International Journal of Mathematical, Engineering and Management Sciences
Vol. 9, No. 5, 1163-1184, 2024
https://doi.org/10.33889/IJMEMS.2024.9.5.061

Ram Arti Publishers

An Overview on Image Segmentation Techniques for Reversible Data Hiding

Rasika Gupta
Department of Computer Science & Engineering,
Delhi Technological University, Delhi, India.
E-mail: rashika99@gmail.com

(Received on March 10, 2024; Revised on April 24, 2024 & May 10, 2024; Accepted on May 24, 2024)

Abstract
The fields of image processing and computer vision have witnessed significant growth due to the proliferation of digital images across diverse domains. Image Segmentation is the fundamental task in digital image processing, finding applications in pivotal areas such as medical imaging, covert communication, autonomous driving, satellite imaging, among others. One particularly intriguing application of image segmentation lies in Reversible Data Hiding (RDH), where the delineation of the main Region of Interest (ROI) and Non-Region of Interest (NROI) using segmentation plays a crucial role for effective data encryption in the images. Over the last two decades, various studies focussed on developing an efficient data hiding approach, which can embed secret data within ROI and NROI part of image while ensuring its quality. A comprehensive survey has been conducted that meticulously examines different segmentation techniques, along with its usage in reversible data hiding. The main objective of this survey is to compare the performance metrics of reversible data hiding after applying different image segmentation techniques. The image segmentation techniques have been categorized systematically into three main classes: i) Traditional segmentation techniques, encompassing a spectrum of approaches like thresholding, region-based and edge detection based techniques, ii) Machine Learning (ML) based approach consisting of Clustering, Support Vector Machine (SVM) and iii) Deep Learning (DL) based technique, propelled by Convolutional Neural Networks (CNNs) that have emerged as a transformative paradigm, revolutionizing segmentation tasks with their ability to learn complex images. The survey finds out that PSNR value of data embedded images is high after applying deep learning based segmentation technique.

Keywords- Image segmentation, Convolutional neural network, Encoder-decoder model, Dilated convolution model, ROI segmentation, Reversible data hiding.

1. Introduction

Image segmentation has long been recognized as a fundamental challenge in the field of computer vision (Stockman & Shapiro, 2001), dating back to its inception. Image segmentation has been defined as a process of dividing an image into homogeneous segments with an objective to get the Region of Interest (ROI) from its surrounding background (Non ROI). The segmented image is easier to analyse and can be used for different domains (Forsyth & Ponce, 2002). This ROI can be used to embed secret information as far as possible and to increase visual quality of an image by contrast enhancement (Yang et al., 2018). More data can be embedded in the NROI, as it is the less critical region of an image. The selective embedding modifies only the less important or irrelevant region of image, without much affecting the visually critical region. The embedding capacity of an image can be increased by segmenting an image (Yang et al., 2018). A low contrast image has poor visual quality after data embedding. Therefore, image pre-processing operations such as segmentation and normalization are combined with RDH so as to improve the image quality. This technique of hiding data creates a balance between embedding capacity, imperceptibility and security of images. The technique can be used in many information securing fields like medical, military, and satellites based remote sensing studies etc. The paper aims to study the various developments in these techniques through the literature review, to learn about the works done so far and to identify gap areas so that further research work can be planned for improving its accuracy and applications to different fields.

Image segmentation has broadly been categorized into three types: i) separating the boundaries of different objects in an image (instance segmentation), ii) labelling each pixel with its corresponding semantic labels (semantic segmentation) and iii) combination of both (panoptic segmentation). Semantic segmentation gives class label to each pixel and then instance segmentation identify each object instance in the image.

Different image segmentation techniques have been developed over a period of time for each type of image segmentation, beginning with the classic traditional methods which primarily concentrated on extracting information from images, often required human intervention. But extracting high-level semantic information from images proved to be challenging. Hence, it was followed with application of Machine Learning (ML) algorithms that helped in preserving semantic information. But these algorithms were unable to provide good results with large data sets. In recent years, Deep Learning (DL) methods have gained popularity due to large size image datasets and provided an opportunity to deal with such challenges.

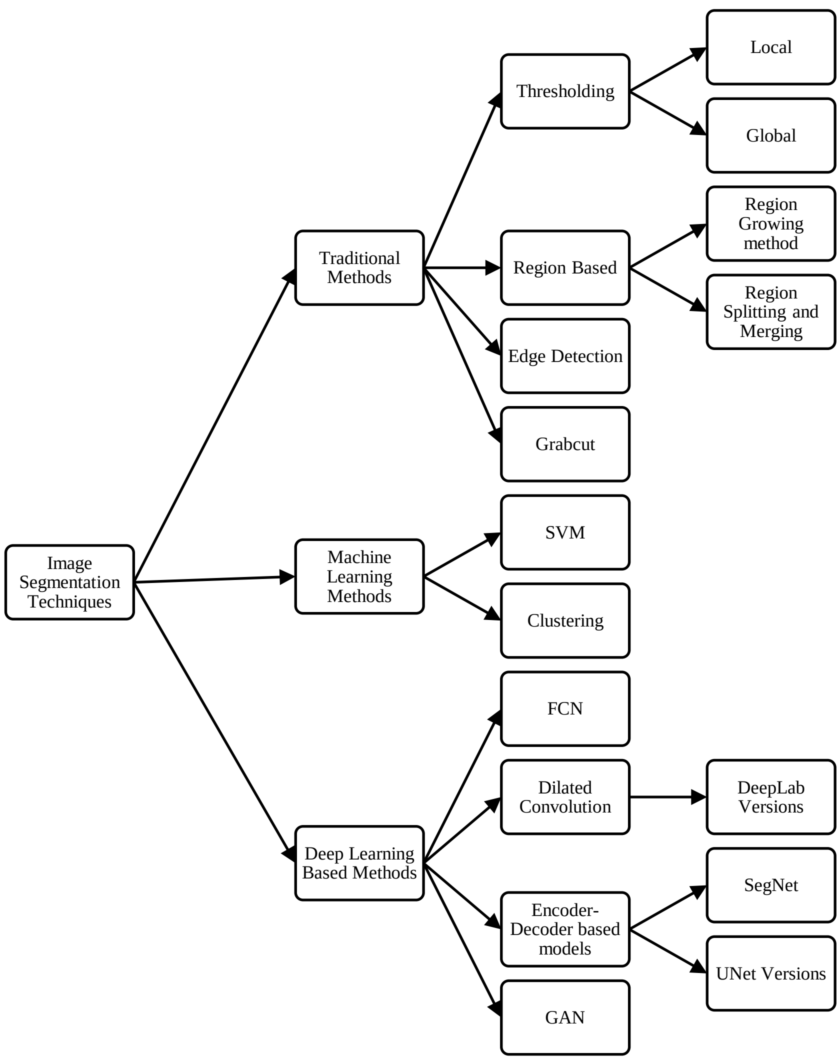

Figure 1. Image segmentation techniques.

This survey has comprehensively summarized the image segmentation methods (**Figure 1**), spanning from conventional approaches to the recent deep learning-based methods, while elucidating their applications in Reversible Data Hiding as explored by various researchers.

The main contributions of this paper may be highlighted as follows:
(i) No existing survey is present which focuses on the combined approach of image segmentation and reversible data hiding. Through this paper, an overview has been presented to showcase how image segmentation techniques have been employed by various researchers in the domain of data hiding.
(ii) The present paper has also discussed about the various image segmentation techniques and has made a comparison between them across different metrics.
(iii) A comprehensive Table has been developed to illustrate the advantages, limitations of different segmentation techniques and their performance in reversible data hiding images.

The paper has systematically discussed the traditional methods in Section 2, machine learning methods in Section 3, and the deep learning methods in Section 4. Section 5 deals with results and discussions, followed by the concise conclusions in Section 6.

2. Traditional Methods
The traditional methods can be broadly categorized into: i) Thresholding Methods, ii) Region Based Segmentation, iii) Edge Detection, and iv) GrabCut Segmentation. These segmentation techniques depend upon two essential aspects within images, i.e. discontinuity and similarity. The different types of traditional techniques have been briefly described below.

2.1 Thresholding
It is one of the primary methods for image segmentation. A threshold value is used to segment different regions of an image.
(i) *Global Thresholding*- In this method, only a single threshold value is used, as the difference between the foreground and background pixel is very large.
(ii) *Local Thresholding*- In this method, different regions of an image are segmented by using different threshold values.

Otsu (1979) method is an example of the implementation of the thresholding technique. But it is biased towards the data class with the larger number of data when the variances between the classes are quite different. Therefore, Adaptive Threshold Detector (ATD) by Pai et al. (2011) computes the threshold value for each pixel by considering the local neighbourhood around it. It considers the trade-off between data size, class interval and standard deviation of a class.

Gao et al. (2017) used method given by Otsu (1979) for segmenting ROI and background region of Computed Tomography (CT) images. Then, the data has been embedded in the ROI region of an image. This approach of Reversible Data Hiding has helped in preserving the visual quality of the background and the contrast enhancement of ROI. However, Pai et al. (2012) and Yang et al. (2018) used ATD method to automatically segment ROI and NROI regions. The embedding rate of data in an image by this method has been found to be higher than Gao et al. method.

2.2 Region Based Segmentation
This method works by comparing pixels with its neighbouring pixels on the basis of some parameters like color, texture, intensity and then divides the region based on similarities.

(i) *Region Growing Method*- In this method, some pixels are selected as seed particles. Then, the seed particles are assigned to the same region, if they have the same characteristics. This process continues until all the seed particles which have the same value assigned to the same region (Kwon, 2007). Parah et al. (2017) employed a region-based segmentation approach, wherein the image has been partitioned into blocks. Subsequently, each block has further been divided into seed and non-seed pixels, with data hidden within the non-seed pixel regions.

(ii) *Region Splitting and Merging*- Quad-tree based segmentation is an example of this method, where the region is partitioned into four quadrants and then merging is done until no further splitting is possible. Salman (2015) used this method which hierarchically partition input image into blocks, such that each block has maximum capacity. The secret data and the partition information are embedded inside each block using the histogram shifting approach of Reversible Data Hiding, where minimum and maximum peak pixel points are calculated, then histogram shifting is performed to embed secret data bits.

Lin and Li (2011) divided input image into variable size pixel blocks using Quad-Tree Segmentation according to the maximal capacity for each block. This hierarchical segmentation enables a more effective distribution of embedding capacity. For the data hiding, author has used multiple local histograms instead of single histogram, derived from pixels in locally contiguous sub-image areas. This approach has enabled the algorithm to exploit the simple local distribution of pixel intensities, leading to higher embedding capacity.

Chaithanya and Srujana (2022) used Quad-tree segmentation. In this, the blocks have been derived through partitioning the input image and have been arranged in a quad-tree structure. Leveraging the histogram shifting technique, blocks with lower contrast can accommodate a greater number of secret data bits, whereas blocks with higher contrast conceal a relatively smaller amount of secret data bits. Consequently, higher contrast blocks necessitate subdivision into smaller sub-blocks. These sub-blocks tend to exhibit smoother characteristics, collectively offering a larger hiding capacity.

2.3 Edge Detection

Objects in an image are outlined with clear edges and corners, making sure each object has its own space separate from others, with noticeable lines or borders in between. The objective of edge detection method (Mahmood et al., 2015) is to find these edges. Laplacian of Gaussian (Marr & Hildreth, 1980), Canny edge detector (Canny, 1986), Shen-Castan detector (Shen & Castan, 1992), Sobel edge detector (Kanopoulos, 1988) are some of the techniques to find out the edges.

Phadikar and Maity (2010) have generated mask by using ‘Sobel’ edge detector to find out the image gradient. Morphological ‘dilation’ operations used for filling linear gaps in edges and the holes in the foreground area have been closed using morphological ‘closing’ operations. The spurious lines in the background have been removed using morphological ‘erode’ operations. The erosion operation may also affect the foreground region, to handle this, morphological ‘thickening’ operation has been used to thicken the boundary of the foreground region. The resultant ROI has been obtained by multiplying the original image with the resultant mask. A Half toning technique has been used to extract image digest from ROI. The image digest contained important details about the ROI that must be kept during error concealment. The information from the image digest has then been embedded into the Region of Background (ROB) using the Quantization Index Modulation (QIM) data hiding method (Chen & Wornell, 2001). This method modulates quantization levels of the host image to embed additional information without any significant loss of visual quality of an image.

Lakshmanan and Rani (2018) have used this segmentation technique to find out the ROI of the image. Then, they used the Difference Expansion Technique to hide data in the NROI and ROI regions. In the difference expansion technique, the secret data has been embedded into the difference between adjacent pixel values such that the difference between the modified pixel value and the original pixel has been statistically distributed. If the difference between two adjacent pixels is large, more data can be embedded without causing any noticeable distortions. Hiding more data in the NROI region increases the payload capacity and also the lossless recovery of image and data is guaranteed.

2.4 GrabCut Segmentation

GrabCut (Rother et al., 2004) is a widely utilized graph-based image segmentation technique renowned for its simplicity and accuracy in implementation. Its fundamental concept involves representing an image as a network diagram (Graph) as described in **Figure 2**. Initially, the user is required to designate foreground and background samples. Then, the method calculates the distance between the image points based on the user provided foreground and background pixels as well as the distance between the image points themselves. Further, the segmentation energy weight values for each of the edges of a graph are calculated by combining these distances linearly. Then, the minimum energy value for splitting the edges within the graph has been found out to get the resultant segmented graph. As a result of this methodology, GrabCut tends to produce segmentation outcomes that are more comprehensive and exhibit natural boundaries compared to other methods.

Wu et al. (2020) adopted GrabCut technique to extract the ROI region from the image. For some of the images whose gray level between the ROI and background is close, this method outperforms Otsu's method (1979). The research has also highlighted the benefits of enhancing this ROI region to improve the visual effects. The high accuracy of GrabCut segmentation has the significant impact on the performance of data hiding.

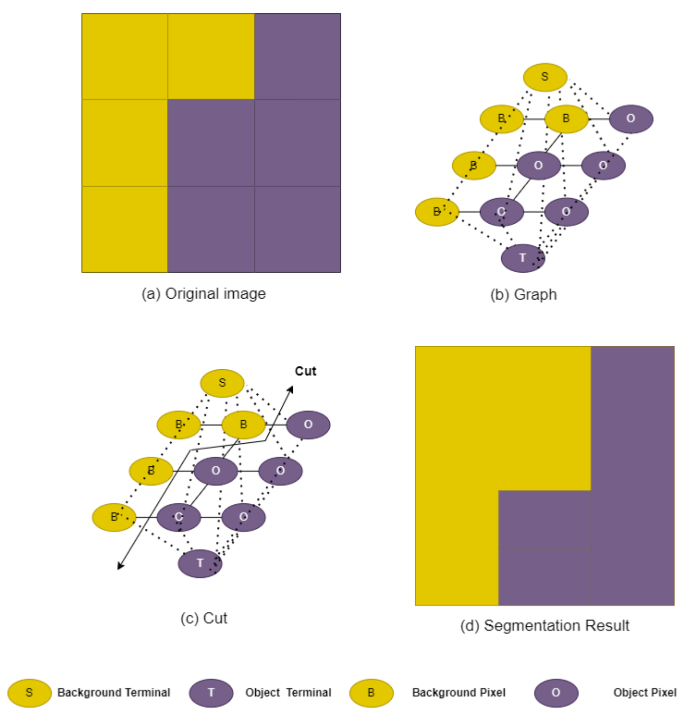

Figure 2. GrabCut segmentation.

3. Machine Learning Methods
Due to the ability of Machine Learning (ML) algorithms to learn from training data and automatically extract meaningful features. The two most used models, i.e. Support Vector Machine (SVM) and Clustering, have gained importance in the field of image segmentation. SVM is a supervised learning algorithm which requires labelled data whereas Clustering is an unsupervised learning algorithm which identifies patterns and groups within unlabelled datasets.

3.1 Support Vector Machine (SVM)
SVM classifier as shown in **Figure 3** can be used for image segmentation. Class 1 can be the object in an image and Class 2 can be the background of the image. Support Vectors are the data points that lie closest to the hyperplane between the two classes. These are used for determining the optimal hyperplane that can separate the two classes in the feature space. SVM is trained using features vectors like gradient information, texture descriptors, color histograms of images and prepared a labelled dataset with corresponding ground truth. A suitable kernel function is selected based on the data characteristics. For every pixel, SVM classifier determines whether it is the foreground/background pixel. Once trained, this classifier can be used to determine segmentation masks for new images. Liu et al. (2011) used modified SVM to increase its efficiency by taking only the main features for segmenting brain images obtained through Magnetic Resonance Imaging (MRI) and obtained good results.

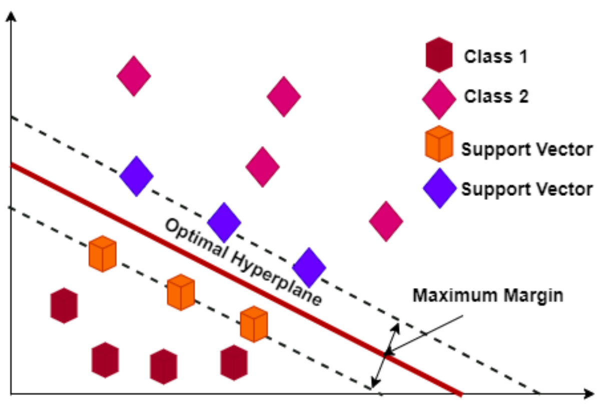

Figure 3. SVM classifier.

Rai and Singh (2017) employed an SVM classifier to classify medical images into ROI and NROI, so as to embed watermark data into the NROI using Discrete Wavelet Transform (DWT) - Singular Value Decomposition (SVD) technique. DWT divides the image into non-overlapping coefficients bands. Then the watermark is embedded into the selective coefficients. SVD protects image from attacks and small variation/change in the Singular Values of an image do not have much impact on visual quality.

Chowdhuri et al. (2023) used SVM to separate ROI and NROI of a medical image and then embed data in NROI using Integer Wavelet Transform, as NROI does not have any significance in diagnosis purpose. Multiple frequency bands are generated from cover image using Integer Wavelet Transform. This decomposition generates low-frequency and high-frequency coefficients. The embedding process involves modifying Least Significant Bit of the wavelet coefficients to embed secret data, with minimum distortion and maximum embedding capacity. The model has been tested on MRI images and the study has indicated that the SVM prediction of ROI and NROI performs better than other machine learning methods like K-Nearest Neighbour and Naïve Bayes classifier.

3.2 Clustering

Clustering is a technique where entities are grouped together into clusters based on their shared characteristics. Each cluster represent subset of data points that are more similar to each other compared to points in other clusters.

Kaur and Josan (2014) have used clustering approach to separate ROI from NROI and then have increased the brightness of ROI followed by hiding the data in it. Fuzzy based clustering by Ji et al. (2015) has combined geometric analysis with fuzzy logic. Radial lines have been drawn from a central point within an image. Along these lines, the significant changes in image characteristics such as intensity, color, or texture have been found using the membership function of fuzzy logic, which assigned a degree of belonging to each point based on these characteristics. These changes have helped in separating boundaries or Regions of Interest. This approach has enhanced partition compactness in segmentation results by ensuring contiguous regions based on spatial relationships and similarity to characteristic features (Balasamy & Suganyadevi, 2021).

Jiang et al. (2022) have used Simple Linear Iterative Clustering (SLIC) to partition image pixels into K irregular blocks, known as super-pixels, based on their similarity. Once the ROIs have been approximated, then these super-pixels have been used for embedding data using the rhombus prediction technique. In this technique, four adjacent neighbouring pixels have been used to calculate the prediction error. SLIC segmentation has eliminated the visual distortion of the background.

In Li et al. (2020) paper, the original image has been first pre-processed using the fuzzy C-means clustering algorithm to produce a number of highly linked classes. After that, each class has been split into two halves based on the appropriate threshold. One has been encrypted using a stream cypher, while the other has used compressive sensing technology to simultaneously encrypt and compress data for simple data embedding by the information hider. With the data-hiding key and encryption key, the recipient can retrieve the original image and extract additional data. In comparison to existing state-of-the-art techniques, experiments and analysis has shown that the suggested scheme can obtain a greater embedding rate about new data and better visual quality of the recovered image.

4. Deep Learning Methods

In the last few years, Deep Learning (DL) models have become a revolutionary breakthrough in the image segmentation field with significant performance gain. This paper has reviewed some of the State-Of-The-Art (SOTA) DL models used for image segmentation as discussed below.

4.1 Fully Convolutional Network (FCN)

Long et al. (2015) have introduced FCN for semantic segmentation, which is made up of convolutional layers that generate the segmentation map corresponding to input image of same size. It has been used as a baseline model to develop other deep learning-based segmentation neural networks. The most common FCN network used for image segmentation is AlexNet as displayed in **Figure 4**. The first two layers L1 and L2 are the convolutions, ReLu, Max Pooling and normalization. The third L3 and fourth L4 layers performed only convolution and ReLu operations. L5 is the same as L1 except it does not perform normalization. It also converts the output to the long vector for the L6 and L7 layers, which are the fully connected layers. The last layer L8 is the Softmax layer which generates labels for each pixel of input image.

FCN-AlexNet is not able to capture contextual information effectively, at different scales. FCN network is computationally expensive.

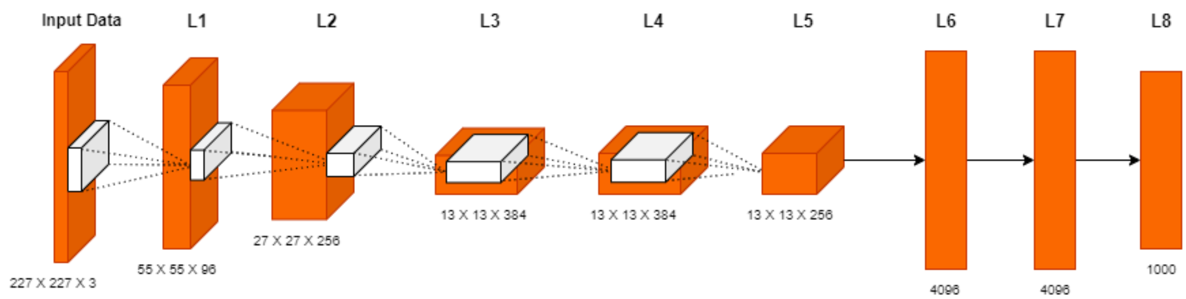

Figure 4. FCN-AlexNet architecture.

Meng et al. (2020) have used mask region-based convolution neural network (R-CNN), a state-of-the-art segmentation technique to generate mask of foreground object. Mask R-CNN uses a Fully Convolutional Network for pixel- level classification by feeding feature maps into the network. A Least Significant Bit (LSB) steganographic algorithm is added along with pixel classification to generate stego-image. This stego-image contains hidden secret data.

Luo et al. (2020) used ResNet, a deep convolutional neural network to extract semantic features from the images, capturing high-level information about the content of the image so as to make it easy to identify suitable object areas for information hiding. Mask R-CNN has been utilized for precise object detection and instance segmentation. This approach has allowed for the extraction of less object area, reducing the chances of attacks on the hidden information. Hash algorithms have been used to generate hash sequences corresponding to the selected object areas. These hash sequences are used for information mapping and embedding. The hidden information has been embedded in the selected object areas, and stego-images have been generated.

Saidi et al. (2024) have deployed Mask R-CNN model for segmentation to get the ROI of an image. Mask R-CNN has been designed to classify pixels within an image individually while accurately outlining object boundaries. This unique capability has made it well-suited for identifying regions suitable for data embedding. After segmentation, the proposed model has used Discrete Cosine Transform (DCT) for embedding of information into the frequency components, generated by transforming the spatial domain of an image into the frequency domain. This approach of data hiding has made the changes to the image that are imperceptible to the human eye.

4.2 Dilated Convolutional Segmentation Model

In this model, a dilation rate gets added to the convolutional layer. This increases the receptive field size without any increase in computational cost. One of the most prominent dilated convolutional models used for image segmentation is DeepLab family. Chen et al. (2017) discusses about DeepLabV1 and DeepLabV2 for image segmentation. These models are the combination of convolutional network and probabilistic graphical models like Conditional Random Fields (CRFs) and Markov Random Fields (MRFs).

Pan et al. (2020) have used MobilenetV2, developed based on DeepLab CNN for pixel-level image segmentation. The secret information is embedded inside the semantic features. It becomes difficult for the attacker to restore secret information from the semantic level with varying neural network parameters.

4.3 Encoder-Decoder Models

These are the convolution network consists of two parts- (i) the encoder part, made up of convolution and pooling layer which extract high-dimensional features maps containing the semantic information and (ii) decoder part generates segmentation masks using deconvolutional and unpooling.

(i) *SegNet*- It is a convolutional encoder-decoder model given by Badrinarayanan et al. (2017). It is used for pixel-wise semantic segmentation. SegNet as shown in **Figure 5**, consists of encoder network, made up of simple convolution, followed by Batch Normalization and Rectified Linear Unit (ReLu) activation function. ReLu helps in eliminating vanishing gradient problem occur due to the deep architecture of neural network. Dropout Layer has been used in the model to prevent overfitting. Each encoder block has a corresponding decoder block. At the end of the network, a Softmax layer is used for final pixel-wise classification to produce the segmentation mask. MaxPooling layer maintains translation invariance with respect to small shifts in the spatial position of features within the input image. During the UpSampling, the output of MaxPooling layer from the corresponding encoder block have been used. This has helped in preserving the spatial relationships captured during downsampling and accurately reconstructed during the UpSampling phase. It retains the integrity of high frequency information but during the downsampling of low resolution feature maps, it ignores neighboring information.

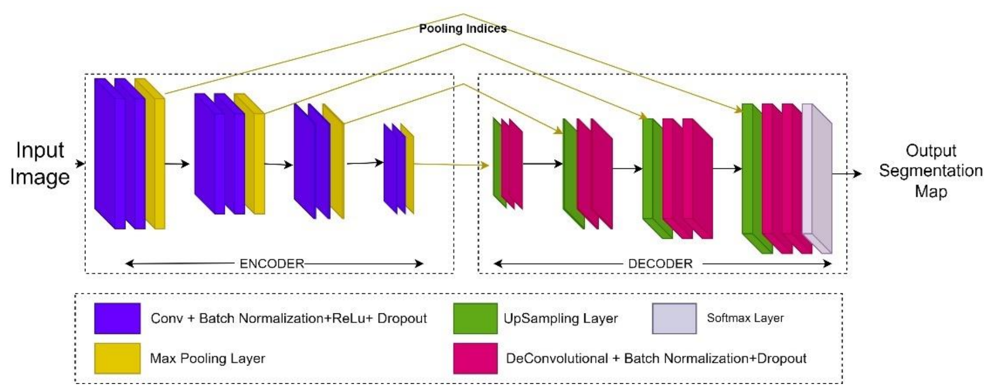

Figure 5. SegNet architecture.

(ii) *UNet*- To enhance the accuracy of pixel positioning, skip connections have been introduced in encoder-decoder model. As deep neural networks grow in depth, their performance tends to decrease. UNet is a novel approach called the long skip connections model that transmit and cascades features from encoder layers to corresponding decoder layers, enabling the capture of fine-grained image details. UNet by Ronneberger et al. (2015) has become widely adopted in medical image segmentation research. The U-Shaped structure help in capturing fine details at low-level spatial dimension. U-Net can be used for low training data, because the skip connections allow the use of feature maps from the contracting path directly into the expansive path.

Since its introduction in 2015, the UNet architecture has undergone numerous variations and improvements, leading to the development of several model variants up to the present day. Here is an elaboration on the evolution of UNet models over this period.

- *Base UNet*- UNet has two parts: the first part of the network is the contracting path, which consists of block of 3*3 convolution operation followed by the ReLu activation function. The output of these block is pass through a MaxPooling layer to handle spatial relationships. This arrangement of block is repeated several times in U-Shaped architecture as shown in **Figure 6**. The second part is the expanding path, which UpSample the feature map from the lower level block using 2*2 Up Convolutional layer. This model differs from SegNet, by utilizing the entire feature maps instead of only pooling indices from corresponding contracting blocks and followed by the two 3*3 Convolutional blocks with ReLu activation function. The final layer is the 1*1 Convolutional layer so as to generate the required number of channels from the feature map. The output of this layer is the segmented image. This model propagates the contextual information throughout the network, which helps in more accurate segmentation results.

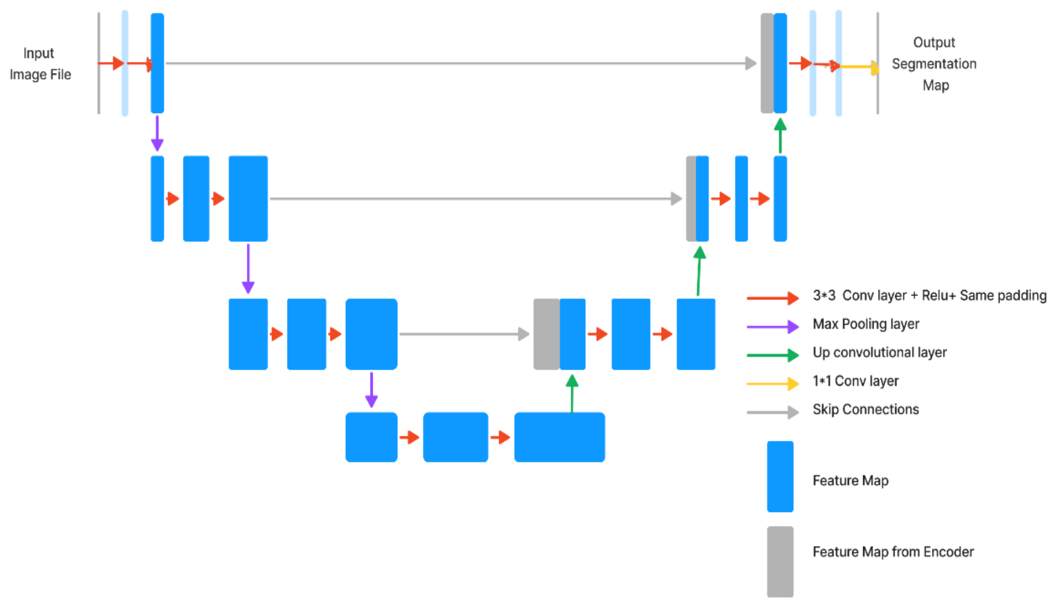

Figure 6. Base UNet architecture.

- *3D UNet*- In this model, all the 2D operations have been replaced with the 3D operations. This has enabled the segmentation of volumetric data. These models have been mainly used in medical image segmentation like CT scans and MRIs. It provides segmentation results of high accuracy.

- *Residual UNet*- The residual approach redefines layers as learning residual functions relative to their input, improving the flow of information inside a neural network. This method successfully tackles the issue of degradation in deeper networks. Multiple residual (ResNet) blocks, each with stacked layers like batch normalisation (BN), ReLu activation, and a weight layer (like a convolutional layer), make up a deep residual network. Shortcut connections (SC) are used to skip one or more layer. The SC pass through a convolutional block to maintain the dimensionality of the main convolutional block. The basic unit design of ResNet block is given in **Figure 7**. These residual blocks can be added in the UNet network in both contracting and expanding path by stacking at suitable position, depending on the application.

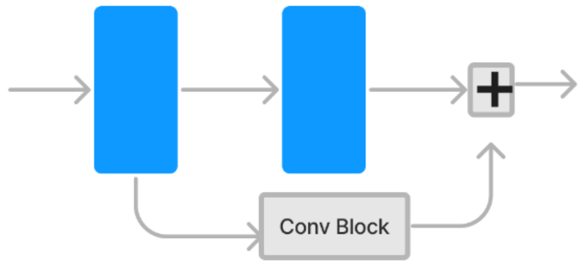

Figure 7. ResNet Block with 1*1 conv block.

- *Attention UNet*- This type of UNet model uses an Attention module to focus on specific important details while ignoring the irrelevant or unnecessary details. Attention Modules are capable of learning useful features information and neglect non-useful background information. These modules consist of attention gates that trims the irrelevant features before passing it to the next stage. The most popular attention module is the Convolutional Block Attention Module (CBAM). It consists of a Channel and Spatial attention map as displayed in **Figure 8**.

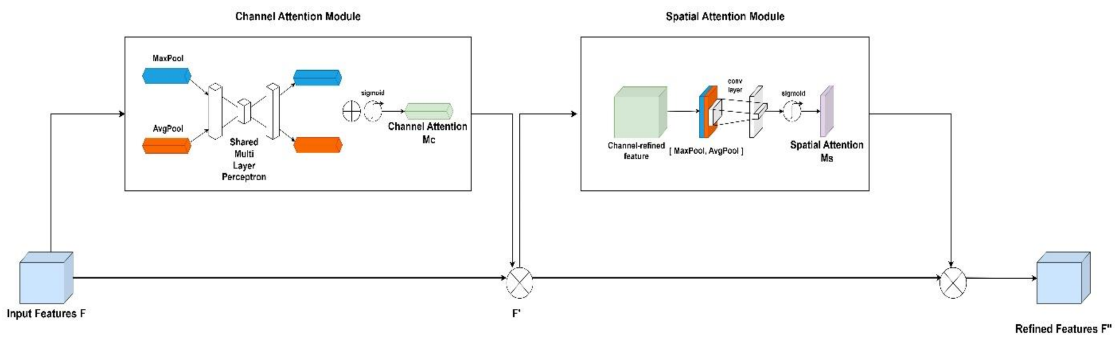

Figure 8. CBAM attention module.

The input to the CBAM module is the intermediate feature map F ∈C*H*W, where C represents the number of channels, H represents Height and W is the width. CBAM sequentially computes a one-dimensional channel attention map Mc ∈ C*1*1 and a two-dimensional spatial attention map Ms with dimension 1*H*W.

$$F' = Mc(F) \otimes F, \tag{1}$$

$$F'' = Ms(F') \otimes F' \tag{2}$$

Here, ⊗ denotes element-wise multiplication . F'' is the attention refined output.

Channel Attention Module consist of Max Pooling and Average Pooling layer to get two different spatial information $F^c_{max}$and F^c_{avg}respectively. These are passed to the Multi-Layer Perceptron with only one hidden layer of dimension C/r, where r is the reduction ratio to produce Mc ∈ C*1*1. This output of perceptron is passed to each spatial information, then both the output feature vectors are concatenated using element-wise summation. At last, the feature vector is given to the sigmoid function to filter out less important channels.

$$Mc(F) = \sigma\left(MLP\big(AvgPool(F)\big) + MLP\big(MaxPool(F)\big)\right) \tag{3}$$

$$= \sigma\left(W_1\left(W_0 F^c_{avg}\right)\right) + W_1(W_0(F^c_{max}))) \quad (4)$$

Here, σ is the Sigmoid Function, W_0 and $W_1$are the MLP weights, where $W_0 \in C/(rxC)$, and $W_1 \in CxC/r$.

Spatial Attention Module emphasizes on 'where' the most important information is present. It utilizes the inter-spatial relationship of features map. In this, first average pooling and max pooling are applied to the channel refined feature map F' and concatenated them to produce an efficient feature descriptor. This feature descriptor is passed to the small convolutional layer of 7*7 to create a 2D spatial attention map Ms (F') ∈ 1*H*W.

$$Ms(F') = \sigma(f^{7*7}([AvgPool(F); MaxPool(F)])) \quad (5)$$

$$= \sigma(f^{7*7}(\left[F^s_{avg}; F^s_{max}\right\})) \quad (6)$$

Here, f^{7*7} represents a convolution operation of kernel size 7*7, F^s_{avg} is the average pooled features and F^s_{max} is the max pooled feature.

These two attention modules can be arranged in both sequential and parallel manner. But, Woo et al. (2018) found that sequential arrangement gives better results than parallel one. After obtaining Mc and Ms, they are combined elementwise to produce the final attention map. Since Mc and Ms have different dimensions, resulting in a final map of C*H*W. This CBAM block can be added in the expanding path to focus on important features from the corresponding block in the contracting path. It significantly improves the performance of segmentation process without adding any excessive computational complexity.

- *UNet++* - It overcomes the semantic gap between the contacting path and extracting path by using convolution layers on the path of skip connections as displayed in **Figure 9**. It also improves the gradient flow by using dense skip connection network. This dense connection in skip pathway facilitates the capturing of multi-scale contextual information, thereby augmenting the model's capability to segment objects of varying size. Each block in the skip connection path at the same level receives feature maps from all the previous units. Deep supervision in the network helps in monitoring the performance of the model during training. This model has been given by Zhou et al. (2018).

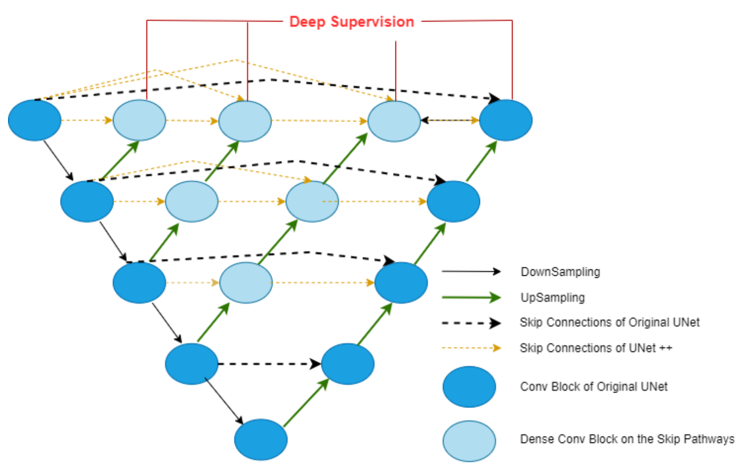

Figure 9. UNet++ architecture.

- *Deformable UNet*- Most of the input images given to the segmentation model have irregular shapes and sizes. Jin et al. (2019) integrated deformable convolution network as shown in **Figure 10** in the UNet architecture to handle geometric transformations. Some of the normal convolution layers in the encoder-decoder part of UNet have been replaced with the deformable convolution block. Each block consists of convolution offset, convolution layer, batch normalization followed by an activation function. The offset field and sampling locations of the deformable layer are adjusted according to the input data. Each deformable block is followed by a normal convolution block, as deformation adds computational cost to the model.

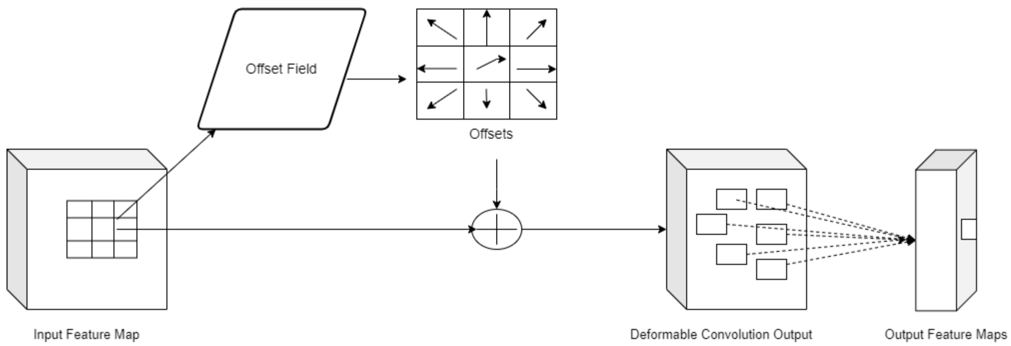

Figure 10. Illustration of deformable convolution.

The normal convolution operation can be represented as

$$y(m_0) = \sum_{m_i \in G} w(m_i).x(m_0 + m_i) \tag{7}$$

where, y is the output feature map for each location m_0, x is the input feature map, G represents the convolution grid, m_i represents the locations in G, w means the weights of samples value.
In deformable convolution, the formula changes to

$$y(m_0) = \sum_{m_i \in G} w(m_i).x(m_0 + m_i + \Delta m_i) \tag{8}$$

where, Δm_i is the value of the sampled points after migration, detemined by using bilinear interpolation.

- *UNet 3+*- UNet ++ has not been able to explore information at full scale despite of having bulky network. Therefore, Huang et al. (2020) redesigned UNet++ by changing the dense connection to the one shown in **Figure 11**. This full scale skip connection network capture the semantic information at different level. At each level of UNet, different types of information gets exploited, the lower levels feature map capture better spatial information and the higher levels capture relative positioning of objects better. The inter encoder-decoder skip connection passes the low-level detailed information while each of the inter- decoder skip connections helps in preserving high-level semantic information, leading to the more accurate segmentation result. UNet3+ uses lesser parameters than UNet++ and have better performance. UNet 3+ is known for its ability to accurately localize objects and produce coherent boundaries in intricate structures too. UNet 3+ model has been the popular choice for medical image segmentation.

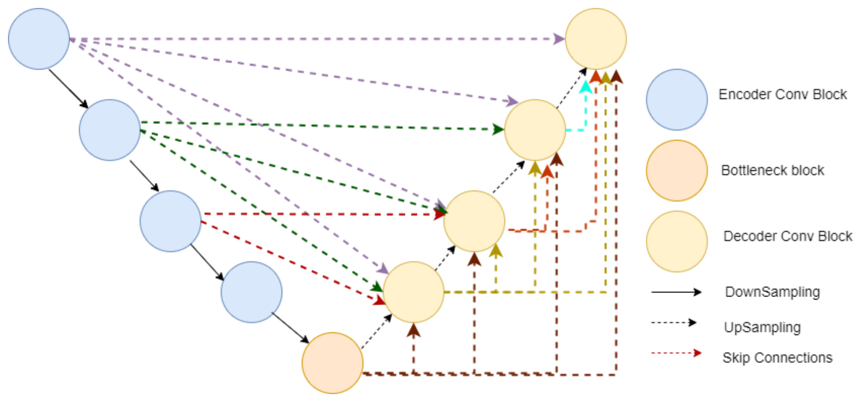

Figure 11. UNet3+ architecture.

Garcia-Nonoal et al. (2023) have used UNet to accurately identify and extract key features from retinal images. These features have then been combined with other image components to create a secure zero-watermarking code. This code has been used to perform Reversible Data Hiding in retinal images.

Gao et al. (2023) have used enhanced UNet architecture by reducing redundant skip connections, named as UNet3+ by Huang et al. (2020) with CBAM attention mechanism by Woo et al. (2018) to segment the image into ROI and NROI. The secret data is embedded in a reversible manner within the Region of Interest (ROI), using stretched ROI histogram. When the ROI reaches its embedding limit, any remaining secret data is then embedded into the Non-Region of Interest (NROI). Additionally, the side information necessary for image recovery is embedded into the NROI simultaneously.

Amrit et al. (2023) used customized UNet3+ architecture to segment the cover image into ROI and NROI. Doctor's signature and MAC address have been embedded in ROI region after encrypting it. Other information like the hospital logo and patient reports have been embedded in the NROI region. The watermark image has then been obtained by combining embedded ROI and embedded NROI images.

Recently, Shi et al. (2024) have constructed a new model named as J-Net which is the combination of UNet and AlexNet. In this model, a UNet model has been followed by the AlexNet with a bridge layer in between. Bridge layer helps in changing the dimension of the output tensor from the UNet to match the shape to the input tensor of the AlexNet. The ReLu activation function of AlexNet has been replaced with the Leaky ReLu to address the problem of 'Dying ReLu'. The ROI region has been encrypted using plain text encryption to prevent it from unauthorized access and skipping hiding strategy has been used to embed secret information into the NROI. Skipping hiding finds out the skipping position in the histogram of an image. The experimental results of this paper have been found to be significantly good and can be a break-through in the field of image segmentation based reversible data hiding.

4.4 Generative Adversarial Neural Models

Adversarial Neural Models consists of two networks as shown in **Figure 12**: a Generator and a Discriminator, which compete with each other in order to improve their performance. The generator takes the input and generate synthetic data samples and the discriminator, which is the binary classifier distinguishes between real data sample and fake data samples. During training, the generator tries to fool the discriminator by providing indistinguishable samples, while the discriminator tries its ability to differentiate between fake and real samples. GAN proved as a powerful framework to work with realistic data.

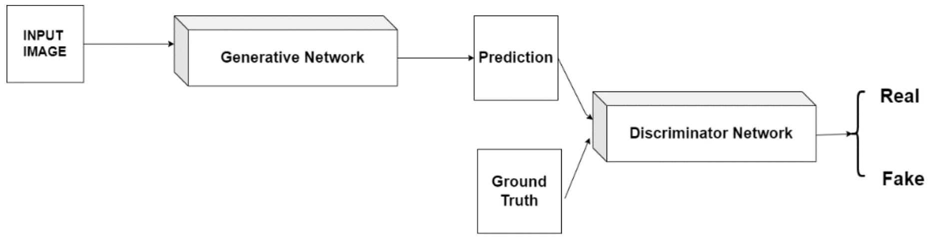

Figure 12. GAN block diagram.

These Generative and Discriminator network can be built on any suitable SOTA architecture depending on the application and performance. The Generator of the Adversarial UNet is based on the UNet design while the discriminator is made of standard CNN layers.

Luc et al. (2016) introduced an adversarial training method for semantic segmentation. This adversarial network has been tasked with distinguishing between ground-truth segmentation maps and those generated by the segmentation network. The study has indicated that the adversarial training does not succeed in increasing accuracy much on PASCAL VOC 2012 dataset. But after fine tuning the GAN models and changing their architecture, GAN improves the accuracy of image segmentation task.

Annadurai et al. (2023) used a convolutional GAN model as shown in **Figure 13**, for the segmentation and classification of MRI and CT watermarked images.

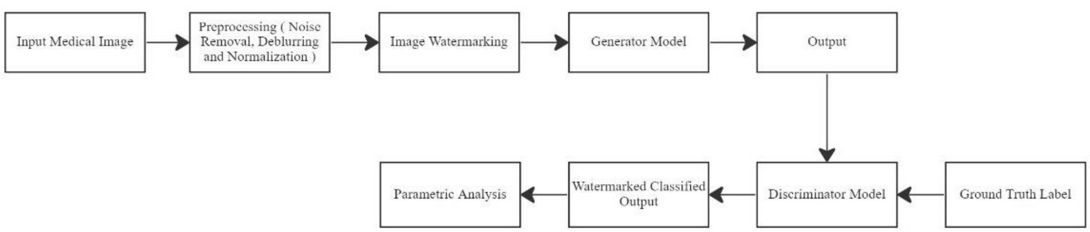

Figure 13. GAN based segmentation.

5. Results and Discussions

The literature review indicates that presently the deep learning techniques are highly used in Reversible Data Hiding (RDH) and hence, the author has compared the above mentioned deep learning techniques for image segmentation in **Table 1**.

The evaluation metrics used for comparing image segmentation technique are as follows:

1) Accuracy – It is defined as the total number of correctly predicted samples against the total number of samples, which is given by

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (9)$$

where, TP= True Positive, TN= True Negative, FP= False Positive and FN=False Negative.

2) Mean Intersection over Union (IOU) / Jaccard Index – It is average of the area of intersection between the predicted segmentation map A and the ground truth map B divided by the area of the union between the two maps over all classes.

$$IOU = J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{10}$$

Table 1. Comparison of deep learning based image segmentation techniques.

Technique	Cited paper authors (year of publication)	Dataset used	Accuracy (in %)	Mean IOU (in %)
FCN	Long et al. (2015)	PASCAL VOC (PASCAL Visual Object Classes Challenge)	90.3	62.7
FCN-AlexNet	Kaymak et al. (2020)	Skin Lesion dataset	94.81	-
Dilated Convolution	Chen et al. (2017)	PASCAL VOC 2012	-	79.70
SegNet	Badrinarayanan et al. (2017)	CamVid11(Cambridge-driving Labelled Video Database	-	60.10
Base UNet	Ronneberger et al. (2015)	ISBI cell Dataset	-	92.03
3D UNet	Çiçek et al. (2016)	Xenopus kidney Embryos	-	86.3
Residual UNet	Mostafiz et al. (2018)	DRIVE (Digital Retinal Images for Vessel Extraction) Dataset	98.18	-
Attention UNet	Arora et al. (2021))	ISIC 2018 Skin Lesion Dataset	95	83
UNet++	Zhou et al. (2018).	LiTS (Liver Tumor Segmentation) 2017	-	82.90
Deformable UNet	Jin et al. (2019)	DRIVE	95.66	95.66
UNet 3+	Huang et al. (2020)	LiTS 2017	-	95.50
GAN	Sarkar et al. (2021)	ISBI Cell Dataset	97.61	81.98

There is a growing need of image segmentation based on Reversible Data Hiding in certain application such as medical image analysis as discussed above. In health sector, most of the patient data and their diagnostic images are stored on the server, so it will be easy to manage this data if it gets embedded inside these images.

However, most of the medical images contain ROI, which is critical for diagnosis, and directly applying RDH may affect these areas. To resolve this issue, the mentioned image segmentation techniques in the literature review have been employed by various researchers to separate ROI and NROI, and then apply suitable RDH techniques. This would ensure the integrity of ROI part of the medical image. The following **Table 2** highlights the advantages, limitations and their performance for different RDH techniques.

The Performance Metrics used in Reversible Data Hiding are:

1) Max Embedding Rate – It is the maximum amount of data that can be embedded or hidden within a cover media while maintaining its acceptable perceptual quality.

$$Max\ Embedding\ Rate = \frac{total\ no\ of\ bits\ that\ can\ be\ embedded}{total\ no\ of\ pixels\ in\ an\ image}\ bpp \tag{11}$$

where, bpp=bits per pixel

2) Peak Signal to Noise Ratio (PSNR) - It measures the difference between the reconstructed image compared to the original image by calculating the amount of noise or distortion introduced during the reconstruction process.

$$PSNR = 10 * log_{10}\frac{MAX^2}{MSE}\ dB \tag{12}$$

where, MAX is the maximum possible pixel value of the image, MSE= Mean Squared Error, dB=Decibel.

Table 2. Advantages, limitations and performance evaluation of image segmentation techiques in RDH.

Image segmentation techniques	Advantages	Limitations	Reversible data hiding methods used	Performance metrics evaluated for reversible data hiding
Thresholding	Require low computational cost with less storage space.	1. Doesn't work well with uneven distribution of data. 2. Sensitive to noise	LSB substitution in Non-ROI and high-payload frequency based RDH in ROI part by Pai et al. (2011)	Max Embedding Rate= 0.5 bpp Mean PSNR value= 47.95 dB
Region Based	1. Used for images with good characteristics features, where it is easy to define similarity criteria 2. Noise Resistant	1. Different seed points may give different results 2. High computational cost	Histogram Shifting by Salman (2015)	PSNR value >=40dB
Edge Detection	1. Works well with images having prominent edges 2. The boundaries of images with high contrast can be preserved	1.Human intervention is required 2. not suitable for low contrast images and images with too many edges	Difference Expansion Technique by Lakshmanan et al. (2018)	Data Not Available
GrabCut	1. Iterative refinement improves segmentation results 2. Works well with images having complex background	Quality of segmentation result depends on the initial estimates provided by the user	Histogram Shifting by Wu et al. (2020)	Max Embedding Rate= 0.5 bpp PSNR value= 33.04 dB, In GrabCut row under performance metrics evaluated for Reversible Data Hiding column
SVM	1. Can work with complex boundaries images 2. SVMs are inherently robust to noise and outliers in data	Struggle with imbalanced dataset, where one class is superior to others, leading to suboptimal segmentation results for minority class.	Discrete Wavelet Transform – Singular Value Decomposition by Rai & Singh (2017)	PSNR value= 52.18 dB
			Integer Wavelet Transform by Chowdhuri et al. (2023)	Max Embedding Rate= 0.44 bpp PSNR value = 64 dB
Clustering	Fuzzy based clustering is better than k-means clustering as it does not use distance as the evaluation metric. It uses membership function, which is more suitable for real world problems.	1. Determining membership function is difficult 2. It can be affected by noise	Discrete Wavelet Transform by Balasamy et al. (2021)	Max Embedding Rate= 1.25 bpp PSNR value=49.5 dB
Fully Convolutional Network	1. It preserves spatial information throughout the network. 2. Can operate on large datasets	1. Overfitting in the model 2. Huge Training data and time is required 3. Can Suffer from vanishing gradient problem	Least Significant Bit substitution by Meng et al. (2020)	PSNR value >=50dB
			Discrete Cosine Transform by Saidi et al. (2024)	Max Embedding Rate= 0.50 bpp PSNR value= 115.53 dB
Encoder-Decoder Based Model	1. Model can learn complex features. 2. Incorporate skip connections that preserves fine grained details by combining low-level encoder features with high-level decoder features	1. Computational cost of these models is high. 2. Difficulty in long range dependencies.	Histogram Shifting by Gao et al. (2023) Max Embedding Rate= 0.98 bpp	Mean PSNR value= 34.24 dB
			Skipping Hiding by Shi et al. (2024)	Embedding Capacity= 0.6 bpp PSNR value= 43dB
Generative Adversarial Network	1. Improve segmentation quality by using the discriminator, which evaluates the ground truth segmentation maps with the one produced by the generator 2. Robust to noise due to its generative nature.	1. High computational resource requirement 2. Hyper tuning of model parameters is difficult.	Discrete Wavelet Transform by Amrit el al. (2023)	Mean PSNR value=60.8dB

It would be worth to mention here that the deep learning based segmented images exhibit higher PSNR value compared to those generated using traditional and machine-based techniques.

Image segmentation helps in Reversible Data Hiding to find out the regions within an image for selective embedding. To further reduce the distortion, the local complexity of ROI pixels can be calculated to find out the smooth pixels for embedding, so as to minimize the impact of these modifications. However, this area needs some more exploration to find out its effectiveness and potential in the field of Image segmentation based RDH.

Efforts can be made to combine deep learning method with traditional methods to improve the segmentation performance. CNN in combination with probabilistic graphical methods like CRFs have been explored but its integration with other method needs further studies. Deep learning models can be combined like JNet mentioned earlier, to further increase the segmentation accuracy of the model. Integrating deep learning architectures like VGG (Simonya and Zisserman, 2014), InceptionNet (Szegedy et al., 2015), ResNet (He et al., 2016), DenseNet (Huang et al., 2017), EfficientNet (Tan and Le, 2019) with the different SOTA models for image segmentation like UNET versions, DeepLab, mask R-CNN can offer more advancement in the field of image segmentation, leading to more robust segmentation model for wide range of applications. Ensemble techniques like boosting and bagging can be used by combining the outputs of multiple segmentation model to further explore the performance.

Most of the authors have used image segmentation based RDH technique for medical images. Efforts can also be made to explore this for other applications, especially in the domains of high security such as forensic analysis, cryptographic signatures within images, legal proceedings etc.

6. Conclusions

The comprehensive survey has provided valuable insights into various segmentation techniques and how these can be used in the field of data hiding. The study has revealed that image segmentation is being used as the fundamental step in Reversible Data Hiding. It minimizes visual distortions and maximizes embedding capacity by delineating ROIs and NROIs within an image. The survey has also highlighted the improvements that can be made on image segmentation techniques. Overall, the insights gained from this survey will serve as a valuable foundation for advancing the field of reversible data hiding and its integration with image segmentation techniques in real-world applications.

Conflict of Interest

There is no conflict of interest with anyone, regarding authorship and sources of data and information in the paper.

Acknowledgments

The author would like to express sincere thanks and gratitude to the research supervisor Dr. Rajeev Kumar, Faculty of Computer Science and Engineering, Delhi Technological University, Delhi for his kind guidance and support in writing the manuscript. The author would also like acknowledge and appreciate the reviewers of the paper for their kind comments and suggestions to improve its quality and usability.

References

Amrit, P., Singh, K.N., Baranwal, N., Singh, A.K., Singh, J.P., & Zhou, H. (2023). Deep learning-based segmentation for medical data hiding with Galois field. *Neural Computing and Applications*, 1-16. https://doi.org/10.1007/s00521-023-09151-2.

Annadurai, C., Nelson, I., Devi, K.N., Manikandan, R., & Gandomi, A.H. (2023). Image watermarking based data hiding by discrete wavelet transform quantization model with convolutional generative adversarial architectures. *Applied Sciences*, *13*(2), 804. https://doi.org/10.3390/app13020804.

Arora, R., Raman, B., Nayyar, K., & Awasthi, R. (2021). Automated skin lesion segmentation using attention-based deep convolutional neural network. *Biomedical Signal Processing and Control*, *65*, 102358.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481-2495.

Balasamy, K., &Suganyadevi, S. (2021). A fuzzy based ROI selection for encryption and watermarking in medical image using DWT and SVD. *Multimedia Tools and Applications*, *80*(5), 7167-7186.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-8*(6), 679-698.

Chaithanya, P. & Srujana V. (2022). Reversible image data hiding using quad tree segmentation. *International Journal of Emerging Technologies and Innovation Research*, (9), d521-d523.

Chen, B., & Wornell, G.W. (2001). Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information tTheory*, *47*(4), 1423-1443.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A.L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834-848.

Chowdhuri, P., Pal, P., & Si, T. (2023). A novel steganographic technique for medical image using SVM and IWT. *Multimedia Tools and Applications*, *82*(13), 20497-20516.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., &Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Proceedings, Part II 19* (pp. 424-432). Springer International Publishing. Athens, Greece.

Forsyth, D.A., & Ponce, J. (2002). *Computer vision: a modern approach*. Prentice Hall professional technical reference.

Gao, G., Wan, X., Yao, S., Cui, Z., Zhou, C., & Sun, X. (2017). Reversible data hiding with contrast enhancement and tamper localization for medical images. *Information Sciences*, *385*, 250-265.

Gao, G., Zhang, H., Xia, Z., Luo, X., & Shi, Y.Q. (2023). Reversible data hiding-based contrast enhancement with multi-group stretching for ROI of medical image. *IEEE Transactions on Multimedia*, *26*, 3909-3923.

Garcia-Nonoal, Z., Mata-Mendoza, D., Cedillo-Hernandez, M., & Nakano-Miyatake, M. (2024). Secure management of retinal imaging based on deep learning, zero-watermarking and reversible data hiding. *The Visual Computer*, *40*(1), 245-260.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K.Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700-4708).

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W., & Wu, J. (2020). Unet 3+: A full-scale connected unet for medical image segmentation. In *IEEE international conference on acoustics, speech and signal processing* (pp. 1055-1059). IEEE. Barcelona, Spain.

Ji, Z., Xia, Y., Sun, Q., Cao, G., & Chen, Q. (2015). Active contours driven by local likelihood image fitting energy for image segmentation. *Information Sciences*, *301*, 285-304.

Jiang, X., Tang, Z., Ou, B., & Xiong, J. (2022). Dynamic reversible data hiding for edge contrast enhancement of medical image. In *2022 International Conference on Automation, Robotics and Computer Engineering* (pp. 1-6). IEEE. Wuhan, China.

Jin, Q., Meng, Z., Pham, T.D., Chen, Q., Wei, L., & Su, R. (2019). DUNet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, *178*, 149-162.

Kanopoulos, N., Vasanthavada, N., & Baker, R. L. (1988). Design of an image edge detection filter using the Sobel operator. *IEEE Journal of Solid-State Circuits*, *23*(2), 358-367.

Kaur, A., & Josan, H.S. (2014). Data hiding in medical images by preserving integrity of ROI using semi-reversible watermarking technique. *International Journal of Engineering Research and Applications*, *4*, 84-88.

Kaymak, R., Kaymak, C., &Ucar, A. (2020). Skin lesion segmentation using fully convolutional networks: A comparative experimental study. *Expert Systems with Applications*, *161*, 113742.

Kwon, D. (2007). A image segmentation method based on improved watershed algorithm and region merging. *IEEE Transactions on Circuits and Systems for Video Technology*, *17*, 517-529.

Lakshmanan, S., & Rani, M. (2018). Reversible data hiding in medical images using edge detection and difference expansion technique. *Journal of Computational and Theoretical Nanoscience*, *15*(6-7), 2400-2404.

Li, M., Wang, L., & Fan, H. (2020). Privacy-preserved data hiding using compressive sensing and fuzzy C-means clustering. *International Journal of Distributed Sensor Networks*, *16*(2), https://doi.org/10.1177/1550147720908748.

Lin, Y.C., & Li, T.S. (2011). Reversible image data hiding using quad-tree segmentation and histogram shifting. *Journal of Multimedia*, *6*(4), 349-358.

Liu, Y.T., Zhang, H.X., & Li, P.H. (2011). Research on SVM-based MRI image segmentation. *The Journal of China Universities of Posts and Telecommunications*, *18*, 129-132.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3431-3440).

Luc, P., Couprie, C., Chintala, S., & Verbeek, J. (2016). Semantic segmentation using adversarial networks. *arXiv preprint arXiv:1611.08408*.

Luo, Y., Qin, J., Xiang, X., Tan, Y., He, Z., & Xiong, N.N. (2020). Coverless image steganography based on image segmentation. *Computers, Materials and Continua*, *64*(2), 1281-1295.

Mahmood, N., Shah. A., Waqas, A., ABUBAKAR, A., KAMRAN, S., ZAIDI, S.B. (2015). Image segmentation methods and edge detection: An application to knee joint articular cartilage edge detection. *Journal of Theoretical and Applied Information Technology*, *71*(1), 87-96.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*(1167), 187-217.

Meng, R., Cui, Q., Zhoul, Z., Yuan, C., & Sun, X. (2020). A novel steganography algorithm based on instance segmentation. *Computers, Materials & Continua*, *63*(1), 183-196.

Mostafiz, T., Jarin, I., Fattah, S. A., & Shahnaz, C. (2018). Retinal blood vessel segmentation using residual block incorporated U-Net architecture and fuzzy inference system. In *2018 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)* (pp. 106-109). IEEE. Chonburi, Thailand.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62-66.

Pai, P.Y., Chang, C.C., Chan, Y.K., & Liu, C.M. (2012). An ROI-based medical image hiding method. *International Journal of Innovative Computing, Information and Control*, *1*(8), 4521-4533.

Pai, P.Y., Chang, C.C., Chan, Y.K., & Tsai, M.H. (2011). An adaptable threshold detector. *Information Sciences*, *181*(8), 1463-1483.

Pan, N., Qin, J., Tan, Y., Xiang, X., & Hou, G. (2020). A video coverless information hiding algorithm based on semantic segmentation. *EURASIP Journal on Image and Video Processing*, *2020*, 1-18. https://doi.org/10.1186/s13640-020-00512-8.

Parah, S.A., Ahad, F., Sheikh, J.A., Loan, N.A., & Bhat, G.M. (2017). A new reversible and high capacity data hiding technique for E-healthcare applications. *Multimedia Tools and Applications*, *76*, 3943-3975.

Phadikar, A., & Maity, S.P. (2010). Roi based error concealment of compressed object based image using QIM data hiding and wavelet transform. *IEEE Transactions on Consumer Electronics*, *56*(2), 971-979.

Rai, A., & Singh, H.V. (2017). SVM based robust watermarking for enhanced medical image security. *Multimedia Tools and Applications*, *76*(18), 18605-18618.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Proceedings, part III 18* (pp. 234-241). Springer International Publishing, Munich, Germany.

Rother, C., Kolmogorov, V., & Blake, A. (2004). " GrabCut" interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)*, 23(3), 309-314.

Saidi, H., Tibermacine, O., &Elhadad, A. (2024). High-capacity data hiding for medical images based on the mask-RCNN model. *Scientific Reports*, *14*(1), 7166.

Salman, A.G. (2015). Reversible data hiding technique on jpeg image by quad-tree segmentation and histogram shifting method based on android. *Procedia Computer Science*, *59*, 530-539.

Sarker, M.M.K., Rashwan, H.A., Akram, F., Singh, V.K., Banu, S.F., Chowdhury, F.U., Choudhury, K.A., Chambon, S., Radeva, P., Puig, D., & Abdel-Nasser, M. (2021). SLSNet: Skin lesion segmentation using a lightweight generative adversarial network. *Expert Systems with Applications*, *183*, 115433.

Shen, J., & Castan, S. (1992). An optimal linear operator for step edge detection. *CVGIP: Graphical models and Image Processing*, *54*(2), 112-133.

Shi, H., Zhou, Z., Qin, J., Sun, H., & Ren, Y. (2024). A separable privacy-preserving technique based on reversible medical data hiding in plaintext encrypted images using neural network. *Multimedia Tools and Applications*, 1-26. https://doi.org/10.1007/s11042-024-18600-6.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Stockman, G., & Shapiro, L.G. (2001). *Computer vision*. Prentice Hall PTR.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9).

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105-6114). Long beach, California.

Woo, S., Park, J., Lee, J.Y., & Kweon, I.S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision* (pp. 3-19).

Wu, H. T., Huang, Q., Cheung, Y. M., Xu, L., & Tang, S. (2020). Reversible contrast enhancement for medical images with background segmentation. *IET Image Processing*, *14*(2), 327-336.

Yang, Y., Zhang, W., Liang, D., & Yu, N. (2018). A ROI-based high capacity reversible data hiding scheme with contrast enhancement for medical images. *Multimedia Tools and Applications*, *77*, 18043-18065.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 3-11). Springer International Publishing, Granada, Spain.