International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Comparative Study of Various Interactive Image Segmentation Models

Jasbeer Kaur[1], Vidhi Sharma[2]

[1]Research Scholar, Department of Computer Science, SSIET, IKGPTU, Punjab, India
[2]Assistant Professor, Department of Computer Science, SSIET, IKGPTU, Punjab, India

Abstract
Image segmentation is the process of dividing an image into multiple segments to extract meaningful information. It is a fundamental step in various computer vision tasks, including object detection, image recognition, and scene understanding. The goal of image segmentation is to divide an image into regions that are homogeneous with respect to certain characteristics, such as color, intensity, or texture. Interactive image segmentation (IIS) methods are usually evaluated in terms of segmentation performance vs. number of clicks (NoC). However, the automatic evaluation depends on a clicking procedure and its relation to the procedure used for training. In this work we compare qualitatively and quantitatively three state-of-the-art IIS methods that report the best performances but have not been compared against each other. From the research work it will be showed which method is better, Various objective performance evaluation parameters Intersection Over Union (IOU), Accuracy, Dice, F-Score, Precision and Recall values are used for providing worthiness of various interactive image segmentation of digital images. Matlab is used as a execution software for the proposed research work.

Keywords: Interactive Image Segmentation, Accuracy, Dice, Intersection Over Union, Precision, Recall

1. Introduction

Interactive image segmentation refers to segmenting an image with user input. Unlike traditional methods, interactive segmentation allows users to provide guidance to refine the segmentation results, enhancing accuracy and efficiency.

Interactive graph cut-based segmentation is a sophisticated technique in image processing that allows users to segment images by specifying a few seed points and letting the algorithm automatically find the optimal segmentation. This method is particularly useful for segmenting complex images with irregular shapes or multiple objects.

At the core of interactive graph cut-based segmentation is the graph representation of the image. In this representation, each pixel in the image is a node in the graph, and there are two special nodes, the source and sink. The source node is connected to the pixels inside the foreground region, and the sink node is connected to the pixels inside the background region. The edges between nodes are weighted based on the similarity of pixel intensities, with higher weights for pixels that are more similar.

The user starts by providing seed points, typically one inside the object to be segmented (foreground) and one outside (background). These seed points are used to define two initial sets of pixels, one for the foreground and one for the background. The algorithm then iteratively adjusts the segmentation boundary by cutting the graph to minimize the energy function, which is a combination of the data term

(related to pixel intensities) and the smoothness term (related to the smoothness of the segmentation boundary).

The interactive aspect of this method comes into play when the user can adjust the segmentation by adding or removing seed points and observing the effect on the segmentation result. This iterative process allows the user to fine-tune the segmentation until they are satisfied with the result.

One of the key advantages of interactive graph cut-based segmentation is its ability to handle complex image structures and object shapes. Unlike traditional thresholding or region-growing methods, which may struggle with irregular shapes or occlusions, graph cut-based segmentation can accurately segment objects even in challenging scenarios.

Example: In robotics, interactive graph cut segmentation can be used to segment objects in a cluttered environment. The user would add or remove edges between pixels to guide the segmentation algorithm. [2][3]

2. Literature Survey

R. Zou et al. [1] proposed an interactive image-segmentation technique for static images based on multi-level semantic fusion. The method used user-guidance information both inside and outside the target object to segment it from the static image, applicable to both 2D and 3D sensor data. The proposed method introduced a cross-stage feature aggregation module, enabling the effective propagation of multi-scale features from previous stages to the current stage. This prevented the loss of semantic information caused by multiple up-sampling and down-sampling of the network, allowing the current stage to make better use of semantic information from the previous stage. Additionally, a feature channel attention mechanism was incorporated to address the issue of rough network segmentation edges. This mechanism captured richer feature details from the feature channel level, leading to finer segmentation edges. In experimental evaluation on the PASCAL Visual Object Classes (VOC) 2012 dataset, the proposed method demonstrated an intersection over union (IOU) accuracy approximately 2.1% higher than the currently popular interactive image segmentation method in static images.

J. Sun et al. [2] In this paper, authors introduced a novel deep interactive image segmentation network that incorporates a feature-aware attention module to fuse human-click information with semantic features. This module was designed to seamlessly integrate with existing deep image segmentation networks, enabling these models to leverage user input for refining segmentation outcomes. Current experimental results demonstrated that the proposed module enhances the performance of image segmentation networks by improving the delineation of segmented objects. Notably, this method achieved significant performance gains when targeting specific objects from a selection of multiple targets with minimal user input. Moreover, current interactive approach was highly adaptable to various network backbones, resulting in substantial performance improvements across different architectures.

Q. Liu et al. [3] introduced PseudoClick, a novel framework that enhances existing segmentation networks by predicting candidate next clicks. These pseudo clicks mimic human input and are utilized to refine segmentation masks. PseudoClick was built upon established segmentation backbones, leveraging its click prediction mechanism to boost performance. Proposed framework was evaluated on 10 publicly available datasets spanning various domains and modalities, showcasing superior performance compared to existing methods and demonstrating robust generalization across different domains. Authors achieved state-of-the-art results on several benchmarks, notably surpassing the existing state-of-the-art by reducing the required number of clicks by 12.4% to achieve an 85% IOU on the Pascal dataset.

T. Kontogianni et al. [4] presented an interactive method for 3D instance segmentation, allowing users to collaborate iteratively with a deep learning model to segment objects directly in a 3D point cloud. Existing techniques for 3D instance segmentation typically relied on fully-supervised training, requiring extensive and expensive labeling efforts and often struggling to generalize to unseen classes. Limited research had explored obtaining 3D segmentation masks through user interactions, with most methods relying on 2D image-based feedback, necessitating frequent switching between 2D and 3D views and complex architectures for modal fusion, hindering integration with standard 3D models. Current approach enabled user to interact directly with 3D point clouds by selecting objects of interest or background elements, facilitating interactive segmentation in an open-world scenario.

K. Sofiiuk et al. [5] In this study, authors conducted a thorough evaluation of different design choices for interactive segmentation and discover that significant improvements in segmentation quality can be achieved without additional optimization schemes. Authors introduced a straightforward feedforward model for click-based interactive segmentation, which leverages segmentation masks from previous steps. This model not only facilitated the segmentation of entirely new objects but also allows for the refinement of external masks. Authors analysis revealed that the choice of training dataset significantly influences the performance of interactive segmentation models. Models trained on a combined dataset of COCO and LVIS, characterized by diverse and high-quality annotations, exhibit superior performance compared to existing models. Even this proposed baseline model, utilized the HR-Net18+OCR backbone, surpasses previous methods.

3. Problem Definition and Research Methodology

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract.

3.1 Problem Statement

Interactive image segmentation (IIS) methods play a crucial role in extracting objects of interest from images, but their performance is highly dependent on the number and quality of user clicks. Evaluating and comparing these methods is challenging due to the lack of standardized procedures and the sensitivity of results to clicking strategies. There is a need for a comprehensive comparative study to determine the effectiveness and robustness of different IIS methods in various scenarios and datasets. The research focuses on comparative analysis of various interactive image segmentation techniques.

3.2 Objectives

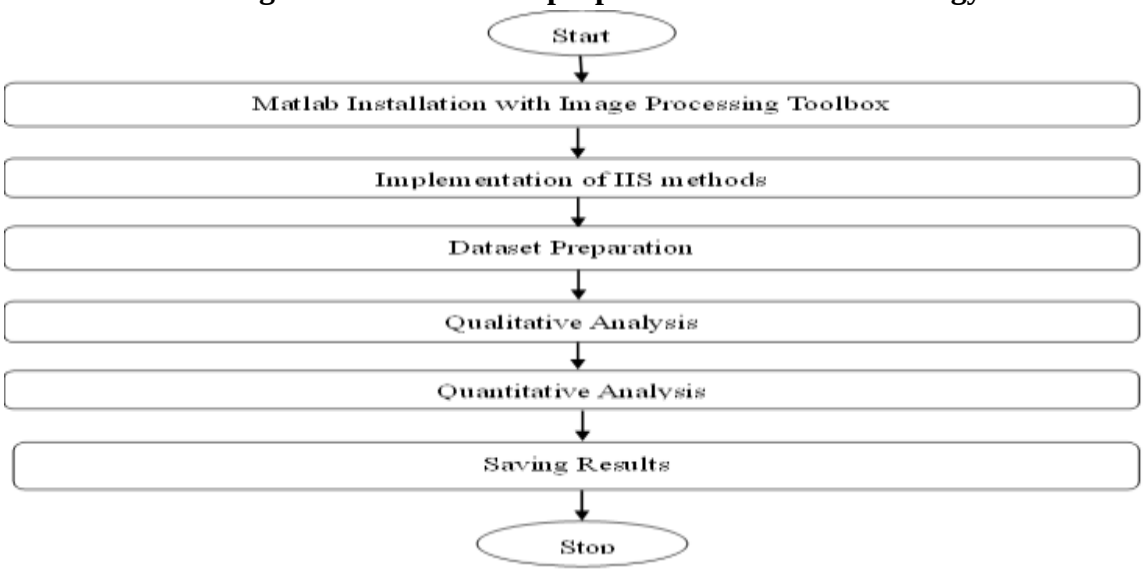
Consequently, objectives of this paper are:

1. To compare the segmentation performance of three state-of-the-art IIS methods across multiple datasets and clicking strategies.
2. To assess the sensitivity of IIS methods to variations in clicking procedures, including random, user-guided, and active learning-based approaches.
3. To determine the robustness of IIS methods to different clicking strategies and dataset characteristics.
4. To investigate the impact of training procedures, such as dataset selection, augmentation, and model architecture, on the performance of IIS methods.

3.3 Research Methodology

Step 1: Implementation of IIS Methods: Implement the three state-of-the-art IIS methods in MATLAB, ensuring they are compatible with the chosen datasets and clicking procedures.
Step 2: Dataset Preparation: Prepare the image datasets for evaluation, ensuring they cover a range of complexities and characteristics to provide a comprehensive assessment of the methods.
Step 3: Qualitative Evaluation: Conduct a qualitative evaluation of the segmentation results by visually inspecting the output masks generated by each method for different images in the dataset.
Step 4: Quantitative Evaluation: Utilize MATLAB's image processing and evaluation functions to quantitatively evaluate the segmentation results. Calculate metrics such as Intersection over Union (IoU), accuracy, and others to measure the performance of each method.
Step 5: Clicking Procedure Simulation: Simulate different clicking procedures, such as random, user-guided, and active learning-based approaches, to assess their impact on the performance of the methods.
Step 6: Robustness Analysis: Analyze the robustness of the methods to different clicking strategies and dataset characteristics. Determine which method performs better under various conditions.
Step 8: Results Comparison: Compare the qualitative and quantitative results obtained from the evaluation steps to determine the strengths and weaknesses of each method.

Figure 1: Flowchart of proposed research methodology



3.4 Implemented Methods

3.4.1 Reviving Iterative Training with Mask Guidance for Interactive Segmentation (RITM)

Step 1: Convolutional block that outputs the tensor of exactly the same shape as the first convolutional block in the backbone does.
Step 2: This tensor is then summed element-wise with the output of the first backbone convolutional layer, which has 64 channels.
Step 3: Choose a different learning rate for new weights without affecting the weights of a pre-trained backbone.

3.4.2 Getting to 99% Accuracy in Interactive Segmentation (GTO99)

Step 1: Take input stream and use ResNet-50 model
Step 2: Take Interaction Stream and apply mask estimates with user clicks

Step 3: Use Fusion Pyramid Pooling
Step 4: Use Decoder (U-Net Style)
Step 5: Use Final Layer along with Deep Guided Filter

3.4.3 Component - Trees

This algorithm is based on two steps:
Step 1: Component-tree computation of the image to be segmented
Step 2: Computation of the segmentation result based on a cost minimization.

3.5 Performance Parameters for Evaluation

A. Intersection Over Union (IOU)

Measures the overlap between the predicted segmentation mask and the ground truth mask.

$$\text{IOU} = \frac{\text{TP}}{\text{TP+FP+FN}} \qquad (4.1)$$

where:
TP = True Positives (intersection of predicted and ground truth masks)
FP = False Positives (pixels predicted as positive but not in ground truth)
FN = False Negatives (pixels not predicted as positive but in ground truth)

B. Accuracy

Measures the overall correctness of the segmentation results.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (4.2)$$

TN = True Negatives (correctly predicted as negative)

C. Precision

Measures the ratio of true positive pixels to the total number of pixels labeled as positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (4.3)$$

D. Recall

Measures the ratio of true positive pixels to the total number of actual positive pixels.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (4.4)$$

E. F1-score

Harmonic mean of precision and recall, provides a balance between the two metrics.

$$\text{F1} = 2 * \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (4.5)$$

F. Dice

$$\text{Dice} = \frac{2 \times \text{TP}}{(\text{TP} + \text{FP}) + (\text{FP} + \text{FN})} \qquad (4.6)$$

The Dice coefficient measures the ratio between the double of the intersection of estimated and ground truth masks and the sum of their areas

G. Visual Quality

Human observers rate the visual quality of the segmentation results based on how well they match the

ground truth and preserve object boundaries.

4. Results

4.1 Result of Flower Image

Figure 2 (a, b, c, d, e): Original flower image, mask image, ritm image, gto99 image, connected component image

(a) **(b)**

(c) **(d)** **(e)**

Table 1: Objective parametric values of flower image

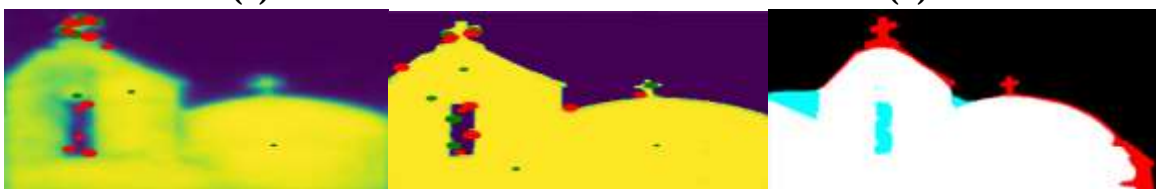
Parameters/Methods	IOU	Accuracy	Dice	F-Score	Precision	Recall
RITM	0.951	0.993	0.78	0.978	0.965	0.992
GTO99	0.976	0.987	0.988	0.988	0.987	0.987
CON-COMPONENTS	0.786	0.80	0.81	0.82	0.815	0.797

The results in the above table 1 shows that GTO99 performed better in comparison to RITM and Connected components methods. The value of parameters Intersection Over Union (IOU), Accuracy, Dice, F-Score, Precision and Recall values are better in comparison to other two techniques.

4.2 Result of Church Image

Figure 3(a, b, c, d, e): Original Church image, mask image, ritm image, gto99 image, connected component image

(a) **(b)**



(c) **(d)** **(e)**

Table 3: Objective Parametric Values of Church Image

Parameters/Methods	IOU	Accuracy	Dice	F-Score	Precision	Recall
RITM	0.992	0.994	0.980	0.98	0.994	0.982
GTO99	0.998	0.997	0.999	0.999	0.995	0.997
CON-COMPONENTS	0.881	0.884	0.894	0.880	0.885	0.889

The results in above table 2 shows that GTO99 performed better in comparison to RITM and Connected components methods. The value of all parameters are better in comparison to other two techniques.

4.3 Result of Star Fish Image

Figure 4(a, b, c, d, e): Original Star Fish image, mask image, ritm image, gto99 image, connected component image

(a) **(b)**

(c) **(d)** **(e)**

Table 3: Objective Parametric Values of Star Fish Image

Parameters/Methods	IOU	Accuracy	Dice	F-Score	Precision	Recall
RITM	0.96	0.99	0.98	0.98	0.978	0.99
GTO99	0.97	0.972	0.98	0.99	0.99	0.981
CON-COMPONENTS	0.812	0.88	0.82	0.81	0.90	0.91

The results in above table 3 shows that GTO99 performed better in comparison to RITM and Connected components methods. The value of parameters is better in comparison to other two techniques.

4.4 Result of Camel Image

Figure 5(a, b, c, d, e): Original Camel image, mask image, ritm image, gto99 image, connected component image

(a) **(b)**

(c) **(d)** **(e)**

Table 4: Objective Parametric Values of Camel Image

Parameters/Methods	IOU	Accuracy	Dice	F-Score	Precision	Recall
RITM	0.84	0.93	0.913	0.914	0.894	0.93
GTO99	0.87	0.95	0.91	0.93	0.91	0.95
CON-COMPONENTS	0.731	0.80	0.77	0.79	0.81	0.86

The results in above table 4 shows that GTO99 performed better in comparison to RITM and Connected components methods. The value of parameters is better in comparison to other two techniques.

5. Conclusion

In the present research work, a detailed comparative analysis is performed among three well known interactive image segmentation methods. From the results it is find that that GTO99 performed better in comparison to RITM and Connected components methods. The value of various parameters Intersection Over Union (IOU), Accuracy, Dice, F-Score, Precision and Recall values are better in comparison to other two techniques. Also, it is analyzed that GTO99 method performed better segmentation in various corners region of all the images. After GTO99, RITM performed better and various objective parameters proves the results. While the Connected Component method also performed will but not up to the level of other two methods.

In the future work other methods can be compared with these methods for performing more deep analysis of interactive image segmentation methods. Also, other objective parameters can also be taken for the detailed comparative analysis.

6. Acknowledgement

I am highly grateful to Dr. G. N. Verma, Co-Ordinator, SSIET, Derabassi for providing this opportunity to carry out the present research work. The constant guidance and encouragement received from Ms. Vidhi Sharma, Asst. Professor, Department of CSE, SSIET as a supervisor has been of great help so carrying the present work and is acknowledged with reverential thanks. I express gratitude to faculty members of CSE Department, SSIET for their intellectual support throughout the course of this work.

7. References

1. R. Zou, Q. Wang, F. Wen, Y. Chen, J. Liu, S. Du, C. Yuan, “An Interactive Image Segmentation Method Based on Multi-Level Semantic Fusion,” MDPI Sensors, vol. 23(14), pp. 6394, 2023.
2. J. Sun, X. Ban, B. Han, X. Yang, C. Yao, “Interactive Image Segmentation Based on Feature-Aware Attention,” Symmetry, pp. 1-13, vol. 14, 2022.
3. Q. Liu, M. Zheng, B. Planche, S. Karanam, T. Chen, M. Niethammer, Z. Wu, “PseudoClick: Interactive Image Segmentation with Click Imitation,” arXiv:2207.05282, 2022.
4. T. Kontogianni, F. Celikkan, S. Tang, K. Schindler, “Interactive Object Segmentation in 3D Point Clouds,” arXiv:2204.07183, 2022.
5. Sofiiuk, I. A. Petrov, A. Konushin, “Reviving iterative training with mask guidance for interactive segmentation,” arXiv:2102.06583, 2021.
6. Z. Lin, Z Zhang, L. Z Chen, M. M. Cheng, S. P. Lu, “Interactive image segmentation with first click attention,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13339–13348, 2020.
7. K. Sofiiuk, K. I. Petrov, O. Barinova, A. Konushin, “f-brs: Rethinking backpropagating refinement

for interactive segmentation," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8623–8632, 2020.