ITM Web of Conferences **73**, 02030 (2025) https://doi.org/10.1051/itmconf/20257302030
IWADI 2024

Trends and Techniques in Medical Image Segmentation for Disease Detection

Xinli Jiang*

School of Computer Science & Technology, Beijing Institute of Technology, 102401, Beijing, China

Abstract. Medical images have become an indispensable and important tool for the diagnosis of medical conditions and surgical guidance. As computer vision technology advances, Medical image segmentation technology has effectively assisted clinicians in making accurate diagnoses and providing personalized treatment. In this paper, some excellent medical image segmentation methods in recent years are summarized, and according to the deep learning method (e.g. Convolutional Neural Network (CNN), U-net, etc.), and traditional methods (such as active contour model, threshold segmentation model, etc.) are sorted out. This paper compares various image segmentation methods, analyzes their similarities and differences, and summarizes and looks forward to the future development of medical image segmentation technology. With the continuous advancement of computer vision models, medical image segmentation is expected to become increasingly accurate and efficient. This will significantly enhance the speed and accuracy of medical image processing, helping doctors to better identify and analyze diseases, thereby providing more accurate clinical diagnoses and treatment plans. With these technological advancements, future medical image segmentation will not only handle more complex images but also enable more intelligent and automated analysis, offering strong support for clinical practice.

1 Introduction

In order to offer a solid foundation for clinical diagnosis, treatment planning, and pathology research, the primary goal of medical image segmentation is to precisely distinguish areas of exceptional clinical significance (also known as regions of interest) from the background regions of medical pictures [1]. Computed tomography (CT), ultrasound, and other medical imaging devices are widely used due to the quick development of medical imaging technology. Medical images are now a crucial tool to help doctors diagnose diseases and make treatment decisions [1]. Medical image segmentation can offer a scientific foundation for further quantification of lesion areas, planning of precise treatment plans, and assessment of surgical risks through accurate extraction of key tissues in the source image. Therefore, the accuracy of medical image segmentation directly impacts the physicians' subjective assessment and judgment in the diagnostic process, including the development of

* Corresponding author: 1120223441@bit.edu.cn

preoperative treatment plans, intraoperative real-time monitoring, postoperative evaluation and other aspects.

Moreover, segmenting medical images is equally significant for pathology research. It can not only help researchers better understand the pathological mechanisms of diseases but also improve the reliability and reproducibility of research data, providing strong support for not only early diagnosis but also personalized treatment of diseases. In the research of complex diseases such as cancer, brain diseases, cardiovascular diseases and so on, accurate medical image segmentation technology has become an indispensable tool, that plays a crucial role in the quantitative analysis of the lesion area and the evaluation of the treatment effect. As a result, one of the key areas of current study in computer vision and medical imaging is how to increase the precision and effectiveness of medical image segmentation.

Many excellent image segmentation models and architectures have been proposed in recent times, and this paper divides them into two categories: medical image segmentation based on traditional methods and medical image segmentation based on deep learning methods. Traditional methods are mainly based on image pixel segmentation, information entropy polarization, threshold setting, etc. as the theoretical basis. Deep learning methods are based on neural networks, with generative networks, convolutional networks, attention mechanisms, etc. as the theoretical basis. Finally, a summary is made based on the techniques mentioned in the paper, and an outlook is made on the future research trend of medical image segmentation to provide corresponding references for future scholars in this research field.

2 Medical Image Segmentation Methods by Traditional Methods

In future research of image segmentation methods, multi-scale feature fusion can be considered to be combined with traditional medical image segmentation methods. By extracting features at different scales and using the pyramid model for feature fusion, the details and global structure in the medical image can be better captured, which in turn reduces the effects of grayscale inhomogeneity and noise in the medical image, and achieves a more accurate segmentation effect.

2.1 Threshold Segmentation

Threshold segmentation is an image segmentation method based on statistics, the basic idea is to consider the gray value of each pixel in the image as a pair of random variables following a specific probability distribution, transformed and added to the noise, or based on histogram analysis, edge information, texture and other image features to establish a threshold model. The method distinguishes different objects or regions in an image by setting one or more thresholds. Among them, the background is usually represented by a gray level of 255, and the output is white, and the target is represented by a gray level of 0, and the output is black. The thresholding models used vary for different medical application scenarios. For example, the Otsu method, as the most widely used thresholding method in the image processing field, is characterized by simplicity, high efficiency, and low parameter count, and has been practically applied in binarization of cellular images, X-ray images, and CT scan images. And when dealing with images with uneven brightness (e.g., pathology slice images), the Sauvola method takes into account the statistical information of the brightness of the local region in its model and thus performs superiorly in images with large variations of illumination within the local region. However, the limitation of the threshold segmentation method is that in the face of low-contrast scenes, backgrounds with similarities complicate the thresholding operation, and noise can seriously affect the selection of thresholds, leading to inaccurate segmentation.

2.2 Active Contour Model

The active contour model is an image segmentation method based on the principle of energy minimization. The basic idea is to combine the internal force of the curve itself and the multi-gradient information in the image as the external force to drive the active contour curve to evolve iteratively. When the energy function reaches a minimal value, the location where the active contour curve is located is considered the target boundary. However, the selection of parameters in the active contour model is more dependent on the personal experience of the researchers, cannot be determined by the model adaptively, and is more sensitive to noise, etc.

Aiming at these problems, Deng et al. proposed an adaptive weighting operator based on the difference between the variance of the local grayscale inside and outside the contour, which solves the problem of the LBF model's sensitivity to the parameters of the energy term, and they added both the mean and variance of the local grayscale to the energy square connotation to control the energy weighting inside and outside the curve, which accelerates the curve's convergence speed. Morphological operators and local similarity factors inside and outside the contour domain are introduced to overcome the negative effects of gray scale inhomogeneity and noise in medical images, and accurate segmentation is achieved on parotid gland images [2]. Tamboli et al. on the other hand, proposed an optimized active contour model, which is optimized by fine-tuning the weighting factors and the maximum number of iterations of the active contour model for the bee mating algorithm, and the regions of interest and non-interested regions using ISPIHT algorithm and DCT model for feature extraction respectively, and fusion coding to reduce the number of bits required for transmission and to reduce the noise effect [3].

2.3 Edge Detection Method

Image segmentation based on edge detection generates target regions by detecting discrete points in the image with strong gray scale changes, according to certain similarity principles or graph theory algorithms. Since the gray value of the boundary between different regions changes more drastically, the gradient partial derivative information of the edge detection function is usually used to determine the location of the edge points. The first or second order partial derivative detection methods include the Roberts gradient operator, Wills operator, Prewitt operator, Canny operator and Hessian operator. However, for complex medical images, especially those containing multiple tissues, edge detection may not be effective in capturing the boundaries of all tissues. In addition, the high sensitivity and poor continuity of edge detection may lead to discontinuous edge points in the segmentation results, thus affecting the accuracy of segmentation.

For this reason, Soheila et al. improved issues such as sensitivity to weak edges by smoothing 3x3 pixel blocks, feature extraction and elongating the vectors [4]. Fatimah Al-Hafiz et al. proposed a red blood cell segmentation model optimized for edge detection by combining the Canny operator and morphological operator red blood cell contouring and eliminating the interference of leukocytes [5]. The region of interest is enhanced by an expansion operation and a median filter is induced. Additionally separated the overlapping erythrocytes.

3 Medical Image Segmentation Methods Based on Deep Learning Approaches

3.1 CNN-based Segmentation Framework

CNN constructs a deep network by stacking multiple convolutional and pooling layers, and each convolutional layer can extract features (e.g., edges, texture, etc.) at different levels of the image. The input information is propagated forward from the input layer to the output layer along the deep network, after which the gradient of the computational error is gradually back-propagated from the output layer to the input layer, which in turn redistributes the corresponding weights.

Jiang et al. developed a dual path complementary (MC-DC) network based on MLP-CNN, which presents dynamic cyclic focus loss DCLF to solve the problem of focus category imbalance, improves the accuracy and sensitivity of segmentation, and optimizes the Multi Headed Compressible Self-Attention (MHSA) mechanism to efficiently fuse multilevel features from the MLP and the CNN, and at the same time reduces the amount of computation [6]. A Dice coefficient of 91.69% and an average surface distance (ASSD) of 9.52 mm are achieved on the ISIC2018 dataset. Dice coefficients of 91.6% and 94.4% were achieved on the Kasir-SEG dataset and CVC-ClinicDB dataset, respectively. Zhang et al. proposed a medical image segmentation framework (TC-Net) that utilizes local perception and interdependence over vast distances in medical imaging [7]. On IDRiD and DDR datasets, TC-Net achieved scores of 0.6985 and 0.5171 in terms of average pixel accuracy, respectively. In addition, on the skin image database, TC-Net achieved an average pixel accuracy of 0.886. Alqazzaz et al. proposed a SegNet-based automatic brain tumor segmentation model. Segmentation was performed by training four SegNet models combining their feature maps and original image pixel values. The decision tree classifier used in this segmentation model has higher accuracy than single modality MR image segmentation and forms the reference feature vector needed for segmentation through feature fusion [8].

3.2 U-Net Based Segmentation Framework

U-NeT is a network model with a U-shaped symmetric structure, which mainly consists of an encoder-decoder structure. U-Net fuses multi-scale features through jump connections, combines contextual information and speeds up training, efficiently maintaining detailed information and spatial data, thus improving network performance [9]. Zhang et al. based on the U-Net model proposed Dense Inception U-Net (DIU-Net) medical image segmentation model and combined DenseNet and GoogLeNet in the U-Net architecture of the model. By using convolution kernels of different sizes and residuals as the backbone structure, DIU-Net performs multilevel feature extraction and fusion on images [10]. The model achieves 95.8% Dice coefficient and 96.6% accuracy in a mixed dataset (DRIVE2 + STARE3 + CHASE_DB14). Despite the strong segmentation and scalability of the U-Net model, its lack of long-range modeling capability due to the limitation of its local connectivity leads to its limitation in effectively learning multi-scale information.

3.3 Segmentation Framework Based on Transformer

Attention mechanism support to establish transformer module. In recent years, Transformers has made great achievements in the field of natural language processing, which in turn has given rise to the use of Transformers for computer vision models [11, 12]. With its powerful attention mechanism, Transformer excels in extracting global information. However, the Transformer is deficient in learning local information. To solve this problem, Liu et al. proposed a polyp segmentation algorithm that fuses Transformer and convolution. The algorithm utilizes a Transformer to extract image global features and introduces convolutional operation to enhance the network's ability to process polyp details, thus refining the boundary segmentation effect [13]. Finally, the features extracted by Transformer and convolution are deeply fused to achieve feature complementarity. The similarity coefficients of this algorithm on CVC-ClinicDB and Kvasir-SEG datasets are 95.4% and 93.2%, respectively, and the average intersection and merger ratios are 91.3% and 88.6%, respectively. Fu et al. designed a CNN- and Transformer-based multiscale network Hierarchical Multi-Scale U- Net (HmsU-Net) [14]. This network further improves the accuracy and effectiveness of image segmentation by introducing a multiscale feature fusion (MFF) module that fuses features from parallel networks of CNN and Transformer components.

4 Dataset and Evaluation Criteria

4.1 Medical Image Segmentation Datasets

In the field of medical image segmentation, several public datasets have been created to promote the research and development of related algorithms. These datasets provide researchers with rich samples and annotations and become important tools for training and evaluating various medical image segmentation models. The following are three classical datasets widely used for medical image segmentation: the CVC-ClinicDB, BraTS and ISIC datasets. They cover different medical fields such as colorectal polyps, gliomas and skin lesions, respectively, and provide valuable data support for various image segmentation tasks. CVC-ClinicDB is a colorectal polyp dataset released in 2015 by the Polyp Segmentation Challenge organized. The dataset contains 612 images from 31 colonoscopy sequences and is primarily used to evaluate and enhance polyp segmentation techniques in colonoscopy images. Due to its wide range of applications, CVC-ClinicDB has become an important standard dataset in the field of colon polyp segmentation, facilitating the development of computer-aided diagnostic techniques. Brain Tumor Segmentation (BraTS) is a large-scale brain multimodal MRI glioma segmentation dataset containing 8,160 MRI scans from 2,040 patients. The BraTS dataset provides researchers with different types of glioma cases, helping to develop and evaluate deep learning for glioma segmentation and diagnostic models. The multimodal nature of the dataset allows researchers to explore the fusion of different imaging techniques for more accurate brain tumor segmentation and classification. The International Skin Imaging Collaboration (ISIC) dataset is published by the International Skin Imaging Collaboration to advance research in automated diagnosis of skin cancer. The ISIC dataset contains 2,000 dermoscopic images for training and their true segmentation (2,000 binary masked images), which include multiple types of skin lesions, such as melanoma. The ISIC dataset is widely used in the development of algorithms for skin lesion segmentation and skin cancer detection and has become one of the key benchmark datasets in the field.

4.2 Evaluation Criteria

This paper summarizes several quantitative evaluation metrics commonly used in medical image segmentation tasks, including Similarity Coefficient (Dice), Intersection and Union Ratio (IoU), Precision, Sensitivity and Specificity:

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{1}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

Where TP refers to the number of samples correctly segmented by the model; FP refers to the number of samples incorrectly segmented by the model; FN refers to the number of samples incorrectly segmented by the model as non-target regions (i.e., background or non-diseased areas), and TN refers to the number of samples where the model accurately identifies image regions as non-target areas. Higher values of Dice, IoU, precision rate, and sensitivity indicate better segmentation performance.

5 Conclusion

This study organized and classified recent medical image segmentation approaches based on their underlying principles. This paper examined both traditional and deep learning-based techniques in the field of medical imaging. These methods are crucial for assisting doctors in developing effective treatment plans for patients. The ability to automatically, quickly, and accurately segment images is vital for outlining target regions in various types of medical images. Utilizing computer models not only enhances the accuracy of image segmentation but also improves efficiency and conserves human resources. Currently, deep learning-based models have shown significant progress and practical applications in the realm of medical image segmentation, making them a popular research area moving forward. However, challenges such as insufficient hardware capabilities and uneven computational resource distribution still hinder the accuracy and stability of these models. As science and technology continue to advance, it's reasonable to expect that more computational methods tailored to clinical needs with practical value will emerge in the future.

References

1. Y. Q. Song, Digital Medical Imaging (Tsinghua University Press, Beijing, 2008)
2. X. Deng, T. Lan, M. Zhang, et al., Fast adaptive active contour model for segmenting parotid duct based on local gray level differences, J. South China Univ. of Tech., **12** (2018)
3. S. S. Tamboli, R. Butta, T. S. Jadhav, et al., Optimized active contour segmentation model for medical image compression, Biomed. Signal Process. Control, **80**, 104244 (2023)
4. S. E. Soheila, E. Zahra, A robust edge detection technique based on Matching Pursuit algorithm for natural and medical images, Biomed. Eng. Adv. **4**, 100052.0992 (2022)
5. F. Al-Hafiz, S. Al-Megren, H. Kurdi, Red blood cell segmentation by thresholding and Canny detector, Procedia Comput. Sci. **141**, 327-334 (2018)

6. X. Jiang, Y. Zhu, Y. Liu, et al., MC-DC: an MLP-CNN based dual-path complementary network for medical image segmentation, Comput. Methods Programs Biomed. **242**, 107846 (2023)
7. Z. Zhang, G. Sun, K. Zheng, et al., TC-Net: A joint learning framework based on CNN and vision transformer for multi-lesion medical image segmentation, Comput. Biol. Med. **161**, 106967 (2023)
8. S. Alqazzaz, X. F. Sun, X. Yang, et al., Automated brain tumor segmentation on multi-modal MR image using SegNet, Comput. Vis. Media. **5**(2), 209-219 (2019)
9. T. Zhou, Y. Dong, B. Huo, et al., Review of U-net network applications in medical image segmentation, J. Chin. Soc. Image Graph. **26**(9), 2058-2077 (2021)
10. Z. Zhang, C. Wu, S. Coleman, et al., DENSE-Inception U-net for medical image segmentation, Comput. Methods Programs Biomed. **192**, 105395 (2020)
11. Q. Jin, H. Hou, G. Zhang, et al., FEGNet: A feedback enhancement gate network for automatic polyp segmentation, IEEE J. Biomed. Health Inform. **27**(7), 3420-3430 (2023)
12. A. Vaswani, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017)
13. H. B. Liu, D. Gu, Fusion of Transformer and convolution for colorectal polyp segmentation algorithm, J. Chin. Med. Phys., **41**(3), 316-322 (2024)
14. B. Fu, Y. Peng, J. He, et al., HmsU-Net: A hybrid multi-scale U-net based on CNN and transformer for medical image segmentation, Comput. Biol. Med., **170**, 108013 (2024)