J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 6859--6876

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on Medical Image Processing Combining Edge Detection and Morphological Features in Deep Learning Framework

Changhong Xie[1,✉]

[1] College of Information, Mechanical and Electrical Engineering, Shanghai Normal University (SHNU), Shanghai, 201418, China

ABSTRACT

In order to improve the accuracy and efficiency of medical image segmentation, this paper designs and proposes a medical image visualization method containing Sobel edge detection operator and 3D U-Net network based on deep learning and edge detection. The 3D U-Net network is used to capture the morphological and edge features of medical images on the public dataset, and the image binarization is performed on the result of its operation. The binarized image processed by corrosion and expansion algorithms is multiplied by the corresponding elements of the matrix with the medical image to obtain the visualization of the medical image. Different comparison algorithms and data sets are selected to verify the effectiveness of the optimized 3D U-Net network module and feature fusion module. Parameter settings are carried out, and the LIDC-IDRI dataset is used as the algorithm training base data to analyze the segmentation accuracy of the image processing method that fuses the edge detection operator with the 3D U-Net network. The algorithm ablation experiments are carried out according to different pruning degrees and training methods. The algorithm in this paper can achieve more than 80% segmentation accuracy on LIDC-IDRI dataset, in which the segmentation accuracy of liver reaches 97.1%.

Keywords: sobel edge detection, 3D U-Net network, erosion and inflation algorithm, feature fusion

1. Introduction

With the arrival of the digital era, the amount of data of all kinds of images grows rapidly, how to quickly and accurately image processing has become an urgent need, as an important part of computer vision in image processing, intelligent image segmentation and image classification technology has

✉ Corresponding author.
E-mail address: 1000526994@smail.shnu.edu.cn (C. Xie).
Received 20 February 2024; Revised 25 April 2024; Accepted 20 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-381

been widely used in various fields, and to a certain extent, it replaces the complex technology and instruments, providing people's lives with many convenience [1-4].

Image segmentation is one of the basic steps for image analysis and pattern recognition, and it plays a crucial role in computer vision [5]. As an important part of image understanding and also one of the most challenging problems in image processing, image segmentation determines the quality of the final analysis results, and plays an important role in medical analysis and diagnosis, automated driving, industrial quality inspection, security systems, and other fields [6-8]. Based on features such as gray scale, geometric shape, color, spatial texture, etc., image segmentation divides an image into blocks with the aim of grouping pixels with similar or consistent features into the same region and making the pixel features between different regions significantly different, so as to extract the important information in the image and provide a basis for subsequent analysis and processing [9-11]. Due to the importance and difficulty of the image segmentation problem, many researchers have been working on the problem since the 1970s [12]. However, due to its complexity and diversity, although there is no universally applicable and perfect image segmentation method, researchers have gained a certain understanding of the general principles of image segmentation, on the basis of which many research results and methods have been proposed [13-14].

As an important problem in the field of computer vision, image classification aims to classify digital images into different categories, which are often used to recognize and classify different objects, scenes or situations [15]. Image classification techniques have been widely used in various fields, such as medical image diagnosis, security monitoring, automatic driving, intelligent transportation, aerial remote sensing, and face recognition. Early image recognition and classification techniques mainly relied on the a priori knowledge of experts and required human identification and manual coding to adapt to specific data types and domain characteristics [16-18]. However, there are some problems with this approach, such as inefficiency, subjective evaluation results, and poor reliability, especially when facing massive data. In addition the effectiveness of manual classification may be greatly limited when facing complex data, such as high-dimensional data and nonlinear data [19-20]. With the non-stop exploration of deep learning techniques by researchers, the research on image classification has made considerable progress, in which deep learning models such as convolutional neural networks have become the mainstream methods for image classification research, and more significant results have been achieved. In addition, there are many other methods, such as support vector machine, decision tree, etc., which have also been widely used in the research of image classification [21-22].

With the progress of medical technology, various medical imaging devices occupy an irreplaceable position in diagnosis, treatment, and rehabilitation, and medical image processing is one of the important research directions in image processing [23]. The types of medical imaging that are currently widely used include magnetic resonance imaging, computed tomography, positron emission tomography, X-ray and ultrasound imaging. In addition, some microscopic examinations and retinal fundus images are commonly included. In medical image processing, both medical image segmentation and medical image classification are key components [24-25].

In the process of clinical diagnosis and treatment, medical image is an important auxiliary diagnostic means and plays an important role in clinical diagnosis and treatment. The difference between subjective judgment and objective understanding makes this diagnostic method require high clinical level and experience of doctors, otherwise the correctness of the diagnostic results of direct imaging, and the diagnostic process of medical imaging usually consumes a large amount of physician's time, and the diagnostic efficiency is not high [26-27]. Subject to the limitations of medical imaging technology, in the process of data acquisition of medical imaging equipment, medical images will inevitably introduce noise signals compared to natural images, resulting in low image quality, low contrast, noise and other characteristics [28]. For example, there will be speckle noise in CT images, artifacts and blurring due to mechanical vibration and magnetic field effects brought about in MRI images, as well as complex textures and blurred boundaries in medical images. Therefore, by utilizing preprocessed medical images, clearer organ edge features can be obtained, which can more effectively help doctors diagnose and understand the relevant information of the disease. The quality of preprocessed images has a direct impact on the processing effect of the subsequent images, which is of great significance for the further processing of medical images in computer vision algorithms [29].

In this paper, the results of the first two layers of convolution operation of the 3D U-Net network segmentation medical image processing process are utilized to extract medical morphological and edge features. The binarized image based on the underlying features of 3D U-Net network is processed using erosion and expansion algorithms, and the matrix corresponding element multiplication operation is performed to obtain the visualization results of the medical image. Combining the advantages of Sobel's edge detection algorithm, a medical image visualization method fusing the edge detection operator and the underlying features of 3D U-Net network is designed and proposed. The effectiveness of the improved 3D U-Net network and feature fusion algorithm is tested respectively. Set up the experimental environment and experimental parameters, analyze the segmentation performance of the algorithm, and conduct three sets of ablation experiments on the algorithm.

2. Medical image visualization methods

2.1. Medical Image Visualization Methods Fusing Edge Detection and Underlying Features

In this paper, 3D U-Net network is utilized for medical image segmentation on LIDC-IDRI dataset, and the result of convolution operation of the first two layers of 3D U-Net is obtained [30]. The result of this convolution operation is successively binarized, eroded, and expanded, and then do matrix multiplication operation with the corresponding elements of the original image to get the image containing only the subject O. The main work in this chapter is to obtain the edge feature image E of medical and use the fusion algorithm to fuse the O and E to obtain the final medical visualization image. The experimental flow of this paper is shown in Fig. 1.

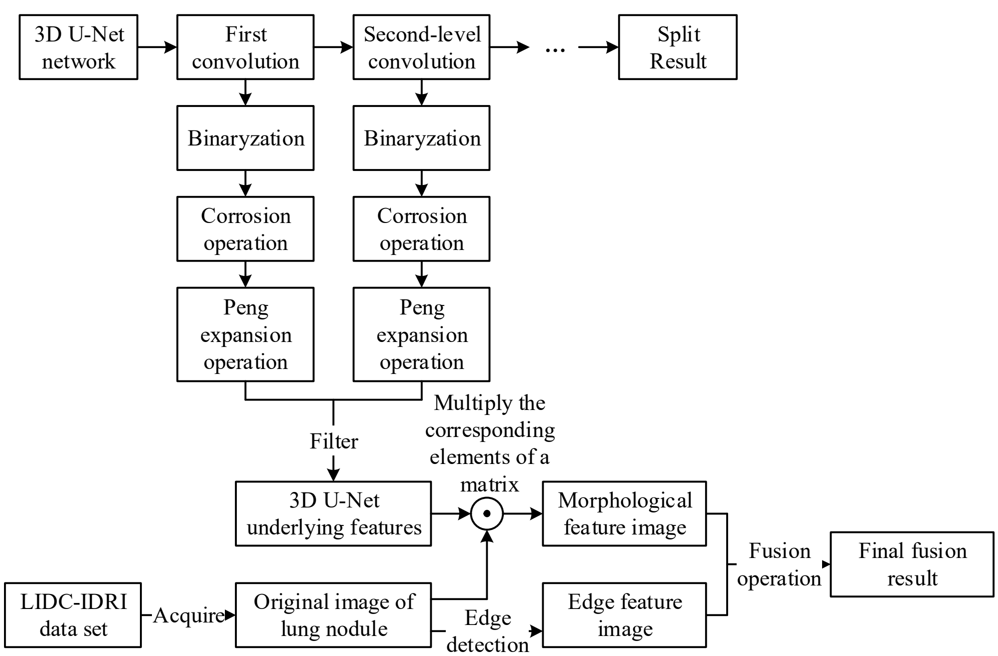


Fig. 1. The experimental process of this article

2.1.1. LIDC-IDRI dataset. The LIDC-IDRI dataset contains 1018 low-dose CT images of the lungs. Three types of regions are labeled in this dataset, including nodules larger than 3 mm in diameter, nodules smaller than 3 mm in diameter, and regions of non-nodular lung aberrations.

In order to study early cancer detection, the Cancer Research Institute has collected chest medical image files, such as CT radiographs, and labeled these images, i.e., formed the LIDC-IDRI dataset. A portion of the images from this dataset will be used for validation in the study of this paper. The dataset consists of 1018 low-dose CT images of the lungs, and the annotation of each instance was performed by four chest radiologists. The work was divided into two phases: in the first phase, each physician independently labeled the location of the patient and three types of regions, including nodules larger than 3 mm in diameter, nodules smaller than 3 mm in diameter, and regions larger than 3 mm in diameter that were non-nodular but had lung aberrations. In the second stage, each physician independently viewed the labeling results of the other physicians and presented his or her own judgment in order to arrive at the final results.

The composition of the LIDC-IDRI dataset is shown in Table 1. The format of the images was adopted as the standard format for medical images, i.e., dicom format, where an image has a total of 512 × 512 pixel points. Some other information such as serial number, image generation time, image type etc. are stored in this format. In this dataset, there is also xml format file which stores the results of the doctor's annotation.

Table 1. Composition of the LIDC-IDRI data set

Project	Quantity/size
Data size	140GB
Image type	CT:256354 DX:CR
Image number	245963
Series number	1018CT;300CR/DX
Research number	1500

2.1.2. Data set pre-processing. Given that there is some redundant information in the CT image processing process, such as images of the human torso, blood, and bones, this information is not needed in lung nodule segmentation. Therefore, if the information of these parts is removed before segmentation, it will greatly reduce the workload of the subsequent network computation.The lung images obtained from CT scans are calculated based on the X-ray attenuation values in Heinz units (HU), where the HU value of water is 0 and the HU value of air is -1000.The HU values of other objects can be calculated using this information with the formula:

$$HU = (\mu_x - \mu_{\text{water}}) / (\mu_x - \mu_{\text{air}}) * 1000 \tag{1}$$

where μ means the linear attenuation coefficient and the value is related to the intensity of the X rays.

The images of the lungs obtained from CT scans are pixel values based on a linear transformation of the Heinz unit (HU) values. Different CT devices have different transformation criteria, so they may produce different pixel values. However, under the same X-ray conditions, all devices produce the same HU values.

Since the HU value of the lung is about -500, only areas within the range [-1000,+400] (from air to bone) are retained when data preprocessing is performed. Regions outside this range were not considered relevant for lung disease surveillance and were therefore discarded. The generated lung parenchyma image was obtained by setting the portion of the image smaller than the -1000 HU value to -1000 and the portion larger than 400 to 400, then scaling to 0 to 255, and finally adding Mask.

2.2. Sobel edge detection operator

The Sobel extracted image edge operator is computed based on the first order derivative operator which contains two 3×3 detection templates. One template to extract the edges in the horizontal direction of the image and this template acts on the edges in the horizontal direction. The other one to extract the edges in the vertical direction of the image and it acts large for the edges in the vertical direction [31-32]. The detection template is as follows:

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} (x-1,y-1) & (x,y-1) & (x+1,y-1) \\ (x-1,y) & (x,y) & (x+1,y) \\ (x-1,y+1) & (x,y+1) & (x+1,y+1) \end{bmatrix} \tag{3}$$

In extracting the edges using the template in Eq. (2), the gradient values are computed within a 3 × 3 neighborhood by doing the operation with each pixel point in image $f(x,y)$ with the template first sub in each of these two directions. For the desired pixel point (x,y) at the center element of the template is placed at its position and a convolution operation is done within the image neighborhood to approximate the partial derivatives in these two directions. For the image neighborhood of Eq. (3), (x,y) is the center element, and the expressions of the Sobel detection operator for the sought edges in the horizontal and vertical directions are shown in Eqs. (4) and (5):

$$\begin{aligned} f_x(x,y) = {} & f(x+1,y-1) + 2f(x+1,y) + f(x+1,y+1) \\ & -f(x-1,y-1) - 2f(x-1,y) - f(x-1,y+1) \end{aligned} \tag{4}$$

$$\begin{aligned} f_y(x,y) = {} & f(x-1,y+1) + 2f(x,y+1) + f(x+1,y+1) \\ & -f(x-1,y-1) - 2f(x,y-1) - f(x+1,y-1) \end{aligned} \tag{5}$$

The result can be obtained as Eq. (6), which can be calculated by Eq. (7) for computational simplicity:

$$f(x,y) = \sqrt{f_x^{\,2}(x,y) + f_y^{\,2}(x,y)} \tag{6}$$

$$f(x,y) = |f_x(x,y)| + |f_y(x,y)| \tag{7}$$

After obtaining the result, a suitable threshold value T is selected and compared with the threshold value. If the result obtained is greater than the set threshold T, then this pixel point is recognized as an edge point. If it is less than T, then this pixel point is set to zero. As a result, the edges of the image can be obtained.Sobel operator extracts the edges, which is easy to compute and removes some of the noise, but there are pseudo edges in the detection results.

2.3. Medical image visualization based on 3D U-Net underlying features

With the penetration of deep learning in the field of medical image processing, many excellent neural network structures have been applied to the segmentation of medical images to obtain ideal segmentation results. In recent years, the study of medical images has also gradually expanded from 2D slice images to 3D holistic images. One of the main advantages of using 3D holistic images for medical image segmentation is that it enables the network to capture the spatial relationship between the slices and thus obtain more accurate segmentation results. In addition, 3D convolutional neural networks can better preserve the 3D structure of an image, which is especially important in medical imaging, where precise anatomical structures are critical for diagnosis and treatment planning.

3D CNNs have become essential tools for many image processing tasks, especially in the field of medical images. 3D CNNs have the additional advantage of being able to establish spatio-temporal correlations in medical images compared to traditional 2D CNNs. 3D CNNs allow the temporal dimension of data to be analyzed by sharing convolutional kernels and generating multiple feature mappings, thus improving the network's ability to capture both spatial and temporal information about the input image. This is particularly important in the field of medical image processing. In medical images by analyzing the correlation information between the overall structure of the organ and the slices is essential to make an accurate diagnosis. It has been shown that the use of 3D CNNs in medical image segmentation significantly improves feature extraction, demonstrating the importance of capturing temporal information in 3D medical images.

The development of 3D image processing techniques and algorithms is crucial for accurate and efficient analysis of biomedical images. In addition, the use of 3D images in medical applications such as diagnosis, treatment planning, and surgical navigation is becoming increasingly important, as it can help to enable more precise and personalized medical technology tools.

In 3D U-Net, each convolutional layer operates in 3D space and extracts spatial relationships between voxels within a 3D volume. Compared to 2D U-Net, 3D U-Net has more parameters and requires more computational resources and a larger dataset for training. And 3D U-Net can learn more comprehensive and accurate features from 3D input images, which leads to better segmentation performance. In addition, 3D convolution can capture the spatial relationships of 3D medical images more efficiently, resulting in more accurate segmentation results, and the ability to process and extract the 3D structure of an image is very important for practical needs in many medical fields.

2.3.1. 3D U-Net Underlying Feature Extraction. The application of deep learning in medical image analysis has received widespread attention, among which medical feature extraction is one of the important applications. In this paper, we use 3D U-Net network to segment the results of the first two layers of convolutional operations of the medical image processing process for the extraction of medical morphological and edge features.

This experiment uses the publicly available dataset LIDC-IDRI, and in the training process, a medical image dataset with categorized labels is used, which enables the 3D U-Net network to learn features such as morphology and edges of medical images. The results of the first two layers of convolution operation are output during the execution of medical image segmentation by the 3D U-Net network, and the results of this convolution operation are saved in a file in Python data format.

Based on the convolution operation results of the first and second layers of the 3D U-Net network, assuming that the feature map of the intermediate channel output from the convolution layer is $D \in R^{D \times H \times W}$ and the pixel points in the feature map are $a \in D$. Assuming that the visualization map of the result of the convolution operation is $A^{'} \in R^{H \times W}$ and the pixel points in the visualization map are $a^{'} \in A^{'}$, this section proposes a method for obtaining the visualization image of the result of the convolution operation, i.e., taking the minimum of the pixel positions corresponding to the feature maps of each channel, Eq:

$$a^{'} = \min(a_1, a_2, ..., a_D) \tag{8}$$

where $(a_1, a_2, ..., a_D)$ is in the same coordinate position as $a^{'}$ with respect to H and W.

2.3.2. Image binarization. The binarization of grayscale image is an important preprocessing step in image processing, and is also an indispensable step in this experiment, which converts the grayscale image into a binary image to pave the way for the subsequent corrosion and expansion operations. Considering the running complexity of the above algorithms and the image effect, this paper uses the adaptive thresholding method to process medical grayscale images.

Adaptive thresholding method is an improved algorithm based on local thresholding. Different from the local thresholding method, instead of dividing the image into multiple small blocks, the adaptive thresholding method determines the threshold value of each pixel point based on the gray value around the point.

Specifically, for each pixel point, a threshold is calculated based on the mean and variance of the pixels around it, and then it is determined whether the gray value of the point is greater than the threshold, and if it is, it is categorized as foreground, otherwise it is categorized as background. Adaptive thresholding method can effectively cope with uneven illumination, complex background, etc., and at the same time can retain the detailed information of the image, which is a more commonly used binarization method.

Assuming that m_{xy} and σ_{xy} are the mean and standard deviation of the set of pixels contained in the neighborhood S_{xy} of the image centered on (x, y), the adaptive threshold of the image is shown in Equation (9):

$$T_{xy} = a \times \sigma_{xy} + b \times m_{xy} \tag{9}$$

where a and b are non-negative constants, the computed binary image $g(x, y)$ is shown in Eq. (10):

$$g(x,y) = \begin{cases} 1, f(x,y) > T_{xy} \\ 0, f(x,y) \le T_{xy} \end{cases} \tag{10}$$

where $f(x, y)$ is the input image, the segmented binary image can be obtained by applying equation (10) to all pixels in the image.

2.3.3. Pre-processing of medical morphology information. A corrosion and expansion algorithm is used to process a binarized image based on the underlying features of a 3D U-Net network.

Assuming that A denotes a binarized image, B denotes a binarized template, C is the set of pixels with value 1 (target) in image A, and structuring element D is the set of pixels with value 1 in template B, the corrosion operation of B on A is defined as follows:

$$A\,!\ \ B = \{z\} \tag{11}$$

z denotes a pixel. D_i denotes the set of pixels that are A at B the position 1 (structural element D) when the center of the template B is translated to the position z. At pixel z, the corrupted image takes the value 1 at pixel z position if the condition $D_z \subseteq C$ is satisfied, i.e., A the pixel values at all $|D|$ positions corresponding to the structural element D are 1, otherwise it takes the value 0. $|D|$ denotes the total number of elements in the set D.

Similarly, the mathematical definition of the non-inverted structural element version of the expansion operation is as follows:

$$A \oplus B = \{z\} \tag{12}$$

When the center of the template B is translated to the z position, A the inflated image takes the value of 1 at the pixel z position as long as one of all $|D|$ the pixels at the corresponding position of the structuring element has the value of 1. Otherwise, it is 0.

The convolution of erosion and inflation is implemented as follows:

Let the size of the binary image A be $M \times N$ and the size of the template $\hat{B}$ be $m \times n$, $m = 2a+1, n = 2b+1$. The (linear) convolution between them is defined as follows:

$$(A * \hat{B})(x,y) = \sum_{s=-a t=-b}^{a} \hat{B}(s,t) A(x-s, y-t) \tag{13}$$

The above convolution operation can be summarized in eight words: invert, translate, phase (dot) multiply, and accumulate.

First, the template is inverted around the center origin of template $\hat{B}$. Then, the inverted template is translated to (x, y) and the pixel values at the corresponding $m \times n$ positions are multiplied together. Finally, the $m \times n$ product terms are added together to get the output value of the convolution at (x, y). In this paper, "linear convolution" is referred to as "convolution".

From the definitions of convolution, erosion and expansion, we can see that the common point is that they all need translation operations. Their differences are reflected in the following two aspects:

1) The convolution operation inverts and then translates template $\hat{B}$, while the erosion and expansion operations directly translate template $\hat{B}$ without inversion.

2) Erosion and expansion are defined using set relations, while convolution is defined using algebraic operations. Therefore, to address these two differences, the convolution operation and the output image of the operation are designed as follows: ① $\hat{B} = imrotate(B, 180°, center)$, i.e., templates $\hat{B}$ and B are inverse to each other, which ensures that the convolution operation uses the same templates as the corrosion and expansion operations in the translation operation. ② The result $A * \hat{B}$ of convolution is binarized to obtain the binary image E:

$$E(x,y) = \begin{cases} 1 & (A*\hat{B})(x,y) \geq T \\ 0 & (A*\hat{B})(x,y) < T \end{cases} \tag{14}$$

If the threshold T takes the value 1, the binary image E is equal to the inflated image, $E = A \oplus B$. If the threshold T takes the value $|D|$ (with respect to the image A, the region where there is no convolution out-of-bounds point). Or $|D \cap A|$ is taken near the four boundaries of image A (with respect to image A, the region where the convolution out-of-bounds point exists), then the binary image E is equal to the corrupted image, $E = A!\ B$. In this way, morphological corrosion and inflation are realized through the convolution operation and subsequent processing.

In summary, the main steps to realize corrosion and expansion by spatial domain convolution operation are as follows:

Step1: Input binary image A and template B.

Step2: Perform inversion on template B to get the inverted cross-panel $\hat{B}$.

Step3: Perform convolution operation on A and $\hat{B}$ to get convolution result $G_p = conv2(A, \hat{B})$.

Step4: Perform cropping on G_p, keeping the middle part of the same size as A, denoted as G.

Step5: Perform binarization on G to obtain the binary image E. When the thresholds are taken as 1 and $|D|$ (when the template is out of bounds, $|D \cap A|$), respectively, the corresponding E are the inflated image and the corrupted image respectively.

The cropping operation aims to obtain an image of the same size as the input binary image A. If this step is omitted, an output image of full size $(M+m-1) \times (N+n-1)$ is obtained. The same result is obtained by exchanging the order of steps 4 and 5.

2.3.4. Image visualization and analysis. The processed binarized image is multiplied by the corresponding elements of a matrix with the medical image, where the pixels in the medical part of the binarized image have a value of 1 and the pixels in the non-medical part of the image have a value of 0. Matrix multiplication by corresponding elements is a type of operation known as element-by-element multiplication, which is also known mathematically as the Hadamard product. The basic principle of this operation is that the elements in corresponding positions are multiplied together to obtain a new matrix. Let two images are A and B, both of size $m \times n$, and their corresponding points have pixel values $a_{i,j}$ and $b_{i,j}$, respectively, then their corresponding elements of the element-by-element multiplication (element-by-element product) C = AOB are shown in Equation (15):

$$c_{i,j} = a_{i,j} \times b_{i,j} \tag{15}$$

The result of this operation is a matrix of a new image with the same size as the input matrix. Since the images in this paper are all of the same size, the problem of resizing the input image is not required. The resultant image of the 3D U-Net network convolution operation is thus obtained.

In this section, we propose new method for visualization of medical images by using edge detection operator to obtain the edge features of the original image and fusing this result with the image obtained based on the underlying 3D U-Net features to obtain an image containing medical morphological features and edge features.

3. Deep learning-based medical image segmentation

When evaluating the segmentation effect of medical images, the visualization method is firstly used to compare the segmented pictures with the lesion areas labeled by professional physicians, and observe their number, edge information and segmentation area to obtain a more intuitive evaluation of the segmentation effect. In terms of objective evaluation, quantitative analysis is performed based on four specific assessment indexes, including precision (PRE), recall (SEN), Dice similarity coefficient/DI coefficient, and volume overlap error (VOE).

The Dice similarity coefficient, also known as the ensemble similarity measure function, is identical to the similarity index and is one of the most important evaluation metrics for medical image segmentation. It is used to measure the ratio of the size of the intersection of two sets to the sum of their total number of elements, and its value ranges between [0, 1], with higher scores indicating higher overlap. The calculation formula is shown in (16):

$$Dice = \frac{2\left|I_i \cap I_i\right|}{\left|I_i\right| + \left|I_i\right|}, i, j \in U_{image} \tag{16}$$

3.1. Effectiveness Analysis of 3D U-Net Network Improvement

In order to verify that the medical image segmentation algorithm based on 3D U-Net underlying feature extraction can be effectively used for the segmentation task of medical images, it is compared with CSFFNet, Attentiou Unet, Resunet, mU-Net and ARCNet.

Among them, CSFFNet is a 3D medical image segmentation algorithm based on cross-scale feature fusion, and CSFFNet includes three parts: shared coding module, segmentation decoding module, and reconstruction decoding module.

Attention Unet is the introduction of attention mechanism in U-net, by adding the attention module before feature fusion between the encoder's features and the decoder's corresponding features, so that the network can improve the performance of the network by suppressing the irrelevant regions of the input image during the training process, and strengthening the regions related to the segmentation task.

ARCNet is a 3D medical image segmentation algorithm based on multi-task learning, which uses the 3D U-Net backbone network as the basic network framework and combines the image reconstruction assistance task and the image segmentation task for multi-task learning.

The segmentation results of BraTS2018 dataset, LITS2017 dataset on different networks are shown below respectively.

1) BraTS 2018 dataset. In this paper, the BraTS 2018 dataset is used, which contains the brain image data of 285 glioma patients, and each patient's image has four MRI sequences of FLAIR, T1, T1GD, and T2, and contains the corresponding real segmentation labels. Among them, glioma, as one of the more common types of brain tumors, has an incidence rate of more than forty percent of brain tumor patients. Usually gliomas can be classified into four grades, according to which the severity of the patient's condition can be judged. In addition, the MRI images of each patient in the dataset are three-dimensional data with a size of 240 × 240 × 155.

2) LITS 2017 dataset. The liver medical imaging dataset used in this study was provided by the Liver Segmentation 2017 Challenge, which consists of 131 samples of abdominal enhanced CT images

containing authentic manually labeled labels. The data were obtained from patients with liver tumor diseases in hospitals and research institutions in five countries, including Germany. In this case, each CT sample image was labeled by three independent, specialized radiologists for specific parts of the liver and tumor prior to collection in order to eliminate inconsistencies in the results of labeling labels due to the subjective factors of the labelers. The size of each patient's liver 3D image in this dataset was 512 × 512 × 75 voxels.

The results of BraTS 2018 comparison experiments are shown in Table 2. From the segmentation results in the table, it is obvious that the algorithm proposed in this paper has better segmentation advantages. For example, taking the WT region segmentation results as an example, the performance of Improved 3D U-Net compared with CSFFNet, Resunet, mU-Net, Attention Unet, and ARCNet segmentation is improved by 0.0109, 0.0385, 0.0663, 0.0915, and 0.0243, respectively.

Table 2. The brats 2018 compare the results of the experiment

Method	Lessons		
	WT	TC	ET
CSFFNet	0.8827	0.7511	0.7382
Resunet	0.8551	0.6639	0.6639
mU-Net	0.8273	0.6758	0.6531
Attention Unet	0.8021	0.7236	0.7582
ARCNet	0.8693	0.7594	0.6325
Improved 3D U-Net	0.8936	0.8213	0.7906

The results of the LITS 2017 comparison experiments are shown in Table 3, demonstrating the results of several performance evaluation indexes of the medical image segmentation algorithms based on 3D U-Net underlying features and several existing segmentation networks on the LITS 2017 dataset.

Taking the DI results as an example, the segmentation result of the data on the optimized 3D U-Net network reaches 95.12%, which is an improvement of 0.0733 compared to Attention Unet.

From the table results, it can be concluded that the performance of the medical image segmentation algorithm with 3D U-Net underlying features proposed in this paper outperforms the existing image segmentation algorithms on both types of image datasets, MRI and CT.

Table 3. Lits 2017 compares experimental results

Method	Lessons		
	DI	JA	MSD
CSFFNet	0.9021	0.8364	46.82
Resunet	0.8364	0.8916	55.07
mU-Net	0.8695	0.8542	42.31
Attention Unet	0.8779	0.8669	38.62
ARCNet	0.8937	0.8745	37.45
Improved 3D U-Net	0.9512	0.9013	34.58

3.2. Feature Fusion Module Effectiveness Analysis

In order to verify the effectiveness of the feature fusion module, in this section, U-net-rec and U-net-rec-att networks are used as the base networks, and the feature fusion module is added to the two networks respectively to demonstrate the facilitating effect of feature fusion on the segmentation network. Where U-net-rec, U-net-rec-att, and U-net-rec-cff refer to the addition of reconstruction module to the improved U-Net (with the addition of the null-space convolutional pooling pyramid module), the addition of reconstruction module and attention module to the improved U-Net, and the addition of reconstruction module and feature fusion module to the improved U-Net, respectively. 3D U-net-sobel-feature-cff refers to adding Sobel edge detection module, underlying feature module, and feature fusion module on the improved U-Net, i.e., the medical image segmentation algorithm proposed in this paper that fuses edge detection operator and 3D U-Net underlying feature extraction.

The effect of feature fusion module on BraTS 2018 segmentation is shown in Table 4. By comparing U-net-rec and U-net-rec-cff, it can be found that after adding the feature fusion module, the DI value of the network is improved on all three segmentation regions of the data.

Taking the U-net-rec-att network as the base network for verifying the validity of the module as an example, the DI of the three segmentation regions of WT, TC, and ET reached 89.93%, 85.98%, and

80.24% on the 3D U-net-sobel-feature-cff network, which is an improvement of 1.39% over that on the U-net-rec-att network, 3.41%, and 1.69%.

Table 4. Feature fusion module affects brats 2018 segmentation

Method	Lessons		
	WT	TC	ET
U-net-rec	0.8576	0.8012	0.7747
U-net-rec-cff	0.8799	0.8364	0.7968
U-net-rec-att	0.8854	0.8257	0.7855
3D U-net-sobel-feature-cff	0.8993	0.8598	0.8024

The effectiveness of the feature fusion module on the LITS dataset is shown in Table 5. The table shows that the feature fusion module still has effectiveness on the CT dataset. In the U-net-rec-cff network experiment, for example, the segmentation performance of the result of the evaluation index DI reaches 90.76%, which is 1.45% better than the segmentation result of the U-net-rec network.3D In the U-net-sobel-feature-cff network experiment, the segmentation result of the evaluation index DI is better than the segmentation result of the U-net-rec network by 6.56%.

The feature fusion module proposed in this paper can effectively improve the segmentation results of the network and has a certain promotion effect on the segmentation task. To a certain extent, it solves the problems of existing networks such as information loss and insufficient information acquisition.

Table 5. The effectiveness of the feature fusion module in the LITS data set

Method	Lessons		
	DI	JA	MSD
U-net-rec	0.8931	0.8632	43.65
U-net-rec-cff	0.9076	0.8715	42.04
U-net-rec-att	0.9361	0.8936	50.72
3D U-net-sobel-feature-cff	0.9587	0.9122	36.51

4. Fusion of edge detection and underlying features for medical image processing

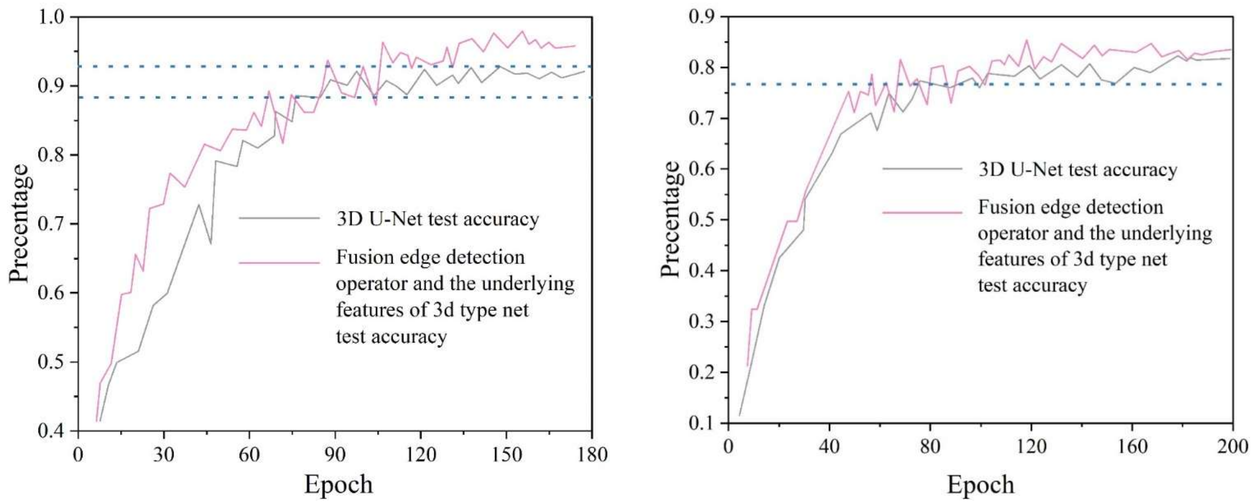
4.1. Optimization of experimental parameters

The hardware environment used in this experiment is: Intel(R) Xeon(R) Silver 4110 CPU, memory (RAM) 100G, and two graphics cards Geforce GTX 1080Ti. The operating system used: Ubuntu 16.04.6, the programming language used is python, the framework is tensorflow, and the cuda acceleration platform is installed.

4.1.1. Parameter setting. In this paper, the initial learning rate in the hyperparameters is set to 0.00006, the gradient computation is set to (0.9,0.999), the eps is set to 1e-8, and the weight decay is set to 0. In order to prevent network fitting, the dropout value is set to 0.6.

In segmenting the CT images of liver, through several trials, the training set was finally divided into 20 groups of CT images of liver and liver tumors of 10 people each, with epoch set to 180. Since the number of liver tumor samples is relatively small compared to the number of liver samples and the shape is irregular, which makes the segmentation difficult, the epoch for segmentation of liver tumor CT images is set to 200, and other parameters are kept unchanged.

4.1.2. Algorithm training effect. The liver and liver tumor segmentation training curves are shown in Fig. 2. Among them, Fig. (a) shows the liver segmentation training curve, and it can be seen that the network model reaches a stable state around 120 network iterations. The segmentation accuracy of the liver reaches 97.1% when the training network is stable, and the 3D U-Net network medical image segmentation method that fuses the edge detection operator and the underlying features improves the liver segmentation accuracy by about 3% over the 3D U-Net network. Figure (b) shows the training curve of liver tumor segmentation, from the training curve, it can be seen that the network iterates about 160 times to enter the stable state, and the value of the segmentation accuracy of the liver tumor reaches 82.9%, and the accuracy of the liver tumor segmentation is improved by about 1%.



(a) Hepatic segmentation (b) Liver tumor segmentation

Fig. 2. Liver and liver tumor segmentation training curve

4.2. Comparative experimental analysis of different models

In order to demonstrate the model performance of the improved 3D U-Net algorithm, the proposed method, i.e., the 3D U-Net network incorporating edge detection operators and underlying features, and the original U-Net were implemented, tested, and compared with the same set of CT image samples from the LIDC-IDRI dataset.

The variation of training loss of different models is shown in Fig. 3, which demonstrates the training process of different deep learning segmentation algorithms on the LIDC-IDRI dataset. Comparing the loss curves of all model training in the figure, it can be seen that the convergence speed of the improved U-Net network during the training process is very fast and the final training loss value is the lowest. In this paper, the proposed algorithm enters the state of minimum loss value when the number of iterations is 40, at this time, the loss value is less than 0.1.

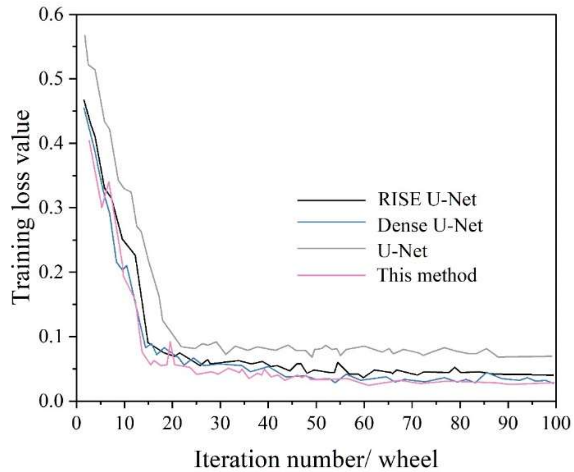


Fig. 3. Different model training loss changes

4.3. Analysis of ablation experiments

In order to verify the effectiveness of the method proposed in this paper, the controlled variable method is used to discuss and analyze the experiments on the LIDC-IDRI dataset, and the specific experiments will be divided into the following three groups.

4.3.1. Depth-supervised ablation-based experiments. In order to verify the effectiveness of the underlying feature extraction function of the network for lung parenchyma segmentation after the introduction of deep supervision, therefore, the performance of training time, IOU metrics and segmentation accuracy of the improved 3D U-Net on the lung dataset under different pruning levels were investigated, using D (D = 1, 2, 3, 4) to denote the model with different pruning levels.

The training metrics of the model under depth supervision are shown in Figure 4. As the depth of the model changes, the segmentation effect also fluctuates continuously, and it is found that the training time of the model with depth 3 is reduced by 0.215 on average while the IOU metrics are only reduced by 0.0079 when compared with the model with depth 4. It is also found that the training time can be further reduced by pruning the model, but the bad thing brought about by this is the significant decrease in the IOU metrics.

Segmentation experiments were conducted using the same dataset, and the final average accuracies of lung parenchyma segmentation for D_1, D_2, D_3, and D_4 were 0.750, 0.814, 0.903, and 0.895, respectively.

It can be seen that as the depth of the model increases, the segmentation effect is not better the deeper it is. When the depth is 3, the segmentation effect is optimal. When the depth is 4, the segmentation accuracy decreases instead, which is because the importance of features at different levels is different for different datasets. The original 4-layer U-Net is optimal for cellular dataset segmentation. However, when performing lung data segmentation, the shallower network can easily access some simple features of the image, such as texture, boundary and color. While the deeper network can acquire more abstract features of the image due to the expansion of the sensory field and the increase of the convolution operation, which leads to the feature redundancy problem and reduces the segmentation accuracy, so the lung parenchyma segmentation effect can be improved by trimming the model depth.

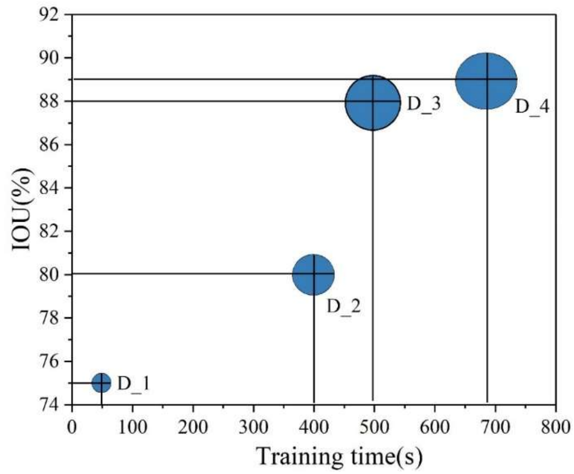


Fig. 4. The training indicators of the model under depth supervision

4.3.2. Ablation experiments based on dense jump connections. The improved 3D U-Net is better optimized than U-Net, which can be confirmed by the learning curve of the lung parenchyma segmentation task. A comparison of the improved 3D U-Net and U-Net validation loss values is shown in Fig. 5. Due to the newly designed dense jump connections in the middle layer, the improved 3D U-Net starts to converge at epoch=48, while the validation loss value of U-Net is still fluctuating at this time. Thus, the addition of dense jump connections helps to speed up the convergence and yields a lower validation loss.

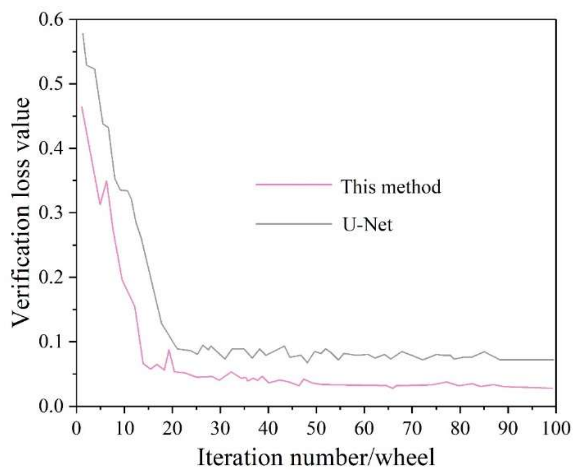


Fig. 5. Improved 3D U-Net and U-Net verified loss value comparison

4.3.3. Ablation experiments based on training modalities. With the same medical image dataset, the overall training scheme was compared with the isolated training scheme with different levels of pruning, and a comparison of the different training mode metrics is shown in Figure 6.

Since no model pruning was initially performed, the network with depth 4 naturally reached the same level of performance in both the isolated and embedded training modes. When the complete network is pruned to a depth of 1, it can be observed that the embedded training has about 2% more IOUs than the separate training scheme. This finding proves that the proposed architecture design improves the performance of each shallower network embedded in the network. The performance of the embedded shallower networks has a significant improvement when pruned compared to the same networks trained individually, meaning that supervised signals from the deeper layers are able to train higher performance shallow models. Simultaneous training of networks of different kinds of depths embedded in the improved 3D U-Net architecture can stimulate collaborative learning between them, improving segmentation performance more than training them individually.

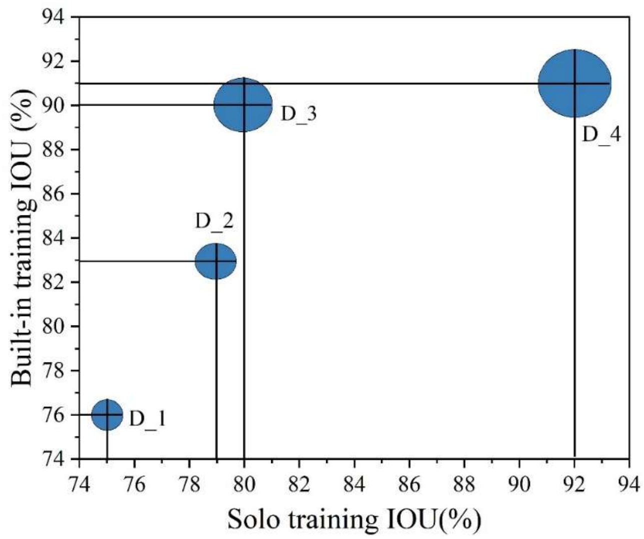


Fig. 6. Comparison of different training mode indicators

5. Conclusion

In this paper, Sobel edge detection algorithm and 3D U-Net network are fused to form a medical image visualization method. The 3D U-Net network is trained with medical image feature morphology, the results of the first two layers of convolution operation are output, the convolution results are processed, and the fusion algorithm is processed to obtain the visualization of medical images.

1) The segmentation performance of the improved 3D U-Net network module is judged by three evaluation indexes, DI, JA and MSD, and the medical image segmentation algorithm based on 3D U-Net underlying feature extraction has more segmentation performance than the comparison algorithms on BraTS 2018 dataset and LITS 2017 dataset.

The segmentation impact of the feature fusion algorithm on BraTS 2018 dataset and LITS2017 dataset are all improved, and taking the U-net-rec-att network as the base network for verifying the validity of the module as an example, the feature fusion algorithm proposed in this paper has improved the performance of the segmentation algorithm in the three segmentation regions of the BraTS 2018 dataset, namely, WT, TC, and ET, on the 3D U-net-sobel- feature-cff network achieves DIs of 89.93%, 85.98%, and 80.24%, which are 1.39%, 3.41%, and 1.69% higher than on the U-net-rec-att network.

2) Setting the initial learning rate to 0.00006, the fusion of Sobel edge detection algorithm and 3D U-Net network medical image processing method proposed in this paper achieves 97.1% segmentation accuracy of liver images and 82.9% segmentation accuracy of liver tumor images on LIDC-IDRI dataset. The improvement produces increased segmentation accuracy and the algorithm proposed has validity.

References

[1] Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps: Automation of decision making, 323-350.

[2] Zhou, S. K., Greenspan, H., & Shen, D. (Eds.). (2023). Deep learning for medical image analysis. Academic Press.

[3] Li, J., Jiang, P., An, Q., Wang, G. G., & Kong, H. F. (2024). Medical image identification methods: A review. Computers in Biology and Medicine, 169, 107777.

[4] Maier, A., Syben, C., Lasser, T., & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. Zeitschrift für Medizinische Physik, 29(2), 86-101.

[5] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[6] Birkfellner, W. (2024). Applied medical image processing: a basic course. CRC Press.

[7] Tian, D., Jiang, S., Zhang, L., Lu, X., & Xu, Y. (2023). The role of large language models in medical image processing: a narrative review. Quantitative Imaging in Medicine and Surgery, 14(1), 1108.

[8] Ramesh, K. K. D., Kumar, G. K., Swapna, K., Datta, D., & Rajest, S. S. (2021). A review of medical image segmentation algorithms. EAI Endorsed Transactions on Pervasive Health and Technology, 7(27), e6-e6.

[9] Carin, L., & Pencina, M. J. (2018). On deep learning for medical image analysis. Jama, 320(11), 1192-1193.

[10] Tournier, J. D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., ... & Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. Neuroimage, 202, 116137.

[11] Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. Sustainability, 13(3), 1224.

[12] Bhattacharya, S., Maddikunta, P. K. R., Pham, Q. V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., & Piran, M. J. (2021). Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable cities and society, 65, 102589.

[13] Chow, L. S., & Paramesran, R. (2016). Review of medical image quality assessment. Biomedical signal processing and control, 27, 145-154.

[14] Jasti, V. D. P., Zamani, A. S., Arumugam, K., Naved, M., Pallathadka, H., Sammy, F., ... & Kaliyaperumal, K. (2022). Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis. Security and communication networks, 2022(1), 1918379.

[15] Abdou, M. A. (2022). Literature review: Efficient deep neural networks techniques for medical image analysis. Neural Computing and Applications, 34(8), 5791-5812.

[16] Wang, J., Zhu, H., Wang, S. H., & Zhang, Y. D. (2021). A review of deep learning on medical image analysis. Mobile Networks and Applications, 26(1), 351-380.

[17] Du, J., Li, W., Lu, K., & Xiao, B. (2016). An overview of multi-modal medical image fusion. Neurocomputing, 215, 3-20.

[18] Li, Y., Zhao, J., Lv, Z., & Li, J. (2021). Medical image fusion method by deep learning. International Journal of Cognitive Computing in Engineering, 2, 21-29.

[19] Yan, F., Huang, H., Pedrycz, W., & Hirota, K. (2024). Review of medical image processing using quantum-enabled algorithms. Artificial Intelligence Review, 57(11), 300.

[20] Suganyadevi, S., Seethalakshmi, V., & Balasamy, K. (2022). A review on deep learning in medical image analysis. International Journal of Multimedia Information Retrieval, 11(1), 19-38.

[21] Lu, L., Zheng, Y., Carneiro, G., & Yang, L. (2017). Deep learning and convolutional neural networks for medical image computing. Advances in computer vision and pattern recognition, 10, 978-3.

[22] Hu, M., Zhong, Y., Xie, S., Lv, H., & Lv, Z. (2021). Fuzzy system based medical image processing for brain disease prediction. Frontiers in Neuroscience, 15, 714318.

[23] Alnaggar, O. A. M. F., Jagadale, B. N., Saif, M. A. N., Ghaleb, O. A., Ahmed, A. A., Aqlan, H. A. A., & Al-Ariki, H. D. E. (2024). Efficient artificial intelligence approaches for medical image processing in healthcare: comprehensive review, taxonomy, and analysis. Artificial Intelligence Review, 57(8), 221.

[24] Sagheer, S. V. M., & George, S. N. (2020). A review on medical image denoising algorithms. Biomedical signal processing and control, 61, 102036.

[25] Gong, M., Chen, S., Chen, Q., Zeng, Y., & Zhang, Y. (2021). Generative adversarial networks in medical image processing. Current pharmaceutical design, 27(15), 1856-1868.

[26] Roth, H. R., Shen, C., Oda, H., Oda, M., Hayashi, Y., Misawa, K., & Mori, K. (2018). Deep learning and its application to medical image segmentation. Medical Imaging Technology, 36(2), 63-71.

[27] Suzuki, K. (2017). Overview of deep learning in medical imaging. Radiological physics and technology, 10(3), 257-273.

[28] Obuchowicz, R., Strzelecki, M., & Piórkowski, A. (2024). Clinical applications of artificial intelligence in medical imaging and image processing—A review. Cancers, 16(10), 1870.

[29] Song, G., Han, J., Zhao, Y., Wang, Z., & Du, H. (2017). A review on medical image registration as an optimization problem. Current Medical Imaging, 13(3), 274-283.

[30] Mohammad Amin Izadi,Nafiseh Alemohammad,Parham Geramifar,Ali Salimi,Zeinab Paymani,Roya Eisazadeh... & Zaynab Sabouri. (2024). Automatic detection and segmentation of lesions in 18F-FDG PET/CT imaging of patients with Hodgkin lymphoma using 3D dense U-Net.. Nuclear medicine communications.

[31] Taoufik Saidani,Refka Ghodhbani,Mohamed Ben Ammar,Marouan Kouki,Mohammad H Algarni,Yahia Said... & Eman H. Abd Elkawy. (2024). Design and Implementation of a Real-Time Image Processing System Based on Sobel Edge Detection using Model-based Design Methods. International Journal of Advanced Computer Science and Applications (IJACSA)(3).

[32] K. N. Vijeykumar,Silambarasan Dharmalingam,Vineel Talluri Jessy & V. Aravind. (2024). ASIC Design of Low Power Sobel Edge Detection Filter: An Analog Approach. Journal of Circuits, Systems and Computers(11).