International Journal of Scientific Research in Engineering & Technology
Volume4, Issue3 (May-June 2024), PP: 72-77.
www.ijsreat.com ISSN No: 2583-1240

Reconnoitering Image Segmentation Methods: Techniques, Challenges, and Trends

Srishty Jain[1], Meenakshi Arora[2], Rohini Sharma[3]
[1]P.G. Student, Department of CSE, Sat Kabir Institute of Technology and Management, Bahadurgarh, Haryana, India.
[2]Assistant Professor, of CSE, Sat Kabir Institute of Technology and Management, Bahadurgarh, Haryana, India.
[3] Assistant Professor, CS, GPGCW, Rohtak, Haryana, India.

To Cite this Article: *Srishty Jain[1], Meenakshi Arora[2], Rohini Sharma[3], "Reconnoitering Image Segmentation Methods: Techniques, Challenges, and Trends", International Journal of Scientific Research in Engineering & Technology Volume 04, Issue 03, May-June 2024, PP: 72-77.*

Abstract: *Image segmentation is a fundamental process in computer vision and image analysis, involving the partitioning of an image into meaningful regions or segments. This review provides a comprehensive overview of various image segmentation techniques, highlighting their methodologies, applications, strengths, and limitations. Over the years, various segmentation methods have been developed, each with its advantages and drawbacks. This review covers traditional methods, machine learning approaches, and modern deep learning techniques.*

Key Word: *Image Segmentation, Thresholding, Pixel, Handcrafted Methods, Machine Learning Based Approaches.*

I.INTRODUCTION

The goal of image segmentation (IS), a subfield of digital image processing and computer vision, is to classify related areas or segments of an image under the appropriate labels. Forming segments is the same as combining pixels because the entire procedure is digital and provides a pixel-by-pixel description of the analog image. Image segmentation is an advancement of image classification in which localization is done as well as classification[1]. Hence, image segmentation is a subset of image classification, wherein the model uses the borders of an object to identify where the associated object is present. According to the quantity and nature of information they communicate, image segmentation tasks can be divided into three categories. Instance segmentation, which does not know which class an object belongs to, generates a segment map for each thing it examines in the image, whereas semantic segmentation segments off a wide boundary of objects belonging to a specific class. As the combination of instance and semantic segmentation tasks, panoptic segmentation is by far the most informative[2]. With panoptic segmentation, we can obtain segment maps of every object in the image belonging to a specific class.

Semantic Segmentation is a subfield of image segmentation where the goal is to classify each pixel in an image into a predefined category. Unlike object detection or instance segmentation, which aim to detect objects or individual instances of objects, semantic segmentation focuses on labeling regions of the image that belong to the same class.

Instance segmentation is a complex and advanced task in computer vision that involves not only classifying each pixel in an image but also distinguishing between different instances of the same object class.

Panoptic segmentation is an advanced computer vision task that aims to unify the concepts of semantic segmentation and instance segmentation. In panoptic segmentation, each pixel in an image is assigned a semantic label (like semantic segmentation) and, for object classes, also an instance ID (like instance segmentation).

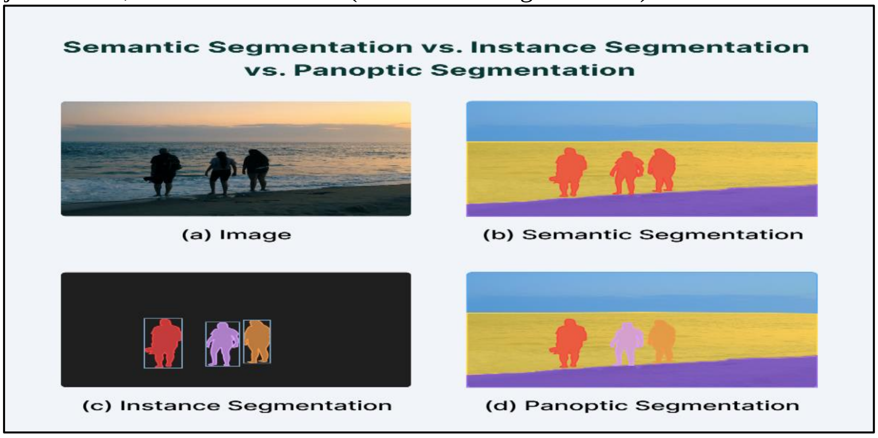

Figure 1: Different types of Image Segmentations [3]

II.RESEARCH BACKGROUND

A hybrid method using complex networks and dynamic image processing was presented in [1] to find welding faults in oil and gas pipeline radiography that repeatedly show up as pictures. Radiographic images are interpreted by qualified interpreters, and nondestructive testing relies heavily on the investigation of welding in gas and oil pipelines. The area growth technique has limits in photos with fewer subject variety, but it is useful for segmenting images and identifying welding faults. The suggested approach employs a histogram to ascertain the beginning and ending images of the welding range, after which a number of conventional techniques are used to detect flaws. The image's key spots are retrieved, and the matching complicated dynamic network is sketched and its computations carried out[4].

The authors of [5]offer a novel method for segmenting images that is based on social network community identification techniques. The authors suggest an approach for community detection in graphs that makes use of super pixels and methods. Super-pixel techniques minimize the number of nodes inside the graph, whereas community identification algorithms yield more precise segmentation in comparison to conventional methods. The approach is contrasted with the deep learning and prior work-based image segmentation technique. The approach gives more accurate segmentation, according to the experimental findings. A straightforward framework for community detection was presented in [6], and it addresses everything from the process of creating a graph from feature vectors created from non-graph data to the application and assessment of community detection techniques on such a graph. The framework is further tested on the invariant pattern clustering of images problem, which is essentially clustering the images associated with each object given a series of image objects acquired from multiple positions, angles, or orientations.

III. CATEGORIZATION OF IMAGE SEGMENTATION METHODS

The two main categories of currently used image segmentation techniques are handcrafted features and Deep Learning (DL) algorithms.

Table 1: A comparison of HandCraft based IS Approaches

Reference	Feature Type	Application Domain	Method Description	Strengths	Limitations
[7]	Textural Features	General Image Classification	captures spatial links by extracting 14 textural elements depending on the gray-level co-occurrence matrix (GLCM).	acquires spatial information and has a robust texture description.	dependent on the direction and size of the window; theoretically costly for huge photos.
[8]	Gabor Descriptors	Aerial Image Classification	captures textural information at many scales and orientations by using Gabor filters.	Efficient for multi-scale, multi-orientation evaluation, and texture and pattern recognition.	Highly efficient and filter parameter-sensitive.
[9]	Spatial Pyramid Co-occurrence	General Image Classification	preserves spatial relationships and texture at many scales by combining co-occurrence matrices with spatial pyramids..	catches both local and global characteristics, adapting well to changes in the orientation and scale of the image.	significant computational difficulty; precise parameter adjustment is necessary.
[10]	Multifeature Probabilistic LSA	High Spatial Resolution Remote Sensing Images	identifies semantic scene structures by applying probabilistic latent semantic analysis (pLSA) to a variety of features.	combines several feature kinds and obtains highly sophisticated semantic data.	costly to compute and in need of a lot of training data.
[11]	Fisher Kernel Coding	High Spatial Resolution Scene Classification	combines local features into a global description for scene classification using Fisher Kernel coding..	Robust against changes in image content, efficient at capturing intricate scene structures.	has significant memory and processing needs, as well as a complicated training procedure.

IV. DEEP LEARNING-BASED IMAGE SEGMENTATION

Deep learning has significantly advanced the field of image segmentation, offering robust and highly accurate methods for various segmentation tasks. This review covers the main deep learning-based techniques for semantic segmentation, instance segmentation, and panoptic segmentation.

Table 2: A comparison of Deep Learning based IS Approaches

Method	Task	Architecture	Key Features	Strengths	Limitations
FCNs [12]	Semantic	Convolutional	End-to-end learning, dense predictions	High spatial resolution, simple architecture	Limited context understanding
U-Net [13]	Semantic	Encoder-decoder with skips	Skip connections for spatial information	Effective for small datasets, preserves details	High memory usage
SegNet [14]	Semantic	Encoder-decoder	Pooling indices for up sampling	Efficient memory usage, fast inference	Reduced feature map accuracy
DeepLab [15]	Semantic	Atrous convolutions, CRFs	Multi-scale context, boundary refinement	High accuracy, refined boundaries	Complex, computationally expensive
Mask R-CNN [16]	Instance	Faster R-CNN with mask branch	Bounding box, class, and mask prediction	High accuracy, handles overlaps well	Computationally intensive, complex training
YOLACT [17]	Instance	Prototype masks, localization	Decoupled mask and localization	Fast inference, simple training	Lower accuracy compared to Mask R-CNN
SOLO [18]	Instance	Location-based	Location prediction for masks	Simple, fast	Limited scale handling
Panoptic FPN [19]	Panoptic	FPN with dual branches	Combines semantic and instance segmentation	High accuracy, unified approach	Computationally intensive, complex
Panoptic-Deep Lab [20]	Panoptic	Extended DeepLab	Dual heads for semantic and instance tasks	Powerful backbone, high accuracy	High computational resources required
DETR [21]	Panoptic	Transformer-based	Global context, unified object detection/segmentation	Simplified architecture, global understanding	High computational demand, large datasets needed

FCNs replace the filly connected layers in traditional CNNs with convolutional layers, enabling pixel-wise prediction. The network is trained end-to-end, producing dense predictions for segmentation. **U-Net** is an encoder-decoder network with symmetric skip connections that transfer spatial information from the encoder to the decoder, improving segmentation accuracy, especially for biomedical images. **SegNet** is an encoder-decoder architecture where the encoder is identical to a standard convolutional network and the decoder uses pooling indices from the encoder for upsampling, reducing the computational load. **DeepLab** employs atrous (dilated) convolutions to capture multi-scale context and Conditional Random Fields (CRFs) for post-processing, refining the boundaries. **Mask R-CNN** extends Faster R-CNN by adding a branch for predicting segmentation masks in parallel with the bounding box and class prediction branches. **YOLACT** decouples mask prediction from localization, generating a set of prototype masks for each image and combining them linearly to produce instance masks. **SOLO** segments objects by predicting locations and masks simultaneously, treating instance segmentation as a dense prediction problem. **Panoptic FPN** combines the outputs of a semantic segmentation branch and an instance segmentation branch, merging them to produce panoptic predictions. **Panoptic-Deep Lab** extends the DeepLab architecture to panoptic segmentation by incorporating both semantic and instance segmentation heads. **DETR** uses a transformer-based architecture for object detection and segmentation, capturing global context and handling both tasks in a unified framework.

V. PIXEL-BASED IMAGE SEGMENTATION

Pixel-based image segmentation is a fundamental technique in computer vision that involves partitioning an image into distinct regions at the pixel level. Pixels are the fundamental building block of picture analysis, according to several recent image segmentation algorithms; nonetheless, the majority of these algorithms have neglected the spatial connection between pixels, producing subpar image border segmentation. In the meantime, the super-pixels technique can match the created hyperparameter region's perimeter to the edge of a substance or the image's backdrop.

Table 3: A comparison of Pixel based IS Approaches

Method	Key Features	Strengths	Limitations
[22]	SLIC stands for Simple Linear Iterative Clustering. employs k-means clustering in the five-dimensional space of image coordinates and color.	yields superpixels that are small and nearly uniform. computationally effective. Simple to put into practice.	Effectiveness is contingent upon the starting parameters. difficulties in areas with a lot of texture.
[23]	integrates superpixel segmentation with deep learning. makes use of superpixels to make complicated histology pictures simpler.	high precision in digital histology image nuclei detection. simplifies calculation by utilizing superpixels.	big annotated datasets are necessary for training. heavy on computations when in training.
[24]	employs superpixels in semantic segmentation to facilitate active learning. aims to lower the cost of annotations.	Expands annotation efficiency. Reduces the cost of manual labeling.	Depending on how well the created superpixel is quality. It might need to be adjusted for various datasets.
[25]	integrates superpixel feature extraction using contractive autoencoder. intended to identify changes in SAR photos.	efficient in identifying alterations in SAR pictures. robust to changes and noise in picture data .	It might not translate well to other kinds of photos. need the autoencoder to be carefully adjusted.
[26]	blends superpixel efficiency with deep learning. specifically designed to separate rice panicles in field photos	robust segmentation of panicles of rice under complicated field circumstances. High precision as a result of optimizing super pixels.	specifically designed for use in agriculture. domain-specific tweaking is necessary.

Table 4: A comparison of Traditional IS Approaches

Method	Key Features	Strengths	Limitations
Thresholding	Using a preset intensity value as a guide, binary conversion is used to segment images. Easy to use and quickly implemented	Computationally inexpensive. Applicable for high-contrast images.	sensitive to changes in illumination. Not useful for photographs with low contrast or complexity.
Region-Based Segmentation	divides an image into sections according to predetermined standards, like closeness in intensity. Methods include region splitting and merging, as well as region expansion.	useful for dividing areas that are homogeneous. maintains region-wide connectivity.	susceptible to the initial areas or seed spots selected. computationally demanding in the case of big photos.
Edge	finds discontinuities in pixel intensity to detect	Excellent for drawing the	susceptible to

Segmentation	borders. Sobel, Prewitt, and Canny operators are examples of common techniques.	edges of objects. helpful in identifying structures and forms	both texture and noise. may result in false positives or missing edges.
Clustering-based Segmentation	clusters pixels according to shared features, including color or texture. K-means and mean shift are examples of common methods.	efficient for a range of picture formats. capable of managing various feature dimensions.	susceptible to the choice of the number of clusters. Computationally expensive for large datasets.

VI. CONCLUSION

Each image segmentation method has its unique advantages and drawbacks, making them suitable for different applications and types of images. Every meticulously designed picture segmentation technique possesses distinct advantages and is appropriate for particular application areas. For texture analysis, Haralick's texture characteristics and Gabor descriptors work well; for capturing multi-scale and intricate scene structures, spatial pyramid co-occurrence and Fisher Kernel coding work well. Several feature types are integrated using techniques such as multi-feature probabilistic LSA, which offer a full representation but at the expense of higher computational costs and data needs. Thresholding is ideal for simple, high-contrast images, while region-based segmentation is beneficial for homogeneous regions. Edge segmentation excels in detecting boundaries and shapes, and clustering-based segmentation offers flexibility and effectiveness across various image types. The choice of method should be guided by the specific requirements of the segmentation task, including the complexity of the images, computational resources, and the desired accuracy. This review highlights the diversity in pixel-based image segmentation methods, spanning from traditional approaches to advanced deep learning techniques. Deep learning-based image segmentation has made significant strides, with different methods excelling in various segmentation tasks. Each method has its unique strengths and limitations, making them suitable for different applications and datasets. The choice of method depends on specific requirements, such as accuracy, computational efficiency, and the nature of the images being segmented.

REFERENCES

[1] *R. S. Nitesh Vashist, Parvesh, Kirti Bhatia, "AN OVERVIEW: IMAGE SEGMENTATION USING REGION AND CLUSTERING METHODS," Int. Res. J. Mod. Eng. Technol. Sci., vol. 5, no. 6, pp. 2258–2264, 2023.*

[2] *R. S. Nitesh Vashist, Kirti Bhatia, Parvesh, "An Innovative Region Growing and Region Merging Image segmentation Method," Int. J. Adv. Res. Arts, Sci. Eng. Manag., vol. 10, no. 3, pp. 1938–1944, 2023.*

[3] *H. Bandyopadhyay, "An Introduction to Image Segmentation: Deep Learning vs. Traditional." https://www.v7labs.com/blog/image-segmentation-guide*

[4] *A. Ebrahimi, K. Mirzaie, and A. M. Latif, "A hybrid approach of dynamic image processing and complex network to identify repetitive images of welding defects in radiographs of oil and gas pipelines," Int. J. Nonlinear Anal. Appl., vol. 14, no. 1, pp. 1671–1682, 2023.*

[5] *I. Gammoudi and M. A. Mahjoub, "Brain tumor segmentation using community detection algorithm," in 2021 International Conference on Cyberworlds (CW), 2021, pp. 57–63.*

[6] *L. M. Freitas and M. G. Carneiro, "Community detection to invariant pattern clustering in images," in 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), 2019, pp. 610–615.*

[7] *R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," IEEE Trans. Syst. Man. Cybern., no. 6, pp. 610–621, 1973.*

[8] *V. Risojević, S. Momić, and Z. Babić, "Gabor descriptors for aerial image classification," in Adaptive and Natural Computing Algorithms: 10th International Conference, ICANNGA 2011, Ljubljana, Slovenia, April 14-16, 2011, Proceedings, Part II 10, 2011, pp. 51–60.*

[9] *Y. Yang and S. Newsam, "Spatial pyramid co-occurrence for image classification," in 2011 International Conference on Computer Vision, 2011, pp. 1465–1472.*

[10] *Y. Zhong, M. Cui, Q. Zhu, and L. Zhang, "Scene classification based on multifeature probabilistic latent semantic analysis for high spatial resolution remote sensing images," J. Appl. Remote Sens., vol. 9, no. 1, p. 95064, 2015.*

[11] *F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in 2007 IEEE conference on computer vision and pattern recognition, 2007, pp. 1–8.*

[12] *J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.*

[13] *O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, 2015, pp. 234–241.*

[14] *V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, 2017.*

[15] *L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, 2017.*

[16] *K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision,*

2017, pp. 2961–2969.

[17] *D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "Yolact: Real-time instance segmentation," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 9157–9166.*

[18] *X. Wang, T. Kong, C. Shen, Y. Jiang, and L. Li, "Solo: Segmenting objects by locations," in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16, 2020, pp. 649–665.*

[19] *A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 6399–6408.*

[20] *B. Cheng et al., "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 12475–12485.*

[21] *N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in European conference on computer vision, 2020, pp. 213–229.*

[22] *R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274–2282, 2012.*

[23] *S. Sornapudi et al., "Deep learning nuclei detection in digitized histology images by superpixels," J. Pathol. Inform., vol. 9, no. 1, p. 5, 2018.*

[24] *L. Cai, X. Xu, J. H. Liew, and C. S. Foo, "Revisiting superpixels for active learning in semantic segmentation with realistic annotation costs," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 10988–10997.*

[25] *N. Lv, C. Chen, T. Qiu, and A. K. Sangaiah, "Deep learning and superpixel feature extraction based on contractive autoencoder for change detection in SAR images," IEEE Trans. Ind. informatics, vol. 14, no. 12, pp. 5530–5538, 2018.*

[26] *X. Xiong et al., "Panicle-SEG: a robust image segmentation method for rice panicles in field based on deep learning and superpixel optimization," Plant Methods, vol. 13, pp. 1–15, 2017.*