Abstract: Medical image segmentation is a core task in the field of medical image processing. It aims to separate areas of interest such as organs, tissues, and lesions from the background in medical images to facilitate further analysis, diagnosis, and treatment planning. Accurate image segmentation is crucial for improving the accuracy of disease diagnosis, assessing disease progression, and developing personalized treatment plans. However, fully supervised segmentation methods face the challenge of high annotation costs. With the emergence of the U-Net architecture, semi-supervised medical image segmentation based on U-Net is receiving increasing attention. This article reviews semi-supervised segmentation methods, the design concept and structure of the U-Net network, how it has been extended, and its application in semi-supervised medical image segmentation. The article also identifies the challenges faced by semi-supervised medical image segmentation techniques based on U-Net and speculates on possible future research directions. In conclusion, the article summarizes the potential of semi-supervised medical image segmentation technology based on U-Net as an accurate and efficient tool for medical diagnosis and treatment.

Keywords: Semi-Supervised, U-Net, Medical Image Segmentation.

1. Introduction

Image segmentation technology plays a crucial role in the medical field, enabling doctors to precisely identify and quantitatively analyze various anatomical structures and pathological states from complex medical images (Chaudhury et al., 2022; Ryalat et al., 2023). This technology significantly enhances the accuracy of diagnosis and treatment and greatly optimizes the medical workflow, from diagnostic support to treatment planning and disease monitoring (Saïd, Alsheikhy, Shawly, & Lahza, 2023). Although fully supervised learning methods have achieved remarkable success in the field of medical image segmentation, these methods heavily depend on a large amount of manually annotated data. Acquiring such data is both time-consuming and expensive, especially in specialized medical fields requiring deep expertise.

Semi-supervised learning, as a method that utilizes a large amount of unlabeled data along with a small amount of labeled data, offers a new perspective for addressing the data scarcity problem in medical image segmentation. This approach not only challenges the traditional reliance on extensively annotated data but also opens up an effective pathway for high-precision segmentation using unlabeled medical images. With the rapid development of deep learning technologies, semi-supervised medical image segmentation methods based on deep learning have quickly risen to prominence (Jiao et al., 2023). By incorporating innovative network architectures, loss functions, and training strategies, these methods effectively utilize the rich information contained in unlabeled data, thereby achieving remarkable performance in various medical image segmentation tasks (Hu, Zeng, Xu, & Shi, 2021; Luo, Chen, Song, & Wang, 2021; Zhang, Zhang, Tian, Lukasiewicz, & Xu, 2023). Despite facing challenges such as data heterogeneity, model generalization capabilities, and the design of learning strategies, semi-supervised learning continues to demonstrate immense potential, bringing new breakthroughs to the field of medical image segmentation.

With the rapid advancement of deep neural networks, significant progress has been made in medical image segmentation techniques based on deep learning. Deep learning models, particularly Convolutional Neural Networks (CNNs), have become central technologies in the field of medical image segmentation. These models are capable of learning complex image features, offering higher accuracy and robustness compared to traditional machine learning methods. U-Net is a deep learning model originally proposed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015, specifically designed to address medical image segmentation challenges (Ronneberger, Fischer, & Brox, 2015). This model features a U-shaped architecture that efficiently captures both contextual and locational information. Due to its exceptional performance and flexible structural design, U-Net has played a significant role in the field of medical image segmentation. It has not only enhanced the accuracy and efficiency of segmentation tasks but also laid a solid foundation for further research and applications (Cao et al., 2022; Su, Zhang, Liu, & Cheng, 2021; Zhou, Rahman Siddiquee, Tajbakhsh, & Liang, 2018). With the advancement of models like 3D U-Net, deep learning has become increasingly effective at processing volumetric medical images, including CT and MRI scans. This progress allows for more accurate segmentation results, especially in complex anatomical structures (Çiçek, Abdulkadir, Lienkamp, Brox, & Ronneberger, 2016).

In recent years, variants of the U-Net network have made significant progress in the field of medical image segmentation (Zhou et al., 2018) (Huang et al., 2020) (Ibtehaz & Rahman, 2020). Semi-supervised medical image segmentation utilizing U-Net as the backbone network is...
garnering increasing attention. This article aims to review the latest advancements in semi-supervised medical image segmentation based on the U-Net architecture, encompassing key technologies and principal algorithms. By conducting a thorough analysis of current research outcomes and challenges, this review serves to guide and inspire future research directions, aiming to promote further development and innovation in this area.

2. Semi-Supervised Medical Image Segmentation Methods

Semi-supervised learning is a crucial methodology in machine learning that lies between supervised and unsupervised learning, utilizing both labeled and unlabeled data concurrently. This approach showcases unique advantages in scenarios where labels are scarce or costly to obtain. It is particularly significant in the field of medical image segmentation, where the high cost and time consumption of professional medical annotations make semi-supervised methods especially important.

There are various prevalent semi-supervised learning techniques, including pseudo-labeling, semi-supervised learning with unsupervised regularization, and semi-supervised learning with knowledge priors. Pseudo-labeling uses the model's predictions on unlabeled data as "pseudo-labels" to enlarge the training set, thus enhancing model performance. Semi-supervised learning with knowledge priors leverages unlabeled images to impart the model with a priori knowledge of the target's shape and location, improving representational power for medical image segmentation. Semi-supervised learning with unsupervised regularization involves adding one or more regularization losses based on unlabeled data during training, in addition to supervised losses from labeled data, training models with unsupervised regularization by utilizing both labeled and unlabeled images.

Mean Teacher is a widely utilized learning strategy in semi-supervised medical image segmentation, categorized as a type of consistency regularization technique(Tarvainen & Valpola, 2017). It employs two neural networks: a student model and a teacher model, both sharing the same architecture but initialized differently. In the training phase, the student model learns directly from the labeled data, whereas the teacher model's weights are the exponential moving average (EMA) of the student model's weights(Tarvainen & Valpola, 2017). This configuration enhances the stability of the teacher model, reducing its sensitivity to variabilities in the training process, such as initialization and mini-batch selection.

The core idea behind the Mean Teacher strategy is to enforce consistency in the predictions of the student and teacher models on unlabeled data(Tarvainen & Valpola, 2017). This consistency loss encourages the student model to generate stable and reliable predictions even for unlabeled samples. In the domain of medical image segmentation, where annotated data are typically scarce and expensive to acquire, the Mean Teacher approach effectively utilizes unlabeled images to enhance segmentation performance without the need for extensive labeled datasets(Tarvainen & Valpola, 2017). Figure 2.1 depicts the framework of the Mean Teacher strategy, divided into two network branches: the teacher and the student models. For an input unlabeled image, the teacher model provides its gradient-free prediction as a target for the student model, which then learns and optimizes its parameters based on this target. The Mean Teacher model employs an exponential moving average (EMA) for calculating the teacher model's parameters. The teacher model does not undergo direct gradient optimization; instead, its parameters are averaged with the student model's parameters before each backpropagation step, serving as a temporal ensemble of the student model—essentially a collection of the student model's parameters over time. The consistency cost is defined as the expected discrepancy between the weighted, noisy predictions of the student model and those of the teacher model(Tarvainen & Valpola, 2017).

Figure 2.1 Mean Teacher framework diagram(Tarvainen & Valpola, 2017)

Table 2.1 describes the research progress of three commonly used semi-supervised medical image segmentation methods.
Semi-supervised learning methods effectively harness the information within unlabeled data, offering a potent tool for medical image segmentation. Nevertheless, semi-supervised medical image segmentation encounters several challenges,
such as designing more effective strategies to utilize unlabeled data, addressing inconsistencies in data distribution, and enhancing model adaptability across different types of medical images. Future research must delve into more innovative methods to surmount these challenges and further improve the performance of semi-supervised medical image segmentation.

3. U-Net Network and Its Variants

Since its inception, U-Net (Ronneberger et al., 2015) has been extensively utilized in various medical image segmentation tasks such as cell segmentation, organ localization, and lesion detection, thus becoming an essential tool in the domain of medical image analysis (Anand, Gupta, Koundal, & Singh, 2023; Du et al., 2020). U-Net is distinguished by its unique U-shaped architecture, as illustrated in Figure 3.1, which includes a contracting pathway for capturing contextual information and a symmetric expanding pathway for precise localization. This configuration enables the network to effectively capture features across different resolutions from high to low. U-Net enhances segmentation accuracy by incorporating skip connections that merge feature maps from the contracting pathway with those from the expanding pathway, not only facilitating the transmission of contextual information deep within the network but also restoring image details.

The success of U-Net is not only due to its inherent performance but also because it has inspired a series of research focused on improvements and variants of the U-Net structure, thus driving the rapid advancement of medical image analysis technologies. Table 3.1 offers a comparative explanation of U-Net and its variants.

<table>
<thead>
<tr>
<th>Network</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-Net (Ronneberger et al., 2015)</td>
<td>By adding skip connections between the encoder and decoder, it captures feature information lost due to downsampling.</td>
<td>The skip connection method is too simplistic and cannot fully extract feature information.</td>
</tr>
<tr>
<td>U-Net++ (Zhou et al., 2018)</td>
<td>By employing a series of nested dense skip connections and deep supervision, the semantic gap between the encoder and decoder has been reduced.</td>
<td>It fails to filter important feature information, merely concatenating multiple features, which in turn increases the network's burden.</td>
</tr>
<tr>
<td>Attention U-Net (Oktay et al., 2018)</td>
<td>Attention gates have been incorporated into the skip connections, focusing on the regions of interest in the input image for segmentation, while suppressing other areas, thereby highlighting the spatial and geometric information of the segmentation targets.</td>
<td>Extraction of shallow-level information is minimal.</td>
</tr>
<tr>
<td>R2U-Net (Alom, Hasan, Yakopcic, Taha, &amp; Asari, 2018)</td>
<td>The depth of the network has been increased.</td>
<td>The feature fusion strategy is overly simplistic, leading to redundancy in feature information.</td>
</tr>
<tr>
<td>U-Net3+ (Huang et al., 2020)</td>
<td>Learn features from full-scale.</td>
<td>Insufficient generalization capability.</td>
</tr>
<tr>
<td>MultiResU-Net (Ittehaz &amp; Rahman, 2020)</td>
<td>Reduced the semantic gap between low-level and high-level features.</td>
<td>Insufficient utilization of the spatial information in features.</td>
</tr>
<tr>
<td>MSRF-Net (Srivastava et al., 2021)</td>
<td>The network employs multiple dense fusion blocks to achieve the integration of high-level and low-level feature information.</td>
<td>The large number of network parameters makes training difficult to deploy.</td>
</tr>
<tr>
<td>Single level UNet3D (Akbar, Fatichah, Suciati, &amp; Sciences, 2022)</td>
<td>Multi-path residual attention blocks have been introduced into the encoder, combining fusion attention with dilated convolution to effectively enhance the network's ability to capture small tumor features.</td>
<td>The network has an excessively large number of parameters, increasing the network's burden.</td>
</tr>
<tr>
<td>SCAU-Net (Liu et al., 2023)</td>
<td>Embed external attention into the skip connections to better utilize the encoder's capability for semantic upsampling.</td>
<td>The model's performance in enhancing tumor (ET) region segmentation needs improvement.</td>
</tr>
</tbody>
</table>

Although encoder-decoder structured networks used for medical image segmentation, such as U-Net, U-Net++, and R2U-Net, can achieve high segmentation accuracy on some medical datasets, the field of medical image segmentation still faces issues of low feature information utilization, insufficient generalization, and robustness of segmentation models.
4. Semi-Supervised Medical Image Segmentation Method Based on U-Net

Following the success of U-Net in medical image segmentation, researchers have begun investigating how to effectively apply U-Net and its variants to semi-supervised learning to address the issue of scarce annotated data. By incorporating semi-supervised learning strategies like self-training, pseudo-labeling, and consistency regularization, U-Net-based semi-supervised methods are able to effectively utilize a vast amount of unlabeled data, thereby improving segmentation performance.

Ahmed Iqbal and Muhammad Sharif proposed an enhanced version of U-Net for the semi-supervised segmentation of breast tumor images, improving the accuracy of breast tumor segmentation in situations with limited standard datasets (Iqbal & Sharif, 2023). Ashwini and others developed an improved U-Net model (Mod-U-Net), which enhanced the traditional training process of the standard 2-D U-Net (Upadhyay, Bhandari, & Sciences, 2023). They utilized this improved U-Net within a semi-supervised segmentation framework, validating the superiority of their method across two COVID-19 CT segmentation datasets and actual pulmonary CT volumes. Additionally, Chen and his team introduced a semi-supervised learning method using an improved U-Net deep learning network for vascular structure segmentation. Extensive experiments analyzing this semi-supervised learning method demonstrated its ability to overcome the lack of sufficient manual labels while achieving satisfactory performance (Chen, Ao, & Liu, 2020).

Although semi-supervised medical image segmentation based on U-Net and its variants has achieved high accuracy in many applications, several issues persist. The U-Net architecture typically involves a large number of parameters, necessitating significant memory and computational power, particularly for processing 3D images. This complicates the deployment of U-Net in resource-constrained environments.

U-Net enhances information transmission via skip connections, yet its performance may still be limited when processing images with complex structures or those requiring broader contextual information. This is because U-Net's fundamental design is geared more towards capturing local features. In actual medical settings, image quality may be compromised by various factors such as differences in equipment and operational methods, leading to noise and artifacts in images. Without adequate preprocessing or enhancement strategies, U-Net may be particularly sensitive to these noises and artifacts. Considering these issues, further research is needed to improve and optimize semi-supervised medical image segmentation based on U-Net, aiming to boost its performance, adaptability, robustness, and interpretability.

5. Challenges and Future Directions

Despite significant progress in semi-supervised medical image segmentation based on U-Net, several challenges remain:

5.1. Scarcity and Imbalance of Annotated Data
In medical image segmentation tasks, high-quality annotated data are often difficult to obtain, and in some cases, there can be a severe imbalance between different classes of annotated data, such as between images of normal tissue and pathological tissue. This imbalance can negatively impact the training effectiveness of the model.

5.2. Noise and Diversity in Unlabeled Data
Unlabeled data may contain noise or irrelevant information, and data from different sources may vary in distribution. Effectively extracting useful information from these unlabeled data and minimizing the impact of noise on the model are critical issues that semi-supervised learning must address.

5.3. Design of Consistency Constraints
In semi-supervised medical image learning using U-Net, designing effective consistency constraints is crucial to ensure that the model learns from unlabeled data. These constraints help maintain stable outputs in response to minor input variations and are key to improving model generalization.
5.4. Model Generalization and Overfitting

Due to the limitations of labeled data, semi-supervised medical image segmentation models based on U-Net sometimes struggle to generalize to unseen data, especially when there is a significant discrepancy between the distributions of unlabeled and labeled data. Moreover, models may overfit to the small amount of labeled data available, affecting their performance in practical applications.

5.5. Difficulties in Evaluation and Validation

In semi-supervised learning, since most data are unlabeled, traditional performance metrics such as accuracy and recall may not adequately reflect the model's performance. Developing new evaluation metrics or validation methods to accurately assess model performance on unlabeled data is a current research challenge.

5.6. Cross-Modality and Cross-Domain Adaptability

Medical imaging data come from diverse sources, and different imaging techniques (such as CT and MRI) and clinical scenarios pose higher demands on segmentation models. Designing semi-supervised learning models that can operate across different modalities and domains is a direction for future research.

In summary, research in the field of semi-supervised medical image segmentation based on U-Net must continue to overcome these challenges to develop more accurate and generalizable segmentation models that meet clinical application needs.

6. Conclusion

Semi-supervised medical image segmentation technology based on U-Net has evolved from its original model architecture and basic semi-supervised learning strategies to now encompass a variety of innovative algorithms and broadly applied research fields. As new technologies continue to emerge and medical imaging data grows, this area is expected to maintain a rapid development pace, offering increasingly accurate and efficient tools for medical diagnostics and treatment. In response to the challenges and adjustments facing semi-supervised medical image segmentation based on U-Net, researchers need to persist in exploring new improvements to enhance the resolution of issues within this technology, driving its evolution towards greater efficiency, precision, and intelligence.

Acknowledgement

The authors are grateful to the Management and Science University in Malaysia, Guangzhou College of Commerce in China, and the University of Mosul in Iraq for their supports. This work is funded by Guangdong Provincial Education Science Foundation [grant number: 2023GJJK407], the Scientific Research Foundation of Guangzhou [grant number: 202201011667], the Philosophy and Social Science Foundation of Guangzhou [grant number: 2021GZGJ145], the Guangdong Province Undergraduate Teaching Quality and Teaching Reform Engineering Construction Project (2022SJJXGG992), the school level scientific research project of Guangzhou Business School (2023XJYB33).

References


