, N. M., El-Wakad, M. T., & Al-Atabany, W. (2024). Robust detection and refinement of saliency identification. Scientific Reports, 14(1), 11076.

[8]. Lad, B. V., Hashmi, M. F., & Keskar, A. G. (2022). Boundary preserved salient object detection using guided filter based hybridization approach of transformation and spatial domain analysis. IEEE Access, 10, 67230-67246.

[9]. Zhang, Q., Xiao, T., Huang, N., Zhang, D., & Han, J. (2020). Revisiting feature fusion for RGB-T salient object detection. IEEE Transactions on Circuits and Systems for Video Technology, 31(5), 1804-1818.

[10]. Ren, Q., Lu, S., Zhang, J., & Hu, R. (2020). Salient object detection by fusing local and global contexts. IEEE Transactions on multimedia, 23, 1442-1453.

[11]. Kousik, N., Natarajan, Y., Raja, R. A., Kallam, S., Patan, R., & Gandomi, A. H. (2021). Improved salient object detection using hybrid Convolution Recurrent Neural Network. Expert Systems with Applications, 166, 114064.

[12]. Ahmed, K., Gad, M. A., & Aboutabl, A. E. (2022). Performance evaluation of salient object detection techniques. Multimedia Tools and Applications, 81(15), 21741-21777.

[13]. Huo, L., Hou, J., Feng, J., Wang, W., & Liu, J. (2024). Global and Multiscale Aggregate Network for Saliency Object Detection in Optical Remote Sensing Images. Remote Sensing, 16(4), 624.

[14]. Zheng, P. (2024). Discriminative Consensus Mining with A Thousand Groups for More Accurate Co-Salient Object Detection. arXiv preprint arXiv:2403.12057.

[15]. Akram, T., Khan, M. A., Sharif, M., & Yasmin, M. (2024). Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features. Journal of Ambient Intelligence and Humanized Computing, 1-20.

[16]. Wang, W., Lai, Q., Fu, H., Shen, J., Ling, H., & Yang, R. (2021). Salient object detection in the deep learning era: An in-depth survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6), 3239-3259.

[17]. Zhuge, M., Fan, D. P., Liu, N., Zhang, D., Xu, D., & Shao, L. (2022). Salient object detection via integrity learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(3), 3738-3752.

[18]. https://www.kaggle.com/datasets/balraj98/duts-saliency-detection-dataset/data