

The Application of MI Based Prediction of The Deep Learning CNN Model To Enhancement And Scheduling The GPU Utilization of The DS

^[1]Ms.P.Sivakamasundari ,^[2] Santhohkumar.R

^[1] Assistant Professor, Department of Computer Science and Engineering, Adhiparasakthi Engineering College, Melmaruvathur, Tamilnadu, India.

^[2] Department of Computer Science and Engineering, Adhiparasakthi Engineering College, Melmaruvathur, Tamilnadu, India.

Abstract A Segmentation process is labeling an image used for obtaining more meaningful information. In biomedical images, it plays has an important role in helping pathologist for conducting advanced analysis. After the Graphical Processing Unit (GPU) introduced not only for graphical necessary but also for general computing, segmentation process which is computationally expensive can be potentially improved. The good accuracy detection and segmentation result provides morphological information for the pathologist. Consequently, more advanced approaches were developed to ensure the good performance of detection and segmentation such as deep learning approach. The Convolutional Neural Network (CNN) is one of deep learning architecture with complex computation. This paper presents an overview of utilization of Convolutional Neural Network as prominent deep learning architecture under Graphical Processing Unit platform and proposes an approach of using GPU as potential further parallel techniques in CNN.

Keyword Image segmentation; deep learning; convolutional neural network; medical image, GPU speed.:

1. INTRODUCTION

Image segmentation is an important and difficult part of image processing. It has become a hotspot in the field of image understanding. This is also a bottleneck that restricts the application of 3D reconstruction and other technologies. Image segmentation divides the entire image into several regions, which have some similar properties. Simply put, it is to separate the target from the background in an image. At present, image segmentation methods are developing in a faster and more accurate direction. By combining various new theories and new technologies, we are finding a general segmentation algorithm that can be applied to kind of images[1]. With the advancement of medical treatment, all kinds of new medical imaging equipment are becoming more and more popular. The types of medical imaging widely used in clinic are mainly CT, MRI, PET, X-ray and UI. In addition, it also includes some common RGB images, such as microscopy and fundus retinal images. There is very useful information in medical images. Doctors use CT and other medical images to judge the patient's condition, which has gradually become the main basis for doctors' clinical diagnosis [2]. Therefore, the research on medical image processing has become the focus of attention in the field of computer vision with the rapid development of artificial intelligence, especially deep learning, image segmentation methods based on deep learning have achieved good results in the field of image segmentation. Compared with traditional machine learning and computer vision methods, deep learning has certain advantages in segmentation accuracy and speed. Therefore, the use of deep learning to segment medical images can effectively help doctors confirm the size of diseased tumors, quantitatively evaluate the effect before and after treatment, greatly reducing the workload of doctors.

In order to better summarize the various methods, we searched the keywords "medical image processing" or "deep learning" from Google Scholar and ArXiv to obtain the latest literature. In addition, the top medical image processing conferences are also good places for us to obtain materials, such as MICCAI ,ISBI and IPMI. The papers we selected are mainly based on deep learning methods. We guarantee that all the results of the papers are verified. Different from the existing reviews [4–6], this survey reviews the recent progress, advantages, and disadvantages in the field of medical image segmentation from the perspective of deep learning. It compares and summarizes related methods and identifies the challenges for successful methods of deep learning to medical imaging segmentation

task in the future work. In this paper, we conduct a comprehensive review of medical imaging DL technology in recent years, mainly focusing on the latest methods published in the past three years and the classic methods in the past. First, it focuses on the application of deep learning technology in medical image segmentation in the past three years. A more in-depth study is carried on its network structure and methods. At the same time, its strengths and weaknesses are analyzed. Second, some state-of-the-art segmentation methods are summarized according to the characteristics of different organs and tissues. Third, we shared many evaluation metrics and data sets of medical image segmentation for readers to evaluate and train the network. The structure of the article is as follows: Section 2 examines what is medical image segmentation. In Section 3, we explained the concept of deep learning and the application of deep learning. Sections 4 and 5 are the main body of the reviewed literature. Section 4 introduces the three network structures, FCN, U-Net and GAN based on deep learning medical image segmentation. Section 5 introduces the segmentation methods of different organs and tissues. Section 6 is the sharing of evaluation metrics and data sets, which are derived from the influential medical image analysis challenges.

1.1. MEDICAL IMAGE SEGMENTATION

Image segmentation based on medical imaging is the use of computer image processing technology to analyze and process 2D or 3D images to achieve segmentation. It divides the image into several regions based on the similarity or difference between regions. Doctors can perform qualitative or even quantitative analysis of lesions and other regions of interest through this method, thereby greatly improving the accuracy and reliability of medical diagnosis. Currently, the main variety, tissues and organs of the image cells are used as object. Generally, medical image segmentation can be described by a set theory model: given a medical image I and a set of similarity constraints C_i ($i = 1, 2, \dots$), the segmentation of I is to obtain a division of it, namely:

$$N \times U = 1 \quad R_x \cap R_y = \emptyset, \quad \forall x \neq y, \quad x, y \in [1, N] \quad (1)$$

where R_x satisfies both sets of all pixels in communication similarity constraint C_i ($i = 1, 2, \dots$), i.e., the image areas. The same is true for R_y . x, y are used to distinguish the different regions. N is a positive integer not less than 2, indicating the number of regions after division.

Obtain medical imaging data set, generally including training set, validation set, and test set. When using machine learning for image processing, the data set is often divided into three parts. Among them, the training set is used to train the network model, the verification set is used to adjust the hyperparameters of the model, and the test set is used to verify the final effect of the model. Preprocess and expand the image, generally including standardization of input image, perform random rotation and random scaling on the input image to increase the size of the data set use appropriate medical image segmentation method to segment the medical image, and output the segmented images. In order to verify the effectiveness of medical image segmentation, effective performance indicators need to be set to be verified.

1.2. IMAGE SEGMENTATION

Image segmentation is a classic problem in computer vision research and has become a hotspot in the field of image understanding. The so-called image segmentation refers to the division of an image into several disjointed areas according to features such as grayscale, color, spatial texture, and geometric shapes. So that these features show consistency or similarity in the same area, but between different areas shows a clear difference. Image segmentation is divided into semantic segmentation, instance segmentation and panoramic segmentation according to the different coarse and fine granularity of segmentation. At present, there are more and more research branches of image segmentation, such as satellite image segmentation, medical image segmentation, autonomous driving [8,9], etc. With the large increase in the proposed network structure, the image segmentation method is improved step by step to obtain more and more accurate segmentation results. However, for different segmentation examples, there is no universal segmentation algorithm that is suitable for all images.

Traditional image segmentation methods can no longer be compared with the segmentation methods based on deep learning in effect, but the ideas are still worth learning [10–12]. Like the proposed threshold-based segmentation method [13], region-based image segmentation method [14], and edge detection-based segmentation method [15]. These methods use the knowledge of digital image processing and mathematics to segment the image. At present, methods based on deep learning have made remarkable achievements in the field of image segmentation. Their segmentation accuracy has surpassed traditional segmentation methods. The fully convolutional network was the first to successfully use deep learning for image semantic segmentation. The authors proposed the concept of full convolutional networks. Then there are outstanding segmentation networks such as U-Net, Mask R-CNN [16], RefineNet [17], and DeconvNet [18]. This template, modified in MS Word 2007 and saved as a “Word 97-2003 Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to

electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-level equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

II. DEEP LEARNING

2.1. OVERVIEW OF DEEP LEARNING

Deep learning is a research trend in the rise of machine learning and artificial intelligence. It uses deep neural networks to simulate the learning process of the human brain and extract features from large-scale data (sound, text, images, etc.) in an unsupervised manner [19]. A neural network is composed of many neurons. Each neuron can be regarded as a small information-processing unit. The neurons are connected to each other in a certain way to form the entire deep neural network. The emergence of neural networks makes end-to-end image processing possible. When the hidden layers of the network develop to multiple layers, it is called deep learning. In the field of computer vision, deep learning is mainly used in data dimensionality reduction, handwritten number recognition, pattern recognition and other fields. Such as image recognition, image repair, image segmentation, object tracking, scene analysis, etc., showing very high effectiveness [20].

2.2. CONVOLUTIONAL NEURAL NETWORK

The convolutional neural network (CNN) [21] is a classic model produced by the combination of deep learning and image-processing technology.

As one of the most representative neural networks in the field of deep learning technology, it has made many breakthroughs in the field of image analysis and processing including image feature extraction and classification, pattern recognition, etc. The convolutional neural network is a deep model with supervised learning. The basic idea is to share the weights of feature mapping in different positions of the previous layer network, and to reduce the number of parameters by using spatial relative relationships to improve training performance. The proposal of receptive fields and neurocognitive machines in human visual information is an important theory in the embryonic stage of theory.

In 1962, Hubel et al. [22] showed through biological research that the transmission of visual information in the brain from the retina is accomplished through multilevel receptive field excitation. This is the first proposed the concept of receptive field. In 1980, Fukushima [23] proposed a neurocognitive machine based on the concept of receptive fields. It is regarded as the first implementation network of convolutional neural networks. In 1998, LéCun et al. [24] proposed LeNet5 using a gradient-based backpropagation algorithm for supervised training of the network, which entered the experimental development stage. The academic circle's attention to convolutional neural networks also began with the proposal of the LeNet5 network and successfully applied to handwriting recognition. The AlexNet proposed by Krizhevsky et al. [25] was the most successful at image classification of the training set of ImageNet, making convolutional neural networks become the key research object in computer vision, and this research continues to deepen.

2.2.1. 2D CNN

CNN consists of an input layer, an output layer, and several hidden layers. Each layer in the hidden layer performs a specific operation, such as convolution, pooling, and activation.

The input layer is connected to the input image, and the number of neurons in this layer is the pixel of the input image. The middle convolutional layer performs feature extraction on the input data through a convolution operation to obtain a feature map. The result of the convolution operation depends on the setting of the parameters in the convolution kernel. The pooling layer behind the convolutional layer filters and selects feature maps, simplifying the computational complexity of the entire network, the general convolutional neural network is 2D CNN.

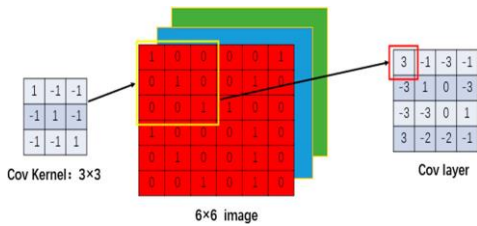


Fig1. Two-dimensional convolutional neural network (2D CNN) convolution.

2.2.2. 3D CNN

Most images in medical images are usually 3D, such as CT and MRI. Although the CT image we usually see is a 2D image, it is just a slice of it. Therefore, if you want to segment some diseased tissues, you must use a 3D convolution kernel. For example, the convolution kernel used by the segmentation network 3D U-Net is 3D. It changed the 2D convolution kernel in the U-Net network to a 3D convolution kernel, which is suitable for 3D medical image segmentation [28]. 3D CNN can extract a more powerful volume representation on the three axes of X, Y, and Z.

The use of three-dimensional information in segmentation makes full use of the advantages of spatial information. The 3D convolution kernel has one more depth than the 2D convolution kernel, which means the number of 2D slices of medical images. Given a 3D image $C \times N \times H \times W$ where C , N , H and W represent the number of channels, the number of slice layers, the height and width of the convolution kernel.

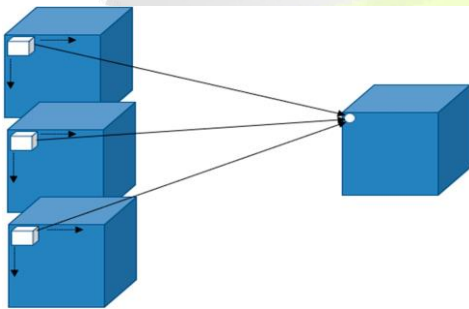


Fig 2. 3D CNN convolution.

2.2.3. BASIC DL ARCHITECTURES FOR SEGMENTATION

The segmentation network is also changed in the common CNN structure. The first segmentation network was to change the last two fully connected layers for the classification network to convolutional layer. The bone of the medical image segmentation network is based on the deep structure like VGG and ResNet as well as the encoder-decoder structure. LeNet and AlexNet are early network models. The two network structures are relatively similar and belong to shallow networks. AlexNet has many more parameters than LeNet network. Its idea of adding a pooling layer after the convolutional layer is still popular now. An improvement of VGG over AlexNet is to deepen the number of network layers. It used several consecutive 3x3 convolution kernels to replace the larger convolution kernel in AlexNet. Under the condition of ensuring the same receptive field, the depth of the network and the effect of feature extraction are advanced and this can also cause problems, such as overfitting and vanishing gradients.

In response to these problems, GoogleNet [29] improved from another perspective, dividing the evacuation network structure into modules. The inception structure is proposed to increase depth and width of the network while reducing parameter of the network. Inception uses multiple convolution kernels of different sizes and adds pooling. Then the result of convolution and pooled are together in series. The depth of the entire network reached 22 layers. The CNN network has developed from the seven layers of AlexNet to the 19 layers of VGG, followed by 22 layers of GoogleNet. When the depth reaches a certain number of layers, the further increase cannot improve the performance of classification, but will cause the network to converge slowly. In order to train a deeper network with good results, He et al. [26] proposed a new 152-layer network structure—ResNet. ResNet solves this problem by using shortcut, which is composed of many residual blocks. Each module consists of a number of consecutive layers and a shortcut. This shortcut connects the input and output of the module together, adding them before ReLU (rectified linear unit) activation.

The resulting output is then send to the ReLU activation function to generate the output of this block. Besides, there are network

structural units like squeeze-and-excitation blocks, which improve the expressive ability of the network model from the perspective of the new network model, the channel relationship, to design [30]. Combining the front-end-based CNN encoder and the back-end-based decoder together, this is the encoder-decoder architecture. It is also the basic structure of a semantic segmentation network.

The structure of the encoder in the segmentation task is similar, and most of them are CNNs for classification tasks. It extracts image features from the input image, and compacts the features by encoding to produce the low-resolution feature map. The decoder maps the low-resolution discriminative feature map learned by the encoder to the high-resolution pixel space to realize the category labeling of each pixel. SegNet [31] is a classic encoding-decoding structure. Its encoder and decoder correspond one-to-one, both have the same spatial size and number of channels. The innovation of semantic segmentation network mainly comes from the continuous optimization of the encoder and decoder structure and the improvement of its efficiency. In particular, the effect and complexity of the decoder are very large for the result of the entire segmentation network.

2.3. APPLICATION OF DL IN IMAGE SEGMENTATION

Deep learning has been driving the development of the image field, including image classification and image segmentation. Image segmentation is different from image classification. Image classification only shows which class or classes the entire image belongs to, while image segmentation needs to identify the information of each pixel in the image.

Fig3. The structure of the fully convolutional network (FCN)

The study of the fully convolutional network [32] for semantic segmentation was the first article that applied deep learning to image segmentation and achieved outstanding results. After that, many models of image segmentation have borrowed from FCN. This network is inspired by the VGG network structure. FCN does not require the size of the input image. It is a novel point that all layers are fully convolutional. However, the result obtained after FCN segmentation is still not fine enough, relatively blurry and smooth. It is not sensitive to details in the image. Later, Ronneberger et al. [33] proposed U-Net for the lack of training images in biomedical images.

This network has two advantages: first, the output result can locate the position of the target category. Second, the input training data are patches, which is equivalent to data augmentation and solves the problem about a small number of biomedical images. SegNet [31] builds an encoder-decoder symmetric structure based on the semantic segmentation task of FCN to achieve end-to-end pixel-level image segmentation. Zhao et al. [34] proposed the pyramid scene parsing network (PSPNet). Through the pyramid pool module and the proposed pyramid scene parsing network, it aggregates the ability to mine global context information based on the context information of different regions. Another important segmentation model is Mask R-CNN. Faster R-CNN [35] is a popular target detection framework, and Mask R-CNN extends it to an instance segmentation framework. These are used for image segmentation very classic network model. Furthermore, there are other methods of construction, such as those done by RNN (recurrent neural network), and the more meaningful weakly-supervised methods.

2.3.1. MEDICAL IMAGE SEGMENTATION BASED ON DL

When performing image segmentation operations, convolutional neural networks have excellent feature extraction capabilities and good feature expression capabilities. It do not require manual extraction of image features or excessive preprocessing of images. Therefore, CNN has been used in medical image segmentation in recent years. It has achieved great success in the field and auxiliary diagnosis. This section summarizes the existing classic research results and divides the existing deep-learning-based medical image segmentation methods into three categories: FCN, U-Net, and GAN. Each category is separately introduced. The advantages and disadvantages of each method are compared.

2.4. FULLY CONVOLUTIONAL NEURAL NETWORKS

FCN is the pioneering work of the most successful and advanced deep learning technology for semantic segmentation. In this section, the advantages and limitations of FCN networks are introduced. The variants of FCN and its applications are presented. For general classification CNN networks, such as VGG and ResNet, some fully connected layers are added at the end of the network. The category probability information can be obtained after the softmax layer, but this probability information is one-dimensional. That is, only the category of the entire image can be identified, not the category of each pixel. So, this fully connected method is not suitable for image segmentation. Long et al. [32] proposed the fully convolutional network in response to the above problems. In the usual CNN structure, the first five layers are convolutional layers.

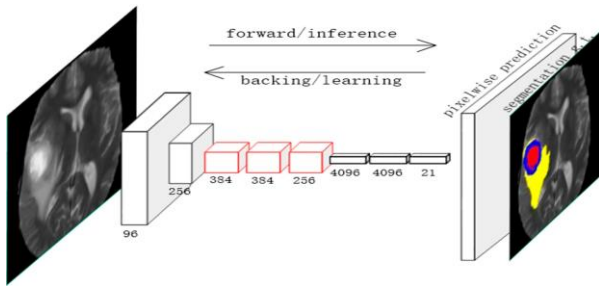


Fig3.The structure of the fully convolutional network (FCN)

The sixth and seventh layers are fully connected layers with a length of 4096 (one-dimensional vector). The eighth layer is a fully connected layer with a length of 1000, corresponding to the probability of 1000 categories. FCN changes the three layers from layer 5 to 7 into convolution layers whose convolution kernel sizes are 7 7, 1 1, and 1 1, so as to obtain a two-dimensional feature map of each pixel. Then it is followed by a softmax layer to obtain the classification information of each pixel. The fully convolutional network can accept input images of any size. FCN uses the deconvolution layer to up- sample the feature map of the last convolution layer and restore it to the same size of the input image. Thus, a prediction can be generated for each pixel, while retaining the spatial information in the original input image. Finally, pixel-by-pixel classification is performed on the upsampled feature map to complete the final image segmentation. According to the magnification of upsampling, it is divided into FCN-32s, FCN-16s, and FCN-8s.

2.4.1. DEEPLAB V1

However, the shortcomings of FCN are also very prominent. First, the results of its upsampling are relatively fuzzy and insensitive to the details of the image, resulting in the segmentation results not being fine enough. Second, the idea of segmentation is essentially to classify each pixel without full consideration. The relationship between pixels and pixels lacks spatial consistency. In order to get a denser score map in FCN, the authors added padding to the first convolutional layer. The padding size is equal to 100, which will bring a lot of noise. Chen et al. [36] proposed DeepLab v1, the size of the pooled image is not reduced and the score map result obtained is denser than that of FCN.

DeepLab v1 is rewritten based on the VGG-16 network, removing the last fully connected layer of the VGG network and using full convolution instead because using too many pooling layers will result in the feature layer size being too small. The features contained are too sparse, which is not conducive to semantic segmentation. The authors removed the last two pooling layers and added atrous convolution. Compared with traditional convolution, the receptive field can be expanded without increasing the amount of calculation and the density of features can be increased. Finally, DeepLab v1 uses conditional random field (CRF) [37] to improve the accuracy of segmentation boundaries.

2.4.2. DEEPLAB V2

DeepLab v2 is an improvement based on DeepLab v1. DeepLab v2 [38] solved the difficulty of segmentation caused by differences of the same object scale in the same image. When the same thing has different sizes in the same image or different images, the traditional method is to force the image to the same size by resizing. But this will cause some features to be distorted or disappear. In DeepLab v2, authors switched to the more complex and expressive ResNet-101 network. The continuous pooling and downsampling of deep convolutional neural network (DCNN) cause the resolution to decrease. DeepLab v2 removes downsampling in the last few maximum pooling layers. It instead uses atrous convolution to calculate feature maps with a higher sampling density. They also removed the fully connected layer in the network and replaced it with a fully convolutional layer, using a conditional random field to improve accuracy of the segmentation boundary. In addition, DeepLab v2 uses a fully connected CRF. The deep neural network has a high accuracy rate for classification, which means that it has obvious advantages in high-level semantics. However, pixel-level classification belongs to low-level semantic information, so it appears very vague in local details. Therefore, the author hopes to optimize the detailed information through CRF.

2.4.3. DEEPLAB V3 AND DEEPLAB V3+

DeepLab v3 [39] continued to use the ResNet-101 network. Aiming at the problem of multiscale target segmentation, a cascaded or

parallel atrous convolution module is designed. It adopted multiple atrous rates to capture multiscale context. In addition, the authors added the previously proposed ASPP module. This module detects convolutional features on multiple scales and uses image-level features to encode the global context to further improve performance. Finally, DeepLab v3 began to remove CRF. The experimental results showed that the model has a significant improvement over the previous DeepLab version. However, DeepLab v3 also has some shortcomings. For example, the zooming effect of output image is not good and there is too little information. DeepLab v3+ [40] extended DeepLab v3. It added a simple and effective decoder module to refine the segmentation results, especially the segmentation results along target boundary. In order to improve the effect of the output image, DeepLab v3+ used a feature map of the middle layer to enlarge the output image. The Xception model is used in the semantic segmentation task. The depthwise separable convolution is used in ASPP and the decoding module to improve the running speed and robustness of the encoder-decoder network.

2.4.4. SegNet

SegNet [31] builds an encoder-decoder symmetric structure based on the semantic segmentation task of FCN to achieve end-to-end pixel-level image segmentation. The network is mainly composed of two parts: the encoder and the decoder. The encoder is a network model that continues to use VGG16, mainly for analyzing object information. The decoder corresponds the parsed information into the final image form, that is, each pixel is represented by the color or label corresponding to its object information. The novelty lies in the way that the decoder upsamples its input feature map with lower resolution.

The difference of SegNet is that decoder uses a larger pooling index (position) transmitted from the encoder to nonlinearly upsample its input, so that upsampling does not require learning and a sparse features map is generated. Then, a trainable convolution kernel is used for convolution operation to generate a dense feature map. When feature maps are restored to original resolution, they are sent to the softmax classifier for pixel-level classification. This helps maintain integrity of high-frequency information, improves edge characterization, and reduces training parameters, but, when depooling low-resolution feature maps, it will also ignore adjacent information.

2.4.5. OTHER FCN STRUCTURES

Zhou et al. [41] used FCN in a 2.5D approach for the segmentation of 19 organs in 3D CT images. This technology uses a three-dimensional volume two-dimensional slice for pixel-to-label training, and designs a separate FCN (three FCNs in total) for each two-dimensional profile. Finally, the segmentation result of each pixel is merged with results of other FCNs to obtain final segmentation output. The accuracy of this technology on large organs such as the liver is higher than that of small organs such as the pancreas. Christ et al. [42] proposed superimposing a series of FCNs. Each model using context features extracted from the prediction map of the previous model can improve accuracy of segmentation. This method is called cascaded FCN (CFCN). Zhou et al. [43] proposed the application of focal loss on FCN to reduce number of false positives in medical images due to imbalance in the ratio of background and foreground pixels

2.5. U Net

2.5.1. 2D U-Net

Based on FCN, Ronneberger et al. [33] designed a U-Net network for biomedical images, which was widely used in medical image segmentation after it was proposed. Due to its excellent performance, U-Net and its variants have been widely used in various sub-fields of computer vision (CV). This approach was presented at the 2015 MICCAI conference and has been cited more than 4000 times. So far, U-Net has had many variants. There are many new design methods of convolutional neural network. But many of them still cited the core idea of U-Net, adding new modules or integrating other design concepts.

U-Net network is composed of U channel and skip-connection. The U channel is similar to the encoder-decoder structure of SegNet. The encoder has four submodules, each of which contains two convolutional layers. After each submodule, there is a max pool to realize down sampling. The decoder contains four submodules. The resolution is increased successively by upsampling. Then it gives predictions for each pixel. The network structure is shown in Figure 4. The input is 572 572, and the output is 388 388. The output is smaller than the input mainly because of the need for segmentation in the medical field, which is more accurate. It can be seen from the figure that this network has no fully connected layer, only convolution and downsampling.

The network also uses a skip connection to connect the upsampling result to the output of submodule with the same resolution in the encoder as the input of next submodule in the decoder. The reason why U-Net is suitable for medical image segmentation is that its structure can simultaneously combine low-level and high-level information. The low-level information helps to improve accuracy. The high-level information helps to extract complex features.

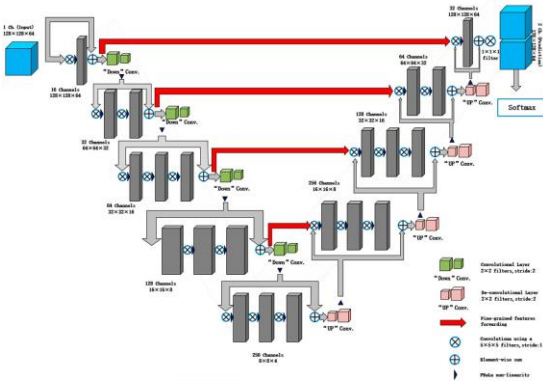


Fig 4. The structure of the U-Net

2.5.2. 3D U-Net

The improvement of U-Net has become a research hotspot in medical image segmentation. Many variants have been developed on this basis. Çiçek et al. [44] proposed a 3D U-Net model. This model aims to make the U-Net structure have richer spatial information. Its network structure is shown in Figure 5. The network structure is similar to U-Net, with one encoding path and one decoding path. Each path has four resolution levels. Each layer in the encoding path contains two 3x3 convolutions, followed by a ReLU layer. It uses a maximum pooling layer to reduce dimensionality.

The network realizes 3D image segmentation by inputting a continuous 2D slice sequence of 3D images. This network can not only train on a sparsely labeled data set and predict other unlabeled places on this data set, but also train on multiple sparsely labeled data set and then predict new data. Compared with U-Net input, the input is a stereo image (132 132 116) and it has three channels. The output image size is 44 44 28. 3D U-Net retains the excellent original features of FCN and U-Net. Its advent is of great help to volumetric images.

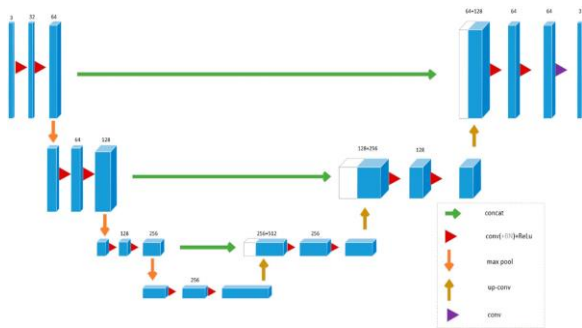


Fig 5. The structure of the 3D U-Net

2.5.3. V-Net

Milletari et al. proposed a 3D deformation structure V-Net of the U-Net network structure. Its network structure is shown in Figure 6. The V-Net structure uses the Dice coefficient loss function instead of traditional cross-entropy loss function. It uses a 3D convolution kernel to convolve image and reduces the channel dimension through a 1 1 1 convolution kernel. On the left side of the network is a gradually compressed path, which is divided into many stages. Each stage contains one to three convolutional layers. In order to make each stage learn a parameter function, the input and output of each stage are added to obtain learning of residual function. The size of the convolution kernel used in each stage of the convolution operation is 5 5 5. The convolution operation is used to extract features of data, while, at the same time, at the end of each “stage”, through the appropriate step size, the resolution of the data is reduced.

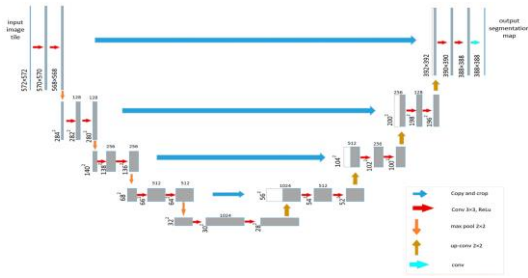


Fig 6. The structure of the V-Net

On the right side of the network is a gradually decompressed path. It extract features and expand the spatial support of lower resolution feature maps to collect and combine necessary information to output dual-channel volume segmentation. The final output size of network is consistent with the original input size.

2.5.4. OTHER U-NET STRUCTURES

Res-UNet (Weighted Res-UNet) [46] and H-DenseUNet (hybrid densely connected UNet) [47] are inspired by residual connections and dense connections, respectively. Each submodule of U-Net is replaced with a residual connection and dense connection. Res- UNet is used for image segmentation about retinal blood vessels. In the segmentation of retinal vessels, we often encounter problems of missing small blood vessels and poor segmentation of optic disc. The structure of retinal blood vessels is similar to the bifurcation structure of trees. When blood vessels are too thin to detect, this structure is difficult to maintain. For these challenges, Xiao et al. proposed a weighted Res-UNet. Based on the original U-Net model, a weighted attention mechanism is added. This allows the model to learn more for distinguish characteristics of blood vessels and nonvascular pixels, and to better maintain retinal vessel tree structure. H-DenseUNet is used to segment liver and liver tumor from the contrast-enhanced CT volumes.

The network takes each 3D input and transforms the 3D volume into 2D adjacent slices through the transformation processing function F proposed in the article. Then these 2D slices are sent to 2D DenseUNet to extract the intraslice features. The original 3D input and predicted result after 2D DenseUNet conversion are concat sent to 3D network for extracting interslice features. Finally, the two features are fused and result is predicted through the HFF layer. Ibtehaz et al. [48] proposed MultiResUNet that based on probable scopes for improvement to analyze the U-Net model architecture. The authors proposed a MultiRes block to replace sequence of two convolutional layers. In addition to introduction of the MultiRes block, the common shortcut connection is replaced with proposed Res path. Finally, the authors conducted experiments on public medical image data sets of different modes. The results showed that MultiResUNet has a high accuracy rate. Wang et al. [50] proposed a new U-Net model based on self- attention, called nonlocal U-Nets. A new up/down sampling method is proposed: global aggregation block, which combines self-attention and up/down sampling. It considers the full image information while up/down sampling, so as to obtain a more accurate segmentation image while reducing parameters.

2.6. GENERATIVE ADVERSARIAL NETWORK

A new method of training generative models to generate adversarial networks has recently been introduced. Goodfellow et al. [51] proposed an adversarial method in 2014 to learn a deep generative model, GAN. Its structure is shown in the Figure 7 and consists of two parts. The first part is the generation network, which receives a random noise z (random number) and generates an image through this noise.

The second part is to fight against the network, which is used to judge whether an image is “real”. Its input parameter is x (an image), and output $D(x)$ represents the probability that x is a real image. Simply put, it is through training to make two networks compete with each other. Generation network generates fake data, and the adversarial network uses a discriminator to determine authenticity. Finally, it is hoped that data generated by the generator can be fake.

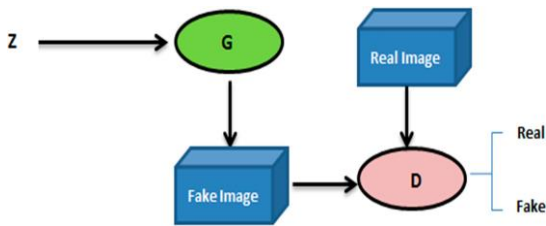


Fig7. The structure of the GAN.

2.6.1. FIRST GAN FOR SEGMENTATION

Combining the requirements of semantic segmentation and characteristics of GAN, Luc et al. [52] trained a convolutional semantic segmentation network and an adversarial network. This paper was the first time that GAN ideas were applied to semantic segmentation. The loss function of this network is:

$$n=1$$

$$l(\theta_s, \theta_a) = \sum_N \text{lmce}(s(x_n), y_n) - \lambda [\text{lbce}(a(x_n), 1) + \text{lbce}(a(x_n), 0)] \quad (2)$$

Among them, θ_s and θ_a represent parameters of the segmentation model and adversarial model respectively. N is the size of data set. x_n are training images and corresponding label maps y_n . $a(x, y)$ is the scalar probability of the ground truth label map y being x predicted by adversarial model. So, $s(\cdot)$ is a label map produced by the segmentation model. lbce and lmce are binary and multiclass cross-entropy losses, respectively. Segmentor is a traditional CNN-based segmentation network. Segmentor is a traditional CNN-based segmentation network, which attempts to generate a segmentation map that is close to ground truth so that it looks more realistic. The adversarial network is the discriminator in GAN. The training process is classic game idea, which mutually improves the network's ability to improve segmentation accuracy and discrimination ability.

2.6.2. SEGMENTATION ADVERSARIAL NETWORK (SegAN)

Xue et al. [53] proposed the U-Net structure as the generator of GAN, called segmentation adversarial network (SegAN). For medical image segmentation, U-Net cannot effectively solve the problem of unbalanced pixel categories in the image. Based on the above problem, authors designed a new segmentation network based on the ideas of GAN, and proposed a multiscale L1 loss to optimize the segmentation network. Its network structure is divided into two parts: segmentor network S and critic network C . Segmentor network S is a common U-Net structure. We use the convolutional layer with kernel size 4×4 and stride 2 for downsampling, and perform upsampling by image resize layer with a factor of 2 and convolutional layer with kernel size 3×3 stride 1. The experiment is on the BRATS (brain tumor segmentation) brain tumor segmentation data set is more effective and stable for segmentation task. Compared with single-scale loss function, the multiscale loss function multiscale L1 loss proposed by the authors optimizes the entire segmentation network.

2.6.3. STRUCTURE CORRECTION ADVERSARIAL NETWORK (SCAN)

Chest X-ray (CXR) is the most common X-ray used to diagnose various cardiopulmonary abnormalities in daily clinical practice. Due to the low cost and low dose radiation of CXR, it accounts for more than 55% of the total number of medical images. Therefore, it is important to develop computer-aided detection methods that support chest X-rays to support clinicians. Dai et al. [54] proposed a structure correction confrontation network (SCAN) to segment the lung field and heart in CXR images. This network adopted idea that Luc et al. first used GAN for image segmentation. The difference is that both the segmentation network and discriminant network use a fully convolutional network. During the training process, the critic network learned to distinguish ground truth organ annotations from a mask synthesized by the segmentation network. Through this confrontation process, the critic network learns higher-order structures and instructs the segmentation model to achieve realistic segmentation results. In addition, SCAN simplified the downsampling module based on the particularity of CXR images.

2.6.4. PROJECTIVE ADVERSARIAL NETWORK (PAN)

Three-dimensional medical image segmentation has always been a problem to be solved. Khosravan et al. [55] proposed a new segmentation network PAN to capture 3D semantics in an efficient and computationally efficient way. PAN integrates high-level 3D information through 2D projection, without relying on 3D images or enhancing the complexity of segmentation. The network backbone

is a segmentor and two adversarial networks. The segmentor contains 10 convolution layers in the encoder and 10 convolution layers in the decoder. The input of segmentor is a two-dimensional grayscale image. The output is a pixel-level probability map. The goal of designing adversarial networks is to compensate for missing global relations and correct the high-order inconsistencies caused by the loss of a single pixel. An adversarial signal is generated by these networks and applies it to the segmentor as part of the overall loss function. The adversarial network is only used in the training phase to improve performance of the segmentor without increasing its complexity.

2.6.5. DISTRIBUTED ASYNCHRONIZED DISCRIMINATOR GAN (AsynDGAN)

GAN can not only improve performance of medical image segmentation, but also contribute to data processing of medical image segmentation. The privacy of medical data is a very important issue, which leads to very few medical data sets. However, training a successful deep learning algorithm for medical image segmentation requires sufficient data. Data enhancement can alleviate this problem slightly. We can use GAN-based data enhancement as a data expansion method for medical image segmentation. In CVPR 2020, Chang et al. [56] proposes a data privacy-preserving and communication efficient distributed GAN learning framework named distributed asynchronized discriminator GAN (AsynDGAN). AsynDGAN is composed of a central generator and multiple distributed discriminators located in different medical entities. The central generator accepts the input of a specific task and generates a composite image to fool the discriminator. AsynDGAN does not need to share data, protect data security, and achieve a distributed GAN learning framework for efficient communication. It realizes the use of a distributed discriminator to train a central generator. The generated data can be used for segmentation model training, which improves segmentation accuracy.

2.6.6. OTHER GAN STRUCTURES

Zhao et al. [57] proposed Deep-supGAN to map the 3D MR data of the head to its CT image to facilitate segmentation of craniomaxillofacial bony structure. In order to obtain better conversion results, they proposed a deep-supervision discriminator, which uses the feature representation extracted by the pretrained VGG-16 model to distinguish between real and synthetic CT images. It provides gradient updates to the generator. The first block in the structure is used to generate high-quality CT images from MRI. In the case of segmenting 3D multimodal medical images, such as the PAN mentioned earlier there are often very few label examples used for training, resulting in insufficient model training. Using the application of antagonistic learning in semisupervised segmentation, Arnab et al. [58] proposed to use generative adversarial learning for a few-shot 3D multimodal medical image segmentation. Based on the advantages about the combination of adversarial learning and semisupervised segmentation, a new method of generating adversarial networks is used to train segmentation models with labeled and unlabeled images. Compared with the advanced segmentation network trained in a fully supervised manner, the performance of this network is greatly improved. It is worth studying to train an effective segmentation model using unannotated images.

In addition to segmentation, the application of generative adversarial networks in medical images also plays an important role in image enhancement. In the training of medical image segmentation model, the model is overfitted due to the insufficient data set. This problem is very common in medical image analysis. A solution to insufficient training data set is data augmentation. The GAN-based data enhancement technology for segmentation tasks is widely used in different medical images. Conditional GANs (cGAN) [61] and CycleGANs [62] have been used in various ways to synthesize certain types of medical images. Wolterink et al. [65] used CycleGAN to convert 2D MR images into CT images. No matching image pairs are required, and training brings better results.

2. THE SEGMENTATION METHOD FOR VARIOUS HUMAN ORGAN AREA

The human body has multiple organs and tissues. Different parts have their specificities. For example, the segmentation area for diagnosing brain tumors and lung nodules is relatively large, while retinal blood images require segmentation of blood vessels. The latter requires higher segmentation accuracy. Researchers extract ideas from these messages and design segmentation algorithms for different organs to improve accuracy of segmentation. The best way to segment different organs will be introduced below. Through reading the literature, we summarized the segmentation methods of brain, eyes, chest, abdomen, heart and other parts besides, and drew Tables 1–6.

3.1. BRAIN

The analysis of brain-related diseases generally requires MRI. Brain imaging analysis is widely used to study brain diseases such as Alzheimer's disease [66], epilepsy, schizophrenia, multiple sclerosis, cancer, and neurodegenerative diseases. Myronenko et al. [67]

proposed a deep learning network 3D MRI brain tumor segmentation based on asymmetric FCN and combined with residual learning. It won the first place in the 2018 challenge. Nie et al. [68] obtained T1, T2 and diffusion weighted modal neural images of 11 healthy infants.

The authors conducted network optimization by integrating contextual semantic information and fusing features of different scales, and segmented multimodal brain MRI images using 3D FCN. Wang et al. [69] proposed a CRF-based edge-sensing FCN, which achieved more accurate edge segmentation by adding edge information into the loss function. The accuracy of the model was up to 87.31%, far higher than that of FCN-8S and other basic semantic segmentation networks. Borne et al. [70] selected 62 healthy brain images from different heterogeneous databases as the training set, and segmented them using 3D U-Net. The result was 85% correct. Casamitjana et al. [71] proposed the cascaded V-Net segmentation of brain tumor, the segmentation of entire tumor and the division of different tumor regions. There are a lot of segmentations using GAN. For example, Moeskops et al. [72] used adversarial training to improve the segmentation performance of brain MRI in fully convolutional and a network structure with dilated convolutions.

Rezaei et al. [73] used cGAN to train a semantic segmentation convolutional neural network, which has a superior ability for brain tumor segmentation. Focusing on the segmentation task of MRI brain tumors, Giacomello et al. [74] proposed SegAN-CAT, a deep learning architecture based on a generative adversarial network. They apply a trained model to different modalities through transfer learning. SegAN-CAT is different from SegAN in that the loss function is extended, a dice loss term is added. The input of the discriminator network is composed of MRI image stitching and segmentation. By training several brain tumor segmentation models on the BRATS 2015 and BRATS 2019 data sets for testing, SegAN-CAT has better performance than SegAN.

TABLE 1. Segmentation CNN-Based Methods for the Brain.

Reference	Object	Modalities	Network Type	Data Set
Myronenko et al. [67]	Brain	MRI	FCN	BRATS2018
Nie et al. [68]	Brain	MRI	3D FCN	Infant brain images
Wang et al. [69]	Brain	MRI	FCN	ANDI data set and NITRC data set
Borne et al. [70]	Brain	MRI	3D U-Net	62 healthy brain images
Casamitjana et al. [71]	Brain	MRI	V-Net	BRATS2017
Moeskops et al. [72]	Brain	MRI	GAN	MRBrainS13
Rezaei et al. [73]	Brain	MRI	cGAN	BRATS 2017
Giacomello et al. [74]	Brain	MRI	SegAN-CAT	BRATS2015, BRATS2019

TABLE 2. Segmentation CNN-Based Methods for The Eye.

Reference	Object	Modalities	Network Type	Data Set
Leopold et al. [75]	Eye	Funduscopy	PixelBNN	DRIVE, STARE, CHASE DB1
Zhang et al. [76]	Eye	Funduscopy	U-Net	DRIVE, STARE, CHASE DB1
Jaemin et al. [77]	Eye	Funduscopy	GAN	DRIVE, STARE
Edupuganti et al. [78]	Eye	Funduscopy	FCN	Drishti-GS data set
Shankaranarayana et al. [79]	Eye	Funduscopy	FCN	RIM-ONE
Xiao et al. [46]	Eye	Funduscopy	Res-U-Net	DRIVE

3.2. EYE

Retinal blood image segmentation is a challenging subject in the research of retinal pathology. The problem of missing small and weak blood vessels or oversegmentation has not been solved. Leopold et al. [75] proposed a fast architecture for retinal vessel segmentation,

a fully-residual autoencoder batch normalization network (PixelBNN). It is based on U-Net, PixelCNN. It also uses skip connections and batch normalization within FCN. Finally, the model is trained, tested and cross-tested on the DRIVE (Digital Retinal Images for Vessel Extraction), STARE (Structured Analysis of the Retina) and CHASEDB1(Child Heart Health Study in England) retinal blood vessel segmentation data sets. The test time and performance are relatively good.

Zhang et al. [76] used U-Net with residual connection to detect vessels, and introduced an edge-sensing mechanism to add additional labels to the boundary area to improve accuracy. They conducted experiments on STARE, CHASEDB1 and DRIVE. Jaemin et al. [77] proposed a method that uses generative adversarial training to generate precise segmentation of retinal blood vessels. This method proposes that the segmented blood vessels are clear and sharp, with fewer false positives. It finally achieved the most advanced performance on the two public data sets DRIVE and STARE. In Section 4, we introduced Res-UNet, which can also be used for retinal vessel segmentation.

3.3. CHEST

Because chest X-ray examination is quick and easy, it is the most common medical image in medicine. Chest X-rays use very small doses of radiation to produce images of the chest. In chest X-rays, we can realize the segmentation of the lung area [80]. It can be used to help diagnose and monitor various lung diseases, such as pneumonia and lung cancer. The SCAN mentioned in Section 4 is used for lung fields and the heart segmentation in chest X-ray. The proposed framework was extensively evaluated on the JSRT (Japanese Society of Radiological Technology) and Montgomery data sets, and it was proved that this method can perform high-precision and realistic segmentation of lung fields and heart in CXR images.

Novikov et al. [81] made some modifications to U-Net for overfitting the model and the number of parameters, and proposed an all-convolutional modification of the original U-Net. By replacing the pool with strided convolutions to solve simplification problem of convolutional networks, the parameters are reduced by about ten times, while maintaining accuracy and achieving better results. The models are trained and tested on the JSRT database, and the performance exceeds expert observations of the lungs and heart. In CT and MRI image studies of the chest, Anthimopoulos et al. [82] used FCN with atrous convolution structure and multiscale feature fusion to segment lung parenchyma, healthy tissue, micronodules and honeycomb structures in lung CT images. Finally, it was verified on 172 high-resolution CT images collected from multiple medical institutions. A fully convolutional network was used to construct multiple shared representations between CT and MRI. Jue et al. [83] developed a learning method derived from cross-modality, using MR information derived from CT for hallucination MRI to improve CT segmentation.

TABLE 3. Segmentation CNN-Based Methods for The Chest.

Reference	Object	Modalities	Network Type	Data Set
Dai et al. [54]	Chest	CXR	SCAN	JSRT, Montgomery
Novikov et al. [81]	Chest	CXR	U-Net	JSRT
Anthimopoulos et al. [82]	Chest	CT	FCN	A data set of 172 sparsely annotated CT scans
Jue et al. [83]	Chest	CT, MRI	U-Net, dense-FCN	TCIA, NSCLC

TABLE 4. Segmentation CNN-Based Methods for The Abdomen.

Reference	Object	Modalities	Network Type	Data Set
Christ et al. [84]	Liver	CT, MRI	FCN	3DIRCADb and other
Han et al. [85]	Liver	CT	DCNN	LITS
Oktaç et al. [49]	Pancreas	CT	Attention U-Net	TCIA
Yang et al. [60]	Liver	CT	DI2IN-AN	1000 CT volumes
Huo et al. [86]	Spleen	MRI	SSNet	60 clinically acquired abdominal MRI scans

TABLE 5. Segmentation CNN-Based Methods in Cardiology.

Reference	Object	Modalities	Network Type	Data Set
Tran et al. [87]	Left and right ventricles	MRI	FCN	SCD, LVSC, RVSC
Xu et al. [88]	The whole heart	CT	CFUN	MM-WHS2017
Dong et al. [89]	Left ventricles	3D echocardiography	VoxelAtlasGAN	60 subjects on 3D echocardiography
Zhang et al. [90]	Cardiac	MRI	LU-Net	ACDC Stacom 2017
Ye et al. [91]	The whole heart	CT	3D U-Net	MICCAI 2017 whole-heart
Xia et al. [92]	Left atrium	MRI	3D U-Net	LASC2018

ABDOMEN

In CT and MRI abdomen images, we can segment the liver, spleen, kidney and other organs. Christ et al. [84] proposed cascaded fully convolutional neural networks (CFCNs) to automatically segment liver and lesions in CT or MRI abdomen images. This network is composed of two FCNs cascaded. The first FCN segments the liver ROI area used as the input of the second FCN. The second FCN is only for lesions within the liver ROIs in the first FCN. The experiment was implemented on an abdominal CT data set comprising 100 hepatic tumor volumes and 3DIRCADb data set. Han et al. [85] developed a deep convolutional neural network method, which belongs to the category of “fully convolutional neural networks”. The DCNN model takes a bunch of adjacent slices as input and generates a segmentation map corresponding to the central slice, so it works in 2.5D. It is 2% to 3% higher than other models in the dice score indicator. It is essential in many clinical applications of liver segmentation in 3D medical images. GAN is also used more in the segmentation of organs about the abdomen.

Yang et al. [60] proposed a segmentation of liver method that using an adversarial image to image network (DI2IN-AN). The generator generates segmentation predictions. The discriminator classifies predictions and ground truth during the training process. When segmenting the spleen on an MRI image, the size and shape of the spleen cause vast false positive and false negative labeling. Huo et al. [86] proposed the splenomegaly segmentation network (SSNet) for this. The cGAN framework is introduced into SSNet. In order to reduce false negatives and false positives, the generator uses a global convolutional network (GCN), and Markovian discriminator (PatchGAN) is used to replace the general generator.

3.5. CARDIOLGY

The heart is an important organ in our body. However, various heart diseases also seriously threaten the lives of many people. It is necessary to realize automatic segmentation of the heart region to solve practical problems in the field of cardiac medical treatment. For the first time, Tran et al. [87] applied a fully convolutional neural network architecture to pixel classification for cardiac magnetic resonance imaging. The proposed FCN architecture achieves the most advanced semantic segmentation in short-axis cardiac MRI. The authors conducted experiments to segment the left and right ventricles on the SCD (Sunnybrook cardiac data), LVSC (Left ventricle segmentation challenge), and RVSC (Right Ventricle Segmentation Challenge) data sets. Xu et al. [88] combined Faster R-CNN with fast detection capabilities and 3D U-Net with powerful segmentation capabilities, and proposed a CFUN to obtain the results of the whole heart segmentation.

This is the first time that cGAN has been used for 3D left ventricle segmentation on echocardiography. Zhang et al. [90] proposed an improved U-Net named LU-Net, in order to solve the problem of U-Net’s low accuracy in cardiac ventricular segmentation. LU-Net has been improved in three aspects: the effectiveness of extracting original image features, the degree of pixel location information loss, and the traditional U-Net segmentation accuracy. In order to obtain a finer whole-heart segmentation, Ye et al. [91] proposed a new deep-supervised 3D U-Net, which is applied to the original network in multiple depths to better extract context information. Xia et al. [92] proposed a fully automated two-stage segmentation framework that included the first 3D U-Net for roughly locating the atrial center from downsampled images. The second 3D U-Net for accurately segmenting the atrial catheters in the original images at full resolution.

3.6. OTHER ORGANS AND LESION SEGMENTATION

CNN-based semantic segmentation networks also have important applications in other biomedical image segmentation fields [94,95]. Liu et al. [96] used SegNet structure as the core network to segment muscles, cartilages and bones from 100 groups of labeled knee

MRI images in the MICCAI Challenge data set, so as to provide rapid and accurate segmentation methods of cartilage and other tissues for clinical osteoarthritis research. In addition, SegNet is also used for cell segmentation under the microscope. Tran et al. [97] used the SegNet structure to segment red blood cells and white blood cells in microscopic blood smear images. Similarly, Han et al. [99] proposed the application of Spine-GAN to multiple tasks and multiple targets bone marrow segmentation. V-Net combines MRI images using different equipment to achieve an end-to-end prostate segmentation process. The network outputs segmentation results while calculating the prostate volume for subsequent clinical analysis. Rundo et al. [30] proposed to merge the squeeze-and-excitation (SE) blocks into U-Net as a new convolutional neural network, USE-Net. The introduction of this structure is expected to enhance the representation ability by modeling the channel dependence of convolutional features. The author conducted experiments on multiple heterogeneous MRI data sets of prostate.

The experiments show that the model enhances the segmentation performance and improves the generalization ability. Kohl et al. [100] proposed a fully convolutional network to detect aggressive prostate cancer. Different from the general FCN, the author first used an adversarial network to distinguish between expert annotations and generated annotations to train FCNs for semantic segmentation. Finally, MRI images of 152 patients were used to segment aggressive prostate cancer. Wang et al. [104] modified the proposed contour segmentation deep learning model by adopting an adversarial training strategy, and proposed the basal membrane segmentation method for the diagnosis of cervical cancer.

TABLE 6. Other Segmentation CNN-Based Methods.

Reference	Object	Modalities	Network Type	Data Set
Liu et al. [96]	Musculoskeletal	MRI	SegNet	MICCAI Challenge data set
Tran et al. [97]	Cell	Microscopic	SegNet	ALL-IDB1 database
Sekubovjina et al. [98]	Spines	CT	Btrfly Net	302 CT scans
Han et al. [99]	Spines	MRI	Spine-GAN	253 multicenter clinical patients
Milletari et al. [45]	Prostate	MRI	V-Net	PROMISE2012
Rundo et al. [30]	Prostate	MRI	USE-Net	three T2-weighted MRI data sets
kohl et al. [100]	Prostate	MRI	FCN	MRI images of 152 patients
Taha et al. [101]	Kidney	CT	Kid-Net	236 subjects
Izadi et al. [102]	Skin	Dermoscopy	GAN	DermoFit
Mirikharaji et al. [103]	Skin	Dermoscopy	FCN	ISBI 2017
Wang et al. [104]	Basal membrane	Histopathology	GAN	IPMCH

4. SEGMENTATION EVALUATION METRICS AND DATASETS

4.1. EVALUATION METRICS

Evaluating the quality of an algorithm requires a correct objective indicator. In medical segmentation algorithms, doctors' hand-drawn annotations are usually used as the gold standard (ground truth, GT for short). Other results of the algorithm segmentation are the prediction results (Rseg, SEG for short). The segmentation evaluation of medical images is divided into pixel-based and overlap-based methods.

Dice index: The dice coefficient is a function for evaluating similarity. It is usually used to calculate the similarity or overlap between two samples. It is also the most frequently used. Its value range is 0 to 1. The closer the value is to 1, the better the segmentation effect. Given two sets A and B, the metrics is defined as:

$$\text{Dice}(A, B) = 2 \frac{|A \cap B|}{|A| + |B|} \quad (3)$$

Jaccard index: Jaccard index is similar to the dice coefficient. Given two sets A and B, the metrics are defined as:

$$\text{Jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

Segmentation accuracy (SA): The area of accurate segmentation accounts for the percentage of the real area in the GT image. Among them, R_s represents the reference area of the segmented image manually drawn by the expert. T_s represents the real area of the image obtained by the algorithm segmentation. $R_s - T_s$ indicates the number of pixels that are incorrectly segmented.

$$SA = 1 - \frac{|R_s - T_s|}{R_s} \times 100\% \quad (5)$$

Over segmentation rate: The ratio of pixels that are divided into the reference area of the GT image is calculated as follows:

$$OR = \frac{O_s}{R_s} \quad (6)$$

The pixels in O_s appear in the actual segmented image, but do not appear in the theoretical segmented image R_s . R_s represents the reference area of the segmented image manually drawn by the expert.

Undersegmentation rate: The ratio of the segmentation result to the missing pixels in GT image. Calculated as follows:

$$UR = \frac{U_s}{R_s + O_s} \quad (7)$$

The pixels in U_s appear in the theoretical segmented image R_s , but do not appear in the actual segmented image. R_s , O_s have the same meaning as above.

Hausdorff distance: This describes a measure of the degree of similarity between two sets of points, that is, the distance between the two boundaries of ground truth and the segmentation result input to the network. Sensitive to the divided boundary.

$$H = \max \min(d(i, j)), \max \min(d(i, j)) \quad (8)$$

where, i and j are points belonging to different sets. d represents the distance between i and j .

4.2. DATASETS FOR MEDICAL IMAGE SEGMENTATION

For any model segmentation based on deep learning, it is crucial to collect enough data into the data set. The quality of the segmentation algorithm depends on the high-quality image data by the experts and the corresponding label-standardized data set, which enables fair comparison between systems. This section will introduce some public data sets frequently used in the field of medical image segmentation.

Medical segmentation decathlon (MSD): Simpson et al. [105] created a large, opensource, hand-annotated medical image data set of various anatomical parts. This data set can objectively evaluate general segmentation methods through comprehensive benchmarks, and make the access to medical image data public. The data set has a total of 2633 three-dimensional medical images, involving real clinical applications of multiple anatomical structures, multiple models, and multiple sources (or institutions). It is divided into ten categories:

1. Task01_BrainTumour: There are a total of 750, and the labels are divided into two categories: Glioma (necrotic/active tumor), edema. It is an MRI scan obtained in routine clinical practice.
2. Task02_Heart: There are a total of 30, and the label is the left atrium. These data come from the Left Atrial Segmentation Challenge (LASC). Images were obtained on a 1.5T Achieva scanner with voxel resolution 1.25 1.25 2.7 mm³.
3. Task03_Liver: There are 201 sheets in total, with labels divided into liver and tumors. The type of imaging is CT. The images were provided with an in-plane resolution of 0.5 to 1.0 mm, and slice thickness of 0.45 to 6.0 mm.
4. Task04_Hippocampus: There are a total of 394, and the labels are hippocampus, head and body. The type of imaging is MRI. The data set consisted of MRI acquired in 90 healthy adults and 105 adults with a nonaffective psychotic disorder.
5. Task05_Prostate: There are a total of 48, and the labels are: Prostate central gland, peripheral zone. The type of imaging is MRI. The prostate data set consisted of 48 multiparametric MRI studies provided by Radboud University (The Netherlands) reported in a previous segmentation study.
6. Task06_Lung: There are a total of 96, and the label is lung tumor. The type of imaging is CT. The lung data set was comprised of patients with non-small-cell lung cancer from Stanford University. The tumor region was denoted by an expert thoracic radiologist on a representative CT cross section using OsiriX.
7. Task07_Pancreas: There are a total of 420, with labels divided into pancreas and pancreatic mass (cyst or tumor). The type of imaging is CT. The pancreas data set consisted of patients whose pancreatic masses were removed.
8. Task08_HepaticVessel: There are a total of 443, and the labels is liver vessels. The type of imaging is CT. This second liver

data set consisted of patients with various primary and metastatic liver tumors.

9. Task09_Spleen: There are a total of 61, and the label is the spleen. The type of imaging is CT. The spleen data set comprised of patients undergoing chemotherapy treatment for liver metastases at Memorial Sloan Kettering Cancer Center.

10. Task10_Colon: There are a total of 190, and the label is colon cancer. The type of imaging is CT.

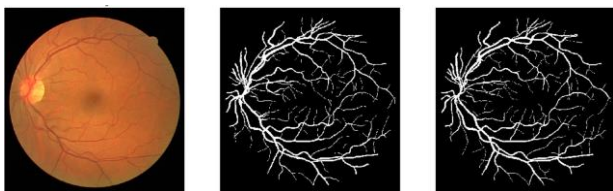
Segmentation in Chest Radiographs (SCR): All chest radiographs are taken from the JSRT database. The SCR database was created to simplify the comparative study of lung field, heart and clavicle segmentation in standard posterior chest radiographs [106]. All data in the database are manually segmented to provide reference standards. The image is scanned from film to 2048 2048 pixels, with a spatial resolution of 0.175 mm/pixel and a gray scale of 12 bits. Each of the 154 images have a lung nodule, and the other 93 images have no lung nodules. Brain tumor segmentation (BRATS): This data set is a brain tumor segmentation competition data set, which is combined with the MICCAI conference [107]. In order to evaluate the best brain tumor segmentation methods and compare different methods, it has been held every year since 2012. For this reason, the data set is published. There are five types of labels: healthy brain tissue, necrotic area, edema area, tumor enhancement and nonenhancement area. New training sets are added every year.

Digital database for screening mammography (DDSM): DDSM [108] is a resource used by the mammography image analysis research community and is widely used by researchers. The database contains approximately 2500 studies. Each study includes two images of each breast, as well as some relevant patient information and image information. Ischemic stroke lesion segmentation (ISLES): This provides MRI scans containing a large number of accurate stroke samples and related clinical parameters. This challenge is organized to evaluate stroke pathology and clinical outcome prediction in accurate MRI scan images.

Liver tumor segmentation (LiTS): These data and segmentations are provided by different clinical sites around the world for the segmentation of liver and liver tumors. The training data set contains 130 CT scans, and the test data set contains 70 CT scans [109]. Prostate MR image segmentation (PROMISE12): This data set is used for prostate segmentation. These data include patients with benign diseases (such as benign prostatic hyperplasia) and prostate cancer. These cases include a transversal T2-weighted MR image of the prostate.

Lung image database consortium image collection (LIDC-IDRI): The data set is composed of chest medical image files (such as CT, X-ray) and corresponding diagnosis result lesion labels. The purpose is to study early cancer detection in high-risk populations. A total of 1018 research examples are included. For the images in each example, four experienced thoracic radiologists performed a two-stage diagnosis and annotation [110].

Open Access Series of Imaging Studies (OASIS): This is a project aimed at enabling the scientific community to provide brain MRI data sets free of charge. A third generation has been released. OASIS-3 is a retrospective compilation of more than 1000 participants' data collected from several ongoing projects through WUSTL Knight ADRC over the past 30 years. OASIS-3 is a longitudinal neuroimaging, clinical, cognitive, and biomarker data set for normal aging and Alzheimer's disease. Participants included 609



cognitively normal adults and 489 people at various stages of cognitive decline, ages 42 to 95 [111].

Digital retinal images for vessel extraction (DRIVE): This data set is used to compare the segmentation of blood vessels in retinal images. The photos in the DRIVE database came from a diabetic retinopathy screening project in the Netherlands, and 40 photos were randomly selected. Among them, 33 cases had no signs of diabetic retinopathy and seven cases had signs of mild early diabetic retinopathy. Each image is captured with 768 584 pixels with 8 bits per color plane. The field of view of each image is circular with a diameter of approximately 540 pixels. Figure 8 is a sample of the DRIVE data set and its ground truth [112].

Mammographic Image Analysis Society (MIAS): MIAS is a breast cancer X-ray image database created by a British research organization in 1995. Each pixel has a grayscale of 8 bits. The MIAS database contains left and right breast images of 161 patients, with a total of 322 images, including 208 healthy images, 63 benign breast cancer and 51 malignant breast cancer images.

(a) (b) (c)
Fig 8. Digital retinal images for vessel extraction (DRIVE) sample diagram and manual labeling sample. (a) The blood vessels in retinal RGB image; (b) manual annotation 1 of sample; (c) manual annotation 2 of sample.

The boundary of the lesion area has also been calibrated by experts [113]. Sunnybrook cardiac data (SCD): It also known as the 2009 cardiac MR left ventricle segmentation challenge data, and consists of 45 cine-MRI images from a mixed of patients and pathologies: healthy, hypertrophy, heart failure with infarction and heart failure without infarction [114]. In addition to the several data sets

commonly used for medical image segmentation described above, there are also many competition data sets that verify the superiority of the algorithm provided by the famous medical image challenge competition.

Grand Challenges in Biomedical Image Analysis: It was designed to help people solve global health and development issues. It covers all challenges in the field of medical image analysis, including medical image processing.. Liver Tumor Segmentation Challenge: The purpose of this competition is to encourage researchers to study liver lesion segmentation methods. The data and slices of the challenge competition are provided by different clinical sites around the world. The training data set contains 130 CT scans, and the test data set contains 70 CT scans. 2019 Kidney and Kidney Tumor Segmentation Challenge (KiTS19): The KiTS19 challenge is the semantic segmentation of kidneys and kidney tumors in contrast-enhanced CT scans. The data set consisted of 300 patients with preoperative arterial-phase abdominal CTs annotated by experts. Two-hundred and ten (70%) of these were released as a training set and the remaining 90 (30%) were held out as a test set. Table 7 is the medical image data sets for segmentation.

TABLE 7. Medical Image Data Sets for Segmentation.

Data Set	Modalities	Objects	URL
MSD	MRI, CT	Various	http://medicaldecathlon.com/
BRATS	MRI	Brain	https://www.med.upenn.edu/sbia/brats2018/data.html
DDSM	Mammogra	Breast	http://www.eng.usf.edu/cvprg/Mammography/Database.html
ISLES	MRI	Brain	http://www.isles-challenge.org/
LITS	CT	Liver	https://competitions.codalab.org/competitions/17094
PROMISE12	MRI	Prostate	https://promise12.grand-challenge.org/
LIDC-IDRI	CT	Lung	https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI
OASIS	MRI, PET	Brain	https://www.oasis-brains.org/
DRIVE	Funduscopy	Eye	https://drive.grand-challenge.org/
STARE	Funduscopy	Eye	http://homes.esat.kuleuven.be/~mblaschk/projects/retina/
CHASEDB1	Funduscopy	Eye	https://blogs.kingston.ac.uk/retinal/chasedb1/
MIAS	X-ray	Breast	https://www.repository.cam.ac.uk/handle/1810/250394?show=full
SCD	MRI	Cardiac	http://www.cardiacatlas.org/studies/
SKI10	MRI	Knee	http://www.ski10.org/
HVSMR2018	CMR	Heart	http://segchd.csail.mit.edu/

5. CONCLUSIONS

Although research into medical image segmentation has made great progress, the effect of segmentation still cannot meet the needs of practical applications. The main reason is that the current medical image segmentation research still has the following difficulties and challenges:

1. Medical image segmentation is a cross-disciplinary field between these two disciplines span. Clinical medical pathology conditions are complex and diverse. However, artificial intelligence scientists do not understand clinical needs. Clinicians do not understand the specific technology of artificial intelligence. As a result, artificial intelligence cannot well meet the specific clinical needs. In order to promote the application of artificial intelligence in the medical field, extensive cooperation between clinicians and machine learning scientists should be strengthened. This cooperation will solve the problem that machine learning researchers cannot obtain medical data. It can also help machine learning researchers develop deep learning algorithms more in line with clinical needs and apply them to computer-aided diagnosis equipment, thereby improving diagnosis efficiency and accuracy.
2. Medical images are different from natural images. There are differences between different medical images. This difference also affects the adaptability of the deep learning model during segmentation. The noise and artifacts of medical images are also a major problem in data preprocessing.
3. Limitations of existing medical image data sets. The existing medical image data sets are small in scale. The training of deep learning algorithms requires a large amount of data set support, which leads to the problem of overfitting in the training process of deep learning models. One way to solve the insufficient amount of training data is data enhancement, such as geometric transformation, color space enhancement. GAN uses original data to synthesize new data. Another method is based on a meta-learning model to study medical image segmentation under small sample conditions.
4. The deep learning model has its own flaws. It mainly focuses on three aspects: network structure design, 3D data segmentation model design and loss function design. The design of the network structure is worth exploring. 3D medical data can more accurately capture the geometric information of the target, which may be lost when the 3D data is sliced slice by slice. Therefore, a researchable direction is the design of 3D convolution models to process 3D medical image data.
5. The design of loss function has always been a difficult point in deep learning research.

FUTURE DIRECTIONS

For medical image segmentation, deep learning has performed very well. More and more new methods are used to continuously

improve the accuracy and robustness of segmentation. Diagnosing various diseases through artificial intelligence realizes the idea of sustainable medical treatment. It becomes a powerful tool for clinicians. But it is still an open problem, so we can expect a series of innovations and research results in the next few years.

Author Contributions: Methodology, X.L.; writing—original draft preparation, L.S.; writing—review and editing, S.L.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Hunan Province with No.2020JJ4434, Key Scientific Research Projects of Department of Education of Hunan Province with No.19A312; Hunan Provincial Science & Technology Project Foundation (2018TP1018, 2018RS3065); Scientific Research Fund of Hunan Provincial Education(14C0710).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

- [1] Lateef, F.; Ruichek, Y. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* 2019, 338, 321–348. [CrossRef]
- [2] Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 2017, 19, 221–248. [CrossRef]
- [3] Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, UK, 2016.
- [4] Almeida, G.; Tavares, J.M.R.S. Deep learning in radiation oncology treatment planning for prostate cancer: A systematic review. *J. Med. Syst.* 2020, 44, 1–15. [CrossRef]
- [5] Hesamian, M.H.; Jia, W.; He, X.; Kennedy, P. Deep learning techniques for medical image segmentation: Achievements and challenges. *J. Digit. Imaging* 2019, 32, 582–596. [CrossRef] [PubMed]
- [6] Altaf, F.; Islam, S.M.S.; Akhtar, N.; Nanjua, N.K. Going deep in medical image analysis: Concepts, methods, challenges, and future directions. *IEEE Access* 2019, 7, 99540–99572. [CrossRef]
- [7] Hu, P.; Cao, Y.; Wang, W.; Wei, B. Computer Assisted Three-Dimensional Reconstruction for Laparoscopic Resection in Adult Teratoma. *J. Med. Imaging Health Inform.* 2019, 9, 956–961. [CrossRef]
- [8] Ess, A.; Müller, T.; Grabner, H.; Van Gool, L. Segmentation-Based Urban Traffic Scene Understanding. *BMVC* 2009, 1, 2.
- [9] Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The kitti vision benchmark suite. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 16–12 June 2012; pp. 3354–3361.
- [10] Ma, Z.; Tavares, J.M.R.S.; Jorge, R.M.N. A review on the current segmentation algorithms for medical images. In *Proceedings of the 1st International Conference on Imaging Theory and Applications*, Lisbon, Portugal, 5–8 February 2009.
- [11] Ferreira, A.; Gentil, F.; Tavares, J.M.R.S. Segmentation algorithms for ear image data towards biomechanical studies. *Comput. Methods Biomech. Biomed. Eng.* 2014, 17, 888–904. [CrossRef]
- [12] Ma, Z.; Tavares, J.M.R.S.; Jorge, R.N.; Mascarenhas, T. A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Comput. Methods Biomech. Biomed. Eng.* 2010, 13, 235–246. [CrossRef]
- [13] Xu, A.; Wang, L.; Feng, S.; Qu, Y. Threshold-based level set method of image segmentation. In *Proceedings of the Third International Conference on Intelligent Networks and Intelligent Systems*, Shenyang, China, 1–3 November 2010; pp. 703–706.
- [14] Cigla, C.; Alatan, A.A. Region-based image segmentation via graph cuts. In *Proceedings of the 2008 15th IEEE International Conference on Image Processing*, San Diego, CA, USA, 12–15 October 2008; pp. 2272–2275.
- [15] Yu-Qian, Z.; Wei-Hua, G.; Zhen-Cheng, C.; Tang, J.-T.; Li, L.-Y. Medical images edge detection based on mathematical morphology. In *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, China, 17–18 January 2006; pp. 6492–6495.
- [16] He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
- [17] Lin, G.; Milan, A.; Shen, C.; Reid, I. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 1925–1934.
- [18] Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, Las Condes, Chile, 11–18 December 2015; pp. 1520–1528.
- [19] Zhu, X.J. *Semi-Supervised Learning Literature Survey*; University of Wisconsin: Madison, WI, USA, 2005.
- [20] Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *Proceedings of the European*

Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 818–833.

- [21] Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shshroudy, A.; Shuai, B.; Liu, I.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* 2018, 77, 354–377. [CrossRef]
- [22] Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 1962, 160, 106. [CrossRef] [PubMed]
- [24] Fukushima, K.; Miyake, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and Cooperation in Neural Nets*; Springer: Berlin, Germany, 1982; pp. 267–285.
- [25] Lécun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *IEEE* 1998, 86, 2278–2324. [CrossRef]
- [26] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 2012, 60, 1097–1105. [CrossRef]
- [27] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 16 June–1 July 2016; pp. 770–778.
- [28] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, arXiv:1409.1556.
- [29] Qiu, Z.; Yao, T.; Mei, T. Learning spatio-temporal representation with pseudo-3d residual networks. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017; pp. 5533–5541.
- [30] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 1–9
- [31] Rundo, L.; Han, C.; Nagano, Y.; Zhang, J.; Hataya, R.; Militello, C.; Tangherloni, A.; Nobile, M.S.; Ferreti, C.; Besozzi, D.; et al. USE-Net, Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing* 2019, 365, 31–43. [CrossRef]
- [32] Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 2481–2495. [CrossRef] [PubMed]
- [34] Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
- [35] Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 5–9 October 2015; pp. 234–241.
- [36] Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
- [37] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 2015, 39, 91–99. [CrossRef]
- [38] Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* 2014, arXiv:1412.7062.
- [39] Krähenbühl, P.; Koltun, V. Efficient inference in fully connected crfs with gaussian edge potentials. *Adv. Neural Inf. Process. Syst.* 2011, 24, 109–117.
- [41] Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 40, 834–848. [CrossRef]
- [42] Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* 2017, arXiv:1706.05587.
- [43] Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 8–14 September 2018; pp. 801–818.
- [44] Zhou, X.; Takayama, R.; Wang, S.; Hara, T.; Fujita, H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Med. Phys.* 2017, 44, 5221–5233. [CrossRef] [PubMed]
- [45] Christ, P.F.; Elshaer, M.E.A.; Ettliger, F.; Tatavarty, S.; Bickel, M.; Bilic, P.; Rempfler, M.; Armbruster, M.; Hoffman, F.; D'Anastasi, M.; et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Athens, Greece, 17–21 October 2016; pp. 415–423.

- [46] Zhou, X.Y.; Shen, M.; Riga, C.; Yang, G.-Z.; Lee, S.-L. Focal fcn: Towards small object segmentation with limited training data.
- [47] arXiv 2017, arXiv:1711.01506.
- [48] Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016; pp. 424–432.
- [49] Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
- [50] Xiao, X.; Lian, S.; Luo, Z.; Li, S. Weighted Res-UNet for high-quality retina vessel segmentation. In Proceedings of the 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018; pp. 327–331.
- [51] Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.-W.; Heng, P.A. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans. Med. Imaging* 2018, 37, 2663–2674. [CrossRef] [PubMed]
- [52] Ibtihaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation.
- [53] *Neural Netw.* 2020, 121, 74–87. [CrossRef] [PubMed]
- [54] Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. arXiv 2018, arXiv:1804.03999.
- [55] Wang, Z.; Zou, N.; Shen, D.; Ji, S. Non-Local U-Nets for Biomedical Image Segmentation. In Proceedings of the AAAI, New York, NY, USA, 7–12 February 2020; pp. 6315–6322.
- [56] Luc, P.; Couprie, C.; Chintala, S.; Verbeek, J. Semantic segmentation using adversarial networks. arXiv 2016, arXiv:1611.08408.
- [57] Xue, Y.; Xu, T.; Zhang, H.; Long, L.R.; Huang, X. SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation. *Neuroinformatics* 2018, 16, 383–392. [CrossRef]
- [58] Dai, W.; Dong, N.; Wang, Z.; Liang, X.; Zhang, H.; Xing, E.P. Scan: Structure correcting adversarial network for organ segmentation in chest x-rays. In *Mining Data for Financial Applications*; Springer: Cham, Switzerland, 2018; pp. 263–273.
- [59] Khosravan, N.; Mortazi, A.; Wallace, M.; Bagci, U. Pan: Projective adversarial network for medical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–18 October 2019; pp. 68–76.
- [60] Chang, Q.; Qu, H.; Zhang, Y.; Sabuncu, M.; Chen, C.; Zhang, T.; Metaxas, D.N. Synthetic Learning: Learn from Distributed Asynchronous Discriminator GAN Without Sharing Medical Image Data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 13856–13866.
- [61] Mondal, A.K.; Dolz, J.; Desrosiers, C. Few-shot 3d multi-modal medical image segmentation using generative adversarial learning.
- [62] arXiv 2018, arXiv:1810.12241.
- [63] Zhang, Y.; Yang, L.; Chen, J.; Fredericksen, M.; Hughes, D.P.; Chen, D.Z. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 10–14 September 2017; pp. 408–416.
- [64] Yang, D.; Xu, D.; Zhou, S.K.; Georgescu, B.; Chen, M.; Grbic, S.; Metaxas, D.; Comaniciu, D. Automatic liver segmentation using an adversarial image-to-image network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 10–14 September 2017; pp. 507–515.
- [65] Mirza, M.; Osindero, S. Conditional generative adversarial nets. arXiv 2014, arXiv:1411.1784.
- [66] Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
- [67] Bayramoglu, N.; Kaakinen, M.; Eklund, L.; Heikkila, J. Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 64–71.
- [68] Dar, S.U.H.; Yurt, M.; Karacan, L.; Erdem, A.; Erdem, E.; Çukur, T. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans. Med. Imaging* 2019, 38, 2375–2388. [CrossRef]