International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Unified Hybrid Segmentation: Combining Classical Techniques with State-of-the-Art Deep Learning Models

Sandeep Kumar Dubey[1], Bineet Kumar Gupta[2], Ashish Rastogi[3], Saiyed Faiayaz Waris[4], Pratibha [5], Vishal Vikram Singh[6]

Submitted:11/03/2024 **Revised**: 26/04/2024 **Accepted**: 03/05/2024

Abstract: In recent years, image segmentation has seen remarkable advancements through the development of various deep learning models. It plays a pivotal role in numerous computer vision applications, such as medical imaging, autonomous driving and scene understanding. This paper presents a novel hybrid segmentation approach that integrates the strengths of GrabCut, Mask R-CNN, U-Net, FCN and DeepLab v3 models to achieve superior segmentation performance. GrabCut provides an efficient graph-cut based foreground extraction, which serves as a refined initial mask for subsequent deep learning models. Mask R-CNN improves object detection and instance segmentation functionalities, while U-Net's encoder-decoder architecture excels in segmenting images with limited annotated data, making it particularly effective for medical imaging tasks. FCN contributes by enabling pixel-wise segmentation, ensuring comprehensive coverage of image details. Finally, DeepLab v3's atrous convolution and spatial pyramid pooling enable capturing multi-scale context, enhancing segmentation accuracy in complex scenes. The proposed hybrid approach is evaluated on multiple benchmark datasets, showing substantial improvements in segmentation accuracy and robustness compared to standalone models. Experimental results demonstrate that our hybrid model surpasses state-of-the-art methods in terms of evaluation metrics. This research covers the way for future advancements in image segmentation by combining the strengths of classical and deep learning-based techniques, offering a comprehensive solution for diverse segmentation challenges.

Keywords: *Image Segmentation, GrabCut, Mask R-CNN, U-Net, Fully Convolutional Network (FCN), Deeplab V3, Hybrid Model, Deep Learning*

1. Introduction

Image segmentation is a core task in computer vision, essential for various applications such as medical imaging, autonomous driving, and scene understanding. Accurate segmentation allows for the precise delineation of objects within an image, facilitating tasks ranging from object detection and recognition to image analysis and enhancement. Despite significant advancements in recent years, achieving robust and accurate segmentation remains a challenging problem, especially in complex scenes with intricate details and overlapping objects.

Traditional segmentation methods, such as the GrabCut algorithm, have been widely used for their efficiency and effectiveness in foreground extraction. GrabCut, which utilizes a graph-cut approach, excels in providing an initial mask that separates foreground from background based on color and texture information. However, its performance is often limited by the complexity of the scene and the quality of the initial user input.

On the other hand, deep learning-based segmentation models have revolutionized the field, offering remarkable improvements in accuracy and robustness. Among these, Mask R-CNN, U-Net, FCN and DeepLab v3 have emerged as leading techniques, each contributing unique strengths to the segmentation task. Mask R-CNN is renowned for its ability to perform precise object detection and instance segmentation, making it highly effective in identifying and segmenting individual objects within an image. U-Net, with its encoder-decoder architecture, is particularly adept at handling images with limited annotated data, which is a common scenario in medical imaging. FCN enables pixel-wise segmentation, ensuring that the entire image is comprehensively segmented. DeepLab v3 improves segmentation accuracy by utilizing dilated convolutions and spatial pyramid pooling, capturing multi-scale contextual information and tackling the challenges of complex scenes.

Despite their individual successes, these models have limitations when used in isolation. To overcome these limitations and harness the complementary strengths of these techniques, this paper proposes a novel hybrid segmentation method. By integrating GrabCut, Mask R-CNN, U-Net, FCN and DeepLab v3, the proposed method

1Department of Computer Science and Engineering, Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India
2Department of Computer Science & Information Systems, Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India
3Department of Information Technology ,University of Technology and Applied Sciences, Muscat, Oman
4Department of Artificial Inteligence and Data Science, Koneru Lakshmaiah Education Foundation , Vaddeswaram , Andhra pradesh, India
5Department of Electronics & Communication Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, India
6Research Scholar, Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India

aims to achieve superior segmentation performance. GrabCut serves as the initial step, providing a refined foreground extraction that improves the input for subsequent deep learning models. Mask R-CNN enhances the segmentation by accurately detecting and segmenting individual objects. U-Net contributes its robust performance in scenarios with limited data, while FCN ensures comprehensive pixel-wise coverage. Finally, DeepLab v3 captures the multi-scale context necessary for accurately segmenting complex scenes.

The hybrid approach has been assessed across various benchmark datasets, illustrating substantial enhancements in segmentation accuracy and resilience when compared to employing singular models. Experimental findings indicate that the suggested model surpasses current state-of-the-art techniques in Intersection over Union (IoU) and mean pixel accuracy, particularly in contexts involving complex and overlapping objects. This research highlights the potential of combining classical techniques with advanced deep learning models to create a comprehensive and effective solution for diverse segmentation challenges. The integration of these methods not only leverages their individual strengths but also addresses their limitations, paving the way for future advancements in image segmentation.

2 Related Works

Zhaobin Wang et.al [1] proposing Grab-Cut, a graph theory-based image segmentation method, has garnered significant attention due to its simplicity and effectiveness. This review begins with the original GrabCut model and explores recent improvements such as superpixel-based preprocessing, saliency maps, and non-interactive enhancements. The study also evaluates GrabCut alongside improved algorithms like LazySnapping and Deep GrabCut using five performance metrics. Finally, it identifies current challenges and suggests directions for future research.

Hao Wu et.al [2] proposing to streamline the labor-intensive process of Mask R-CNN object detection; we introduce a GrabCut-based approach for automatic mask generation. This two-stage method first utilizes GrabCut for initial mask generation, followed by Mask R-CNN for refinement and object detection using the generated mask. Experimentation on Berkeley Instance Recognition Dataset objects yielded over 95% mean average precision (mAP) for segmentation, showcasing the method's simplicity and efficiency.

Hyungjoon Kim et al. [3] proposing a hybrid model in urban environment analysis, leveraging diverse strengths for pixel-based object grouping in images. Addressing challenges of dataset variance by pre-processing and fusion of segmentation results from multiple models. Evaluated with Google Street View images, trained with cityscapes dataset, showcasing superior performance through quantitative and qualitative assessments. Demonstrating potential for precise urban analysis, surpassing existing methods with efficient computer vision technology.

Tiara Lestari Subaran et al. [4] a calorie estimation system employs computer vision for food image analysis, relying on detection and segmentation processes. To enhance accuracy, Mask R-CNN and Grab-Cut algorithms are fused, forming a new mask for calorie calculation. With image augmentation, the method achieves less than 10% error and F1 scores surpassing 90% in all tests.

Ronneberger et al. [16] We propose a deep learning network that leverages data augmentation to efficiently train with a limited number of annotated samples. The architecture includes a contracting path for context capture and a symmetric expanding path for precise localization. Trained end-to-end with only a few images, the network surpasses previous methods in segmenting neuronal structures in electron microscopy. Additionally, it processes 512x512 images in under a second on a modern GPU.

Rother et al. [17] proposing efficient foreground/background segmentation in still images is crucial for image editing. Traditional tools use either texture (color) or edge (contrast) information. A recent approach using graph-cuts combines both types effectively. This paper enhances the graph-cut method in three ways: by developing a more powerful iterative optimization, simplifying user interaction, and creating a robust "border matting" algorithm for estimating the alpha-matte and foreground colors around object boundaries. The proposed method outperforms other tools in moderately difficult cases.

3. Methodology

This section covers the algorithm, a detailed explanation of threshold segmentation, and the proposed framework for hybrid segmentation. In the histogram thresholding method, the threshold value is derived from the histogram of intensity differences to select real-time frames. The evaluation of methods is based on Precision, Recall, F1 Score, IoU, Boundary IoU, and Dice Coefficient. Experimental results demonstrate the effectiveness of our approach.

3.1 Overview

Mask R-CNN (Region-based Convolutional Neural Networks) enhances Faster R-CNN by incorporating an additional branch for predicting segmentation masks for each Region of Interest (RoI), alongside the existing branches for classification and bounding box regression [13]. This allows Mask R-CNN to efficiently detect objects in an image while simultaneously generating high-quality segmentation masks for each instance.

- High accuracy in instance segmentation.
- Effective for complex scenes with multiple objects.

DeepLab is a deep learning architecture for semantic image segmentation, it employs dilated convolutions to control the resolution of feature responses and capture multi-scale contextual information without losing resolution [14]. DeepLab has evolved through various versions (DeepLabv1, DeepLabv2, DeepLabv3, DeepLabv3+), each incorporating improvements encoder-decoder structures for better segmentation accuracy.

- Excellent performance on semantic segmentation tasks.
- Effective at capturing contextual information at multiple scales.

Fully Convolutional Networks (FCN), FCNs convert traditional classification networks to fully convolutional ones for dense prediction tasks like semantic segmentation. This architecture allows end-to-end training for pixel-wise prediction, using de-convolution layers (also known as transposed convolutions) to up sample the low-resolution feature maps back to the input image resolution [15].

- Simple and effective for many segmentation tasks.
- Faster inference time compared to more complex models.

U-Net is a convolutional network specifically considered for complex images; it features a symmetric U-shaped architecture with an encoder-decoder structure. The encoder path captures context using down sampling, while the decoder path enables precise localization through up sampling. Skip connections between corresponding layers in the encoder and decoder paths enable the network to combine coarse and fine features, thereby enhancing segmentation accuracy [16, 18].

- High accuracy and efficiency for medical image segmentation.
- Effective at capturing fine details and boundaries.

Suitability with Mask R-CNN is high. Combining Mask R-CNN with GrabCut can enhance instance segmentation by refining the masks predicted by Mask R-CNN using GrabCut's energy minimization.

Suitability with DeepLab is moderate to High. DeepLab's segmentation can be refined with GrabCut, especially in complex scenes where boundary accuracy is critical.

Suitability with FCN is moderate. FCN can provide an initial segmentation that GrabCut can refine. This combination can be useful for simpler tasks where speed is important.

Suitability with U-Net is high. U-Net's precise segmentation, combined with GrabCut, can yield very accurate results, especially in complex imaging.

3.2 Proposed Framework

Combining traditional image segmentation techniques like GrabCut with modern deep learning models leverages the strengths of both approaches. GrabCut, an interactive algorithm based on graph cuts, is known for its efficient boundary refinement capabilities. It refines an initial rough segmentation (bounding box) by modeling the foreground and background with Gaussian Mixture Models (GMMs) and minimizing energy through graph cuts. When used in conjunction with deep learning models, it can enhance segmentation accuracy, particularly at object boundaries [17,19].

In this paper we are proposing the masks from a U-Net model and the GrabCut algorithm to produce a hybrid mask. The script then iterates through each image in a specified directory, applying the hybrid segmentation process and saving the resulting masks. It evaluates the performance of the hybrid segmentation against ground truth masks using metrics such as IoU, precision, recall, F1 score, Dice coefficient, and boundary IoU, and records the runtime for each image. The results are appended to lists, displayed in tabular format, and stored in a DataFrame for further analysis. The final DataFrame is displayed to show the collected metrics for the images processed.

Step 1: Intially get_grabcut_mask(img_path): Loads, resizes an image and applies the GrabCut algorithm to generate a binary mask.

Step 2: Uses a U-Net model to create a mask, combines it with the GrabCut mask and returns the hybrid mask.

Step 3: The script iterates through image files in a specified directory, applies transformations and uses the hybrid segmentation function.

It saves the resulting masks, applies them to the original images and evaluates them against ground truth masks.

Step 4: Determines IoU, precision, recall, F1 score, Dice coefficient and boundary IoU.Records the runtime for processing each image.

Step 5: Appends metrics to lists and displays them in a tabular format for each image.

Creates a Data Frame from the collected metrics and displays the top results.

This approach integrates traditional image processing techniques with deep learning models, combining their strengths to improve segmentation quality and evaluates their performance comprehensively.

4. Experimental Evaluation for segmentation Methods

This paper primarily emphasizes the Segmentation part to make the foreground image extractable, as well as to satisfy the Quality for image data comparison. For comparing and choosing one of the best segmentation techniques for our research work, it is very crucial to know about the necessary evaluation metrics for segmentation methods.

4.1 Precision Score

The Precision Score evaluates the accuracy of the model's positive predictions. It is the ratio of true positive (TP) predictions to the total number of true positive and false positive (FP) predictions.

$$\text{Precision Score} = \frac{TP}{(FP+TP)} \quad (1)$$

4.2 Recall Score

It estimates the ability of model to identify all relevant instances. It is defined as the ratio of true positive (TP) predictions to the total number of true positive and false negative (FN) predictions.

$$\text{Recall Score} = \frac{TP}{(FN+TP)} \quad (2)$$

4.3 Intersection over Union (IoU)

The Intersection over Union (IoU), also known as the Jaccard Index, estimates the overlap between the predicted segmentation mask and the ground truth mask. It the ratio of intersection of the predicted and ground truth areas to the union of these areas.

$$\text{Intersection over Union} = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

4.4 F1-Score

The F1 Score is the harmonic mean of Precision and Recall, offering a unified metric that accounts for both false positives and false negatives.

$$\text{F1 Score} = \frac{2\times \text{Precision Score} \times \text{Recall Score}}{(\text{Precision Score} + \text{Recall Score})} \quad (4)$$

4.5 Dice Coefficient

The Dice Coefficient also referred to as the F1 Score in binary classification, evaluates the similarity between two sets. It is defined as twice the area of overlap between the predicted and ground truth masks, divided by the total area of both masks.

$$\text{Dice Coefficient} = \frac{2*|A \cap B|}{|A|+|B|} \quad (5)$$

4.6 Boundary IoU

Boundary IoU evaluates the intersection over union specifically for the boundary regions of the predicted and ground truth masks. This is done by computing the intersection and union of the dilated boundary regions of both masks.

$$\text{F1 Score} = \frac{\textit{Intersection of Boundary Regions}}{\text{Union of Boundary Regions}} \quad (6)$$

5 Simulation Results and Analysis

The following analysis is presented in this paper:

- Capture real-time images from a camera and computes the histogram of the equalized image and then calculates its entropy as a quality score (based on a specified threshold) for each frame using the defined function, as illustrated in Figures 2 and 3.
- Apply deep learning segmentation models, analyze the results and calculate precision, recall, F1 score, IoU and boundary IoU.
- After obtaining the initial experimental evaluation metrics, implement GrabCut segmentation on each model and select two hybrid proposed approach to demonstrate the updated values. Compare these updated values.

The proposed segmentation has been shown below.

The segmentation's performance is evaluated using open-source software. Real-time image frame acquisition is tested using software with a webcam. The input image sequences considered for implementation are "real-time self-images" with dimensions of 512x512, captured from real-time video. Various segmentation techniques can be applied to the frames selected for further processing.

Data Set 1: Near-field Capture with Unobstructed Background

Fig 1- Capturing and Selecting High-Quality Frames Using Adaptive Histogram-Based Selection for Unobstructed Background

Data Set 2: Distant Capture with Obstructed Background

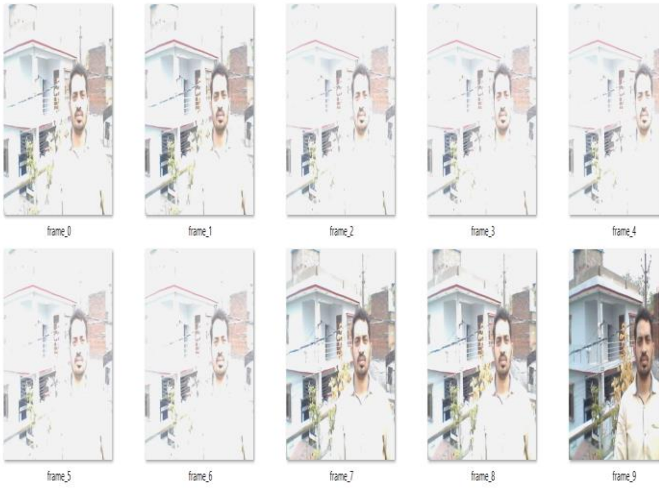

Fig 2- Capturing and Selecting High-Quality Frames Using Adaptive Histogram-Based Selection for Obstructed Background

The experiments are conducted for all Segmentation techniques on real time set of image frames.

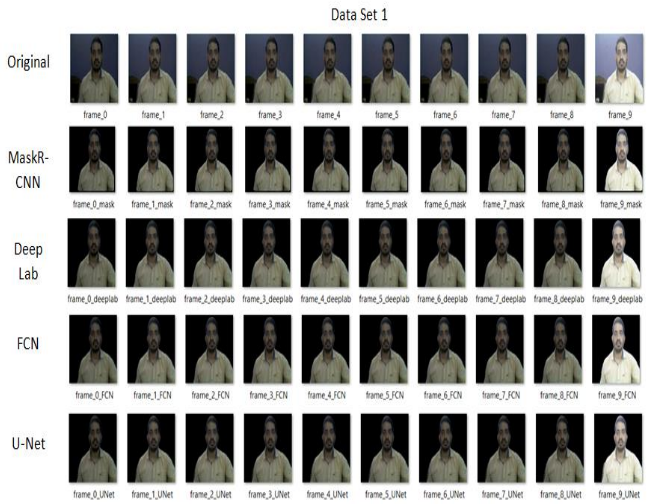

Fig 3: Foreground Detection Analysis for Near-field Imaging with Unobstructed Background

Fig 4: Outcomes of Foreground Detection in Capture with Obstructed Background

Table 1- Comparative Analysis of Segmentation Methods Across Diverse Metrics

Background Invariant Data Set	Experimental Evaluation	Segmentation Technique (Mean Value of Ten Image Frames)			
		Mask R-CNN	Deep Lab	FCN	U-Net
Data Set 1	Precision	0.9986	0.9999	0.9999	0.9995
	Recall	0.9141	0.9512	0.9510	0.9607
	F1 Score	0.9545	0.9749	0.9748	0.9797
	Intersection-over-Union (IoU)	0.9129	0.9511	0.9509	0.9602
	Boundary IoU	0.4686	0.4551	0.4538	0.4512
	Dice Coefficient	0.7595	0.7488	0.7486	0.7558
Data Set 2	Precision	0.6846	0.6599	0.6829	0.6699
	Recall	0.8604	0.9680	0.9052	0.9408
	F1 Score	0.7465	0.7590	0.7619	0.7613
	Intersection-over-Union (IoU)	0.6263	0.6444	0.6477	0.6470
	Boundary IoU	0.8051	0.8132	0.8777	0.8240
	Dice Coefficient	0.9045	0.8986	0.8484	0.8778

Table-1 presents the following analysis;

Dataset 1: The U-Net method generally performs the best in terms of Recall, F1 Score and IoU, indicating its robustness in segmenting the images accurately. However, the Experimental Evaluation method has the highest Boundary IoU and Dice Coefficient, suggesting it may be better at maintaining boundary accuracy.

Dataset 2: It shows the highest Precision and Boundary IoU but significantly lower Recall and F1 Score, indicating it may have high accuracy but misses many true positives. The Experimental Evaluation method performs well in terms of Boundary IoU and Dice Coefficient.

The selection of a segmentation technique is based on the specific needs of the task. If maintaining high precision and boundary accuracy is crucial, the Proposed Method might be suitable. However, for a balanced performance in Recall, F1 Score, and IoU, techniques like U-Net and Mask R-CNN are preferable. The Experimental Evaluation method shows strong performance in boundary metrics.

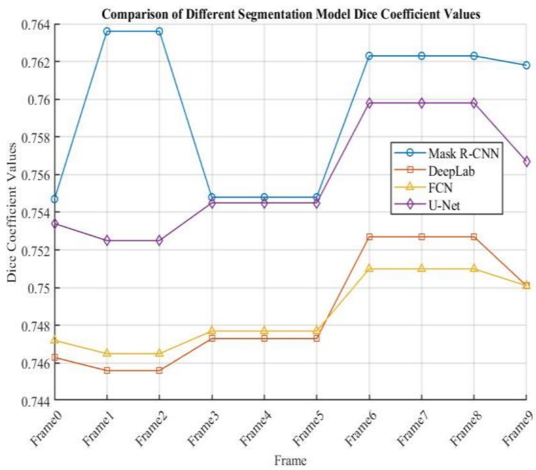

Fig 5- Analyzing Dice Coefficient Scores for Different Segmentation Methods

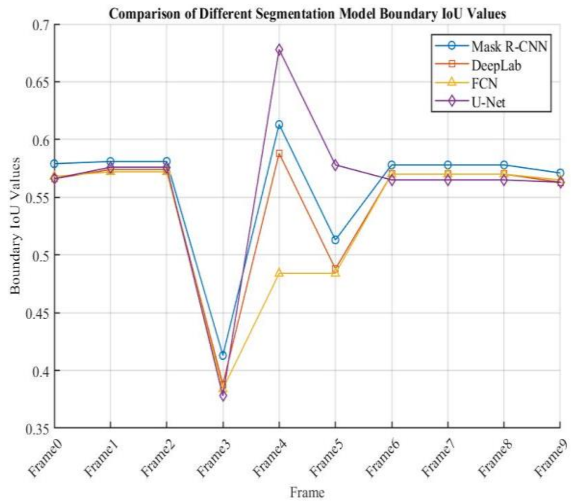

Fig 6- Comparison of Boundary IoU Values Across Various Segmentation Techniques

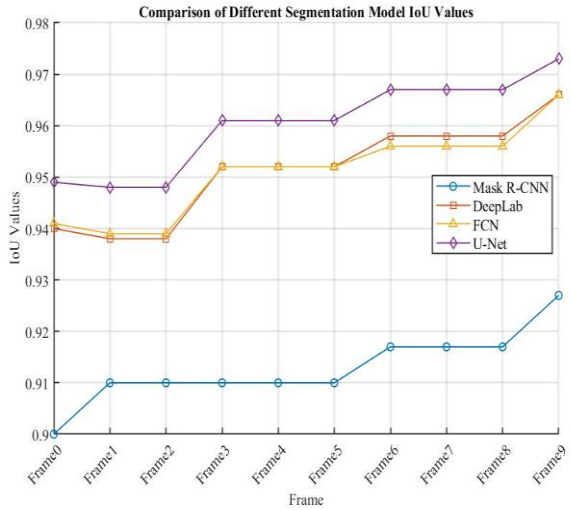

Fig 7- Comparison of IoU Values Across Various Segmentation Techniques

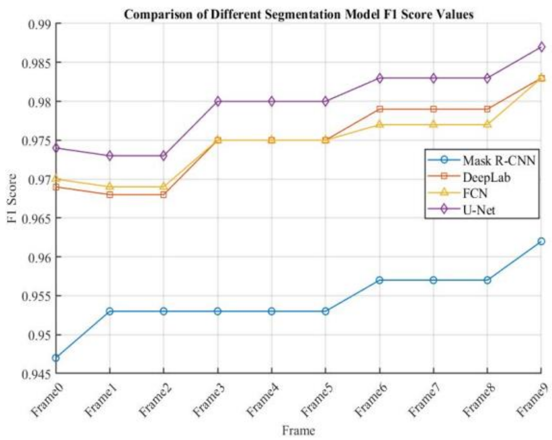

Fig 8- Analyzing F1-Score for Different Segmentation Methods

For a hybrid approach with GrabCut, Mask R-CNN and U-Net stand out as the most suitable candidates due to their high-quality initial segmentations and the potential for GrabCut to refine boundaries effectively. DeepLabv3 can also be a strong contender, especially with its multi-scale context capture, while FCN offers a simpler and faster alternative for less complex tasks.

U-Net is the best method overall for most applications due to the following reasons:

- Highest Recall (0.9607): It identifies the most relevant instances with the fewest false negatives.
- Highest F1 Score (0.9797): It balances precision and recall most effectively.
- Highest IoU (0.9602): It provides the best overlap between the predicted and actual regions.

However, Mask R-CNN performs better in terms of Boundary IoU and Dice Coefficient, indicating:

- Highest Boundary IoU (0.4686): It has better accuracy at the boundaries.
- Highest Dice Coefficient (0.7595): It suggests better similarity between the predicted and actual regions.

If boundary accuracy and similarity are more critical for your application, Mask R-CNN would be the preferred choice. However, for a balanced performance and overall best results in segmentation accuracy, U-Net is the recommended method.

6 Proposed Method Results and Analysis

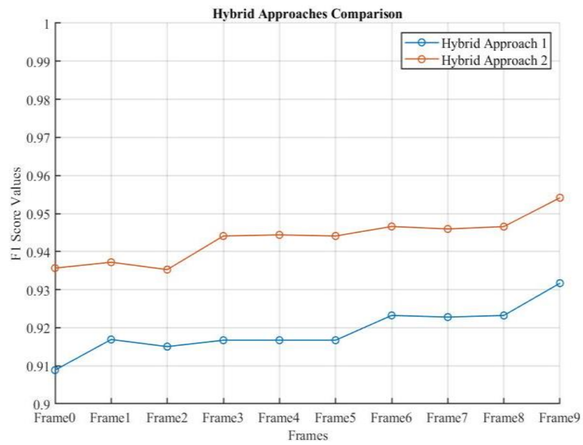

Fig 9 Assessment of F1 Score Values Performance in Hybrid Approaches 1 versus 2

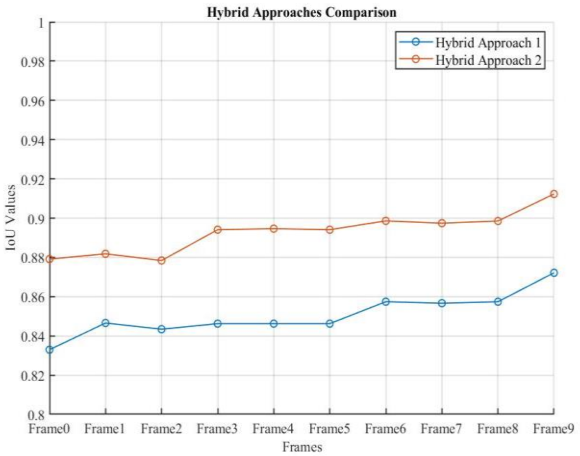

Fig 10- Assessment of IoU Values Performance in Hybrid Approaches 1 versus 2

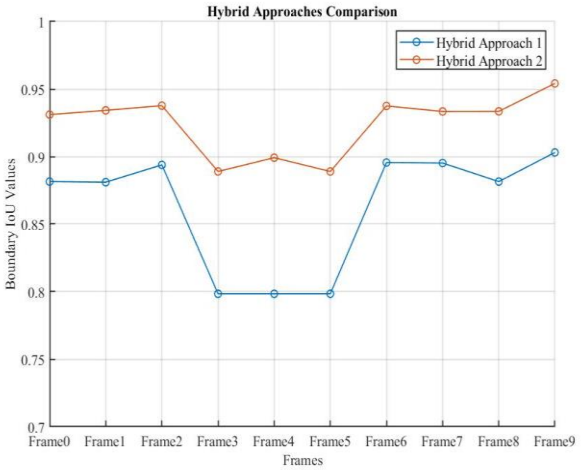

Fig 11- Assessment of Boundary IoU Values Performance in Hybrid Approaches 1 versus 2

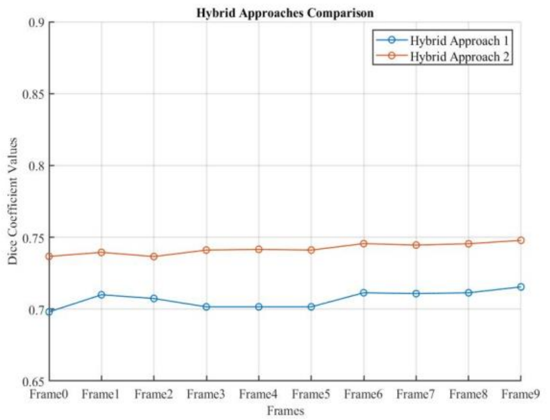

Fig 12- Assessment of Dice Coefficient Values Performance in Hybrid Approaches 1 versus 2

The provided graph in Fig 9 to Fig 12 compares the performance of two hybrid approaches on various frames using several evaluation metrics and explains below:

- Precision: This is the ratio of true positive results to the total number of positive results, which includes both true positives and false positives. A higher precision means fewer false positives.
- Recall: This is the ratio of true positive results to the actual number of positives. A higher recall indicates fewer false negatives
- F1 Score: This metric is the harmonic mean of precision and recall. It balances both metrics and is particularly useful for assessing performance when class distributions are imbalanced.
- IoU (Intersection over Union): This evaluates the overlap between the predicted segmentation and the ground truth. A higher IoU indicates more accurate segmentation.
- Boundary IoU: Similar to IoU but focuses on the boundaries of the segments. This is crucial for tasks where precise boundary delineation is important.
- Dice Coefficient: Similar to IoU, this metric focuses on the degree of similarity between the predicted segmentation and the ground truth.

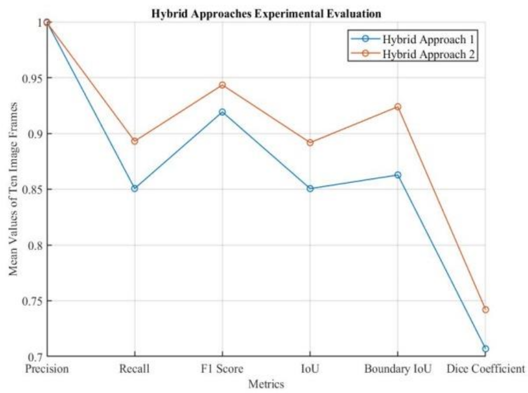

Fig 13- Comparative Study of Mean (Ten Image Frames) Evaluation Metrics for Hybrid Approach 1 and Hybrid Approach 2

Based on the provided metrics, Hybrid Approach 2 is the better method. It has higher recall, F1 score, IoU, Boundary IoU and Dice Coefficient, all of which are crucial metrics for evaluating the performance of a segmentation method. Although Hybrid Approach 1 has a slightly higher precision, the improvement in other metrics by Hybrid Approach 2 indicates a more balanced and overall better performance. This means Hybrid Approach 2 is better at capturing relevant instances while maintaining a good balance between precision and recall and it also shows superior performance in both region and boundary accuracy.

8 Conclusion and Future Work

Incorporating GrabCut hybridization in a research paper on segmentation offers a compelling narrative of how combining traditional algorithms with modern deep learning techniques can address limitations, improve performance and yield more accurate and reliable segmentation results. This approach highlights innovation in leveraging complementary methods to push the boundaries of what's achievable in image segmentation.

This article offers a comprehensive examination of five frequently used foreground detection algorithms that strive to achieve illumination invariance. Furthermore, we have presented an innovative approach for observation. Our primary objective was to evaluate the effectiveness of these algorithms in detecting foreground objects under varying illumination circumstances, including both sudden and gradual changes. In order to objectively evaluate their performance.

Based on the detailed evaluation metrics provided in the table, it is evident that both hybrid approaches achieve high performance in terms of precision. However, Hybrid Approach 2 consistently outperforms Hybrid Approach 1 in recall, F1 score, IoU, Boundary IoU and Dice Coefficient across all frames. This suggests that while both approaches are effective, Hybrid Approach 2 is more robust and reliable, particularly in terms of recall and segmentation accuracy. Therefore, Hybrid Approach 2 is the preferred method for tasks requiring high accuracy and precision in segmentation.

The above exploration establishes a strong foundation for further research in optimizing algorithm recall, enhancing scalability and generalization across diverse datasets and implementing real-time applications such as medical imaging and autonomous vehicles.

Future work could involve developing an enhanced hybrid approach leveraging advanced techniques like deep learning, conducting detailed error analysis for robustness, incorporating user feedback for iterative improvements and exploring cross-domain applications to demonstrate versatility and effectiveness in fields such as medical diagnosis, environmental monitoring and industrial inspection.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Zhaobin Wang, Yongke Lv, Runliang Wu & Yaonan Zhang, "Review of GrabCut in Image Processing," Mathematics, MDPI, vol. 11(8), pp. 1-41, April 2023, https://doi.org/10.3390/math11081965 .

[2] Hao Wu , Yulong Liu, Xiangrong Xu and Yukun Gao “Object Detection Based on the GrabCut Method for Automatic Mask Generation,” Micromachines, vol. 13, no. 12, pp. 2095, 2022. https://doi.org/10.3390/mi13122095.

[3] Hyungjoon Kim , Jae Ho Lee and Suan Lee “Hybrid Image Segmentation Method for Accurate Measurement of Urban Environments,” Electronics, vol. 12, pp. 1845, 2023, https://doi.org/10.3390/electronics 12081845

[4] Tiara Lestari Subaran, Transmissia Semiawan and Nurjannah Syakrani “Mask R-CNN and GrabCut Algorithm for an Image-based Calorie Estimation System,” Journal of Information Systems Engineering and Business Intelligence, vol. 8, no. 1, pp. 1-10, 2022, http://dx.doi.org/10.20473/jisebi.8.1.1-10.

[5] C Rother, V Kolmogorov, A. Blake, “GrabCut - Interactive foreground extraction using iterated graph cuts,” ACM Trans. Graph, vol. 23, pp. 309-314, 2004, https://dl.acm.org/doi/pdf/10.1145/1015706. 1015720.

[6] S Chouhan, A Kaul and U Singh, “Image segmentation using computational intelligence techniques: Review,” Archives of Computational Methods in Engineering, vol. 26, no. 3, pp. 533–596, 2018, https://doi.org/10.1007/s11831-018-9257-4.

[7] Y Li, X Liu, X Gong and M Wang, “A multi-view features hinged siamese U-net for image Co-segmentation,” Multimedia Tools and Applications, vol. 80, no. 15, pp. 22965-22985, 2021. https://doi.org/10.1007/s11042-020-08794-w.

[8] E Shelhamer, J Long and T Darrel, “Fully convolutional networks for semantic segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640-651, 2017, https://doi.org/10.1109/TPAMI.2016.2572683.

[9] K Liang, Y Jiang, M Long and G Liang, “Adaptive dual threshold based moving target detection algorithm,” IEEE 4th Information Technology and Mechatronics Engineering Conference, pp. 1111-1115, 2018, https://doi.org/10.1109/ITOEC.2018.8740402.

[10] X Chen and L Pan, “A survey of graph cuts/graph search based medical image segmentation,” IEEE Reviews in Biomedical Engineering, vol. 11, pp. 112-124, 2018, https://doi.org/10.1109/RBME.2018.2798701.

[11] F An and J Liu, "Medical image segmentation algorithm based on multilayer boundary perception-self attention deep learning model” Multimedia Tools and Applications, vol. 80, pp. 15017-15039, 2021, https://doi.org/10.1007/s11042-021-10515-w.

[12] A Farahani and H Mohseni, “Medical image segmentation using customized U-net with adaptive activation functions,” Neural Computing and

Applications, vol. 33, pp. 6307-6323, 2021, https://doi.org/10.1007/s00521-020-05396-3.

[13] K He, G Gkioxari, P Dollár and R Girshick, Mask R-CNN. IEEE international conference on computer vision, pp. 2961-2969, 2017, doi: 10.1109/ICCV.2017.322.

[14] L C Chen, G Papandreou, I Kokkinos, K Murphy and A L Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834-848, 2017.

[15] J Long, E Shelhamer and T Darrell, "Fully convolutional networks for semantic segmentation," IEEE conference on computer vision and pattern recognition, pp. 3431-3440, 2015.

[16] O Ronneberger, P Fischer, and T Brox, "U-Net: Convolutional networks for biomedical image segmentation," Medical image computing and computer-assisted intervention, pp. 234-241, 2015.

[17] C Rother, V Kolmogorov and A Blake, "GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts," ACM Transactions on Graphics, vol. 23, no. 3, pp. 309-314, 2004

[18] Xuchao Huang, Shigang Wang, Xueshan Gao , Dingji Luo, Weiye Xu , Huiqing Pang and Ming Zhou, "An H-GrabCut Image Segmentation Algorithm for Indoor Pedestrian Background Removal," Sensors, vol. 23, pp. 7937, 2023, https://doi.org/10.3390/s23187937.

[19] Peng Shi , Mengmeng Duan , Lifang Yang , Wei Feng , Lianhong Ding and Liwu Jiang, "An Improved U-Net Image Segmentation Method and Its Application for Metallic Grain Size Statistics," Materials MDPI, vol. 15, pp. 4417, 2022, https://doi.org/10.3390/ma15134417.