Image segmentation model of pneumothorax disease based on Unet

~~Yiyang Niu,~~
~~China Jiliang University,~~
~~Hangzhou , China,~~
~~dc43yiyang@foxmail.com,~~

Abstract—It is very meaningful to use deep learning methods to analyze medical images to assist diagnosis and reduce the misdiagnosis rate~~probability of misdiagnosis~~. In this paper, we propose a new method for predicting pneumothorax disease and segmenting the range of pneumothorax in medical images. We use~~d~~ the chest radiological image data set provided by the Society of Medical Imaging Informatics (SIIM) to build a network structure based on Unet and SE_ResNeXt for feature extraction. Due to the difficulty of obtaining medical image data, the sample size is small, ~~and~~To address this, ~~the~~ the network structure we designed can combine the information of the bottom and high layers of the image, with a small amount of parameters, and can effectively avoid over-fitting when the number of samples is small. In order to comprehensively evaluate the performance of ~~our~~ the proposed method, we choose the classic network architecture as a comparison, and use the same evaluation indicators and data sets for comparison. Experiments show that compared with DenseNet121, ResNet50, RestNet101 and Vgg16, our Seresnext50 model can ~~mine~~ extract features of different dimensions more effectively, and has ~~stronger~~ better image recognition and image segmentation capabilities.

Index Terms—Deep learning; image segmentation; pneumothorax disease; image enhancement

I. INTRODUCTION

In the medical field, in order to meet the needs of disease diagnosis and treatment plan formulation, patients are often scanned to determine the condition of their internal organs~~it is often necessary to scan the patient to determine the condition of each internal organ~~. At present, th~~e~~is process is mainly done directly by ~~the~~ doctors. Although the judgment accuracy of experienced doctors is high, they still need to consume a lot of valuable human resources. Therefore, it is very meaningful to analyze medical images with the help of deep learning methods to assist diagnosis and reduce the misdiagnosis rate ~~probability of misdiagnosis~~.

This paper uses the chest radiological image data provided by the Society of Medical Imaging Informatics (SIIM) to build an image segmentation model. The model can judge whether there is pneumothorax disease and identify the area where pneumothorax exists.

Pneumothorax is a state where air enters the pleural cavity, resulting in pneumothorax. Most of them are due to lung disease or external force, rupture of the lung tissue and visceral pleura, or the rupture of fine emphysema near the surface of the lung, causing the air in the lungs and bronchi to escape into the pleural cavity. This disease is one of the pulmonary emergencies, severe cases can be life-threatening, but timely treatment can be cured. Therefore, timely identification of pneumothorax disease is of great significance for treatment.

This paper aims to use the chest radiological image data provided by the Society of Medical Imaging Informatics (SIIM) to predict pneumothorax disease and segment the image area where the disease exists. In this paper, we propose a new method for predicting pneumothorax disease and segmenting the range of pneumothorax in medical images. First, we perform~~ed~~ data cleaning and data enhancement operations on the data, and then we built a network structure based on Unet and SE_ResNeXt for feature extraction. Experiments show that compared with DenseNet121, ResNet50, RestNet101 and Vgg16, our Seresnext50 model can ~~mine~~ extract features of different dimensions more effectively, and has stronger recognition and segmentation capabilities.

Figure 1. Original image, mask, and composite image

In summary, the contributions of this paper are as follows:

- We propose a new method for predicting pneumothorax disease and segmenting the range of pneumothorax in medical images. ~~This~~ The method

performs well in identifying the pneumothorax and segmenting the image area where the pneumothorax exists.

● We analyzed the characteristics of the original data set, and obtained the relationship between gender and the distribution of pneumothorax disease, which provided us with guiding suggestions for future research.

● Before building the deep learning model, we first performed operations such as data cleaning and image enhancement on the original data set to obtain a higher quality data set. When training the model, we used methods such as optimizing the learning rate and adding regularization to speed up the model's convergence speed and obtain the optimal model.

II. Related Work

The deep neural network has multiple non-linear mapping function conversions, which can adapt to very complex functions. The application of deep neural network is also very extensive [4]. After the development of deep neural networks for a period of time, due to the limitation of the size of the data set and the depth of the network, the design of the network structure has encountered a bottleneck. [5] Krizhevsky et al. proposed a very ingenious solution, training a large network with 8 layers and millions of parameters on the ImageNet dataset to guide people to build larger and deeper network structures. Different from other image segmentation tasks, medical image segmentation has the characteristics of small sample size and uneven sample distribution on the data set. Therefore, the sliding sampling method can solve this problem [2].

In this paper, we follow the idea and method of "fully convolutional network" [3], we and use a network architecture that combines encoder and decoder. On this basis, we extend and improve our model to have stronger segmentation ability and expand and improve on this basis to make our model stronger. The segmentation ability. Due to the difficulty in collecting medical image data and the small sample size Because the collection of medical image data is difficult and the sample size is small, we performed data enhancement operations on the data before model training [1], thereby increasing the distribution of data and improving the generalization ability of the model.

A very important part of the image segmentation model is the feature extraction of the image. Some specific network structures have strong feature extraction capabilities. For example, using multiple branches to extract high-level features with different levels of abstraction is can be very effective, and it can enrich the expressiveness expressive power of the network. Among them, The the Inception models [6, 7, 8] are successful multi-branch architectures where each branch is carefully customized. The ResNets [9] network structure also has the ability to extract features from multiple branches. [10] The network structure of packet convolution also provides a constructive direction for enhancing the feature extraction capability of network the network to enhance the ability of feature extraction. In the encoding process of the image segmentation network, it is actually to compress the image features. It is very necessary to use adopt a specific method to reduce the redundancy of network parameters [12, 13, 14, 15].

III. Data Preprocessing

The data set in this paper comes from the chest radiological image data provided by the Society for Medical Imaging Informatics (SIIM). When using the deep learning method to train a model, the larger the scale and the higher the quality of the data, the better the generalization ability of the model, and That is, the data greatly affects the upper limit of model learning. However, in actual engineering, it is difficult for the collected data to cover all the scenes, such as the lighting illumination conditions of the image. The pictures taken in the same scene may be very different due to different lighting, so we should add lighting when training the model. Aspects of data enhancement. On the other hand, As for data enhancement, even if we have a lot of data, we should also perform data enhancement perform data enhancement, which helps to increase the amount of data in the relevant data set, prevent the model from learning unwanted models, and avoid overfitting.

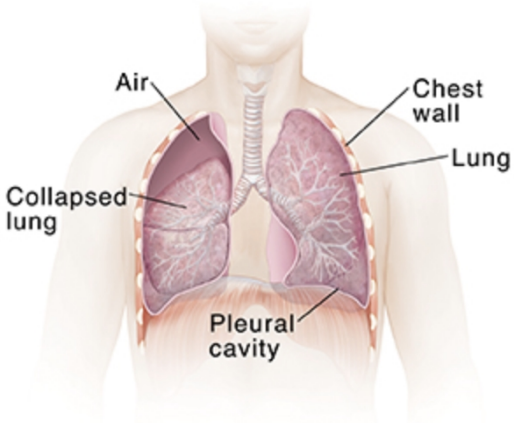

Figure 2. Human structure diagram

.

There are two specific methods of us to perform ing data enhancement. One is to perform all conversions in advance, which is essentially to enhance increase the size of the data set. This method, which is called offline enhancement. It is more suitable for smaller data sets, and will eventually increase the amount of data by a certain multiple, which depends on the

number of converted pictures. The other is mini-batch conversion when sending data to the deep learning model. This method is called online enhancement, which is more suitable for large data sets.
Before training the model ~~training~~, ~~in this paper,~~ we first enhance~~s~~ the data set. The following data enhancement methods are mainly used:

(1) Flip and rotate
Both the flip and rotation ~~Flipping and rotating both~~ transform the original image pixels in the positional space~~position space~~. The image flip is the mirror operation of the original image. Mirror operation is a common operation~~often used~~ in data enhancement and plays a very important role. It includes~~Including~~ horizontal mirror flipping, vertical mirroring flipping and origin mirroring flipping. In specific use, we select the corresponding flip operation in combination with the data form. For example, the data set is about car image ~~data~~, and the training set and test set are all pictures taken normally. At this time, only ~~use~~ h~~H~~orizontal mirror operation is used, and if we add vertical or origin mirror flip, it will interfere with the original image.
The angle rotation operation is opposite to the image mirroring. It mainly performs arbitrary angle transformation along the center of the screen. The transformation is ~~realized~~ achieved by multiplying the original image ~~and~~ by ~~the~~ an affine transformation matrix. In order to realize the center rotation of the image, ~~in addition to~~besides knowing the rotation angle, the ~~amount of~~ translation must be calculated to make the effect of the affine transformation equivalent to the center of the picture on the rotation axis.

(2) Color jitter
The pictures obtained in the actual project have background interference. In order to eliminate the difference of images in different backgrounds, some color dithering operations are usually performed to expand the data set. Color dithering is mainly to enhance the color of the image, ~~mainly to~~ adjust the brightness, saturation and contrast of the image.

(3) Tailoring
The cropping operation mainly considers the width and height disturbance of the original image. There are two ways to crop. One is to ~~directly~~ crop the large-size image directly according to the size that needs to be sent to the network. For example, the resolution size of the original image is 256x256. The pixel size of the image that needs to be input is 224x224, so you can crop a 224x224 pixel image directly from the original image~~so you can directly crop an image of 224x224 pixel size directly on the original image~~, so that a picture can be expanded to 32x32 pictures; the other is to adjust the size of image to the network required through an interpolation algorithm The size required. Since the data size is usually different in the data set, the latter is often used.~~usually used more~~.

In order to obtain a high-quality data set, and thus speed~~thereby speeding~~ up the model fitting, we clean~~ed~~ the data set. Data cleaning is an indispensable part of the entire data analysis process, because the quality of the data set is directly related to the effect of the model ~~the model effect~~ and the final conclusion. Data cleaning can be divided into ~~the~~ following steps:

(1) Missing value cleaning
Missing values are the most common data problem, and there are many ways to deal with missing values. We first determine the range of missing values, calculate the proportion of missing values for each field, and then formulate strategies according to the missing proportion and field importance. Then we remove the unnecessary fields. ~~When~~ As for filling in missing content, we choose to fill in missing values with the ~~mean~~ average of the same indicator.

(2) Format content cleaning
If the data is collected manually or filled in by users, the potential for problems with format and content is high~~there is a high possibility that there are some problems in the format and content~~. Simply put, there are the following types of format and content problems:1. The display formats of time, date, numeric value, and half-width are inconsistent. This kind of problem is usually related to the input, and may also be encountered when integrating data from multiple sources. We process it into a consistent format. 2. Some characters should not appear in the content~~There are characters that shouldn't exist in the content~~. In this case, we use a semi-automatic check and a semi-manual method to find possible problems and remove unnecessary characters.
The format content issue is a more detailed issue, ~~but~~ and if it is not handled properly, it will seriously affect the subsequent model training.

(3) Logic error cleaning
This part of the work is to remove some data that can be found directly ~~found using~~ through simple logical reasoning to prevent deviation of the analysis results. First, we delete duplicate values in the data set, and then we delete unreasonable data in the data set. If there is conflicting content in the data set, we will ~~also~~ delete it as well.

Since in the original data set, pictures with pneumothorax only accounted for about 28.37%, the positive and negative samples of the image data are not balanced. Therefore, we implemented a sliding sampling rate strategy on the data during training.

Specifically, in different training stages, the data is sampled differently to make the positive and negative samples reach a certain ratio.

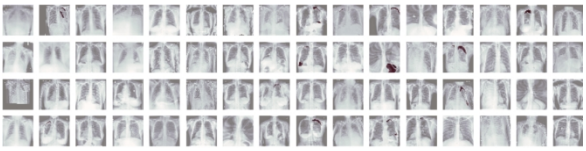

Figure 3. data set

IV. IMAGE SEGMENTATION ALGORITHM

Image segmentation refers to the division of an image into several disjoint areas according to its grayscale, color, spatial texture, geometric shape and other characteristics, so that the features in the same area show consistency or similarity, and the features in different areas show obvious differences~~these features show consistency or similarity in the same area, but in different areas Shows~~ shows ~~a clear difference between~~. Simply put, it is to separate the target from the background in an image. For grayscale images, the pixels within the region generally have grayscale similarity, and generally have grayscale discontinuities on the boundary of the region. Regarding image segmentation technology, due to the importance and difficulty of the problem itself, the image segmentation problem has attracted a lot of researchers since the 1970s ~~to~~ who make great efforts and contribute a lot of research results and methods.

A. Traditional image segmentation method

Before the deep learning method was proposed, ~~we~~ researchers used the knowledge of digital image processing, topology, mathematics, etc. to realize image segmentation. Although with the increase of computing power and the continuous development of deep learning, some traditional segmentation methods can no longer be compared with segmentation methods based on deep learning in effect, ~~but~~ some genius ideas are still worth learning. Traditional image segmentation algorithms mainly include the following:

(1) Segmentation method based on threshold

The basic idea of the threshold method is to calculate one or more gray thresholds based on the gray features of the image, and compare the gray value of each pixel in the image with the threshold, and finally sort the pixels into appropriate categories ~~based on~~ the comparison results~~.~~. Therefore, the most critical step of this method is to solve the ~~best~~ optimal gray threshold according to a specified function. The threshold method is particularly suitable for images with different gray ranges of target and background.~~in which the target and background occupy different gray-scale ranges.~~

If the image has only two categories: target and background, then only one threshold value needs to be selected for segmentation. This method is called single threshold segmentation. However,~~But~~ if there are multiple targets in the image that need to be extracted, the result of a single threshold segmentation will be wrong ~~a single threshold segmentation will be wrong~~. In this case, multiple thresholds need to be selected to separate each target. This segmentation method is correspondingly called multi-threshold segmentation.

The advantage of threshold segmentation method is simple calculation and high efficiency. However, this method only needs to consider the grayscale characteristics of ~~the characteristics of the gray value of~~ the pixel itself, and generally does not consider the spatial characteristics, so it is more sensitive to noise and not robust.

(2) Region-based image segmentation method

The region-based segmentation method is a segmentation technique based on the direct search of regions~~directly finding the region~~. The region-based image segmentation method has two basic forms: one is region growth, which starts from a single pixel and gradually merging~~s~~ to form the required segmentation region; the other~~One~~ is to start from the global~~overall~~ ~~situation~~ and gradually cut to the required segmentation area.

The first form of region growth is to start from a set of seed pixels representing different growth regions, and then merges the eligible pixels in the seed pixel neighborhood into the growth region represented by the seed pixels, ~~and use~~using the newly added pixels as the new seed pixels. The merging process of seed pixels continues until no new pixels that meet the conditions are found. The key to this method is to select a suitable initial seed pixel and a reasonable growth criterion.

The region growing algorithm needs to solve three problems: 1. Select or determine a set of seed pixels that can correctly represent the desired region. 2. Determine the criteria for including adjacent pixels in the growth process. 3. Specify the conditions or rules to stop the growth process.

The second form of region splitting and merging can be ~~said to be~~consider as the inverse process of region growth. Starting from the entire image, it continuously splits to obtain each subregion, and then merges the foreground regions to obtain the foreground target that needs to be segmented, and then achieve the extraction of the target. The advantage of the region splitting and merging algorithm is that it has a good effect on complex image segmentation, but this algorithm is more

complex~~complicated~~ and computationally expensive, and the split may destroy the boundary of the region.

In practical applications, the region growing algorithm and the region splitting and merging algorithm are usually used in combination.

(3) Watershed algorithm

The watershed segmentation method is a mathematical morphology segmentation method based on topological theory. Its basic idea is to regard an image as a geodetic topological landform. The gray value of each pixel in the image represents the altitude of the point. Each local minimum and its affected area ~~is~~are called a catchment basin, and the boundary of the catchment basin forms a watershed. The concept and formation of watershed can be illustrated by simulating the immersion process. On the surface of each local minimum, pierce a small hole, and then slowly immerse the entire model in the water. As the immersion deepens, the influence area of each local minimum gradually expands outward~~the area of influence of each local minimum slowly expands outwards~~. A dam is built at the confluence of the basins to form a watershed.

The ~~W~~watershed has a good response to weak edges. Noise in the image and subtle gray-scale changes on the surface of the object may cause excessive segmentation, but it can ~~also~~ still ensure ~~that~~ a closed continuous edge~~ is obtained~~. At the same time, the closed collection basin obtained by the watershed algorithm also provides information for analyzing the regional characteristics of the image.

(4) Segmentation method based on edge detection

The image segmentation algorithm based on edge detection solves the segmentation problem by detecting edges containing different regions. Usually the gray values of pixels on the boundaries of different regions vary greatly~~change drastically~~. If the image is transformed from the spatial domain to the frequency domain through Fourier transform, the edge corresponds to the high frequency part. This is a very simple edge detection algorithm.

Edge detection technology can usually be divided into serial edge detection and parallel edge detection according to the processing technology. The idea of serial edge detection is to determine whether the current pixel is a point on the detected edge, ~~it~~ which depends on the verification result of the previous pixel. The idea of parallel edge detection is that whether a pixel belongs to a point on the detection edge depends on the pixel currently being detected and some ~~neighboring~~ pixels adjacent to ~~with~~ the pixel.

The advantages of the segmentation method based on edge detection are accurate edge location and fast speed. However, this method cannot guarantee the continuity and closure of the edges. There are a large number of broken edges in the high-detail area, ~~and it~~which is difficult to form a large area, ~~but~~ it is not suitable to divide the high-detail area into small fragments.

(5) Image segmentation method based on wavelet analysis and wavelet transform

Wavelet transform is a widely used mathematical tool. It has ~~localized properties~~locality in the time domain and frequency domain, and can unify the time domain and frequency domain to study signals. In addition, wavelet transform has multi-scale characteristics ~~and~~ which can analyze signals on different scales, so it has good applications in image segmentation.

Binary wavelet transform has the ability to detect local mutations of binary functions, ~~so it~~and can be used as an image edge detection tool. The edge of the image appears at the local gray level discontinuity of the image, which corresponds to the modulus maximum point of the binary wavelet transform. The edge of the image can be determined by detecting the modulus maximum point of the wavelet transform. The wavelet transform is located ~~on~~ at each scale, and the wavelet transform ~~on~~ at each scale can provide certain edge information, so multi-scale edge detection can be performed to obtain ~~a~~ more ideal ~~The edge~~edges of the image.

(6) Segmentation method based on active contour model

Active contour~~s~~ is an important method of image segmentation. It has a unified open description form and provides an ideal framework for the research and innovation of image segmentation technology. When ~~realizing~~ constructing the active contour model, the constraint force, initial contour and scope can be ~~flexibly~~ selected flexibly to obtain a better segmentation effect.

~~This~~ The method is a kind of method that uses curve evolution to detect targets in a given image, based on which the accurate edge information can be obtained. The basic idea is to ~~first~~ define the initial curve C first, and then obtain the energy function according to the image data~~,~~. By minimizing the energy function, the curve changes gradually approach the target edge, and finally find the target edge~~and induce the curve change by minimizing the energy function to gradually approach the target edge, and finally find the target edge~~. The edge curve obtained by this dynamic approximation method has the advantages of being closed and

smooth.

Traditional active contour models are roughly divided into parametric active contour models and geometric active contour models. The parametric active contour model expresses the deformation of a curve or surface in a parametric form. Its characteristic is to place the initial curve near the target area without without the need for artificial evolutionartificially setting the evolution of the curve to shrink or expand. The advantage of this method is that it can directly interact with the model, and the model expression is compact, and the realization speed is fast. Its The disadvantage is that it is difficult to deal with changes in the model topology. The geometric active contour method using level set just solves this problem.

(7) Image segmentation based on genetic algorithm

Genetic Algorithms (GA) was proposed by the American professor Holland in 1973. It is a randomized search algorithm that draws on the natural selection and natural genetic mechanisms of the biological world. It is the application of bionics in the field of mathematics. The basic idea is to simulate the evolution process of a biological population controlled by some gene strings, and apply the principle of this process to the search algorithm to improve the speed and quality of finding the optimal solution. The entire search process is to iterates from one set of solutions to another set of solutions. The method of processing multiple individuals in the group simultaneouslyat the same time, which reduces the possibility of falling into a local optimal solution and is easy to be parallelized. The search process uses probabilistic transition rules to guide the search direction, instead of using deterministic search rules, and does not have any special requirements for the search space. It only uses adaptive information without other auxiliary information such as derivatives, and has a wide range of adaptability adaptations.

Genetic algorithm is good at global search, with insufficient local search abilitybut the local search ability is insufficient, so genetic algorithm is often used in combination with other algorithms. The application of genetic algorithm to in image processing mainly considers that genetic algorithm has the ability to of fast and ramdom search search quickly and randomly regardless of the problem domain. The search starts from the group, has potential parallelism, can compare multiple individuals at the same time, and can effectively accelerate the speed of image processing. But the genetic algorithm also has its shortcomings: the design of the evaluation function used in the search and the selection of the initial population are dependent to some extenthave certain dependence.

B. Image feature extraction

The process of image segmentation is inseparable from image feature extraction. In the field of image feature extraction, VggNet and ResNet are two very influential methods.

VggNet is a deep convolutional neural network developed by researchers from the Computer Vision Group of Oxford University and Google DeepMind. It explores the relationship between the depth of the convolutional neural network and its performance. By repeatedly stacking 3*3 small convolution layerskernels and 2*2 maximum pooling layers, it successfully builds a volume of 16 to 19 layers successfully. Product neural network. VggNet won the runner-up of the ILSVRC 2014 competition and the champion of the positioning project, with an error rate of 7.5% on the top5. So far, VggNet is still a common method of image feature extractionstill often used to extract image features.

In the VggNet network structure, since the parameters are mainly concentrated in the last three convolutional layers, the depthdeepening of the network will not bring about the problem of parameter explosion. The VggNet network structure replaces the receptive fields of multiple small-kernelcore convolutional layers with a large-kernelcore convolutional layer, so the amount of parameters is far less than that of the large-kernelcore convolutional layer, and the nonlinear operation is more than the latter, which makes its learning ability sStronger. However, the VggNet network structure also has shortcomings. The last three fully connected layers in its network have more parameters, which causes it to occupy more memory.

For a long time, the deep learning model has a crucial probleman important problem. When the deep learning network is stacked to a certain depth, the gradient disappears, resulting in worse error increasing effectwhich causes the error increase effect to become worse, and the gradient cannot be fed back to the front when propagating backwardsduring backward propagation. The network layer makes it difficult to update the parameters of the front network layer, and the training effect becomes worse. The emergence of ResNet solved solves this problem and becames an important turning point in the development of deep learning.

ResNet was is proposed by four Chinese including Kaiming He from Microsoft Research. They successfully trained a 152-layer neural network through their own ResNet Unit unit and won the championship in the ILSVRC2015 competition. ResNet is the most popular and widely used neural network in the field of semantic segmentation. Its

core idea is to introduce identity mapping in the network, allowing the original input information to be transferred directlyto be directly transmitted to the subsequent network layer, and only the residual of the last network output can be learned during the learning process, so ResNet is also known ascalled the residual network.

The advantage of ResNet is that it introduces a new network structure, namely the residual learning module, forming a new network structure, which can deepen the network as much as possible. At the same time, it enables the forward propagation algorithm and the back propagation algorithm to proceed smoothly, making the network structure simpler.

In this paper, the SE-ResNeXt network structure we used is an improved structure based on Resnet. SE-ResNeXt is the encoder part of our proposed pneumothorax disease image segmentation algorithm, which combines the characteristics of the SeNet network structure and the ResNeXt network structure.

The characteristic of SeNet is that it can automatically obtain the importance of each feature channel through learning, and then promote the useful features based on this importance according to this importance to promote useful features and suppress features that are useless not useful for the current task. In the SeNet network structure, Squeeze and Excitation are two very critical operations. The first is the Squeeze operation. We perform feature compression along the spatial dimension, turning each two-dimensional feature channel into a real number. The real number has a global receiver domain to some extentThis real number is to some extent It has a global receptive field, and the output dimension matches the input characteristic channel number. It shows that the layer close to the input can also obtain the global receptive field, which is very useful in many tasks. Then there is the Excitation operation, which is a mechanism similar to the gate in a recurrent neural network. The parameter w is used to generate weights for each feature channel, where the parameter w is learned to explicitly model the correlation between the feature channels. Finally, the operation of Reweight, we regard the weight of the output after the Excitation operation as the importance of each feature channel after feature selection, and then weight the previous feature channel by channel through multiplication to complete the channel dimension Recalibration of original features. The above are the compositions is the composition of the SE module, which can be embedded in the ResNet model containing Residual Connection.

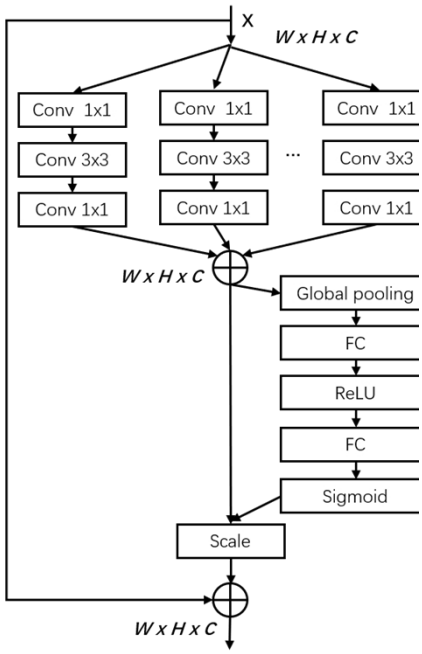

Figure 4. SE-ResNext block

The advantage of the ResNeXt network structure is to improve the performance of the network without increasing the complexity of the parameters, while also reducing the number of hyperparameters. ResNeXt combines the advantages of Vgg and Inception, and proposes a simple architecture: a strategy that repeats the same network layer and a split-transform-merge strategy in a simple, extensible mannerthe strategy of repeating the same network layer is adopted, and the split-transform-merge strategy is continued in a simple and scalable way. The building blocks of the entire network are the same. There is no need to adjust the hyperparameters of each building block in each stage. Only one building blocks with of the same structure can be stacked repeatedly to form the an entire network.

The SE-ResNeXt we used in this paper is to follow the SeNet network structure, and stack the SE module and Residual Connection to form the final building block of the SE_ResNeXt network. Repeated stacking can form the entire ResNeXt.

B. Image segmentation algorithm based on Unet

The segmentation of medical images has the following characteristics:

(1) The image semantics is are relatively simple and the structure is relatively fixed. For brain imaging, brain CT and brain MRI are used. Chest X-ray

effects only use chest X-ray CT. They are all images of a fixed organ~~, not~~ instead of the whole body. Since the structure of the organ itself is fixed and the semantic information is not particularly rich, both the high-level semantic information and the low-level features are very important.

(2) The amount of data is small. Medical image data acquisition is a difficult problem~~The data acquisition of medical imaging is difficult~~. Therefore, the model we design should not be too ~~large~~big. Too many parameters can easily lead to overfitting.

(3) It has the characteristics of multi-modality. Compared with natural images, a special aspect~~one of the special aspects~~ of medical images is that medical images have multiple modalities, which requires the use of a network with a specific structure to extract features of different modalities.

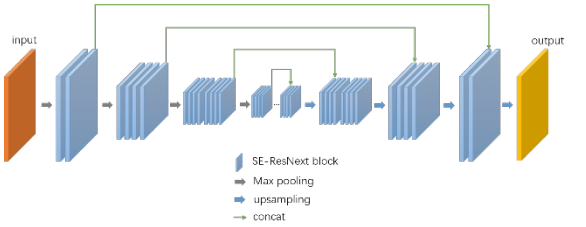

Figure 5. UNet

Based on the above characteristics, the Unet network structure shows its advantages for medical image segmentation. The skip connection and U-shaped structure of UNet can combine top-down information~~information from the bottom and the top~~. At the same time, the parameter of Unet is small. The parameter of the original UNet is about 28M, and the parameter of UNet with transposed convolution is about 31M. If the number of channels is doubled, the model can be smaller. After shrinking twice, the UNet parameter volume is 7.75M. A four-fold reduction can reduce the amount of model parameters to within 2M, which is very lightweight, so that overfitting can be avoided when the amount of data is small.

Fully Convolutional Network (FCN) is a well-known image segmentation network, while U-Net is more widely used in the medical image processing direction~~while in the medical image processing direction, U-Net is more widely used~~. Compared with FCN, UNet's first feature is complete symmetry, while FCN's decoder is relatively simple, using only ~~a~~ one deconvolution operation, and no convolution structure is added afterwards. The second difference is skipping connection. FCN uses summation and U-Net uses concatenation.

In this paper, we propose an image segmentation model based on Unet. In the feature extraction part of the model, we use the network structure of SE_ResNeXt. The first half of our model is mainly used for feature extraction, and the second half of the network is mainly used for upsampling. At the same time, we continue to use Unet's feature fusion method to splice the features together in the channel dimension to form ~~a~~ thicker features.

V. EXPERIMENTS

A. Loss Function

In the semantic segmentation task, different loss functions can be selected according to the distribution of the data to perform numerical operations on the network output and labels to achieve a better training effect. Particularly in the case of unbalanced data samples and different sample difficulties~~In particular, when the data sample is not balanced and the sample difficulty is different~~, choosing FocalLoss and DiceLoss can often achieve a multiplier effect with half the effort. In this paper, we use a joint loss function, mainly FocalLoss and DiceLoss.

Focal loss is proposed by He Yuming's team for the imbalanced training samples and different sample difficulty~~for the imbalance of training samples and the different difficulty of samples~~. It is a variant of the cross-entropy loss function. We know that One-Stage's target detector usually produces 10k-order boxes, but only a few are positive samples, and the number of positive and negative samples is very unbalanced. We usually add a parameter in front of the binary cross-entropy loss to solve this problem. Although this method balances the number of positive and negative samples, in fact, a large number of candidate targets ~~in target detection~~ are easily distinguishable samples in target detection. The loss of these samples is very low, but due to the extremely unbalanced number, the number of easily distinguishable samples is relatively ~~too~~ large, which ultimately dominates the total loss. Therefore, we believe that samples with high confidence have very little effect on the improvement of the model, and the model should mainly focus on samples that are difficult to distinguish. Focal Loss solves this problem by reducing the loss of high-confidence samples.'

The Dice coefficient is a ~~measurement~~ metric function used to measure the similarity of sets, and is usually used to calculate the pixels between two samples. Dice Loss is used when the sample is extremely unbalanced, because Dice Loss is an area-related loss. Area correlation refers to the fact that~~means that~~ the loss of the current pixel is not only related to the predicted value of the current

pixel, but also related to the values of other points. The form of Dice Loss intersection can be understood as a mask operation, so no matter how big the picture is, the loss calculated for the area of the fixed-size positive sample is the same, and the supervision contribution to the network will not vary with the size of the picture. ~~Variety.~~ In general, Dice Loss has a good performance for scenes with severe imbalance between positive and negative samples, and the training process focuses more on mining the foreground area.

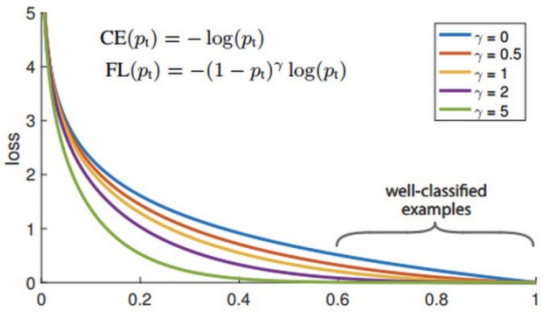

Figure 64. Probability of ground truth class

B. *Experimental Steps*

First, we preprocess the data set, including data cleaning and image enhancement. Data cleaning is to improve the quality of the data and make the data more suitable for training models. We use information completion and other methods to ensure the data integrity, and delete duplicate data to ensure the data uniqueness~~data~~ uniqueness. Image enhancement is to expand the breadth of data distribution, so that the trained model has stronger generalization ability.

Then, we built a network structure based on Unet and SE_ResNeXt for feature extraction. The process of model training is mainly divided into three steps. We first use a larger sampling rate and learning rate to train the model so as to quickly obtain an effective model with less time consumption~~a model with good price comparison~~. Then we load the best model ~~in~~ from the first step to continue training, reduce the sampling rate to 0.6, and set the initial learning rate to 1e-5. Finally, we further reduce the sampling rate for model training.

In the process of model training, in order to accelerate~~speed up~~ the convergence speed and obtain the optimal model, we adopted the following methods:

(1) Set reasonable initialization weights and biases

Deep learning ~~uses~~ utilizes forward propagation and backward propagation to continuously adjust parameters to extract optimal features to achieve the purpose of prediction. The parameters to be adjusted are weighted and biased. Model training is essentially ~~the~~ a process of adjusting weights and biases. If they are initialized to ~~a~~ reasonable values, the convergence speed can be accelerated.

(2) Optimize the learning rate

In the process of continuously adjusting weights and biases, the magnitude of each adjustment is the learning rate. During model training process, if the learning rate is too small, it will increase the number of iterations and increase the training time. If the learning rate is too large, it is easy to overstep the local optimum and reduce the accuracy. Therefore, we set a larger learning rate at the beginning of training to speed up the convergence. At the end of the training period, we will adjust the learning rate slightly ~~a little bit~~, so that it falls steadily into the local optimal solution~~so as to stably fall into the local optimal solution~~. We use an adaptive optimization algorithm when training the model, which can realize the adaptive adjustment of the learning rate, thereby ensuring the accuracy and accelerating the convergence speed~~speeding up the convergence~~.

(3) Regularization of the input value of the network node

In the process of model training, the distribution of input values in each layer changes ~~When the model is trained, the input value distribution of each layer is changing~~. Regardless of whether the input value is large or small, our learning rate is ~~the same~~fixed, which is a waste of efficiency. And when the input value is very small,

the learning rate should not be set too high to allow fine tuning~~in order to ensure the fine adjustment of it, the learning rate cannot be set too large~~. Using regularization can make the input value fall within a certain range, thereby speeding up the convergence speed of the model.

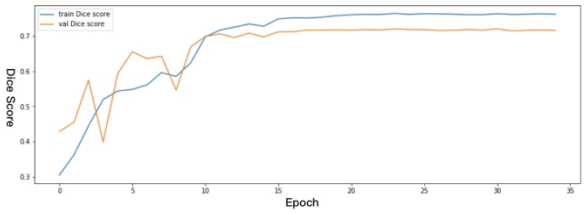

Figure 75. Training process

C. *Evaluation Index*

Deep learning has achieved remarkable achievements~~obvious results~~ in image semantic segmentation, and has produced many models and benchmark data sets focusing on image semantic segmentation. These benchmark data sets provide a unified set of critical model standards. Most of the time we evaluate the performance of a model~~we evaluate a model The the performance will be considered~~ in terms of execution time, memory usage, and algorithm accuracy. Sometimes the

evaluation indicators are different depending on the application scenarios of the model. Accuracy is prioritized for some harsh ~~use~~ usage scenarios, and speed is prioritized for some real-time application scenarios. For semantic segmentation models, the following methods are usually used to evaluate the effectiveness~~effect~~ of the algorithm.

(1) Execution time
Running time is a very critical indicator, especially in the inference stage after model deployment. In some application scenarios, we know that the time required to train the model is also meaningful, but it is usually not very important. The main reason is that the training model is not a real-time requirement, unless the training time is extremely long or the running speed during training is extremely slow. Another problem is that execution time is easily affected by various hardware resources, so it is generally difficult to measure uniformly, regardless of hardware resources, and it is unfair to measure the quality of the model by one-sided execution time.

(2) Memory usage
For deep learning models, memory is another important factor. Although memory can be expanded at any time in most scenarios, it can be very precious on some embedded devices~~memory is also very precious on some embedded devices~~. Even with high-end GPU cards, memory is not unlimited. Consumption, so the network's memory consumption is also an indicator for evaluation and consideration.
(4) Accuracy
Accuracy is the most important technical indicator for evaluating image segmentation networks. Different tasks have different accuracy estimation methods, ~~but~~ and they can be divided into two categories. One is based on pixel accuracy, and the other is based on IOU. The evaluation of the most common semantic segmentation methods currently is based on pixel tags.

D. Analysis Of Results

In order to comprehensively evaluate the performance of our proposed method, we choose the classic network architecture as a comparison, and use the same evaluation indicators and data sets for comparison. In this paper, we use the Dice Score metric. The higher the value of Dice Score, the better the prediction effect of the model.
Table 1 shows the experimental results of the classic image segmentation model and our model. It is easy to find that our Seresnext50 model has the highest Dice Score. Specifically, the Dice Score of our Seresnext50 model is 0.02 higher than ~~the~~ DenseNet121 model, 0.013 higher than ~~the~~ ResNet50 model, 0.01 higher than ~~the~~ RestNet101 model, and 0.026 higher than ~~the~~ Vgg16 model. Experiments show that we use ~~the~~ Seresnext50 model to perform well in the task of identifying pneumothorax and segmenting pneumothorax images.

Models	Dice Score	Loss
Seresnext50	0.721	1.88
DenseNet121	0.701	1.98
ResNet50	0.708	1.92
RestNet101	0.711	1.93
Vgg16	0.695	2.09

Table 1. Results of different models on the dataset used for pneumothorax recognition and image segmentation tasks.

In the course of the experiment, we also explored additional feature information of the data, such as the relationship between gender and pneumothorax distribution. These data features provide a guidance~~guiding direction~~ for our further research.

Figure 86. The relationship between gender and pneumothorax distribution

VI. CONCLUSIONS

In this paper, we propose a new method for predicting pneumothorax disease and segmenting the range of pneumothorax in medical images. The data set we used comes from the chest radiological image data provided by the Society for Medical Imaging Informatics (SIIM). First, we performed the data cleaning and data enhancement operations ~~on the data~~, and then we built a network structure based on Unet and SE_ResNeXt for feature extraction. Experiments show that compared with DenseNet121, ResNet50, RestNet101 and Vgg16, our Seresnext50 model can ~~mine~~ extract features of different dimensions more effectively, ~~and has~~ with ~~stronger~~ better recognition and segmentation capabilities. The network we proposed has strong scalability and can be applied to a variety of medical image segmentation scenes. In addition, we analyzed the characteristics of the data set and obtained some constructive guidance..

~~ACKNOWLEDGEMENT~~

REFERENCES

[1] Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: NIPS (2014).

[2] Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS. pp.2852–2860 (2012)

[3] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation (2014), arXiv:1411.4038 [cs.CV]

[4] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.,Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4), 541–551 (1989).

[5] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. pp. 1106–1114 (2012).

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed,D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In CVPR,2015.

[7] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inceptionv4, inception-resnet and the impact of residual connections on learning. In ICLR Workshop, 2016..

[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In CVPR, 2016.

[9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016.

[10] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[11] L. Sifre and S. Mallat. Rigid-motion scattering for texture classification. arXiv:1403.1687, 2014..

[12] E. Denton, W. Zaremba, J. Bruna, Y. LeCun, and R. Fergus. Exploiting linear structure within convolutional networks for efficient evaluation. In NIPS,2014.

[13] Y. Ioannou, D. Robertson, R. Cipolla, and A. Criminisi. Deep roots: Improving cnn efficiency with hierarchical filter groups. arXiv:1605.06489, 2016

[14] Y.-D. Kim, E. Park, S. Yoo, T. Choi, L. Yang, and D. Shin. Compression of deep convolutional neural networks for fast and low power mobile applications. In ICLR, 2016

[15] M. Jaderberg, A. Vedaldi, and A. Zisserman. Speeding up convolutional neural networks with low rank expansions. In BMVC, 2014