International Journal on Semantic Web and Information Systems
Volume 20 • Issue 1

A Review of Semantic Medical Image Segmentation Based on Different Paradigms

Jianquan Tan, Key Laboratory of Intelligent Computing Research and Application, Yili Normal University, Yining, China & School of Network Security and Information Technology, Yili Normal University, Yining, China

Wenrui Zhou, Key Laboratory of Intelligent Computing Research and Application, Yili Normal University, Yining, China & School of Network Security and Information Technology, Yili Normal University, Yining, China

Ling Lin, Key Laboratory of Intelligent Computing Research and Application, Yili Normal University, Yining, China & School of Network Security and Information Technology, Yili Normal University, Yining, China*

Huxidan Jumahong, Key Laboratory of Intelligent Computing Research and Application, Yili Normal University, Yining, China & School of Network Security and Information Technology, Yili Normal University, Yining, China

ABSTRACT

In recent years, with the widespread application of medical images, the rapid and accurate identification of these regions of interest in a large number of medical images has received widespread attention. This article provides a review of medical image segmentation methods based on deep learning. Firstly, an overview of medical image segmentation methods was provided in the relevant knowledge, segmentation types, segmentation processes, and image processing applications. Secondly, the applications of supervised, semi supervised, and unsupervised methods in medical image segmentation were discussed, and their advantages, disadvantages, and applicable scenarios were revealed through the application of a large number of specific segmentation examples in practical scenarios. Finally, the commonly used medical image segmentation datasets and evaluation indicators were introduced, and the current medical image segmentation methods were summarized and prospected. This review provides a comprehensive and in-depth understanding for researchers in the field of medical image segmentation, and provides valuable references for the design and implementation of future related work.

KEYWORDS

Biomedical Image Processing, Deep Learning, Image Segmentation, Medical Imaging

INTRODUCTION

Medical images can intuitively reflect the anatomical structure and tissue function and extract large amounts of rich pathological information for medical image segmentation, classification, and disease detection (Johny et al., 2021). This assists doctors in treating diseases, surgery planning, and rehabilitation monitoring. Wang et al. (2018) proposed a cascaded U-Net network combined with a

DOI: 10.4018/IJSWIS.345246 *Corresponding Author

graphical model to segment the aorta, pulmonary artery, myocardium, left and right ventricles, and left and right atria of the heart and then performed similarity shape analysis for image comparison. In recent years, researchers have successfully implemented deep learning (DL) for brain (Zhuang et al., 2022), ear (Xu et al., 2019), liver (Fogarollo et al., 2023), spleen (Sharbatdaran et al., 2022), lung (Johny et al., 2021), kidney (Song et al., 2022), and multi-organ (N. Shen, et al., 2023) segmentation. DL has since been widely used in clinical applications. Due to the ability to extract and analyze biomedical information and apply image segmentation techniques to help doctors and researchers better understand diseases and human physiological conditions, medical image segmentation with DL has become a hot research topic.

Representative literature in the field of medical image segmentation includes the work of Ramesh et al. (2021) and Liu et al. (2021), who focused on classification of the fully convolutional network (FCN), U-Net, and mask R-CNN, but they only touched on the use of DL. Fu et al. (2020) combined image alignment and DL but derived only a few relevant combination techniques. Asgari et al. (2021) introduced FCNs, U-Net, and guided convolutional neural networks (CNNs) and classified them purely from the perspective of DL methods, with little elaboration. With the goal of addressing the shortcomings of previous research, in this study, we sought to provide a systematic exposition of DL-based medical segmentation methods. The study described in this paper contributes to the current literature in the following ways:

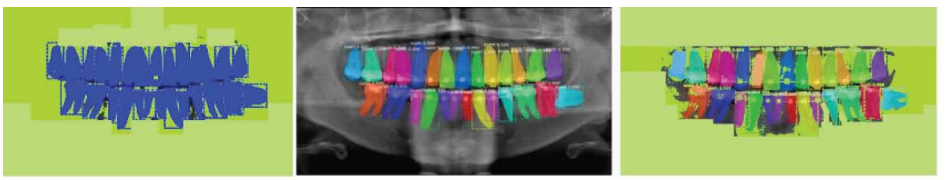
1) This study focused on DL-based medical segmentation methods. It systematically elaborates on basic concepts, basic methodological processes, and learning paradigms, with emphasis on relevant technologies used in medical segmentation and the latest technological improvements in the field.
2) Systematically discussed are the basic ideas of DL segmentation methods in the literature, a summary of the basic methods and main technologies of DL segmentation, and an analysis of the advantages and disadvantages of these technologies.
3) Common medical image segmentation datasets and evaluation metrics are comprehensively summarized. Then described are the data collection and annotation, model interpretability, multi-modal medical image segmentation, semi-supervised learning, unsupervised learning, and optimization of network structure.

RELATED KNOWLEDGE

Overview of Medical Image Segmentation

Medical image segmentation is a technique for separating regions of interest (ROI) in an image (e.g., tissues, organs, lesions) from the background. Segmentation is an important step in image processing and is widely used for medical diagnosis, treatment planning, and assisted surgery. DL methods can automatically learn features of medical images by training neural network models and applying them to segmentation tasks. The types of image segmentation are semantic, instance, and panoramic. Dental X-ray (Z. Shen, et al., 2023) images (Figure 1) are good examples for demonstrating the different

Figure 1. Comparison of Application of Different Tooth Segmentation Types (Note: From left to right the figure shows semantic segmentation, instance segmentation, and panoramic segmentation)



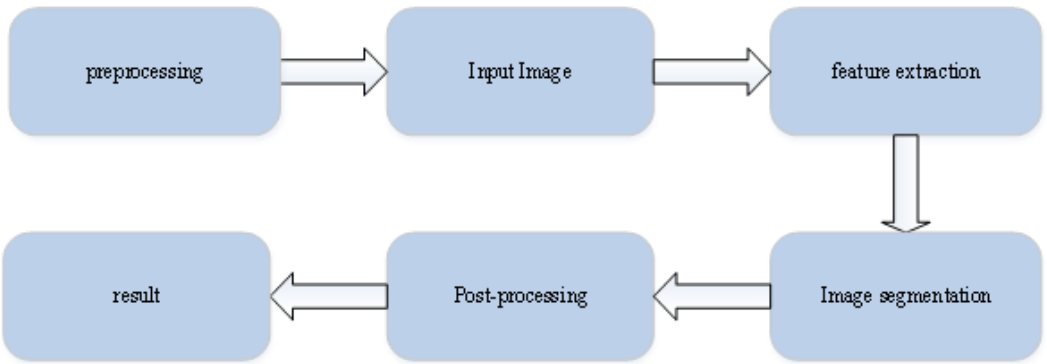
segmentation types. Semantic segmentation, also known as pixel-level segmentation, divides different classes of objects in an image. Instance segmentation implements pixel-level classification and distinguishes instances according to specific classes. Panoramic segmentation requires the assignment of two labels to each pixel of an image: semantic label and instance label. Panoramic segmentation combines semantic segmentation and instance segmentation.

Medical image segmentation requires several steps to distinguish the target area of interest from the background and usually involves image preprocessing, feature extraction, image segmentation, and post-processing. Image preprocessing may include cropping, enhancing contrast, denoising, resizing, cropping edges, rotation, etc. Traditional image segmentation methods often rely on pixel based techniques, region growth, or contour detection. The general segmentation process of deep learning based segmentation techniques is shown in Figure 2. Classifiers and algorithms are usually used to segment medical images, where convolutional layers and max pooling layers play a role in extracting representative feature vectors from preprocessed images.

In recent years, with the continuous development of information technology (IT), computing power has been continuously enhanced. In the face of the shortage of public health resources in recent years and the imbalance of resources caused by regional differences, it has become a trend for machines to replace manual labor for repetitive work. Digital images have become an important type of multimedia data, and it is widely used in modern life (Zheng et al., 2015). Patient electronic medical records on medical data clouds make periodic health examination (PHE) reports accessible to the general public. Onyebuchi et al. (2022) built data warehouses that acquire data from numerous heterogeneous sources and transform, clean, and process it into applicable data repositories for implementation across healthcare organizational settings. Mandle et al. (2022) proposed brain tumor classification based on VGG19 convolutional neural network (CNN), which solved the problem of automatic tumor identification.

In terms of digital image security, in order to solve the complexity of images, Wang et al. (2020) proposed a two-stage image multi-feature fusion paradigm for multimedia data processing. Digital image watermarking technology is used to identify suspicious watermark signals (Li et al., 2019) and improve detection efficiency (Jelušić et al., 2022). Yu et al. (2018) proposed a four-image encryption scheme based on quaternion Fresnel transform (QFST), computer-generated holograms, and two-dimensional logically adjusted sine mapping (LASM). Xu et al. (2021) proposed a secure and efficient certificateless public audit scheme for cloud-assisted medical wireless sensor networks, which not only supports dynamic data sharing and privacy protection, but also achieves efficient group user revocation. Masud et al. (2020) proposed a lightweight physically secure mutual authentication and key establishment protocol that uses a physical unclonable function (PUF) to enable network devices to authenticate the doctor and the user before establishing session keys. Regarding sensor

Figure 2. General Flow of Medical Segmentation Method



node legitimacy, the Internet of medical things (IoMT) enables doctors to remotely diagnose patients, control medical equipment, and monitor quarantined patients through their own digital devices.

In terms of digital synthetic image processing, Qian et al. (2022) believed that the texture clarity of images obtained by most methods is low, resulting in insufficient details of IST. To this end, the authors proposed a new IST method based on an enhanced GAN with a priori recurrent local binary pattern (LBP). They utilized the circular LBP in the GAN generator as texture and then improved the detailed texture of the generated style image. Chopra et al. (2022) generated images through iterative refinement through a stack of two GANs, making the images more realistic after appropriate training.

Many scholars have further conducted semantic research. Chu et al. (2022) used the rich knowledge obtained from labeled 2D images to organize unlabeled 3D models and performed feature extraction on the images and models. Then, the semantic information of multiple clusters on the 3D features and 3D model features were clustered to obtain a more reliable target pseudo label. Zheng et al. (2022) proposed a multi-scale, multi-level ViT model, which can effectively improve the accuracy of fine-grained image classification through data augmentation technology. Nhi and Le (2022) identified the desired images from a large and diverse image dataset. The low-level semantic content of the image included color, shape, and texture. The basic idea of this method is to map low-level features to high-level semantic structures.

Overall, the widespread application of machine learning methods in medical image segmentation and related technologies provides powerful tools and methods for improving healthcare outcomes and addressing challenges in the field of medical imaging. Subsequently, we explored in this article the technical applications of machine learning in the field of medical image segmentation, providing theoretical guidance for specific practical technical applications.

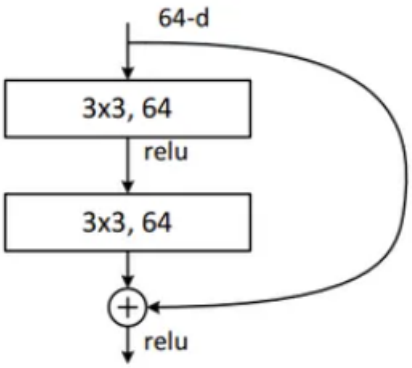
DL-Based Medical Image Segmentation Methods

Image Segmentation Methods for Supervised Learning

Supervised machine learning (ML) is generally more accurate than other paradigms because it trains the model with existing data samples and the corresponding type labels to facilitate the mapping of each data sample to each type label. In recent years, CNNs have been widely used in the field of medical image segmentation. Convolutional computational neural networks acquire image features by learning specific convolutional kernels and subsequently obtain more accurate and efficient segmentation results. With the increasing influence of computational resources, CNNs usually use many smaller-scale convolutions to overlap computational layers. Down sampling is used to reduce the image spatial scale and further increase the CNN depth perception field, thus achieving multi-level feature extraction to improve the segmentation results. Although the supervised learning segmentation method has a high accuracy rate, it also requires a large number of labels for training, and the labeling of data increases the workload. In this subsection, we introduced representative methods applied to image segmentation and summarized the technical characteristics of each method.

U-Net. U-Net is a typical network framework in the field of medical image segmentation. It was proposed by Ronneberger et al. (2015), based on the fully convolutional network (FCN) technique for segmentation. The network can handle smaller training sets and produce more precise segmentation results by modifying and expanding upon the FCN. One main advantage of U-Net is that it preserves a large number of channels during the up sampling step, allowing the network to transmit contextual signals to higher resolutions. Additionally, its symmetrical contracting and expanding paths form a U-shaped architecture, with skip connections used to merge feature maps from different stages. In the 3D U-Net proposed by Setio et al. (2017), after all convolution layers, there are normalization layers and nonlinear activations. The decoder part of the network is almost symmetrical to the encoder. Spatial up sampling is performed using

Figure 3. 64-Channel 3x3 Convolution



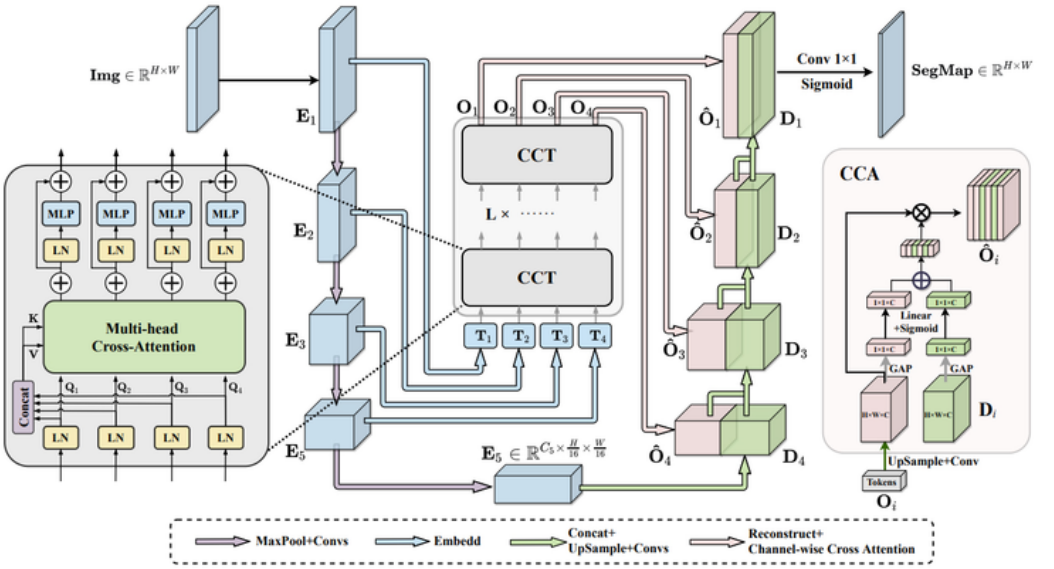
trilinear interpolation between stages. The authors found that down sampling the feature maps using dilation techniques would result in irreversible loss of spatial information.

NnU-Net. Medical image segmentation is a rapidly developing field, with many new network architectures proposed every year. However, certain networks sometimes do not perform effectively in distinct organ or pathology segmentation. This is mainly because there are significant differences in medical datasets such as data size, pixel size, and grayscale values. To overcome this problem, Isensee et al. (2020) proposed NnU-Net, a computational architecture for medical image segmentation based on U-Net, also called 3D U-Net. NnU-Net follows the architectural design pattern of U-Net, but it further focuses on enhancing network training techniques by improving the network technology aspects of U-Net and 3D U-Net. NnU-Net uses processing techniques such as clipping, resampling, normalization, and information reinforcement in the preprocessing stage of simulated input information and sets its own hyper-parameters (e.g., batch size and patch size) according to the characteristics of the information system. It also introduces a fivefold cross-validation exercise in U-Net, 3D U-Net, and two other 3D U-Net cascade models. NnU-Net has the advantage of being applicable to different medical image datasets and can effectively solve the problem of variability among datasets. Differences between datasets thus improve the generalization ability and performance of the model. This scheme has been used in a variety of medical image segmentation tasks such as brain, heart, liver, and other organ segmentation.

ResUNet. ResUNet is a type of residual network with a neural network structure as proposed by He et al. (2016). The increase in bandwidth and depth of progression in neural networks is prone to problems such as step decrement or step explosion, which leads to degradation of network performance. The basic design idea of a residual network system is to introduce a spanning connection in each network layer (i.e., a residual connection). This combination enables network layers to connect to one another and directly transfer the input information to the last level. The residual block (Figure 3) also attempts to understand and fit the residuals to ensure that the number of network layers increases without reducing the expressiveness of the network system. By introducing residual connections, residual network performs well on deep networks and is less prone to step disappearance or step explosion, thus improving the properties of the network system. This approach also enables the network to learn complex features, achieving good results in various computer vision tasks.

UCTransNet. Wang et al. (2021) first proposed the UCTransNet network architecture. It is a deep neural network model based on a U-shaped CNN and transformer structure. The design purpose is to bridge the semantic gap between the encoder and decoder. The literature proposes channel transformers (CTrans) to replace skip connections (Figure 4) in U-Net that consists

Figure 4. Schematic Diagram of UCTransNet (Note: The two components composed of CTrans, which replaced the original skip connections, are channel-wise cross-fusion transformer (CCT) and channel-wise cross-attention (CCA))



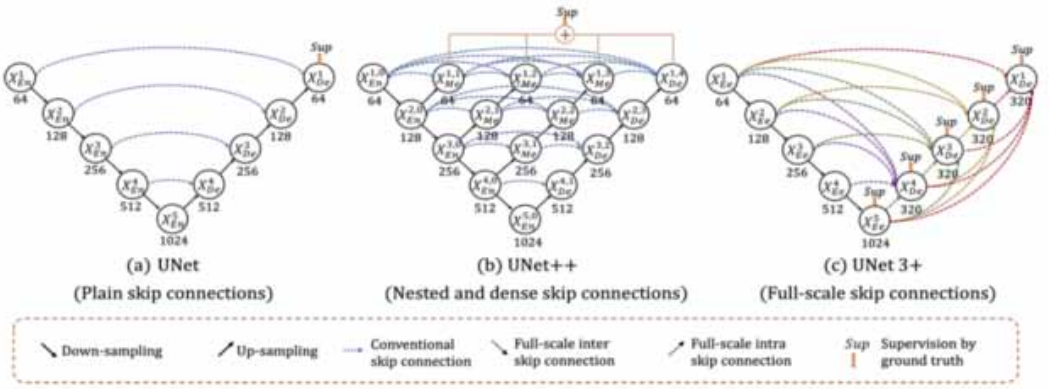
of two modules, namely, the channel-wise cross fusion transformer (CCT) for multi-scale encoder feature fusion and the channel-wise cross-attention (CCA) for decoder feature and enhanced CCT feature fusion. The proposed connection composed of CCT and CCA can replace the original skip connection and solve the semantic gap problem. The multi-head attention mechanism can also improve the model performance by introducing global context information in the transformer structure.

ScaleFormer. Huang et al. (2022) from Tsinghua University proposed the ScaleFormer network architecture. The overall structure is similar to that of the transformer but with modifications and optimizations for the image segmentation task. The model employs a global self-attentive mechanism to capture global contextual information. It also uses a depth separable convolution based on grouped convolution to reduce the number of parameters. The computational effort of the model can handle images of various resolutions, does not require down samplings to reduce the image resolution, simultaneously processes multiple scales of feature maps, and combines them by performing specific fusion methods to improve segmentation accuracy.

MISSFormer. MISSFormer was proposed by X. Huang et al. (2021) at Beijing University of Posts and Telecommunications in 2021. They used hierarchical encoder-decoder in the model. The feedforward network was redesigned using the proposed enhanced transformer block—which makes features adaptively aligned and enhances remote dependencies and local context— and the remote dependencies and local context of multi-scale features generated by the hierarchical transformer encoder were modeled. MISSFormer also introduces an attention mechanism for different scales, dividing the image into multiple scales, and feature vectors within each scale are only associated with feature vectors at other locations within that scale. This reduces computational effort and improves segmentation efficiency.

U-Net++ and U-Net3+. These models were proposed by Zhou et al. (2018). As shown in Figure 5, the main advantage is the addition of more skip connection paths and up sampling convolutional blocks, which establish a direct connection between the decoder and decoder, allowing the former to access more high-rise characteristic information. UNet++ also introduces deep training monitoring technology, adding a branch to the hidden layer in the network to monitor

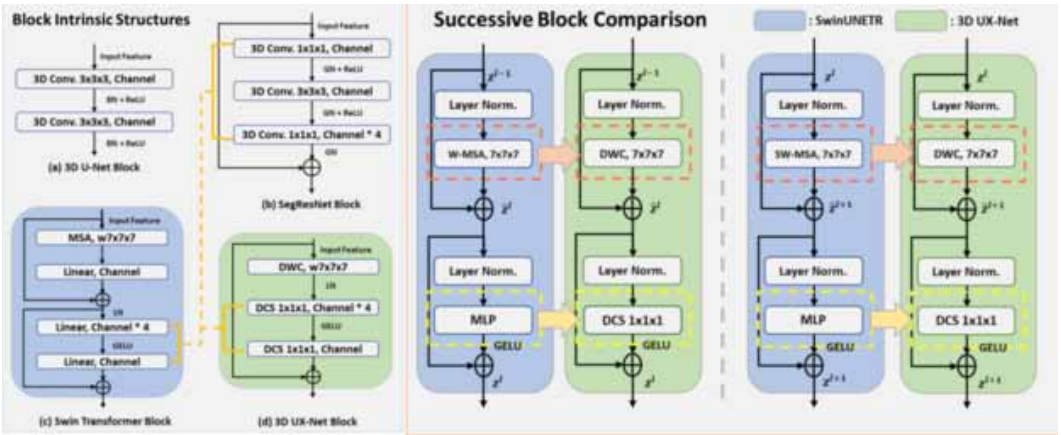
Figure 5. Comparison of U-Net, U-Net++, and U-Net3+ Structural Diagrams



the training of the entire network, and thus overcomes the problem of gradient disappearance. In addition, network-pruning techniques can reduce the number of model parameters, achieving higher operational efficiency of the model by reducing the inference time. Huang et al. (2020) proposed a new U-Net framework called U-Net3+. U-Net3+ can better address the aforementioned problem via full-size skip connection and deep monitoring compared with U-Net++ (Figure 5), which was proposed in 2018 to overcome the inability to obtain full information at multiple sizes. Full-size skip connections originate from fusions of high-level and underlying semantics in feature graphs of different sizes, while deep monitoring shows feature graphs aggregated by different sizes and depths. In addition, U-Net3+ provides a hybrid loss function and a classification bootstrap to increase organ boundaries and reduce the over segmentation of non-organ regions, allowing the model to improve the accuracy of segmentation information. U-Net3+ has higher accuracy and a more flexible network structure than U-Net++ (Figure 5) and can significantly reduce the covariance within a tolerable accuracy range.

3D UX-Net. A large kernel volume convolutional neural network was proposed by Lee et al. (2022). 3D UX-Net (Figure 6), based on the 3D U-Net architecture, adapts the features of a layered transformer to a pure ConvNet module for medical image segmentation in a 3D network architecture. This

Figure 6. 3D UX-Net (Note: This article takes several points from (b) and (c) simultaneously, performs 7 * 7 * 7 convolution using DWC, and then performs depth-wise convolutional scaling (DCS))



scheme has four encoder stages, which is similar to standard 16-fold down sampling. Each stage consists of several but varying 3D UX-Net blocks, and the multi-scale output of each stage connects to the ConvNet decoder through a long skip connection, forming a U-shaped network-like structure for downstream segmentation. Three challenging common data sets of brain and abdomen volume imaging show the fastest rate of convergence in the limited sample training (FeTA2021) and transfer learning (AMOS2022) scenarios. The training with increasing sample size (FLARE2021) resulted in convergence speeds comparable to SwinUNETR, and both their accuracies were higher than the comparative algorithms. 3D UX-Net aims to reduce the number of parameters and computational workload, thereby improving the computational efficiency and inference speed of the model while improving segmentation accuracy. This model has broad application prospects.

Dilated-Unet. As an improved U-Net model, Dilated-Unet was proposed by Azad et al. (2022), introducing the idea of dilated convolution. The convolution operation of expanding the receptive field without increasing the number of parameters is achieved by introducing voids in the convolution kernel. This operation can help the network capture a larger range of contextual information, thereby improving segmentation performance. Both the encoding and decoding stages use dilation convolution. The encoder gradually reduces the size of the feature map and extracts high-level semantic features through multi-layer convolution and pooling operations. The decoder gradually restores the size of the feature map through up sampling and dilation convolution operations, and fuses detail information with contextual information. It performs well in image segmentation tasks in complex scenes.

CS-Unet. CS-Unet is a model based on ViT, which relies too heavily on pretrained data and has inductive bias, making it unable to effectively generalize small datasets. To address this issue, Liu et al (2022) designed CS-Unet, which combines convolutional blocks with multi-head self-attention mechanisms and feedforward networks to provide the required local spatial context information, improve induction bias, introduce conditional parameters to enhance the flexibility and generalization ability of the model, and dynamically adjust the behavior of the input image features to improve the accuracy of segmentation results and detail expression ability. The encoder is responsible for extracting abstract features from the input image, while the decoder gradually restores the original image size and generates segmentation results. By using skip connections, CS-Unet can retain feature information at different levels, which helps to better capture details and contextual information in images. W. Zhang et al. (2023) achieved good results in heart segmentation on the ACDC dataset.

Table 1 shows the medical image segmentation methods based on supervised learning addressed in this section.

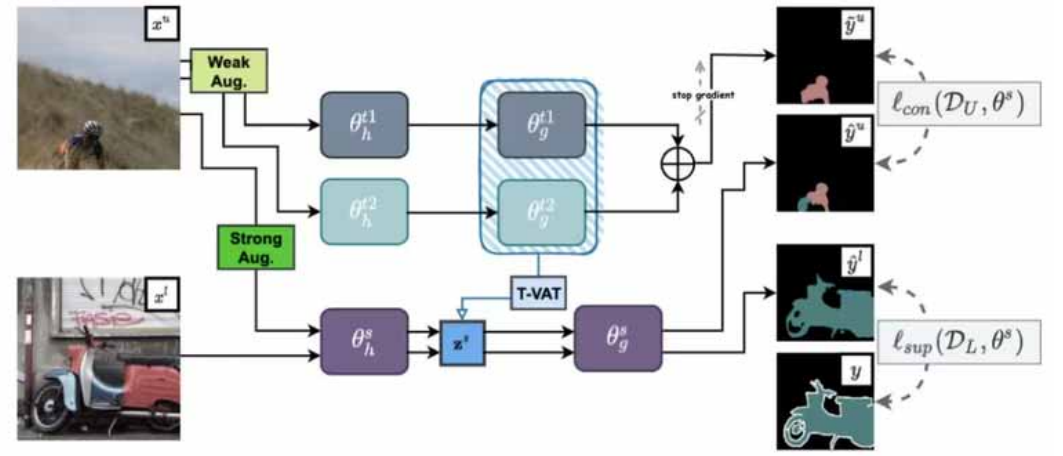
Image Segmentation Methods for Semi-Supervised Learning

The use of deep neural networks is an approach that has recently attained significant progress in the field of medical image analysis. Deep neural networks usually require a large amount of labeled data to provide for model learning, which is particularly difficult for tasks such as medical image segmentation. Medical images are semantically complex and often have 3D information, leading to expensive and time-consuming labeling of image datasets, severely limiting the further development of DL algorithms in this field. With the goal of overcoming these difficulties, researchers have given increased attention to semi-supervised learning methods. During model training, semi-supervised learning realizes the labeling of data by using a large number of pseudo labels. Compared with supervised learning methods, semi-supervised learning methods not only reduce the dependence of labels but also reduce the cost. These features explain why semi-supervised learning methods have wide application prospects in the field of medical image analysis. Large amounts of medical image data are much easier to obtain compared with annotation work, and it can provide more data support

Table 1. Medical Image Segmentation Methods Based on Supervised Learning

Method	Main Technique	Advantage	Shortcoming	Year
U-Net (Isensee et al., 2019)	Encoder and decoder architecture, deconvolution and skip connections	The traditional classification network is transformed into a segmentation network, and the network structure is relatively simple.	The up sampling result is coarse and insensitive to image details.	2015
NnU-Net (Yu et al., 2018)	Data set adaptation, multi-branch networks and deep supervision	Dynamic adaptation to different data sets	Training time is too long, high computational complexity and demanding dataset	2018
ResUNet (Xu et al., 2021)	Residual connection, convolution and batch normalization	Alleviates the gradient disappearance problem in deep networks and accelerates the convergence of the network	The network structure is more complex and requires more training data and computational resources to train the network.	2018
UCTransNet (Masud et al., 2020)	Atreus convolution and global average pooling	Improves the contextual information of the network, reduce the risk of overfitting, make the training and inference process simpler	There are some problems of spatial location information loss.	2022
ScaleFormer (Qian et al., 2022)	Hierarchical encoder structure and multi-scale self-attentive mechanism	Capable of handling multi-scale features effectively	High performance computing equipment required for large computational volumes	2022
MISSFormer (Chopra et al., 2022)	Hierarchical encoder structure and multi-scale self attentive mechanism	Large and small scale features in medical images can be processed.	High performance computing equipment required	2021
U-Net++ (Chu et al., 2022)	Dense connection and deep supervision	Enhancing the gradient flow during training, the network converges, generalizes more easily	Longer training time, redundant training data and consumes more video memory	2018
U-Net3+ (Zheng et al., 2022)	Feature pyramid poo-ling, dense connection and residual connection	Use the underlying features and up sampling features to improve segmentation performance and make the network more robust and generalizable	Longer training time, high memory consumption and high computational complexity	2020
3D UX-Net (Nhi et al., 2022)	The large receptive field brought by nonlocal self attention	Spatial perception ability and richer feature representation	High calculation cost and overfitting risk	2023
Dilated-Unet (Azad 2021)	Dilated convolution	Expanding convolutional operations can increase receptive fields and design Dilated blocks to achieve sparse global attention	Large number of parameters and high computational complexity	2021
CS-Unet (Liu 2023)	Using compressed sensing theory to reconstruct high quality images from a small amount of sampled data	Deep convolutional neural networks and hyper pixel segmentation and multi-level feature extraction and up sampling methods	High computational resource requirements and complex model	2023

Figure 7. Dual-Teacher Architecture



for the semi-supervised learning methods. In this section we summarized a number of semi-supervised segmentation methods and discussed some variant models.

PS-MT. This model proposed by Liu et al. (2021), is a semi-supervised semantic segmentation model based on the teacher algorithm. It is composed of two mean teacher models (Figure 7), namely, the perturbed mean teacher (PMT) model and the strict mean teacher (SMT) model with label constraints. The PMT model introduces a noise perturbation mechanism to enhance the generalization ability by perturbing the input data, whereas the SMT model improves the robustness by imposing constraints on the prediction results. The teacher algorithm generates pseudo labels for training unlabeled data and replacing the MSE loss of MT with the confidence-weighted CE loss (Conf-CE) function, resulting in stronger convergence and better overall training accuracy.

ST++. ST++ is an enhanced self-training model, introduced at Nanjing University, Tencent and Southeast University by Yang et al. (2022). This model retrains samples according to the reliable and unreliable sets. Because of the self-training label method and due to the alternating method between self-training labeled data and unlabeled data for training, errors generated by pseudo labeling are amplified during the training process, leading to the performance degradation of the self-training method. Therefore, two strategies are used to reduce pseudo-labeling errors: A noise suppression mechanism is used for unlabeled data and a new dual CNN design is implemented that further reduces pseudo-labeling errors by integrating deep and shallow features. The specific process of training the ST++ model entails three stages. The first is to train labeled images to obtain an initial teacher model. The second stage involves predicting one-hot pseudo labels for unlabeled images using teacher models. Finally, labeled images, unlabeled images, and pseudo labels, are mixed and a student model is retrained for final testing. The model uses easy-to-hard and reliable-to-unreliable approaches to select unlabeled images and their pseudo labels at the image level, utilizing unlabeled images incrementally. In contrast to the general practice of selecting pixels with high confidence, ST++ selects reliable images based on the stability of the pseudo labels in the first stage of training. Strong data augmentation of unlabeled images in the retraining phase can learn a richer representation based on the teacher model.

RCPS. In 2023, Zhao et al. (2023) proposed correction and contrast pseudo supervision, bidirectional voxel comparison loss, and a confidence negative sampling strategy, which combine correction pseudo-supervision technology and voxel level comparative learning to improve

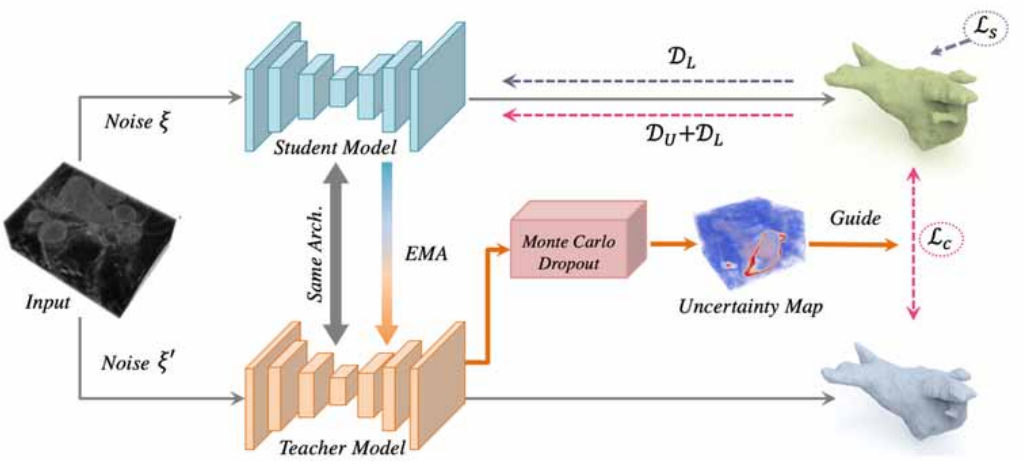
the performance of semi-supervised segmentation models. Correction pseudo-supervision technology improves the segmentation performance of semi-supervised models by learning the robust representations of different segmentation objects in the image space. Bidirectional voxel comparison loss and a confidence negative sampling strategy are used to improve semantic separability between different categories in the feature space. This is a new strategy based on uncertainty estimation and the consistency regularization pseudo-supervision method to reduce the noise impact in pseudo labels. By introducing bidirectional voxel comparison loss to ensure intra-class consistency and inter-class discrimination in the feature space, the class separability in image segmentation is increased.

SCP-Net. Consistency learning effectively utilizes limited labeled and unlabeled data in semi-supervised medical image segmentation. However, its effectiveness and efficiency are challenged by predictive diversity and training stability, as the limited amount of labeled data used for training is often insufficient to form the internal compactness and inter-class differences of pseudo labels. To address these issues,Zhang et al. (2023) proposed a self-perception cross-sample prototype learning method (SCP-Net) to improve the diversity of predictions in consistency learning by utilizing broader semantic information obtained from multiple inputs. It is a self-perceived consistency learning method that utilizes unlabeled data to improve the compactness of pseudo labels in each class. The consistency learning method introduces the dual loss weight loss weighting method to improve the reliability and stability of the model and reduce the negative impact of noise on pseudo labels.

Uncertainty-Aware Mean Teacher (UAMT). UAMT is often used for semi-supervised learning. It uses standard data to improve a more robust model, usually to constrain statistical consistency before and after various data perturbations. Defining additional subtasks provides corresponding invariants to assist in optimizing the network. In contrast to the mean teacher approach, UAMT (Figure 8) uses the dropout technique to generate the uncertainty estimates of the model, using entropy as a metric for screening reliable pseudo labels. Yu et al. (2019) applied the UAMT to a semi-supervised 3D left atrial segmentation task and achieved good results.

FixMatch. Sohn et al. (2020) developed FixMatch in 2020. It combines two common semi-supervised learning methods: the consistency regularization technique and the pseudo-labeling technique. FixMatch generates virtual labeled data, mixes them with other labeled data, and then regularizes

Figure 8. Pipeline of Our Uncertainty-Aware Framework for Semi-Supervised Segmentation (Note: D_U is unlabeled data, D_L is labeled data, and $\mathcal{L}_C$ is consistency loss)



the model via consistency regularization to generate two versions of the input data for model prediction. If the predictions of these two versions are the same, then the sample is consistent and regarded as virtual labeled data with high confidence. This confidence estimation method helps to reduce the effect of mislabeling and improve the generalization ability of the model while simplifying the process of semi-supervised learning. As demonstrated by Luo et al. (2021), FixMatch successfully segmented a prostate gland from MRI images. Unlabeled MRI images were used for virtual label generation and data expansion, extending the labeled dataset and providing new ideas and methods for research and practice in medical image segmentation.

ReMixMatch. This method was developed by Berthelot et al. (2020), whose core idea was to implement distribution alignment and data augmentation anchoring. First, they established the model to align data distribution between training data and unlabeled data. Then, they combined training data with unlabeled data by forcing the enhanced versions of both to use the same labels for learning. Z. Huang et al. (2021) used the adaptive cross-entropy loss based on ReMixMatch, a loss function that can adjust different weighting coefficients according to the perplexity of the samples. This scheme enables hard-to-classify samples to receive more attention during training, thereby improving the performance of the model. In addition, a new data augmentation strategy called sharpness-aware augmentation (SAA) was proposed. This method can improve the robustness of the model by minimizing the entropy of its prediction distribution to select the optimal data augmentation method. Compared with the original MixMatch method, the improved method is more robust and scalable and achieves better results in semi-supervised learning tasks.

Uncertainty-Guided Cooperative Mean Teacher (UCMT). The UCMT-based algorithm, advocated by Tohoku University, Fujian Normal University, and the University of Alberta through the work of Z. Shen et al. (2023), involves semi-supervised semantic segmentation with high confidence pseudo labeling. UCMT consists of two main components: collaborative mean teacher (CMT) and uncertainty-guided region mixing (UMIX). UMIX operates on the input image according to the uncertainty mapping of CMT, whereas CMT is trained collaboratively under the supervision of pseudo labeling of UMIX images. UCMT combines the advantages of UMIX and CMT, preserving model consistency in co-training segmentation and improving the quality of pseudo labeling.

Table 2 outlines the medical image segmentation methods for semi-supervised learning described in this section.

Image Segmentation Methods for Unsupervised Learning

Medical image segmentation based on unsupervised learning refers to a machine learning method that does not rely on manual labeling. It solves the problems of expensive medical image datasets and time-consuming labeling work. The learning process of unsupervised learning involves evaluating and selecting the optimal prediction model from the hypothesis space (model set), and then predicting the results through computational methods such as clustering, dimensionality reduction, and probability estimation. Due to the complex knowledge involved in medical images, the generalization of the model determines the segmentation results. At present, the segmentation effect of many organs is not ideal. The main reason is that the relatively uniform tissue far away from the ROI has serious noise interference, so it is very challenging to segment insensitive structures. Following are several models for unsupervised learning.

TricycleGAN. The TricycleGAN model is a generative adversarial network (GAN)-based image generation model proposed by Baruhov and Gilboa (2020) of the University of Waterloo, Canada. The TricycleGAN model uses a tricycle structure of three generators and three discriminators. It uses multiple loss functions (i.e., generator loss, discriminator loss, and reconstruction loss)

Table 2. Medical Image Segmentation Methods Based on Semi-Supervised Learning

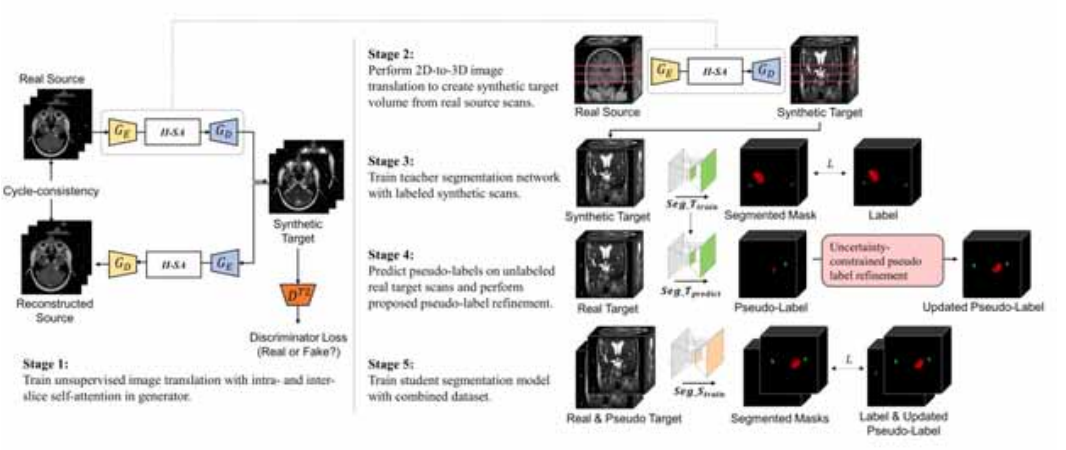
Method	Main Technique	Advantage	Shortcoming	Year
PS-MT (Liu et al., 2021)	Average teacher model and confidence weighting	Ability to learn features automatically	Cannot handle nonparallel voice and text data	2021
ST++ (Yang et al., 2022)	Two-stream convolutional neural network	Improved speech recognition accuracy	Need more computing resources	2022
RCPS (Zhao et al., 2023)	Correction strategy for pseudo-supervised methods of uncertainty estimation and consistency regularization	Higher data utilization and reduced error propagation	Increased complexity and difficulty in parameter tuning	2023
SCP-Net (Z. Zhang et al., 2023)	Self perception and cross sample	Reduce noise in low contrast areas	Dependent on data quality	2023
UAMT (Yu et al., 2019)	Adversarial training, multi-task learning	Multi-task learning scenarios allow for efficient use of unlabeled data, improved generalization of models, and application to different domains and tasks.	Need more computing resources	2021
FixMatch (Sohn et al., 2020)	Confidence estimation	Easy to use, good results	Manual setting of hyperparameters	2020
ReMixMatch (Berthelot et al., 2020)	Multiple data enhancements	Good robustness and scalability	Costly to calculate	2019
UCMT (Z. Shen et al., 2023)	Uncertainty oriented collaborative mean tester with high confidence pseudo labels	Reduce annotation costs	Algorithm complexity	2023

to improve the stability of the model and image quality. In addition, TricycleGAN uses a batch normalization method, thus avoiding pattern collapse.

SDC-UDA. Most medical image segmentation UDA methods use 2D UDA (Figure 9), which can cause inconsistent predictions in the slice direction when stacked together. Shin et al. (2023) led Yonsei University, Naver AI Laboratory, Naver Cloud, Probe Medical Inc., and other units in a collaboration to launch the effective SDC-UDA framework for cross modal medical image segmentation. The joint venture focused the model in the direction of continuous slices, which combines self-attention image conversion within and between slices, pseudo-label optimization with uncertainty constraints, and volume-based self-training. Volume information is considered in the translation and segmentation process, thereby improving the continuity of segmentation results in the slice direction. The difference from previous medical image segmentation UDA methods is that it can achieve continuous segmentation in the slice direction, ensuring higher accuracy and potential in clinical practice.

SSL-ALPNet. Ouyang et al. (2020) proposed a network architecture based on self-supervised learning. The network consists of three main components: a self-encoder network, an auxiliary task network, and a semantic segmentation network. The self-encoder network learns the low dimensional representation of images, the auxiliary task network learns the rotation angle of images, and the semantic segmentation network segments images to complete the process of learning medical images. The introduction of auxiliary tasks and training with label-free data

Figure 9. UDA Framework

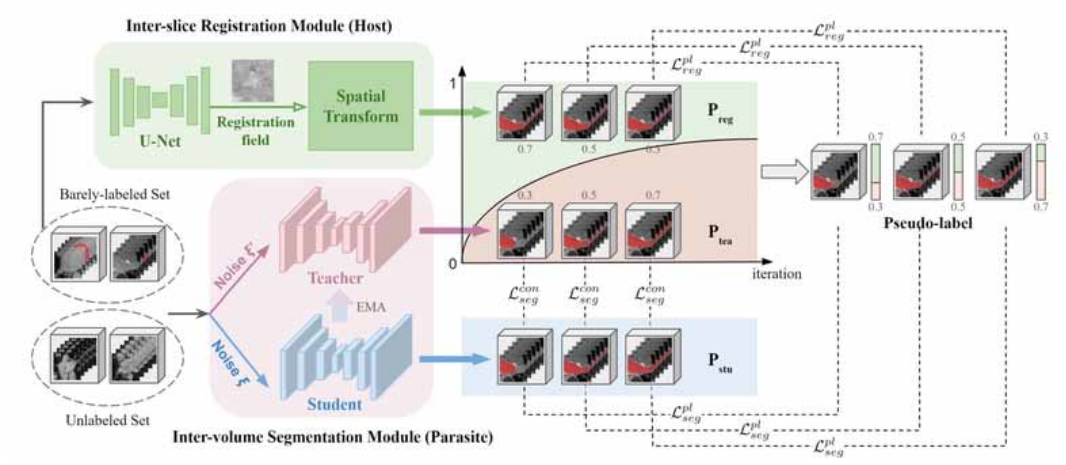


can improve the generalization ability and robustness of the model. SSL-ALPNet also adopts a learning strategy of gradually introducing auxiliary tasks, which enables the model to better learn semantic information and improve segmentation accuracy and stability.

BT-Unet. The BT-Unet method was developed by Sugimoto and Agarwal (2021). Its framework has two phases: pretraining and fine-tuning. BT-Unet is a semantic segmentation model based on U-Net and the bi-tempered logistic loss function, which is an improved cross-entropy loss function that can handle category imbalance and noisy data. The BT-Unet framework applies various advanced U-Net models, such as attention U-Net. Punn and Agarwal (2022) proposed the use of the Barlow Twins method in conjunction with BT-Unet. The Barlow Twins method trains the model by minimizing the difference between two different representations to generate a model with a high-quality representation.

PLN. The PLN method was proposed by Li et al. (2022) and uses a parasitic-like network (or parasitic network) to extract features from raw medical images (Figure 10). These features are then fed into specific model training. In this model, a parasitic-like network with an alignment module

Figure 10. Overview of Our Proposed PLN Framework



(as a host) and a semi-supervised segmentation module (as a parasite) address inter-slice label propagation and inter-volume segmentation prediction, respectively. This method utilizes a small amount of annotated data for training. Furthermore, it generates high-quality features through auxiliary tasks of self-supervised learning. The generalization ability and performance of the model in this manner are improved.

UniMiSS. This model, called the medical self-supervised learning model, breaks down dimensional barriers. University of Adelaide and Northwestern Polytechnical University jointly developed UniMiss with Xie et al. (2022) at the helm. The multi-dimensional medical image data is converted into one-dimensional data and then input into the masked auto-encoder for training. As the masked self-encoder is an unsupervised learning method, the feature representation of the data trains the model by masking some random regions in the data. A pyramidal U-shaped medical switchable patch-embedding module and a transformer composition provide pretrained models and custom training options for fast medical image segmentation. UniMISS supports a wide range of medical image formats and datasets. Visualization tools view the segmentation results and allow for necessary adjustments and editing.

Segmentation methods based on unsupervised learning described in this section are detailed in Table 3.

MEDICAL IMAGE DATASETS AND EVALUATION CRITERIA

Medical Image Datasets

In medical image segmentation, training during network modeling often requires large amounts of labeled information, but the collection process entails many problems. First, considerable time and

Table 3. Medical Image Segmentation Methods Based on Unsupervised Learning

Method	Main Technique	Advantage	Shortcoming	Year
TricycleGAN (Baruhov & Gilboa, 2020)	Tricycle structure with three generators and three discriminators	High quality and efficiency in image conversion tasks between multiple domains	Sensitive to hyperparameters, high computational cost and unstable training	2021
SDC-UDA (Shin et al., 2023)	Intra-slice and inter-slice self-attentive image translation, uncertainty constrained and volumetric self-training	Capture 3D feature information	The difference between the source and target domains is significant, and performance may still be affected.	2023
SSL-ALPNet (Ouyang et al., 2020)	Atreus convolution, self-attention mechanism, residual network, and multi-scale pyramid pooling	A small amount of labeled data can be used for medical image segmentation with strong generalization ability.	Larger computational resources required and longer training time	2021
BT-Unet (Sugimoto & Agarwal, 2021)	Dual logistic loss function, residual connection and self-attention mechanism	Can handle complex medical images	Experiments were conducted on specific data sets only.	2021
PLN (Li et al., 2022)	Sequence-to- sequence models, self-attention and transfer learning	Generate high-quality features with a small amount of labeled data	Requires longer training time and larger computing resources	2022
UniMiSS (Xie et al., 2022)	U-shaped structure of pyramid-shaped	Easy-to-use framework and support for multiple segmentation algorithms	The splitting algorithm is not very accurate.	2022

expertise are required for annotation work when correctly labeling each pixel. The annotation process of the dataset also usually requires large investments in terms of time and human resources. Second, medical image data acquisition requires the use of professional medical imaging equipment; however, the cost of acquisition equipment, such as CT and MRI, is high. Medical institutions must invest huge human and material resources in acquiring, preserving, and managing the relevant data. Furthermore, without patient approval, medical institutions would not disclose relevant medical image data due to privacy requirements and other reasons. Therefore, the datasets currently available to research and development personnel are mainly open source datasets launched by well-known research teams in conjunction with major medical institutions. These include catalogues of categories for medical image data collection (e.g., relevant parts, dataset, pathological mode, image mode, opening year, data format, data content). Also, depending on the imaging mode, the camera forms an intuitive reflection of the organ area in the image, while CT imaging technology scans the human body 360 degrees to calculate the degree of X-ray absorption in the body. It usually provides much higher image quality (bones, lungs, and other areas), faster rates, and higher-dose displays, which are suitable for patients requiring rapid treatment. Meanwhile, MRI uses strong magnetic fields and radio waves to generate pixels with different signal intensities to form images, but it has a relatively slow imaging rate and a lower exposure dose. MRI is mainly suitable for imaging soft tissues such as the brain, liver, kidney, muscle, tumors, and brain tissue and the nervous system. Common image datasets are outlined in Table 4.

Evaluation Criteria

The complexity of the actual evaluation divides the ordinary loss function and the hybrid loss function. In the field of medical image segmentation, the commonly used evaluation metrics are accuracy (AC), recall (SE), specificity (SP), Dice similarity coefficient (DSC), and Jaccard index (JAC). TP is the number of positive samples correctly classified; TN is the number of negative samples incorrectly labeled as positive samples; TP is the number of negative samples correctly classified; and FN is wrong identification as negative samples. In medical image segmentation, A is the predicted data information, while B is the real data information. The expression of the formulae is as follows:

Table 4. Commonly Used Medical Image Datasets

Position	Dataset	Pathological Type	Imaging Modality	Data Format	Year(s)
Heart	ImageCHD (Xu et al., 2019) ACDC (Peng et al., 2021) MMWHS (Peng et al., 2021) HVSMR (Peng et al., 2021)	Coronary heart disease, Congenital heart disease, etc. Heart disease -	CT MRI MRI MRI	NIFTI DICOM NIFTI DICOM	2017-2020 2017 2017 2016
Liver	SLIVER07 (Peng et al., 2021) MICCAI Liver (Di et al., 2022) LiTS Liver (Peng et al., 2021) LiTS2017 (Peng et al., 2021)	Liver disease, liver cancer Liver disease, liver cancer Liver disease, liver cancer Liver disease, liver cancer	CT MRI, CT CT CT	DICOM DICOM, NIFTI NIFTI NIFTI	2007 2007 to present 2017 2017
Lungs	CT Chest (Gozes et al., 2020) LUNA16 (Peng et al., 2021) CTVIE19 (Peng et al., 2021) COVID-19-20 (Peng et al., 2021)	Lung disease, lung cancer Lung disease, lung cancer New crown New crown	CT CT CT CT, X-ray	DICOM DICOM DICOM DICOM, PNG	2020 2016 2020 2020
Spleen	CHAOS (Peng et al., 2021)	-	MRI, CT	DICOM, NIFTI	2019
Kidney	KiTS19 (Peng et al., 2021)	Kidney disease, kidney cancer	CT	DICOM, NIFTI	2019
Multiple organs	TCIA (Huang et al., 2018) MSD((Peng et al., 2021))	Various cancers Various diseases	CT, MRI, PET MRI, CT	DICOM, NIFTI NIFTI	2012 to present 2019
Brain	BraTS (Peng et al., 2021) ISLES (Wang et al., 2019) ADNI (Chen et al., 2019)	Brain disease, brain tumor stroke, Ischemic stroke Alzheimer's disease	MRI MRI MRI, PET	NIFTI NIFTI DICOM, NIFTI	2012 to present 2015 to present 2004 to present

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$SE = \frac{TP}{TP + FN} \tag{2}$$

$$SP = \frac{TN}{TN + FP} \tag{3}$$

$$DSC = \frac{2\left(A \cap B\right)}{A + B} \tag{4}$$

$$JAC = \frac{A \cap B}{A \cup B} \tag{5}$$

The problem of category imbalance is common in medical image segmentation. Dice calculates the similarity of two collections, allowing the problem of category imbalance to be effectively circumvented (i.e., the weights of true examples better measure the accuracy and reliability of segmentation results). Jaccard loss, known as the intersection over union (IoU), focuses more on the segmentation details of the results. The values of the two coefficients are in the range of [0, 1]. The higher the Dice coefficient and Jaccard coefficient are, the better the segmentation effect is. The relationship between the two coefficients is as follows:

$$SP = \frac{2JAC}{1 + JAC} \tag{6}$$

In practice, individual evaluation metrics have difficulty achieving results in complex segmentation tasks; hence, the specific loss function uses specific tasks to improve segmentation accuracy. For example, people with congenital heart disease distribute in different age groups, and this disease has a special complexity. Whole heart segmentation involves many complex segmentation parts, and with the common loss function it is difficult to achieve ideal results. The hybrid loss function proposed by Yang et al. (2018) performed better in whole heart segmentation. It consisted of two parts: voxel size-weighted cross-entropy and multi-class Dice similarity coefficient. In this scheme, the error between the target output and the actual output of the neural network can be determined. Then, the weight of the two loss functions control the proportion of the functions in the total loss function. The commonly used loss function is the cross-entropy-based loss function, but it can hardly achieve the desired effect in cardiac segmentation because the volume of the left ventricular blood chamber, myocardium, and other substructures is often smaller than that of other substructures. The imbalance of simple loss function is more obvious during whole heart segmentation. Furthermore, the loss function based on cross-entropy simply summarizes the error of each pixel and voxel. It may not

capture the specific characteristics of the region of interest, which will lead to the focus on the main region of interest, while the network monitors the region of uninterested, which leads to researchers to focusing on the primary ROI and the network supervising the non-ROI.

The formula of voxel size weighted cross-entropy and multi-class Dice similarity coefficient (DSC) follow.

The complementary hybrid loss function is show in Equation (7). wCross and mDSC can both reduce class imbalance. wCross usually guides the network to retain complex boundary details but introduces considerable noise, whereas mDSC tends to generate more compact and clear boundaries but ignores branch details. The complementary loss approach combines the advantages of both functions to obtain detail-enhanced segmentation results.

$$\varepsilon_{hybrid} = \varepsilon_{wCross} + \alpha\varepsilon_{mDSC} \tag{7}$$

Voxel-size weighted cross-entropy:

The expression of block-by-block weighted cross-entropy is shown in equation (8), where χ denotes the training sample $\left(y_i = \ell\left(\chi_i\right) \mid \chi_i; W\right)$ correspond to sample, and sample corresponding to the target class label $\ell\left(\chi\right)_i$ with the probability of $\left|\chi^{\ell(\chi_i)}\right|$ is determined by the size of the proportion of the target class label $\ell\left(\chi\right)_i$ in the training sample $\chi\eta_\ell\left(\chi_i\right)$ to obtain the $\eta_\ell\left(\chi_i\right)$ weights.

$$\varepsilon_{mDSC}\left(\chi; W\right) = \sum_{\chi_i \in} -\eta_\ell\left(\chi_i\right)\log_p\left(y_i = \ell\left(\chi_i\right) \mid \chi_i; W\right), \eta_\ell\left(\chi_i\right)1 - \frac{\left|\chi^{\ell(\zeta_i)}\right|}{\left|\chi\right|} \tag{8}$$

Multi-class Dice similarity coefficient:

The Dice similarity coefficient (DSC) is a function to alleviate gradient imbalance (Thada & Jaglan, 2013), which is measured by the global shape similarity and based on the differentiable multi-class Dice similarity coefficient (mDSC) as in formula (9), using the loss function to balance multi-class training, where $\mathcal{G}$ is ground truth, $\mathcal{P}$ is probability, i is voxel, and c is volume:

$$\varepsilon_{mDSC} = -\sum_{c \in C} \frac{\frac{2}{N}\sum_i^N \mathcal{G}_c^i \mathcal{P}_c^i}{\mathcal{P}_i^N \mathcal{G}_c^i \mathcal{G}_c^i + \sum_i^N \mathcal{P}_c^i \mathcal{P}_c^i} \tag{9}$$

SUMMARY AND OUTLOOK

Combined with the intensive research in machine science that presently drives the widespread development of CNNs, the DL image segmentation algorithm has the ability to obtain information independently; it can help quickly handle complex medical segmentation tasks. It has a broad application prospect in the field of biomedical image segmentation, but there are also many problems. In the future, researching more efficient, accurate, and stable image segmentation methods to improve clinical application effectiveness still faces a series of challenges. This paper provides a summary and outlook from the following aspects.

Complexity of Medical Image Segmentation

The essence of medical image segmentation is to minimize error analysis, requiring a large number of datasets and high-performance GPUs to provide computational power in deep networks, or spending

more time training the network to compensate for insufficient computational power. High precision will inevitably lead to model complexity, increased parameter count, and excessive reliance on high-performance GPUs. However, convolutional neural networks are usually developed under a fixed budget, and not all research institutions can install high-performance GPUs, which has a relatively low penetration rate. Therefore, from a practical application perspective, the model needs to fully consider the balance between accuracy and efficiency at the beginning of design and cannot blindly add common modules. It is necessary to independently innovate mechanisms that comply with model learning according to actual needs. Medical images are more complex than natural images. Medical images usually contain richer structures and features such as texture, color, and shape. Different diseases, organs, and tissues also vary greatly in morphology and texture. In addition, medical image acquisition entails interference from a variety of factors such as imaging techniques, noise, and artifacts, which affect the accuracy and robustness of segmentation algorithms. Improving the resolution and sensitivity of the probe and software can directly improve the imaging methods. For example, Archibald and Gelb (2002) interpolated segmentation linear functions and corrected an image to reduce the effect of Gibbs ring artifacts. Super-resolution reconstruction (Dong et al., 2015) is another technique for addressing noise issues, and new artifact enhancement techniques can provide new ideas. DL algorithms can also avoid the tedious process of manual design. The automatic learning of features can be adapted to different datasets and tasks, resulting in better portability and robustness.

Label Mark Date

Medical image segmentation usually requires the use of pixel-level annotation data, with each pixel labeled as a target region or a background. However, label marking is a time-consuming and intricate task. The structures and tissues in medical images are diverse and complex, and the label mark is oftentimes susceptible to human factors, as they require specialized medical knowledge and experience to perform accurate labeling. DL techniques can automatically perform medical image segmentation and annotation by training and learning from large-scale data, thus reducing the burden of manual labeling of marks, unbalanced categories, low volume of data annotation, and inconsistent data distribution. Xiao et al. (2020) used semi-supervised domain adaptive methods with pretrained models and migration learning techniques to achieve significant performance gains on several datasets. Self-supervised learning and self-encoder methods and clustering algorithms can also improve the generalization ability and efficiency of DL algorithms and expand their application scenarios.

Model Generalization

At present, medical image segmentation models are mainly designed for single organ segmentation, and the models' generalization abilities are insufficient. When transferred to other organ datasets for training, the performance significantly lags. The fundamental reason for the poor generalization of such deep models is the difference in feature distribution between the training data and the unknown dataset, resulting in poor performance of the originally excellent model when substituted into other datasets. Although training more annotated data is the ideal method to improve generalization ability and help the model fully learn, considering the difficulty of medical image annotation, this method can be implemented. Transfer learning can adapt to other datasets by fine-tuning some parameters of the pretrained network, but it also requires annotating the data. The process of collecting a large amount of data from new fields to retrain the network is expensive. The model lacks the ability to perform global modeling for medical image segmentation and extract detailed features; and has high computational complexity and insufficient generalization ability. To solve the problems in the above models, measures such as enhancing feature extraction, enhancing spatial information extraction, accelerating convergence, eliminating gradient vanishing, avoiding overfitting, expanding receptive fields, achieving global attention, controlling parameter quantity, and reducing computational complexity can be taken to improve the usability of the model.

Interpretability of Models

In deep learning models, the hidden layer contains numerous neurons and parameters. These complex internal structures make the interpretability of the results unclear, making it difficult to enhance the models' performance with the help of doctors' knowledge. Therefore, it is crucial to explain the segmentation process of the model from a clinical perspective. By visualizing feature maps and decision logic at different levels, doctors can understand the sources of pixel-level label maps and master how to form these labels. At the same time, utilizing the saliency map of the derivative direction in the conceptual space of the network, they can predict the trend of disease development; provide explanations for the cause, pathological features, and diffusion process of lesions; and reveal potential indicator markers. These methods enable doctors to comprehensively grasp the working principles and intermediate outputs of the model, thereby applying their prior knowledge to guide the design of the model structure, optimizing decision-making based on clinical experience and identifying potential problems in specific organ segmentation tasks.

Research on Large Models

In the field of image segmentation, the open source of the large segment anything model (SAM) has promoted the development of MedSAM. Although there is still a significant gap in performance between MedSAM and professional models in some tasks, this has not affected researchers' interest in large models. In the future, there may be more advanced network architectures with stronger representation capabilities. The interpretability and visualization level of model predictions will also be enhanced, which will help us better understand the process and quality of manual annotation by doctors.

CONCLUSION

In this article, we summarized the key tasks of medical image segmentation, which involve accurately identifying lesion areas in medical images. We discussed various methods used in the field of medical image segmentation and explored improved approaches, metrics, and the advantages and disadvantages of related techniques. Additionally, we introduced commonly used datasets and evaluation metrics, as well as the challenges and potential solutions in medical segmentation tasks. This research plays an important role in assisting doctors in quickly and accurately identifying areas of interest and has a significant impact on evaluating segmentation quality. In the future, we will combine more domain knowledge and technical means to further summarize the application of convolutional neural networks in the field of medical image segmentation, laying a solid foundation for further research.

CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

FUNDING STATEMENT

This research was supported by National Natural Science Foundation of Xinjiang Province [grant number 2021D01C466], and School-level key project of Yili Normal University [grant number 2023YSZD006, 2020YSZD002], and National Natural Science Foundation Project [grant number 62266046].

PROCESS DATES

Received: 2/26/2024, Revision: 4/7/2024, Accepted: 4/10/2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Ling Lin (China, linling_xj@163.com)

REFERENCES

Archibald, R., & Gelb, A. (2002). A method to reduce the Gibbs ringing artifact in MRI scans while keeping tissue boundary integrity. *IEEE Transactions on Medical Imaging*, *21*(4), 305–319. doi:10.1109/TMI.2002.1000255 PMID:12022619

Asgari, T., Abhishek, K., & Cohen, J. (2021). Deep semantic segmentation of natural and medical images: A review. *Artificial Intelligence Review*, *54*(1), 137–178. doi:10.1007/s10462-020-09854-1

Azad, R., Al-Antary, M. T., Heidari, M., & Merhof, D. (2022). Transnorm: Transformer provides a strong spatial normalization mechanism for a deep segmentation model. *IEEE Access : Practical Innovations, Open Solutions*, *10*, 108205–108215. doi:10.1109/ACCESS.2022.3211501

Baumgartner, C. F., Koch, L. M., & Pollefeys, M. (2018). *An exploration of 2d and 3d deep learning techniques for cardiac MR image segmentation*. Springer. doi:10.1007/978-3-319-75541-0_12

Chen, X., Liu, P., Sun, Y., & Shen, X. (2019). Research on disease prediction models for imbalanced medical datasets. *IT Chinese Journal of Computers*, *3*, 596–609.

Chopra, M., Singh, S. K., Sharma, A., & Gill, S. S. (2022). A comparative study of generative adversarial networks for text-to-image synthesis. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–12. doi:10.4018/IJSSCI.300364

Chu, J., Zhao, X., Song, D., Li, W., Zhang, S., Li, X., & Liu, A. A. (2022). Improved semantic representation learning by multiple clustering for image-based 3D model retrieval. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–20. doi:10.4018/IJSWIS.297033

Di, S. H., Yang, W. H., & Liao, M. (2022). Segmentation of liver tumors in CT images based on RA-unet. *Journal of Instrument and Apparatus*, *43*(8), 65–72.

Dong, C., Chen, C. L., & He, K. (2015). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307. doi:10.1109/TPAMI.2015.2439281 PMID:26761735

Fogarollo, S., Bale, R., & Harders, M. (2023). Towards liver segmentation in the wild via contrastive distillation. *International Journal of Computer Assisted Radiology and Surgery*, *18*(7), 1–7. doi:10.1007/s11548-023-02912-3 PMID:37145251

Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2020). Deep learning in medical image registration: A review. *Physics in Medicine and Biology*, *65*(20), 1–8. doi:10.1088/1361-6560/ab843e PMID:32217829

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

Huang, B., Chen, Z., Wu, P. M., Ye, Y., Feng, S.-T., Wong, C.-Y. O., Zheng, L., Liu, Y., Wang, T., Li, Q., & Huang, B. (2018). Fully automated delineation of gross tumor volume for head and neck cancer on PET-CT using deep learning: A dual-center study. *Contrast Media & Molecular Imaging*, *2018*, 11–19. doi:10.1155/2018/8923028 PMID:30473644

Huang, H., Lin, L., & Tong, R. (2020). Unet 3+: A full-scale connected unet for medical image segmentation. In *Proceedings of the IEEE International Conference on Acoustics*. doi:10.1109/ICASSP40776.2020.9053405

Huang, X., Deng, Z., & Li, D. (2021). MISSFormer: An effective medical image segmentation transformer. *IEEE Transactions on Medical Imaging*, *40*(7), 236–248. PMID:37015444

Huang, Z., Long, G., & Wessler, B. (2021). A new semi-supervised learning benchmark for classifying view and diagnosing aortic stenosis from echocardiograms. In *Proceedings of the Machine Learning for Healthcare Conference,* 2021.

Isensee, F., Jaeger, P., Kohl, S., Petersen, J., & Maier-Hein, K. H. (2020). Nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, *17*(2), 203–211. doi:10.1038/s41592-020-01008-z PMID:33288961

Jelušić, P. B., Poljičak, A., Donevski, D., & Cigula, T. (2022). Low-frequency data embedding for DFT-based image steganography. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–11. doi:10.4018/IJSSCI.312558

Johny, D., Subramanyam, K., & Baikunje, N. (2021). Cardiac tamponade and massive pleural effusion in a young COVID-19- positive adult. *Case Reports*, *14*(9), e244518. PMID:34518185

Li, D., Deng, L., Bhooshan Gupta, B., Wang, H., & Choi, C. (2019). A novel CNN based security guaranteed image watermarking generation scenario for smart city applications. *Information Sciences*, *479*, 432–447. doi:10.1016/j.ins.2018.02.060

Li, S., Cai, H., Qi, L., Yu, Q., Shi, Y., & Gao, Y. (2022). PLN: Parasitic-like network for barely supervised medical image segmentation. *IEEE Transactions on Medical Imaging*, *42*(3), 582–593. doi:10.1109/TMI.2022.3211188 PMID:36178993

Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A review of deep-learning based medical image segmentation methods. *Sustainability*, *13*(3), 1–29. doi:10.3390/su13031224 PMID:34676112

Liu, Y., Tian, Y., & Chen, Y. (2022). Perturbed and strict mean teachers for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR52688.2022.00422

Luo, X., Chen, J., & Song, T. (2021). Semi-supervised medical image segmentation through dual-task consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v35i10.17066

Mandle, A. K., Sahu, S. P., & Gupta, G. P. (2022). CNN-based deep learning technique for the brain tumor identification and classification in MRI images. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–20. doi:10.4018/IJSSCI.304438

Masud, M., Gaba, G. S., Alqahtani, S., Muhammad, G., Gupta, B. B., Kumar, P., & Ghoneim, A. (2020). A lightweight and robust secure key establishment protocol for internet of medical things in COVID-19 patients care. *IEEE Internet of Things Journal*, *8*(21), 15694–15703. doi:10.1109/JIOT.2020.3047662 PMID:35782176

Nhi, N. T. U., & Le, T. M.Thanh The Van. (2022). A model of semantic-based image retrieval using c-tree and neighbor graph. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–23. doi:10.4018/IJSWIS.295551

Onyebuchi, A., Matthew, U. O., Kazaure, J. S., Okafor, N. U., Okey, O. D., Okochi, P. I., & Matthew, A. O. (2022). Business demand for a cloud enterprise data warehouse in electronic healthcare computing: Issues and developments in e-healthcare cloud computing. [IJCAC]. *International Journal of Cloud Applications and Computing*, *12*(1), 1–22. doi:10.4018/IJCAC.297098

Ouyang, C., Biffi, C., & Chen, S. (2020). Self-supervision with super pixels: Training few-shot medical image segmentation without annotation. In *Proceedings of the Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK*.

Peng, J., Luo, H., & Zhao, G. (2021). A review of medical image segmentation algorithms based on deep learning. *Computer Engineering and Applications*, *57*(3), 44–57.

Punn, N. S., & Agarwal, S. (2022). BT-unet: A self-supervised learning framework for biomedical image segmentation using Barlow twins with U-net models. *Machine Learning*, *111*(12), 107–108. doi:10.1007/s10994-022-06219-3

Qian, W., Li, H., & Mu, H. (2022). Circular LBP prior-based enhanced GAN for image style transfer. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(2), 1–15. doi:10.4018/IJSWIS.315601

Ramesh, K. K., Kumar, G. K., & Swapna, K. (2021). A review of medical image segmentation algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*, *7*(27), 1–10.

Ronneberger, O., Fischer, P., & Brox, T. (2015). Medical Image Computing and Computer-Assisted Intervention: Vol. 234-241. *U-net: convolutional networks for biomedical image segmentation*.

Serra, J. (1982). Image analysis and mathematical morphology. *Biometrics*, *39*(2), 536.

Setio, A., Traverso, A., & Bel, T. D. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Elsevier, 89, 117-128.

Sharbatdaran, A., Romano, D., Teichman, K., Dev, H., Raza, S. I., Goel, A., Moghadam, M. C., Blumenfeld, J. D., Chevalier, J. M., Shimonov, D., Shih, G., Wang, Y., & Prince, M. R. (2022). Deep learning automation of kidney, liver, and spleen segmentation for organ volume measurements in autosomal dominant polycystic kidney disease. *Tomography : a Journal for Imaging Research*, *8*(4), 1804–1819. doi:10.3390/tomography8040152 PMID:35894017

Shen, N., Wang, Z., Li, J., Gao, H., Lu, W., Hu, P., & Feng, L. (2023). Multi-organ segmentation network for abdominal CT images based on spatial attention and deformable convolution. *Expert Systems with Applications*, *211*, 1–15. doi:10.1016/j.eswa.2022.118625

Shin, H., Kim, H., & Kim, S., Jun, (2023). SDC-UDA: Volumetric unsupervised domain adaptation framework for slice-direction continuous cross-modality medical image segmentation. In *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR52729.2023.00716

Shu-Long, Z. (2002). Image fusion using wavelet transform. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *34*(4), 552–556.

Sohn, K., Berthelot, D., Li, C. L. (2020). *Fixmatch: Simplifying semi-supervised learning with consistency and confidence.*

Song, Y., Zheng, J., Lei, L., Ni, Z., Zhao, B., & Hu, Y. (2022). CT2US: Cross-modal transfer learning for kidney segmentation in ultrasound images with synthesized data. *Ultrasonics*, *122*, 1–9. doi:10.1016/j.ultras.2022.106706 PMID:35149255

Sugimoto, M. P., & Agarwal, S. (2021). BT-unet: A self-supervised learning framework for biomedical image segmentation using Barlow Twins with U-Net models. *arXiv e-prints*, arXiv2112.

Thada, V., & Jaglan, V. (2013). Comparison of Jaccard, Dice, Cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *International Journal of Innovations in Engineering and Technology*, *2*(4), 202–205.

Tong, Q., Ning, M., & Si, W. (2017). 3D deeply-supervised u-net based whole heart segmentation: Statistical atlases and computational models of the heart. In *Proceedings of the ACDC and MMWHS Challenges: Held in Conjunction With MICCAI.*

Wang, H., Cao, P., & Wang, J. (2021). Uctransnet: Rethinking the skip connections in u-net from a channel-wise perspective with transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Wang, H., Li, Z., Li, Y., Gupta, B. B., & Choi, C. (2020). Visual saliency guided complex image retrieval. *Pattern Recognition Letters*, *130*, 64–72. doi:10.1016/j.patrec.2018.08.010

Wang, P., Gao, C., & Zhu, L. (2019). Ischemic stroke lesion segmentation algorithm based on 3D deep residual network and cascaded U-Net. *Jisuanji Yingyong*, *39*(11), 6.

Xiao, L., Xu, J., & Zhao, D. (2020). Self-supervised domain adaptation with consistency training. *IEEE Transactions on Image Processing*, *29*, 574–576.

Xie, Y., Zhang, J., & Xia, Y. (2022). UniMiSS: Universal medical self-supervised learning via breaking dimensionality barrier. In *Proceedings of the European Conference on Computer Vision*. Springer, Cham. doi:10.1007/978-3-031-19803-8_33

Xu, X., Wang, T., & Shi, Y. (2019). Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. In *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer, Cham. doi:10.1007/978-3-030-32245-8_53

Xu, Z., He, D., & Vijayakumar, P. (2021). Certificateless public auditing scheme with data privacy and dynamics in group user model of cloud-assisted medical WSNs. *IEEE Journal of Biomedical and Health Informatics*. PMID:34788225

Yang, L., Zhuo, W., & Qi, L. (2022). ST++: Make self-training work better for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR52688.2022.00423

Yang, X., Bian, C., Yu, L., Ni, D., & Heng, P.-A. (2018). Hybrid loss guided convolutional networks for whole heart parsing. *Lecture Notes in Computer Science*, *10663*, 215–223. doi:10.1007/978-3-319-75541-0_23

Yu, C., Li, J., Li, X., Ren, X., & Gupta, B. B. (2018). Four-image encryption scheme based on quaternion Fresnel transform, chaos and computer generated hologram. *Multimedia Tools and Applications*, *77*(4), 4585–4608. doi:10.1007/s11042-017-4637-6

Yu, L., Wang, S., & Li, X. (2019). Uncertainty-aware self-assembling model for semi-supervised 3D left atrium segmentation. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention-MICCAI 2019: 22nd International Conference, Shenzhen, China.*

Zhang, W. Z., Qian, Y., & Su, J. S. (2023). From U-Net to Transformer: A review of the application of deep models in medical image segmentation. *Computer Applications (Nottingham)*.

Zhang, Z. X. (2023). Self-aware and cross-sample prototypical learning for semi-supervised medical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. doi:10.1007/978-3-031-43895-0_18

Zhao, X. Y. (2023). RCPS: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, ●●●, 1–12. PMID:37801388

Zheng, Y. H., Jeon, B., Xu, D. H., Wu, Q. M., & Zhang, H. (2015). Image segmentation by generalized hierarchical fuzzy c-means algorithm. *Journal of Intelligent & Fuzzy Systems*, *28*(2), 961–973. doi:10.3233/IFS-141378

Zheng, Z., Zhou, J., Gan, J., Luo, S., & Gao, W. (2022). Fine-grained image classification based on cross-attention network. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–12. doi:10.4018/IJSWIS.315747

Zhou, Z., Rahman, S., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain*.

Zhuang, X., Xu, J., Luo, X., Chen, C., Ouyang, C., Rueckert, D., Campello, V. M., Lekadir, K., Vesal, S., RaviKumar, N., Liu, Y., Luo, G., Chen, J., Li, H., Ly, B., Sermesant, M., Roth, H., Zhu, W., Wang, J., & Li, L. et al. (2022). Segmentation on late gadolinium enhancement MRI: A benchmark study from multi-sequence cardiac MR segmentation challenge. *Medical Image Analysis*, *81*, 1–14. doi:10.1016/j.media.2022.102528 PMID:35834896