International Journal of Advances in Electrical Engineering

E-ISSN: 2708-4582
P-ISSN: 2708-4574
IJAEE 2023; 4(2): 12-29

www.electricaltechjournal.com
Received: 07-04-2023
Accepted: 09-05-2023

Devendra Tanaji Rane
Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunūn, Rajasthan, India

Dr. Prashant Kumbharkar
JSPMs Rajarshi Shahu College of Engineering, Pune, Maharashtra, India

Dr. Archana T Bhise
Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunūn, Rajasthan, India

Correspondence
Devendra Tanaji Rane
Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunūn, Rajasthan, India

Improved image segmentation technique for foreground extraction

Devendra Tanaji Rane, Dr. Prashant Kumbharkar and Dr. Archana T Bhise

DOI: https://doi.org/10.22271/27084574.2023.v4.i2a.41

Abstract
With the new era of Computer Technology, Computer vision has made tremendous progress. It has made a lot of improvements in Digital Image Processing. Digital Image processing techniques are typically categorized into three categories include Image Generation, Image Enhancement, and Image Restoration. Among multiple phases of Image processing, Segmentation Procedure is the area in which Image gets partitioned into its constituent parts or objects [1]. Image Segmentation is a method in which a digital image is broken down into various subgroups/regions called Image Segments which helps in reducing the complexity of the image to make further processing or analysis of the image simpler [3]. Many Image segmentation techniques are available based on two basic approaches Similarity/Region Approach and Discontinuity/Boundary Approach. Image Segmentation techniques are widely used in document processing, object recognition, remote sensing image, biomedicine, and many other aspects like factory production automation, Computed Tomography (CT) Images etc. Among existing foreground extraction techniques, the graph-based method Grabcut can effectively extract the foreground according to edges and appearance models [6]. Gaussian Mixture Model (GMM) is used for modelling the foreground and the background. Though this method is easy to use it has multiple manual interactions and iterations to achieve final segmentation. Accuracy and performance are the main problem areas with current GrabCut method of foreground extraction in image segmentation. Objective here is to develop Image pre-processing and initializing stage using a combination of binarization of mask and then pass it to GrabCut to establish a model of Image Background and Foreground to avoid manual interaction and to achieve more accuracy in segmentation.

Keywords: Digital Image Processing, Image Segmentation, Graphcut, Gaussian Mixture Model, Histogram Shape Analysis, Retinex Algorithm

Introduction

Background study

Digital image processing techniques are typically classified into three categories/steps. These categories include image generation, enhancement, and restoration. Generation techniques help project and recognize a scanned image, while the process of enhancing an image involves improving contrast, brightness and hue. Restoration techniques help eliminate and correct errors that do not accurately reflect the original picture [1].

An Image is defined as a two-dimensional function F(x, y), where x and y are spatial coordinates, and the amplitude of F at any pair of coordinates (x, y) is called the intensity of that image at that point. When x, y, and amplitude values of F are finite, we call it a digital image [2]. Digital images are represented in rows and columns as Matrix.

Phases of Image Processing

1. **Acquisition:** It could be as simple as being given an image which is in digital form. The main work involves: a) Scaling b) Color conversion (RGB to Gray or vice-versa)
2. **Image enhancement:** Used to extract some hidden details from an image and is subjective.
3. **Image Restoration:** It also deals with appealing to an image but it is objective (Restoration is based on mathematical or probabilistic model or image degradation).
4. **Colour Image Processing:** It deals with pseudocolour and full-colour image processing colour models are applicable to digital image processing.

5. **Wavelets and multi-resolution processing:** It is foundation of representing images in various degrees.
6. **Image Compression:** It involves in developing some functions to perform this operation. It mainly deals with image size or resolution.
7. **Morphological Processing:** It deals with tools for extracting image components that are useful in the representation & description of shapes.
8. **Segmentation Procedure:** It includes partitioning an image into its constituent parts or objects. Autonomous segmentation is the most difficult task in Image Processing.
9. **Representation & Description:** It follows output of the segmentation stage, choosing a representation is only the part of solution for transforming raw data into processed data.
10. **Object Detection and Recognition:** It is a process that assigns a label to an object based on its descriptor.

Overlapping Fields with Image Processing

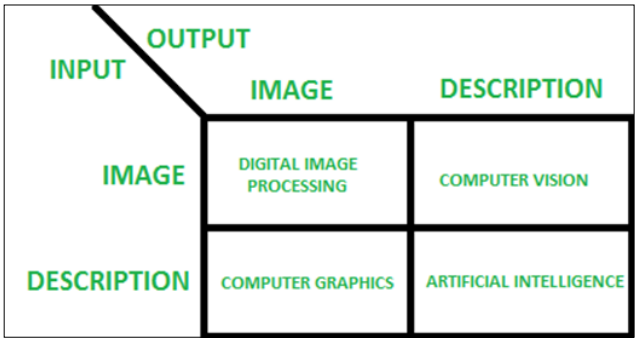


Fig 1: Image Processing Fields

Image segmentation is a method in which a digital image is broken down into various subgroups called Image segments which helps in reducing the complexity of the image to make further processing or analysis of the image simpler. Segmentation in easy words is assigning labels to pixels. All picture elements or pixels belonging to the same category have a common label assigned to them [3].

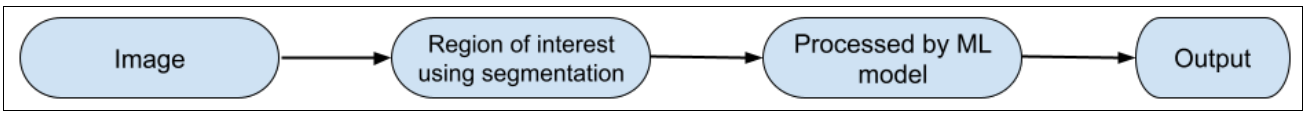


Fig 2: Image Segmentation

Approaches in Image Segmentation

- **Similarity approach (Region Approach):** This approach is based on detecting similarity between image pixels to form a segment, based on a threshold. ML algorithms like clustering are based on this type of approach to segment an image.
- **Discontinuity approach (Boundary Approach):** This approach relies on the discontinuity of pixel intensity values of the image. Line, Point, and Edge Detection techniques use this type of approach for obtaining intermediate segmentation results which can be later processed to obtain the final segmented image.

Image Segmentation Techniques

Based on the image segmentation approaches and the type of processing we have following techniques for segmentation

1. Threshold Based Segmentation.
2. Edge-Based Segmentation.
3. Region-Based Segmentation.
4. Clustering-Based Segmentation.
5. Watershed-Based Segmentation.
6. Artificial Neural Network Based Segmentation.

Threshold Based Segmentation

Image threshold segmentation is a simple form of image segmentation. It is a way to create a binary or multi-colour image based on setting a threshold value on the pixel intensity of the original image. In this threshold process, the intensity histogram of all the pixels in the image are considered. Then threshold is set to divide the image into sections.

Various threshold techniques are

1. Global threshold.
2. Manual threshold.
3. Adaptive threshold.
4. Optimal Threshold.
5. Local Adaptive Threshold.

Edge Based Segmentation

Edge-based segmentation relies on edges found in an image using various edge detection operators. These edges mark image locations of discontinuity in grey levels, colour, texture, etc. When we move from one region to another, the grey level may change. So if we can find that discontinuity, we can find that edge.

Region Based Segmentation

A region can be classified as a group of connected pixels exhibiting similar properties. The similarity between pixels can be in terms of intensity, colour, etc. In this type of segmentation, some predefined rules are present which have to be obeyed by a pixel in order to be classified into similar pixel regions. Region-based segmentation methods are preferred over edge-based segmentation methods in case of a noisy image. Region-Based techniques are further classified into 2 types based on the approaches they follow [4].

1. Region growing method
2. Region splitting and merging method

Clustering Based Segmentation

Clustering is a type of unsupervised machine-learning algorithm. It is highly used for the segmentation of images. One of the most dominant clustering-based algorithms used for segmentation is K-Means Clustering. This type of clustering can be used to make segments in a coloured image.

Watershed Based Segmentation

Watershed is a ridge approach, also a region-based method,

which follows the concept of topological interpretation. We consider the analogy of geographic landscape with ridges and valleys for various components of an image. The slope and elevation of the said topography are distinctly quantified by the grey values of the respective pixels – called the gradient magnitude. Based on this 3D representation which is usually followed for Earth landscapes, the watershed transform decomposes an image into regions that are called "catchment basins". For each local minimum, a catchment basin comprises all pixels whose path of steepest descent of grey values terminates at this minimum [5].

Artificial Neural Network Based Segmentation

The approach of using Image Segmentation using neural networks is often referred to as Image Recognition. It uses AI to automatically process and identify the components of an image like objects, faces, text, hand-written text etc. Convolutional Neural Networks are specifically used for this process because of their design to identify and process high-definition image data [5].

Literature Review

Image segmentation technology is a key step in the process of digital image processing and computer vision, and plays an important role in image processing technology. On the one hand, it can extract the object in the image, which has a very important impact on image recognition. On the other hand, based on the segmentation, recognition, characterization and measurement of statements, the target statement can transform the original image into the abstract form of the image, so as to analyse and understand the high-resolution image. So far, thousands of image segmentation methods have been developed. Image segmentation technology is also widely used in document processing, object recognition, remote sensing image and biomedicine and many other aspects. It also plays a very vital role in factory automation production. Digital image segmentation technology is mainly based on the similarity of some aspects and functions of the image itself to reshape the image. In the process of image segmentation, planning at a certain rate can improve the clarity of image pixels, and the image quality can be significantly improved [9]. In addition, it is important to establish a proper connection for the segmented image, and on this basis, it cannot be accessed and repeated. At the same time, it is important to ensure that the segmented image is highly consistent and the image will not change. Image segmentation and feature extraction transform the original image into abstract form for advanced image analysis and understanding, which lays a good foundation for better application of image segmentation technology [10].

Most the image segmentation techniques are based on two characteristics: discontinuity and similarity. In discontinuity-based algorithms, the image is partitioned based on change in the intensity. In the latter approach image is partitioned based on similar regions based on predefined criteria. Image segmentation can be very useful for line detection, point detection and edge detection which are available in many literatures [10]. An Image segmentation technique based on morphological tools is discussed by Hai Gao, *et al*. [11]. The main focus of their work is on image segmentation in video processing. It is a three-stage process including simplification, marker extraction and boundary decision to detect object.

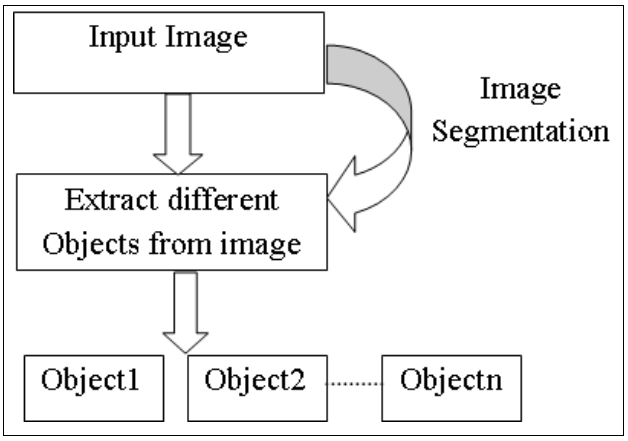


Fig 3: Block diagram of image segmentation process [10]

To detect fat content from beef images Lucia Ballerini *et al* [12] proposed an image segmentation technique based on Nuclear Magnetic Resonance (NMR), which is useful for that specific purpose. It contains three steps: background suppression (histogram thresholding), no uniformity removal (median filtering and subtraction), fat extraction (convolved image thresholding), for fat content extraction. An image segmentation technique for ultrasound images based on boundary extraction is presented by P. Abolmaesumi and MR Sirouspour [13]. The authors were able to achieve 98% accuracy of segmentation in less than 1 second. A wrapper-based approach presented by Farmer, ME and Anil K Jain [14] is based on region-based image segmentation technique. The authors achieved 91% accuracy among 2000 images. Some segmentation techniques are able to extract only one object from the image but there are some techniques in which multiple objects can be extracted. Ana C Teodoro *et al*. [15] developed image segmentation technique for extracting Douro River Plume Size from Medium Resolution Imaging Spectrometer (MERIS) data. This method uses two techniques: watershed and region based for extracting Douro River plum size. Some techniques are useful in medical imaging and can be very handy for special purposes. A system proposed by Mouloud Adel *et al*. [16] is based on Maximum a posteriori (MAP) probability criterion. It is used for detecting blood vessels from medical images of the retina. The authors were able to achieve 0.8 of maximum true positive rate (TPR) and corresponding false positive rate (FPR) as 0.094. Chitsaz, Mahsa and Seng Woo [17] proposed a system for detecting object from medical images. They developed software agent with Reinforcement Learning Approach for extracting several objects simultaneously from Computed Tomography (CT) images. Chirag Patel and Dr. Atul Patel proposed a threshold-based image binerization technique for vehicle number plate segmentation [9]. They converted original image to grey scale and then adaptive threshold technique is applied over it to convert image to binary form.

Over the years, computer-assisted algorithms have been used to aid the radiologists for interpreting the ultrasound images. The presence of speckle adversely affects the ultrasound image quality because of which accurate segmentation of tumours has become a challenging task. Kirti, JitendraVirmani and Ravindra Agarwal in their work [18], have listed various machine learning (ML) and deep

learning (DL) based approaches designed for segmenting breast ultrasound images have been reviewed over the past two decades using a characterization approach in terms of (a) datasets used, (b) pre-processing methods, (c) augmentation methods, (d) segmentation methods and (e) evaluation metrics used for the segmentation algorithms along with their brainstorming diagrams.

Greig *et al*. first introduced the theory of graph cut to the field of image processing in the late 1980s [19]. He proposed to minimize the energy function in computer vision by using the min-cut/max-flow algorithm in combinatorial optimization theory. Then Boykov and Jolly proposed an effective method of interactive image segmentation based on graph cut in 2001, which uses the mouse to click on some foreground pixels and background pixels, and realizes the global optimization with the help of graph cut technology [20]. Based on this Graph-cuts algorithm, Rother proposed the Grab-cut algorithm in 2004, by introducing the Gaussian mixture model (GMM) of colour pixels instead of the grey histogram model, the idea of iteration, and turning interactive operation from selecting seed point into surrounding the foreground with a rectangular frame [21]. Thus the segmentation of the grey image is successfully extended to the colour image segmentation. Xv Han *et al* [22] proposes to add the edge information of the image to the smoothness terms. Firstly, Gaussian filter is used to smooth the image, then the Sobel operator is used to perform edge detection on the smoothed image. Edge detection relies on the change of the entire surrounding neighbourhood pixels, instead of just the difference between two adjacent pixels, and on this account the influence of noise point on the constraint value can be reduced.

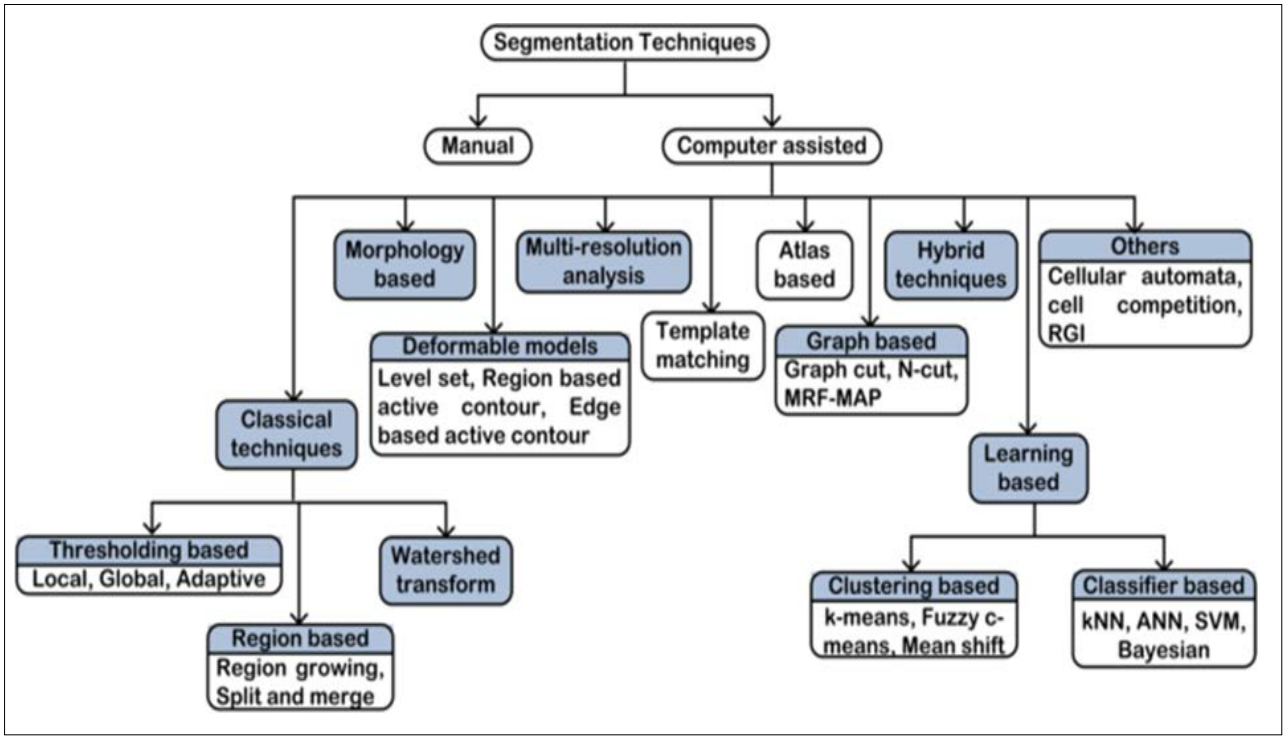


Fig 4: Segmentation Techniques and Algorithms [18]

Kun He *et al*. [6] in his work propose a novel appearance model which is formulated as maximum likelihood rather than the weighted sum of Gaussian. In this appearance model, the optimal number of Gaussians is estimated by the histogram shape analysis method, in which the number is automatically adjusted according to the intensity distributions of an image. Combining edges and appearance models, the foreground extraction is formulated as a joint optimization for the foreground extraction and appearance parameters with Graph Cut method.

Many attempts are made so far to improve the efficiency of the Grab-Cut algorithm, Zhai Li proposed a Grab-Cut algorithm based on super-pixels and improvement of features [23]. Zhao Yuan proposed a SAR image segmentation method based on Grab Cut and two-dimensional entropy algorithm [24]. At the same time, some scholars proposed an image segmentation method combining the Grab Cut algorithm and the watershed algorithm, and the universality and stability of the algorithm has been improved [25]. Tao Binjiao proposed an algorithm for image foreground extraction based on Grab Cut and the region growing algorithm. YaWei Yu *et al*. [26] proposes a foreground extraction algorithm for the target detection based on deep learning in combination of the sub-block region growing algorithm and the Grab Cut algorithm. Finally, the algorithm proposed takes the advantages of Grab Cut algorithm and the region growing algorithm, avoiding the shortcomings of the two algorithms. Another alternative to the sub-block region growing method is Graph-based image segmentation. Graph-based image segmentation, a classical segmentation algorithm, was presented by Felzenszwalb in 2004 [27]. The idea of the algorithm is very easy to understand, and the implementation is as well simple. And it is a reference by many computer vision, such as object recognition [28]. It works on pixel clustering.

NhatBaoSinh Vu *et al*. [29] present a set of novel image segmentation algorithms that utilize high-level semantic priors available from specific application domains. These priors are incorporated into the segmentation framework to further constrain the results to a more semantically meaningful solution space. Their algorithms are formulated

using Random Field models and employ combinatorial graph cuts for efficient optimization.

ImanAganj *et al.* [30] present a new atlas-based method for soft (i.e., fuzzy or probabilistic) segmentation of images, which – instead of attempting to determine a single correct label – produces the expected value of the label at each voxel of the new image, while considering the probability of possible atlas-to-image transformations. This is accomplished without either explicitly sampling from the transformation distribution (which would be intractable) or running the costly deformable registration in training or testing stages. It creates a single image from the training data, which is called the *key*. Then, for a new image (after affine alignment, if necessary), it computes the *expected label value* (*ELV*) map simply via convolution with the key, which is efficiently performed using the fast Fourier transform (FFT). Fuzzy ELV map is therefore a robust combination of labels suggested by atlas-to-image transformations, weighted by a measure of the transformation validity. This soft segmentation can be further used to initiate a subsequent hard-segmentation procedure.

Andrzej Brzoza *et al.* [31] propose a new approach to the segmentation of images based on shortest paths in a graph representation (SPG). A new texture descriptor is based on the spatial distribution of intensity levels in a neighbourhood. The local image region, statistics or property are considered over the textured region. This means that characterization by invariance of local attributes are distributed over a region of an image.

Vishal Lonarkar *et al.* [32] use 3D Colour Histogram and K-Mean clustering for segmentation. It uses the region-based histogram. For each region, it plots a separate histogram for better feature extraction. Histogram returns a density of pixel intensity in an image. In short, histogram finds the probability of pixel p of colour g occurring in the image I. For plotting the histogram important term is bin selection i.e. for better feature extraction proper selection of bins are required. If we take less number of bins then the histogram contain less components and it is unable to differentiate between two images, and if we take a large number of bins then more component present in the histogram so it will reject a very similar image. Also, it returns those images which are not similar. So ideal number of bins selection are required by a number of observations.

Main Contribution

- Experiment with combinational methods approaches for segmentation using histogram shape analysis, Grey-scale image and Binarization, K-nearest neighbour and gradient-based approach with Grabcut Method to improve on accuracy and performance.
- Avoid manual steps and iterations in the Grabcut approach by setting up prior information as per image category. This should improve segmentation accuracy.

Methodology

Working of an already existing algorithm

It is necessary to systematically summarize the existing segmentation methods, especially the state-of-the-art methods [34]. We elaborate on the working mechanisms of these methods and enumerate some influential image segmentation algorithms, and introduce the essential techniques of semantic segmentation systematically, as shown in Figure.

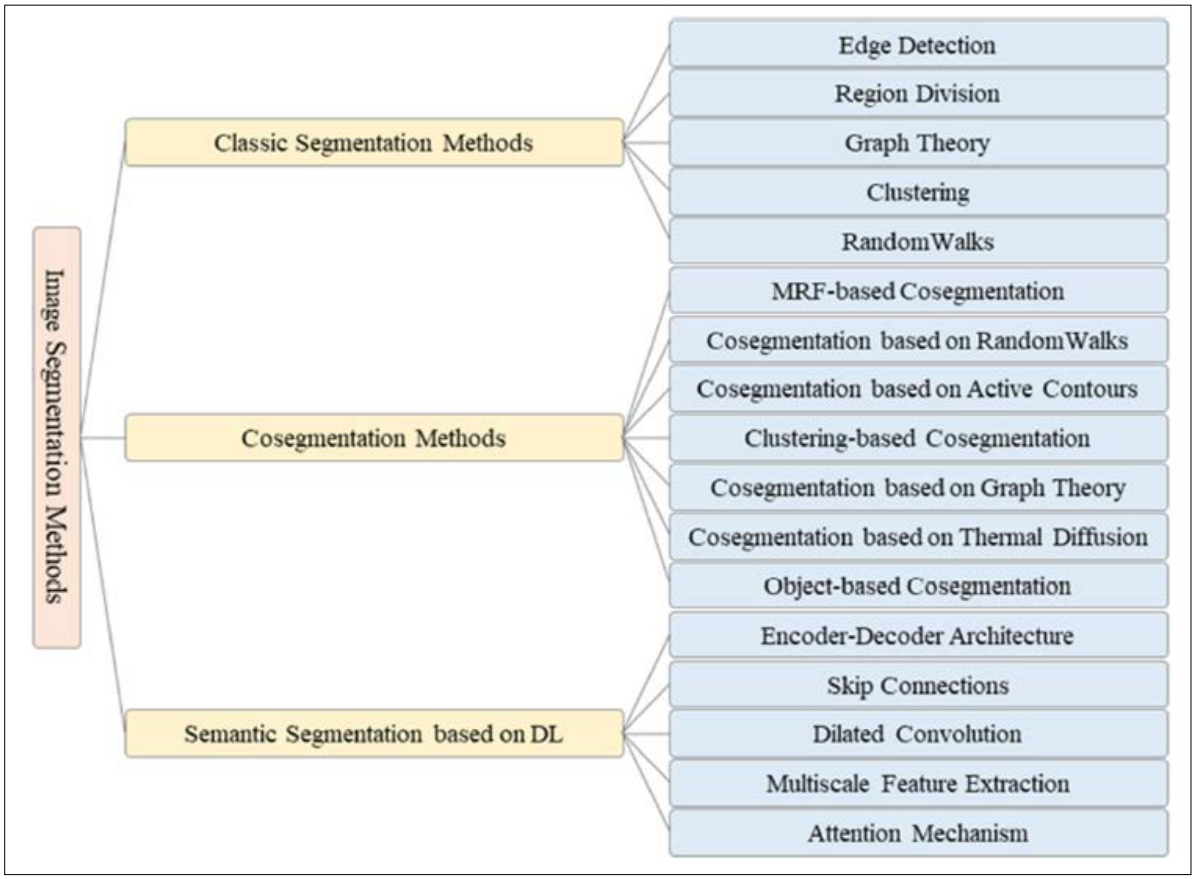


Fig 5: Categories of Image Segmentation

Classic Segmentation Methods

For grayscale images, the classic segmentation methods were suggested, which primarily take into account grey-level discontinuity in distinct regions and grey-level similarity within the same region. In general, grey-level similarity is the foundation for region division, while grey-level discontinuity is the foundation for edge detection. Using the similarity between pixels, colour image segmentation divides the image into various sections or superpixels, which are subsequently combined. The techniques in this segmentation category are listed below.

Edge Detection

Finding the spots on these boundaries is the goal of edge detection. One of the earliest segmentation techniques is edge detection, commonly known as the parallel boundary technique. To find the evident changes at the boundary, use the grey level's derivative or differential.

Region Division

Both serial and parallel region divisions are used in the region division approach. A typical parallel region division algorithm is thresholding. The grey histogram's trough value typically serves as the threshold, with the histogram's troughs being processed to make them deeper or turn them into peaks. The zeroth-order or first-order cumulant moment of the grey histogram can be used to calculate the ideal grayscale threshold in order to optimize categorization.

Graph Theory

The graph theory-based method for segmenting images transfers an image to a graph that encodes pixels or regions as graph vertices and the similarity between vertices as edge weights. Based on graph theory, image segmentation is defined as the division of vertices in the graph, weighted graph analysis using the principle and method of graph theory, and optimal segmentation using global graph optimization (e.g., the min-cut).

Clustering

The two most used techniques for clustering are K-means and GMM. Unsupervised clustering methods include K-Means and the Gaussian Mixture Model (GMM). Utilizing distance from the cluster centroid, K-Means organizes data points. GMM assigns data points to clusters probabilistically. While GMM takes into account both the mean and the variance of the data, k-means only examines the mean while updating the centroid. There are numerous K-means and GMM variations, such as Mean-shift, which uses density estimation to fit the image feature space to the probability density function, and SLIC (Simple Linear Iterative Clustering), which employs K-means to produce superpixels.

Random Walks

Random walks is a segmentation algorithm based on graph theory that is commonly used in image segmentation, image denoising, and image matching. By assigning labels to adjacent pixels in accordance with predefined rules, pixels with the same label can be represented together to distinguish different objects.

Co-Segmentation Methods

It is challenging to retrieve the high-level semantic information of an image when using the classical segmentation approaches, which typically concentrate on the feature extraction of a single image. The first time the idea of collaborative segmentation was put forth was by Rother *et al*. in 2006. Co-segmentation, also known as collaborative segmentation, is the process of automatically extracting the common foreground areas from a set of images in order to gain previous knowledge.

It is important to use a classical segmentation approach to extract the foreground elements of one or more photos (the seed image (s) as prior information in order to achieve co-segmentation. Then, using the prior knowledge, a collection of images containing the same or similar objects can be processed. The techniques in this segmentation category are listed below.

MRF-Based Co-Segmentation

A Markov Random Field (MRF) is a graph whose edges represent desired local impacts between pairs of random variables and whose nodes represent random variables. It is an undirected graphical model that depicts the interdependence of the random variables and aids in calculating their combined probability distribution. In order to address the problematic issues in multiple image segmentation, Rother *et al*. enhanced the MRF segmentation and made use of past information. The co-segmentation based on MRF has good universality and is frequently used in interactive image editing and video object identification and segmentation.

Co-Segmentation Based on Random Walks

Collins *et al*. developed a professional CUDA library to calculate the linear operation of the image sparse features, further exploited the quasiconvexity to optimize the segmentation algorithm, and expanded the random walks model to address the co-segmentation problem. By using a super voxel rather than a single voxel in their proposed optimized random walks algorithm for 3D voxel image segmentation, Fabijanska *et al*. significantly reduced the amount of time and memory required for computation. In order to combine subRW with other random walks methods for seed picture segmentation, Dong *et al*. devised a sub-Markov random walks (subRW) approach with previous label information. This algorithm successfully segmented photos with thin objects.

Random walk-based co-segmentation techniques offer considerable flexibility and robustness. They have had success in various medical image segmentation techniques, particularly 3D medical image segmentation.

Co-Segmentation Based on Active Contours

The energy function minimization by level set problem was resolved by Meng *et al*. by extending the active contour approach to co-segmentation, building an energy function based on foreground consistency between images and background inconsistency within each image, and using this energy function to segment images. In order to solve the problem of segmenting the brain MRI image, Zhang *et al*. proposed a deformable co-segmentation algorithm that converted the prior heuristic information of brain anatomy contained in multiple images into the constraints controlling the brain MRI segmentation and acquired the minimum energy function by level set.

Although the co-segmentation techniques based on active contours are effective at extracting the boundaries of

complicated structures, their unidirectional movement property significantly restricts their flexibility, making it difficult to identify and handle objects with weak edges.

Clustering-Based Co-Segmentation

An expansion of the clustering segmentation of a single image is clustering-based co-segmentation. A co-segmentation technique based on spectral clustering and discriminative clustering was proposed by Joulin *et al*. To achieve co-segmentation, they first employed spectral clustering to segment a single image based on local spatial information, and discriminative clustering to spread the segmentation findings across a group of images. The image was segmented into superpixels by Kim *et al*., who then employed spectral clustering to achieve co-segmentation. They used a weighted graph to express the relevance of the superpixels and then transformed the weighted network into an affinity matrix to describe the relationship of the intra-image.

Co-Segmentation Based on Graph Theory

An image is divided into a digraph via co-segmentation, which is based on graph theory. Meng *et al*. created a digraph by employing the local regions of each image as nodes, as opposed to superpixels or pixels, as they did in the previously stated digraph, which separated each image into multiple local areas based on object detection. Directed edges join nodes together, and the weight of those edges indicates how similar and important each object is to its surroundings. The issue of finding the shortest path on the digraph was then applied to the image co-segmentation problem. Finally, they used the dynamic programming (DP) approach to find the shortest route.

Co-Segmentation Based on Thermal Diffusion

By moving the heat source, thermal diffusion image segmentation increases the temperature of the system. Its aim is to locate the heat source's ideal location for the best segmentation results.

Object-Based Co-Segmentation

An object-based measurement technique was put forth by Alexe *et al*. to determine how likely it is that an image window will contain objects of any type. The highest scoring window was utilized as the feature calibration for each category of items in accordance with the Bayesian theory after calculating the likelihood that each sampling window contains an object. When items had distinct spatial bounds, the technique could differentiate between them.

Semantic Segmentation Based on Deep Learning

The richness of image details and the diversity of objects (e.g., scale, posture) have greatly increased with the ongoing advancement of image acquisition technology. The higher generalization ability of image segmentation models is advocated because low-level features, such as colour, brightness, and texture, are challenging to segment well and feature extraction techniques based on manual or heuristic rules cannot handle the complex requirements of current image segmentation.

Before deep learning was applied to the field of picture segmentation, semantic texton forests and random forest approaches were typically employed to build semantic segmentation classifiers. Deep learning algorithms have been used more and more in segmentation tasks over the last few years, and both the segmentation effect and performance have greatly increased. The original method uses small portions of the image to train a neural network, which then categorizes the pixels. The fully linked layers of the neural network require fixed-size images, hence this patch classification approach has been selected.

In order to allow for the input of any image size, Long *et al*. presented fully convolutional networks (FCNs) in 2015. The architecture of the FCN is depicted in Figure below. FCNs establish a basis for deep neural networks in semantic segmentation by demonstrating that neural networks can perform end-to-end semantic segmentation training. The FCN paradigm was used to advance subsequent networks.

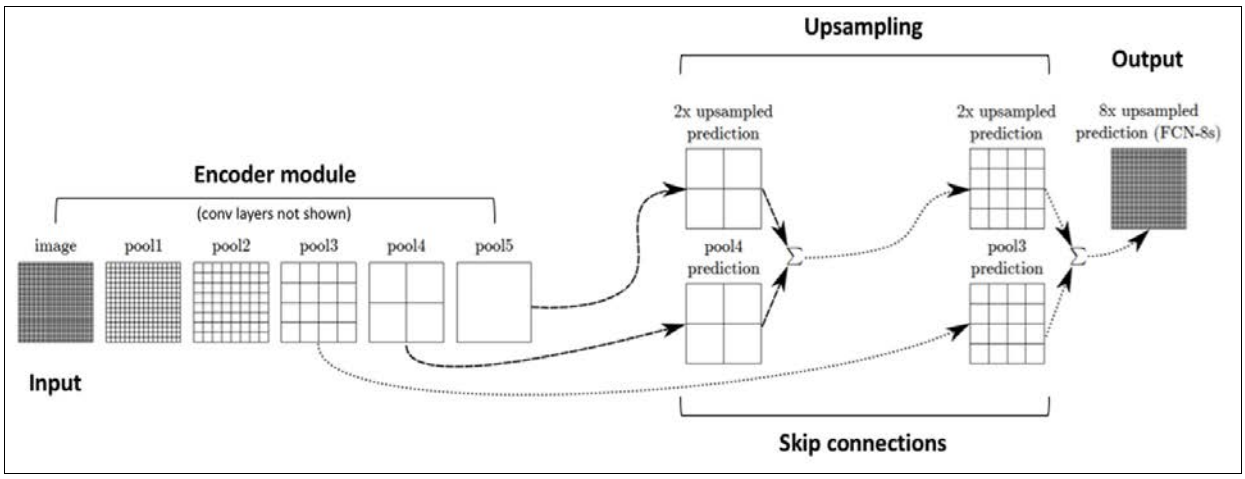


Fig 6: Fully Convolutional Networks architecture

Encoder-Decoder Architecture

On FCNs, encoder-decoder architecture is based. Convolutional neural networks (CNNs), whose output layers are the categories of images, such as LeNet-5, AlexNet, and VGG, obtained good results in image classification before FCNs. However, after gathering high-level semantic data, semantic segmentation must match the high-level features back to the original image size.

Convolution and pooling algorithms are mostly used at the encoder stage to extract high-dimensional features with semantic data. The convolution operation entails multiplying and adding the image-specific region pixel-for-

pixel using various convolution kernels, and then modifying the activation function to produce a feature map. The pooling operation entails sampling a predetermined area (the pooling window) and utilizing a predetermined sampling statistic as the region's representative characteristic. VGG, Inception, and ResNet are the three backbone blocks frequently utilized in segmentation network encoders.
In the decoder stage, an operation is done to turn the high-dimensional feature vector into a semantic segmentation mask. Up-sampling is the process of remapping the multi-level characteristics that the encoder extracted to the original image.

Skip Connections
To enhance pixel placement, skip connections and shortcut connections were created. A degradation concern with deep neural network training is that as the depth grows, performance declines. In ResNet and DenseNet, various skip connection architectures have been suggested as a solution to this issue. UNet, on the other hand, suggested a fresh long skip connection.

Dilated Convolution
To create dilated convolution, also referred to as atrous convolution, holes are inserted into the convolution kernel in order to increase the receptive field and decrease the computation required during down-sampling. To preserve the receiving field of the corresponding layer's receiving field and the high resolution of the feature map in FCN, the max-pooling layers are replaced with dilated convolution.

Multiscale Feature Extraction
The rich and deep level of features in the data is extracted using a multi-scale feature extraction method. The technique comprises of three basic feature extraction blocks that are structurally similar and vary mainly in the convolution kernel size. The final feature representation combines the features retrieved at various scales from each feature extraction block.

Attention Mechanisms
Some techniques frequently used in the field of natural language processing (NLP) have been applied to computer vision, with good results in semantic segmentation, to represent the dependency between various regions in an image, especially the long-distance regions, and obtain their semantic relevance. In the field of computer vision, the attention mechanism was initially proposed in 2014. Attention mechanisms became steadily more common in image processing jobs since the Google Mind team chose the recurrent neural network (RNN) model to apply attention mechanisms to picture categorization.
Vaswani *et al*. proposed the transformer in 2017, a deep neural network that completely dispenses with convolutions and repetition and is exclusively based on a self-attention mechanism. Transformer and its variations, such as X-transformer, were then used to the study of computer vision. The enhanced network made some strides thanks to CNN's pre-training model and the transformer's self-attention mechanism. A vision transformer (ViT), suggested by Dosovitskiy *et al*., demonstrated that it could replace CNN in the classification and prediction of picture patch sequences. They separated the image into fixed-sized patches, arranged the patches in a straight line, and then input the patches sequence vector into a transformer encoder (the right-hand design), which alternated between multi-head attention layers and multi-layer perception (MLP), as shown in Figure below.

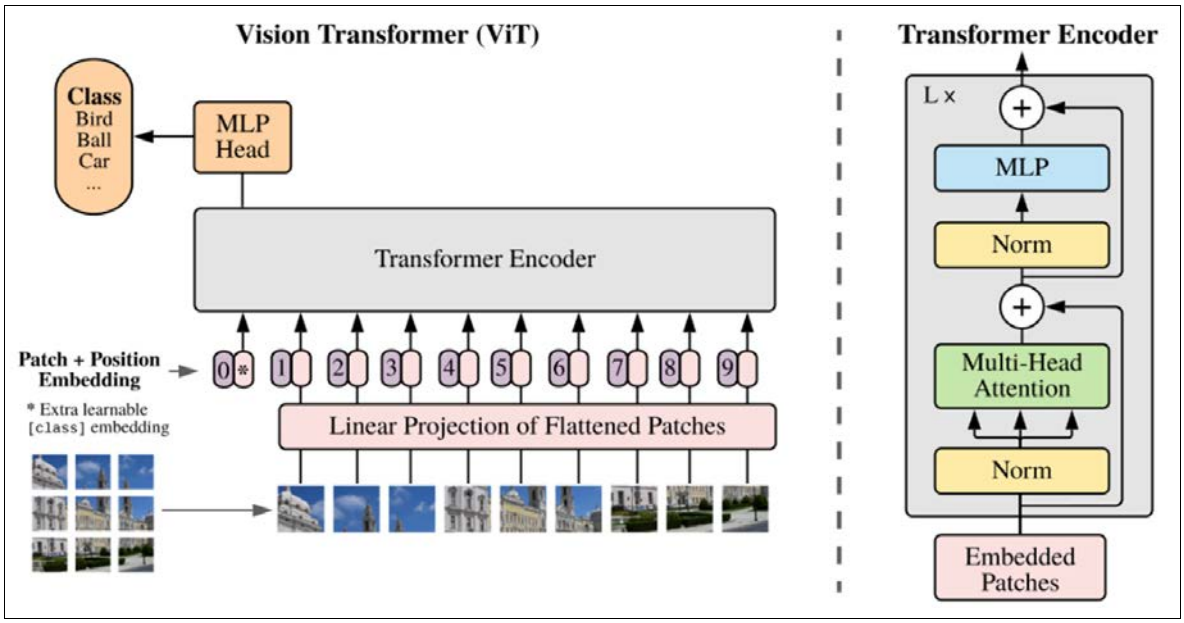


Fig 7: ViT Model [33]

Liu *et al*. developed the swin transformer that has achieved impressive performance in image semantic segmentation and instance segmentation. The swin transformer advanced the sliding window approach, that built hierarchical feature maps by merging image patches in deeper layers, calculated self-attention in each local window, and utilized cyclic-shifting window partition approaches alternatively in the consecutive swin transformer blocks to introduce cross-window connections between neighbouring nonoverlapping windows. The swin transformer network replaced the standard multi-head self-attention (MSA) module in a transformer block with shifted window approach, with the other layers remaining the same, as shown in Figure below.

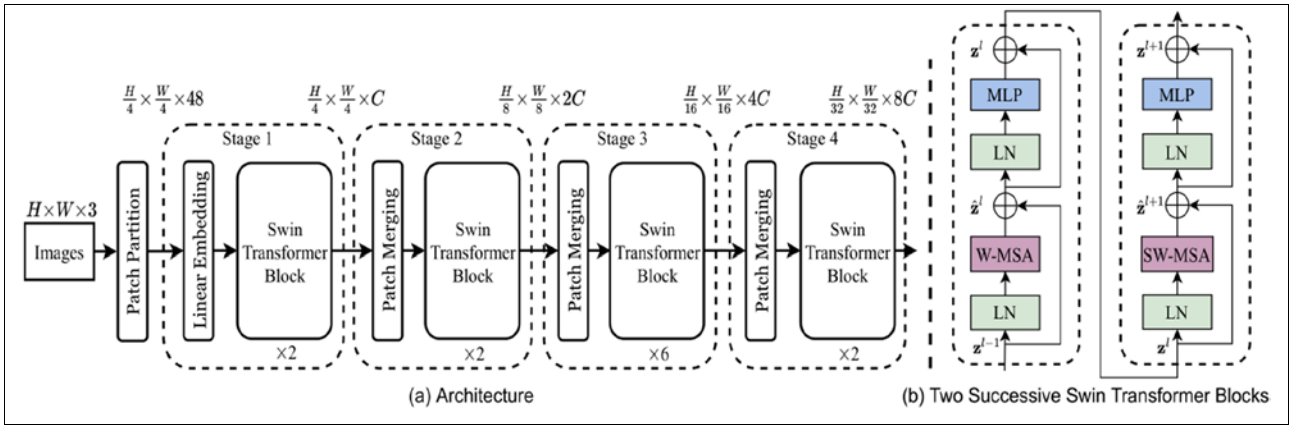


Fig 8: SWIN Transformer Architecture [33]

Problem Statement

Image segmentation, which has become a research hotspot in the field of image processing and computer vision, refers to the process of dividing an image into meaningful and non-overlapping regions, and it is an essential step in natural scene understanding. Despite decades of effort and many achievements, there are still challenges in feature extraction and model design.

Image segmentation is one of the most popular research fields in computer vision, and forms the basis of pattern recognition and image understanding. The development of image segmentation techniques is closely related to many disciplines and fields, e.g., autonomous vehicles, intelligent medical technology, image search engines, industrial inspection, and augmented reality. Image segmentation divides images into regions with different features and extracts the regions of interest (ROIs). These regions, according to human visual perception, are meaningful and non-overlapping. There are two difficulties in image segmentation:

(1) how to define "meaningful regions", as the uncertainty of visual perception and the diversity of human comprehension leads to a lack of a clear definition of the objects, it makes image segmentation an ill-posed problem; and (2) how to effectively represent the objects in an image.

Digital images are made up of pixels, that can be grouped together to make up larger sets based on their colour, texture, and other information. These are referred to as "pixel sets" or "super pixels". These low-level features reflect the local attributes of the image, but it is difficult to obtain global information (e.g., shape and position) through these local attributes.

Since the 1970s, image segmentation has received continuous attention from computer vision researchers. The classic segmentation methods mainly focus on highlighting and obtaining the information contained in a single image that often requires professional knowledge and human intervention. However, it is difficult to obtain high-level semantic information from images.

Co-segmentation methods involve identifying common objects from a set of images that requires the acquisition of certain prior knowledge. Since then image datasets, image segmentation methods based on deep neural networks have gradually become a popular topic. Although many achievements have been made in image segmentation research, there are still many challenges, e.g., feature representation, model design, and optimization. In particular, semantic segmentation is still full of challenges due to limited or sparse annotations, class imbalance, over fitting, long training time, and gradient vanishing.

Among existing foreground extraction techniques, the methods based on the graph cut can effectively extract the foreground according to edges and appearance models [6]. In this algorithm, the region is drawn in accordance with the foreground, a rectangle is drawn over it which encompasses main object. The region coordinates are decided over understanding the foreground mask. But this segmentation is not perfect, as it may have marked some foreground regions as background and vice versa. This problem can be avoided manually. This foreground extraction technique functions just like a green screen in cinematics [7]. Following are the steps are involved in it,

- Region of Interest (ROI) is decided by the amount of segmentation of foreground and background is to be performed and is chosen by the user. Everything outside the ROI is considered as background and turned black.
- Then Gaussian Mixture Model (GMM) is used for modelling the foreground and the background. Then, in accordance with the data provided by the user, the GMM learns and creates labels for the unknown pixels and each pixel is clustered in terms of colour statistics.
- A graph is generated from this pixel distribution where the pixels are considered as nodes and two additional nodes are added that is the Source node and Sink node. All the foreground pixels are connected to the Source node and every Background pixel is connected to the Sink node. The weights of edges connecting pixels to the Source node and to the End node are defined by the probability of a pixel being in the foreground or in the background.
- If huge dissimilarity is found in pixel colour, the low weight is assigned to that edge. Then the algorithm is applied to segment the graph. The algorithm segments the graph into two, separating the source node and the sink node with the help of a cost function which is the sum of all weights of the edges that are segmented.
- After the segmentation, the pixels that are connected to the Source node is labelled as foreground and those pixels which are connected to the Sink node is labelled as background. This process is done for multiple iterations as specified by the user. This gives us the extracted foreground.

As explained in above steps lot of manual interactions and multiple iterations are required in this method to achieve

foreground extraction. An appearance model used is a statistical model for the intensity/ colour distributions. The existing appearance models are broadly categorized into the local histogram and the Gaussian Mixture Models (GMMs), the former can explicitly represent the intensity/ colour distributions of the user-labeled pixels; however, the estimation accuracy of appearance parameters varies with the user interaction. The GMM makes use of the all pixels to estimate appearance parameters in a better way. The accuracy of the GMMs depends on i) the number of Gaussian in each GMM; and ii) the representation form. The original GMM is defined as the weighted sum of Gaussians, which cannot effectively distinguish the small regions in the foreground and background [6].

Thus accuracy and performance are the main problem areas with current graphcut method of foreground extraction in image segmentation.

Classical segmentation methods like Grabcut involves manual steps and on the other hand Deep Learning based segmentation algorithms like Mask R-CNN can automatically predict both the bounding box and the pixel-wise segmentation mask of each object in an input image. The downside is that masks produced by DL based algorithms like Mask R-CNN aren't always "clean". There is always a bit of background that gets mixed into the foreground segmentation which affects the accuracy of segmentation.

Working Flow of Existing Grabcut

Grabcut is an interactive foreground extraction algorithm designed by Carsten Rother, Vladimir Kolmogorov and Andrew Blake in 2004. In Grabcut it labels each pixel as the foreground or the background of the image.

Fig 9: Labelling each part of input image: B-Background, F-Foreground

Graph Modeling [34]

To obtain the segmentation, GrabCut uses the graph-like structure of an image. Each pixel has several links like (1) "n-link" to each of its 4 direct neighbours and (2) "t-link" to the source and sink nodes of the graph which represents the image foreground and background respectively.

After graph construction, the image segmentation task consists of finding the cut of minimal cost that separates foreground and background as shown in figure below,

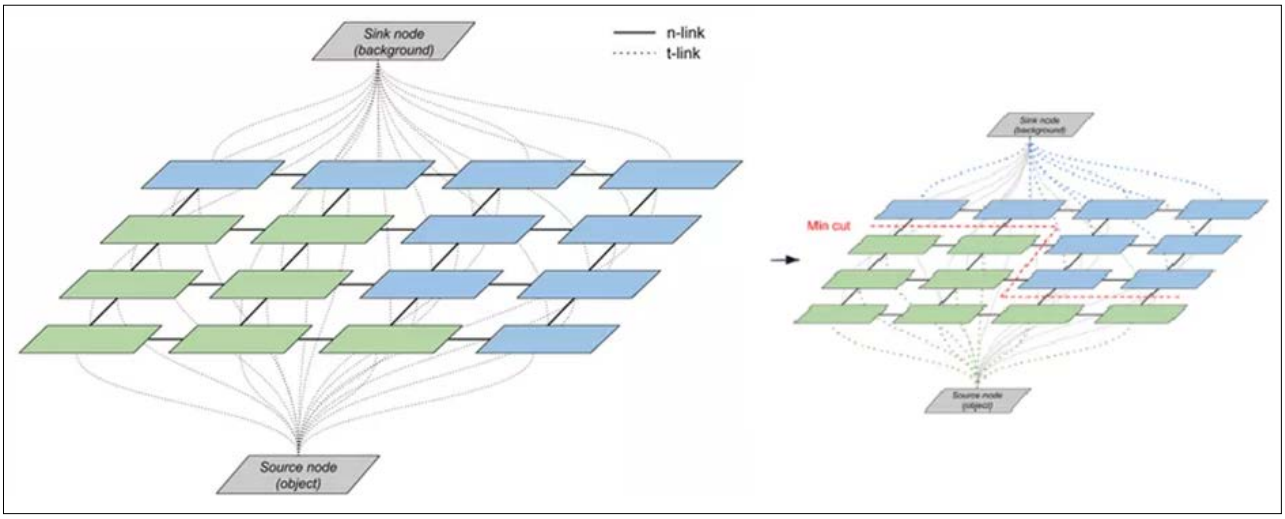


Fig 10: Link each pixel to source/sink node and then use min cut function to separate out foreground

Define the weights for both the links

"n-links" represents information about a pixel and its direct neighbourhood pixels.

$$V(\underline{\alpha},\mathbf{z}) = \gamma \sum_{(m,n)\in\mathbf{C}} [\alpha_n \neq \alpha_m] \exp -\beta \|z_m - z_n\|^2$$

The data term takes into account n-link weights between pixels cut by the segmentation (different alpha values). The more similar the pixels are, the higher the cost. Alpha represents pixels labels and z pixels intensities

"t-links" represent information about colour distribution in the foreground and the background of the image. A t-link weight shows how well a pixel fits the background/foreground model. Gaussian Mixture Models (GMMs) models them in GrabCut.

$$U(\underline{\alpha},\mathbf{k},\underline{\theta},\mathbf{z}) = \sum_n D(\alpha_n,k_n,\underline{\theta},z_n)$$

The smoothness term takes into account background colour modelling. Alpha represents pixels labels, k, and theta GMMs parameter, and z pixels intensities. Value $D(\alpha n, kn,\theta,zn) = -\log p(zn \mid \alpha n, kn,\theta) - \log\pi(\alpha n, kn)$, and $p(\cdot)$ is a Gaussian probability distribution, and $\pi(\cdot)$ are mixture

weighting coefficients [35].
Once the weights are defined, the cost function or energy function E is their sum over the graph:

$$\mathbf{E}(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) = U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) + V(\underline{\alpha}, \mathbf{z}),$$

Where E is the graph energy function. Alpha represents pixels labels, k, and theta GMMs parameter, and z pixels intensities. U is the smoothness term and V the data term, representing respectively global and local information
Once the cost function is defined this is classic graph theory mincut problem which can be easily solved to get foreground extracted. Grabcut is an iterative method so initial foreground segmented image will be used to re-calculate graph weights and then redefine the image segmentation again.

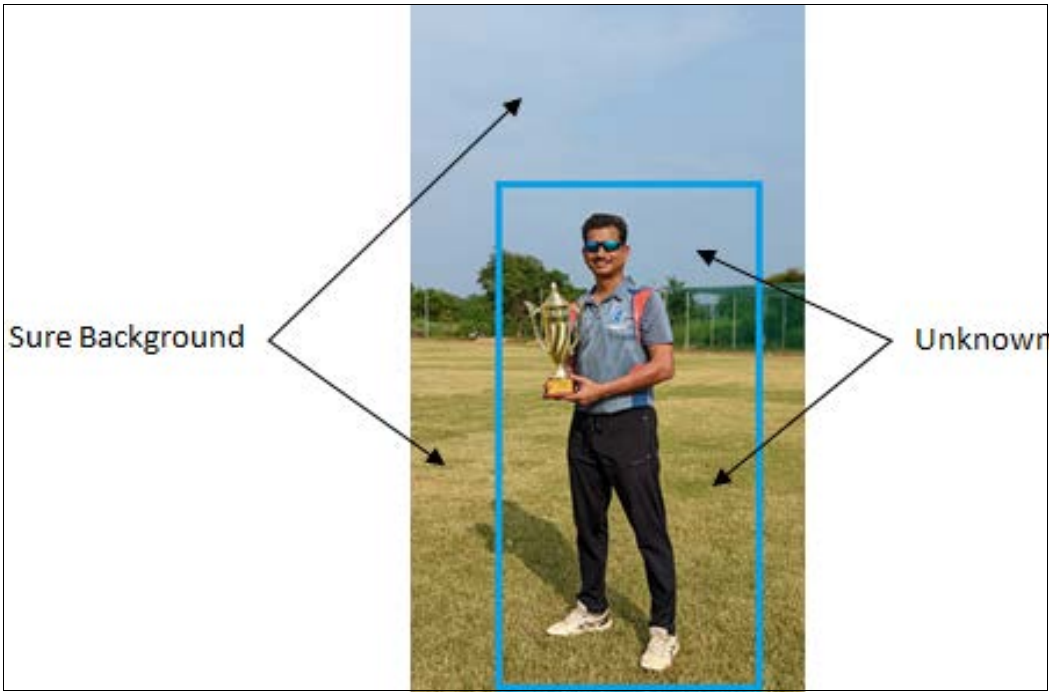


Fig 11: Initial marking of rectangle by user to mark region of interest

Refine the Segmentation
GrabCut is different than the hard segmentation methods of foreground and background segmentation. Once the iterative loop completes, Pixel labels of region of interest (ROI) are refined and classified into four groups like sure background, probable background, probable foreground, and sure foreground.

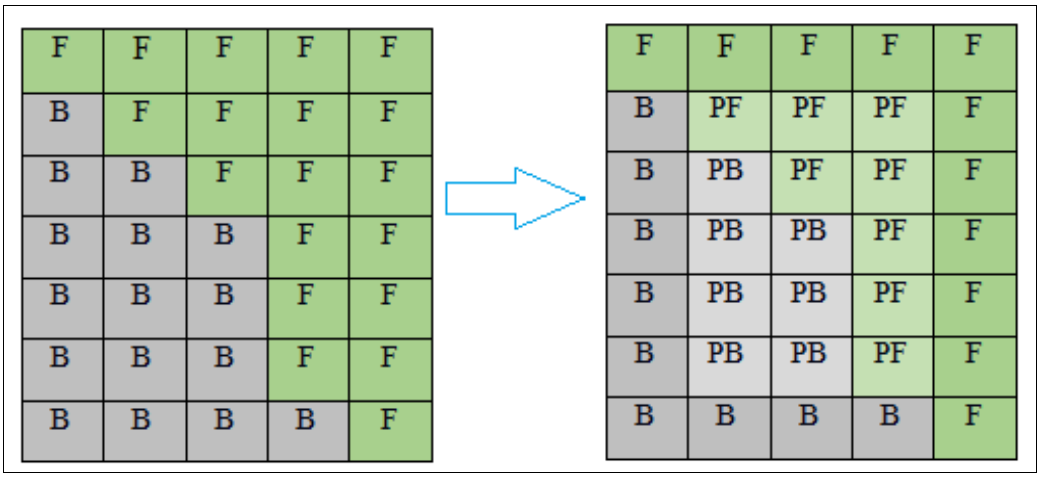
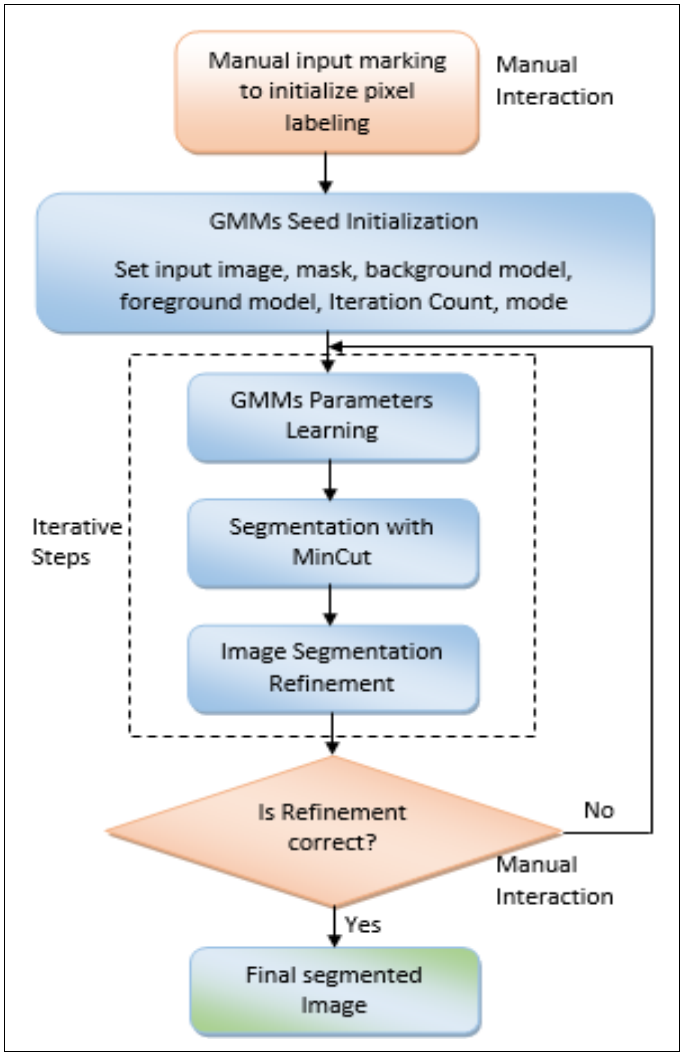


Fig 12: Grabcut segmentation Refinement: PB- Probable Background, PF- Probable Foreground

Grabcut flow



After multiple iterations and manual interactions we get final segmented image with current GrabCut method as shown below.

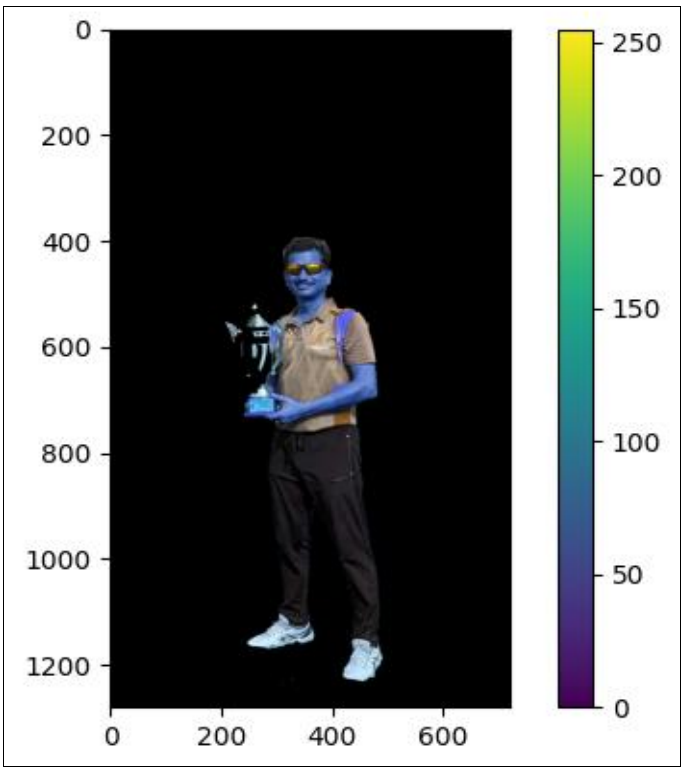


Fig 13: Final segmented Image after multiple iterations and manual interaction with GrabCut (Blue channel image)

Proposed methodology

Original GrabCut doesn't give satisfied output without multiple manual interactions. Check multiple iterations and its outputs below for original image.

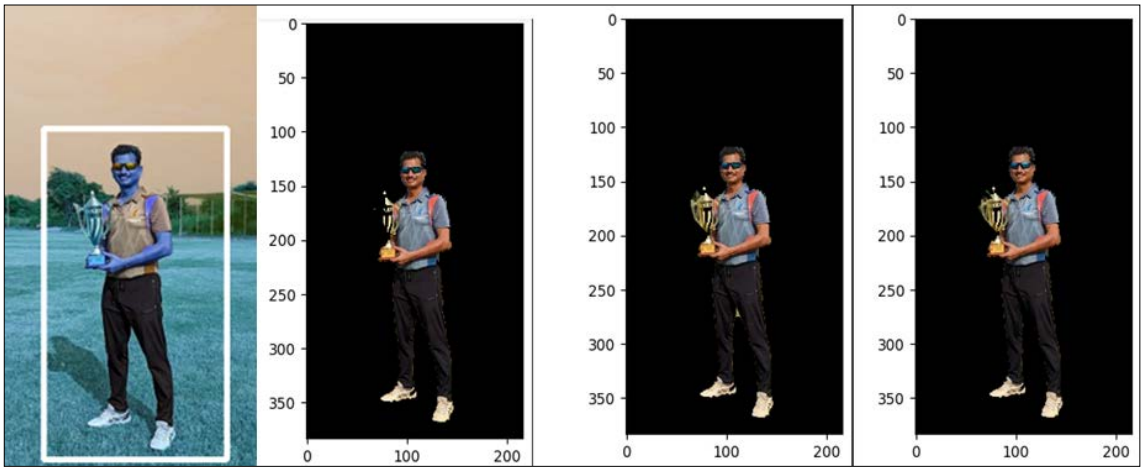


Fig 14: Segmentation with 20 iterations of GrabCut and 3 manual interactions

First image above is an input image. After 5 iterations of GrabCut, it gives 2nd image in which some part of trophy is not visible. We have to mark foreground manually here which will give us 3rd image after 5 iterations of Grabcut. But in 3rd output there is again some unwanted background is seen between the legs which again we have to mark as background and rerun GrabCut for 5 more iterations. Finally after multiple manual interactions and iterations of GrabCut it gives 4th image in which complete foreground is segmented from original image. Also if you observe edges are also not very smooth.

To improve this we propose to add edge-level analysis with grayscale in initialization stage and then apply binarization on it. Deep Learning approaches like Mask-RCNN also can be used with or without edge-level analysis. Deep Learning approaches improves processing speed but accuracy-wise edge level analysis with GrabCut is better approach.

Let's see how each iteration of original GrabCut improves segmentation. See 5 iterations of GrabCut below after inputting image with rectangle manually.

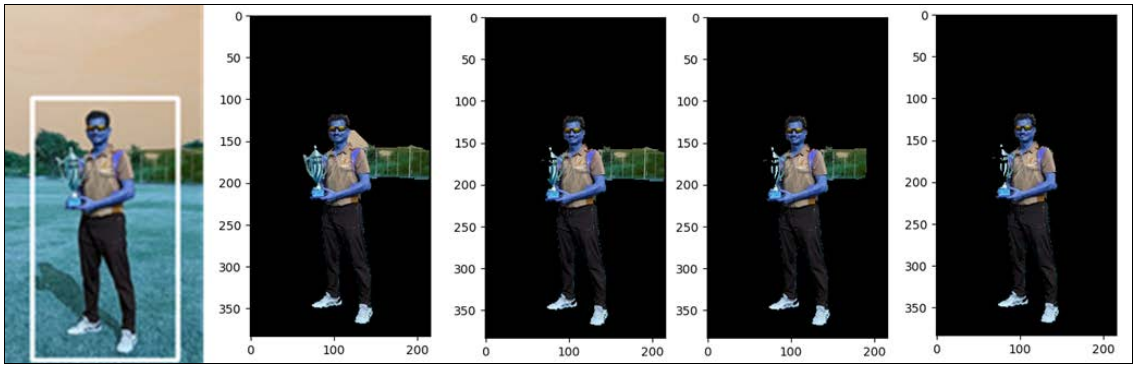


Fig 15: Segmentation with Original GrabCut with single iteration

To avoid number of iterations we can provide accurate input mask instead of input rectangle mask with the help of edge-level analysis and binarization. Below image shows the steps taken to improve on accuracy and to reduce number of iterations.

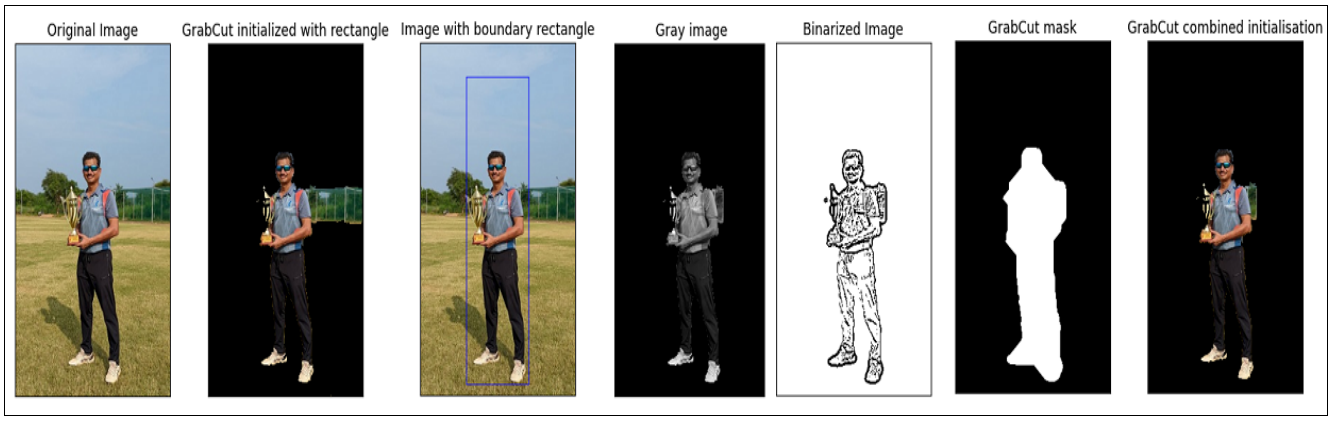
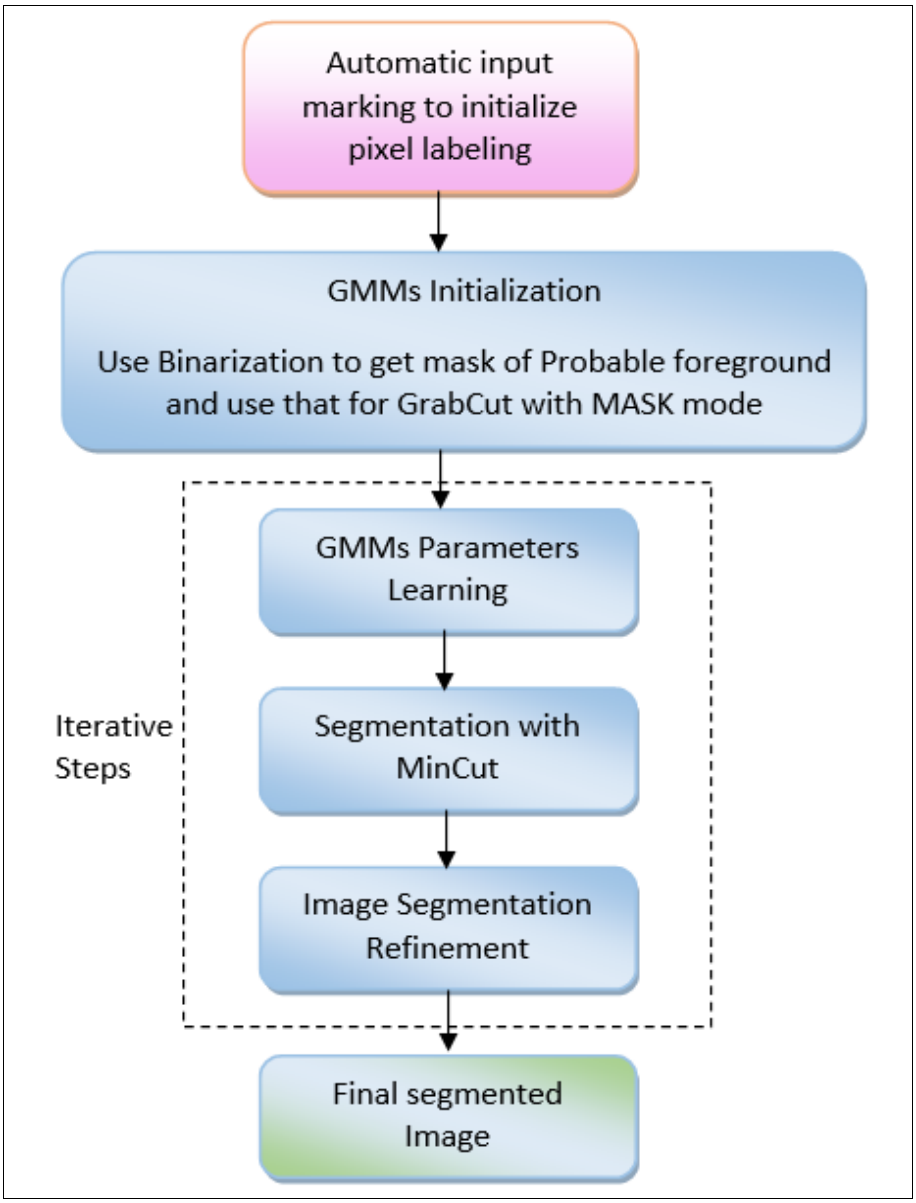


Fig 16: Improved GrabCut Segmentation starts with rectangle initialization and then uses binarized mask to finally arrived at segmented image

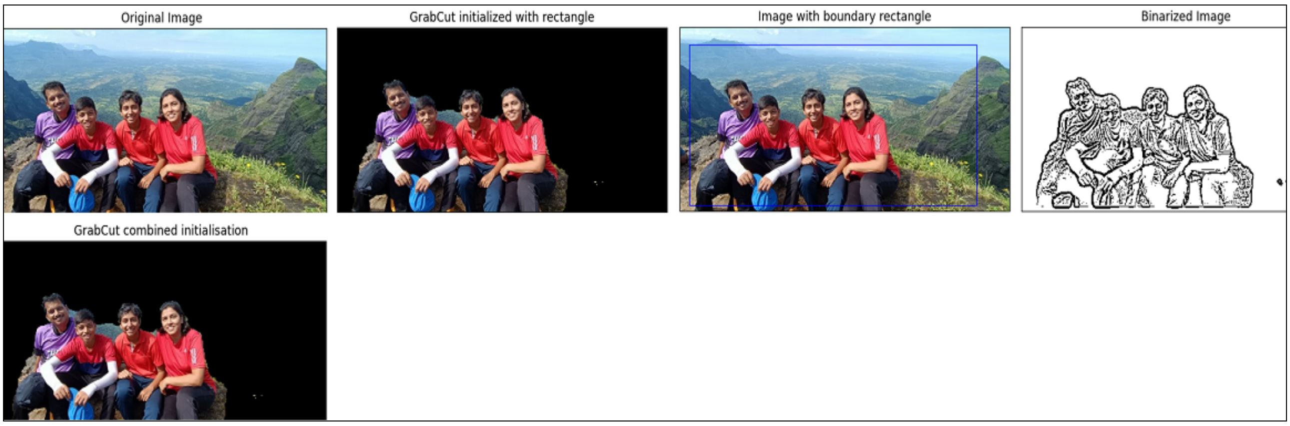
Proposed GrabCut Flow

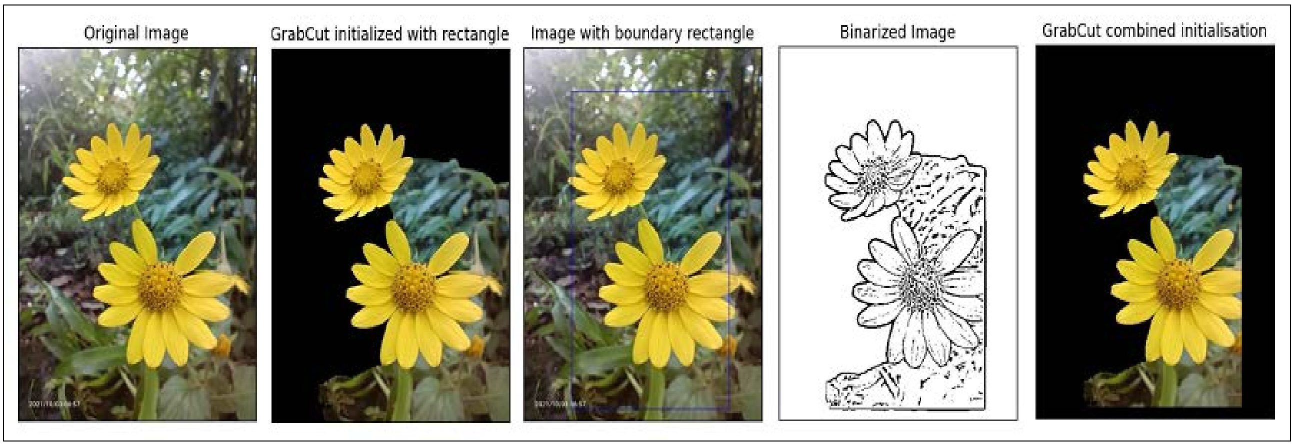


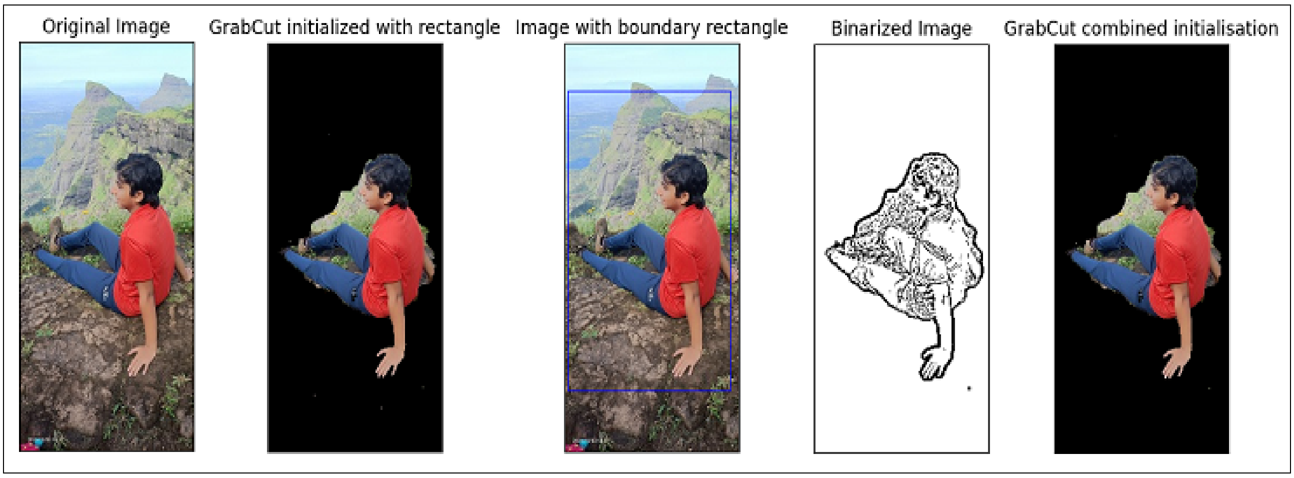
As shown in flow initialization is automatic and with grey_image binarization it sets mask for probable foreground segmented region and same gets passed to GracCut for final MinCut in mask mode.

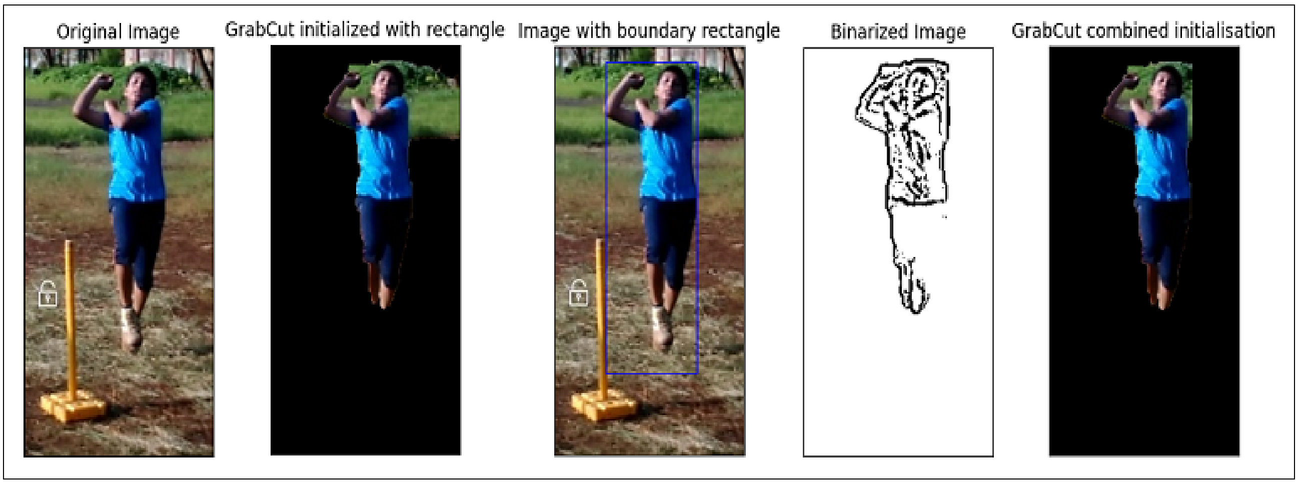
Experimental Results, Analysis and Discussion

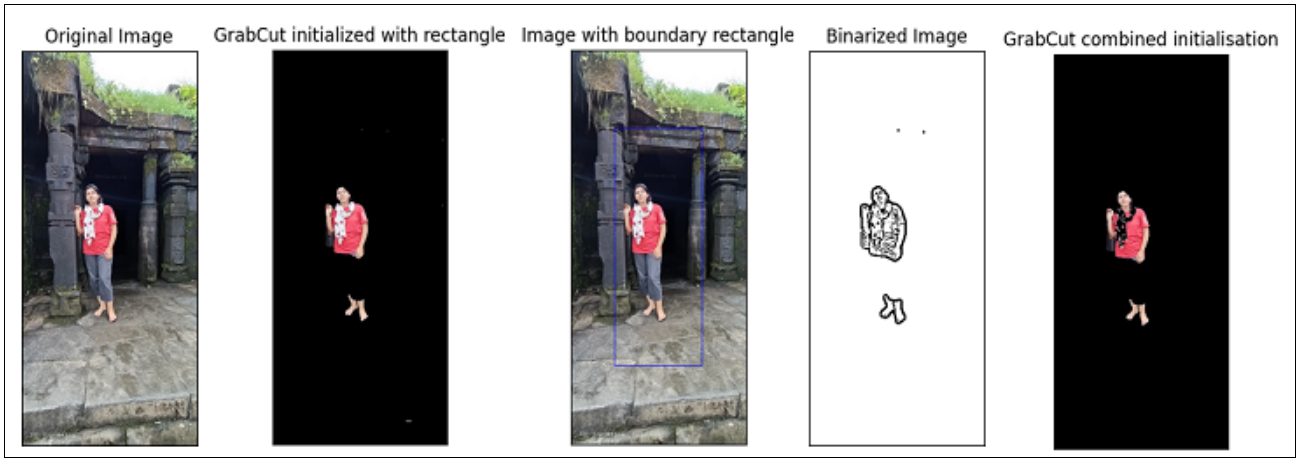
For experimentation we have used OpenCV library which gives out of the box implementations for image operations. OpenCV library is available on Java as well as on Python. Algorithms like GrabCut (with Rectangle and Mask mode), Grey image conversion and Binarization (adaptive Threshold, find Contours) are readily available. We tried different sets of images with initially only with GrabCut in rectangle mode and then with binarization in mask mode. Results are shown below.


Original Image
GrabCut initialized with rectangle
Image with boundary rectangle
Binarized Image
GrabCut combined initialisation


Original Image
GrabCut initialized with rectangle
Image with boundary rectangle
Binarized Image
GrabCut combined initialisation


Original Image
GrabCut initialized with rectangle
Image with boundary rectangle
Binarized Image
GrabCut combined initialisation


Original Image
GrabCut initialized with rectangle
Image with boundary rectangle
Binarized Image
GrabCut combined initialisation



In each of these rows 2nd image is a image with only Grabcut (with rectangle mode) and last i.e. 5th image is the image which is initialized with the binarized and then passed through Grabcut with the mode MASK. It is observed that last image is more accurate and without any manual interaction or manual initialization. There is a scope for improvement here.

Benefits and Limitations

Benefits

- More accurate Segmentation as mask is getting identified in initializing phase itself.
- Less Manual Interaction and less iterations.
- Out of the Box Implementation in OpenCV.
- Very few data about input images is required to use this method.

Limitations

- High Computation time for high resolution Images.
- No previous knowledge about the input.
- Segmentation is only between two zones.

To get away with listed limitations, we will have to adopt deep learning approaches like Mask R-CNN.

Future Enhancements

Along with the binarized mask we can apply deep learning method using Mask R-CNN algorithm. Mask R-CNN is an extension of Faster R-CNN which is used mainly for object detection. Mask R-CNN gives class label and bounding box coordinates for each detected object along with the object Mask. There are many pre-trained models (like MS COCO) available in Mask R-CNN which are already trained on datasets to obtain weights of known classes/objects. When any image is passed to the Mask R-CNN it predicts objects from the images by comparing those with pre-trained dataset. Once we have exact mask we can use GrabCut to have accurate segmentation without much manual interaction.

Conclusion

We have evaluated GrabCut algorithm using rectangle and mask modes to achieve the image segmentation. We have learned the working of GrabCut algorithm. It allows to separate out Pixels of input images into the categories like Background, Foreground, Probable Background and Probable Foreground as per the pixel intensities. It is found that existing GrabCut has multiple manual interactions and iterations. To improve on this, Binarize image as per the segmentation mask and then apply GrabCut to it to achieve some kind of automation. This was implemented using OpenCV and it is observed that this method improves accuracy and also it avoids manual interactions and multiple iterations. Though it is not 100% accurate but it gives base to improve on accuracy as mask is finalized and it is possible to use deep learning approaches to identify objects and so the further enhancement in this method.

We have also gone through the typical solutions for each stage and listed the traditional algorithms with particular influences. Generally speaking, there has been a shift in the development of image segmentation from coarse to fine-grained, from manual feature extraction to adaptive learning, and from segmentation based on a single image to segmentation based on common features of huge data. Deep neural network research has demonstrated advantages in scene interpretation and object recognition since the FCN was first introduced. Image segmentation has moved from the CNN stage to the transformer stage thanks to the swin transformer's breakthrough in the field of computer vision in 2021. The transformer may lead to new developments in the

study of computer vision. Deep learning also has drawbacks, such as the inexplicability of deep learning, which restricts the robustness, dependability, and performance enhancement of its downstream tasks. On the other hand, although graph-based algorithms like GrabCut are incredibly accurate, final segmentation results still require manual intervention. Graph-based segmentation using the Grabcut approach and Mask R-CNN deep learning segmentation could be combined in future study to achieve more accuracy and to improve on performance as it will be using pre-trained models to detect the objects.

References

1. Akers H. what-are-the-different-types-of-digital-image-processing-techniques.html, 21 January 2022. Online Available: https://www.easytechjunkie.com/what-are-the-different-types-of-digital-image-processing-techniques.htm.
2. Gonzalez RC. Digital-image-processing-basics, 30 November 2021. Online Available: https://www.geeksforgeeks.org/digital-image-processing-basics/.
3. Tyagi M. image-segmentation-part-1-9f3db1ac1c50, 18 July 2021. Online Available: https://towardsdatascience.com/image-segmentation-part-1-9f3db1ac1c50.
4. Tyagi M. image-segmentation-part-2-8959b609d268, 24 July 2021. Online Available: https://towardsdatascience.com/image-segmentation-part-2-8959b609d268.
5. Prasad S. Image Segmentation Techniques, 31 May 2020. Online Available: https://www.analytixlabs.co.in/blog/what-is-image-segmentation.
6. He K, Wang D, Wang B, Feng B, Li C. Foreground Extraction Combining GraphCut and Histogram Shape Analysis, IEEE Access Digital Object Identifier. 2019;7:176248-17625. 10.1109/ACCESS.2019.2957504,
7. Singh R, Kumar A. Python-foreground-extraction-in-an-image-using-grabcut-algorithm, 5 January 2022. Online Available: https://www.geeksforgeeks.org/python-foreground-extraction-in-an-image-using-grabcut-algorithm/.
8. Li Y, Zhang J, Gao P, Jiang L, Chen M. Grab Cut Image Segmentation Based on Image Region, 3rd IEEE International Conference on Image, Vision and Computing. 2018;3:311-315. 978-1-5386-4991-6/18.
9. Patel C, Patel DA, Shah DD. Threshold Based Image Binarization technique for number plate segmentation, IJARCSSE ISSN: 2277. 2013;3(7):108-114.
10. Cheng Y, Li B. Image Segmentation Technology and Its Application in Digital Image Processing, IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). 2021;21(978-1-7281-9018-1):1174-1177.
11. Gao H, Siu WC, Hou CH. Improved techniques for automatic image segmentation, IEEE Transactions on Circuits and Systems for Video Technology. 2001;11(12):1273-1280. DOI: 10.1109/76.974681,
12. Ballerini L, Hogberg A, Borgefors GCBAA, Lindgard, Lundstrom K, Rakotonirainy O, Soussi B. A segmentation technique to determine fat content in NMR images of beef meat, IEEE Transactions on Nuclear Science. 2002;49(1):195-199. DOI: 10.1109/TNS.2002.998751,
13. AP, SMR. An interacting multiple model probabilistic data association filter for cavity boundary extraction from ultrasound images, IEEE Transactions on Medical Imaging. 2004;23(6):772-784. DOI: 10.1109/TMI.2004.826954,
14. SWC, BAC, EBL. Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images, IEEE Transactions on Medical Imaging. 2005;24(12):1593-1610. DOI: 10.1109/TMI.2005.859207,
15. CTA, Goncalves H, Veloso-Gomes F, AGJ. Modelling of the Douro River Plume Size, Obtained Through Image Segmentation of MERIS Data, IEEE Geoscience and Remote Sensing Letters. 2009;6(1):87-91. DOI: 10.1109/LGRS.2008.2008446,
16. Adel M, Moussaoui A, Rasigni M, Bourennane S, Hamami L. Statistical-Based Tracking Technique for Linear Structures Detection: Application to Vessel Segmentation in Medical Images, IEEE Signal Processing Letters. 2010;17(6):555-558. DOI: 10.1109/LSP.2010.2046697,
17. Chitsaz Mahsa, Woo S, Chaw. Software Agent with Reinforcement Learning Approach for Medical Image Segmentation", International Journal of Computer Science and Technology; c2011. p. 247-255. DOI: http://dx.doi.org/10.1007/s11390-011-9431-8,
18. Kriti J Virmani, Agarwal R. A Review of Segmentation Algorithms Applied to B-Mode Breast Ultrasound Images: A Characterization Approach, Springer Archives of Computational Methods in Engineering. 28:2567-2606. 10.1007/s11831-020-09469-3,
19. GD, PB, SA. Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistical Society, Series B. 1989;51(2):271-279.
20. BY, JMP. Interactive graph cuts for optimal boundary and region segmentation of objects in ND images, In: Proceedings of the 8th IEEE International Conference on Computer Vision, Vancouver, Canada: IEEE; c2001. p. 105-112,
21. RC, KV, BA. Grab Cut: interactive foreground extraction using iterated graph cuts, In: Proceedings of the ACM SIGGRAPH Conference. Los Angeles, USA: ACM; c2004. p. 309-314,
22. Han X, Qi Y, Sun H, Zi Y, Li Y. Research on colour image segmentation algorithm based on improved Grab-cut. 2018;8(18):923-927. 978-1-5386-8069-8/18 IEEE,
23. ZMDLJ, Zhai L. Foreground segmentation of Grab Cut based on improved super-pixels and features, [J]. Microcomputer Application. 2015;31(11):48-53+3.
24. PBTJH, Zhao Y. A segmentation method for SAR image based on Grab Cut and 2D maximum-entropy, [J]. Engineering of Surveying and Mapping. 2018;27(04):60-64+70.
25. AC, LLM, HYQ. An Improved Adaptive GrabCut Algorithm Based on SLIC, [J]. Process Automation Instrumentation. 2017;38(10):17-20.
26. JZJP, YaWei Yu. Foreground Target Extraction in Bounding Box Based on Sub-block Region Growing and Grab Cut, 978-1-5386-4673-1/18 © 2018 IEEE; c2018. p. 344-349.
27. Felzenszwalb PF, Huttenlocher DP. Efficient graph-

based image segmentation, International Journal of Computer Vision. 2004;59(2):167-181.
28. Wang H, Wang B, Zhou Z, Song L, Li S, Wu S. Region fusion and grab-cut based salient object segmentation, in Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics; c2014. p. 131-135.
29. NBS Vu. Image Segmentation with Semantic Priors: A Graph Cut Approach, University of California, Santa Barbara, Ph.D. Dissertation, September; c2008.
30. Aganj I, Fischl B. Multi-Atlas Image Soft Segmentation via Computation of the Expected Label Value, IEEE Transactions on Medical Imaging. 2021;40(6):1702-1710.
31. Brzoza A, Muszynski G. An approach to image segmentation based on shortest paths in graphs, International Conference on Systems, Signals and Image Processing (IWSSIP), 978-1-5090-6344-4/17 ©2017 IEEE. 2017;4:17.
32. Lonarkar V, ARB. Content-Based Image Retrieval by Segmentation and Clustering, Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017), IEEE Xplore Compliant - Part No: CFP17L34-ART, ISBN: 978-1-5386-4031-9/17. 2017;9(17):771-776.
33. Ying Yu, Chunping Wang, Qiang Fu, Renke Kou, Fuyu Huang, Boxiong Yang, *et al*., Techniques and Challenges of Image Segmentation: A Review, Electronics. 2023;12:1199. https://doi.org/10.3390/electronics12051199, https://www.mdpi.com/journal/electronics
34. Julie Prost https://www.sicara.fr/blog-technique/grabcut-for-automatic-image-segmentation-opencv-tutorial
35. Carsten Rother, Vladimir Kolmogorov, Andrew Blake. UK, GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts', Microsoft Research Cambridge; c2004.