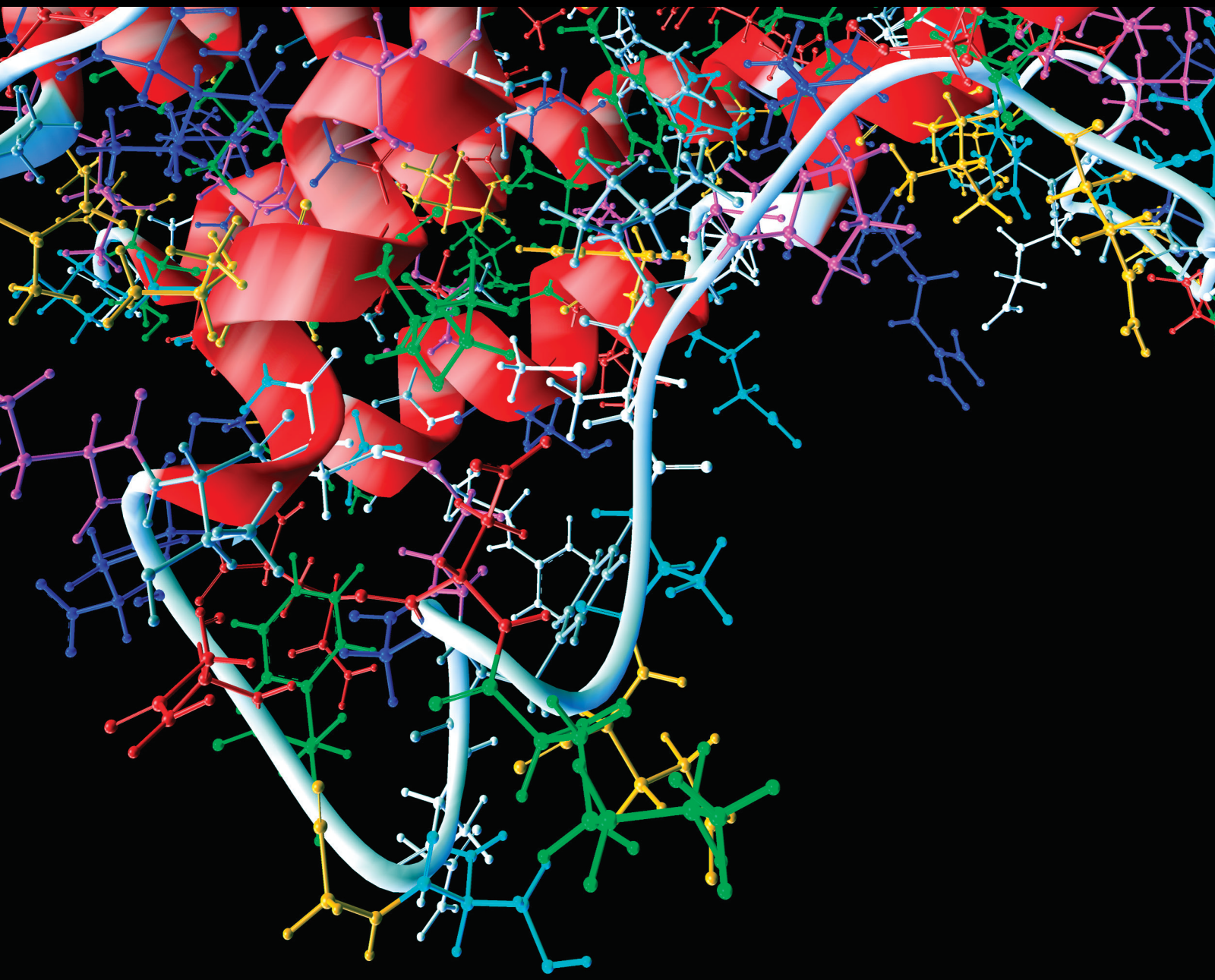
Computational and Mathematical Methods in Medicine

Emerging Challenges of AI for Biomedical Image Analysis

Lead Guest Editor: Shuihua Wang
Guest Editors: Yuankai Huo and Zheng Zhang



Emerging Challenges of AI for Biomedical Image Analysis

Computational and Mathematical Methods in Medicine

Emerging Challenges of AI for Biomedical Image Analysis

Lead Guest Editor: Shuihua Wang
Guest Editors: Yuankai Huo and Zheng Zhang

Contents

A Study on Weak Edge Detection of COVID-19's CT Images Based on Histogram Equalization and Improved Canny Algorithm
Shou-Ming Hou, Chao-Lan Jia, Ming-Jie Hou, Steven L. Fernandes, and Jin-Cheng Guo
Research Article (13 pages), Article ID 5208940, Volume 2021 (2021)

A Segmentation Method of Foramen Ovale Based on Multiatlas
Jiashi Zhao, Huatao Ge, Wei He, Yanfang Li, Weili Shi, Zhengang Jiang, Yonghui Li, and Xingzhi Li
Research Article (10 pages), Article ID 5221111, Volume 2021 (2021)

Robust Blood Cell Image Segmentation Method Based on Neural Ordinary Differential Equations
Dongming Li, Peng Tang, Run Zhang, Changming Sun, Yong Li, Jingning Qian, Yan Liang, Jinhua Yang, and Lijuan Zhang
Research Article (11 pages), Article ID 5590180, Volume 2021 (2021)

Multiclassification of Endoscopic Colonoscopy Images Based on Deep Transfer Learning
Yan Wang, Zixuan Feng, Liping Song, Xiangbin Liu, and Shuai Liu
Research Article (12 pages), Article ID 2485934, Volume 2021 (2021)

LCC-Net: A Lightweight Cross-Consistency Network for Semisupervised Cardiac MR Image Segmentation
Lai Song, Jiajin Yi, and Jialin Peng
Research Article (9 pages), Article ID 9960199, Volume 2021 (2021)

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 5208940, 13 pages
https://doi.org/10.1155/2021/5208940

Research Article

A Study on Weak Edge Detection of COVID-19's CT Images Based on Histogram Equalization and Improved Canny Algorithm

Shou-Ming Hou,[1] Chao-Lan Jia,[1] Ming-Jie Hou,[2] Steven L. Fernandes,[3] and Jin-Cheng Guo[4]

[1]*School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China*
[2]*CT Centre, Jiaozuo People's Hospital, Jiaozuo 454000, China*
[3]*Department of Computer Science, Design & Journalism, Creighton University, Omaha, Nebraska, USA*
[4]*Department of Thoracic Surgery, Jiaozuo Second People's Hospital, Jiaozuo 454000, China*

Correspondence should be addressed to Jin-Cheng Guo; acheng69102@hotmail.com

Received 14 May 2021; Revised 11 June 2021; Accepted 4 October 2021; Published 28 October 2021

Academic Editor: Shuihua Wang

The coronavirus disease 2019 (COVID-19) is a substantial threat to people's lives and health due to its high infectivity and rapid spread. Computed tomography (CT) scan is one of the important auxiliary methods for the clinical diagnosis of COVID-19. However, CT image lesion edge is normally affected by pixels with uneven grayscale and isolated noise, which makes weak edge detection of the COVID-19 lesion more complicated. In order to solve this problem, an edge detection method is proposed, which combines the histogram equalization and the improved Canny algorithm. Specifically, the histogram equalization is applied to enhance image contrast. In the improved Canny algorithm, the median filter, instead of the Gaussian filter, is used to remove the isolated noise points. The K-means algorithm is applied to separate the image background and edge. And the Canny algorithm is improved continuously by combining the mathematical morphology and the maximum between class variance method (OTSU). On selecting four types of lesion images from COVID-CT date set, MSE, MAE, SNR, and the running time are applied to evaluate the performance of the proposed method. The average values of these evaluation indicators are 1.7322, 7.9010, 57.1241, and 5.4887, respectively. Compared with other three methods, these values indicate that the proposed method achieves better result. The experimental results prove that the proposed algorithm can effectively detect the weak edge of the lesion, which is helpful for the diagnosis of COVID-19.

1. Introduction

COVID-19 has caused a health crisis worldwide, impacting all sectors of human life [1]. On 11 March 2020, COVID-19 was declared as an extremely high-risk disease by the World Health Organization (WHO) [2]. Up to April 16, 2021, more than 139.64 million people worldwide were infected, and more than 3 million people had died [3]. CT images contain a lot of important information, which can be used to evaluate disease quantitatively. In COVID-19 diagnosis, the reverse transcription-polymerase chain reaction (RT-PCR) detection is the most common way. However, when test kits are insufficient, especially during the outbreak phase, CT scan is a more effective diagnostic method. CT can show detailed symptoms of clinical diagnosis in COVID-19, especially for patients with moderate to severe [4]. Many hospitals use this technique to scan the lungs of patients and diagnose the illness. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is determined as the cause of COVID-19 [5]. There are several CT image features of COVID-19, such as single ground-glass shadow, diffuse ground-glass shadow of both lungs, large area consolidation of both lungs, multiple patchy consolidations, and paving stone-like[6]. Edge detection algorithms are used to detect

the lesion of CT images, which is helpful for doctors to diagnose the condition. However, the lesion will be missed or incorrectly detected due to poor contrast of the edge, thus affecting patient treatment. Therefore, weak edge detection is an urgent problem to be solved.

In recent years, weak edge detection has been researched by plenty of scholars. In 2012, Ji et al. [7] effectively extracted weak edge information in the skull using a combination of histogram equalization and mean-shift filter (called HMS from now on), but the edge continuity was flawed. In the same year, Xue et al. [8] developed the Laplace operator-based Hessian matrix to accurately detect blood vessels on uneven gray scale blood vessel images, which was only suitable for two-dimensional images. In 2015, Lin et al. [9] proposed a weak edge detection method based on the cumulative change rate, but this method required manual adjustment of the unified threshold. In 2016, Li et al. [10] segmented the prostate with the level set method from ultrasound images with weak edges. In 2017, Kathrin et al. [11] addressed the fuzzy edge detection of biomedical CT images by dynamic programming technology. In the same year, Khadidos et al. [12] proved that the weighted level set evolution method is an effective method in weak edge detection. In 2020, Wang and Xu [13] used fit methods to calculate new local terms of the variation segmentation model. The sensitivity of the method to the initial contour is reduced, and the accuracy of weak edge segmentation is improved. However, the segmentation process relies too much on manual experience. In the same year, Chetna et al. [14] designed an automatic segmentation technology based on energy curves, which used an automatic threshold to reduce step size, and then the breast cancer cells can be accurately identified.

From the experimental results, the above algorithms can extract weak edge, but there are still some defects, for example, the result is not so accurate of the COVID-19 CT image detection. In other words, the detection error needs to be further reduced. In this paper, based on histogram equalization and improved Canny algorithm, we propose an weak edge detection algorithm. Specifically, enhancing image edge is derived from capturing the details by using histogram equalization. Then, to obtain lesion edge, we embed median filter, K-means, mathematical morphology, OTSU algorithm into the Canny algorithm, and namely, the improved Canny algorithm. The proposed algorithm can efficiently eliminate noise of COVID-19 image and extract lesion's detail and edge.

The main contributions of this paper on weak edge detection of CT images are as follows:

(1) The histogram equalization is combined with the improved Canny algorithm to establish an edge detection model. The model benefits from balancing the noise and the edge. In other words, the model can not only reduce the noise influence in COVID-19 CT images but also improve the accuracy of weak edge detection

(2) The median filter is employed to optimize the Canny algorithm. Specifically, the median filter takes the place of the Gaussian filter so that the isolated noise point can be eliminated

(3) In order to automatically obtain threshold, the novelty of this paper is using the OTSU algorithm

This paper is organized as follows. Section 2 describes the data set applied by the proposed algorithm. Section 3 depicts the related theories of the proposed algorithm, including histogram equalization, Canny algorithm, and the proposed algorithm. Experimental evaluation and comparison with other methods are discussed in Section 4. Concluding remarks are given in Section 5.

2. Dataset

In this section, we introduce the dataset used in this paper. From the perspective of data authenticity, COVID-CT is used to verify the effectiveness of the algorithm [15], the dataset contains COVID and non-COVID, and Figure 1(a) and Figure 1(b) are representative image of COVID and non-COVID. From Table 1, COVID contains 349 CT images of clinical manifestation, which are from 216 patients, and the ratio of male to female is 86 : 51. And 463 CT images of healthy people are involved in non-COVID.

3. Methodology

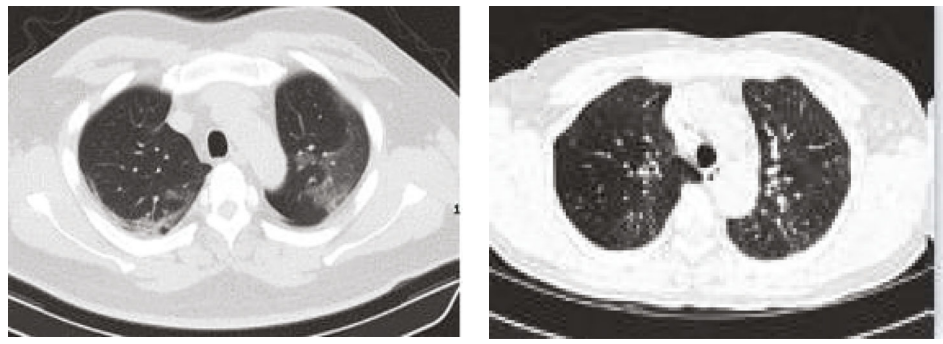
In this section, we introduce the related methods, including histogram equalization, traditional Canny algorithm, the improved Canny algorithm, and the proposed method. Histogram equalization is detailed description firstly. Then, we improve the traditional Canny algorithm, namely, the improved Canny algorithm. Finally, we introduce the proposed method.

3.1. Histogram Equalization. Histogram equalization has been extensively used in image enhancement algorithms. The basic idea is that the original nonuniform probability distribution gray map of the CT image is nonlinearly stretched by the histogram and transformed into a uniform probability distribution map [16]. In other words, the image clarity is enhanced by adjusting the size of the gray value [17]. The theoretical formula of histogram equalization is as follows.

Calculate the normalized gray value of the original image and the normalized gray value of the mapped image, which are r and s, respectively. The relationship between s and r is as follows.

$$s = T(r) = \int_0^r P_r(r)dr, \tag{1}$$

where $P_r(r)$ represents the probability density function of the image, and the range of r, s, and $T(r)$ values is all in $[0, 1]$; also, $T(r)$ increases monotonically in the interval.



(a) COVID image (b) Non-COVID image

Figure 1: Example of CT images for COVID and non-COVID.

Table 1: Comparison of COVID and non-COVID.

	Images	Patients	Male/female
COVID	349	216	86/51
Non-COVID	463	55	

Assume that $P_s(s)$ is the probability density function after gray scale transformation, an the relationship between $P_r(r)$ and $P_s(s)$ is as follows.

$$P_s(s) = P_r(r)\left.\frac{d_r}{d_s}\right|_{r=T^{-1}(r)}. \tag{2}$$

B is expressed as gray level, the image distribution probability is $P(r_B)$, the number of pixels is n_B, the pixel is N, the total gray level is L, and the relationship is as follows.

$$P_r(B) = \frac{n_B}{N}, \tag{3}$$

where $B \in \{0, L\text{-}1\}$. When $\sum_{B=0}^{L-1} P(r_B) = 1$, the image is a uniform probability density function.

Based on the histogram equalization algorithm, the single ground-glass shadow CT image features are enhanced, and the result is shown in Figure 2. Figure 2(a) is the original medical CT image of the single ground-glass shadow, and Figure 2(c) is the original image histogram. Figure 2(b) and Figure 2(d) are the enhanced original image and the enhanced histogram, respectively. We can observe that the weak pixels are enhanced, and the image details tend to be evenly distributed.

3.2. Traditional Canny Algorithm and Existing Problems. The edge detection is helpful towards analyzing CT images. At present, the mainstream edge detection algorithms include morphological processing [18], ant colony algorithm [19], watershed [20], Canny [21], and machine learning [22], in which Canny is one of the most widely used edge detection algorithms. This algorithm has the characteristics of high locating accuracy and effectively suppresses noise. The steps of the traditional Canny edge detection algorithm are as follows.

(1) Gaussian filter. It can remove nonhigh frequency noise and smooth graph

$$G = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{x^2}{2\sigma^2}\right), \tag{4}$$

where σ means the standard deviation of Gaussian filter, and G represents the intensity of the pixel after smoothing.

(2) The gradient magnitude and direction. Traditional edge difference operator, such as Sobel, is used to calculate the gradient magnitude and direction, so as to get thick and bright image edge

$$\begin{aligned} G_x &= [f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1)] \\ &\quad - [f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)], \\ G_y &= [f(x-1, y-1) + 2f(x, y-1) + f(x+1, y-1)] \\ &\quad - [f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1)], \end{aligned} \tag{5}$$

where G_x and G_y represent the gradient components in the horizontal and vertical direction, and $f(x, y)$ represents the pixel value of the image.

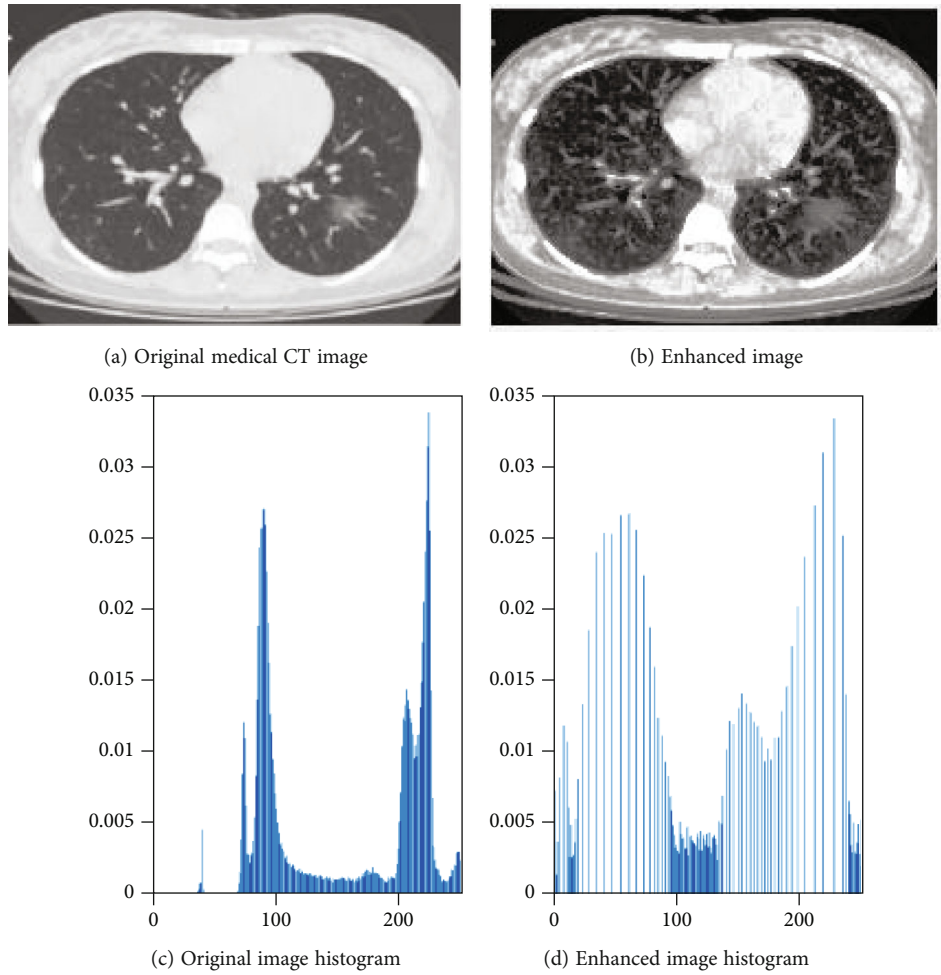
Let M is the gradient magnitude, and the calculation formula of G is as follows.

$$M = \sqrt{G_x^2 + G_y^2}. \tag{6}$$

Assume that J represents the gradient direction, the calculation formula of J is as follows.

$$J = \arctan\left(\frac{G_y}{G_x}\right). \tag{7}$$

(3) Nonmaximum suppression. After calculating the gradient magnitude and direction, the sharpest position is retained to ensure a clearer boundary in the



(a) Original medical CT image (b) Enhanced image

(c) Original image histogram (d) Enhanced image histogram

Figure 2: Comparison of the effect of histogram equalization algorithm.

Algorithm pseudo code
if $P \geq P_1$ and $P \leq P_2$
P is an edge
else
P should be suppressed

Pseudocode 1

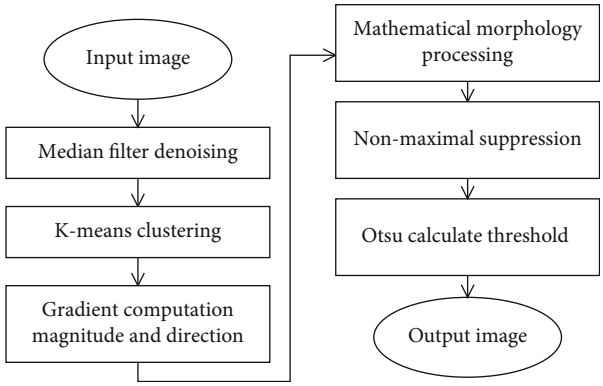
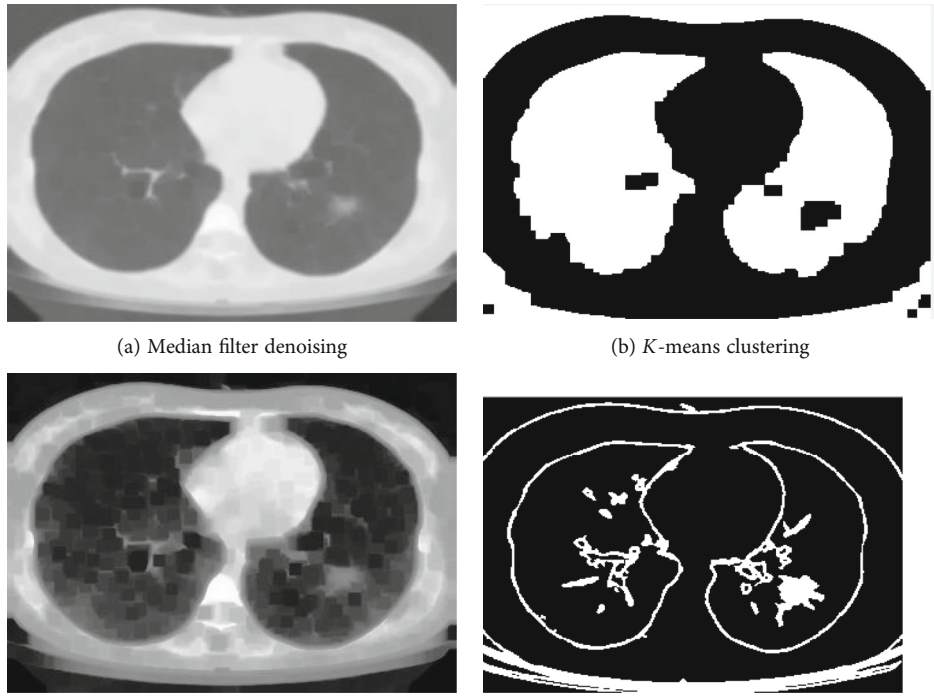


Figure 3: Improved Canny algorithm.

gradient intensity change. Comparison of gradient intensity between current and other directions, the image boundary point is known as the current is greater than the other directions of gradient intensity. Otherwise, the gradient intensity threshold is set to 0, which can eliminate this point. Suppose P is the gradient intensity of the current pixel, P_1 and P_2 are the gradient intensity in positive and negative directions, respectively. The pseudocode for non-maximum suppression is as follows

(4) Dual-threshold and hysteresis boundary track. Dual-threshold means setting high and low thresholds. The high threshold is greater than the pixel gradient value, and it is a strong edge and namely image edge. The low threshold is less than the pixel gradient value, and it is considered not to be the edge of the image and discarded. Weak edge points are defined as the pixel value varies between high and low thresholds. Finally, hysteresis boundary tracking is applied to determine whether the weak edge is an image edge



(a) Median filter denoising

(b) K-means clustering

(c) Mathematical morphology processing

(d) OTSU algorithm

Figure 4: Experimental result images.

Table 2: the original pixel template.

	Column 1	Column 2	Column 3
Row 1	g_{11}	g_{12}	g_{13}
Row 2	g_{21}	g_{22}	g_{23}
Row 3	g_{31}	g_{32}	g_{33}

Table 3: Pixel template after processing.

	Column 1	Column 2	Column 3
Row 1	Max_{11}	Med_{12}	Min_{13}
Row 2	Max_{21}	Med_{22}	Min_{23}
Row 3	Max_{31}	Med_{32}	Min_{33}

The traditional Canny algorithm may causes some problems while detecting the edge of the image.

First of all, the variance of Gaussian filtering determines the image smoothing effect and noise removal effect. Generally speaking, the size of the variance is fixed. If the variance is too large, Gaussian filtering will help to remove noise, but it will cause serious loss of image detail, resulting in blurred image edge; if the variance is too small, the denoising effect of Gaussian filtering will be poor.

Secondly, the Sobel operator is used to calculate the gradient amplitude and direction, and the obtained edge is relatively thick.

Finally, the dual-threshold detection method will generate more false edge, which will break the target edge and lose the edge information.

3.3. Improved Canny Algorithm. We use the median filter, K-means algorithm, mathematical morphology, and OTSU algorithm to improve the classical Canny algorithm.

Figure 3 is a flowchart of the improved Canny algorithm. Firstly, the Gaussian filter in the Canny algorithm is modified to the median filter. Next, based on the K-means, the Sobel operator is utilized to calculate the gradient magnitude and direction. Then, mathematical morphology is applied to thin out the edge. Finally, after nonmaximum suppression processing, the OTSU algorithm is used to determine the threshold automatically. The steps of improvement are as follows.

The experimental results obtained after median filter, K-means clustering, mathematical morphology, and OTSU algorithm are shown in Figures 4(a)–4(d). We can observe that after processing by several methods, the edge of the acquired lesion is very clear.

3.3.1. Median Filter. COVID-19 CT image is complicated to understand; so, we need to remove undesirable portions, including noise [23]. The isolated noise points are eliminated by the median filter [24]. The median filter arranges the pixel values in the neighborhood of a point from large to small and takes the median value as the new pixel value to output, thereby eliminating isolated noise points. The algorithm is as follows.

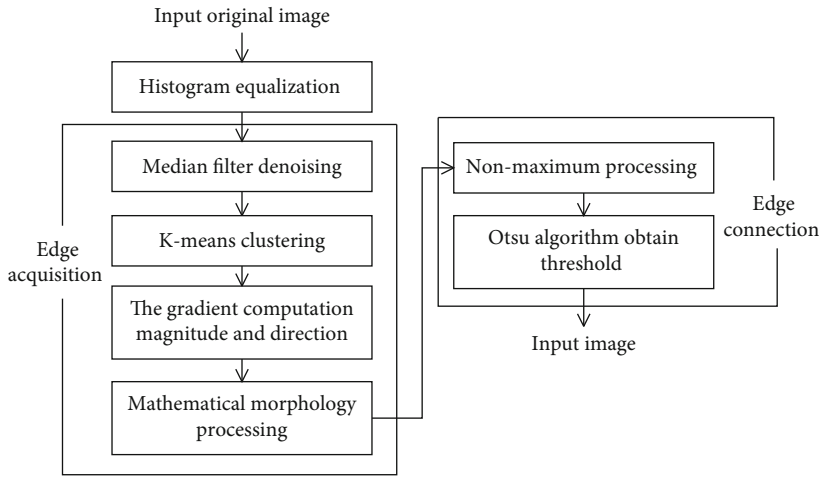
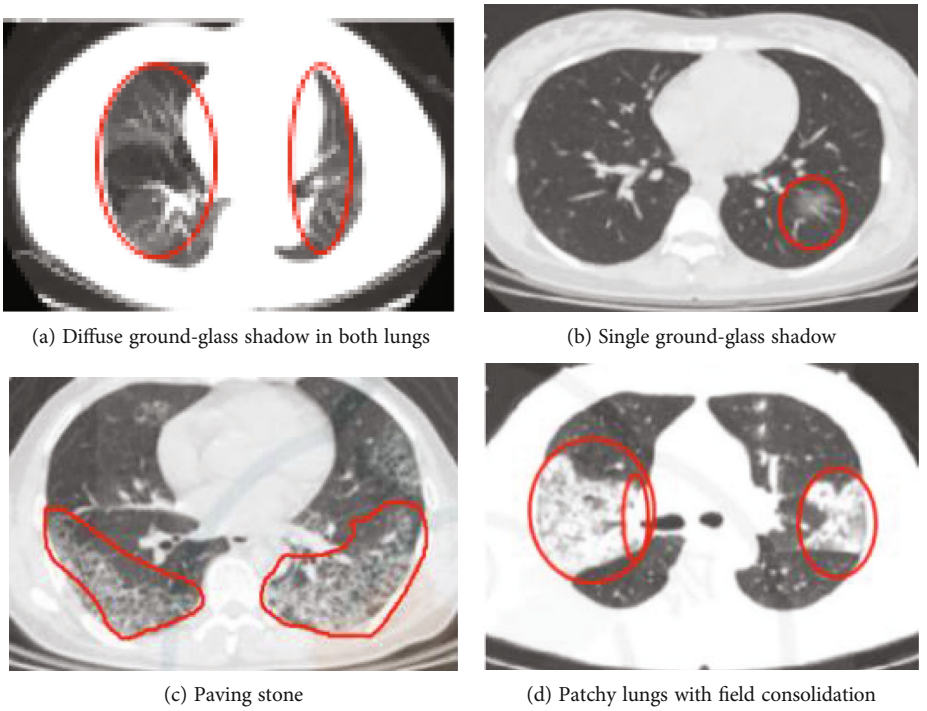


Figure 5: Algorithm flow chart.



(a) Diffuse ground-glass shadow in both lungs

(b) Single ground-glass shadow

(c) Paving stone

(d) Patchy lungs with field consolidation

Figure 6: Partial sample data.

(1) Sort all elements in the array $\rho = [\rho_{11}, \rho_{12}, \rho_{13}, \rho_{21}, \rho_{22}, \rho_{23}, \rho_{31}, \rho_{32}, \rho_{33}]$ from large to small

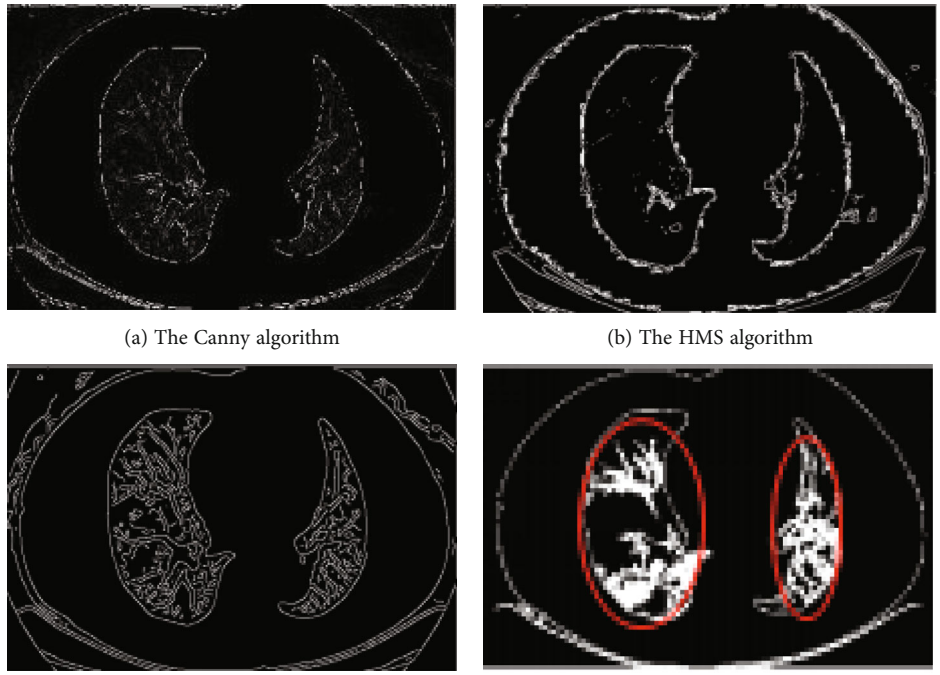
$$\rho = \text{med}\left\{\rho_{1\varphi}, \rho_{2\varphi}, \rho_{3\varphi}\right\}, \tag{8}$$

where $\varphi = [1, 2, 3]$. The commonly used sorting templates are 3×3 and 5×5. In this paper, 3×3 template is cited for processing. The original pixel template is shown in Table 2.

(2) Take the value of each row for comparison and obtain the value of each row maximum, intermediate, and minimum, which are $\text{Max}_{\mu\varphi}$, $\text{Med}_{\mu\varphi}$, and $\text{Min}_{\mu\varphi}$, respectively. At this time, the pixel template is shown in Table 3

(3) Compare the data in each column and get the maximum value, respectively, which is Max_1, Max_2, and Max_3. Finally, by comparing the Max_1, Max_2, and Max_3, the middle value is result

3.3.2. K-Means Algorithm. K-means algorithm is one of the most commonly applied clustering algorithms. By randomly selecting K points as the starting center point, then finding the same characteristic data and letting it be grouped into the same cluster [25]. Cluster ensures that between similar



(a) The Canny algorithm (b) The HMS algorithm

(c) The ETAR algorithm (d) The proposed algorithm

Figure 7: Comparison of the detection of diffuse ground-glass shadow in both lungs.

Table 4: Comparison of detection indexes of diffuse ground-glass shadow in both lungs.

	The Canny algorithm	The HMS algorithm	The ETAR algorithm	The proposed algorithm
MSE	1.1230	8.2723	1.3313	1.0513
SNR	0.0457	0.0475	1.4373	6.6078
MAE	155.2991	157.5637	196.9864	65.0062

sample points are compact, and between different types of samples is scattered so that the clustering effect is optimal.

The main flow of the algorithm is as follows.

(1) Given the sample data w, the set of objects is $A = \{w_1, w_2 \cdots \cdots, w_n\}$, and the initialization center $B = \{C_1, C_2 \cdots \cdots, C_k\}$

According to the principle of the minimum norm of grey value, the sample data point (w, v) is assigned to the cluster closest to it. The calculation formula is as follows.

$$U_i = \min |f(w, v) - C_i|, \tag{9}$$

where U_i is the minimum value of the distance between the sample point $f(w, v)$ and the various centroids C_i.

(2) Calculate the cluster center point again and use it as the new cluster center point T_i, $i = 1, 2 \cdots \cdots k$. Formula is as follows

$$T_i = \frac{\sum_{i\text{-}1}^{m} \{C_i = j\} x_i}{\sum_{i\text{-}1}^{m} \{C_j = j\}}. \tag{10}$$

(3) Repeat steps 2 and 3 until it reaches the maximum number of iterations

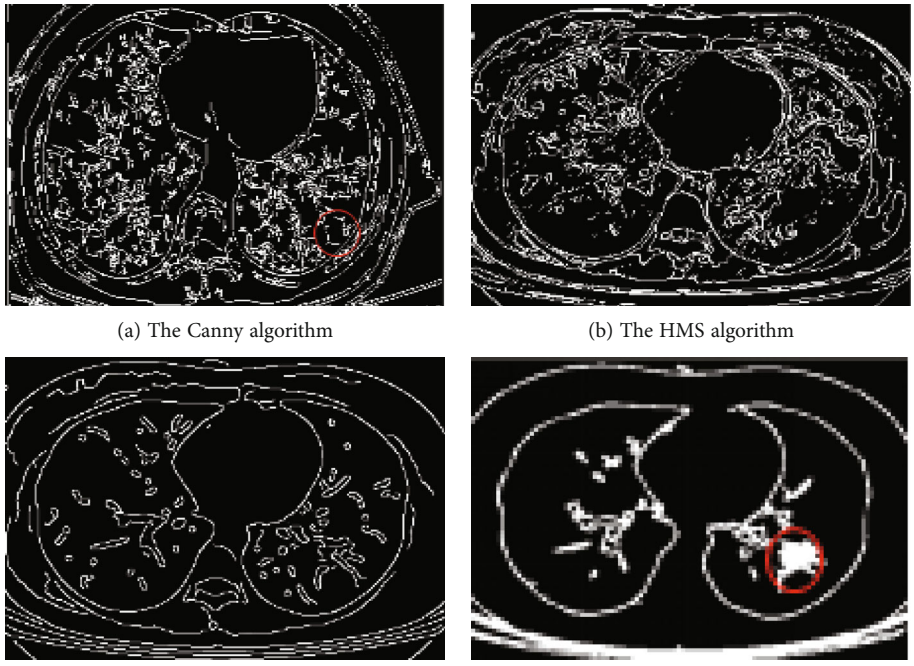
3.3.3. Mathematical Morphology. The basic idea of mathematical morphology is using structural elements to obtain graphical shapes [26]. The main processing processes include expansion, erosion, reconstruction, opening operation, filling, closing operation, and edge extraction. The most basic operations are expansion and corrosion. The most typical operations combining the two are opening and closing operations.

Let T represents an image, $\tau(\alpha, \beta)$ is the gray value at the point (α, β), $g(\sigma, \varepsilon)$ is the structural element of the image, the domains of $\tau(\alpha, \beta)$ and $g(\sigma, \varepsilon)$are D_τ and D_b, and $(\alpha\text{-}\sigma, \beta\text{-}\varepsilon) \in D_\tau$, $(\alpha + \sigma, \beta + \varepsilon) \in D_b$, and $(\sigma, \varepsilon) \in D_b$.

Expansion is used to fill the voids in the graphics and expand the boundaries of the graphics outwards. The calculation formula is as follows.

$$(\tau \oplus g)(\alpha, \beta) = \max \{\tau(\alpha\text{-}\delta, \beta\text{-}\varepsilon) + g(\delta, \varepsilon)\}, \tag{11}$$

where $\oplus$ indicates the expansion operation. The expansion operation is the maximum value of $\tau \oplus g$ in the domain determined by the structural elements.



(a) The Canny algorithm (b) The HMS algorithm

(c) The ETAR algorithm (d) The proposed algorithm

Figure 8: Comparison of single ground-glass shadow detection.

Table 5: Comparison of single ground-glass shadow lesion detection indicators.

	The Canny algorithm	The HMS algorithm	The ETAR algorithm	The proposed algorithm
MSE	5.8147	5.6171	8.3632	2.8982
SNR	2.0091	4.3257	2.3107	7.0879
MAE	157.2183	176.4992	158.5642	55.1343

Corrosion is used to eliminate small, useless edges, and shrink the image boundary inward. The calculation formula is as follows.

$$(\tau \Theta g)(\alpha, y) = \min \{\tau(\alpha + \delta, \beta + \varepsilon) - g(\delta, \varepsilon)\}, \quad (12)$$

where Θ represents the corrosion operation, which is the minimum value of $\tau \Theta g$ in the domain determined by the structural elements.

The step of the open operation is to corrode first and then expand. It can smooth the contour and filter out isolated parts that noncontain structural elements. The calculation formula is as follows.

$$\tau \circ g = (\tau \oplus g) \Theta g. \quad (13)$$

The step of the closed operation is to expand first and then corrode. It can fill in small holes and connect the image boundaries to make them smooth. The calculation formula is as follows.

$$\tau \cdot g = (\tau \Theta g) \oplus g, \quad (14)$$

where $\circ$ represents the open operation, and $\cdot$ represents the closed operation.

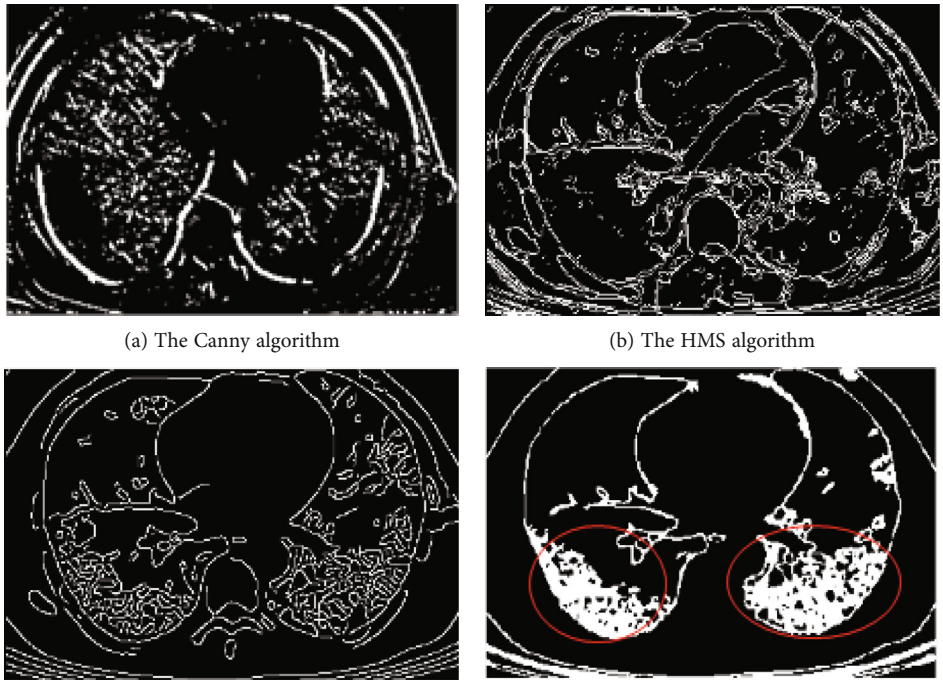
Due to the shrinkage of the open operation and the expanding ability of the closed operation, the results will be biased. In this paper, using equation (15), it is an average method to eliminate the influence of offset.

$$T = \frac{1}{2}(\tau \circ g + \tau \cdot g). \quad (15)$$

3.3.4. OTSU Algorithm. The maximum between-class variance algorithm, also known as the OTSU algorithm, is an algorithm used to automatically determine the image binarization threshold. Because of its simple calculation and not affected by image contrast and brightness, it is widely used in global threshold determination [27]. The basic idea is: according to the gray-scale characteristics of the image, and the image is divided into two parts: the foreground and the background. The difference of value is greater between the foreground and the background, and the effect is better. Based on the input image, the OTSU algorithm automatically calculates the threshold value [28]. In this paper, we firstly set the threshold as 1, then, using the maximum interclass variance to calculate the optimal threshold.

Let γ represent the pixel gray value of a certain point in the image, $[0, R]$ is the pixel gray level, and $n_\gamma$is the number of pixels; thus, the value of the total number of pixels N is as follows.

$$N = \sum_{\gamma=0}^{R} n_\gamma. \quad (16)$$



(a) The Canny algorithm (b) The HMS algorithm

(c) The ETAR algorithm (d) The proposed algorithm

Figure 9: Comparison diagram of paving road shape detection.

Table 6: Comparison table of paving road detection indicators.

	The Canny algorithm	The HMS algorithm	The ETAR algorithm	The proposed algorithm
MSE	9.2358	9.3340	9.3341	1.8103
SNR	0.0460	3.8837	2.7741	7.4247
MAE	173.6431	172.6516	174.6519	67.1646

The probability of each pixel at gray level γ:

$$P_\gamma = \frac{n_\gamma}{N}. \tag{17}$$

The threshold is used to divide the image pixels into two parts: Q_1 and Q_2; then, the probabilities of Q_1 and Q_2 at $[0, L-1]$, and [L-1, R] are D_1 and D_2, and $D_1 + D_2 = 1$.

$$D_1 = \sum_{\gamma=0}^{L-1} P_\gamma, \quad D_2 = \sum_{L}^{R} P_\gamma. \tag{18}$$

The mean values of Q_1 and Q_2 are m_1 and m_2, respectively, and m_θ is the global mean value of the image.

$$m_\theta = D_1 \times m_1 + D_2 \times m_2 \tag{19}$$

The between-class variance expression is as shown in formula (20).

$$\sigma^2 = D_1 \times (m_1\text{-}m_\theta)^2 + D_2 \times (m_2\text{-}m_\theta)^2. \tag{20}$$

3.4. The Proposed Method. In this paper, the proposed weak edge detection method of CT image is mainly divided into three steps. Firstly, enhancing the image detail feature by histogram equalization, secondly, obtaining the weak edge of the lesion on CT image by using median filter, K-means, Sobel operator, and mathematical morphology, and thirdly, combining nonmaximum suppression and OTSU algorithm to connect the edge. The flow of the proposed algorithm is shown in Figure 5.

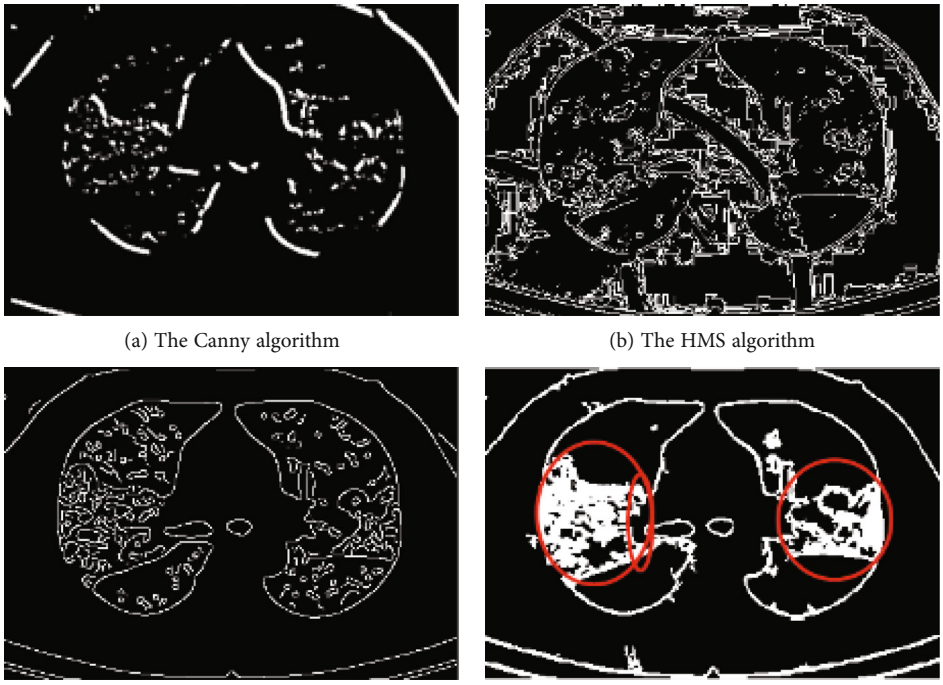
Specific steps are as follows:

(1) Histogram equalization. Performing histogram equalization on the grayscale image to enhance image detail

(2) Edge acquisition. Firstly, using the median filter to preserve the grayscale characteristic of the image and reduce the noise of the image. Then, the K -means algorithm is used to classify the edge of the image for obtaining edge points and nonedge points. After calculating the gradient magnitude and direction, a thick and bright edge is obtained, which further enhances the brightness of the weak edge. Finally, the edge is refined using mathematical morphology to obtain more accurate edge

(3) Edge connection. After nonmaximum value processing, the OTSU algorithm is applied to automatically obtain the best threshold and connect weak edge

3.5. Measure. In order to objectively evaluate the performance of the Canny algorithm, the HMS algorithm, the ETAR algorithm (Benhamza and Seridi [29] proposed



(a) The Canny algorithm (b) The HMS algorithm

(c) The ETAR algorithm (d) The proposed algorithm

Figure 10: Comparison of the detection of patchy lungs with field consolidation.

Table 7: Comparison table of detection indexes for patchy lungs with field consolidation.

	The Canny algorithm	The HMS algorithm	The ETAR algorithm	The proposed algorithm
MSE	1.3192	5.617	1.3313	1.1690
SNR	0.0397	3.2443	2.8746	10.4839
MAE	195.9766	174.4995	196.9860	41.1912

algorithm, called ETAR from now on), and the proposed algorithm, we select mean square error (MSE), signal-noise ratio (SNR), mean absolute error (MAE), and the running time. MSE is used to evaluate the deviation between the observed value and the true value. The value of MSE is larger, and the detection error is greater. SNR refers to the false detection probability of nonedge points. If the value of SNRis larger, indicating that the false detection rate is lower, the detection effect is better. MAE represents the average of the error between the predicted value and the true value. The value of MAE is smaller, which means the error is smaller and the fusion effect is better. The calculation formulas of MSE, SNR, and MAE are as follows.

$$\begin{aligned} \mathrm{MSE}(I_H, I_W) &= \frac{1}{MN}(I_H(i,j) - I_W(i,j))^2, \\ \mathrm{SNR} &= \log \frac{\sum_{i=1}^{M}\sum_{j-1}^{N}(I_w(i,j))^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(I_H(i,j) - I_w(i,j))^2}, \\ \mathrm{MAE}(I_H, I_W) &= \frac{1}{MN}(I_H(i,j) - I_W(i,j)), \end{aligned} \tag{21}$$

where I_H and I_W represent the fused image and the ideal image, respectively, and I_H and I_W are the pixel gray values of the fused image and the ideal image in the ith row and jth column, respectively.

4. Results and Discussion

In this section, firstly, we introduce selected four types of CT lesion images. Then, by MSE, SNR, MAE, and the running time, we verify performance of the proposed algorithm in these images. And the results are shown in the table and figure below. Specifically, we intend to compare the performance of the proposed algorithm with the following algorithms, including the Canny algorithm, the HMS algorithm, and the ETAR algorithm. In the experiment, we use the COVID-CT for verifying the effect of the proposed algorithm on the denoising ability and weak edge detection. Four types of typical sample images selected from the dataset are shown in Figure 6. Figure 6(a) is diffuse ground-glass shadow in both lungs, and the lesion tissue is relatively unclear. Figure 6(b) is single ground-glass shadow, Figure 6(c) is paving stone, and their lesion's edge is relatively weak. Figure 6(d) is patchy lungs with field consolidation, and the lesion's tissue is clear, but it is tightly connected with the lung tissue.

4.1. Verification of the Detection Effect of Diffuse Ground-Glass Shadow in Both Lungs. Compared the edge detection effect of diffuse ground-glass shadow in both lungs, the results are shown in Figure 7. The boundary of the lesion detected by the Canny algorithm is not continuous. The HMS algorithm detection result image contains a lot of

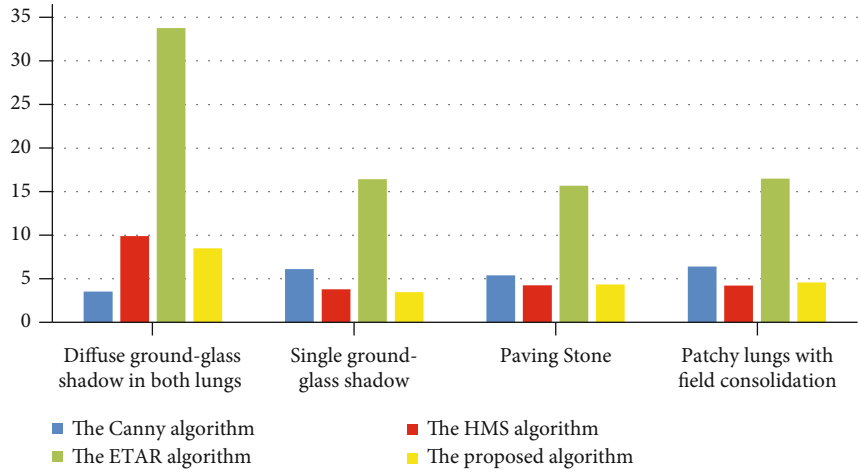


Figure 11: The running time of cases.

noise, and the lesion's detail cannot be detected. The ETAR algorithm can detect the lesion, but the detected lesion feature is different from the actual lesion feature. The proposed algorithm not only effectively removes noise but also clearly shows the characteristics of the lesion.

The MSE, SNR, and MAE of diffuse ground-glass shadows in both lungs are shown in Table 4. The MSE is 1.0513 of the proposed algorithm, and the MSE of other algorithms exceeds 1.1000, which means that the detection error of the proposed algorithm is the smallest. From the SNR values of the four algorithms in Table 4, it can be seen that the SNR of the proposed algorithm is the smallest, indicating detection error is the smallest. The MAE is 65.0062, but the MAE values of all other algorithms exceed 150.0000, suggesting that the average error of the proposed algorithm between the predicted value and the true value is the smallest, and the detection effect is the best.

4.2. Verification of the Detection Effect on Single Ground-Glass Shadow. Figure 8 shows the experimental results of four algorithms for detecting the edge of single ground-glass shadow. We found that the denoising effect and the accuracy of the detected lesion features of the proposed algorithm are better than other three algorithms.

In the experiment, we use three data evaluation indicators to verify the detection effect of the proposed algorithm. We can see from Table 5 that the detection effect of the proposed algorithm is significantly better than the other algorithms. From a data point of view, the MSE, SNR, and MAE of the proposed algorithm are 2.8982, 7.0879, and 55.1343, respectively, while the MSE and MAE of the other algorithms are greater than 5.000 and 150.0000, and the SNR is less than 7.0000, respectively.

4.3. Verification of Detection Effect on Paving Road Shape. For paving road CT image, we compare the edge detection effect of the proposed algorithm with the other three algorithms. The detection effect images are shown in Figure 9. We can see that the lesion feature detected by the proposed algorithm is more clear.

The experimental results are shown in Table 6. After analysis, we can conclude that the experimental results of the proposed algorithm are significantly better than the other three algorithms, which proves that the detection error of the algorithm is the smallest and the detection effect is the best.

4.4. Verification of Detection Effect on Patchy Lungs with Field Consolidation. To verify the effect of edge detection, the Canny algorithm, the HMS algorithm, the ETAR algorithm, and the proposed algorithm are used to carry out experiments on patchy lungs with field consolidation. The result is shown in Figure 10, compared with the other three algorithms, the proposed algorithm has better noise robustness, and the detected lesion feature is more accurate.

In order to evaluate the performance of the proposed algorithm, a comparative analysis is carried out from quantitative perspective. Table 7 shows the data of the four algorithms under the three evaluation indicators. The experimental results show that the proposed algorithm using histogram equalization and improved Canny algorithm has the best denoising effect and the highest accuracy of lesion detection.

4.5. The Running Time. In order to further intuitively evaluate the edge detection effect of the Canny algorithm, the HMS algorithm, the ETAR algorithm, and the proposed algorithm, the edge detection time is calculated of the above cases. Figure 11 gives the result; in terms of the running time of all images, we can observe that the proposed algorithm is shorter as compared to other three algorithms. Therefore, the proposed algorithm is considered to be one of the best edge detection algorithms.

5. Conclusion

In this paper, aiming at the weak edge problem in the edge detection of lesion from COVID-19 CT image, a fusion algorithm is proposed which includes three improvements: (1) usage of histogram equalization, we can obtain the

maximum entropy of the image, thereby improving the clarity of weak edge. (2) The median filter, K-means algorithm, and mathematical morphology are added to the Canny algorithm, which can make the edge of the inspected more accurate. (3) The OTSU algorithm is used to automatically obtain the best threshold. Compared with the Canny algorithm, the HMS algorithm, and the ETAR algorithm, these improvements enable our method to have improved performance. The average values of MSE, MAE, SNR, and the running time of four types of lesion images are 1.7322, 7.9010, 57.1241, and 5.4887, respectively, indicating that our method has better results. Experiments show that the proposed algorithm can effectively extract the weak edge of the lesion.

In the future, we can do further research from the following two aspects: (1) The average running time is 5.1887 seconds, and we needs to further reduce the average running time for the purpose of improving the performance of the proposed method. (2) In order to improve the credibility of the experiment, we will carry on experiment on more data set.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Shou-Ming Hou and Chao-Lan Jia contributed equally to this paper and should be regarded as co-first authors.

Acknowledgments

Henan Key Research and Development Project (182102310629), Royal Society International Exchanges Cost Share Award, UK (RP202G0230), Medical Research Council Confidence in Concept (MRC CIC) Award, UK (MC_PC_17171), and Hope Foundation for Cancer Research, UK (RM60G0680).

References

[1] S. H. Wang, Y. Zhang, X. Cheng, X. Zhang, and Y. D. Zhang, "Psspnn: patchshuffle stochastic pooling neural network for an explainable diagnosis of Covid-19 with multiple-way data augmentation," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 6633755, 18 pages, 2021.

[2] S. Bhattacharya, P. K. Reddy Maddikunta, Q. V. Pham et al., "Deep learning and medical image processing for coronavirus (Covid-19) pandemic: a survey," *Sustainable Cities and Society*, vol. 65, article 102589, 2021.

[3] S. H. Wang, V. V. Govindaraj, J. M. Gorriz, X. Zhang, and Y. D. Zhang, "Covid-19 classification by Fgcnet with deep feature fusion from graph convolutional network and convolutional neural network," *Information Fusion*, vol. 67, pp. 208–229, 2021.

[4] Y. D. Zhang, Z. Dong, S. H. Wang et al., "Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation," *Information Fusion*, vol. 64, pp. 149–187, 2020.

[5] P. Trimboli, C. Camponovo, S. Franscella, E. Bernasconi, and N. Buetti, "Subacute thyroiditis during the covid-19 pandemic: searching for a clinical association with Sars-Cov-2," *International Journal of Endocrinology*, vol. 2021, Article ID 5588592, 4 pages, 2021.

[6] Z. Chen, L. Xu, W. Zhu, and J. Ge, "Cardiovascular manifestations in severe and critical patients withCOVID-19," *Chinese Journal of Frontier Health and Quarantine*, vol. 43, no. 10, pp. 1054–1420, 2020.

[7] F. Ji, X. B. Gao, and S. Y. Xie, "Image weak edge extraction based on mean-shift filtering and histogram enhancement," *Chinese Journal of Image and Graphics*, vol. 17, no. 6, pp. 651–656, 2012.

[8] W. Q. Xue, Z. Y. Zhou, T. Zhang, L. H. Li, and J. Zheng, "Vessel segmentation under non-uniform illumination: a level set approach," *Journal of Software*, vol. 23, no. 9, pp. 2489–2499, 2012.

[9] Q. H. Lin, M. Du, and Y. M. Gao, "Medical image weak edge detection algorithm based on cumulative change rate," *Computer Engineering and Design*, vol. 36, no. 11, 2015.

[10] X. Li, C. Li, A. Fedorov, T. Kapur, and X. Yang, "Segmentation of prostate from ultrasound images using level sets on active band and intensity variation across edges," *Medical Physics*, vol. 43, no. 6, Part1, pp. 3090–3103, 2016.

[11] K. Ungru and X. Jiang, "Dynamic programming based segmentation in biomedical imaging," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 255–264, 2017.

[12] A. Khadidos, V. Sanchez, and C. T. Li, "Weighted level set evolution based on local edge features for medical image segmentation," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1979–1991, 2017.

[13] N. Wang and D. Y. Xu, "Left ventricular image segmentation algorithm for weak edge information," *Computer Engineering and Applications*, vol. 56, no. 23, pp. 211–219, 2020.

[14] C. Kaushal and A. Singla, "Automated segmentation technique with self-driven post-processing for histopathological breast cancer images," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 4, pp. 294–300, 2020.

[15] X. H. He, X. Y. Yang, S. H. Zhang et al., "Covid-Ct-dataset a Ct scan dataset about covid-19," 2020, https://github.com/UCSD-AI4H/COVID-CT.

[16] C. L. Zhao, J. W. Dong, B. Xu, and X. F. Ma, "Research on fog image enhancement algorithm combining histogram equalization and homomorphic filtering," *Journal of Harbin University of Science and Technology*, vol. 24, no. 6, pp. 97–101, 2019.

[17] J. Xiong, D. Yu, Q. Wang et al., "Application of histogram equalization for image enhancement in corrosion areas," *Shock and Vibration*, vol. 2021, Article ID 8883571, 13 pages, 2021.

[18] S. Raza, T. ur Rahman, M. Saeed, and S. Jameel, "Performance analysis of power system parameters for islanding detection using mathematical morphology," *Ain Shams Engineering Journal*, vol. 12, no. 1, pp. 517–527, 2021.

[19] M. Lohvithee, W. Sun, S. Chretien, and M. Soleimani, "Ant colony-based hyperparameter optimisation in total variation reconstruction in X-ray computed tomography," *Sensors*, vol. 21, no. 2, p. 591, 2021.

[20] V. Osma-Ruiz, J. I. Godino-Llorente, N. Sáenz-Lechón, and P. Gómez-Vilda, "An improved watershed algorithm based on efficient computation of shortest paths," *Pattern Recognition*, vol. 40, no. 3, pp. 1078–1090, 2007.

[21] D. Sangeetha and P. Deepa, "FPGA implementation of cost-effective robust canny edge detection algorithm," *Journal of Real-Time Image Processing*, vol. 16, no. 4, pp. 957–970, 2019.

[22] S. Madireddy, D. W. Chung, T. Loeffler et al., "Phase segmentation in atom-probe tomography using deep learning-based edge detection," *Scientific Reports*, vol. 9, no. 1, article 20140, 2019.

[23] W. W. Wu, D. L. Hu, K. An, S. Y. Wang, and F. L. Luo, "A high-quality photon-counting Ct technique based on weight adaptive total-variation and image-spectral tensor factorization for small animals imaging," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.

[24] B. Garg, "Restoration of highly salt-and-pepper-noise-corrupted images using novel adaptive trimmed median filter," *Signal, Image and Video Processing*, vol. 14, no. 8, pp. 1555–1563, 2020.

[25] D. Holilah, A. Bustamam, and D. Sarwinda, "Detection of Alzheimer's disease with segmentation approach using K-means clustering and watershed method of Mri image," *Journal of Physics: Conference Series*, vol. 1725, article 012009, 2021.

[26] S. Rani, "A novel mathematical morphology based edge detection method for medical images," *CSI Transactions on ICT*, vol. 4, no. 2-4, pp. 217–225, 2016.

[27] Q. Lin, W. Jing, Y. L. Chen, and Z. Z. Wang, "Research on the image segmentation of icing line based on Nsct and 2-D Ostu," *International Conference on Modelling*, vol. 57, no. 2, pp. 112–120, 2016.

[28] K. K. Sharma, P. Rai, and K. C. Patidar, "Ct scan image segmentation based on hounsfield unit values using Otsu thresholding method," *Journal of Physics: Conference Series*, vol. 1816, no. 1, article 012080, 2021.

[29] K. Benhamza and H. Seridi, "Canny edge detector improvement using an intelligent ants routing," *Evolving Systems*, vol. 12, no. 2, pp. 397–406, 2021.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 5221111, 10 pages
https://doi.org/10.1155/2021/5221111

Research Article

A Segmentation Method of Foramen Ovale Based on Multiatlas

Jiashi Zhao[1,2] **Huatao Ge**[1] **Wei He**[1,2] **Yanfang Li**[1,2] **Weili Shi**[1,2] **Zhengang Jiang**[1,2] **Yonghui Li**[1] **and Xingzhi Li**[3]

[1]*School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China*
[2]*Zhongshan Institute of Changchun University of Science and Technology, Zhongshan 528436, China*
[3]*Bethune First Hospital of Jilin University, Changchun 130012, China*

Correspondence should be addressed to Weili Shi; shiweili@cust.edu.cn

Received 13 May 2021; Accepted 25 August 2021; Published 20 September 2021

Academic Editor: Yuankai Huo

Trigeminal neuralgia is a neurological disease. It is often treated by puncturing the trigeminal nerve through the skin and the oval foramen of the skull to selectively destroy the pain nerve. The process of puncture operation is difficult because the morphology of the foramen ovale in the skull base is varied and the surrounding anatomical structure is complex. Computer-aided puncture guidance technology is extremely valuable for the treatment of trigeminal neuralgia. Computer-aided guidance can help doctors determine the puncture target by accurately locating the foramen ovale in the skull base. Foramen ovale segmentation is a prerequisite for locating but is a tedious and error-prone task if done manually. In this paper, we present an image segmentation solution based on the multiatlas method that automatically segments the foramen ovale. We developed a data set of 30 CT scans containing 20 foramen ovale atlas and 10 CT scans for testing. Our approach can perform foramen ovale segmentation in puncture operation scenarios based solely on limited data. We propose to utilize this method as an enabler in clinical work.

1. Introduction

Trigeminal neuralgia is a neurological disease that occurs mostly in one or more branches of the facial unilateral trigeminal nerve. The pain is similar to electric strike or tingling-like and is asymptomatic in intermittent periods. It is mainly primary trigeminal neuralgia. In a few cases, trigeminal neuralgia can be secondary to brain tumors or vascular abnormalities [1]. An epidemiological survey study in the United States showed that the incidence of trigeminal neuralgia in men is 2.5 per 100,000, and the incidence in women is 5.7 per 100,000 [2]. The peak prevalence is between 50 and 60 years old, and the prevalence rate increases with age. Among people over 80, the incidence rate is 25.9/100,000 per year [3]. Under normal circumstances, speaking, chewing, brushing teeth, shaving, or even a cool breeze may cause short-term attacks in some patients. The disease causes great trouble to the patients' daily life and easily causes anxiety and depression emotions, even suicide [4].

Clinically, the treatment of trigeminal neuralgia is mainly based on doctors' experience knowledge for puncture and computer-assisted puncture based on radiological information. During the puncture process, the puncture needle needs to pass through the skin and enter the semilunar ganglia from the foramen ovale of the skull. However, the narrow foramen ovale of the skull base in different patients and their different shapes pose great challenges to the surgical process. The key to this operation is to accurately locate the foramen ovale position during the operation [5, 6]. The position of the foramen ovale in the skull base is shown in the red area in Figure 1. Puncture based on empirical knowledge has a high failure rate and a high demand on doctors. Computer-assisted puncture needs to determine the specific location of the foramen ovale before surgery [7]. It takes a long time to manually mark the location and relies on the doctor's personal experience, which is highly subjective. Therefore, precise and rapid segmentation of the foramen ovale at the skull base can effectively improve the success

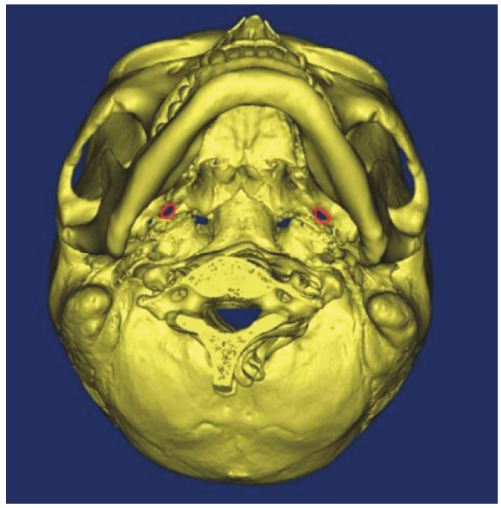


Figure 1: Bottom view of the skull base.

rate of puncture, reduce the length of operation, and relieve the pain of patients.

Medical image segmentation methods widely used domestically and internationally are mainly divided into traditional methods and methods based on deep learning. In traditional methods, the threshold method, region growth method, and map segmentation method are mostly used. The threshold method and the region growing method are simple to implement, but the threshold method is very sensitive to noise and uneven grayscale. The region growing method needs to manually provide a seed point. This method is also very sensitive to noise and may produce discontinuous regions. The two methods are to segment the whole, while our target area is a part of the whole skull. The most popular segmentation method is to use deep learning for segmentation, but deep learning requires a lot of data sets [8], which can explain poorly. In recent years, the atlas segmentation method has gradually become one of the effective methods in the field of medical image segmentation [9]. The work of many scholars has shown that prior knowledge of anatomy can help segment brain images with complex structures, low target area boundary contrast, and large intersubject and intrasubject variance. The atlas-based segmentation method utilizes the most anatomical prior knowledge. In the process of registration, the method minimizes or eliminates the influence of various kinds of noise on the segmentation results and has good robustness. Asim et al. [10] use a multiatlas method to divide the brain according to different atlases and then combine the features extracted from these anatomical units to comprehensively and accurately detect Alzheimer's disease. Bao et al. [11] proposed a multimode and multiatlas feature representation method and used a two-step feature selection method to select the most characteristic features for the classification of schizophrenia. Tor-Díez et al. [12] used a multiatlas segmentation method for the analysis of children's brain structure. The cortex is the region of interest for this problem. They proposed a block-based nonlocal model and iterative optimization scheme, which can provide reliable cortical segmentation. As a result, it is of great significance in predicting children's developmental health information. Su et al. [13] proposed a multiatlas segmentation method optimized for the thalamus, which can accurately quantify the thalamus and volume and can track the development of some neurological diseases in time. Boucher et al. [14] realized the automatic segmentation of the lateral ventricle by using a deformable multiatlas segmentation algorithm for ultrasound and MRI fusion using local linear correlation metrics, which can be used to evaluate the brain development of newborns. van der Heyden et al. [15] used a multiatlas method to automatically segment the healthy tissue around the tumor during radiotherapy, which improved the current situation of clinical doctors manually describing the healthy tissue. Tang et al. [16] introduced the multiatlas segmentation method to the segmentation of brain tumor images and adopted a new low-rank method that uses spatial constraints to obtain restored images containing normal brain regions.

This article comprehensively determines the effectiveness of the multiatlas segmentation technology for the segmentation of the skull base foramen ovale and proposes a segmentation method for skull base foramen ovale based on multiatlas. Through a large number of investigations and studies, we find that we apply the multiatlas segmentation method to the segmentation of the foramen ovale in the skull base for the first time and created the foramen ovale atlas data set of the skull base for the first time. This method selects the 10 atlas images that are most similar to the image to be segmented from the atlas set according to the normalized cross-correlation similarity measure, and then, the method based on multiresolution affine transformation and multiresolution B-spline transformation is used to perform coarse registration and fine registration on the image to be segmented and the image selected from the atlas. Finally, the STAPLE [17] algorithm is used to fuse the label images to obtain the final predicted segmentation results. We also compared the segmentation effects of the MV [18] algorithm and the SIMPLE [19] algorithm. The results show that the segmentation method based on atlas can be applied to the segmentation of the foramen ovale at the base of the skull. We have completed the segmentation of the foramen ovale at the base of the skull under low data conditions, with high accuracy to meet the needs of clinical surgery.

2. Materials and Methods

The multiatlas segmentation method has gradually become one of the commonly used methods in the field of medical image segmentation. This method has three steps: image similarity selection, multiatlas registration, and label fusion.

2.1. Atlas Segmentation Method. The atlas consists of two parts: a grey image and its corresponding manually segmented label image. The segmentation method based on the atlas is equivalent to transforming the segmentation problem into a registration problem. In the image registration, the floating image is matched with the fixed image through deformation. The image to be segmented here is used as the fixed image, and the image selected from the atlas for registration with the fixed image is used as the

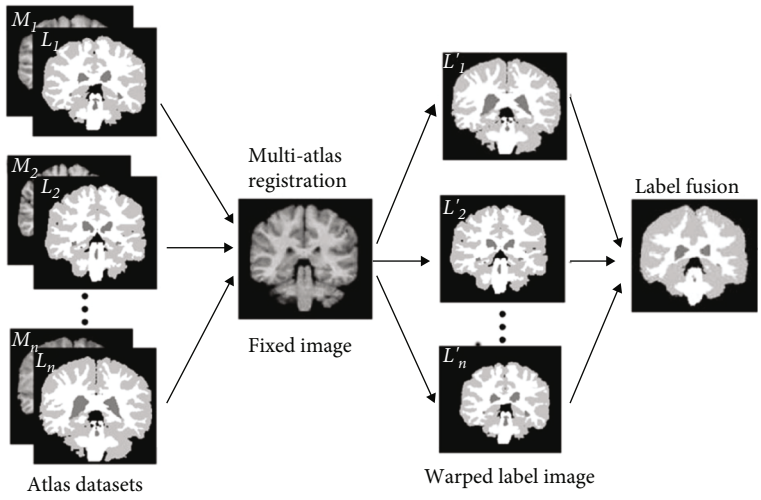


Figure 2: Schematic diagram of multiatlas segmentation method.

floating image. According to the number of atlases required for registration, atlas segmentation is divided into single atlas segmentation and multiple atlas segmentation. The steps of multiatlas segmentation are mainly divided into three steps. Firstly, we find several moving images that are most similar to the fixed image from the atlas. Secondly, it is indispensable to register the selected moving image with the fixed image to obtain the corresponding transformation matrix T and then apply the transformation matrix T to the marked image corresponding to the moving image. Finally, label fusion is performed on all the transformed atlas label images to obtain the final segmentation result. The segmentation process is illustrated Figure 2.

Human brain images are more complex, and the use of multiple atlases for registration and fusion largely compensates for the insufficient registration effect that may be caused by the inappropriate selection of a single atlas. However, Aljabar et al. [20] found that the segmentation accuracy does not completely increase with the increase of the number of atlases, and the more the number of atlases, the time for segmentation calculation will also increase linearly. Awate et al.'s research [21] shows that the most appropriate number of atlases is about 10. Therefore, this article will select 10 moving images from the atlas for registration with the fixed image.

2.2. Registration Technology. Registration is a crucial part of the multiatlas segmentation process. The quality of the registration algorithm has a direct impact on the final segmentation result. The registration process in this article is divided into two steps. The first step is to use a registration method based on multiresolution affine transformation to act on the reference image and the floating image for coarse registration, and the second step uses the registration method based on multiresolution B-spline transformation to perform the fine registration on the fixed image and the moving image.

Multiresolution is a strategy often used in medical image registration. It refers to sampling the image to increase or decrease the resolution of the image, so that it is convenient for further processing of the image. Firstly, the medical image is smoothly processed by a low-pass filter to prevent the image from being interfered by noise during the acquisition and transmission process, improves the quality of the medical image, and obtains an image with a constant scale. Then, downsample the fixed image and the moving image. The image can generate several images with different resolutions to form an image pyramid. Hierarchical registration is essentially a coarse-to-fine registration strategy. At the beginning, the optimal parameters are searched for in the low-resolution layer. Although the image information of this layer is not complete and the registration accuracy is not high, the registration parameters obtained are close to the optimal solution, and the amount of image data of this layer is small, which reduces the time required for registration. After multilevel registration, accurate image registration results can be obtained in the last layer, and at the same time, local convergence problems that occur during single-level registration can be avoided. The multiresolution registration flow chart is shown in Figure 3.

The B-spline transformation function achieves the effect of nonrigid registration by moving the control points, which can control local deformation. The specific displacement of the control point is calculated by the optimization algorithm, so as to achieve the effect of simulating any nonlinear transformation. First, the fixed image is gridded, and the points on the image become control points after gridding. We assume that the position of a control point in the two-dimensional image is $\varphi_{i,j}$ and the grid spacing is $\delta_x \times \delta_y$; based on the consideration of accuracy and efficiency, the uniform third-order B-spline basis function is usually selected for image registration, then the B-spline transformation of any point (x, y) on the moving image can be expressed as

$$T(x,y) = \sum_{m=0}^{3} \sum_{n=0}^{3} B_m(u) B_n(v) \varphi_{i+m,j+n}. \quad (1)$$

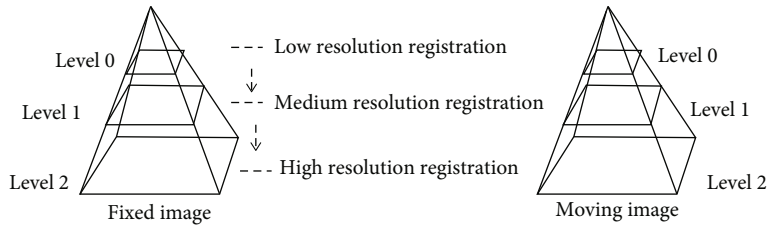


Figure 3: Schematic diagram of multiresolution registration.

In the formula, $\varphi_{i+m,j+n}$ represents the coordinate positions of the nearest 4×4 control points; the i and the j, respectively, represent the position index of the adjacent control points, $i = \lfloor x/\delta_x \rfloor - 1$, $j = \lfloor y/\delta_y \rfloor - 1$; $\lfloor \rfloor$ represents the round-down function; m and n are the order of B-spline basis functions; u and v are the relative unit control grid positions of (x, y), $u = (x/\delta_x) - \lfloor x/\delta_x \rfloor$, $v = (y/\delta_y) - \lfloor y/\delta_y \rfloor$; $B_m(\mathrm{u})$ represents the m-th B-spline basis function; and the expressions are as

$$\begin{cases} B_0(\mathrm{u}) = \dfrac{(1-u)^3}{6}, \\ B_1(u) = \dfrac{3u^3 - 6u^2 + 4}{6}, \\ B_2(u) = \dfrac{-3u^3 + 3u^2 + 3u + 1}{6}, \\ B_3(u) = \dfrac{u^3}{6}. \end{cases} \tag{2}$$

Among them, $0 \le u < 1$, these functions act as weighting functions, and they weight the influence of each control point on $T(x, y)$ according to the distance from the control point to (x, y).

2.3. Label Fusion. This paper uses the STAPLE algorithm to complete the label fusion step. The STAPLE algorithm uses the maximum expectation algorithm iteration to estimate the performance parameters and probability distribution. In the fusion process, it is equivalent to treating each atlas as a weak classifier, using the maximum expectation estimation to set the weight of each classifier and then fusing to obtain the final segmentation result. At the same time, we used the majority voting algorithm (MV) for tag fusion and the SIMPLE method to complete the fusion as a comparison experiment. The MV algorithm is a method to determine the final fusion label value according to the criterion that the minority obeys the majority. The SIMPLE method combines atlas selection and evaluation strategies and gradually reduces the number of maps through selective iteration to achieve a good fusion effect.

3. Results and Discussion

3.1. Construction of Data. In this paper, on the human skull CT, the foramen ovale on the left and right sides of the skull base are segmented separately. The experimental data comes from the Second Hospital of Jilin University, and the inclusion criteria are (1) a complete whole skull and (2) people who are 20 years old and above. A total of 30 CT data are obtained by screening according to the above criteria. Then, under the guidance of professional physicians, 20 data are selected to make the atlas, and the remaining 10 data are used for experimental testing. (Note: all data were obtained with the patient's knowledge and consent.) The preparation steps of the atlas are as follows: firstly, the threshold method and the region growing method are applied to process the CT data. Secondly, the foramen ovale area at the base of the skull was manually segmented. Finally, a slight Gaussian smoothing on the data is performed. A set of atlases contains atlas images and their corresponding label images. One of the sets of the atlas made is shown in Figure 4.

3.2. Experiment and Parameter Setting. In the entire experimental process, firstly, the 10 images with the highest similarity to the fixed image are selected in the atlas using the normalized cross-correlation similarity measurement method for registration. The normalized cross-correlation formula is defined as follows:

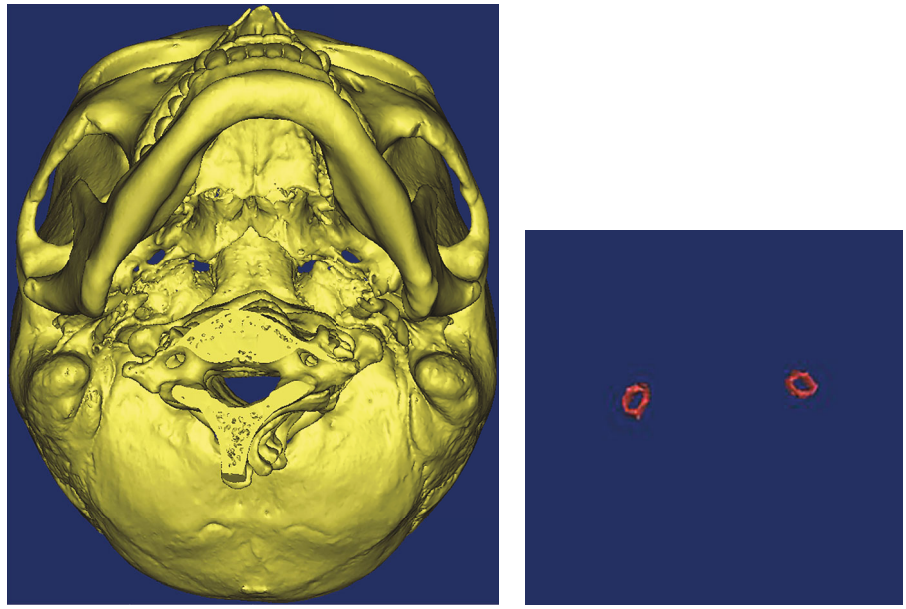
$$\mathrm{NCC}(\tau, \mathrm{TI}, \mathrm{FI}) = \frac{\sum_{i=1}^{n} \left(\mathrm{TI}(x_i) - \overline{\mathrm{TI}}\right) \times \sum_{i=1}^{n} \left(\mathrm{FI}(\tau(x_i)) - \overline{\mathrm{FI}(\tau)}\right)}{\sqrt{\sum_{i=1}^{n} \left(\mathrm{TI}(x_i) - \overline{\mathrm{TI}}\right)^2 \times \sum_{i=1}^{n} \left(\mathrm{FI}(\tau(x_i)) - \overline{\mathrm{FI}(\tau)}\right)^2}}, \tag{3}$$

$$\overline{\mathrm{TI}} = \sum_{i=1}^{n} \frac{TI(x_i)}{n}, \tag{4}$$

$$\overline{\mathrm{FI}(\tau)} = \sum_{i=1}^{n} \frac{\mathrm{FI}(\tau(x_i))}{n}, \tag{5}$$

where $\mathrm{TI}(x_i)/n$ represents the gray value of pixel x_i in the fixed image, $\mathrm{FI}(x_i)/n$ represents the gray value of pixel x_i in the moving image, and n represents the number of image pixels; considering that CT data may come from different imaging equipment, there are nonstandard intensities between images, so it is selected as the atlas selection criterion.

In the registration process, we use the Elastix [22] toolkit to perform registration based on affine transformation and B-spline transformation. In the above two registration processes, a multiresolution strategy is used. The image is first smoothed by Gaussian kernel filtering, and then, downsampling by a factor of 2 is used for each resolution layer. Considering the generation effect and speed, the interpolation method adopts the linear interpolation method, and the



(a) Atlas image (b) Label image

Figure 4: A set of the atlas manually segmented.

interpolation method used to generate the final segmentation result adopts the third-order B-spline interpolation method. During each iteration, a random sampling method is used to randomly select 2000 voxels to calculate the normalized mutual information value between the images, which improves the speed of the registration optimization parameters, and uses the gradient descent optimization algorithm to optimize the normalized mutual information value. For affine transformation registration, each layer is set to 1000 iterations, which is set to 4 layers. For B-spline registration, each layer is set to 3000 iterations, and a grid spacing of 5 mm is used, which is set to 5 layers.

In the label fusion process, the STAPLE algorithm is used to fuse a single prediction result to obtain the final segmentation result. Here, we also use the MV algorithm and the SIMPLE algorithm for comparison experiments.

3.3. Evaluation Index. After obtaining the segmentation results, it is necessary to adopt appropriate evaluation indicators to evaluate the segmentation results of different methods. This paper uses Dice coefficient, 95% Hausdorff distance, and average surface distance (ASD) to verify the accuracy of the segmentation results. The Dice coefficient is used to measure the relative volume overlap between the algorithm segmentation results and the manual segmentation results, and the latter two evaluation standards are used to measure the consistency between the segmentation boundaries. The higher the Dice coefficient, the better the segmentation result. The smaller the Hausdorff distance and the average surface distance, the better the segmentation result. Its definition is as Equations (6), 7, 8, 9, and 10:

$$\text{Dice}(A,B) = \frac{2V(A \cap B)}{V(A) + V(B)}, \tag{6}$$

where $V(A)$ and $V(B)$ represent the predicted segmentation result and the volume of the doctor's manual segmentation result, respectively. $V(A \cap B)$ represents the volume of the above overlapping part.

$$\text{ASD} = \frac{1}{|S(A)| + |S(B)|}\left(\sum_{a \in S(A)} \min_{b \in S(B)} \|a - b\| + \sum_{b \in S(B)} \min_{a \in S(A)} \|b - a\|\right), \tag{7}$$

where $S(A)$ represents the set of surface voxels of the predicted segmentation result, $S(B)$ represents the set of surface voxels of the doctor's manual segmentation result. a and b, respectively, represent a voxel subset of the two voxel sets.

$$H(A,B) = \max\left(h(A,B), h(B,A)\right), \tag{8}$$

$$h(A,B) = \max_{a \in A}\left\{\min_{b \in B}\|a - b\|\right\}, \tag{9}$$

$$h(B,A) = \max_{b \in B}\left\{\min_{a \in A}\|b - a\|\right\}, \tag{10}$$

where $\|\cdot\|$ represents the Euclidean distance between the two points a and b.

Testing on 10 data, we use the MV, STAPLE, and SIMPLE methods to perform experiments on the left and right foramen ovale at the base of the skull and then display the average of average Dice, 95% Hausdorff and average surface distance obtained by different methods on the three-line graph obtained by different methods on an average three-line graph. In order to visually show the difference of the segmentation effect obtained by different methods, we also draw box plots of the three methods on each evaluation index.

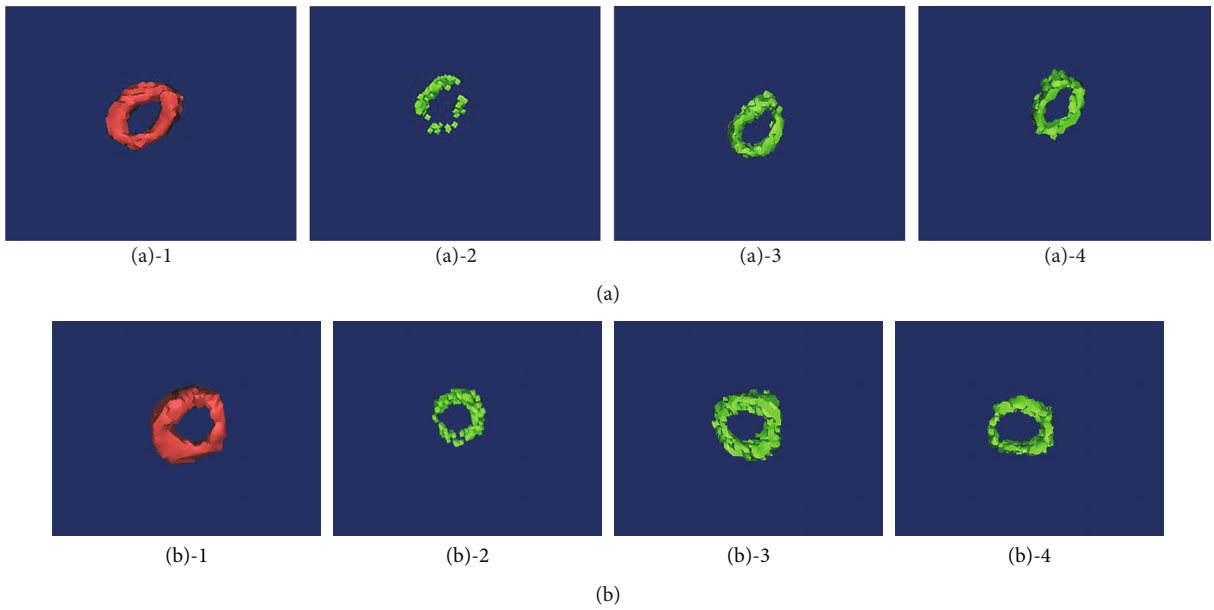


Figure 5: Manual segmentation and the results of segmentation of the foramen ovale structure of the skull base by various algorithms. (a) Segmentation result on the left. (b) Segmentation result on the right. Number 1 is manually segmented images, number 2 is MV algorithm segmentation image, number 3 is STAPLE algorithm segmentation image, and number 4 is SIMPLE algorithm segmentation image.

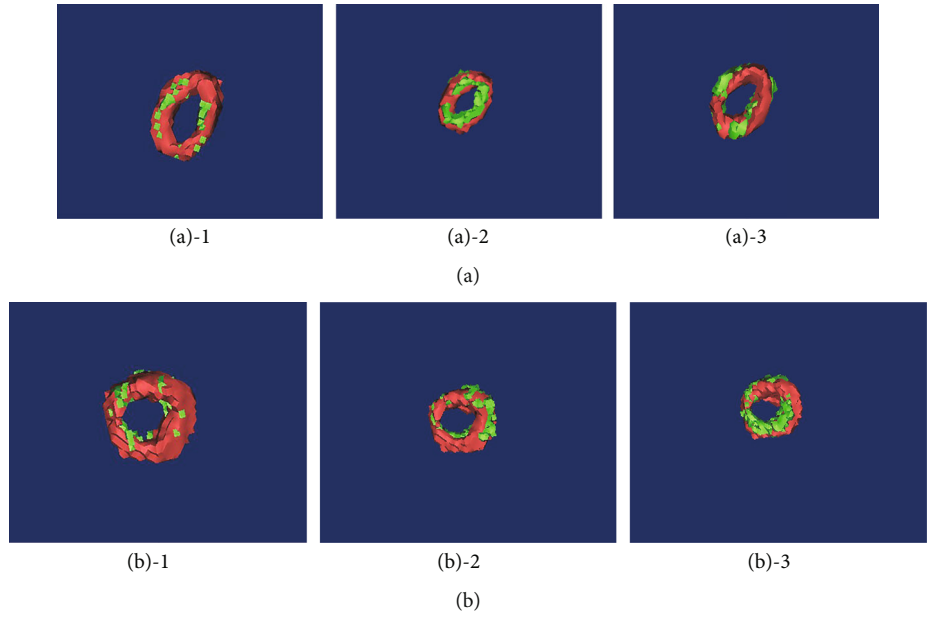


Figure 6: Comparison of the results of each method and manual segmentation. (a) Comparison between segmentation method and manual segmentation on the left. (b) Comparison between segmentation method and manual segmentation on the right. Number 1 is a comparison between MV algorithm and manual segmentation, number 2 is a comparison between STAPLE algorithm and manual segmentation, and number 3 is a comparison between SIMPLE algorithm and manual segmentation.

3.4. Experimental Results. Segmentation results of the foramen ovale at the skull base and manual segmentation results are shown in Figure 5. The experimental results show that the three methods can be used to segment the foramen ovale.

The comparison chart of each method and manual segmentation is shown in Figure 6. Red is the result of manual segmentation, and green is the segmentation result of different methods. We can see that the segmentation effect of the MV algorithm is not good, the segmented oval foramen has a discontinuity problem, and the segmentation result is incomplete, which is quite different from the manual segmentation result. STAPLE and SIMPLE segmentation results are better.

3.5. Data Analysis and Discussion. The average value of Dice which is obtained from 10 groups of data tested by three

methods, MV, STAPLE, and SIMPLE, is shown in Table 1. It can be seen from Table 1 that the average Dice of the foramen ovale on the left of the MV algorithm is 0.790, and the average Dice of the foramen ovale on the right is 0.803. The average Dice of the foramen ovale on the left of the STAPLE algorithm is 0.858, and the right is 0.870. The average Dice of the foramen ovale on the left of the SIMPLE algorithm is 0.853, and the right is 0.871.

The average value of the 95% Hausdorff distance which obtained from 10 groups of data tested by the above three methods is shown in Table 2. From Table 2, it can be indicated that the average 95% Hausdorff distance of the left oval foramen of the MV algorithm is 5.054, and the right is 3.639. The average 95% Hausdorff distance of the left oval foramen of the STAPLE algorithm is 4.274, and the right is 3.452. The average 95% Hausdorff distance of the foramen ovale on the left of the SIMPLE method is 4.644, and right is 3.227.

The average value of ASD which obtained from 10 groups of data tested by the above three methods is shown in Table 3. From Table 3, it can be demonstrable that the average ASD value of the foramen ovale on the left side of the MV algorithm is 1.258, and the average ASD value of the foramen ovale on the right side of the MV algorithm is 0.933. The average ASD value of the foramen ovale on the left side of the STAPLE algorithm is 0.998, and the right is 0.739. The average ASD value of the foramen ovale on the left side of the SIMPLE algorithm is 1.067, and the right side is 0.728. From the data in Tables 1–3, it can be seen that the Dice of MV algorithm segmentation is lower, and the 95% Hausdorff distance and the ASD are higher, indicating that the MV algorithm segmentation effect is poor. The Dice of the STAPLE and SIMPLE algorithms is higher, and the 95% Hausdorff distance and the ASD are lower, indicating that the STAPLE and SIMPLE algorithms have better segmentation effects.

Figures 7–9 sequentially show the left and right foramen ovale Dice box plots, 95% Hausdorff distance box plots, and ASD box plots drawn by the three methods of MV, STAPLE, and SIMPLE for 10 sets of test data. The maximum value, upper quartile, median, lower quartile, and minimum value are displayed on the box plots, which can reflect the overall characteristics of multiple sets of data. The horizontal line in the box plot represents the median number. From the box plots, we can see that the effect of the MV algorithm is relatively poor. The Dice value of the STAPLE algorithm and the SIMPLE method is above the chart, indicating that a better segmentation effect can be achieved for the test image. The median of 95% Hausdorff distance and ASD of STAPLE and SIMPLE methods is below the chart, which also shows that these two methods can achieve better segmentation results for the test image.

Table 1: Dice average of segmentation results of different methods.

Dice	Left foramen ovale	Right foramen ovale
MV	0.790	0.803
STAPLE	0.858	0.870
SIMPLE	0.853	0.871

Table 2: The average value of 95%Hausdorff distance of the segmentation results of different methods.

95%Hausdorff distance	Left foramen ovale	Right foramen ovale
MV	5.054	3.639
STAPLE	4.274	3.452
SIMPLE	4.644	3.227

Table 3: Average surface distance average of segmentation results of different methods.

ASD	Left foramen ovale	Right foramen ovale
MV	1.258	0.933
STAPLE	0.998	0.739
SIMPLE	1.067	0.728

4. Data Analysis and Discussion

From the above chart data, it can be seen that the segmentation results of the STAPLE method and the SIMPLE method are relatively close, and both are significantly better than MV. This is because the MV method compares and selects the pixel values at the same position of each floating image and chooses the pixel value with the most occurrences as the actual pixels of the position; although the prior information of each floating image is fully considered, all image information is treated equally without considering the difference between each floating image and the fixed image. The SIMPLE method and the STAPLE method consider the difference information between the fixed image and each floating image. The STAPLE method uses the expectation maximization algorithm to calculate the weight coefficient of the floating image, and finally, the weighted average is performed to obtain the final segmentation result. The SIMPLE method assigns weights based on the floating images and the performance level of the fusion result obtained after each iteration are, and in the process of estimating the performance level, floating images with poor performance levels are discarded. Theoretically speaking, the SIMPLE algorithm is better than the STAPLE algorithm [19], because the SIMPLE algorithm discards floating images that do not perform well, but in fact, these floating images may also contain useful information. From the overall experimental results, the segmentation effect of the right foramen ovale is better than that of the left foramen ovale. We think that it may be related to the difference in the structure of the foramen ovale on the left and right sides. But we have consulted relevant academic data, and according to the data, the conclusion is that there is no statistical difference in the length, width, and area of the foramen ovale on both sides [23, 24]. There is currently no exact theory to explain this experimental result.

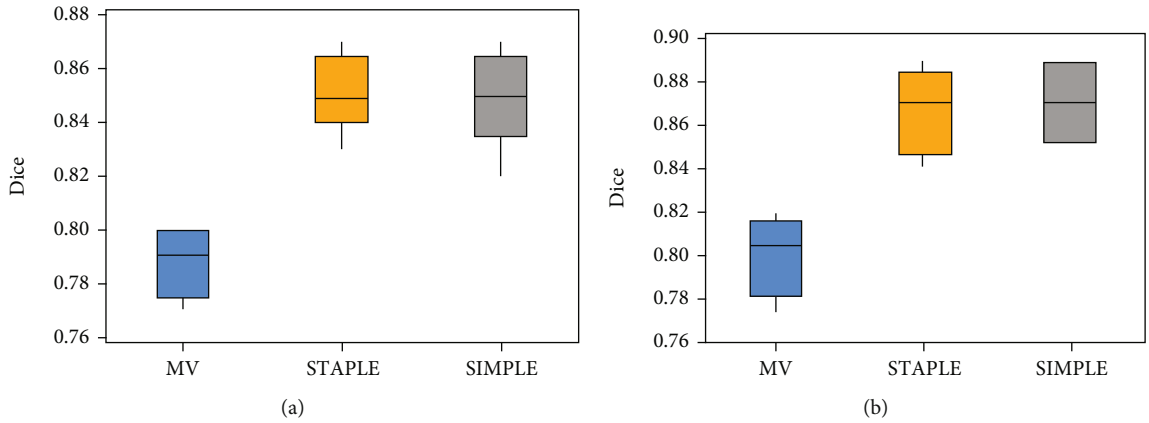


Figure 7: Dice box plots of the foramen ovale on the left and right sides of each method: (a) foramen ovale box plot on the left; (b) foramen ovale box plot on the right.

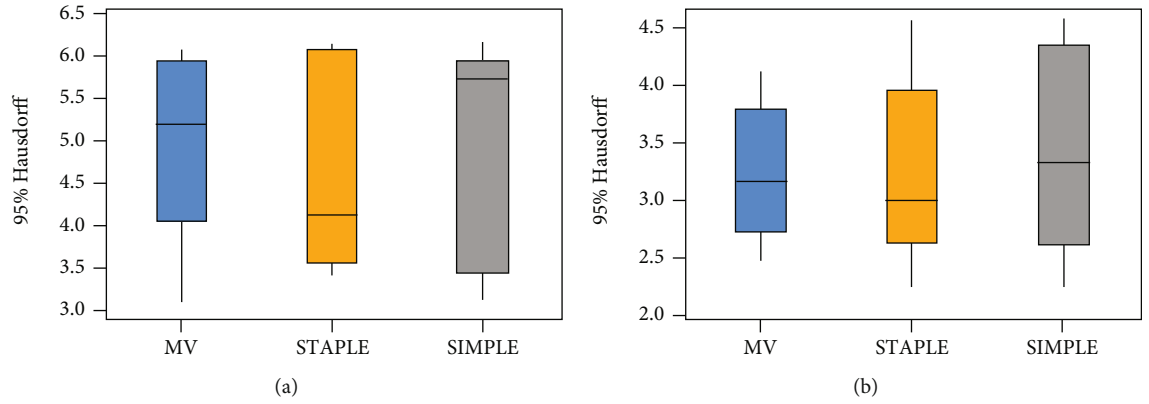


Figure 8: Box plot of the 95% Hausdorff distance of the foramen ovale on the left and right sides of each method: (a) foramen ovale box plot on the left; (b) foramen ovale box plot on the right.

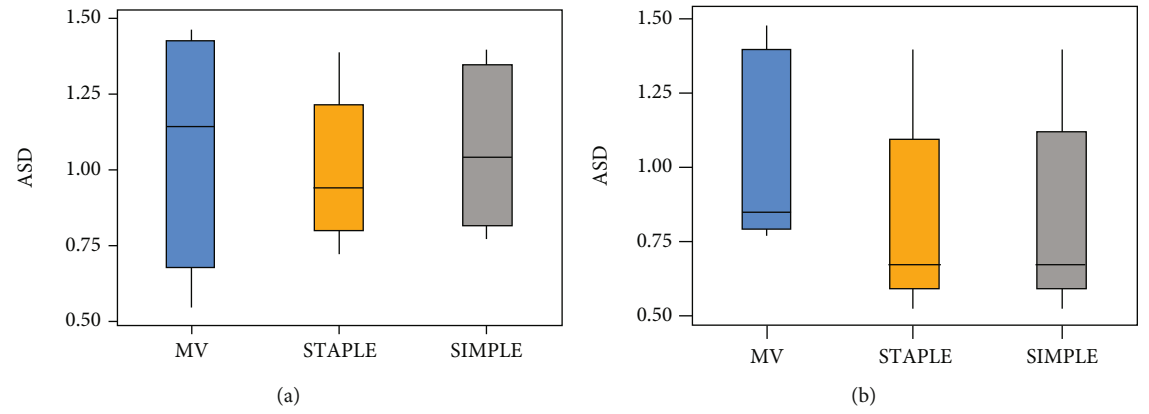


Figure 9: Box plot of the average surface distance of the foramen ovale on the left and right sides of each calculation method: (a) foramen ovale box plot on the left; (b) foramen ovale box plot on the right.

5. Conclusions

Trigeminal neuralgia seriously affects the normal life of patients. In clinical practice, surgical needles are often used to puncture the inside of the foramen ovale at the skull base for radiofrequency ablation. When puncturing the target point, it relies too much on the doctor's personal experience, and the operation is difficult. Computer-assisted puncture based on radiological information can improve the success rate of puncture operations. With the aid of a computer, the doctor performs a puncture based on the information of the foramen ovale region segmented out before the operation and then can accurately reach the target point. During the puncture process, the imaging equipment displays the

position of the puncture needle in real time for the doctor to make judgments. Computer-assisted trigeminal neuralgia puncture surgery reduces the pain suffered by patients and the probability of postoperative complications, lowers the threshold of puncture surgery, and improves the success rate of puncture surgery. The current problem is that the preoperative segmentation takes a long time; the average time is about 30 minutes, but it does not affect the real-time puncture work of the assistant intraoperative doctors. The future work is mainly in two aspects. On the one hand, we optimize our segmentation algorithm to reduce the time required during the registration and fusion process, and on the other hand, we improve the segmentation accuracy.

From what has been discussed above, in this paper, a segmentation method for the foramen ovale based on multiatlas is proposed, which provides an idea for foramen ovale segmentation, and can provide great convenience for computer-assisted puncture surgery. This article also creates the skull base foramen ovale atlas data set for the first time, which provides data support for future research on the skull base foramen ovale. The experimental results show that the segmentation of the foramen ovale by the multiatlas method has high accuracy and good effect, and it is expected to be applied in clinical puncture surgery.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the Key Research and Development Projects of China's Jilin Province Science and Technology Development Plan (No. 20200401078GX).

References

[1] T. M. Cheng, T. L. Cascino, and B. M. Onofrio, "Comprehensive study of diagnosis and treatment of trigeminal neuralgia secondary to tumors," *Neurology*, vol. 43, no. 11, pp. 2298–2302, 1993.

[2] S. Katusic, C. M. Beard, E. Bergstralth, and L. T. Kurland, "Incidence and clinical features of trigeminal neuralgia, Rochester, Minnesota, 1945–1984," *Annals of Neurology*, vol. 27, no. 1, pp. 89–95, 1990.

[3] N. N. Montano, G. Conforti, R. Di Bonaventura, M. Meglio, E. Fernandez, and F. Papacci, "Advances in diagnosis and treatment of trigeminal neuralgia," *Therapeutics & Clinical Risk Management*, vol. 11, pp. 289–299, 2015.

[4] J. M. Zakrzewska, "Insights: facts and stories behind trigeminal neuralgia," *Trigeminal Neuralgia Association*, vol. 202, no. 10, p. 638, 2006.

[5] J. S. Cheng, D. A. Lim, E. F. Chang, and N. M. Barbaro, "A review of percutaneous treatments for trigeminal neuralgia," *Neurosurgery*, vol. 10, no. 1, pp. 25–33, 2014.

[6] B. Huang, M. Yao, Z. Feng et al., "CT-guided percutaneous infrazygomatic radiofrequency neurolysis through foramen rotundum to treat V2 trigeminal neuralgia," *Pain Medicine*, vol. 15, no. 8, pp. 1418–1428, 2014.

[7] T. Dong, D. Wang, and D. Liu, "Clinical study on the treatment of trigeminal neuralgia with radiofrequency catheter ablation guided by neuronavigation system," *Nerve injury and functional reconstruction*, vol. 12, no. 6, 2017.

[8] M. I. Razzak, S. Naz, and A. Zaib, *Deep learning for medical image processing: overview, challenges and the future*, Classification in BioApps, 2018.

[9] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: a survey," *Medical Image Analysis*, vol. 24, no. 1, pp. 205–219, 2015.

[10] Y. Asim, B. Raza, A. K. Malik, S. Rathore, L. Hussain, and M. A. Iftikhar, "A multi-modal, multi-atlas-based approach for Alzheimer detection via machine learning," *International Journal of Imaging Systems and Technology*, vol. 28, no. 2, pp. 113–123, 2018.

[11] S. Bao, C. Bermudez, Y. Huo et al., "Registration-based image enhancement improves multi-atlas segmentation of the thalamic nuclei and hippocampal subfields," *Magnetic resonance imaging*, vol. 59, pp. 143–152, 2019.

[12] C. Tor-Díez, N. Passat, I. Bloch, S. Faisan, N. Bednarek, and F. Rousseau, "An iterative multi-atlas patch-based approach for cortex segmentation from neonatal MRI," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 73–82, 2018.

[13] J. H. Su, F. T. Thomas, W. S. Kasoff et al., "Thalamus optimized multi atlas segmentation (Thomas): fast, fully automated segmentation of thalamic nuclei from structural MRI," *Neuroimage*, vol. 194, pp. 272–282, 2019.

[14] M. A. Boucher, S. Lippé, and A. Damphousse, "Dilatation of lateral ventricles with brain volumes in infants with 3d transfontanelle US," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 557–565, Springer, Cham, 2018.

[15] B. van der Heyden, P. Wohlfahrt, D. B. P. Eekers et al., "Dual-energy CT for automatic organs-at-risk segmentation in brain-tumor patients using a multi-atlas and deep-learning approach," *Scientific Reports*, vol. 9, 2019.

[16] Z. Tang, S. Ahmad, P.-T. Yap, and D. Shen, "Multi-atlas segmentation of MR tumor brain images using low-rank based image recovery," *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2224–2235, 2018.

[17] S. K. Warfield, K. H. Zou, and W. M. Wells, *Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation*, IEEE Trans Med Imaging, 2004.

[18] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *Neuro Image*, vol. 33, no. 1, pp. 115–126, 2006.

[19] T. R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, M. A. Viergever, M. van Vulpen, and J. P. W. Pluim, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE)," *IEEE Transactions on Medical Imaging*, vol. 29, no. 12, pp. 2000–2008, 2010.

[20] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, pp. 726–738, 2009.

[21] S. P. Awate, P. Zhu, and R. T. Whitaker, "How many templates does it take for a good segmentation?: error analysis in multi-atlas segmentation as a function of database size," in *Second International Conference on Multimodal Brain Image Analysis*, 2012.

[22] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[23] M. S. Somesh, L. V. Prabhu, M. S. Gangadhara Swamy et al., "A morphometric study of foramen ovale," *Turkish Neurosurgery*, vol. 21, no. 3, pp. 378–383, 2011.

[24] B. Ray, N. Gupta, and S. Ghose, "Anatomic variations of foramen ovale," *Kathmandu University Medical Journal*, vol. 3, no. 1, pp. 9–64, 2005.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 5590180, 11 pages
https://doi.org/10.1155/2021/5590180

Research Article

Robust Blood Cell Image Segmentation Method Based on Neural Ordinary Differential Equations

Dongming Li,[1,2] **Peng Tang**,[1] **Run Zhang**,[3] **Changming Sun**,[4] **Yong Li**,[3] **Jingning Qian**,[3] **Yan Liang**,[1] **Jinhua Yang**,[2] **and Lijuan Zhang**[3]

[1]*School of Information Technology, Jilin Agricultural University, Changchun 130118, China*
[2]*College of Opto-Electronic Engineering, Changchun University of Science and Technology, Changchun 130022, China*
[3]*College of Computer Science and Engineering, Changchun University of Technology, Changchun, Jilin 130012, China*
[4]*CSIRO Data61, Epping, NSW 1710, Australia*

Correspondence should be addressed to Jinhua Yang; yangjh@cust.edu.cn and Lijuan Zhang; zhanglijuan@ccut.edu.cn

Received 3 February 2021; Revised 10 May 2021; Accepted 27 July 2021; Published 9 August 2021

Academic Editor: Shuihua Wang

For the analysis of medical images, one of the most basic methods is to diagnose diseases by examining blood smears through a microscope to check the morphology, number, and ratio of red blood cells and white blood cells. Therefore, accurate segmentation of blood cell images is essential for cell counting and identification. The aim of this paper is to perform blood smear image segmentation by combining neural ordinary differential equations (NODEs) with U-Net networks to improve the accuracy of image segmentation. In order to study the effect of ODE-solve on the speed and accuracy of the network, the ODE-block module was added to the nine convolutional layers in the U-Net network. Firstly, blood cell images are preprocessed to enhance the contrast between the regions to be segmented; secondly, the same dataset was used for the training set and testing set to test segmentation results. According to the experimental results, we select the location where the ordinary differential equation block (ODE-block) module is added, select the appropriate error tolerance, and balance the calculation time and the segmentation accuracy, in order to exert the best performance; finally, the error tolerance of the ODE-block is adjusted to increase the network depth, and the training NODEs-UNet network model is used for cell image segmentation. Using our proposed network model to segment blood cell images in the testing set, it can achieve 95.3% pixel accuracy and 90.61% mean intersection over union. By comparing the U-Net and ResNet networks, the pixel accuracy of our network model is increased by 0.88% and 0.46%, respectively, and the mean intersection over union is increased by 2.18% and 1.13%, respectively. Our proposed network model improves the accuracy of blood cell image segmentation and reduces the computational cost of the network.

1. Introduction

One of the most basic methods to diagnose diseases is by examining the blood smear through a microscope to check the shape, number, and proportion of red blood cells and white blood cells. However, manual examination of a blood microscope image is a time-consuming and laborious task. In recent years, with the development of computer vision and medical image processing technology, the recognition of medical microscopic cell images has also made considerable progress in the field of medical image processing. The research on medical image processing methods has become an important research direction in image processing and analysis.

Image segmentation is an important stage in the process of image analysis and processing. Traditional medical image segmentation methods mainly include activity contour, intensity thresholding, mathematical morphology, region growing, and watershed algorithm [1–5]. Since the fully convolutional neural network (FCN) [6] was first proposed by Long et al., it has achieved semantic segmentation of natural images from end to end, and it has also indicated the most progressive capacity in image segmentation. And they regard FCN as a foundation and have innovated tremendous

numbers of excellent semantic segmentation networks [7] from different perspectives under the stimulation from all kinds of semantic segmentation challenging competitions. Ronneberger et al. [8] focused on the large size and small quantity of medical images, providing a U-Net network model which adopted a coding-decoding structure. After 4 times of pooling during downsampling, dimensional splicing and fusion are performed with the corresponding scale in upsampling for adding feature information. To construct the pixel weight matrix, the closer it is to the cell's boundary, the larger the pixel weights will be, so it would be trained specifically. Kowal et al. [9] combined a convolutional neural network (CNN) and a seeded watershed algorithm [4] to segment the nucleus in breast cancer cell images, utilizing the accurate nucleus mask produced by CNN to replace the nucleus mask which was defined by normal thresholding. This process generates watershed topographic maps and nucleus seeds, and then, a watershed algorithm was used to separate the overlapping nucleus. Song et al. [10] proposed a multiscale convolutional network (MSCN) and a method based on image partition segmentation of the cervical cytoplasm and nucleus. They extracted features by MSCN and then divided the central region of each pixel. This method can segment all the nuclei in the cell images, but it could not distinguish normal cells and abnormal cells. Araújo et al. [11] used CNN to segment abnormal cells and blocky abnormal cells with high image overlap from digital images of conventional pap smears, filtering input images and eliminating cells that only include background or bad information. They adopted postprocessing to improve segmentation of abnormal cells and sorted the images according to probability of containing abnormal cells in the image. Öztürk et al. [12] proposed a new DCNN structure based on the residual network (ResNet) [13] and the deconvolutional network [14] structure. Semantic segmentation would be launched according to histopathological cell type, and all nuclei would be identified. They were classified as cancerous or normal according to each cell type. Shibuya and Hotta [15] proposed the feedback U-Net [8] network based on convolution long-short-term memory (LSTM). The output of U-Net reports back to the input, and then, it is fed into the second round. They extracted second-round features based on the first-round features by utilizing convolution LSTM [16]. Convolution LSTM that is used to process ordered data is a convolutional version of LSTM [17]. Chen et al. [18] proposed a new neural network that is referred to as neural ordinary differential equations (NODEs). This paper refers to the idea of Chen et al. [18]. We used the latest NODEs to improve the classic medical image segmentation method based on the U-Net network.

We put an ODE-block into a U-Net network model for blood cell image segmentation (named NODEs-UNet). The proposed NODEs-UNet network model can effectively reduce the use of parameters and improve the segmentation effect. NODEs can adapt to the receptive fields (RFs). There is no need to optimize the RFs for various segmentation tasks, and we only need to adjust the error tolerance of ODE-block. The generalization ability of the NODEs-UNet model architecture is strong.

2. Image Preprocessing

The experiment dataset in this paper was provided by the Center for Medical Image and Signal Processing (MISP) and the Department of Pathology, Isfahan University of Medical Sciences [19]. MISP.rar contains 148 clear blood cell smear images with a size of 775×519 pixels. Since the blood cell image is quite large, we picked up appropriate areas for convenient network training. We cropped 100 blood cell images with a size of 256×256 pixels by selecting a suitable area. To ensure the accuracy of the training model, we retained 20 images as the testing set and we used the remaining 80 images to increase the dataset to 800 by data augmentation. Besides, we used a ratio of 3 : 1 as the training set and the validation set. The image label was obtained by manual labeling by using the labeling tool LabelMe. There are three cell types that need labeling: background, white cells, and red cells. They are given the labels of 0, 1, and 2, respectively. Figure 1 shows the original blood cell image and its postvisualization of labels.

This paper employed blood smear images, which contain a small number of white blood cells and a large number of red blood cells. The original blood cell images are in color, and we use the color image for segmentation. We conducted preprocessing to the cell image and enhanced the contrast among cell images for segmenting the targets better. The blood cell images were converted from the RGB color space to the YUV space. The pseudocode is as follows:

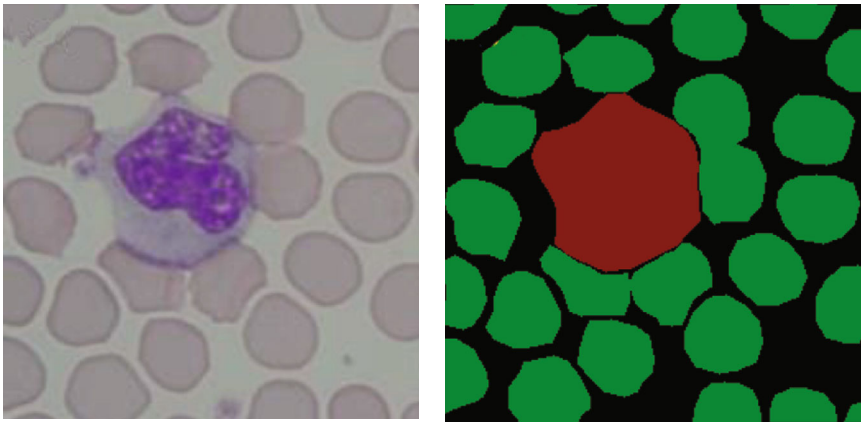
```
Img = Read(Path)
Y,U,V = BGR2YUV(Img)
Y′ = clahe_equalized(Y)
Img = YUV2BGR(Y′,U,V)
```

where "Y" means brightness. "U" stands for the difference between the blue channel and brightness. "V" means the difference between the red channel and brightness. Figure 2 shows the original cell image and the preprocessed image.

3. Methodology

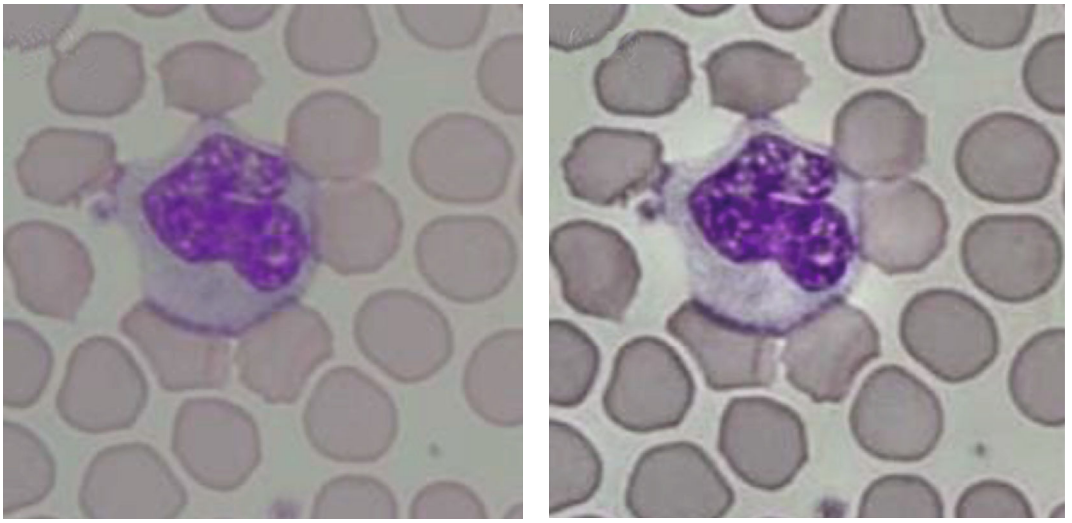
We present a novel segmentation method based on neural ordinary differential equations (NODEs) and U-Net for blood cell image segmentation. Firstly, the NODEs are introduced. Then, based on the classic U-Net network, we imported an ODE-block into the U-Net network architecture and determined the ODE-block location in the network. Finally, the proposed NODEs-UNet network architecture is built. The segmented image is constructed based on the NODEs-UNet network framework.

3.1. Neural Ordinary Differential Equation. Neural ordinary differential equation means a differential equation with a single independent variable. We are supposed to find the general solution of the unknown $f(x)$ for an ordinary differential equation normally. For instance, the general solution of equation $f'(x) = 2x$ is $f(x) = x^2 + C$, where C means an arbitrary constant. But the more common method to solve this problem in practice is by using an ODE-solver. That is, given an initial value $f(x_0)$, this does not have to find the general



(a) (b)

Figure 1: Display of an example image and its labels: (a) a blood cell image; (b) corresponding label shown in color.



(a) (b)

Figure 2: Blood cell image preprocessing: (a) original blood cell image; (b) preprocessed cell image.

solution of $f(x)$ when seeking the unknown value $f(x_1)$ except approaching its value gradually. In terms of the neural network, they are similar to an extreme complicated composite function whether they are a fully connected network, recurrent network, or convolutional network. The number of compositions is equal to the depth of the layers. For instance, a two-level fully connected network could be

$$\begin{aligned} h_{t+1} &= f(h_t, \theta_t), \\ h_{t+2} &= f(h_{t+1}, \theta_{t+1}), \\ h_{t+2} &= f(f(h_t, \theta_t), \theta_{t+1}), \end{aligned} \tag{1}$$

where h_t is the input value of the hidden unit of the t-th layer and f parameterizes the neural network by θ_t. Therefore, every neural network layer is similar to a universal function approximator.

A residual network (ResNet) [13] is a special type of convolutional network. It solved the gradient reversion problem with residual connection, which means that the gradient can still be effectively transmitted back to the input end when the neural network layer is very deep. Figure 3 is the structure of a ResNet-block. The output of the ResNet-block combines the input information and the output information of the internal convolution operation. This residual connection ensures that the accuracy of the deep model is at least not lower than the accuracy of the shallow network.

We can illustrate the ResNet-block above formally as an equation below: $h_{t+2} = h_t + f(f(h_t, \theta_t), \theta_{t+1})$, which stands for the whole ResNet-block above. If we rewrite it in the form of a residual network, that is,

$$h_{t+2} = h_t + f(f(h_t, \theta_t), \theta_{t+1}). \tag{2}$$

We can find that the traditional neural network f is directly parameterized as hidden layers and the residual neural network f parameterizes the residual among hidden layers. But the neural ordinary differential equation in this paper takes another way to use for parameterizing the

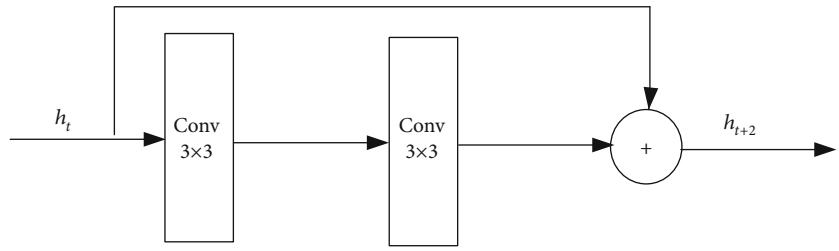


Figure 3: Structure of a ResNet-block.

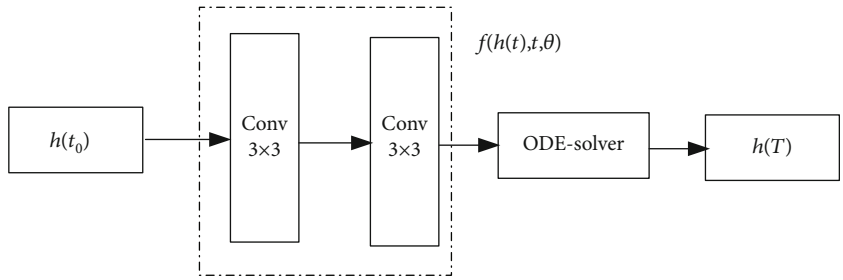


Figure 4: Structure of ODE-block.

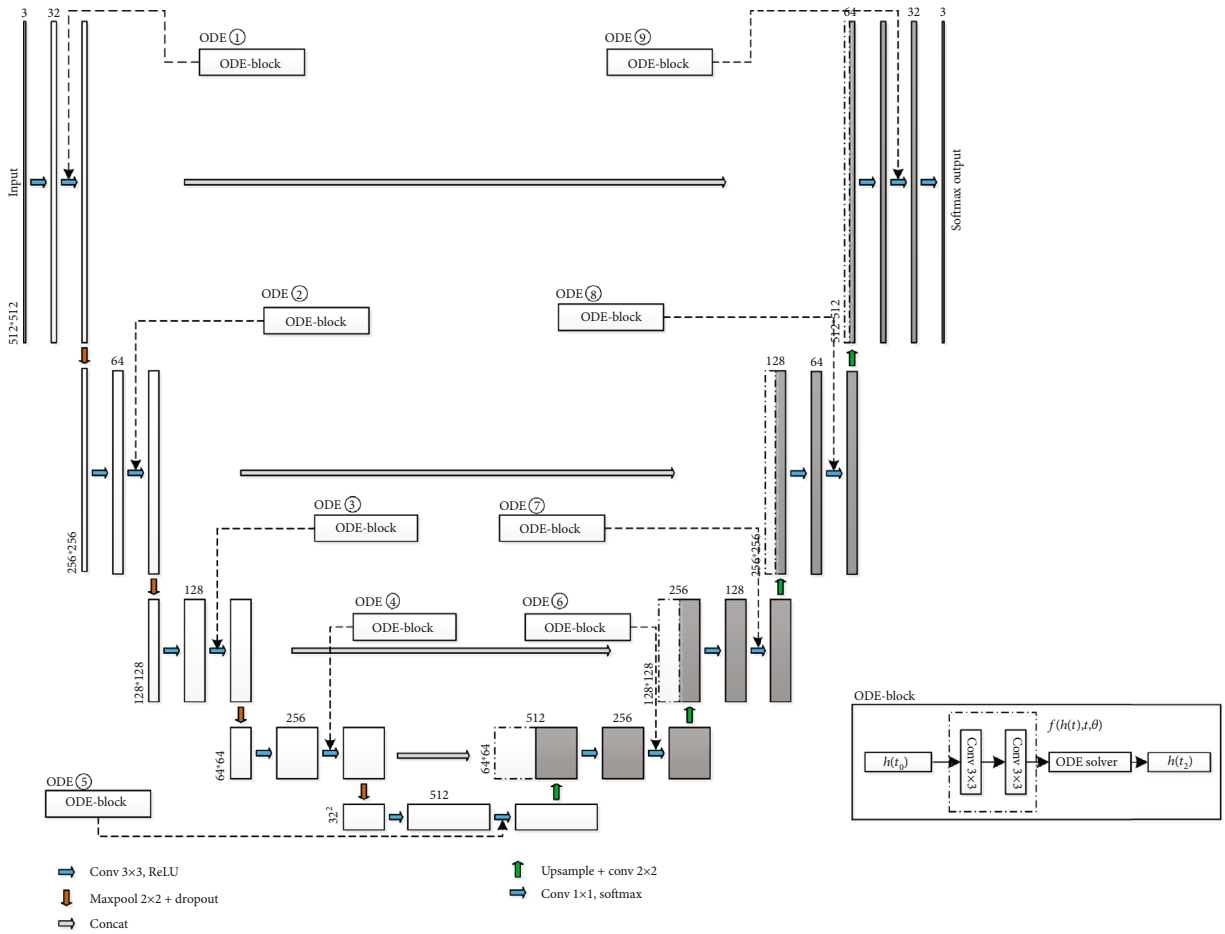


Figure 5: Nine different ODE-block locations.

derivative in hidden states by the neural network. By assuming the discrete layers as continuous layers and parameters, this continuous transformation form can be expressed as a neural ordinary differential equation (NODE):

$$\frac{dh(t)}{dt} = f(h(t), t, \theta), \tag{3}$$

where f is defined as a neural network as before, but now this and its parameter θ are a unit, and t is also fed into the neural network as an independent parameter. From the perspective of derivative definition, when the change of t tends to become infinitely small, the change of the hidden state $dh(t)$ can be modeled by the neural network. When t changes slowly from the initial to the end, the change of $h(t)$ ultimately represents the result of forward propagation. In this way, using the neural network to parameterize the derivative of the hidden layer, the neural network layer is indeed continuous.

If the numerical solution of the ordinary differential equation can be obtained, then it is equivalent to forward propagation. Now, we convert equation (3) to

$$\int_{t_0}^{t_2} dh(t) = \int_{t_0}^{t_2} f(h(t), t, \theta)dt, \tag{4}$$

$$h(t_2) = h(t_0) + \int_{t_0}^{t_2} f(h(t), t, \theta)dt. \tag{5}$$

From equation (5), we can see that the numerical solution of ODE $h(t_2)$ requires the integral of the neural network f from t_0 to t_2. It is a problem about the initial value of ODE. We can obtain the result with an ODE-solver directly. Such an ODE-solver can also control the numerical error so that we can make a contrast between computing ability and model accuracy. Figure 4 is the structure of the ODE-block.

3.2. The Location of ODE-Block. The network architecture of this paper is based on the classic U-Net fully convolutional neural network model in medical image segmentation. Considering the reduction of the computational cost, we decrease the number of convolutional cores in the convolution layers to a half in the original U-Net network. In order to study the influence of a single ODE-block on the network at different positions, we imported an ODE-block in the U-Net network architecture which is shown in Figure 5. The training set, validation set, and testing set of the whole networks are consistent, and the error tolerance of the ODE-block solver is $1e^{-3}$. When we train the network, we input the training set and the validation set, and the training times (epochs) are 50 times. We use a callback function to save the network model with the minimum val_loss of the validation set.

Nine ODE-blocks obtained in the above experiment were tested on the testing set. The cell image segmentation results were evaluated by pixel accuracy (PA), class pixel accuracy (CPA), mean intersection over union (MIoU), and computation time, and the comparison results are shown in Table 1.

From Table 1, we can see that compared with the U-Net network and the nine ODE-block-based networks, it can be seen that after the ODE-block is added, the PA and MIoU have been significantly improved. The computation time is obviously surging, which is the time it takes for the network to segment twenty blood cell images. Compared with the nine ODE-block-based networks, the location where the ODE-block is added has no obvious impact on PA and MIoU, but it has a greater impact on the computation time, so we could conclude that the location where the ODE-block is added goes down as the U-Net "U"-shaped structure goes down. And the time is much shorter when the "U"-shaped structure keeps going down. The more the "U"-shaped structure goes up, the longer the time is.

Table 1: Comparison results of different ODE-block positions.

Algorithm	PA (%)	CPA (%)			MIoU (%)	Time (s)
		Background	Red cells	White cells		
U-Net	94.59	92.55	96.54	94.10	88.58	0.15
ODE-UNet1	95.05	94.25	95.97	93.92	89.80	7.35
ODE-UNet2	95.08	94.23	95.69	95.98	89.68	3.73
ODE-UNet3	95.11	95.14	95.70	91.65	89.61	1.94
ODE-UNet4	95.19	94.08	96.51	93.42	90.17	1.03
ODE-UNet5	95.05	95.62	94.62	94.53	89.62	0.61
ODE-UNet6	95.14	93.78	96.34	95.39	90.09	1.06
ODE-UNet7	95.17	93.80	96.59	94.19	90.02	1.95
ODE-UNet8	95.16	94.33	96.30	92.98	89.91	3.72
ODE-UNet9	95.14	94.33	96.23	93.15	90.03	7.35

3.3. NODEs-UNet Neural Network. In this paper, we present a novel blood cell image segmentation method based on NODEs and U-Net (named NODEs-UNet) neural network framework. It is based on the U-Net network model, and downsampling is performed through the maximum pooling layer. For the coding part, each time it passes through a pooling layer, a new scale is constructed, and there are five scales including the original image. Finally, the convolution results in five scales are fused. The convolutional layer extracts features, and the "same" convolution is used to keep the image size unchanged before and after convolution. Upsampling is performed through bilinear interpolation, and the scale corresponding to the feature extraction part is fused with each upsampling.

From Section 3.2, it can be concluded that the ODE-block location that is added goes down with the U-Net "U"-shaped structure, and the time is shorter. Therefore, we add one ODE-block with error tolerance $1e^{-6}$ and two ODE-blocks with error tolerance $1e^{-4}$ at the bottom of the "U" shape, as shown in Figure 6.

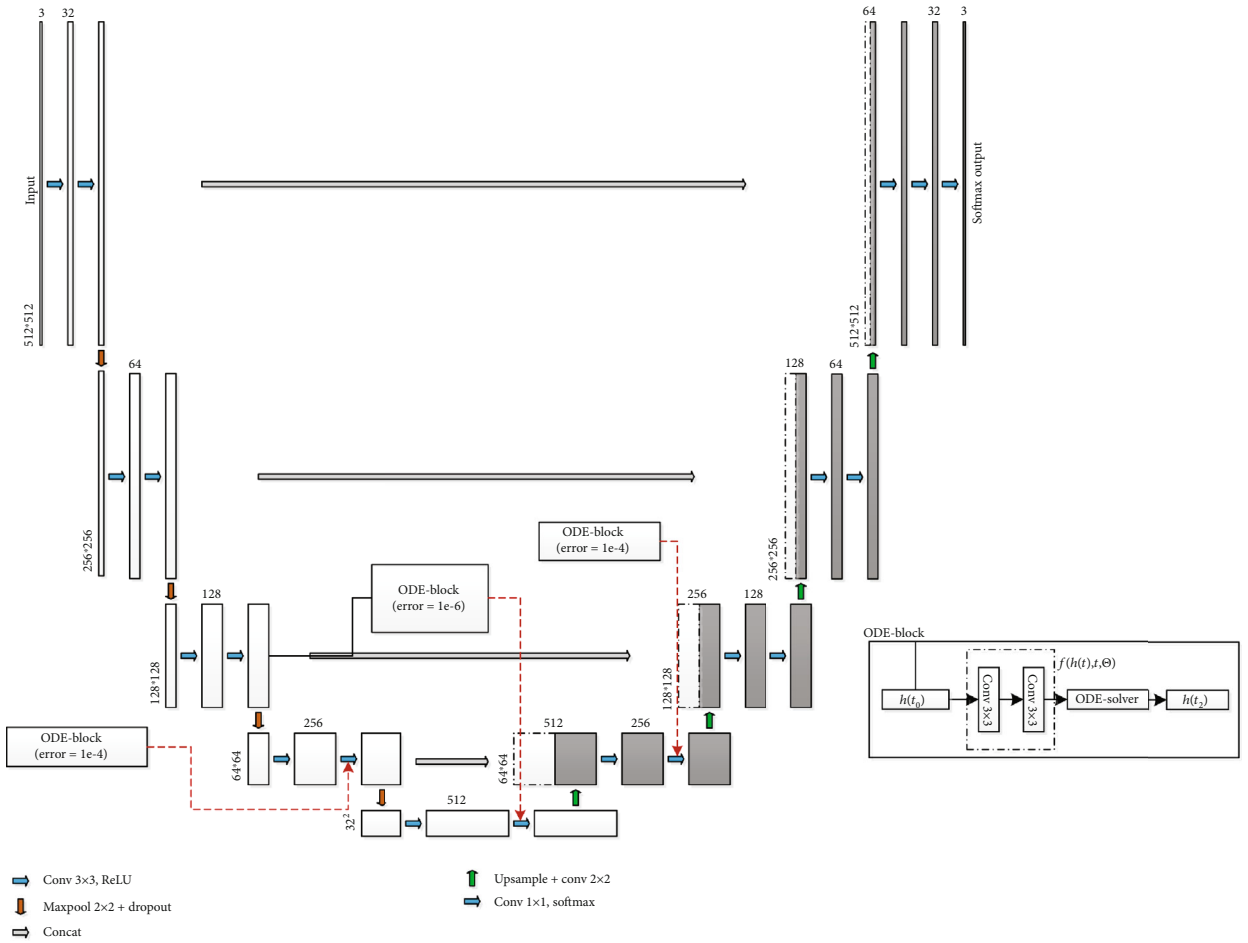


Figure 6: NODEs-UNet neural network framework.

The final prediction result of this network uses the activation function softmax, that is,

$$a = \text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}}, \tag{6}$$

where z_i is the output value of the i-th node and C is the number of output nodes which is the number of classification categories. The output of the multiclass is converted into a probability distribution in the range of $[0, 1]$ through the softmax function, which means the probability that node i belongs to the background, white cells, or red cells. We use categorical cross entropy as the loss function of the network, which is often suitable for multiclass problems and can avoid the problem of reduced learning rate of the mean square error loss function. The equation is as follows:

$$\text{loss} = -\frac{1}{n}(y \ln a + (1 - y) \ln (1 - a)), \tag{7}$$

where y is the probability distribution of the expected output and a is the probability distribution of the actual output of the network. When the value of the cross entropy is smaller, the two probability distributions are closer.

4. Experimental Results and Discussion

The proposed NODEs-UNet network framework was applied to the problem of multiclass blood cell image segmentation, and it is used to evaluate the role of exploiting the ODE-block in segmentation. The real image dataset was chosen from the publicly available dataset on MISP and the Department of Pathology, Isfahan University of Medical Sciences, that contains blood smear microscopic images with red cells and white cells, namely, the MISP01 dataset [19]. The results of this experiment were compared with those of the U-Net network [8] and the ResNet network [13]. The reason we choose these two networks is that the U-Net network is good at semantic segmentation and it is the basis of our proposed NODEs-UNet network. The ResNet network is also based on the reduced version of the U-Net network in this paper, and the residual module is added to the U-Net network. The added ResNet-block location was referred to the D-LinkNet [20] architecture; then, we built the ResNet network model. In the following sections, we give the experimental settings. Then, we compare our method with those two methods and give the statistical results.

4.1. Experimental Settings. In this study, all the experiments are implemented in a Ubuntu 16.04 LTS 64-bit operating

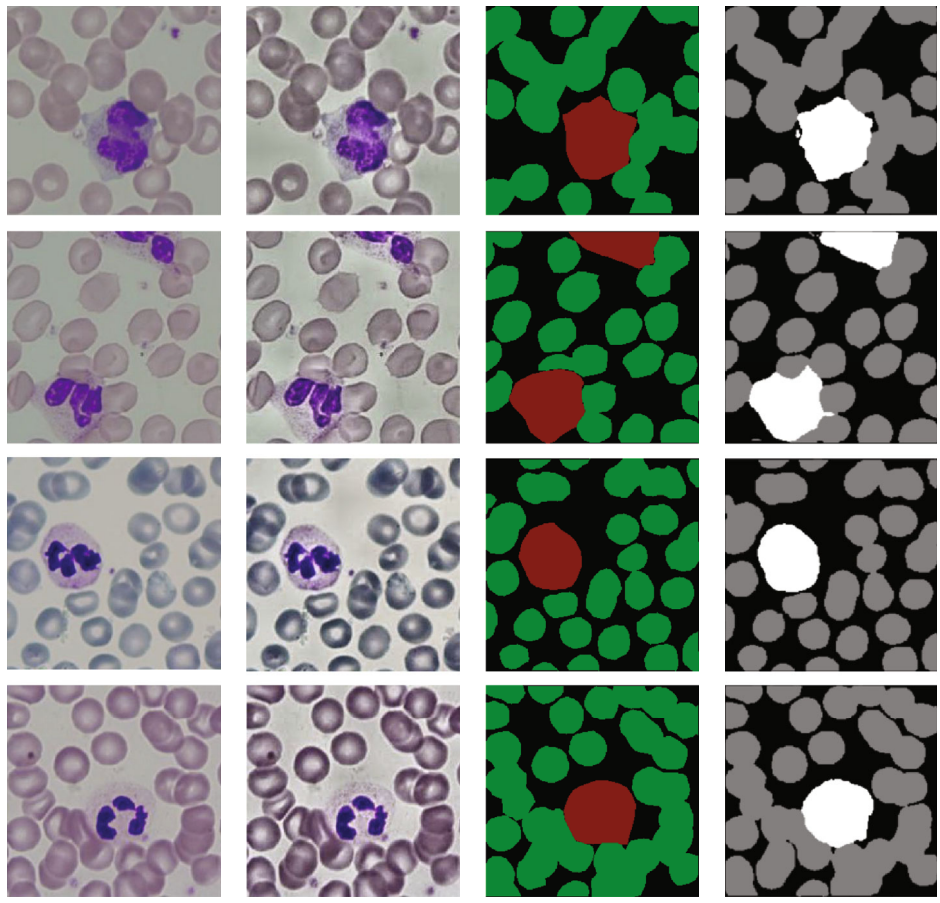


Figure 7: Blood cell image segmentation results: (a) original blood cell images; (b) preprocessed blood cell images; (c) corresponding labels; (d) segmentation results by our algorithm based on the NODEs-UNet network.

system with Intel Xeon E5 64 core CPU and NVIDIA GeForce GTX 1080 Ti 11 G ∗ 4 GPU, based on the Keras deep learning framework equipped with the NODEs-UNet network model. The ODE-block uses TensorFlow. We complete the training and testing for blood cell image segmentation in CUDA 8.0 GPU calculating the platform and cuDNN 7.5 deep learning GPU acceleration library.

In order to increase the training speed, we call the function multi_gpu_model() to copy the model on four GPUs. Each GPU calls its own model, running on its own dataset, and then connects all the running results together. In order to avoid memory overflow, the model is built on the CPU. We input the training set and validation set to train the network, saving the model with the smallest loss (val_loss) in the validation set in a single model and saving the network framework in a HDF5 file.

4.2. Validation on Blood Cell Image Segmentation. For the MISP01 dataset, four randomly selected blood cell images were used for blood cell image segmentation based on the NODEs-UNet framework, and Figure 7 shows the results. Figure 7(a) is the original blood cell images. Figure 7(b) is the preprocessed blood cell images using an adaptive histogram equalization method (see Section 2). Figure 7(c) is the corresponding labels of blood cell images. Figure 7(d) is the segmentation result using our proposed algorithm. From Figure 7, we can see that our method can accurately segment background, red cells, and white cells. It has clear boundary and complete details, and the segmentation results are very close to the ground truth.

In order to further verify our proposed segmentation method based on the NODEs-UNet network in this paper, we compared and analyzed the quality of the segmentation results from our method with the related works developed on the basis of the U-Net network [8] and the ResNet network [13], and the comparison on the segmentation results are shown in Figure 8. Figure 8 shows four randomly selected blood cell image segmentation results using three networks. As shown in Figure 8(a), the original blood cell images are randomly selected from the MISP01 dataset [19] with blurring and noise. Figure 8(b) is the enhanced cell images using an adaptive histogram equalization method (Section 2). Figure 8(c) is their corresponding ground truth. Figure 8(d) shows the segmentation results after applying U-Net to the images. Figure 8(e) is for the segmentation result after applying ResNet to the images. As shown in Figure 8(f) for the result of our proposed segmentation method based on the NODEs-UNet network, we can see that our work can provide more accurate segmentation and more details.

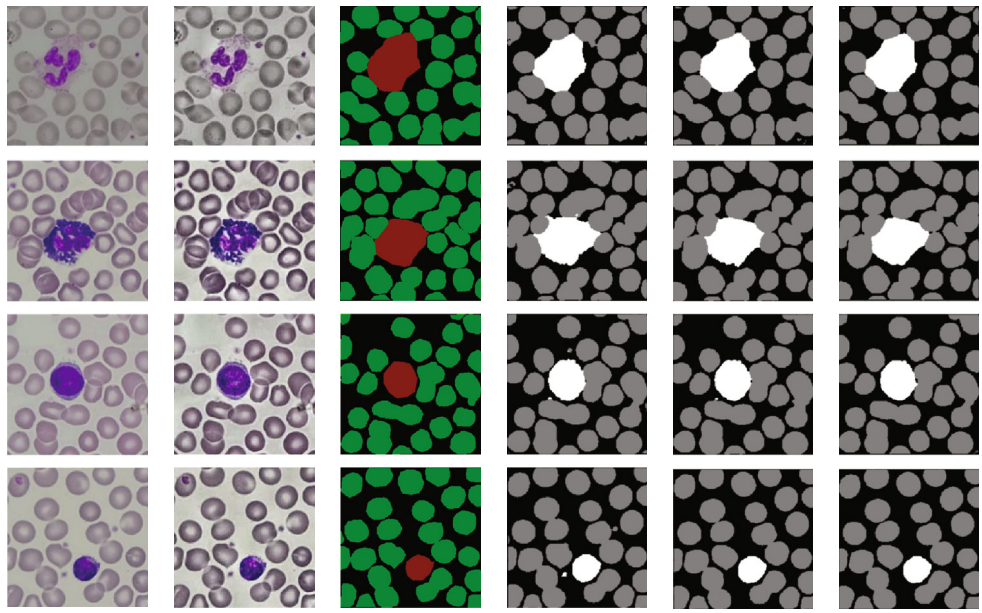


Figure 8: The original images and the comparison results based on three networks on the MISP01 dataset: (a) original image; (b) preprocessed blood cell image; (c) ground truth; (d) segmented method based on the U-Net network; (e) segmented images based on the ResNet network; (f) results from our proposed method based on the NODEs-UNet network.

To quantitatively measure and compare the accuracy of our proposed method with other methods, we applied each of the methods on the segmented dataset and compared it with the ground truth. Then, we counted the truly and falsely detected segmentation results. Effectiveness measures based on PA, CPA, and MIoU are calculated. Table 2 shows the evaluation results of each method on the blood cell image segmentations for the testing set, where parameter M refers to the memory space occupied by the parameter weight of the network model. Figures 9 and 10 show the specific PA and MIoU indicators of the segmentation results of each network in 20 blood cell images.

Observing Table 2, by comparing the objective evaluation data (PA, CPA, and MIoU) of the U-Net network with those of the ResNet and NODEs-UNet networks, we concluded that on the basis of the U-Net network architecture, whether by adding the ResNet-block or the ODE-block, both segmentation results are significantly improved. For the ResNet network, the PA and MIoU have increased by 0.42% and 1.05%, respectively, and the PA and MIoU of the NODEs-UNet network increased by 0.88% and 2.18%, respectively. This is mainly because the output of the ResNet-block and the ODE-block combines the input information with the output information of the internal module operation, and this connection method ensures that in the network model after being added, the accuracy of the module is at least not lower than the accuracy of the initial network model. And due to the limitation of the computational power of the experimental equipment, the number of convolution cores in the convolutional layer in the U-Net network in this paper is twice as small as that in the traditional U-Net network, and the complexity is lower. After adding the ODE-block, the depth of the network is increased, so the accuracy of the network is significantly improved.

Table 2: Comparison results for three networks.

Algorithm	PA (%)	CPA (%)			MIoU
		Background	Red cells	White cells	
U-Net	94.42	92.55	96.54	94.10	88.43
ResNet	94.84	94.55	95.60	92.89	89.48
NODEs-UNet	95.30	94.58	95.94	95.37	90.61

Then, by comparing the indicator data (PA and MIoU) of the NODEs-UNet network and the ResNet network, it can be seen that the ODE-block has more advantages in performance than the ResNet-block, and the PA and MIoU in the NODEs-UNet network has increased by 0.46% and 1.13%, respectively, as compared to those in the ResNet network. This is because the residual network is a special case of ordinary differential equations, which is the discretization of Euler's method. Euler's method is very intuitive for solving ordinary differential equations, that is, $h(t + \Delta t) = h(t) + \Delta t \times f(h(t), t)$. Whenever the hidden layer takes a small step Δt along t, the new hidden layer state $h(t + \Delta t)$ should take a small step in the existing direction. If we walk from t_0 to t_1 in such a small step, then the numerical solution of ODE is obtained. If Δt is equal to 1 every time, then the Euler method of discretization is equal to the expression of the residual module $h(t + 1) = h(t) + f(h(t), t)$. But Euler's method is a basic method adopted to solve ordinary differential equations. Each step will make a little error, and the error will be accumulated.

The ODE-solver in the NODEs-UNet network does not move a fixed step length like Euler's method. It will select

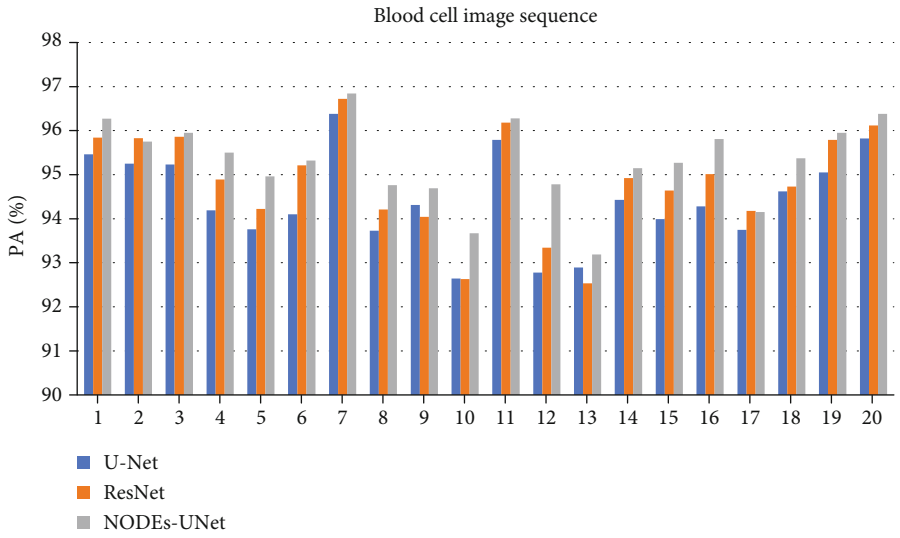


Figure 9: Experimental results for the PA index.

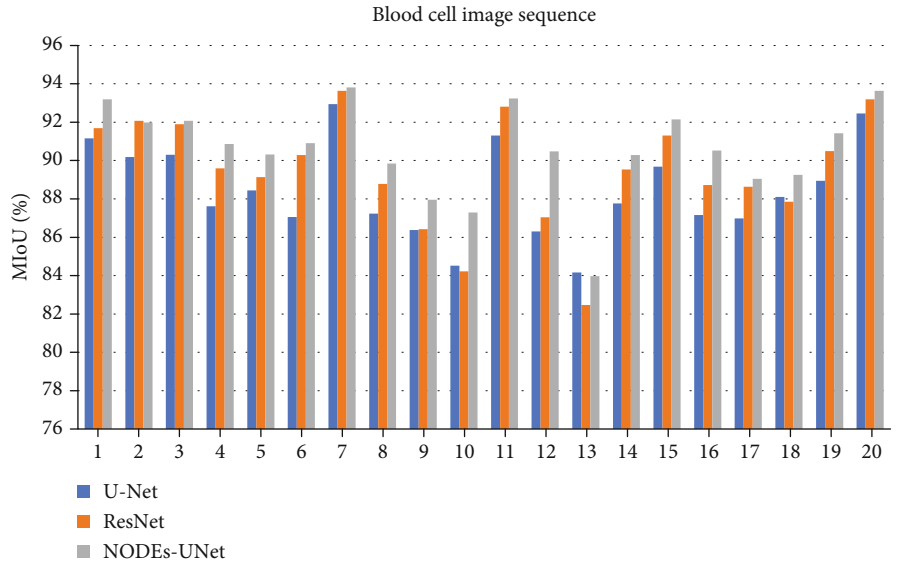


Figure 10: Experimental results for the MIoU index.

an appropriate step length to approximate the real solution according to the given error tolerance. Reducing the error tolerance will increase the number of evaluations of the function, similar to increasing the depth of the model. Therefore, we can change the behavior of the neural network by changing the error tolerance. During training, the error can be reduced, the accuracy rate can be improved, and a better neural network can be learned. During testing, the error can be increased according to the actual computing environment, the number of function evaluations can be reduced, and we can obtain the segmentation result faster. By comparing the memory space occupied by the NODEs-UNet network with the ResNet network parameters, the memory space occupied by the NODEs-UNet network is only 46% of the ResNet network. This is due to the derivative of the parameterized hidden state of the ODE-block, which similarly constructs continuity layers and parameters. There are no intermediate results stored in the forward propagation process, so it only needs approximately constant memory cost.

5. Conclusion

This paper combines the neural ordinary differential equation with the U-Net network to segment blood smear images. Compared with the more common semantic segmentation using fully convolutional networks, this paper does not improve on feature extraction and multiscale fusion, but it is directly based on the U-Net network model. The ODE-block is added to improve the network and improve the network accuracy for cell image segmentation. Utilizing the characteristics of the ODE-block, we use the ODE-solver in the ODE-block to parameterize the derivative of the hidden

state, instead of directly parameterizing the hidden state as usual. This connection method can achieve the same effect as the residual network and can effectively avoid the problem of network degradation in the deep network. Of course, the network layer of this paper is not very deep, and the advantages of this paper have not been fully utilized. The ODE-block can select an appropriate step length to approximate the real solution according to the given error tolerance. Based on these characteristics, reducing the error tolerance will increase the number of evaluations of the function, which is similar to increasing the depth of the model without increasing the parameters of the model. We reduce the error tolerance of the ODE-block in the condition of limited computing resources, and a deep network model can also be built.

The next research plan is to perform convolution with a 1×1 convolution kernel for each scale sampled on the NODEs-UNet network. We will perform multiscale fusion of all outputs, connect them to the fully connected layer, and do linear regression. So we can directly output the number of white blood cells and red blood cells in the blood image.

Data Availability

The source code supporting the study will be available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the National Science Foundation of China (Grant No. 61806024), the Scientific and Technological Research Project of the Education Department of Jilin Province (No. JJKH20210637KJ), and the Scientific Research Project of the Department of Ecological Environment of Jilin Province (No. 202107), China. We thank the authors for the MISP dataset.

References

[1] A. Gertych, Z. Ma, J. Tajbakhsh, A. Velásquez-Vacca, and B. S. Knudsen, "Rapid 3-D delineation of cell nuclei for high-content screening platforms," *Computers in Biology and Medicine*, vol. 69, pp. 328–338, 2016.

[2] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 97–114, 2014.

[3] M. Kowal, M. Skobel, and N. Nowicki, "The feature selection problem in computer–assisted cytology," *International Journal of Applied Mathematics and Computer Science*, vol. 28, no. 4, pp. 759–770, 2018.

[4] C. Koyuncu, E. Akhan, T. Ersahin, R. Cetin-Atalay, and C. Gunduz-Demir, "Iterative h-minima-based marker-controlled watershed for cell nucleus segmentation," *Cytometry Part A*, vol. 89, no. 4, pp. 338–349, 2016.

[5] A. Piorkowski, K. Nurzynska, J. Gronkowska-Serafin, B. Selig, C. Boldak, and D. Reska, "Influence of applied corneal endothelium image segmentation techniques on the clinical parameters," *Computerized Medical Imaging and Graphics*, vol. 55, pp. 13–27, 2017.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, Boston, Massachusetts, 2015.

[7] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[8] 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)*, vol. 19351 of Lecture Notes in Computer Science, , pp. 234–241, Springer, 2015.

[9] M. Kowal, M. Żejmo, M. Skobel, J. Korbicz, and R. Monczak, "Cell nuclei segmentation in cytological images using convolutional neural network and seeded watershed algorithm," *Journal of Digital Imaging*, vol. 33, no. 1, pp. 231–242, 2020.

[10] Y. Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2421–2433, 2015.

[11] F. H. D. Araújo, R. R. Silva, D. M. Ushizima et al., "Deep learning for cell image segmentation and ranking," *Computerized Medical Imaging and Graphics*, vol. 72, pp. 13–21, 2019.

[12] S. Öztürk and B. Akdemir, "Cell-type based semantic segmentation of histopathological images using deep convolutional neural networks," *International Journal of Imaging Systems and Technology*, vol. 29, no. 3, pp. 234–246, 2019.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[14] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1520–1528, Santiago, Chile, 2015.

[15] E. Shibuya and K. Hotta, "Feedback U-net for cell image segmentation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4195–4203, Seattle, WA, USA, 2020.

[16] S. H. Xingjian, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," pp. 1–12, 2015, http://arxiv.org/abs/1506.04214.

[17] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," *In Proceedings Presses universitaires de Louvain*, vol. 89, pp. 89–94, 2015.

[18] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," pp. 6571–6583, 2018, http://arxiv.org/abs/1806.07366.

[19] O. Sarrafzadeh, H. Rabbani, A. Talebi, and H. U. Banaem, "Selection of the best features for leukocytes classification in blood smear microscopic images," in *SPIE Medical Imaging*, p. 90410P, San Diego, California, United States, 2014.

[20] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 182–186, Salt Lake City, UT, USA, 2018.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 2485934, 12 pages
https://doi.org/10.1155/2021/2485934

Research Article

Multiclassification of Endoscopic Colonoscopy Images Based on Deep Transfer Learning

Yan Wang,[1,2] **Zixuan Feng**,[3] **Liping Song**,[4,5] **Xiangbin Liu**,[4,5] **and Shuai Liu**[4,5]

[1]*Department of General Surgery, China-Japan Union Hospital of Jilin University, Changchun 130033, China*
[2]*Department of Pharmaceutics, College of Pharmacy, Shenyang Pharmaceutical University, Shenyang 110116, China*
[3]*College of Computer Science and Technology, Jilin University, Changchun 130012, China*
[4]*College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China*
[5]*Hunan Xiangjiang Artificial Intelligence Academy, Changsha 410081, China*

Correspondence should be addressed to Shuai Liu; cs.liu.shuai@gmail.com

Yan Wang, Zixuan Feng, and Liping Song contributed equally to this work.

Received 6 May 2021; Revised 27 May 2021; Accepted 9 June 2021; Published 3 July 2021

Academic Editor: Shuihua Wang

With the continuous improvement of human living standards, dietary habits are constantly changing, which brings various bowel problems. Among them, the morbidity and mortality rates of colorectal cancer have maintained a significant upward trend. In recent years, the application of deep learning in the medical field has become increasingly spread aboard and deep. In a colonoscopy, Artificial Intelligence based on deep learning is mainly used to assist in the detection of colorectal polyps and the classification of colorectal lesions. But when it comes to classification, it can lead to confusion between polyps and other diseases. In order to accurately diagnose various diseases in the intestines and improve the classification accuracy of polyps, this work proposes a multiclassification method for medical colonoscopy images based on deep learning, which mainly classifies the four conditions of polyps, inflammation, tumor, and normal. In view of the relatively small number of data sets, the network firstly trained by transfer learning on ImageNet was used as the pretraining model, and the prior knowledge learned from the source domain learning task was applied to the classification task about intestinal illnesses. Then, we fine-tune the model to make it more suitable for the task of intestinal classification by our data sets. Finally, the model is applied to the multiclassification of medical colonoscopy images. Experimental results show that the method in this work can significantly improve the recognition rate of polyps while ensuring the classification accuracy of other categories, so as to assist the doctor in the diagnosis of surgical resection.

1. Introduction

Image classification is a task that classifies images into a certain category according to different features in the image. It is the core of computer vision to distinguish different categories of images. Image classification is the basis of other high-level visual tasks in computer vision, such as target detection, image segmentation, and face recognition. Image classification is widely used in many fields, such as payment method for face recognition, toll system for license plate recognition, autonomous driving for traffic safety, and computer-aided system for diagnosis [1]. In the medical world, the number of medical images generated every day is uncountable, and a tomographic slice of the lungs of a patient can generate dozens of CT images [2]. Therefore, classifying massive medical images is an important step of computer-aided diagnosis [3, 4].

Medical image classification refers to taking one or more examination images as input, predicting them through the trained model, and outputting a diagnostic result indicating whether a certain disease is suffering or whether the severity is graded. At present, it has been widely used in epidemic prevention and diagnosis of benign tumors and cancer and to distinguish between different categories of the same disease

and other important clinical events [5]. The object of medical image classification is the image obtained by patients through various kinds of examination equipment, mainly including Computed Tomography (CT), X-ray, Magnetic Resonance Imaging (MRI), and ultrasound image (UI) [6]. In addition, there are other pathological images as well as endoscopic images. When a doctor examines disease in the intestine, the patient generally needs an endoscope to obtain the surface features of the intestine tract to make a diagnosis. The main manifestations of various lesions in the intestine are polyps, inflammation, and cancer. Polyps are spherical or oval pedicled masses protruding from the mucosal surface of the large intestine. It is a common benign tumor. Inflammation is manifested as colonic mucosa with extensive hyperemia, edema, and erosion under a colonoscopy, and it is easy to bleed when touched, with pus and blood and exudate on the surface. Cancer is a common malignant tumor in the gastrointestinal tract. Cancer protrudes into the intestinal lumen and has hemorrhage and necrosis on the surface. The classification of medical colonoscopy images is mainly divided into four categories: polyps, inflammations, cancer, and normal [7].

The research of image classification technology has always attracted attention. In recent years, lots of advanced intelligent classification methods have emerged, and classification accuracy has been continuously improved. However, there are still many problems to be further studied, such as the difficulty of classification and recognition caused by image quality and the inapplicability of feature extraction methods to different images. Traditional medical image classification methods are mainly based on the neural network, Bayesian network, decision tree, and other single modes. With the development of Artificial Intelligence (AI) and the demand for application in the medical field, deep learning has become the mainstream image classification framework. Deep learning simulates the human brain for analytical learning and uses human brain mechanisms to interpret data. Convolutional neural networks (CNN) are the most commonly used network model for deep learning. It sends images into the network for training and classifies image data according to image features. Deep learning requires a large number of data sets to achieve better classification results when training network structures, so as to prevent model overfitting. Due to the complexity and inconvenience of endoscopy, the requirements for doctors and patients in the process of examination are extremely high, which leads to difficulty in the collection of colonoscopy data sets. Aiming at the problem of model overfitting caused by the small amount of colonoscopy images trained by deep learning, this work proposes the use of transfer learning to solve the problem of deep network training requiring a large number of data sets. The contributions of this paper are as follows:

(1) Four-classification tasks of the intestinal image are proposed. In order to accurately distinguish all kinds of pathologically similar lesions in the intestinal tract, while improving the classification accuracy of polyps, the colonoscopy images are divided into four categories: normal, polyp, inflammation, and cancer, which increased the diversity of disease types in the classification task

(2) Transfer learning is proposed to optimize classical deep learning, and a multiclassification model of endoscopic colonoscopy images based on deep transfer learning was obtained. The pretrained network model on the natural image is used to fine-tune its network with the intestinal image to solve the problem that deep learning training requires a large number of data sets

(3) Finally, the feasibility of the proposed method was verified by experiments. Inputting the intestinal image into the model proposed in this paper, four classification results were obtained, which significantly improved the polyp recognition rate while ensuring the classification accuracy of other categories

The following is the structure of this paper. In Section 2, deep transfer learning and intestinal image classification tasks are discussed. Section 3 introduces the intestinal data set and its classification methods used in this work. In Section 4, the classification methods mentioned are experimented and compared. Section 5 summarizes the work of this paper and looks forward to future work.

2. Related Work

2.1. Medical Image Classification. Medical image classification is the first major contribution made by deep learning in the field of medical image processing. Medical image processing initially focused on unsupervised pretraining and network structures, such as Stacked Autoencoder (SAES) and Deep Belief Networks (RBMS). As early as 2013, Suk et al. [8] proposed a feature representation and stacked autoencoder based on deep learning, which combined potential information with original low-level features and improved the classification accuracy of Alzheimer's disease (AD)/Mild Cognitive Impairment (MCI). With the rise of large-scale visual recognition challenges such as ImageNet, several excellent deep neural networks have emerged, which promote the development of deep learning. At present, CNN has always been the current standard technology in medical image classification. The pretraining of CNN on natural images has shown amazing results, challenging the authority of human experts in certain classification tasks. Many researchers use classic deep learning networks to improve classification accuracy. For example, Gao et al. [9], in order to grade and evaluate the severity of nuclear cataract, first clustered the images to obtain local filters, then sent the learned filters to the convolutional neural network and recursive neural network for further feature extraction, and finally used Support Vector Machine (SVM) for classification. The results verify that the model is superior to the latest progress in clinical cataract classification. Jiao et al. [10] proposed a breast mass classification framework based on deep features to solve the problem of poor performance of the underlying features. The framework mainly includes a convolutional neural network and a decision mechanism, which combines enhanced information and deep features to simulate the process of doctor diagnosis to improve classification accuracy. Lin et al. [11]

proposed a recurrent neural network with an attention model for sequence labeling; in this test, a hierarchical structure to incorporate character-level and word-level information is proposed by applying an attention mechanism to both levels, and the experiment proves its effectiveness. Since generative adversarial networks (GAN) were proposed by Goodfellow and others in 2014, it has occupied a large part of the deep learning network framework and is one of the most promising methods for unsupervised learning on complex distributions in recent years. GAN is composed of two models, generator and discriminator, and the two models play with each other to produce relatively good output [12]. The application of GAN in the field of medical image processing not only is limited to medical image classification but also has a great contribution to segmentation, detection, and enhancement of medical images. In the field of classification, Madani et al. [13] used the semisupervised learning characteristics of the generative adversarial networks. First, the authors solved the problems of medical labeling data scarcity and data overfitting and then used the deep generative adversarial network to learn the vision of the chest X-ray structure to classify abnormal and normal samples. Frid-Adar et al. [14] proposed a network for the classification of liver diseases. First, classical data enhancement is used to enlarge the data sets, and then, GAN technology is used to further expand the size and diversity of the data. Finally, 182 CT images of liver lesions are used for verification, and the sensitivity and specificity of the network are improved after adding synthetic data enhancement. Ma et al. [15] proposed a blood cell image classification framework based on Deep Convolutional Generative Adversarial Network (DC-GAN) and Residual Network (ResNet).

2.2. Deep Learning. As one of the most cutting-edge scientific and technological fields, deep learning has always been leading the progress of science and technology. It simulates the mechanism of the human brain to interpret data and establish a neural network for analytical learning. With the advent of the era of AI, deep learning has begun to shine in the field of machine learning research and application. The most notable applications are the fields of computer vision and natural language processing [16, 17], implemented by CNN, GAN, and RNN, respectively. Lecun et al. invented CNN as early as the 1990s. With the success of the ImageNet competition, more and more researchers have turned their attention to CNN [18]. Krizhevsky et al. [19], the champion of the ImageNet competition in 2012, first used the ReLU activation function, local response normalization, Dropout, and other tricks; improve the generalization ability of the network; and avoid overfitting of the model. Subsequently, Simonyan and Zisserman [20] proposed VGG, which is improved on the basis of AlexNet. The entire network uses the same size of 3×3 convolution kernel, which simplifies the structure of the neural network. In addition, VGG increases the number of layers of convolution since increasing network depth can improve network performance. GoogLeNet continues this feature. In order to prevent problems such as gradient disappearance and overfitting caused by deepening the number of network layers, the authors introduced the Inception structure. Inception improves the training results from another perspective and can extract more features with the same amount of calculation, thereby improving the training results [21]. In order to solve the problem of the inability to train when the level is deepened, He et al. [22] proposed ResNet. The authors use the residual block to learn the representation of the residual between input and output, and the internal residual block uses shortcut connections, which alleviate the problem of gradient disappearance caused by increasing depth in the deep neural network. Huang et al. [23] proposed Densenet in 2017. Inspired by ResNet and Highway Networks, the authors transmit information through a direct connection with a later layer. Its core idea is to establish the connection relationship between different layers, make full use of the characteristics of the image, and further alleviate the problem of gradient disappearance. The recurrent neural network (RNN) is one of the most common deep learning algorithms; it has taken the whole world by storm. Almost all state-of-the-art performance in natural language processing or understanding is attributed to variants of RNN [24].

2.3. Transfer Learning. Most tasks in medical image processing require doctors to annotate the original image, namely, GroundTruth. Labeling data is a tedious and expensive task. Since medical data is related to the privacy of patients, it is more difficult to build medical data sets. As one of the most effective methodologies in the current deep learning field, transfer learning alleviates the above problems. Transfer learning utilizes the similarity between data, tasks, and models to apply models trained on natural image classification tasks to medical image classification [25].

Long et al. [26] proposed a new adaptive method for the deep network domain, which can jointly learn adaptive classifiers and transferable features from labeled data in the source domain and unlabeled data in the target domain. By inserting several layers in the deep network to learn the residual function of the reference target classifier, classifier adaptation was realized. Chang et al. [27] proposed a multiscale convolutional sparse coding (MSCSC) method with unsupervised transfer learning ability. By strengthening scale specificity and combining automatic learning filter banks at different scales, the basic knowledge of transferability can be learned, and finally, the target task can be fine-tuned. Aiming at the problem of poor application of deep learning in the diagnosis and treatment of multiple retinal lesions, Choi et al. [28] adopted the random forest migration learning method based on the VGG19 structure. The authors found that transfer learning combined with ensemble classifiers can improve classification performance to detect multiclassification retinal diseases. Kaur and Gandhi [29] discussed the ability of the pretrained model DCNN VGG16 with transfer learning ability to classify pathological brain images and replaced the last few layers of the VGG16 model to adapt to new images in current applications. The verification on the test set showed good results in sensitivity, specificity, and accuracy. Hosny et al. [30] proposed a high-performance automatic skin lesion classification system using transfer learning and pretrained deep neural network, AlexNet. The transfer learning is achieved by fine-tuning the weight of

the architecture and replacing the classification layer with the softmax layer. Talo et al. [31] proposed deep transfer learning based on ResNet34 to automatically classify normal and abnormal brain MRI images. The authors use data augmentation, optimal-learning rate finder, and fine-tuning to train the model. The model obtained 100% of 5 times classification accuracy on 613 MR images, helping radiologists to perform daily MR imaging examinations.

3. Materials and Methods

3.1. Materials. Compared with other medical data sets, colonoscopy images are more difficult to collect, mainly due to the complexity and inconvenience of the inspection method. Endoscopy is shown in the form of video, and colonoscopy data sets must be viewed from beginning to end to capture frames with key information. The popular medical images of enteroscopy mainly include the Kvasir-SEG intestinal polyp data set (https://www.simula.no/publications/kvasir-seg-segmented-polyp-dataset) [32], CRCHisto Phenotypes datasets (https://warwick.ac.uk/fac/cross_fac/tia/data/crchistolabelednucleihe) [33], and CVC-Clinic data sets (http://www.cvc.uab.es/CVC-Colon/index.php/databases/) [34]. Most of these data sets are used for image segmentation, including intestinal diseases mainly polyps and tumors, and the data types are relatively single.

The experimental data in this paper are from real cases of patients in the anorectal department of a hospital. The colonoscopy images of patients were obtained by endoscopy. First, screening took place to eliminate the unusable data caused by insufficient bowel preparation and the examination process, then data cleaning; finally, a small data set is constructed. The data sets collected 430 intestinal images of patients suffering from various diseases, which are mainly divided into four categories: normal, inflammation, polyps, and cancer. Among them, there are 120 colonoscopy images for normal, 110 colonoscopy images for inflammation, 108 colonoscopy images for polyps, and 92 colonoscopy images for tumors. The category label of each image was obtained by the joint diagnosis of the doctors in the anorectal department, and the subdivided category and characteristics were discussed to obtain the final data sets with the label. Figure 1 is an example diagram of these data sets.

In addition, we have enhanced the data set to increase the diversity of samples, which also made the trained deep learning framework have high generalization ability and strong robustness. Through the four data enhancement methods of brightness enhancement, contrast enhancement, image flip, and angle rotation, 480 images of normal category, 440 images of inflammation category, 432 images of polyp category, and 368 images of tumor category were obtained, respectively. Figure 2 is the result of data enhancement after randomly selecting one of the images. We randomly select 80% of the pictures from the data set as the training set for training and 20% of the pictures as the test set for testing. The distribution structure is shown in Table 1.

3.2. Methods

3.2.1. Multiclassification Task Based on Transfer Learning. Transfer learning refers to the process of knowledge transfer in two different domains. The knowledge learned in the source domain S is used to help the learning task in the target domain T, and the number of training samples in the source domain is generally far greater than that in the target domain. Generally speaking, according to different ways of transfer, transfer learning is divided into inductive transfer learning and transduction transfer learning. This paper uses inductive transfer learning [35]. The specific implementation ideas are summarized as follows: Firstly, a pretraining model with good generalization performance is obtained by training the deep learning model on the large-scale image classification data set, ImageNet 1000. Then, inductive transfer learning was used to transfer these pretraining models to the intestinal multiclassification model to reduce the training burden of insufficient colonoscopy images in our task. Finally, the model was adjusted and updated by using the established colonoscopy four-classification images to improve the classification accuracy.

(1) Pretraining Model Processing. The classic classification deep learning model is used as the classification backbone network to generate the corresponding pretraining model. The essence of generating the pretraining model is to train the data set of the source domain from the beginning using the classification network model. For different layers of CNN, the image features learned by each layer are different. The features learned in the shallower layer are more general, and the features learned in the deeper layer are more relevant to specific tasks. In Figure 3, the shallowest common feature "lines" are the same for the classification task of faces, cars, elephants, and chairs [36]. ImageNet 1000 classification is a subset of ImageNet, with a training set of about 1.2 million pieces, a verification set of 50,000 pieces, and a test set of 100,000 pieces. It belongs to 1,000 different categories, and each image is strictly manually screened and labeled. Multiclassification of a variety of images is implemented on the ImageNet data set, which is more suitable for the target task of this paper. Therefore, we chose to use the model trained on ImageNet 1000 classification to save it as a pretraining model.

(2) Transfer the Pretraining Model. The pretraining model used in this paper is a deep neural network, in which the mobility of each layer is not the same. In general, the lower layers of the network learn some general low-level features, the middle or higher layers learn abstract high-level semantic features, and the last layers generally learn task-specific features. Therefore, according to the feature of the target task and its relevance to the source domain, different layers of the pretraining model can be selected to transfer to the target task. In network training, the initial weight and bias of the convolutional layer are randomly assigned, and the quality of this value directly affects the final performance of the model to a large extent. Therefore, we give the convolutional layer an initial weight and bias and then continuously adjust to the task itself based on the feedback of the network. In this paper, the convolutional layer of other pretrained models is transferred to the colonoscopy image classification task, and the fully connected layer is retrained.

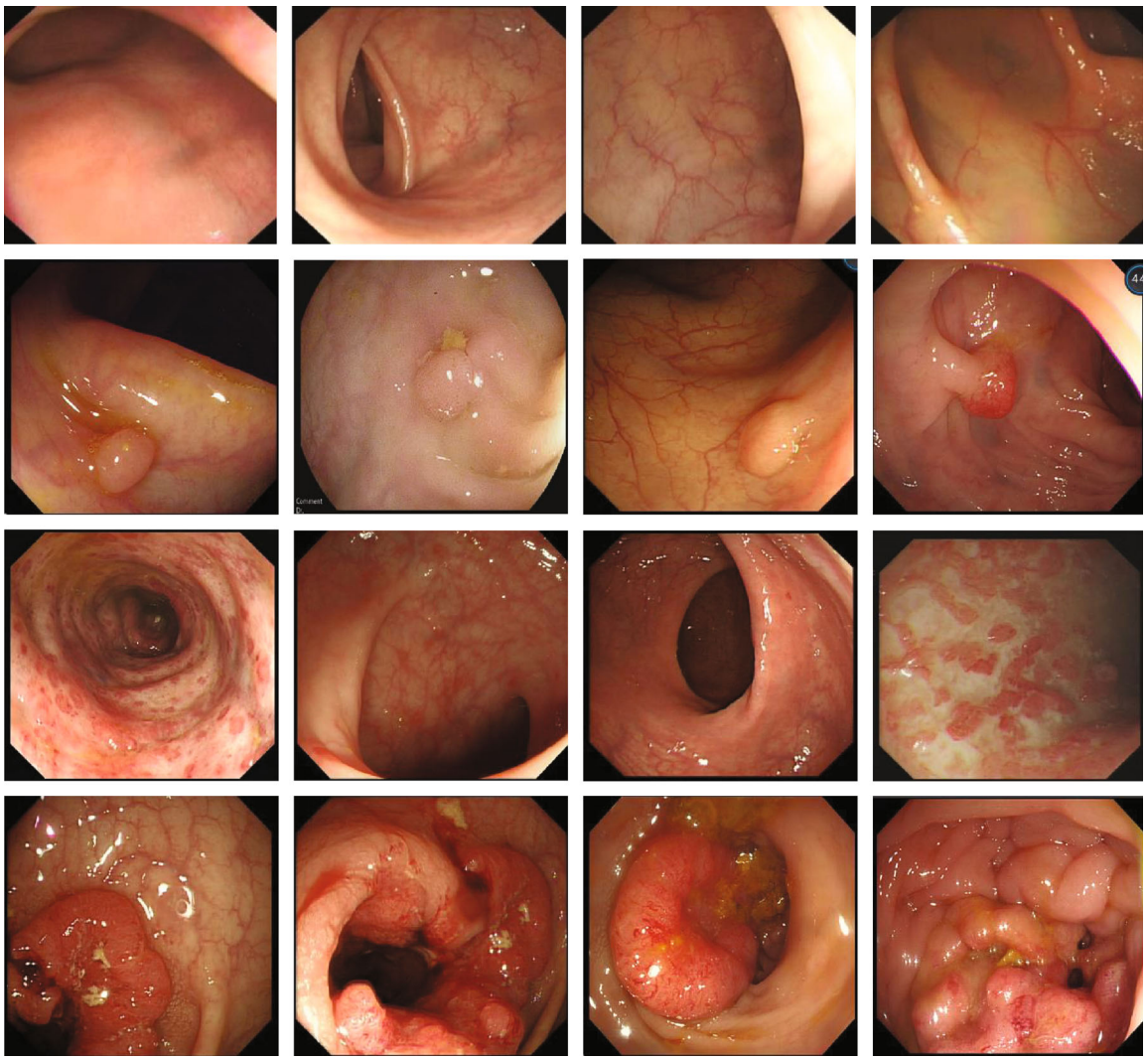


Figure 1: Sample images of four types of colonoscopy in the data set.

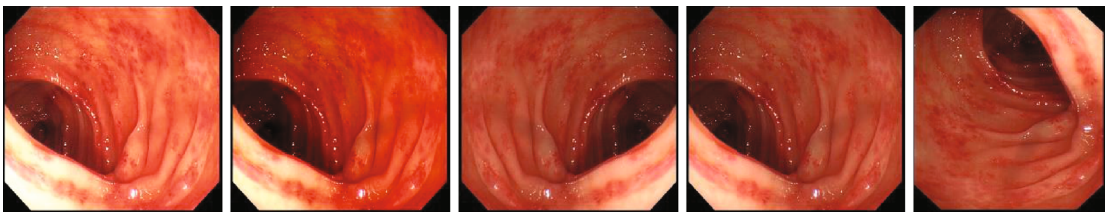


Figure 2: The effect of data enhancement.

Table 1: Distribution of test and training data.

Data category	Normal	Inflammation	Polyp	Cancer	Total quantity
Number of test sets	120	110	110	95	435
Number of training sets	480	440	430	365	1715
Total quantity	600	550	540	460	2150

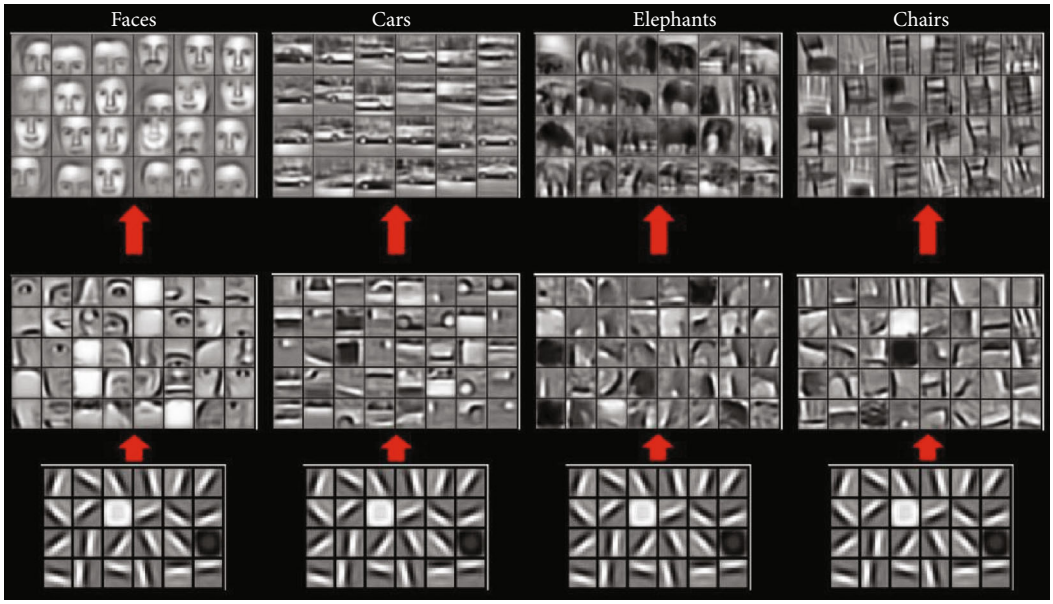


Figure 3: Features learned from different object classes.

(3) Fine-Tuning. Fine-tuning is essential for a successful deep transfer learning model. The parameters of the pretraining model were adjusted by the similarity of the data between the pretrained data set and the colonoscopy image data set. During this process, we made several attempts. As shown in Figure 4, the ImageNet 1000 classification model is on the left and the colonoscopy classification model is on the right. In this paper, the weight and bias of the initial convolutional layer of the pretraining model remain unchanged, and only the fully connected layer related to the final output classification result is modified. The training model was adjusted by using the four-classification data of colonoscopy to find an optimal parameter of the full connective layer. Finally, according to the classification results, the best match between the convolutional layer and the fully connected layer is found, and the best parameters of the fully connected layer are saved to obtain an optimal complete model.

3.2.2. Deep Learning Model Based on Transfer Learning. The mainstream deep learning model in medical image classification tasks is mainly CNN. Since the development of CNN, there have been many types of excellent network frameworks. Among them, several classic models have milestone significance in the development of CNN, such as LeNet, AlexNet, VGG, and GoogLeNet. For the data sets and tasks used in this paper, we choose AlexNet, VGG, and ResNet as the backbone network from the depth of the network. This section discusses the deep convolutional network we use based on transfer learning.

The AlexNet network itself uses many modern deep convolutional network technical methods, and its network structure is shown in Figure 5. It uses Dropout to alleviate the occurrence of overfitting problems and make the model more generalized. The GPU is used for parallel training, so that the model can be trained quickly. In addition, AlexNet amplifies the data set by making small changes to the original image and uses data enhancement to improve the accuracy of the model. Therefore, AlexNet embodies unique advantages in solving large-scale data classification problems [19]. We used AlexNet for the transfer learning of ImageNet 1000 classification, and the AlexNet parameters trained by ImageNet were transferred to the intestinal image training task. When initializing the convolutional layers of the first five layers, the existing parameters are used, while random initialization is used in the full connection layer of the last three layers, and then, adjustments are made according to the intestinal data set.

According to the size of the established data set, AlexNet is beneficial to improve classification accuracy. In addition, AlexNet uses ReLU as the activation function to solve the situation where the gradient disappears when the Sigmoid function is backpropagated, which is conducive to the training of the model. The formula is as follows:

$$f(x) = \max(0, x). \tag{1}$$

When VGG is used for large-scale image classification, it can be extended to various tasks and data sets. According to the characteristics of our task, VGG, which is suitable for arbitrary classification, is applied to the four classifications of intestinal image. VGG emphasizes the depth in the design of convolutional neural networks, and the accuracy of ImageNet 1000 classification is also different for different network depths. VGG uses a smaller convolution kernel to replace a large-size convolution, which saves computing resources, thereby setting aside resources for deepening the network. There are 5 convolutions in VGG; each convolutional layer is followed by a pooling layer. The network structure is shown below. The expansion performance of its network is more prominent; there are mainly two commonly used layers, VGG16 and VGG19. We use VGG16 and VGG19 to set an initial value for the convolution layer in

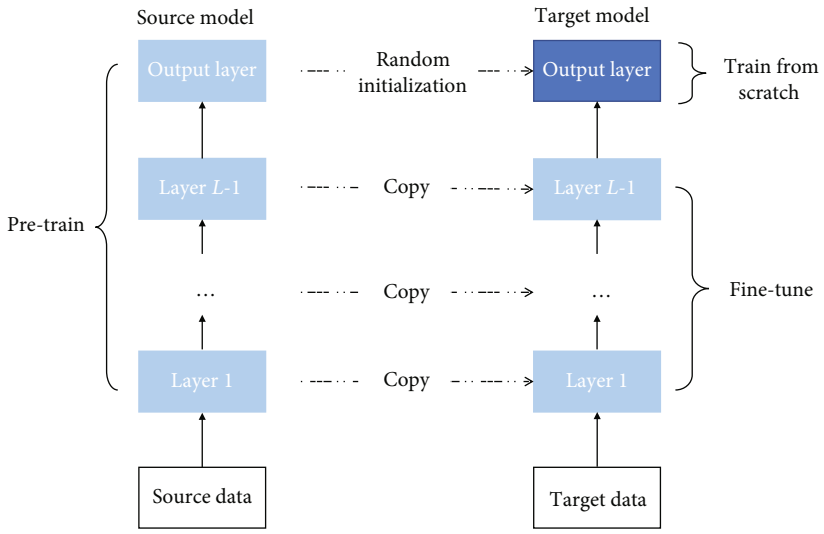


Figure 4: Schematic diagram of fine-tuning in deep learning.

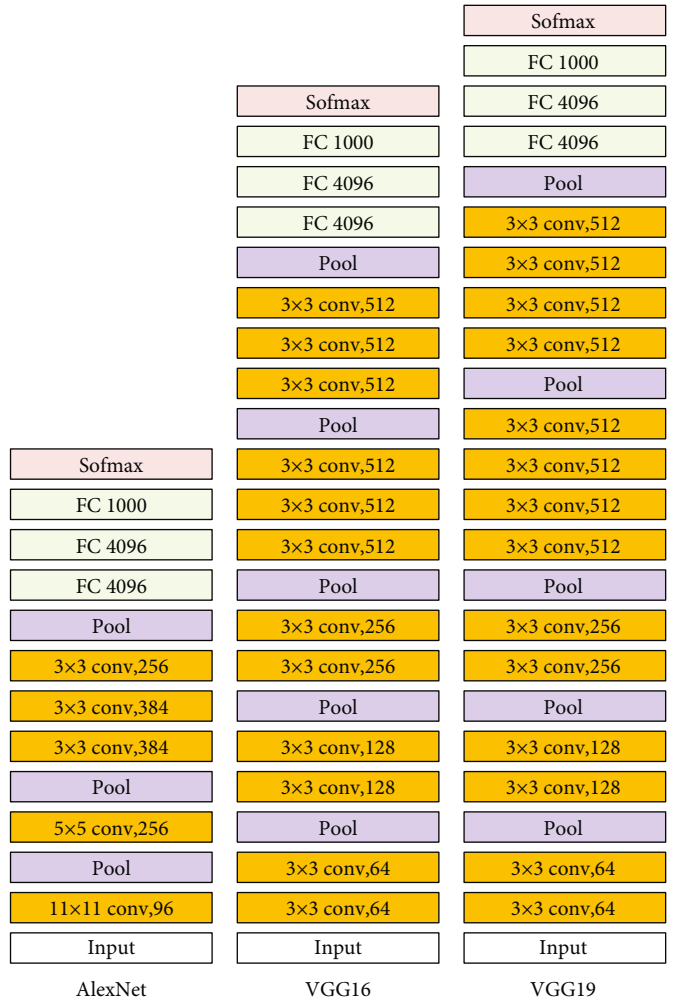


Figure 5: The network structure of AlexNet, VGGNet16, and VGGNet19.

Table 2: Classification results on the traditional deep learning model.

Model	Normal (T/total)	Inflammation (T/total)	Polyp (T/total)	Cancer (T/total)	Accuracy (%)
AlexNet	3/12	5/11	8/11	2/9	41.86
VGG16	10/12	4/11	10/11	8/9	74.42
VGG19	11/12	9/11	7/11	1/9	65.12
ResNet50	11/12	10/11	10/11	5/9	83.72
ResNet101	9/12	9/11	6/11	8/9	74.42

the two models by training on ImageNet 1000 classification and do not deal with other layers. The concise structure of VGG16 and VGG19 makes its migration performance good, and the process of migrating to other data sets is simple [20].

Considering that the gradient disappears more obviously with the increase in the number of layers of the neural network during the training of intestinal image, we considered using ResNet with a residual structure to obtain better results [22]. ResNet was trained by using ImageNet, the pretraining data set we selected. Since its network is no longer a simple stacked structure, it solves the problem of less and less obvious gradients caused by the increase in the number of network layers. The formula is defined as follows:

$$y = F(x, \{w_i\}) + x, \tag{2}$$

where x and y are the input and output vectors of this layer. The $F(x, \{w_i\})$ function represents the remaining mappings to be learned. The obtained ResNet pretraining model solves the problems of information loss and loss to some extent. By directly passing the input information to the output, the integrity of the information is protected and the learning objectives and difficulty are simplified.

4. Experimental Results and Analysis

In this section, we discuss the results obtained when the proposed multiclassification method based on deep transfer learning is applied to the proposed colonoscopy image data set. In order to prove that our proposed method is feasible, we selected multiple commonly used deep learning networks to compare the performance of each network. The parameter setting of the experiment is introduced in detail below, and results of the experiment are analyzed at last.

4.1. Setting of Experimental Parameters. This experiment is based on Tensorflow and runs under the Ubuntu 16.04.5 LTS system. The hardware includes AMD Ryzen 5 1600 Six-Core Processor, NVIDIA GeForce GTX 1080 GPU.

(1) *Size Selection*. The size of the data set is not uniform. The average width, maximum width, and minimum width are 780, 886, and 750, respectively. The average height, maximum height, and minimum height are 655, 729, and 599, respectively. The original shape of the intestinal wall will be destroyed after cutting from the colonoscopy image. However, without cutting, the size of the colonoscopy image cannot be unified, and it is difficult to use deep learning for classification learning. Considering that the background of colonoscopy images is a mostly round-like intestinal wall, the original image size is changed to a size of 224×224 through the resize operation. The bit depth of the image is uniformly modified to 24. In order to maintain the integrity of the intestinal wall in the image while making the data meet the input requirements of the model in this paper, we unified the data size to $224 \times 224 \times 3$.

(2) *Function Selection*. As a classic loss function in classification tasks, cross-entropy can avoid gradient dispersion when performing gradient descent calculations, resulting in a decrease in the learning rate. It makes the learned model distribution closer to the real data distribution, thereby improving the accuracy of multiclassification. ReLU is selected as the activation function. Compared with Sigmoid and Tanh, the actual convergence speed is faster and there is no complex exponential operation, which makes the network training faster. After comparing several optimization functions: ADAM, RMSProp, and SGD, we finally choose SGD as our optimization function.

(3) *Migration Layer Selection*. Considering that all the layer features in the original task do not promote the target task, only partial layer migration is carried out. We choose to use all the convolutional layer parameters obtained on the pretrained model to train the fully connected layer in the target domain after migration.

(4) *The Other*. During model training, the batch size was 64, the epoch was equal to 10000, and the learning rate was set to 0.0001.

4.2. Analysis of Experimental Results. To demonstrate the effectiveness of our proposed model, we evaluated the performance of the target model through repeated validation. We divide the experiment into four groups, do not add transfer learning and data enhancement, add data enhancement without transfer learning, only add transfer learning without data enhancement, add transfer learning and data enhancement, and compare and analyze the impact of each strategy on accuracy. The classic deep learning models used in the experiment include AlexNet, VGG16, VGG19, ResNet50, and ResNet101.

In this experiment, we compared the classification accuracy of the traditional deep learning model and the proposed

Table 3: Classification results on the deep transfer learning model.

Model	Normal (T/total)	Inflammation (T/total)	Polyp (T/total)	Cancer (T/total)	Accuracy (%)
AlexNet-tl	6/12	6/11	8/11	4/9	55.81
VGG16-tl	10/12	7/11	10/11	7/9	79.07
VGG19-tl	9/12	8/11	8/11	5/9	69.77
ResNet50-tl	11/12	10/11	10/11	8/9	90.70
ResNet101-tl	10/12	9/11	8/11	8/9	81.40

Table 4: Classification results of traditional deep learning model based on data enhancement.

Model	Normal (T/total)	Inflammation (T/total)	Polyp (T/total)	Cancer (T/total)	Accuracy (%)
AlexNet	81/120	47/110	69/110	23/95	50.57
VGG16	112/120	67/110	79/110	76/95	76.78
VGG19	101/120	47/110	83/110	61/95	67.13
ResNet50	109/120	97/110	96/110	79/95	87.59
ResNet101	106/120	88/110	90/110	63/95	79.77

Table 5: Classification results of deep transfer learning model based on data enhancement.

Model	Normal (T/total)	Inflammation (T/total)	Polyp (T/total)	Cancer (T/total)	Accuracy (%)
AlexNet-tl	89/120	56/110	79/110	31/95	58.62
VGG16-tl	119/120	72/110	87/110	82/95	82.76
VGG19-tl	106/120	52/110	91/110	67/95	72.64
ResNet50-tl	115/120	106/110	102/110	88/95	94.48
ResNet101-tl	111/120	95/110	98/110	72/95	86.44

Table 6: Comparison of classification models on traditional deep learning models and deep transfer-based learning.

Model	AlexNet		VGG16		VGG19		ResNet50		ResNet101	
	Traditional	Our	Traditional	Our	Traditional	Our	Traditional	Our	Traditional	Our
Acc (%)	41.86	55.81	74.42	79.07	65.12	69.77	83.72	90.70	74.42	81.40
Acc-DA (%)	50.57	58.62	76.78	82.76	67.13	72.64	87.59	94.48	79.77	86.44

Table 7: Comparison of transfer learning and standard machine learning.

Learning type	Sample space	Probability distribution
Standard machine learning	$x_S = x_T, y_S = y_T,$	$p_S(x, y) = p_T(x, y)$
Transfer learning	$x_S \neq x_T$ or $y_S \neq y_T$	or $p_S(x, y) \neq p_T(x, y)$

solution. The experimental results obtained are as follows. The models proposed in this paper have better performance than traditional deep learning models.

First of all, we trained each classical deep learning model from beginning to end to apply the four classifications of colonoscopy images without any data enhancement and transfer learning in the textual colonoscopy data set. Secondly, in order to compare the effect of transfer learning, we used the model of transfer learning training to get another set of results. Due to the small amount of data, if the data set is divided into 8 training sets and 2 test sets, the experimental results show that the model is easy to overfit. Therefore, 90% of the 430 data sets were randomly selected in this paper as the training set and 10% as the test set, and the results are obtained as shown in Tables 2 and 3, respectively. Tables 2 and 3 describe the number and accuracy of the correct classification of the four categories on each model, where T represents that the actual category is identical to the predicted category.

On the basis of data enhancement of the data set, we use the multiclassification model based on deep transfer learning proposed in this paper for training. By comparing the data enhancement strategy training model and the transfer learning combined with the data enhancement strategy training model, the effectiveness of transfer learning is verified. Since the amount of data has increased after data enhancement, this paper divides the data set according to 8 : 2 of the training set to the test set. The results shown in Tables 4 and 5 were obtained, respectively.

The experimental results under the two models are compared, as shown in Table 6. Due to the small number of data

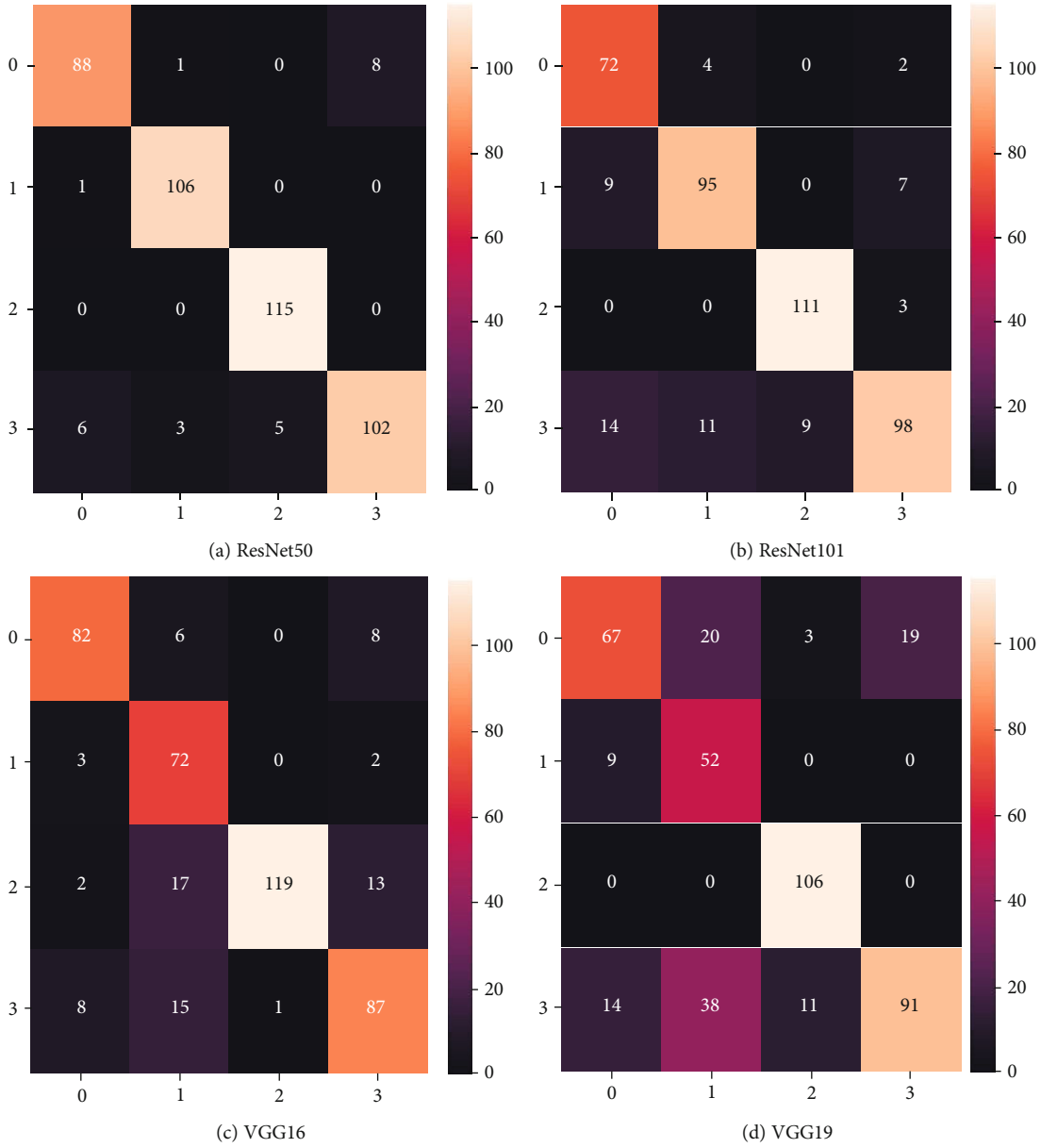


Figure 6: Confusion matrix diagram of four classifications of each model.

sets and the difference in the number of classifications from the widely used diconomies in the conventional models, the classification effect of the traditional deep learning model is poor. In the same data set, it can be seen that accuracy of the model using transfer learning is greatly improved compared with the traditional deep learning model.

Deep learning requires a large number of data sets for training to prevent overfitting of the model, and such a large amount of data is more beneficial to transfer learning. Suppose the sample space of a machine learning task T is $x \times y$, where x is the input space and y is the output space, and its probability density function is $p(x, y)$. Learn from the statistical point of view; a machine learning task T is defined as a modeling problem of conditional probability $p(y \mid x)$ in a domain D. Table 7 shows a comparison between transfer learning and standard machine learning. Therefore, the results of transfer learning are better than traditional deep learning models.

We output the confusion matrix of the four classification results of each model in the intestine, as shown in Figure 6. Among them, 0 represents cancer, 1 represents inflammation, 2 represents normal, and 3 represents polyps. Among the five transfer learning models, ResNet50-tl has the best performance compared with the other models, with a performance of 94.48%. ResNet50 improves the accuracy of the network by using residuals to deepen the network structure. VGG only increases the depth of the network by simply superimposing the convolutional layer, which leads to the disappearance of the model gradient; thus, the classification accuracy is not high. Compared with the scale of the data set used in this paper, ResNet101 has too many layers in the network, which easily leads to overfitting of the model.

Among the four types of colonoscopy images, polyps were correctly diagnosed more often than the other two types of diseases except for normal types. The characteristics of polyp lesions are outstanding, and feature extraction is used. Under the condition of ensuring the classification accuracy of other categories, the model can improve the number of polyp diagnosis and assist doctors in making judgments. Therefore, it is confirmed through experiments that the ResNet four-classification of intestines based on transfer learning alleviates the problem of small data sets in medical image processing. Transfer learning makes the model more robust and generalizable, which is of great clinical value.

5. Conclusion and Prospect

At present, with the continuous improvement of medical technology, computer-aided diagnosis has occupied an important position in the medical industry. In particular, the combination of deep learning and medical image processing has been deeply explored by many researchers. In this interdisciplinary field of integrated medical imaging, mathematical modeling, digital image processing, and so on, there are still many complex and diverse problems that need to be solved urgently.

In medical image analysis, the cost of data acquisition and annotation is high, and it is very difficult to construct a large-scale standard data set, which hinders the development of medical image diagnosis. The common solution is to use data enhancement techniques common in deep learning, such as geometric transformation and color transformation. This paper proposes a deep learning framework based on transfer learning for multidisease colonoscopy image classification. The network model with better training results on natural images is transferred to the specified classification task, and the existing data sets are used to fine-tune the tasks. Finally, the feasibility of the proposed method is verified by experiments. Using this method can help us save training time and improve learning accuracy. Generally, pretrained models are trained on a large data set, which invisibly expands our training data sets and makes the model more robust and capable of generalization.

With the development of deep learning, deep transfer learning will be widely used to solve many challenging problems. However, transfer learning also has its shortcomings. For example, categories that exist only in the source domain but not in the target domain will have a negative transfer impact on the transfer result. Therefore, we need to go one step further and develop different strategies of transfer learning for the deep network.

Data Availability

The colonoscopy image data used to support the findings of this study are available from the first author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yan Wang, Zixuan Feng, and Liping Song contributed equally to this work and are co-first author.

Acknowledgments

This work is partly funded by the Scientific Research Project of Education Department of Jilin Province with No. JJKH20211062KJ, Natural Science Foundation of Hunan Province with No. 2020JJ4434, Key Scientific Research Projects of Department of Education in Hunan Province with No. 19A312, Hunan Provincial Science & Technology Project Foundation (2018TP1018 and 2018RS3065), Innovation and Entrepreneurship Training Program of Hunan Xiangjiang Artificial Intelligence Academy, and Educational Reform Project of Hunan Xiangjiang Artificial Intelligence Academy.

References

[1] Y. D. Zhang, Z. Dong, S. H. Wang et al., "Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation," *Information Fusion*, vol. 64, pp. 149–187, 2020.

[2] M. F. Ijaz, M. Attique, and Y. Son, "Data-driven cervical cancer prediction model with outlier detection and over-sampling methods," *Sensors*, vol. 20, no. 10, p. 2809, 2020.

[3] G. Alfian, M. Syafrudin, M. F. Ijaz, M. Syaekhoni, N. Fitriyani, and J. Rhee, "A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing," *Sensors*, vol. 18, no. 7, p. 2183, 2018.

[4] M. F. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, "Hybrid prediction model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, synthetic minority over sampling technique (SMOTE), and random forest," *Applied Sciences*, vol. 8, no. 8, p. 1325, 2018.

[5] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, 2021.

[6] J. C.-W. Lin, Y. Djenouri, G. Srivastava, U. Yun, and P. Fournier-Viger, "A predictive GA-based model for closed high-utility itemset mining," *Applied Soft Computing*, vol. 108, article 107422, 2021.

[7] S. H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y. D. Zhang, "Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network," *Information Fusion*, vol. 67, pp. 208–229, 2021.

[8] H. I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. MICCAI 2013*, pp. 583–590, Springer, Berlin, Heidelberg, 2013.

[9] X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 11, pp. 2693–2701, 2015.

[10] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, 2016.

[11] J. C. W. Lin, Y. Shao, Y. Djenouri, and U. Yun, "ASRNN: a recurrent neural network with an attention model for sequence labeling," *Knowledge-Based Systems*, vol. 212, article 106548, 2021.

[12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[13] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Semi-supervised learning with generative adversarial networks for chest X-ray classification with ability of data domain adaptation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1038–1042, Washington, DC, USA, 2018.

[14] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using GAN for improved liver lesion classification," in *2018 IEEE 15th International Symposium on Biomedical Imaging*, pp. 289–293, Washington, DC, USA, 2018.

[15] L. Ma, R. Shuai, X. Ran, W. Liu, and C. Ye, "Combining DC-GAN with ResNet for blood cell image classification," *Medical & Biological Engineering & Computing*, vol. 58, no. 6, pp. 1251–1264, 2020.

[16] J. Tamang, J. D. Dieu Nkapkop, M. F. Ijaz et al., "Dynamical properties of ion-acoustic waves in space plasma and its application to image encryption," *IEEE Access*, vol. 9, pp. 18762–18782, 2021.

[17] C. L. Chowdhary, P. V. Patel, K. J. Kathrotia, M. Attique, K. Perumal, and M. F. Ijaz, "Analytical study of hybrid techniques for image encryption and decryption," *Sensors*, vol. 20, no. 18, p. 5162, 2020.

[18] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations*, pp. 1–14, San Diego, CA, USA, 2015.

[21] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, 2015.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, Honolulu, HI, USA, 2017.

[24] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, https://arxiv.org/abs/1409.2329.

[25] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International Conference on Artificial Neural Networks*, pp. 270–279, Rhodes, Greece, 2018.

[26] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *30th Conference on Neural Information Processing Systems*, pp. 1–9, Barcelona, Spain, 2016.

[27] H. Chang, J. Han, C. Zhong, A. M. Snijders, and J. H. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1182–1194, 2018.

[28] J. Y. Choi, T. K. Yoo, J. G. Seo, J. Kwak, T. T. Um, and T. H. Rim, "Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database," *PLoS One*, vol. 12, no. 11, article e0187336, 2017.

[29] T. Kaur and T. K. Gandhi, "Automated brain image classification based on VGG-16 and transfer learning," in *2019 IEEE International Conference on Information Technology*, pp. 94–98, Bhubaneswar, India, 2019.

[30] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with Alex-net," *PLoS One*, vol. 14, no. 5, article e0217293, 2019.

[31] M. Talo, U. B. Baloglu, Ö. Yıldırım, and U. Rajendra Acharya, "Application of deep transfer learning for automated brain abnormality classification using MR images," *Cognitive Systems Research*, vol. 54, pp. 176–188, 2019.

[32] D. Jha, P. H. Smedsrud, M. A. Riegler et al., "Kvasir-seg: a segmented polyp dataset," in *MultiMedia Modeling. MMM 2020*, pp. 451–462, Springer, Cham, Switzerland, 2020.

[33] K. Sirinukunwattana, S. E. A. Raza, Y. W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[34] G. Fernández-Esparrach, J. Bernal, M. López-Cerón et al., "Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps," *Endoscopy*, vol. 48, no. 9, pp. 837–842, 2016.

[35] J. Lin, G. Srivastava, Y. Zhang, Y. Djenouri, and M. Aloqaily, "Privacy-preserving multiobjective sanitization model in 6G IoT environments," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5340–5349, 2021.

[36] S. H. Wang, X. Wu, Y. D. Zhang, C. Tang, and X. Zhang, "Diagnosis of COVID-19 by wavelet Renyi entropy and three-segment biogeography-based optimization," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 1332–1344, 2020.

Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 9960199, 9 pages
https://doi.org/10.1155/2021/9960199

Research Article

LCC-Net: A Lightweight Cross-Consistency Network for Semisupervised Cardiac MR Image Segmentation

Lai Song,[1] **Jiajin Yi**,[1] **and Jialin Peng**[1,2]

[1]*College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China*
[2]*Xiamen Key Laboratory of Computer Vision and Pattern Recognition, Huaqiao University, Xiamen 361021, China*

Correspondence should be addressed to Jialin Peng; 2004pjl@163.com

Received 23 March 2021; Revised 22 April 2021; Accepted 29 April 2021; Published 17 May 2021

Academic Editor: Yuankai Huo

Semantic segmentation plays a crucial role in cardiac magnetic resonance (MR) image analysis. Although supervised deep learning methods have made significant performance improvements, they highly rely on a large amount of pixel-wise annotated data, which are often unavailable in clinical practices. Besides, top-performing methods usually have a vast number of parameters, which result in high computation complexity for model training and testing. This study addresses cardiac image segmentation in scenarios where few labeled data are available with a lightweight cross-consistency network named LCC-Net. Specifically, to reduce the risk of overfitting on small labeled datasets, we substitute computationally intensive standard convolutions with a lightweight module. To leverage plenty of unlabeled data, we introduce extreme consistency learning, which enforces equivariant constraints on the predictions of different perturbed versions of the input image. Cutting and mixing different training images, as an extreme perturbation on both the labeled and unlabeled data, are utilized to enhance the robust representation learning. Extensive comparisons demonstrate that the proposed model shows promising performance with high annotation- and computation-efficiency. With only two annotated subjects for model training, the LCC-Net obtains a performance gain of 14.4% in the mean Dice over the baseline U-Net trained from scratch.

1. Introduction

Medical image analysis plays an increasingly important role in routine clinical work. Magnetic resonance imaging (MRI) is a noninvasive technique for investigating cardiac structures, thus widely used in clinical diagnosis and treatment. Segmentation of the left ventricle (LV), right ventricle (RV), and the myocardium (MYO) from cardiac MR images can provide crucial diagnostic parameters about the cardiac. Recently, convolutional neural networks (CNNs), mostly fully convolutional networks (FCNs) [1, 2], have made substantial progress for cardiac image segmentation [3]. However, the current supervised-learning models rely heavily on a large amount of manually labeled data for model training to achieve competitive performance. Unfortunately, manually labeling cardiac MR images is time-consuming and labor-intensive and requires strong domain knowledge from experts. Moreover, most of the top-performing methods are deep and wide convolutional neural networks involving a massive number of training parameters, which not only increases the chance of overfitting but also hinders their applications in clinical routines. To address the above problems, we introduce a lightweight deep network for semisupervised segmentation of cardiac images. Our model is trained only on a few labeled subjects and a more considerable number of unlabeled subjects.

There are generally two paradigms to make use of unlabeled data. The first one is unsupervised or self-supervised pretraining, followed by fine-tuning on a small set of labeled data. The second paradigm is to jointly use the labeled data and unlabeled data through pseudo labeling [4] or consistency regularization [5–8]. Since there is an obvious gap between the objectives of the unsupervised pretraining and the downstream segmentation, the effect of unsupervised pretraining is not always significant. In this study, we follow the second paradigm and make use of the unlabeled data by enforcing consistency regularization on the supervised model, aiming to improve the generalization ability of the

supervised trained model and reduce the risk of overfitting. Consistency regularization encourages the segmentation prediction to be consistent on the unlabeled examples under different data perturbations or among different models. We follow the studies in [6, 9, 10] and enforce consistency among different models' predictions. Both strong and weak perturbations are applied.

In this study, we propose a lightweight network, LCC-Net, for semisupervised segmentation of cardiac MR images based on consistency training cross models. To be specific, our model, as shown in Figure 1, consists of one shared encoder and three separate decoders: one decoder for supervised learning and the other two decoders for unsupervised consistency learning. Following a similar strategy as in [6], different perturbations are injected on the two unsupervised decoders. We enforce consistency between the predictions of the supervised decoder and unsupervised decoders to make the learned model less sensitive to the extra perturbation. To further improve model robustness and reduce the risk of overfitting, we augment the input data, both the labeled and unlabeled data, with extreme perturbations realizing significant gains. While the previous semisupervised models suffer from a massive scale of parameters and high computational complexity, we lighten our model with the lightweight Ghost module introduced in [11]. Moreover, we validate the proposed method on the ACDC [12] dataset.

The rest of the paper is organized as follows. In Section 2, we briefly review the related work. Section 3 presents the proposed method, which is evaluated on challenging cardiac segmentation tasks in Section 4. Section 5 concludes this study.

2. Related Work

2.1. Cardiac MR Image Segmentation Methods. For cardiac MR image segmentation, Painchaud et al. [13] presented a postprocessing VAE [14], which converts anatomically invalid cardiac shapes into close but correct shapes for introducing strong anatomical guarantees into the network. Khened et al. [15] proposed Densely Connected Fully Convolutional Network (DFCN), which is based on DenseNets [16]. Yang et al. [17] proposed a general and fully automatic solution to concurrently segment three important ventricular structures, starting from 3D Fully Convolutional Network (3D FCN). Simantiris and Tziritas [18] proposed a different Dilated CNN structure that incorporating domain-specific constraints. Isensee et al. [19] combined 2D U-Net and 3D U-Net, obtaining the best performance on the ACDC dataset. However, due to the combination of two different models, the numbers of model params is enormous. All these methods base on supervised learning proposed a series of efficient methods from different perspectives. When it comes to semisupervised cardiac MR image segmentation methods, there are still limitations for obtaining remarkable performance because cardiac MR image segmentation is a particular issue, including unique data distribution and difficult segmentation tasks.

2.2. Semisupervised Learning Methods. As for general semisupervised learning, many methods are proposed to reduce the burden of pixel-wise manual annotations for images, such as pseudo labeling [1], graph-based methods [20, 21], and entropy minimization [5]. Besides, mean-teacher [9] is another notable paradigm for semisupervised learning, which could be used in medical image segmentation. The mean-teacher model has two subnetworks: the teacher network and the student network, and learn cross-consistency from unlabeled data by exerting different perturbances on the two subnetworks. Yu et al. [22] proposed the uncertainty-aware mean teacher (UA-MT) framework, learning from the meaningful and reliable targets by exploiting the uncertainty information. Adversarial learning [23] methods are aimed at matching labeled and unlabeled images and improving testing time performance. Hung et al. [24] proposed a novel method in semisupervised semantic segmentation by introducing adversarial learning. Nie et al. [25] proposed attention-based semisupervised deep networks (ASDNet), where they integrated adversarial learning by a confidence network. Virtual Adversarial Training (VAT) [26] utilizes adversarial learning from a novel perspective and alters the model's predictions the most by approximating the perturbations. Laine and Aila [10] introduced consistency regularization into semisupervised learning, including π-model [10] and temporal ensembling method [10]. Bortsova et al. [27] proposed a novel semisupervised method that learns to predict segmentations consistent under a given class of transformations on both labeled and unlabeled images. The above methods enforce the consistency between predictions and provide critical data information to the supervised trained model. Besides, a series of strong data augmentation methods are proposed for overcoming the limitation of labeled training data, such as MixUp [28], CutMix [29], and Mosaic [30]. CowMix [31] starts from MixUp and enforces the consistency between the mixed outputs and the prediction over the mixed inputs. All the above data augmentation methods have made efforts to semisupervised learning by increasing training data diversity.

2.3. Lightweight Deep Networks. Current existing lightweight methods for networks can be divided into model compression and lightweight architecture design. We mainly review methods designing lightweight architectures, which are more related to our study. The increasing need to deploy deep models on computationally limited platforms and process large-scale data encourages lightweight architecture design. A series of lightweight convolutional modules have been proposed to balance the model performance and computational complexity. In particular, depth-wise convolution [32] and group convolution [33, 34] have gained much attention and have been building blocks for many lightweight architectures. MobileNet [35] used depth-wise separable convolution [32], a combination of depth-wise convolution and point-wise convolution, to build a lightweight model. ShuffleNet [36] is presented with point-wise group convolution and channel shuffle, which improves the information flow exchange between channel groups. Recently, Han et al. [11] proposed GhostNet with a novel Ghost module, which utilizes group convolution to further explore correlation and redundancy between feature maps. The GhostNet has shown higher

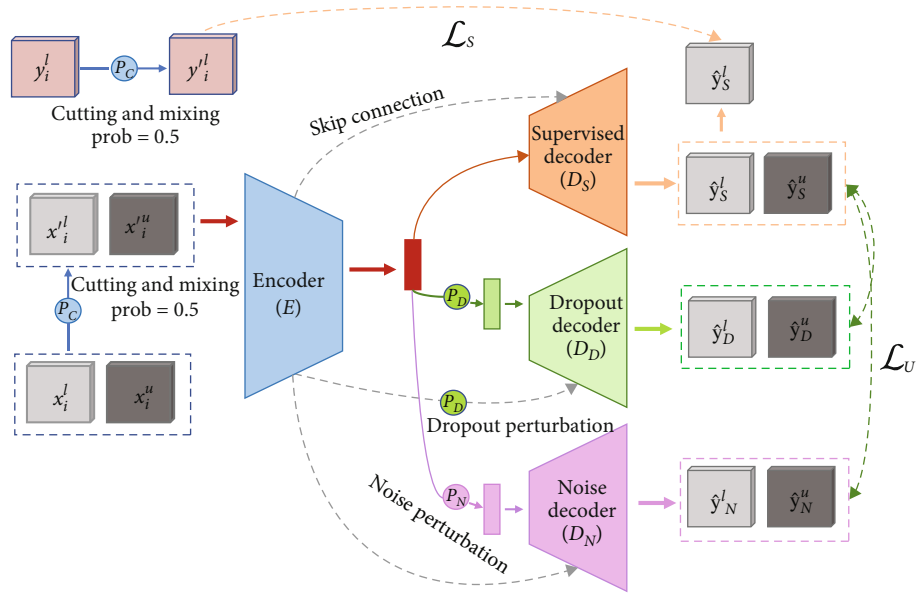


FIGURE 1: Overview of our LCC-Net for semisupervised segmentation. The network contains one supervised decoder and two unsupervised decoders. While the supervised decoder is trained with the labeled data, the two auxiliary decoders are trained with both labeled data and unlabeled data using unsupervised consistency losses. We inject dropout perturbation and noise perturbation in the feature space and inject cutting and mixing perturbation in the input image space.

recognition performance in natural images but has not been applied in medical image segmentation tasks.

3. Methods

3.1. Problem Formulation. We aim to develop a deep network model for semantic segmentation of cardiac MR images with only a few annotated subjects and a larger set of unlabeled subjects. We segment cardiac MR sequences in a slice-by-slice manner. Assume $\mathcal{D}_l = \{X^l, Y^l\}$ denote the labeled data, in which $X^l = \{x_1^l, x_2^l, \cdots, x_n^l\}$ contains n image slices, and $Y^l = \{y_1^l, y_2^l, \cdots, y_n^l\}$ is ground truth. $\mathcal{D}_u = \{x_1^u, \cdots, x_m^u\}$ denotes m unlabeled examples. Usually, the number of unlabeled slices is much larger than labeled ones ($m \gg n$). Making better use of unlabeled data is a critical part of training a semisupervised segmentation network with better generalization ability on unseen data.

An overview of the proposed LCC-Net is demonstrated in Figure 1. We leverage the unlabeled data during supervised segmentation model learning and encourage segmentation consistency on all data under different perturbations with two unsupervised consistency losses. Our segmentation network is in encoder-decoder architecture. Specifically, the LCC-Net contains a shared encoder E and three independent decoders: the supervised decoder D_S, the dropout decoder D_D, and the noise decoder D_N. The encoder E and the decoder D_S constitute the segmentation network $f_S = D_S \circ E$. While the supervised decoder D_S is trained with the labeled data, the two auxiliary decoders are trained with both labeled data and unlabeled data.

We inject perturbations in both the feature space, i.e., the output of the feature encoder E and the input image space.

(i) For perturbations in the feature space, we use two perturbations: dropout perturbation P_D and noise perturbation P_N. The dropout decoder D_D and noise decoder D_N are used to decode the two perturbed versions of features, respectively. We enforce the consistency of predictions between the supervised decoder D_S and the auxiliary decoders D_D and D_N with unsupervised consistency losses. These two auxiliary decoders together with the encoder and feature perturbations constitute the two auxiliary networks $f_D = D_D \circ P_D \circ E$ and $f_N = D_N \circ P_N \circ E$. In the experiments, we use Gaussian noises for the noise perturbation P_N and 10%-40% spatial random dropout for the dropout perturbation P_D

(ii) For perturbations in the image space, we use a stronger perturbation P_C to achieve better model robustness. Specifically, we exploit an adapted version of the Cutmix [29], as illustrated in Figure 2. Given two input images, we first split the images into four blocks of equal size. Then, we randomly exchange one or two blocks on the corresponding positions between the two images. When the two input images are labeled, the corresponding operations are also applied to their label images

We apply the cutting and mixing perturbation on both the labeled data and unlabeled data as a data augmentation to the original data. In addition to the (augmented) unlabeled data, we also feed the perturbed labeled data to the auxiliary networks and enforce cross-model consistency.

3.2. Supervised Training on Few Labeled Data. The segmentation network $f_S = D_S \circ E$ is trained with the (augmented)

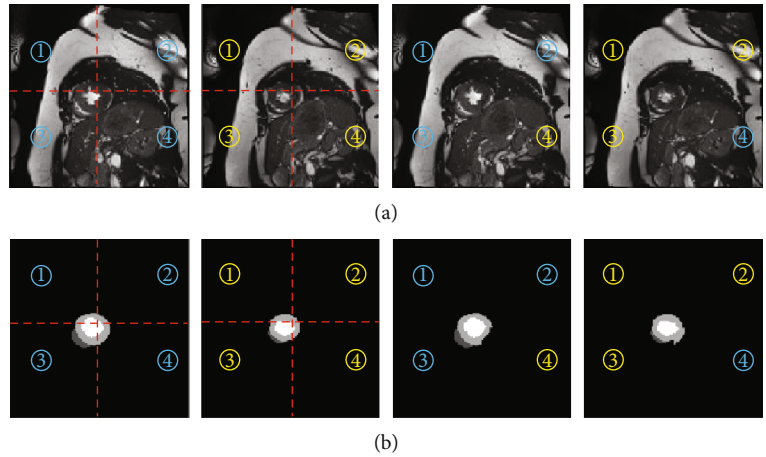


FIGURE 2: Illustration of the input space perturbation P_C used in our study. (a) The cardiac MR images. From left to right: original image A, original image B, the perturbed image A', and the perturbed image B'. (b) Their corresponding label images. We evenly split an image into four blocks and then randomly exchange one or two blocks on the corresponding positions. The same operations are applied on their ground truth label images, as in (b). Best view in color.

labeled data using a cross-entropy- (CE-) based supervised loss. We also denote the augmented labeled data as $\mathscr{D}_l \cup \mathscr{D}_l'$, where $\mathscr{D}_l'$ is generated by perturbing the images in $\mathscr{D}_l$ using cutting and mixing P_C.

$$\mathscr{L}_S = \frac{1}{\left|\mathscr{D}_l \cup \mathscr{D}_l'\right|} \sum_{(x_i,y_i)\in\mathscr{D}_l\cup\mathscr{D}_l'} l_{\mathrm{CE}}(x_i, y_i), \tag{1}$$

where l_{CE} denotes the cross-entropy loss. The input image x_i^l can be the original image and its perturbed version.

3.3. Unsupervised Cross-Consistency Training. As mentioned above, we enforce cross-model consistency between the predictions of the supervised decoder D_S and the auxiliary decoders D_D and D_N with an unsupervised consistency loss. We denote the augmented unlabeled data as $\mathscr{D}_u \cup \mathscr{D}_u'$, where $\mathscr{D}_u'$ is generated by perturbing the images in $\mathscr{D}_u$ using cutting and mixing P_C. The two auxiliary networks f_D and f_N take both the (augmented) unlabeled data $\mathscr{D}_u \cup \mathscr{D}_u'$ and the perturbed labeled data $\mathscr{D}_l'$. The two auxiliary networks are trained with the following loss.

$$\mathscr{L}_U = \frac{1}{\left|\mathscr{D}_u \cup \mathscr{D}_u' \cup \mathscr{D}_l'\right|} \sum_{(x_i,y_i)\in\mathscr{D}_u\cup\mathscr{D}_u'\cup\mathscr{D}_l'} \cdot \left[\mathbf{d}(f_S(x_i), f_D(x_i)) + \mathbf{d}(f_S(x_i), f_N(x_i))\right], \tag{2}$$

where the distance measure $\mathbf{d}$ is used to measure the consistency of the predictions by different models. In the experiments, we use mean squared error (MSE) as the distance measure.

3.4. The Overall Loss. By integrating the supervised loss and unsupervised loss, the loss of our LCC-Net reads

$$\mathscr{L} = \mathscr{L}_S + \lambda \mathscr{L}_U, \tag{3}$$

in which λ is the trade-off parameter. In the experiments, we choose an exp-schedule function as follows:

$$\lambda(\mathrm{epoch}) = \min\left(\lambda_{\max}, \lambda_{\max} \times e^{2\cdot(\mathrm{epoch}/\mathrm{stop}-1)}\right), \tag{4}$$

in which epoch as current training epoch, stop is the max number of epochs to stop increasing λ, and $\lambda_{\max}$ is an upbound of λ.

3.5. The Backbone of the LCC-Net. To avoid overfitting on the small labeled data, we introduce a lightweight segmentation U-Net (L-Unet) as our backbone network, which is demonstrated in Figure 3. The network is an encoder-decoder with skip-connections between the corresponding layers of the encoder and decoder. To lighten the U-Net, we upgrade the U-Net with lightweight convolutional modules. More precisely, we replace the standard convolutions in U-Net with the Ghost module [11], which involves much fewer parameters and computation costs. The Ghost module is shown in Figure 4. For a feature map $F \in \mathbb{R}^{a\times h\times w}$, in which a is the channel number, and $h \times w$ is the spatial size, we first compress F into $F' \in \mathbb{R}^{(b/s)\times h\times w}$ by using a standard 3×3 convolution, where b is the channel number of the final output, and s is the ratio. Then, we apply $s(=4)$ linear transformations, including one identity transform, on each channel of F' separately to generate s groups of new features, each of which contains b/s feature maps. The linear transformations are achieved with 3×3 convolutions. At last, we concatenate all the generate feature maps and obtain the final output $\widehat{F} \in \mathbb{R}^{b\times h\times w}$. Note that the computation costs of the linear transformations are much lower than standard convolutions.

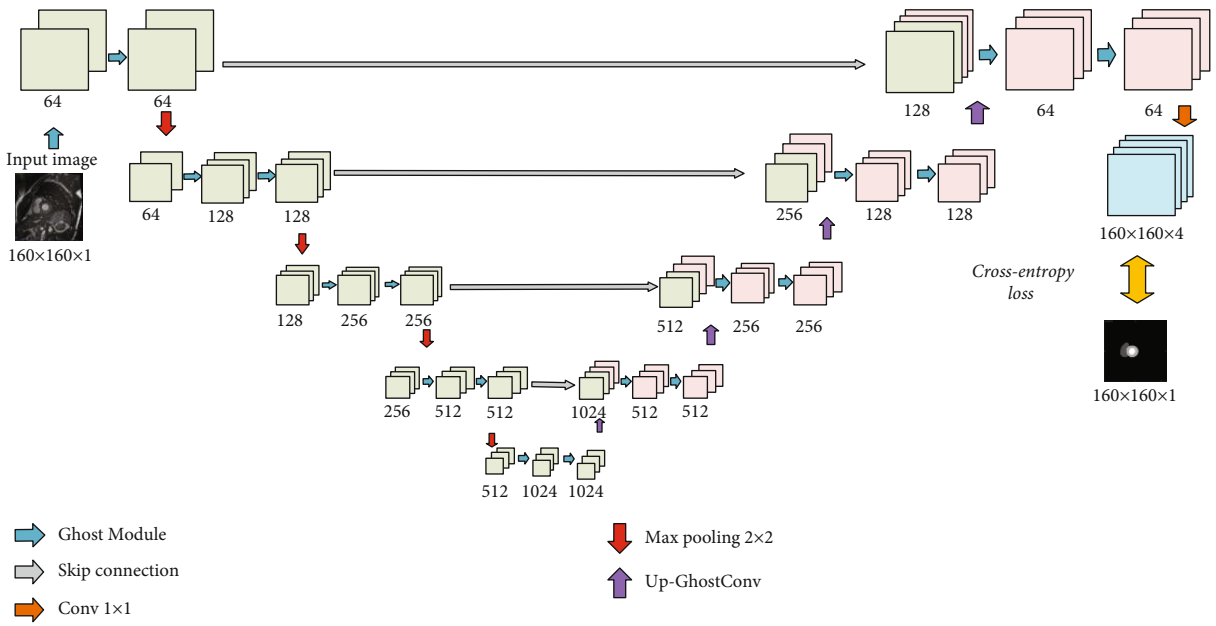


Figure 3: The backbone network of our proposed model, L-Unet. Instead of standard 2D convolutions, the L-Unet uses the Ghost module [11] as the basic building block.

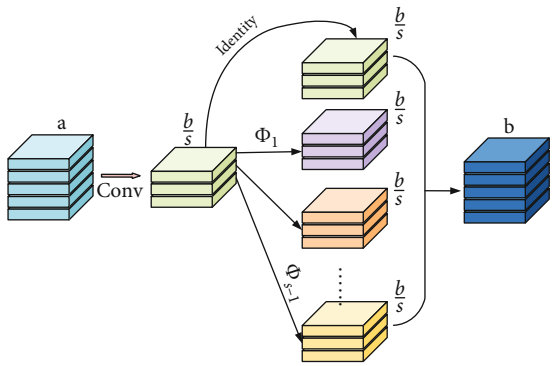


Figure 4: The architecture of Ghost module [11], which uses a series of cheap transformation operations to generate ghost feature maps, which results in significantly reduced computational complexity.

The model size of the L-Unet is only 8.7 M, which is four times less than that of the U-Net (35.5 M).

4. Experiments and Results

In this section, we conduct a series of experiments to evaluate the proposed LCC-Net's performance for semisupervised cardiac segmentation.

4.1. Dataset and Evaluation Measure. ACDC Dataset. We first utilize ACDC (Automated Cardiac Diagnosis Challenge) [12] dataset in our experiments, which belongs to a cardiac MR images segmentation challenge in MICCAI 2017. The ACDC dataset includes the short-axis cine-MRI of 150 subjects acquired from the University Hospital of Dijon using two MR scanners of different magnetic strengths. Left ventricle (LV), right ventricle (RV), and myocardium (MYO) were manually annotated by clinical experts on end-diastolic (ED) and end-systolic (ES) phase instants. The organizer of the ACDC splits the whole dataset into two subsets: (1) 100 subjects with available ground truth and (2) 50 subjects without ground truth for online testing.

We use the 100 labeled subjects (including 1902 image slices) for model evaluation. We randomly selected 20 subjects (containing 380 slices) as the testing set. The remaining 80 subjects are used as the union of the labeled data and unlabeled data. Specifically, we randomly select K subjects (2, 4, 6, and 10) for model training and the remaining 80- K subjects as the unlabeled data.

Evaluation Criteria. Our experiments utilize the Dice Coefficient (DICE) and Hausdorff distance (HD) as the evaluation criteria. Given the ground truth X and the prediction Y, DICE, which evaluates the region overlap of different segmentations, is defined as

$$\mathrm{DICE} = \frac{2 \bullet |X \cap Y|}{|X \cup Y|}. \tag{5}$$

The HD is defined as

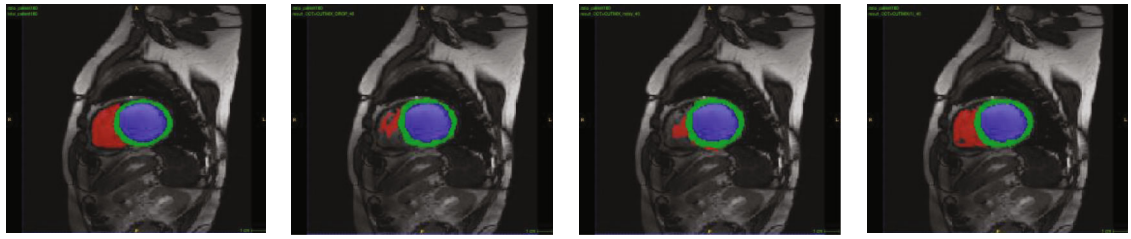
$$\mathrm{HD}(X, Y) = \max\left\{\max_{a \in X} E(a, Y), \max_{b \in Y} E(b, X)\right\}, \tag{6}$$

where $E(a, X)$ is the Euclid distance between a and X.

4.2. Implementation Details. We implemented our experiments on the framework of PyTorch [37] on one GTX 1080 GPU with 8 G memory. We used the adaptive moment estimation (Adam) optimizer with the learning rate of 5×10^{-4} initially, decreasing by 0.5 in epochs 200, 1000, 1500, 1800, and 2100. Moreover, the batch size was set as 4 because of the limitation of the GPU. The maximum epochs of iterations were set as 3000, and $\lambda_{\max}$ was set as 0.4. Data

Table 1: Comparative study of the proposed LCC-Net on the ACDC dataset. We randomly selected 2 subjects as the labeled data and the remaining 78 subjects as the unlabeled data. The models are tested on 20 unseen subjects. P_C denotes the perturbations in the feature space, P_N denotes the noise perturbation in the feature space, and P_D denotes the dropout perturbation in the feature space.

Method	Labeled data	DICE (%)				Hausdorff (mm)			
		LV	RV	MYO	Mean	LV	RV	MYO	Mean
U-Net (upbound)	80 subjects	93.2	85.8	88.9	89.3	2.2	4.8	2.8	3.3
U-Net (baseline)	2 subjects	76.1	24.7	69.1	56.6	9.2	17.0	11.4	12.5
L-Unet		69.9	36.7	60.2	55.6	10.7	19.2	12.7	14.2
LCC-Net w/o P_C		78.2	52.0	69.6	66.6	7.8	16.6	9.8	11.4
LCC-Net w/o P_N		78.0	54.8	70.7	67.8	5.8	11.3	7.8	8.3
LCC-Net w/o P_D		80.6	53.6	73.0	69.0	5.6	11.9	6.9	8.1
LCC-Net		82.0	58.1	73.0	71.0	4.3	13.9	6.3	8.2



(a) Ground truth (b) LCC-Net w/o P_N (c) LCC-Net w/o P_D (d) LCC-Net

Figure 5: Visualization of the performance of the ablated versions of our LCC-Net.

Table 2: The impact of the number of the labeled subjects. The results are tested on the ACDC dataset.

Labeled data	Method	LV (%)	RV (%)	MYO (%)	Mean (%)
2 subjects	U-Net (baseline)	76.1	24.7	69.1	56.6
	LCC-Net	82.0	58.1	73.0	71.0
4 subjects	U-Net (baseline)	79.6	51.7	71.1	67.5
	LCC-Net	85.0	69.3	76.9	77.1
6 subjects	U-Net (baseline)	83.4	62.3	74.0	73.2
	LCC-Net	85.0	74.5	79.3	79.6
10 subjects	U-Net (baseline)	82.1	70.1	76.6	76.3
	LCC-Net	87.4	77.0	82.8	82.4
Fully supervised (80 subjects)	U-Net (baseline)	93.2	85.8	88.9	89.3

augmentation, including affine transform, random rotation, and random intensity shift, was used. All the images were resized to 160×160, and the intensity range of each image was rescaled to [0, 1].

4.3. Segmentation Performance

4.3.1. Comparative Results of the LCC-Net. We first conduct a comparative study to identify the effectiveness of the critical components in the proposed model, including the backbone network L-Unet, the dropout decoder D_D, the noise decoder D_N, and the input space perturbation P_C. Specifically, we randomly select $K = 2$ subjects (40 slices) as the labeled data and the remaining 78 subjects as the unlabeled data, which are used for model training.

Table 1 summarizes the results of the comparative studies. The results of 7 network and data settings are reported: (1) the upbound, i.e., the U-Net trained with all the 80 labeled data; (2) the U-Net as the baseline, which is trained from scratch using the labeled data with standard data augmentations; (3) the L-Unet, which is also trained from scratch using the labeled data with standard data augmentations; (4) the LCC-Net w/o P_C, which is trained on both the labeled and unlabeled data without the input space perturbation P_C; (5) the LCC-Net w/o P_N, which is trained on both the labeled and unlabeled data without the noise decoder D_N; (6) the LCC-Net w/o P_D, which the LCC-Net without the dropout decoder D_D; (7) the full LCC-Net.

As shown in Table 1, when training with only two labeled subjects, the U-Net has a mean performance drop of 32.7% in DICE and 9.2 mm in Hausdorff than the U-Net trained with 80 subjects. Rather than using more labeled data, we exploit the unlabeled data, which is much easier to collect. As illustrated in Table 1, by exploiting unlabeled data, the LCC-

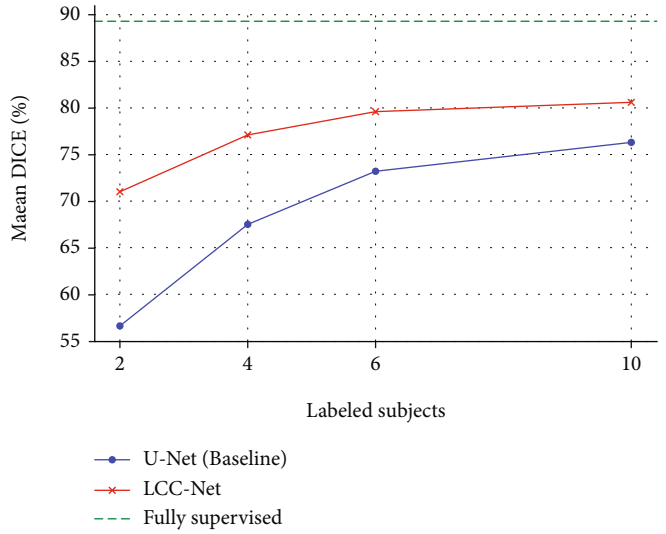


Figure 6: The impact of the number of the labeled subjects.

Table 3: The impact of the selection of the labeled subjects. The results are tested on ACDC dataset. Five samples are randomly selected, where each sample contains two labeled subjects as the labeled data for model training.

	DICE (%)							Hausdorff (mm)						
	(1)	(2)	(3)	(4)	(5)	Mean	Std	(1)	(2)	(3)	(4)	(5)	Mean	Std
LV	82.0	83.6	78.8	80.8	82.1	81.5	1.6	4.3	4.5	6.8	5.6	5.3	5.3	0.9
RV	58.1	64.0	51.0	52.9	54.6	56.1	4.6	13.9	9.1	12.4	11.9	10.1	11.5	1.7
MYO	73.0	74.1	73.6	73.9	75.5	74.0	0.8	6.3	5.8	6. 7	6.4	5.6	6.2	0.4
Mean	71.0	73.9	67.8	69.2	70.7	70.5	2.0	8.2	6.5	8.6	8.0	7.0	7.7	0. 8

Table 4: A comparison of the model complexity of the LCC-Net with different building blocks. The total inference time denotes the inference time on the whole testing set (402 images of 160×160).

LCC-Net with	Model size		FLOPs		Total
	Training	Testing	Training	Testing	Inference time
Standard convolution	81.5 M	35.5 M	329.8 G	102.8 G	20.8 s
Ghost module ($s = 4$)	25.2 M	8.7 M	91.0 G	26.2 G	17.3 s

Net outperforms the U-Net (baseline) by a large margin, i.e., 14.4% in the mean DICE and 4.3 mm in the mean Hausdorff over the three regions. LCC-Net without using the noise perturbation P_N and noise decoder D_N obtains a performance gain of 12.2% in DICE and 5.9 mm in Hausdorff over the L-Unet; LCC-Net without using the dropout perturbation P_D and dropout decoder D_D obtains a performance gain of 13.4% in DICE and 6.1 mm in Hausdorff over the L-Unet. Compared to the full LCC-Net, the LCC-Net without using the input space perturbation P_C shows a mean performance drop of 4.4% in DICE and 3.2 mm in Hausdorff, which indicates the effectiveness of the input space perturbation P_C. However, with only two labeled subjects for model training, the semisupervised model's performance is still significantly lower than the fully supervised U-Net. Figure 5 provides a visual comparison of the LCC-Net without P_N, LCC-Net without P_D, and our LCC-Net. Visually, the LCC-Net shows significantly better results than the other two methods.

4.3.2. The Impact of the Number of the Labeled Subjects. Since our method is a semisupervised method, it is crucial to identify the impact of the size of the labeled training dataset. To this end, we trained our model with different choices of K, i.e., 2, 4, 6, and 10 subjects. Table 2 summarizes the experimental results. The results with U-Net under different settings, including the fully supervised setting (80 labeled subjects), are also reported. As can be expected, with increasingly more labeled data for model training, the performance becomes much higher. With the different choices of K, our semisupervised model consistently outperforms the U-Net. Remarkably, using only four labeled subjects, our model outperforms the U-Net trained on ten labeled subjects. Using ten labeled subjects for training, the LCC-Net achieves a mean performance of 82.4%, which is 6.1% higher in mean DICE than the U-Net. Figure 6 demonstrates a further comparison of the proposed model and the U-Net, which shows the effectiveness of our model.

4.3.3. The Impact of the Selection of the Labeled Subjects. To identify the robustness of the proposed model over the different selections of the label data. To this end, we randomly selected five samples and calculated the mean performance and the standard variance. Here, each sample contains two subjects as the labeled data. The results are reported in Table 3. Although each sample size is very small (2 subjects), our model shows relatively stable performance.

4.4. Model Complexity. Model complexity is typically measured by the number of trainable network parameters (i.e., model size) and the floating-point operations (FLOPs). The model complexity of our model is summarized in Table 4. Our model obtained significantly reduced model size and FLOPs at both the training stage and testing stage by replacing standard convolutions with the lightweight module. Therefore, our model requires less computation cost for each training step and inference step, resulting in higher computational efficiency. The inference time at the testing stage is a critical measure in practical usage. As shown in Table 4, with reduced FLOPs, the proposed LCC-Net involves a shorter inference time than the LCC-Net using standard convolutions.

5. Conclusion

In this paper, we presented a lightweight cross-consistent network for semisupervised cardiac MR image segmentation. We leveraged the unlabeled data during supervised segmentation model learning and encourage segmentation consistency on all data under different perturbations with two unsupervised consistency losses. To achieve a lightweight model, we replaced the standard convolutions with a lightweight module. Extensive comparison experiments with a public dataset demonstrated that our architecture achieved promising performance with only two labeled subjects.

Despite the improved results, there are still more applicable perturbations in semisupervised segmentation. Thus, exploring more efficient perturbations is a significant work in the future.

Data Availability

The data used in our experiments are available at https://www.creatis.insa-lyon.fr/Challenge/acdc/.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the NSFC under Grant 11771160, the Fund of HQU (ZQN-PY411), and by STPF under Grant 2019H0016.

References

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234–241, Springer, 2015.

[3] C. Chen, C. Qin, H. Qiu et al., "Deep learning for cardiac image segmentation: a review," *Frontiers in Cardiovascular Medicine*, vol. 7, 2020.

[4] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on Challenges in Representation Learning*, vol. 3, no. 2, 2013ICML, 2013.

[5] Y. Grandvalet and Y. Bengio, *Semi-Supervised Learning by Entropy Minimization*, Advances in Neural Information Processing Systems, 2005.

[6] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12674–12684, 2020.

[7] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning," in *IEEE Transactions on Neural Networks*, vol. 20no. 3, p. 542, 2009.

[8] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, *Semi-Supervised Learning with Ladder Networks*, Advances in Neural Information Processing Systems, 2015.

[9] A. Tarvainen and H. Valpola, *Mean Teachers Are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results*, Advances in Neural Information Processing Systems, 2017.

[10] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," 2016, https://arxiv.org/abs/1610.02242.

[11] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: more features from cheap operations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1580–1589, 2020.

[12] O. Bernard, A. Lalande, C. Zotti et al., "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[13] N. Painchaud, Y. Skandarani, T. Judge, O. Bernard, A. Lalande, and P.-M. Jodoin, "Cardiac MRI segmentation with strong anatomical guarantees," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 632–640, Springer, 2019.

[14] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proceedings of the International Conference on Learning Representations*, 2014.

[15] M. Khened, V. Alex, and G. Krishnamurthi, "Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 140–151, Springer, 2017.

[16] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densenet: densely connected convolutional networks," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 30, pp. 82–84, 2017.

[17] X. Yang, C. Bian, L. Yu, D. Ni, and P.-A. Heng, "Class-Balanced Deep Neural Network for Automatic Ventricular

Structure Segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 152–160, Springer, 2017.

[18] G. Simantiris and G. Tziritas, "Cardiac mri segmentation with a dilated cnn incorporating domain-specific constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1235–1243, 2020.

[19] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic Cardiac Disease Assessment on Cine-MRI via Time-Series Segmentation and Domain Specific Features," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 120–129, Springer, 2017.

[20] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2016.

[21] B. Liu, Z. Wu, H. Hu, and S. Lin, "Deep metric transfer for label propagation with limited annotated data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.

[22] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 605–613, Springer, 2019.

[23] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[24] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," 2018, https://arxiv.org/abs/1802.07934.

[25] D. Nie, Y. Gao, L. Wang, and D. Shen, "Asdnet: Attention based semi-supervised deep networks for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 370–378, Springer, 2018.

[26] T. Miyato, S.-I. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1979–1993, 2019.

[27] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, and M. de Bruijne, "Semi-supervised medical image segmentation via learning consistency under transformations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 810–818, Springer, 2019.

[28] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: beyond empirical risk minimization," 2017, https://arxiv.org/abs/1710.09412.

[29] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6023–6032, 2019.

[30] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.

[31] G. French, T. Aila, S. Laine, M. Mackiewicz, and G. Finlayson, "Consistency regularization and cutmix for semi-supervised semantic segmentation," 2019, https://arxiv.org/abs/1906.01916.

[32] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, 2017.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[34] T. Zhang, G.-J. Qi, B. Xiao, and J. Wang, "Interleaved group convolutions," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4373–4382, 2017.

[35] A. G. Howard, M. Zhu, B. Chen et al., "Mobilenets: efficient convolutional neural networks for mobile vision applications," 2017, https://arxiv.org/abs/1704.04861.

[36] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: an extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6848–6856, 2018.

[37] A. Paszke, S. Gross, S. Chintala et al., "Automatic differentiation in pytorch," in *31st Conference on Neural Information Processing Systems (NIPS 2017)- Workshop*, 2017.