EAI Endorsed Transactions
on e-Learning

Research Article **EAI.EU**

Applications of Image Segmentation Techniques in Medical Images

Yang-yang Hou[1,*]

1 School of Physics and Information Engineering, Jiangsu Second Normal University, NanJing, China

Abstract

Image segmentation is an important research direction in medical image processing tasks, and it is also a challenging task in the field of computer vision. At present, there have been many image segmentation methods, including traditional segmentation methods and deep learning-based segmentation methods. Through the understanding and learning of the current situation in the field of medical image segmentation, this paper systematically combs it. Firstly, it briefly introduces the traditional image segmentation methods such as threshold method, region method and graph cut method, and focuses on the commonly used network architectures based on deep learning such as CNN, FCN, U-Net, SegNet, PSPNet, Mask R-CNN. At the same time, the application in medical image segmentation is expounded. Finally, the challenges and development opportunities of medical image segmentation technology based on deep learning are discussed.

Keywords: Medical images, Image segmentation, Deep learning, Neural networks

Received on 1 12 2023, accepted on [illegible] 06 2024, published on 19 07 2024

doi: 10.4[illegible]/eetel.444[illegible]

* Corresponding author. Email: hyy12[illegible]715@jssnu.edu.cn

1. Introduction

Image segmentation is a key process in computer vision that involves segmenting an image into different, meaningful regions or segments. The main goal is to group pixels with similar properties such as color, intensity, or texture into homogeneous regions [1]. This technique is fundamental in a variety of applications, enabling computers to interpret and analyze visual information more efficiently [2]. Image segmentation plays a crucial role in tasks such as object recognition, scene understanding, and medical image analysis [3].

At present, medical image processing is deeply valued at home and abroad. When doctors make a diagnosis, they only need to analyze part of the tissues or structures in the medical images, which are called regions of Interest (ROI). These rois usually correspond to different organs, pathologies or other biological structures [4]. The purpose of medical image segmentation is to segment the ROI in the image, remove useless information, to help doctors better observe the organizational structure of things, and to provide more accurate data support for later medical diagnosis.

So far, many medical image segmentation methods have been proposed at home and abroad, and the segmentation methods have experienced the evolution from classical image segmentation methods to improved classical image segmentation methods and then to medical image segmentation methods based on deep learning [5]. As shown in Fig. 1.

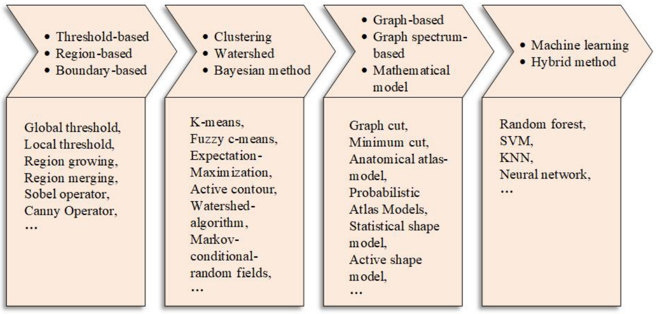

Fig. 1 Classification of image segmentation methods

In the second part of this paper, four classical medical image segmentation methods are briefly introduced, including threshold method, region growing method, edge segmentation method, and graph cut method. The third part focuses on the commonly used medical image segmentation methods based on deep learning, and summarizes the advantages and disadvantages of each network structure. In the fourth section, the specific application tasks of medical image segmentation are given, and in the fifth section, the challenges and opportunities in medical image segmentation are discussed.

2. Traditional medical image segmentation methods

2.1. Threshold segmentation method

The threshold method has the advantages of small amount of calculation, simple implementation and good stability, so it has become the most widely used segmentation technology in the field of medical image segmentation [6]. The basic principle is to divide the pixels into different gray levels of background area and target area by setting different thresholds, so it is more suitable for the image with large difference between background and target gray levels [7]. If the image has only two major categories of object and background, only one threshold value needs to be set, and the transformation of single threshold segmentation is shown in (1).

$$g(x,y)=\begin{cases}1, f(x,y)\geq T\\ 0, f(x,y)<T\end{cases} \quad (1)$$

Where, $f(x,y)$ represents the input image, $g(x,y)$ represents the output image, and T is the threshold. The core technology of threshold method is the selection of threshold, which will affect the final segmentation effect. The most common segmentation method is to calculate the segmentation threshold through the histogram. However, the threshold method is mainly for gray information, if the image background is complex or there are multiple targets, it is difficult to ensure the segmentation effect, so it is often used as preprocessing method [8]. General CT image processing, often used to the threshold segmentation method. According to their own experience, doctors continuously adjust the threshold for segmentation until they achieve the desired effect. For example, the threshold segmentation method for the detection of spiral CT images has high sensitivity, and the sensitivity can reach 96% when detecting tuberculosis lesions.

2.2. Region growing method

The region growing method is to cluster the regions formed by pixels according to their similar properties [9]. This method starts from individual pixels and gradually merges them to form the desired region. Region growing starts with selecting the seed pixel from the region to be segmented, and takes it as the starting point of growing. The eligible pixels in the neighborhood of the seed pixel are merged into the seed pixel to form a new seed pixel, and the merging is continued until no eligible pixels are found [10, 11]. One of the more famous region growing methods is watershed algorithm. The watershed algorithm was proposed by Vincent in 1991. This method simulates the geoglyph in geology. The gray value of the pixel in the image is simulated as the altitude, the local minimum value of the pixel gray value is simulated as the valley bottom, the local maximum value is simulated as the peak, and the boundary between the valley bottom is the watershed. The region segmentation method is simple to implement and can ensure the spatial continuity of the segmented image. It is suitable for segmenting continuous uniform small targets [12, 13]. Its disadvantages are that it needs human participation to select the appropriate seed points for each region, and the algorithm is sensitive to noise, and it is not suitable for the segmentation of large regions, which may lead to over-segmentation or under-segmentation.

2.3. Edge segmentation method

There are many researches on edge segmentation method, and it is also a widely used method. The image segmentation is completed by extracting the edge of the image segmentation target [14]. Edge segmentation methods include five types, which are pixel attributes, deformation templates, mathematical morphology, cost function and edge flow. There are also many existing edge segmentation operators, such as Sobel operator, Roberts operator, Canny operator, Prewitt gradient operator, second order difference operator and so on [15]. Due to the existence of these ready-made operators, the edge segmentation method is also relatively simple to calculate [16].

2.4. Graph cut method

Graph cut method is an image segmentation method based on graph theory, which realizes image segmentation by establishing a probabilistic undirected graph model. Such probabilistic undirected graphical models are also known as Markov random fields [17, 18]. Among the traditional image segmentation methods, the graph cut method is widely used in medical image segmentation because of its good results. The most representative method of graph cuts is graphcut [19]. The core idea is to build a weighted graph, and remove the edges with smaller weight as much as possible, so that the subgraphs are not connected. The graph cut method has high robustness and can also get good results in segmenting more complex images [20, 21]. However, it has high time complexity and space complexity, and is usually used in combination with other traditional segmentation methods.

3. Medical image segmentation methods based on deep learning

In recent years, with the rapid development of deep learning technology, many researchers have successfully applied it to medical image segmentation tasks [8]. Compared with traditional segmentation methods, deep neural networks are designed to imitate the human perception and learning process of things. Without selecting features in advance, deep neural networks can directly learn the relationship between input and output from a large amount of data, fully exploit image information, and achieve accurate and efficient medical image segmentation [4, 22]. Because of this, the intelligent processing and analysis of medical images using deep learning technology is not only beneficial to alleviate the imbalance of doctor-patient ratio and improve the efficiency of doctor diagnosis, but also of great significance to promote the development of computer-aided diagnosis.

At present, the representative classical model frameworks in medical image segmentation tasks are FCN, U-Net, SegNet, PSPNet and Mask R-CNN, and their algorithms are basically based on Convolutional Neural Network (CNN) as the underlying architecture [1].

3.1. Convolutional neural networks (CNN)

CNN is a multi-layer neural network with convolutional structure. In the late 1980s, LeCun et al [23]. proposed LeNet-5 based on the hierarchical neurocognitive machine [24-26], which is regarded as the classical CNN structure, as shown in Fig. 2. Convolutional Layer, Pooling Layers and Fully Connected Layers are the three basic structures that constitute CNN [27].

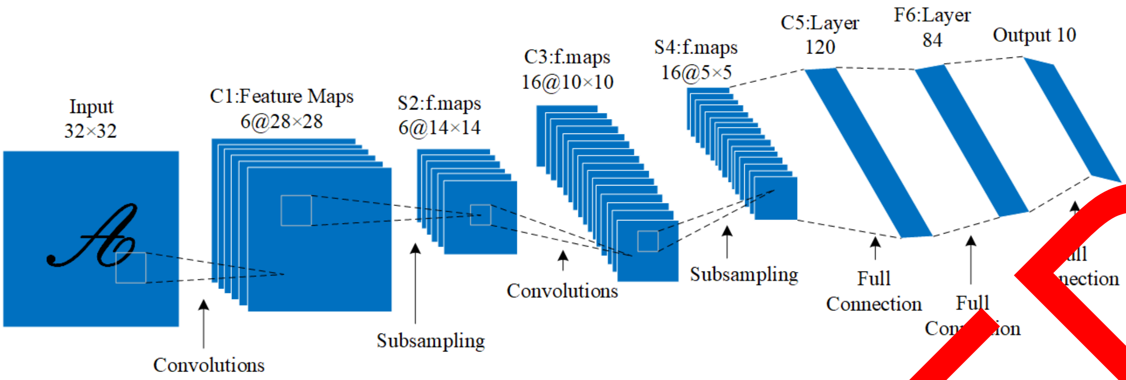

Fig. 2 Architecture of LeNet-5

(1) Convolutional Layer

Convolutional Layer is the basic unit of image feature extraction in image segmentation algorithm [28]. The traditional CNN network extracts image feature information by stacking convolutional layers and pooling layers, and shallow convolutional layers are used to extract simple but detailed features such as edges and textures [29]. Deep convolutions are used to extract more complex deep features [30]. The process can be seen as a matrix operation, and the image features are extracted by moving and calculating the convolution kernel on the image data. Fig. 3 shows the convolution operation process on the two-dimensional tensor in the CNN (ignoring the bias value), and it can be seen that the convolution operation is equivalent to the weighted sum of the input.

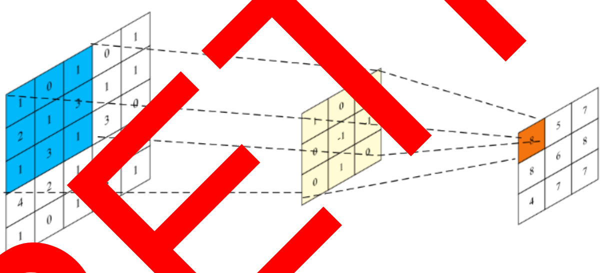

Fig. 3 Schematic diagram of the convolution operation process

(2) Pooling layer

In CNN, the pooling layer is generally after the convolutional layer. The pooling operation is also called down-sampling, which aims to reduce the resolution of the convolutional layer feature map, retain its useful information with less data, remove part of the redundant information, and prevent the network from overfitting [31]. In addition, the switching positions of pixels in the local area of the feature map will not cause the output of the pooling layer to change, so the feature map of the pooling layer has certain scale and spatial invariance, which can enhance the recognition ability of the model for the same object at different scales [27]. The commonly used pooling operations include Max pooling and average pooling. In max pooling, the maximum feature value in the sliding window is selected as the output value. Average pooling, on the other hand, needs to calculate the average of all feature values in the sliding window as the output value. As shown in Fig. 4, assuming that the size of the pooling kernel is 2×2, the size of the input feature map is 4×4, and the sliding step size is 2, the feature map of size 2×2 is obtained by calculating the maximum value or average value of the feature value within the sliding window [32].

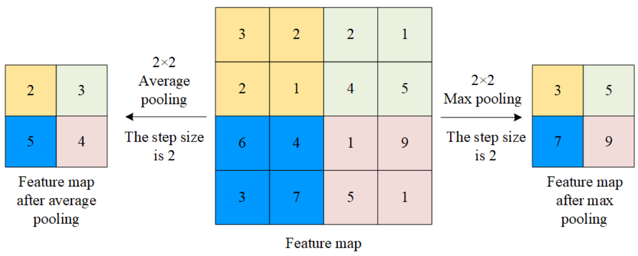

Fig. 4 Schematic of average pooling and max pooling

(3) Fully Connected Layer

The job of the Fully Connected Layer is to classify the input features to that layer. As the name suggests, this layer needs to be connected to all the neurons in the previous network layer to do its job. However, in image segmentation tasks, the fully connected layer is usually replaced by a convolutional layer because it is easy to lose the spatial information of features, which leads to poor segmentation results [33]. In addition, replacing the fully connected layer with the convolutional layer also removes the limitation of the previous neural network for image size, realizes pixel-level classification, simplifies the training process, and inputting image data of any size, the prediction results in the form of images can be obtained [34].

3.2. Fully convolutional Neural Networks (FCN)

When the CNN convolutional neural network model is used for image classification, the fully connected layer at the end will compress the two-dimensional matrix information in the original image, resulting in the loss of spatial information of the image [35], which will have a great impact on the convolutional neural network model for image segmentation [36-38]. The advent of the fully convolutional neural network (FCN) has pioneered the use of convolutional neural networks for image segmentation, and its network architecture is shown in Fig. 5.

In the network structure, the main role of the encoder part is to extract the high-dimensional features of the image, and the spatial dimension of the image is reduced after the convolution layer and the pooling layer [39]. The decoder part up-samples the output feature map, restores the feature map to the same size as the input image, and maps the extracted high-dimensional features to each pixel of the final feature map. The image segmentation at pixel level can be achieved [38-40].

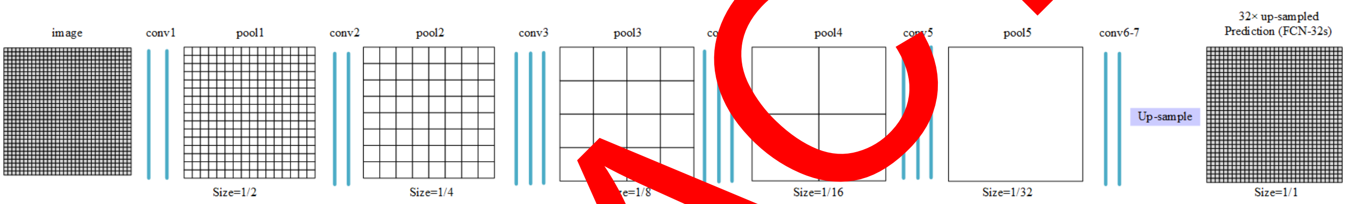

Fig. 5 Architecture of FCN-32s

Ben-Cohen et al [36]. first explored the use of FCN to complete the segmentation task of liver and tumor in CT images. Compared with the CNN convolutional neural network model based on fixed size input, FCN can accept any size input and produce corresponding size output through effective inference and learning. Therefore, the redundant calculation of the network can be eliminated and the results close to the manual cut can be obtained. Yuan et al [41]. trained an end-to-end skin melanoma segmentation method using a 19-layer depth FCN. In order to solve the class imbalance problem in dermoscopic images, the authors also designed a new loss function based on Jaccard distance, which achieved the best segmentation results on the ISBI [illegible] dataset at that time. Dasgupta et al [42]. first introduced FCN to the blood vessel segmentation problem of retinal images, and combined with a structured prediction method, the experimental results on the DRIVE database show the excellent performance of the FCN network.

Compared with CNN network, the advantage of FCN is that it can input images of any size, avoiding the problem of repeated storage and calculation of convolution caused by the use of pixel blocks. The disadvantages are as follows: firstly, the training of the network is troublesome and the segmentation results obtained are not accurate enough, and they are not sensitive to the intrinsic details of the image. Secondly, it does not consider the global context information, ignores the relationship between each pixel, and lacks spatial consistency [43].

3.3. U-Net

In addition to FCN, another classical network in the field of medical image segmentation is the U-Net network proposed by Ronneberger et al [44]. in 2015, which is also the most widely used network in medical image segmentation tasks, aiming to solve the problem of cell segmentation in medical images. U-Net network is an improved version of FCN. Its network structure is similar to the structure of FCN, there is no fully connected layer, and it is composed of a convolutional layer and a pooling layer, which is also divided into an encoder stage and a decoder stage [17, 45]. The structure of U-Net is shown in Fig. 6. The encoder part uses the CNN architecture as a shrinkage path to extract image features and reduce resolution. There are four sub-blocks in the shrinkage path, and each sub-block consists of two consecutive 3×3 convolutions, ReLU activation function, and a max pooling layer for down-sampling [46-48]. Two 3×3 convolution operations can effectively reduce the complexity of the neural network and keep the original segmentation accuracy unchanged [49]. In each down-sampling step, the number of feature channels is doubled. The decoder part consists of convolutional blocks with up-sampling operation to form an extended path to repair the image detail information, locate the boundary of the segmentation object, and gradually restore the spatial resolution of the feature map. In the expansion path, the sub-block contains two consecutive 3×3 convolutions, ReLU activation function, and up-sampling deconvolution layer [50]. Up-sampling expands the feature map to twice its original size and fixes the missing details. Stitching is a unique feature of U-Net, which can clip the low-level detailed features captured by the down-sampling process in the same layer and stitch them into the high-level semantic features extracted by the up-sampling process [48]. Finally, the output segmentation result combines the object class recognition basis provided by low-resolution information and the accurate positioning and segmentation basis provided by high-resolution features, which improves the problem of insufficient up-sampling information and realizes accurate segmentation.

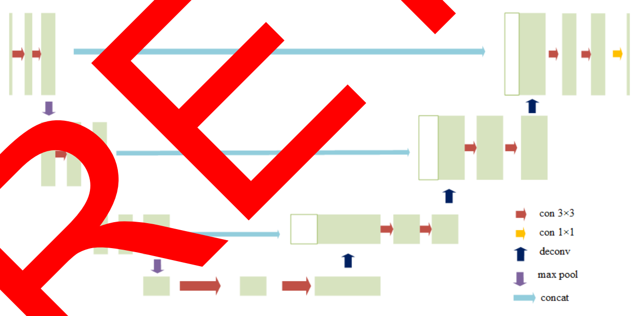

Fig. 6 Architecture of U-Net

Compared with other networks, the advantages of U-Net in medical image segmentation tasks mainly include:

(1) High sensitivity to small amounts of data. A small amount of data can be used to complete the training of the network and achieve good segmentation results.

(2) Design of U-shaped structure. The Encoder fully obtains the image features by stacking the convolutional layer and the down-sampling layer, and uses the skip connection to transfer and splice the shallow features, and then Decoder decodes the fused features.

However, although the multi-layer convolution can extract effective feature information, it also leads to the improvement of the number of parameters and computational cost. In addition, due to the problems such as resolution reduction and detail loss inevitably generated by down-sampling operation, U-Net still has a large room for improvement despite its obvious progress in segmentation results. In view of the excellent performance and existing problems of U-Net model in medical image segmentation tasks, many scholars have improved U-Net from different perspectives based on the U-shaped structure of U-Net to further improve the accuracy of medical image segmentation. At present, the typical U-Net variant models mainly include [illegible] [51], Non-Local U-Nets [45] and V-Net [52].

3.4. SegNet

SegNet is a deep network for image semantic segmentation proposed by Cambridge at *cvpr* in 2016 to solve the problem of autonomous driving or intelligent robots [53]. Its main role is to extract features. The network structure is shown in Fig. 7. The novelty of the overall network is that the decoder up-samples the lower resolution input feature maps. Specifically, the decoder uses the max-pooling index received from the corresponding encoder to perform a nonlinear up-sampling of the input feature map[54]. This approach reduces learning for up-sampling, improves boundary partitioning, and reduces the number of parameters for end-to-end training. The feature maps that become sparse due to up-sampling are subsequently subjected to trainable convolution operations to generate dense feature maps[47]. Finally, the last softmax layer of the network calculates the maximum probability of each pixel of the image in all categories, so as to complete the pixel-level classification of the image. SegNet only stores the max-pooling index and applies it to the decoding network to get better performance.

Therefore, the prominent advantage of SegNet is that it is more efficient than other segmentation networks[55, 56].

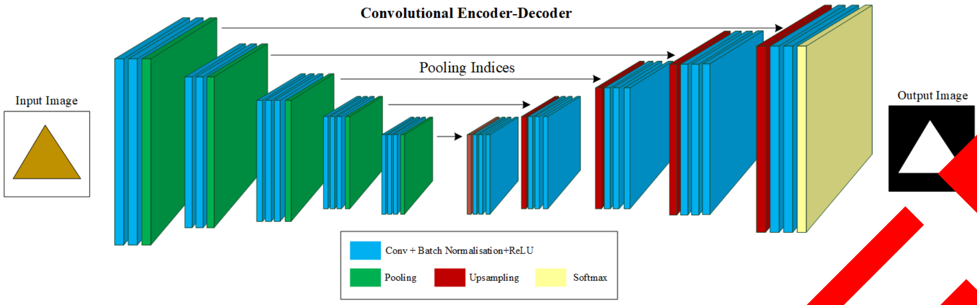

Fig. 7 Architecture of SegNet

3.5. PSPNet

To achieve high-precision segmentation in complex scenes, it is often necessary to use spatial pyramid pooling to obtain global image-level features. In order to incorporate appropriate global features, Zhao et al [57]. proposed Pyramid Scene Parsing Network (PSPNet), which aims to solve the problem of pixel-level classification in image semantic segmentation. The network structure is shown in Fig. 8.

The core idea of PSPNet is to use the pyramid pooling module to capture the context information at different scales to improve the understanding of image semantics and segmentation accuracy [58]. The main features of the model include:

(1) Pyramid Pooling Module: This module captures the global and local context information of the image from different levels by performing pooling operations at different scales [59]. It can effectively expand the receptive field and enable the network to perform fine-grained segmentation of objects and scenes at different scales.

(2) ResNet as the backbone: PSPNet usually uses ResNet as the backbone network to extract image features. ResNet is a kind of deep residual network. It solves the problem of gradient disappearance in deep network training through residual connection, which helps to improve the convergence and performance of the network [60].

(3) Fusion and up-sampling: After the pooling module, PSPNet fuses feature maps from different scales and up-samples them to the dimension of the original image by cascading fusion and up-sampling operations. The fused feature map can represent different semantic regions in the image more accurately.

PSPNet has achieved good performance in image semantic segmentation tasks, which is verified and compared on several public datasets [21]. It is widely used in many computer vision tasks, such as scene understanding, remote sensing image analysis, and autonomous driving.

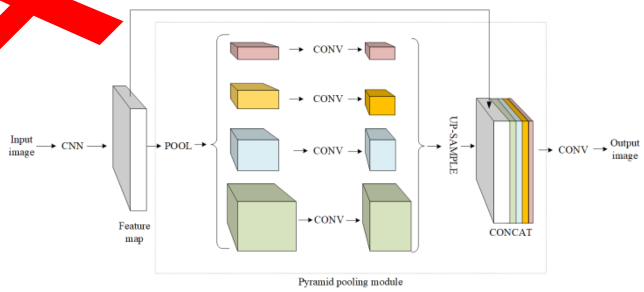

Fig. 8 Architecture of PSPNet

3.6. Mask R-CNN

Mask R-CNN [61] is an instance segmentation framework extended on the basis of Faster R-CNN ,as shown in Fig. 9. It has high object localization accuracy, high-quality pixel masks, and can be used for pixel-level segmentation tasks to obtain geometric properties such as shape, area, and contour. The Mask R-CNN model mainly consists of four parts: backbone network, RPN (region recommendation network), ROI Align (region of interest correction), and output branch.

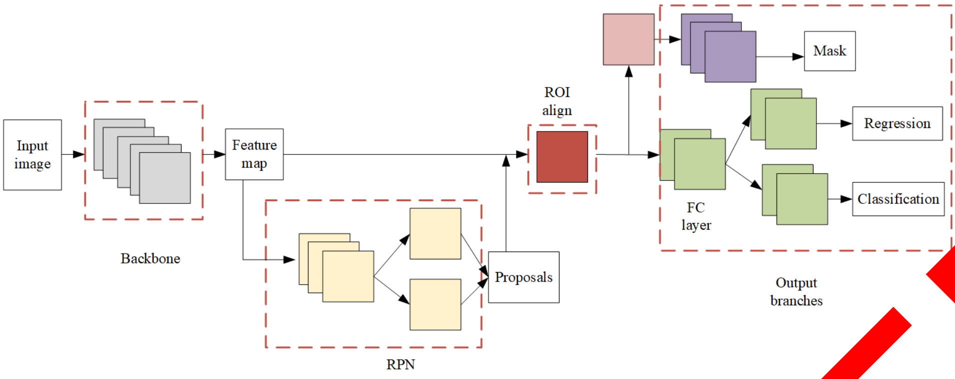

Fig. 9 Architecture of Mask R-CNN

The backbone network is mainly composed of ResNet [62] and FPN (Feature Pyramid network). The residual network ResNet34/50/101 composed of several residual blocks is used for feature extraction to realize cross-layer connection, and the identity mapping is constructed to solve the network degradation problem. FPN performs feature fusion on the feature maps extracted from each stage, and the feature maps of each layer fuse the features of different resolutions and semantic strength, so as to better represent the target at multiple scales [63, 64]. RPN uses a sliding window to scan the feature map to find the area where the object exists, so as to recommend the region of interest. Each sliding window generates a set of three anchor boxes with different aspect ratios ($1:1, 1:2, 2:1$) and sizes ($128, 256, 512$) [illegible]. In one branch, softmax was used for binary classification to determine whether the category of the target in the anchor box was foreground or background and output in the form of score. If there was a target object in the anchor box, it was judged as the foreground, otherwise it was the background [65]. The other branch adjusts the position and size of the anchor box to be close to the target object and outputs the coordinates. The Proposal layer uses non-maximum suppression (NMS) and threshold to filter the anchor box to obtain the ROI. ROI Align performs bilinear interpolation on the four boundary points of the feature map and fixes the pixel values of the coordinates of these four points [66]. Through the ROI Align operation, the output features with the same size can be obtained from the input regions with different feature sizes. Finally, the Mask R-CNN model outputs classification, localization regression, and mask segmentation.

3.7. Comparison of advantages and disadvantages of image segmentation methods

The technical characteristics, advantages and disadvantages of deep learning-based image segmentation methods are shown in Table I.

Table I Performance Comparison of methods

Methods	Characteristics	Advantages	Disadvantages
CNN	Shared convolution kernel; No pressure on high-dimensional data processing.	There is no need to manually select features; after training the weights, the features are obtained.	Parameters need to be tuned and large sample sizes are required.
FCN	Replace fully connected with convolutions; Introducing skip connections.	Images of any size can be input, and high-level features are fused with low-level features.	The global context information is not fully utilized and the segmentation accuracy is insufficient.
U-Net	A symmetric decoding network is added on the basis of the convolutional network.	It makes better use of the whole local context information, and effectively fuses the low-level and high-level information.	The segmentation of object boundaries is coarse.

SegNet	The decoder layer up-samples the feature maps using the max-pooling index stored in the corresponding encoder layer.	It avoids the learning of up-sampling and enhances the accuracy of image boundary location.	The segmentation accuracy is not high enough to meet the real-time needs.
PSPNet	The spatial pyramid module is introduced to aggregate the features of different scales.	The context information of the image is effectively obtained, and the utilization of the local and global features of the image is improved.	The boundary details of the target part will be lost.
Mask R-CNN	ROI Align is used instead of ROI Pooling.	The object is segmented at pixel level with high precision and accuracy.	It takes a long time to train and requires more data to achieve a high degree of generalization ability.

4. Common medical image segmentation tasks

Image segmentation in medical images is a key and challenging task, which plays a pivotal role in the diagnosis and treatment process. Medical image segmentation involves segmenting an image into meaningful regions that accurately depict structures and anomalies [67]. Image segmentation is very widely used in the medical field and spans various modalities, such as magnetic resonance imaging (MRI) [68], computed tomography (CT) [69], ultrasound, etc.

(1) Identification of anatomical structures

One of the main applications of image segmentation in medical images is the identification and delineation of anatomical structures. For example, in neuroimaging, segmentation is essential to separate and study different brain regions, tumors or lesions. In cardiac imaging, segmentation helps to accurately delineate cardiac compartments and blood vessels [70]. The ability to precisely identify anatomical structures is fundamental for quantitative analysis, surgical planning, and monitoring of disease progression [71].

(2). Tumor detection and Characterization

Image segmentation is particularly important for tumor detection and characterization in oncology. By segmenting medical images, clinicians can precisely locate and measure the size and shape of tumors [72]. This information is essential for treatment planning, including radiotherapy and surgery. In addition, segmentation helps to monitor changes in tumor size over time, enabling a more complete understanding of disease progression and treatment efficacy.

(3). Image-guided interventions

Image segmentation provides crucial guidance in medical procedures and interventions. For example, in image-guided surgery, segmentation helps surgeons visualize and navigate anatomical structures in real time. This ensures greater accuracy in targeting specific areas and minimizes the risk of damage to surrounding healthy tissues [73]. Image segmentation is also used in interventions such as radiotherapy, where it helps to pinpoint cancerous tissue while preserving healthy organs.

(4). Quantitative analysis and research

Image segmentation is helpful for quantitative analysis in medical research. Researchers have used segmentation to extract quantitative features and biomarkers from medical images to help study diseases and their progression [74]. Such quantitative data are valuable for understanding the effects of treatment, identifying patterns, and developing predictive models. In addition, segmentation is indispensable in the development of computer-aided diagnosis systems, and automatic algorithms help radiologists interpret medical images to improve diagnostic accuracy and efficiency.

5. Difficulties of medical image segmentation technology

The segmentation of medical images is more complex than that of natural images because of the unique characteristics of medical images. The specific performance is as follows:

(1) Small amount of data

The scale of finely annotated natural image data is very large. In contrast, medical image data is difficult to obtain due to the complex annotation and privacy issues [4]. Training a good model is relatively easy when the data is large and the model does not need to be very interpretable. When the data is small, we need to provide the model with enough prior knowledge to ensure that the model can learn the key features, and control the number of parameters to prevent overfitting.

(2) Aim small

Most of the objects in medical images are very small, with irregular shapes, fuzzy boundaries and complex gradients [75]. The segmentation of medical images requires high precision, so it is necessary to input more high-resolution information to the model to ensure accurate segmentation.

(3) Image semantics is simple

The context information of medical images is very important for the diagnosis of human diseases. However, due to the fixed structure of organs, the semantic information in images is not rich enough, so it is required to make full use of low-resolution information in the training process to ensure the accurate recognition of targets [76].

(4) Multidimensional images

Natural images are two-dimensional data, while medical images are mostly three-dimensional data. Three-dimensional convolution is needed to extract the three-dimensional information in the data, which increases the amount of parameters and is easy to overfitting.

(5) Multi-modal

Compared with natural images, medical images have multi-modal data, including MRI images and PET images [77]. The data of different modalities has its unique characteristics, and the model trained on a certain kind of data is not necessarily applicable to other data, which requires the model to be able to extract the characteristics of different modalities, so as to improve the generalization ability of the model.

These characteristics of medical images determine that medical image segmentation must use an encoder-decoder structure network model. The high difficulty and complexity of medical image segmentation technology is the main reason that medical image segmentation receives special attention in the field of image segmentation [78]. Addressing these challenges requires interdisciplinary collaboration between computer scientists, medical professionals, and imaging experts. Ongoing research focuses on developing more robust and adaptive segmentation algorithms that incorporate domain-specific knowledge and leverage advances in machine learning to improve segmentation performance in the face of these complex challenges.

6. Conclusion

This paper briefly introduces several classical image segmentation algorithms, and introduces the theoretical basis and related technologies of deep learning involved in this paper. Specifically, the typical architecture of CNN models and their components are first explained, and their related components are briefly explained. Then, the commonly used deep learning network models for medical image segmentation based on CNN architecture are introduced, including FCN,U-Net, SegNet, PSPNet, Mask R-CNN. Finally, the specific segmentation tasks in medical images are introduced, and the difficulties in medical image segmentation are discussed.

The development direction of medical image segmentation technology mainly focuses on the following points:

(1) expanding image data through data enhancement; The transfer learning method was used to combine the large data set pre-training and the target data set fine-tuning. Weakly supervised learning was used to effectively combine the advantages of unsupervised pre-training and supervised learning.

(2) batch normalization, regularization and Dropout can be used to solve the problem of intensity inhomogeneity.

(3) Multi-modal medical images are fused to improve the accuracy of analysis by utilizing the complementary information between different images. With the further improvement of computer technology and the continuous optimization and innovation of deep learning algorithms, medical image segmentation technology based on deep learning has great potential for development, and will be more widely used in various fields of medical research and have a more profound impact.

Acknowledgment

We thank all the anonymous reviewers for their hard reviewing work.

Funding

This research was funded by the open project of State Key Laboratory of Millimeter Waves (Grant No. K202218).

Conflict of Interest

The author declares there is no conflict of interest regarding this paper.

Data Availability Statement

There is no data associated with this paper.

References

[1] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET Image Processing,* vol. 16, no. 5, pp. 1243-1267, 2022.

[2] Y. Zhang, "Pathological brain detection in MRI scanning via Hu moment invariants and machine learning," *Journal of Experimental & Theoretical Artificial Intelligence,* vol. 29, no. 2, pp. 299-312, 2017.

[3] S. Wang, "Pathological Brain Detection by a Novel Image Feature—Fractional Fourier Entropy," *Entropy,* vol. 17, no. 12, pp. 8278-8296, 2015.

[4] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle, "Deep Neural Networks for Medical Image Segmentation," *Journal of Healthcare Engineering,* vol. 2022, p. 9580991, 2022.

[5] A. Qi *et al.*, "Directional mutation and crossover boosted ant colony optimization with application to COVID-19 X-ray image segmentation," *Computers in Biology and Medicine,* vol. 148, p. 105810, 2022.

[6] L. Abualigah, A. Diabat, P. Sumari, and A. H. Gandomi, *A Novel Evolutionary Arithmetic Optimization Algorithm for Multilevel Thresholding Segmentation of COVID-19 CT Images* (no. 7). 2021, p. 1155.

[7] L. He and S. Huang, "An efficient krill herd algorithm for color image multilevel thresholding segmentation problem," *Applied Soft Computing,* vol. 89, p. 106063, 2020.

[8] E. H. Houssein, M. M. Emam, and A. A. Ali, "An efficient multilevel thresholding segmentation method for thermography breast cancer imaging based on improved chimp optimization algorithm," *Expert Systems with Applications,* vol. 185, p. 115651, 2021.

[9] S. Zhao, D. M. Zhang, and H. W. Huang, "Deep learning-based image instance segmentation for moisture marks of shield tunnel lining," *Tunnelling and Underground Space Technology,* vol. 95, p. 103156, 2020.

[10] S. L. Bangare, "Classification of optimal brain tissue using dynamic region growing and fuzzy min-max neural network in brain magnetic resonance images," *Neuroscience Informatics,* vol. 2, no. 3, p. 100019, 2022.

[11] Q. Li, J. [illegible]eng, D. [illegible]ong, W. Nie, [illegible] Saffari, and J. Liu, "Automatic recognition of erosion area on the slope of tailings dam using region growing segmentation algorithm," *Arabian Journal of Geosciences,* vol. 15, no. 5, p. 438, 2022.

[12] X. Liu, H. Tian, [illegible] Wang, F. Jiang, and C. Zhang, "Research on Image Segmentation Algorithm and Performance of Power Insulator Based on Adaptive Region Growing," *Journal of Electrical Engineering & Technology,* vol. 17, no. 6, pp. 3601-[illegible], 2022.

[13] F. Poux, C. Mattes, Z. Selman, and L. Kobbelt, "Automatic region-growing system for the segmentation of large point clouds," *Automation in Construction,* vol. 138, p. 104250, 2022.

[14] M. Mushtaq, M. U. Akram, N. S. Alghamdi, J. Fatima, and R. F. Masood, *Localization and Edge-Based Segmentation of Lumbar Spine Vertebrae to Identify the Deformities Using Deep Learning Models* (no. 4). 2022, p. 1547.

[15] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, "ET-Net: A Generic Edge-aTtention Guidance Network for Medical Image Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Cham, 2019, pp. 442-450: Springer International Publishing.

[16] Y. Zhang, "Feature Extraction of Brain MRI by Stationary Wavelet Transform and its Applications," *Journal of Biological Systems,* vol. 18, no. S, pp. 115-132, 2010.

[17] Z. Liu *et al.*, "Liver CT sequence segmentation based with improved U-Net and graph cut," *Expert Systems with Applications,* vol. 126, pp. 54-63, 2019.

[18] D. Lu *et al.*, "Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography

images using a fully convolutional neural network," *Medical Image Analysis,* vol. 54, pp. 100-110, 2019.

[19] M. N. Reza, I. S. Na, S. W. Baek, and K.-H. Lee, "Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images," *Biosystems Engineering,* vol. 177, pp. 109-121, 2019.

[20] Y. Su, M. Jiang, L. Gao, X. You, X. Sun, and P. Li, "Graph-Cut-Based Node Embedding for Dimensionality Reduction and Classification of Hyperspectral Remote Sensing Images," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 1720-1723.

[21] X. Zhu, Z. Cheng, S. Wang, X. Chen, and G. Lu, "Coronary angiography image segmentation based on PSPNet," *Computer Methods and Programs in Biomedicine,* vol. 200, p. 105897, 2021.

[22] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 44, no. 7, pp. 3523-3542, 2022.

[23] G. Wei, G. Li, J. Zhao, and A. He, *Development of a LeNet-5 Gas Identification CNN Structure for Electronic Noses* ([illegible] 1). 2019, p. 217.

[24] T. Kattenborn, J. Leitloff, F. Schiefer, and S. H[illegible], "Review [illegible] Convolutional Neural Networks (CNN) [illegible] vegeta[illegible]n remote sensing," *ISPRS Journal of Phot[illegible]ammetry [illegible] Remote Sensing,* vol. 173, pp. 24-49, 2021.

[25] X. Lei, H. Pan, and X. H[illegible]g, "A Dila[illegible] CNN Model for Image Classification," [illegible] *Access,* vol. [illegible] pp. 124087-124095, 2019.

[26] Y. Li, J. Nie, [illegible] X. Chao, "Do w[illegible]lly need deep CNN for plant dise[illegible]s identification?," *Compu[illegible]s and Electronics in Agric[illegible]re,* vol. 17[illegible] p. 105803, 2020/11/01/ 2020.

[27] J. Lu, L. [illegible] H. Jiang, *R[illegible]iew on Convolutional Neural [illegible]rk (CN[illegible] Applied [illegible] lant Leaf Disease Classification* (no. [illegible] 2021, p. [illegible]

[28] Z. [illegible] and X. Zhang, "MIDCAN: A multiple input deep [illegible]onvolutional attention network for Covid-19 diagnosis based [illegible] chest CT and chest X-ray," *Pattern Recognition Letters,* vol. 1[illegible] pp. 8-16, 2021.

[29] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "Completely Automated CNN Architecture Design Based on Blocks," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 31, no. 4, pp. 1242-1254, 2020.

[30] Y.-D. Zhang, "ANC: Attention Network for COVID-19 Explainable Diagnosis Based on Convolutional Block Attention Module," *Computer Modeling in Engineering & Sciences,* vol. 127, pp. 1037-1058, 2021.

[31] M. F. Aslan, M. F. Unlersen, K. Sabanci, and A. Durdu, "CNN-based transfer learning–BiLSTM network: A [illegible] ap[illegible]ch for COVID-19 infection detection," *Ap[illegible] Soft Computi[illegible]* vol. 98, p. 106912, 2021.

[32] A. Chavda, J. Dsouza, S. Badg[illegible], and A. [illegible]ani, "Multi-Stage CNN Architecture for [illegible]e Mask Detection," [illegible] *202[illegible]th International Conferen[illegible] for Conv[illegible]ence in Tec[illegible]ology (I2CT)*, 2021, pp. 1[illegible]

[33] Y. Ji, H. Zhan[illegible] Zhang, and [illegible]iu, "C[illegible]based encoder-decoder [illegible] for salient [illegible]ct detection: A comprehensive revie[illegible]nd recent advances," *Information S[illegible]* vol. 546, pp. 83[illegible]7, 2021.

[34] [illegible]. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided CNN fo[illegible]age denoising," *Neural Networks,* vol. 124, [illegible]p. 117-129, 202[illegible]

[35] [illegible]an, "A [illegible]vey of Computer-Aided Tumor Diagnosis Based [illegible] Convolutional Neural Network," *Biology*, vol. 10, no. [illegible] doi: 10.3390/biology10111084

[36] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, and H. Greenspan, "Fully Convolutional Network for Liver Segmentation and Lesions Detection," in *Deep Learning and Data Labeling for Medical Applications*, 2016, pp. 77-85: Springer International Publishing.

[37] Y. Chen *et al.*, "Automatic Segmentation of Individual Tooth in Dental CBCT Images From Tooth Surface Map by a Multi-Task FCN," *IEEE Access,* vol. 8, pp. 97296-97309, 2020.

[38] J. Wu, B. Liu, H. Zhang, S. He, and Q. Yang, *Fault Detection Based on Fully Convolutional Networks (FCN)* (no. 3). 2021, p. 259.

[39] Y.-D. Zhang, "Improving ductal carcinoma in situ classification by convolutional neural network with exponential linear unit and rank-based weighted pooling," *Complex & Intelligent Systems,* vol. 7, pp. 1295-1310, 2021.

[40] J. Zhang, C. Lu, J. Wang, L. Wang, and X.-G. Yue, *Concrete Cracks Detection Based on FCN with Dilated Convolution* (no. 13). 2019, p. 2686.

[41] Y. Yuan, M. Chao, and Y. C. Lo, "Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With

Jaccard Distance," *IEEE Transactions on Medical Imaging,* vol. 36, no. 9, pp. 1876-1886, 2017.

[42] A. Dasgupta and S. Singh, "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 248-251.

[43] A. A. Ijjeh, S. Ullah, and P. Kudela, "Full wavefield processing by using FCN for delamination detection," *Mechanical Systems and Signal Processing,* vol. 153, p. 107537, 2021.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234-241: Springer International Publishing.

[45] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," *IEEE Access,* vol. 9, pp. 82031-82057, 2021.

[46] H. Wang *et al.*, "Mixed Transformer U-Net for Medical Image Segmentation," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 2390-2394.

[47] Y.-T. Weng, H.-W. Chan, and T.-Y. Huang, "Automatic Segmentation of Brain Tumor from 3D MR Images Using SegNet, U-Net, and PSP-Net," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Cham, 2020, pp. 226-233: Springer International Publishing.

[48] X. Wu, D. Hong, and J. Chanussot, "UIU-Net: U-Net in U-Net for Infrared Small Object Detection," *IEEE Transactions on Image Processing,* vol. 32, pp. 364-376, 2023.

[49] J. Zhang *et al.*, "LCU-Net: A novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition,* vol. 115, p. 107885, 2021.

[50] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, "DENSE-INception U-net for medical image segmentation," *Computer Methods and Programs in Biomedicine,* vol. 192, p. 105395, 2020.

[51] P. Kaur Buttar and M. K. Sachan, "Semantic segmentation of clouds in satellite images based on U-Net++ architecture and attention mechanism," *Expert Systems with Applications,* vol. 209, p. 118380, 2022.

[52] C. Zhao *et al.*, "Context-aware network fusing transformer and V-Net for semi-supervised segmentation of 3D left atrium," *Expert Systems with Applications,* vol. 214, p. 119105, 2023.

[53] S. Almotairi, G. Kareem, M. Aouf, B. Almutairi, and M. A.-M. Salem, *Liver Tumor Segmentation in CT Scans Using Modified SegNet* (no. 5). 2020, p. 1516.

[54] T. Chen *et al.*, "Pavement crack detection and recognition using the architecture of segNet," *Journal of Industrial Information Integration,* vol. 18, p. 100144, 2020/06/01/ 2020.

[55] N. Yamanakkanavar, J. Y. Choi, and B. Lee, *SM-SegNet: A Lightweight Squeeze M-SegNet for Tissue Segmentation in Brain MRI Scans* (no. 14). 2022, p. 5148.

[56] Z. Yan *et al.*, "PSPNet-based left ventricular MRI segmentation for the diagnosis of cardiac hypertrophy and myocardial infarction," *Computer Methods and Programs in Biomedicine,* vol. 221, p. 107197, 2022.

[57] Y. Qiao *et al.*, *A Diameter Measurement Method of Red Jujubes Trunk Based on Improved PSPNet* (no. 8). 2022, p. 1540.

[58] X. [illegible], Y. Guo, S. Wang, G. Cheng, X. Wang, and L. He, "Rapid detection of incomplete coal and gangue based on improved PSPNet," *Measurement,* vol. 201, p. 111646, 2022.

[59] C. Yang, H. Guo, Y. Xie, J. Tian, and X. X. Zhu, "A Method of Image Semantic Segmentation Based on PSPNet Linking Points With Labels in 3D: A Review of Point Cloud Semantic Segmentation," *Mathematical Problems in Engineering,* vol. 2022, no. 4, pp. 8958154-59, 2022.

[60] L. Yan *et al.*, "PSP net-based automatic segmentation network model for prostate magnetic resonance imaging," *Computer Methods and Programs in Biomedicine,* vol. 207, p. 106211, 2021.

[61] P. Bharati and A. Pramanik, "Deep Learning Techniques—R-CNN to Mask R-CNN: A Survey," in *Computational Intelligence in Pattern Recognition*, Singapore, 2020, pp. 657-668: Springer Singapore.

[62] H. Yu *et al.*, "A multi-stage data augmentation and AD-ResNet-based method for EPB utilization factor prediction," *Automation in Construction,* vol. 147, p. 104734, 2023.

[63] P. Chu, Z. Li, K. Lammers, R. Lu, and X. Liu, "Deep learning-based apple detection using a suppression mask R-CNN," *Pattern Recognition Letters,* vol. 147, pp. 206-211, 2021.

[64] M. Machefer, F. Lemarchand, V. Bonnefond, A. Hitchins, and P. Sidiropoulos, *Mask R-CNN Refitting Strategy for Plant Counting and Sizing in UAV Imagery* (no. 18). 2020, p. 3015.

[65] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Instance segmentation of apple flowers using the improved mask R–CNN model," *Biosystems Engineering,* vol. 193, pp. 264-278, 2020.

[66] X. Xu *et al.*, *Crack Detection and Comparison Study Based on Faster R-CNN and Mask R-CNN* (no. 3). 2022, p. 1215.

[67] J. M. J. Valanarasu and V. M. Patel, "UNeXt: MLP-Based Rapid Medical Image Segmentation Network," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, Cham, 2022, pp. 23-33: Springer Nature Switzerland.

[68] Y. Zhang, "Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine," *SpringerPlus,* vol. 4, no. 1, 2015, Art. no. 716.

[69] Y. D. Zhang and S. Satapathy, "A seven-layer convolutional neural network for chest CT-based COVID-19 diagnosis using stochastic pooling," *IEEE Sensors Journal,* vol. 22, no. 18, pp. 17573 - 17582, 2022.

[70] I. Qureshi *et al.*, "Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends," *Information Fusion,* vol. [illegible], pp. 316-352, 2023.

[71] S. Wang, "Grad-CAM: understanding [illegible] models," *Comp[illegible] Materials & Continua,* vol. [illegible], no. 2, pp. [illegible]-1324, 2023.

[72] M. Arabahmadi, R. F[illegible]khsh, and J. R[illegible]zadeh, *Deep Learning for Smart Healthca[illegible]—A Survey on Brain Tumor Detection from Medical Imaging* (no. 5). 2022, p. 1960.

[73] K. Treb, Y. [illegible]amisch, C. Ullberg, R. Zhang, and K. Li, "Photon counting-energy integrating hybrid flat panel detector systems for image-guided interventions: an experimental proof-of-concept," *Physics in Medicine & Biology,* vol. 68, no. 13, p. 1350[illegible], 2023.

[74] T. [illegible], S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array,* vol. 3-4, p. 100004, 2019.

[75] A. [illegible], B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation," *IEEE Transactions on Instrumentation and Measurement,* vol. 71, pp. 1-15, 2022.

[76] S. Mishra, Y. Zhang, D. Z. Chen, and X. S. Hu, "Data-Driven Deep Supervision for Medical Image Segmentation," *IEEE Transactions on Medical Imaging,* vol. 41, no. 6, pp. 1560-1574, 2022.

[77] D. Karimi, H. Dou, and A. Gholipour, "Medical Image Segmentation Using Transformer Networks," *IEEE Access,* vol. 10, pp. 29322-29332, 2022.

[78] S. Niyas, S. J. Pawan, M. Anand Kumar, and J. Rajan, "Medical image segmentation with 3D convolutional neural networks: A survey," *Neurocomputing,* vol. 493, pp. 397-[illegible], 2022.