International Journal of Advances in Electrical Engineering

E-ISSN: 2708-4582
P-ISSN: 2708-4574
IJAEE 2023; 4(2): 01-11
© 2023 IJAEE
www.electricaltechjournal.com
Received: 03-04-2023
Accepted: 03-05-2023

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

Study of foreground extraction image segmentation techniques

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

DOI: https://doi.org/10.22271/27084574.2023.v4.i2a.40

Abstract
The field of computer vision has advanced significantly in the current era of technology. Digital image processing has undergone significant advancements. Image Generation, Image Enhancement, and Image Restoration are the three categories into which digital image processing techniques are commonly divided. Segmentation Procedure is one of several stages in image processing that divides an image into its individual components or objects [1]. Image segmentation is a technique that divides a digital image into a number of smaller groups or regions known as image segments, thereby lowering the complexity of the image and facilitating easier image processing and analysis [3]. There are numerous image segmentation techniques that are based on the two fundamental concepts of similarity/region and discontinuity/boundary. Techniques for image segmentation are frequently utilized in biomedicine, document processing, object recognition, automated industrial production, computed tomography (CT) images, and many other fields. Here, the goal is to compare and learn about fundamental image pre-processing and accessible segmentation approaches. ML-based algorithms are among the most effective foreground extraction methods currently available for masking images. However, it is discovered that ML approaches are unreliable. According to edges and appearance models, graph-based methods like Graphcut/Grabcut can successfully recover the foreground [6]. Both the foreground and the background are modelled using the Gaussian Mixture Model (GMM). However, with these graph-based techniques, manual creativity is necessary. The main issues with current techniques for foreground extraction in image segmentation are accuracy and performance.

Keywords: Digital image processing, image segmentation, graphcut, grabcut, gaussian mixture model

Introduction
Background study
Digital image processing is the act of applying various computer algorithms to digital images in order to enhance them and extract usable information. Three categories/steps are generally used to group digital image processing processes. Image creation, enhancement, and restoration fall under these areas. A scanned image is projected and recognized with the use of generation processes, while an image is enhanced by enhancing its contrast, brightness, and hue. Techniques for restoration help remove and fix defects that do not accurately reflect the original image [1].

One method for processing digital images that involves organizing an image into a layout is called "image generation". Digital image processing techniques that change a digitized image fall under the wide area of enhancement. Some scanned photos can reveal rips or creases. Replacement techniques can be used to get rid of noise, which can appear as random dots or streaks. Digital filtering may also be used in restoration. Reversing the process that blurred the image is how image restoration is done. This is done by imaging a point source and using the point source image, also known as the Point Spread Function (PSF), to recover the image information that was lost during the blurring process. Image processing techniques are used with the intention of recovering lost resolution and reducing noise.

An image is described as a two-dimensional function F(x, y), where x and y are spatial coordinates. The intensity of an image is defined as the amplitude of F at any given pair of coordinates (x, y). We refer to F as a digital image when its x, y, and amplitude values are all finite [2]. Matrix is a way of displaying digital images as rows and columns.

Phases of Image Processing

- **Acquisition:** It could be as simple as being given an image that is in digital form. The main work involves:
  a) Scaling b) Color conversion(RGB to Gray or vice-versa)
- **Image Enhancement:** It is amongst the simplest and most appealing in areas of Image Processing it is also used to extract some hidden details from an image and is subjective.
- **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).
- **Colour Image Processing:** It deals with pseudo-colour and full-colour image processing color models are applicable to digital image processing.
- **Wavelets And Multi-Resolution Processing:** It is foundation of representing images in various degrees.
- **Image Compression:** It involves in developing some functions to perform this operation. It mainly deals with image size or resolution.
- **Morphological Processing:** It deals with tools for extracting image components that are useful in the representation & description of shapes.
- **Segmentation Procedure:** It includes partitioning an image into its constituent parts or objects. Autonomous segmentation is the most difficult task in Image Processing.
- **Representation & Description:** It follows output of segmentation stage, choosing a representation is only the part of solution for transforming raw data into processed data.
- **Object Detection and Recognition:** It is a process that assigns a label to an object based on its descriptor.

Overlapping Fields with Image Processing

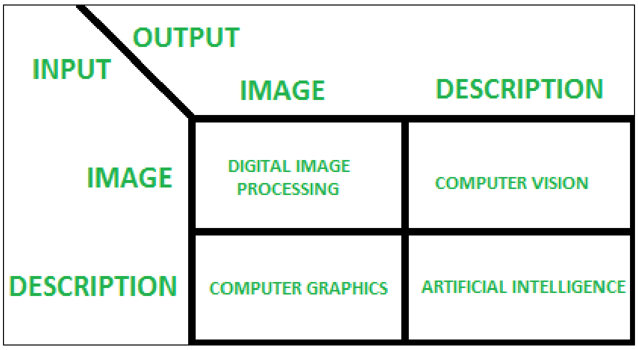

Fig 1: Image Processing Fields

If we focus on block 1 i.e. input is an image and output is also an image and block 2 where input is an image and we get some kind of description as an output then these are the areas where digital image processing and computer vision take place and most important step in these areas is Image Segmentation.

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].

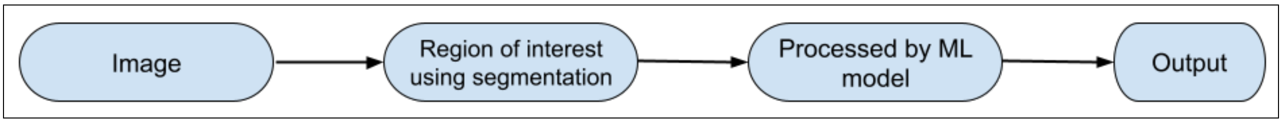

Fig 2: Image Segmentation

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.

The Types of Segmentation Processing Techniques

Based on the above two approaches, there are various forms of techniques that are applied in the design of the Image Segmentation Algorithms. These techniques are employed based on the type of image that needs to be processed and analyzed and they can be classified into three broader categories as below.

Structural Segmentation Techniques

These sets of algorithms require us to firstly, know the structural information about the image under the scanner. This can include the pixels, pixel density, distributions, histograms, colour distribution etc. Second, we need to have the structural information about the region that we are about to fetch from the image – this section deals with identifying our target area, which is highly specific to the business problem that we are trying to solve. Similarity-based approach will be followed in these sets of algorithms.

Stochastic Segmentation Techniques

In this group of algorithms, the primary information that is required for them is to know the discrete pixel values of the full image, rather than pointing out the structure of the required portion of the image. This proves to be advantageous in the case of a larger group of images, where a high degree of uncertainty exists in terms of the required object within an object. ANN and Machine Learning based algorithms that use k-means etc. make use of this approach.

Hybrid Techniques

As the name suggests, these algorithms for image segmentation make use of a combination of structural method and stochastic methods i.e., use both the structural information of a region as well as the discrete pixel information of the image.

Image Segmentation Techniques

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

segmentation

- Threshold Based Segmentation.
- Edge-Based Segmentation.
- Region-Based Segmentation.
- Clustering-Based Segmentation.
- Watershed-Based Segmentation.
- Artificial Neural Network Based Segmentation.

Threshold Based Segmentation

In this threshold process, the intensity histogram of all the pixels in the image is considered. Then threshold is set to divide the image into sections.

Various threshold techniques are

- Global threshold.
- Manual threshold.
- Adaptive threshold.
- Optimal Threshold.
- 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 [4].

- Region growing method.
- 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 [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 analyze 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 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 algorithm, 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.

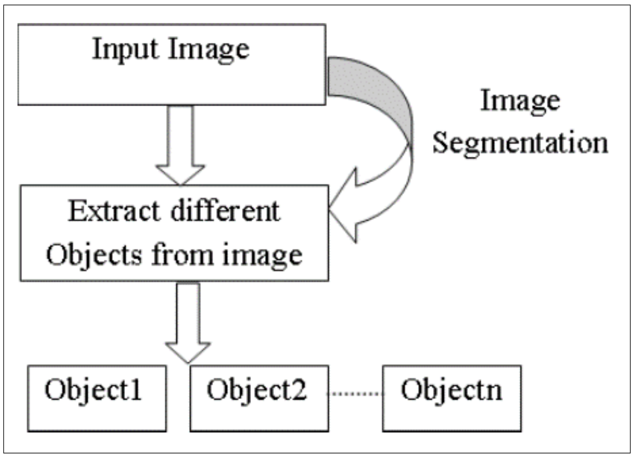

Fig 3: Block Diagram of Image Segmentation Process [10]

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), and 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 a 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 objects 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 a 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, Jitendra Virmani 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 that have been reviewed over the past two decades using 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 cutting 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 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.

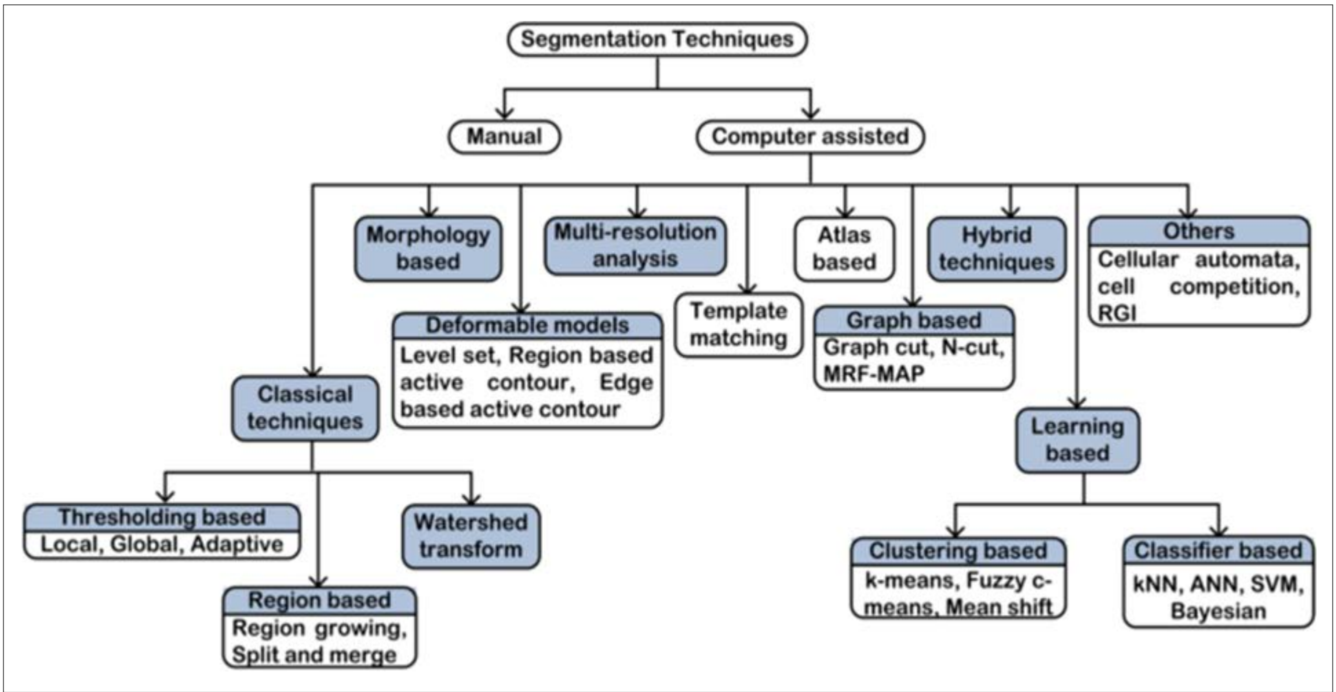

Fig 4: Segmentation Techniques and Algorithms [18]

Kun He *et al.* [6] in his work proposes novel appearance model which is formulated as maximum likelihood rather than the weight sum of Guassians. 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 efficiency of 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 have 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] propose a foreground extraction algorithm for target detection based on deep learning in combination with the sub-block region growing algorithm and the Grab Cut algorithm. Finally, the algorithm proposed takes advantage of the Grab Cut algorithm and the region-growing algorithm, avoiding the shortcomings of the two algorithms. Another alternative to 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.

Iman Aganj *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 as 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 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] uses 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 are 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.

Methodology

Working of already existing image segmentation algorithms

The segmentation techniques now in use, particularly the cutting-edge techniques, need to be comprehensively summarized [33]. We describe the approaches' operating principles, list several noteworthy picture segmentation algorithms, and methodically introduce the key concepts of semantic segmentation, as seen in Figure.

Classic Segmentation Methods

For grayscale images, the classic segmentation methods were suggested, which primarily take into account gray-level discontinuity in distinct regions and gray-level similarity within the same region. In general, the gray-level similarity is the foundation for region division, while gray-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.

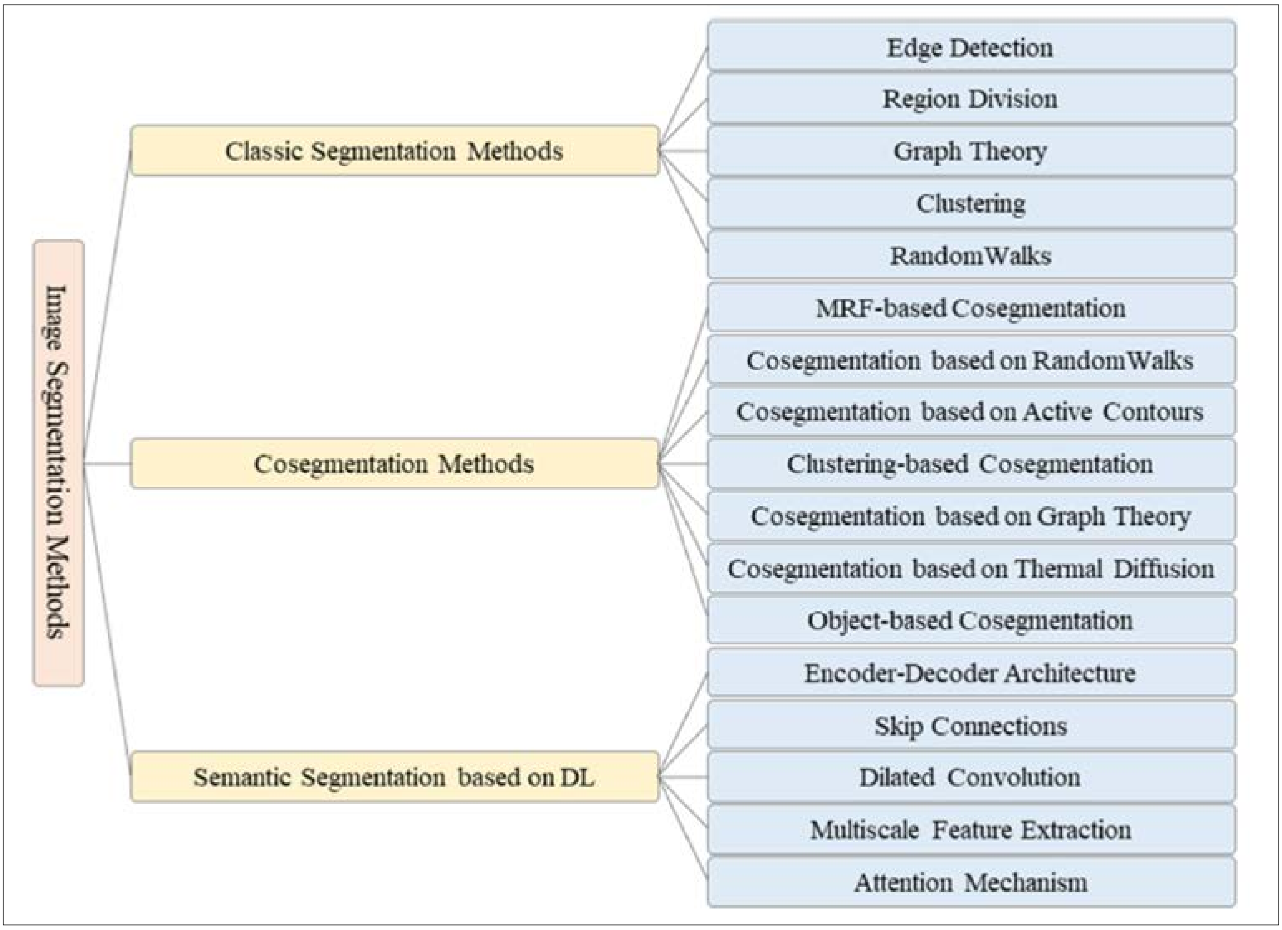

Fig 5: Categories of Image Segmentation

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 division are used in the region division approach. A typical parallel region division algorithm is thresholding. The gray 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 gray 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 text on 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 the 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.

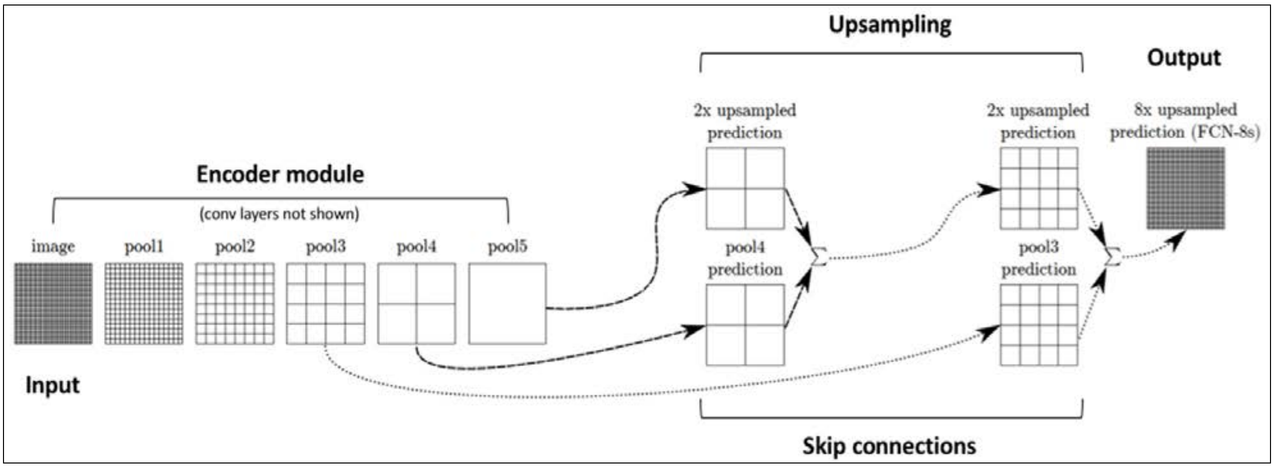

Fig 6: Fully Convolutional Networks architecture

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 the X-transformer, were then used in 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.

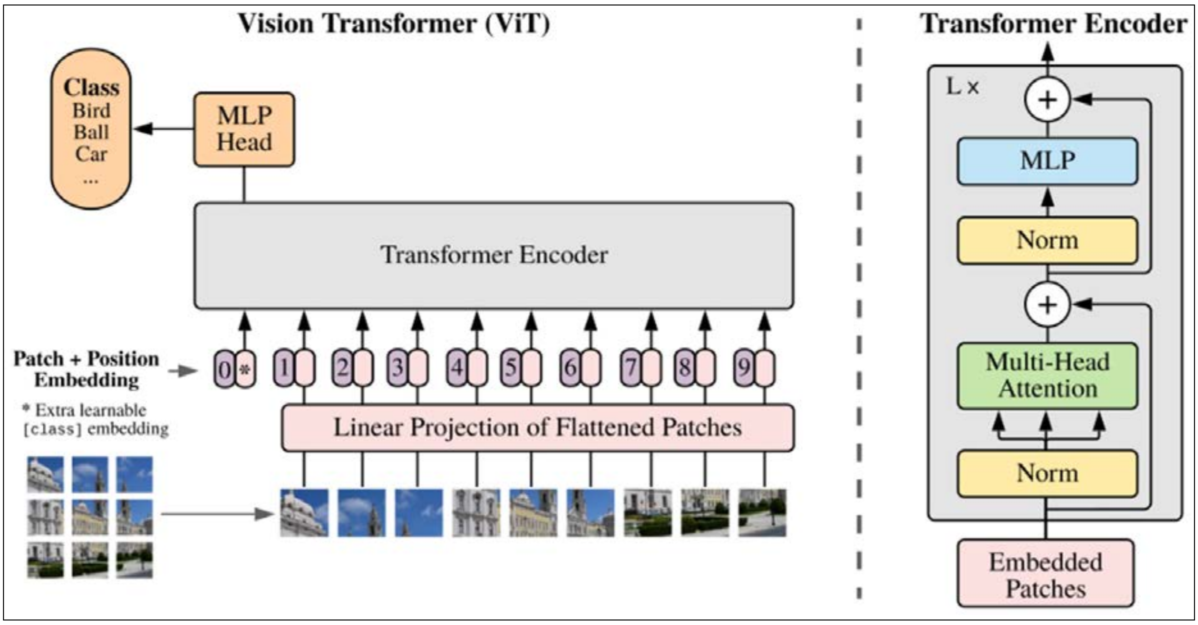

Fig 7: ViT Model [33]

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.

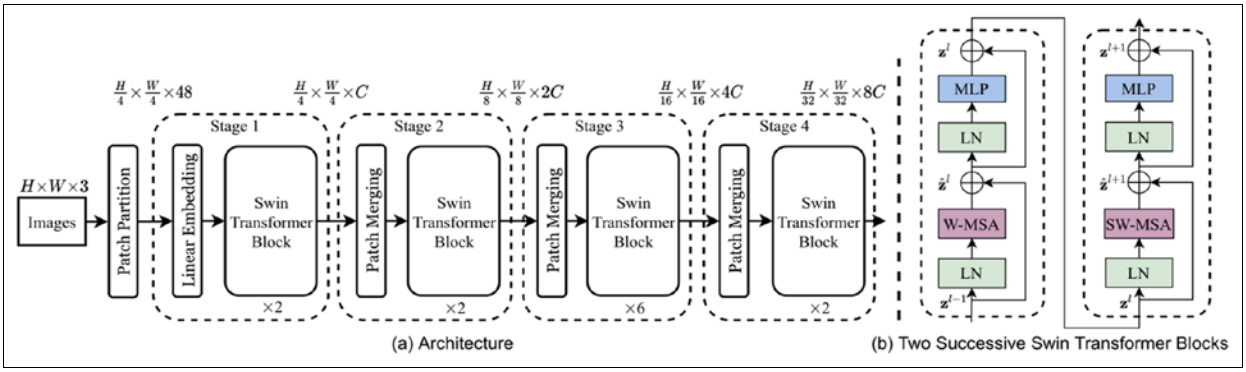

Fig 8: SWIN Transformer Architecture [33]

Challenges of Image Segmentation

- The key targets for image segmentation researchers are still semantic segmentation, instance segmentation, and panoramic segmentation. Building an efficient network to simultaneously identify both huge inter-category differences and minor intra-category differences is a demanding task in panorama segmentation since countable or uncountable examples are difficult to recognize in a single workflow;
- As image-collecting technology has gained popularity (e.g., LiDAR cameras), researchers' focus has shifted to areas like RGBdepth, 3D-point clouds, voxels, and mesh segmentation, which have a wide range of applications in areas like face recognition, autonomous vehicles, VR, AR, architectural modelling, etc. The representation and processing of 3D data, which are unstructured, redundant, disordered, and unevenly distributed, remain a significant challenge; despite some advancements in the research of 3D image segmentation, such as region growth, random walks, and clustering in classic algorithms, and SVM, random forest, and AdaBoost in machine learning algorithms;
- Due to a lack of datasets or fine-grained annotations, it can be challenging to train the network using supervised learning algorithms in some disciplines. The network can be trained on the benchmark dataset first, the lower-level network parameters can then be fixed, and the fully connected layer or some high-level parameters can be trained on the small-sample dataset. In these situations, semi-supervised and unsupervised semantic segmentation can be chosen. Transfer learning does not necessitate a large number of labelled samples. Another option is reinforcement learning, but it is not frequently researched in the area of picture segmentation;
- Deep learning networks' high computational demands during the training phase serve as an example of the deep neural network's high computational complexity.

In several domains, such as video processing, where the human visual mechanism requires at least 25 frames per second, real-time (or almost real-time) segmentation is necessary. However, the majority of contemporary networks operate at far lower frame rates. There has been some progress in segmentation speed thanks to several lightweight networks, but there is still much opportunity for improvement in the trade-off between model correctness and real-time performance.

- Other Graph-based segmentation methods like Grabcut are accurate but repeated manual intervention is required to segment the image which also degrades the performance of segmentation.

Conclusion
We have thoroughly examined the traditional segmentation algorithms as well as the current hot deep learning methods. 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. As image-collecting technology advances, the varieties of images are expanding, which increases the difficulty of segmenting images with various dimensions, scales, resolutions, and imaging modes. 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 Grab are incredibly accurate, final segmentation results still require user intervention. Graph-based segmentation using the Grabcut approach and such manual methods as R-CNN deep learning segmentation could be combined in future studies.

Conflict of interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Akers H. What-are-the-different-types-of-digital-image-processing-techniques.htm, 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 Graph Cut and Histogram Shape Analysis, IEEE Access Digital Object Identifier. 2019;7(10):176248-17625 DOI: 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(978-1-5386-4991-6/18):311-315.
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, DOI: 10.1109/76.974681. 2001;11(12):1273-1280.
12. Ballerini L, Hogberg A, Borgefors GCBA, Lindgard A, 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. CTAH, Goncalves F, Veloso-Gomes, 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. DOI: 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 N-D 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 Grab Cut 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; c2018. p. 344-349. 978-1-5386-4673-1/18 ©2018 IEEE
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; c2008 Sep.
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 IEE. 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 Number: CFP17L34-ART. 2017;9(17):771-776. ISBN: 978-1-5386-4031-9/17.
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