INTERNATIONAL JOURNAL ON ORANGE TECHNOLOGY

https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071

Volume: 5 Issue: 5 | May 2023

# A Study on Image Categorization Techniques

**Renu**

P.G. Student, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana, India

**Princy, Kirti Bhatia**

Assistant Professor, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana, India

**Rohini Sharma**

Assistant Professor, Department of CS, GPGCW, Rohtak, Haryana, India

---------------------------------------------------------------***---------------------------------------------------------------

**Annotation:** Image segmentation is the act of splitting a picture into meaningful and non-overlapping parts. It is a crucial step in comprehending natural scenes and has become a hotbed of research in the fields of image processing and computer vision. Even after decades of work and several successes, feature extraction and model design remain difficult. In this article, we carefully review the development in image segmentation techniques. Three crucial stages of image segmentation—classical segmentation, collaborative segmentation, and semantic segmentation based on deep learning—are primarily examined in accordance with segmentation principles and image data characteristics. We compare, contrast, and briefly discuss the benefits and drawbacks of segmentation models as well as their applicability. We also elaborate on the primary algorithms and critical strategies in each stage.

**Keywords:** Image Segmentation, Clustering, Neural network, Edge Detection.

## I. Introduction

The foundation for pattern recognition and visual comprehension is image segmentation, one of the most well-known study areas in computer vision. The development of picture segmentation techniques is directly related to numerous disciplines and fields, including augmented reality, image search engines, driverless cars, and intelligent medical technologies [1]. Area of interest (AOIs) are extracted from images using image segmentation, which separates images into regions with various attributes. According to human visual perception, these zones are significant and do not overlap. Image segmentation presents two challenges: (1) how to define "meaningful regions" (since there is no universally accepted definition of what constitutes an object due to variability in human comprehension and visual perception), and (2) how to precisely represent each item in an image. Pixels, the building blocks of digital images, can be clustered together to form bigger collections depending on their colour, texture, and other characteristics. "Pixel sets" or "superpixels" are the terms used to describe these. The primary goals of traditional segmentation techniques are to identify and extract the information from a single image, which frequently calls for specialised knowledge and human participation. It is challenging to extract high-level semantic information from photographs, though. The identification of common items from a collection of photos is a step in the co-segmentation process that calls for some prior knowledge. These techniques are categorised as semi-supervised or weakly supervised approaches because the image annotation is not necessary. Deep neural network-based picture segmentation techniques have slowly gained popularity as a result of the enlargement of large-scale fine-grained annotated image datasets. Fig. 1. Shows various types of image segmentation tecchniques.

INTERNATIONAL JOURNAL ON ORANGE TECHNOLOGY
https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071
Volume: 5 Issue: 5 | May 2023

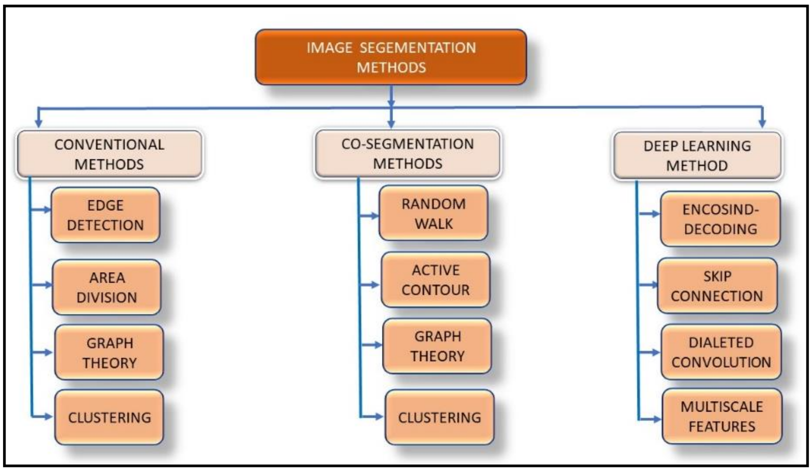


Fig1. Image Segmentation Types

## II. RELATED WORK

Even though picture segmentation research has made significant strides, there are still numerous obstacles to overcome, such as feature representation, model design, and optimisation. Due to inadequate or sparse annotations, class imbalance, overfitting, lengthy training times, and gradient vanishing, semantic segmentation in particular continues to present many difficulties [2-4]. Reviews have not yet sorted and summarised image segmentation algorithms in terms of how the technology in the area of image segmentation has progressed from its inception to the present day. The authors of [5] introduced semantic segmentation methods and commonly used datasets, and [6] analysed metrics for assessment and techniques of semantic segmentation. Consequently, a thorough summary of the current segmentation techniques, particularly the cutting-edge techniques, is required. We examine and reclassify current image segmentation strategies from the standpoint of algorithm growth, explain how they function, list some significant image segmentation techniques, and proceed to establish the fundamental methods of semantic segmentation utilising deep neural networks.

## III. CONVENTIONAL SEGMENTATION METHODS

For grayscale images, the traditional segmentation methods were suggested, which primarily take into account gray-level discontinuities in distinct regions and gray-level resemblance within the same region. Overall, gray-level similarity is the foundation for region division, while gray-level discontinuity is the foundation for edge detection. Utilising the similarity between pixels, colour image segmentation divides the image into various sections or superpixels, which are then combined.

**Edge Detection:** In an image, the edges of several sections are typically where the grey level abruptly changes. Finding the spots on these boundaries is the goal of edge detection. One of the early segmentation techniques is edge detection, commonly known as the parallel border technique. To find the evident changes at the border, use the grey level's derivative or differential. In actuality, the variance estimate for the differential is used to obtain the derivative of the digital image**.**

The serial boundary technique, which concatenates edge points to create a closed border, is an edge detection method. Graph-searching algorithms and dynamic programming algorithms are the two primary types of serial boundary approaches. In graph-searching algorithms, the points on the edges are represented by a graph structure, and the closed borders are found by searching the graph for the path with the lowest cost, which is always computationally expensive. Heuristic rules are used by the dynamic programming algorithm to speed up search

INTERNATIONAL JOURNAL ON ORANGE TECHNOLOGY

https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071

Volume: 5 Issue: 5 | May 2023

computation. The active contours technique finds the closed curve with the least amount of energy by minimising the energy function, and it then matches the closed curve (i.e., the first set of contours determined by the gradient) with the local features of the image to derive the actual contours of the objects. The initialization must be near to the target contour because the procedure is sensitive to the location of the initial contour. Its non-convexity also makes it easy to reach the local minimum, making it challenging to converge to the concave border. A methodology that takes into account the local segmentation energy to evolve outlines was put forth in [7].

**Area Division**: Both serial and parallel region division 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 component of the grey histogram can be used to calculate the ideal grayscale threshold in order to maximise categorization.

Region expanding and region merging are two example processes in the serial region technique, which includes breaking the region segmentation task into many steps to be completed consecutively. A specified growth algorithm is used to combine pixels with the same or comparable properties in the seed neighbourhoods in the areas where the seed is placed until no more pixels can be combined. This process is known as region expanding and entails using numerous seeds (single pixels or regions) as initiation points. The idea behind region merging is the same as that behind region growth, with the exception that region merging determines how similar two regions are by determining whether the difference between the average grey value of the pixels in the region obtained in the previous step and the grey value of its adjacent pixels is less than the specified threshold K. Hard noise loss and object occlusion are problems that can be resolved with region merging. The idea of topography is the foundation of watershed. Dams must be constructed to stop rising water from low areas from reaching the mountain tops. The entire scene is divided into different sections by the dams constructed on the mountain tops. The watershed method has a good processing efficiency and may obtain the closed contour. Building a Gaussian mixture model (GMM) can address the issue of incorrect segmentation that occurs with more complicated images. The enhanced watershed is particularly useful for segmenting medical images including overlapping cells (such as blood cell segmentation), has strong generalisation performance, and is frequently employed in the segmentation of MRI images and digital elevation maps.

**Graph Theory**: The graph theory-based method for segmenting images translates a picture to a graph, where pixels or regions serve as the graph's vertices and weights for the edges signify how comparable the vertices are. Based on graph theory, image segmentation is defined as the separation of vertices in the graph, weighted graph analysis using the principle and method of graph theory, and optimal segmentation using global graph optimisation (e.g., the min-cut).

Rather than utilising fixed merging algorithms in clustering, graph-based area merging uses many criteria to find the best global grouping. After the image was formatted as a graph, authors in [8] utilised the minimal spanning tree (MST) to merge pixels. Probabilistic graphical models (PGMs) are introduced into the region division in MRF-based image segmentation to represent the unpredictability of the lower-level characteristics in the images. It converts the image into an undigraph, where each edge denotes the connection between two vertices and each vertex denotes the feature at the corresponding place in the image. The Markov property of the graph states that each point's feature is exclusively connected to its neighbouring features.

**Clustering**: A unique thresholding segmentation approach called K-means clustering has been devised and is based on the Lloyd algorithm. The following steps make up the algorithm: (i) initialise K points as clustering centres; (ii) determine the distance between each point i in the image and K cluster centres, and choose the distance with the smallest value as the classification ki; (iii) calculate the average of the points in each category (the centroid), and move the cluster centre to the centroid; and (iv) repeat steps (ii) and (iii) until algorithm

INTERNATIONAL JOURNAL ON ORANGE TECHNOLOGY
https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071
Volume: 5 Issue: 5 | May 2023

convergence. K-means, to put it simply, is an iterative procedure for calculating cluster centres. The K-means can only converge to the local optimal solution rather than the global optimum solution, and although it has speedy convergence and noise tolerance, it is not suitable for processing nonadjacent regions.

The clustering approach known as mean-shift [9] models the image feature space to the probability density function. It is based on density estimation. In order to produce more uniform region segmentation, authors in [10] devised a fuzzy C-means method that integrated geographical information into the membership function for clustering.

## IV. CO-SEGMENTATION METHODS

It is challenging to retrieve the high-level semantic information of a picture when using the traditional segmentation approaches, which typically concentrate on the feature extraction of a single image. The first time the idea of collaborative segmentation was put up was by authors in [11]. Co-segmentation, also known as collaborative segmentation, is the process of automatically extracting the common foreground areas from a set of photos in order to gain previous knowledge.

**Random Walks:** Authors in [12] supplied a professional CUDA library to calculate the linear operation of the image sparse features, further utilised the quasiconvexity to optimise the segmentation algorithm, and expanded the random walks model to address the co-segmentation problem. For 3D voxel picture segmentation, Authors in [13] suggested an optimised random walks approach employing a supervoxel rather than a single voxel, which significantly reduced processing time and memory requirements. In order to combine subRW with other random walks techniques for seed picture segmentation, Authors in [14] devised a subMarkov random walks (subRW) approach with previous label information. This algorithm successfully segmented photos with thin objects.

**Active Contours:** In order to co-segment images, authors in [15] extended the active contour approach. They also built an energy function based on foreground coherence between images and background inconsistencies within each image, and they addressed the energy function minimising by level set. In order to solve the problem of segmenting the brain MRI image, authors in [16] put forward a deformable co-segmentation system that converted the prior heuristic data regarding brain anatomy present in numerous pictures into limitations regulating the brain MRI segmentation and acquired the minimum energy function by level set.

**Clustering-Based**: The clustering approach can be used to solve the multi-objective co-segmentation problem if the number of initial cluster centres is not constrained. The algorithm performs the actions listed below. First, using image preprocessing, the image is divided into local regions with several superpixel blocks. The corresponding prior knowledge is then formed by grouping these small regions using a clustering technique. To achieve multi-object co-segmentation, the prior knowledge is conveyed in a set of images. Spectral clustering was utilised by authors in [17] to capture the local information in a single image using a similarity matrix based on feature positions and colour vectors.

**Graph Theory:** An image is divided into a digraph via co-segmentation, which is based on graph theory. Authors in [18] created a digraph by employing the local regions of each image as nodes as opposed to superpixels or pixels as nodes, which is different from the digraph previously stated. Each image was separated into many local areas based on the 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 determining the quickest route on the digraph was then applied to the image co-segmentation challenge. Lastly, they used the dynamic programming (DP) approach to find the shortest route.

## V. DEEP LEARNING BASED METHOD

The richness of details in images and the diversity of objects (e.g., scale, posture) have greatly increased with the ongoing advancement of image acquisition technology. The higher generalisation ability of image segmentation models is advocated because features that are low-level, such as colour, brightness, and texture, are challenging to obtain good segmentation results from and methods for extracting features based on manual or heuristic rules are unable to satisfy the complicated requirements present in image segmentation.

**Encoding–Decoding:** 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 utilising a predetermined sampling statistic as the region's representative characteristic. Segmentation network encoders frequently employ VGG, Inception [19], and ResNet [20] as their backbone blocks.

**Skip Connection:** 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 to address this issue [21]. The new long skip connection proposed by U-Net [22] is depicted in Figure 8 in contrast. To retrieve the fine-grained details of images, U-Net creates jump links and feature cascades from layers in the encoder to the appropriate layers in the decoder. Since then, it has been extensively employed in studies on medical image segmentation. It was initially offered as a solution to the issue of annotations in image segmentation based on biological microscopes.

**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 while down sampling. To preserve the receiving area of the corresponding layer's receiving area and the high resolution of the feature mapping in FCN, the max-pooling layers are substituted with dilated convolution.

To properly address the "gridding" issue brought on by dilated convolution, Authors in [23] presented a hybrid dilated convolution (HDC). The HDC ensures that a square region is completely covered by the final size of the receptive field after a sequence of convolution processes, with no gaps or missing edges. Instead of utilising the same dilation rate for all layers following earlier down-sampling, they used distinct dilation rates for each layer to enable this.

**Multiscale Feature:** It was suggested that spatial pyramid pooling (SPP) be used to address the CNNs' need for fixed-size input pictures. The SPP-net was created by Authors in [24] and its efficacy in semantic segmentation and object detection was confirmed. Authors in [25] created PSPNet with a pyramid pooling module (PPM) to maximise the use of image context information. The PSPNet used PPM to extract and collect various subregion features at various scales, which were then up-sampled and concatenated to generate the feature map, which incorporated both local and global context information, using ResNet as the backbone network. It is important to note that the size of each layer and the number of layers in a pyramid vary depending on the size of the feature.

## VI. CONCLUSION

We have thoroughly sorted the traditional segmentation algorithms and the current hot deep learning methods, explained on the corresponding solutions of each stage, and listed the traditional algorithms with specific influences in accordance with the chronological development of image segmentation technology. Generally speaking, there has been a shift in the development of image segmentation from coarse to fine-grained, from traditional feature extraction to adaptive learning, and from segmentation based on a single image to segmentation based on common characteristics of huge data. Image segmentation has moved from the CNN stage to the

https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071
Volume: 5 Issue: 5 | May 2023

transformer stage thanks to the significant advancements achieved by the swin transformer in the area of computer vision in 2021. The transformer may lead to new developments in the area of computer vision studies. Deep learning does, however, have several drawbacks, such as the inexplicability of deep learning, which restricts the robustness, dependability, and performance optimisation of its subsequent tasks.

## References


1. Anwesh, K., Pal, D., Ganguly, D., Chatterjee, K., Roy, S. Number plate recognition from enhanced super-resolution using generative adversarial network. Multimed. Tools Appl. 2022, 1–17.
2. Yogita, Shalini Bhadola, Kirti Bhatia, Rohini Sharma, A Deep Overview of Image Inpainting Approaches, International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), Volume 11, Issue 6, June 2022, pp.8744-8749.
3. Yogita, Shalini Bhadola, Kirti Bhatia, Rohini Sharma, Development and Analysis of Enhanced Image Inpainting Approach, International Journal of Trend in Scientific Research and Development (IJTSRD), Volume 6 Issue 4, May-June 2022, pp. 1731-1737.
4. Yu Y, Wang C, Fu Q, Kou R, Huang F, Yang B, Yang T, Gao M. Techniques and Challenges of Image Segmentation: A Review. Electronics. 2023; 12(5):1199.
5. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. IEEE Trans. Pattern Anal. Mach. Intell. 2022, 44, 3523–3542.
6. Yu, Y.; Wang, C.; Fu, Q.; Kou, R.; Wu, W.; Liu, T. A Survey of Evaluation Metrics and Methods for Semantic Segmentation. Comput. Eng. Appl. 2023; online preprint.
7. Lankton, S.; Tannenbaum, A. Localizing Region-Based Active Contours. IEEE Trans. Image Process. 2008, 17, 2029–2039.
8. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient Graph-Based Image Segmentation. Int. J. Comput. Vis. 2004, 59, 167–181.
9. Comaniciu, D.; Meer, P. Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Trans. Pattern Anal. Mach. Intell. 2002, 24, 603–619.
10. Comaniciu, D.; Meer, P. Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Trans. Pattern Anal. Mach. Intell. 2002, 24, 603–619.
11. Rother, C.; Minka, T.P.; Blake, A.; Kolmogorov, V. Cosegmentation of Image Pairs by Histogram Matching—Incorporating a Global Constraint into MRFs. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA, 17–22 June 2006; pp. 993–1000.
12. Collins, M.D.; Xu, J.; Grady, L.; Singh, V. Random Walks based Multi-image Segmentation: Quasiconvexity Results and GPU-based Solutions. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1656–1663.
13. Collins, M.D.; Xu, J.; Grady, L.; Singh, V. Random Walks based Multi-image Segmentation: Quasiconvexity Results and GPU-based Solutions. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1656–1663.

https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071

Volume: 5 Issue: 5 | May 2023

14. Collins, M.D.; Xu, J.; Grady, L.; Singh, V. Random Walks based Multi-image Segmentation: Quasiconvexity Results and GPU-based Solutions. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1656–1663.
15. Meng, F.; Li, H.; Liu, G. Image Co-segmentation via Active Contours. In Proceedings of the 2012 IEEE International Symposium on Circuits and Systems (ISCAS), Seoul, Republic of Korea, 20–23 May 2012; pp. 2773–2776.
16. Meng, F.; Li, H.; Liu, G. Image Co-segmentation via Active Contours. In Proceedings of the 2012 IEEE International Symposium on Circuits and Systems (ISCAS), Seoul, Republic of Korea, 20–23 May 2012; pp. 2773–2776.
17. Joulin, A.; Bach, F.; Ponce, J. Multi-class Cosegmentation. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 542–549.
18. Meng, F.; Li, H.; Liu, G.; Ngan, K.N. Object Co-Segmentation Based on Shortest Path Algorithm and Saliency Model. IEEE Trans. Multimed. 2012, 14, 1429–1441.
19. Szegedy, C.; Liu, W.; Jia, Y.Q.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
20. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
21. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
22. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv 2015, arXiv:1505.04597.
23. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.H.; Hou, X.D.; Cottrell, G. Understanding Convolution for Semantic Segmentation. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460.
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 37, 1904–1916.
25. Zhao, H.S.; Shi, J.P.; Qi, X.J.; Jia, J.Y. Pyramid Scene Parsing Network. arXiv 2017, arXiv:1612.01105v2.