Scientific Research Publishing

Health, 2023, 15, 436-456
https://www.scirp.org/journal/health
ISSN Online: 1949-5005
ISSN Print: 1949-4998

Two-Stage Segmentation of Lung Cancer Metastasis Lesions by Fusion of Multi-Resolution Features

Jingwen Zhao[1*], Xinyu Wang[1], Yunlang She[2], Shuohong Wang[3]

[1]School of Electronic and Electric Engineering, Shanghai University of Engineering Science, Shanghai, China
[2]Shanghai Pulmonary Hospital, Shanghai, China
[3]Department of Molecular and Cellular Biology and Center for Brain Science, Harvard University, Cambridge, MA, USA
Email: *jingwen_echo@outlook.com, 13263693196@163.com, wangsh@fas.harvard.edu

How to cite this paper: Zhao, J.W., Wang, X.Y., She, Y.L. and Wang, S.H. (2023) Two-Stage Segmentation of Lung Cancer Metastasis Lesions by Fusion of Multi-Resolution Features. *Health*, **15**, 436-456.
https://doi.org/10.4236/health.2023.155029

Received: April 26, 2023
Accepted: May 23, 2023
Published: May 26, 2023

Abstract

The deep learning method automatically extracts advanced features from a large amount of data, avoiding cumbersome manual feature screening, and using digital pathology and artificial intelligence technology to build a computer-aided diagnosis system to help pathologists quickly make objective and reliable diagnoses and improve work efficiency. Because pathological images are limited by factors such as sample size, manual labeling expertise, and complexity, artificial intelligence algorithms have not been extensively and in-depth researched on pathological images of lung cancer metastasis. Therefore, this paper proposes a lung cancer metastasis segmentation method based on pathological images, to further improve the computer-aided diagnosis method of lung cancer.

Keywords

Transfer Learning, Pathological Image, ACR-UNet, Deep Learning, Cancer Metastasis

1. Introduction

In the process of interpreting pathological maps, clinicians need to constantly move and review under the microscope and adjust low magnification and high magnification for observation and analysis at different resolutions. The image feature information obtained at different resolutions is different. For example, the morphological characteristics and distribution of tissues can be observed at low magnifications, and the morphological characteristics at the cell level can be

observed at high magnifications. Features at different resolutions are crucial for practical diagnosis. Therefore, aiming at the segmentation of lung cancer metastases in pathological images, this paper designs a two-stage segmentation model of lung cancer metastases that combines multi-resolution features. First, the data was preprocessed, and two datasets of high- and low-resolution were produced. In the first stage, in order to reduce the false positive rate and improve the detection speed, the block pathological images are classified, and the pathological images classified as abnormal tissues are input to the next stage. In the second stage, the atrous convolution residual Unet network structure (ACR-Unet) was designed as a segmentation model, and the atrous convolution residual block was added to increase the width and depth of the model while capturing context information, and the training idea of transfer learning was applied. By fusing high and low resolution information, the segmentation contour is refined. The overall structure of the two-stage lung cancer metastasis segmentation model fused with multi-resolution features is shown in Figure 1.

2. Digital Pathology Image Production

Digital pathology images are usually whole-slide images (WSI) made from tissues stained with Hematoxylin and eosin (H & E). The production process is shown in Figure 2. First, fresh tissue blocks were obtained by puncture, and after trimming and sectioning, they were placed on glass slides for H & E staining to make sections. Then put the prepared slices into the scanning device, collect pictures of the target area and store them in the local database [1] [2].

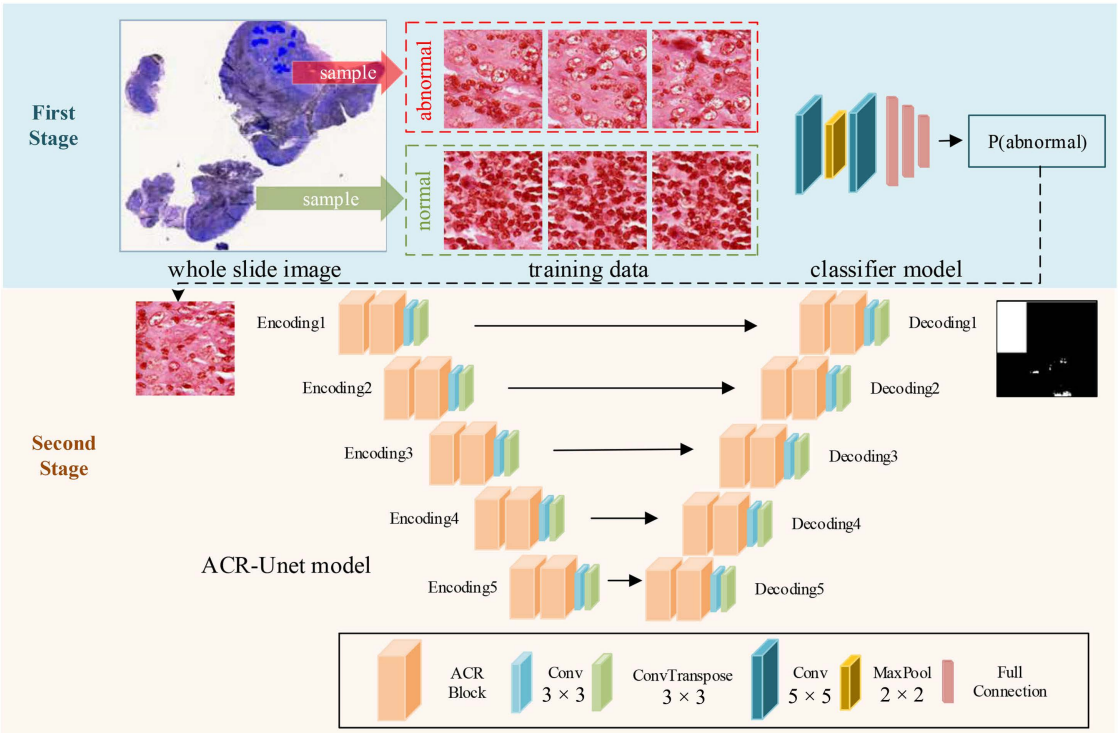

Figure 1. Two-stage lung cancer metastasis lesion segmentation model incorporating multi-resolution features.

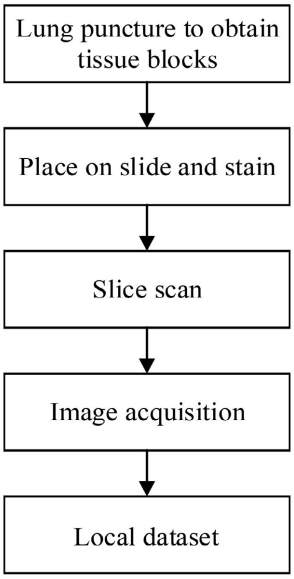

Figure 2. Digital pathology image production process.

In the process of image acquisition, the whole pathological slide is scanned by automatic pathological slide scanning technology. First, the digital microscope is used to scan and image the slice under the low-magnification objective lens, and the micro-scanning platform automatically scans and moves according to the XY axis of the slice, and realizes the automatic focus on the Z axis. Then, the high-efficiency magnification of the optical magnification device is realized by the scanning control software, and high-resolution digital images are obtained according to the program-controlled scanning method. Finally, through image compression and storage software, the images are automatically seamlessly spliced and digitally sliced to generate a complete WSI. The fabricated WSI can be scaled arbitrarily, and viewed and analyzed in any direction, similar to operating an actual optical microscope.

WSI is stored in a pyramid structure, as shown in **Figure 3**, images can be retrieved at different magnifications, and contain a large amount of detailed information from a medical and computational point of view, so the size of each slide is large.

Digital pathology images were visualized, annotated, and automatically analyzed using the Automated Slide Analysis Platform (ASAP). The entire digital pathological image of lung cancer metastasis is stored in svs format. The visualization results at the lowest resolution are shown in **Figure 4**. The blue area is the cancer metastasis area marked by the doctor, in addition to normal tissue areas and large areas. An empty area. ASAP supports viewing most WSI formats, including tif, svs, ASAP can not only zoom in and out to view images in multiple resolutions, but also use the annotation function to create new annotations, and visualize the results of partial annotations in high resolution. As shown in **Figure 5**, the details of cancerous cells and normal cells can be clearly observed at high resolution.

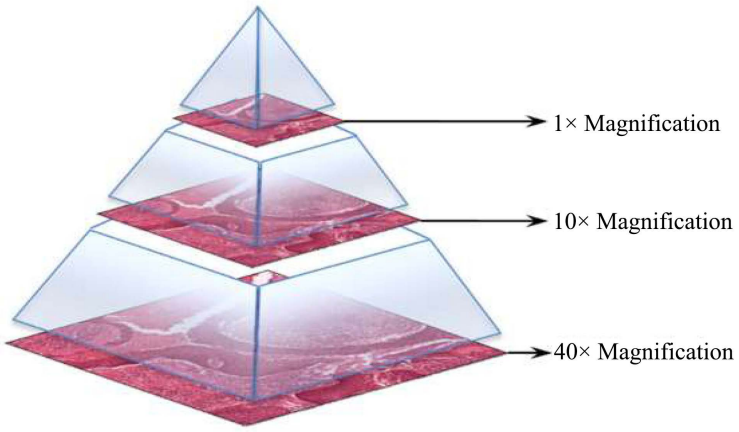

Figure 3. Schematic diagram of pyramid structure.

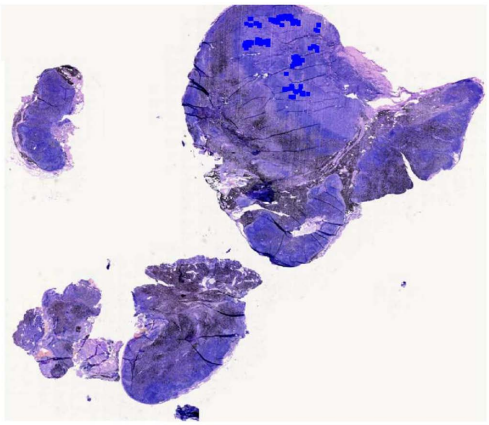

Figure 4. Lung cancer metastasis WSI.

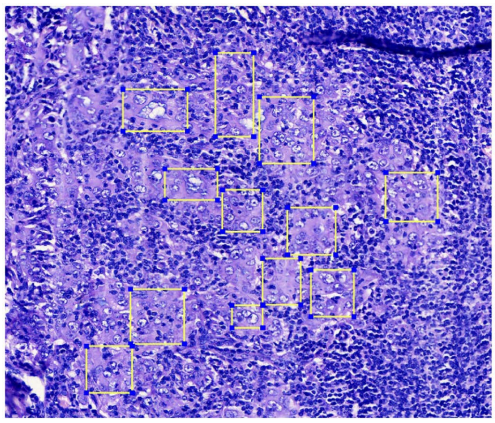

Figure 5. Partial WSI annotation result.

Lung Cancer Metastasis Pathology Image Dataset

1) Introduction to Data and Samples

The TNM staging of lung cancer consists of three stages: the size of the tumor (T stage), the spread of cancer cells to regional lymph nodes (N stage), and whether the cancer cells have metastasized to other parts of the body (M stage).

Lung cancer metastasis is most likely to occur, and the earliest occurrence is regional lymph node metastasis. Regional lymph nodes include intrapulmonary lymph nodes, hilar lymph nodes, and mediastinal lymph nodes. In the experiment, the pathological images obtained by regional lymph node biopsy were selected, and the pathological image data set of lung cancer metastasis was made by the block sampling method.

2) Block sampling method

The block sampling method is also called the image patch-based method. Due to the ultra-high resolution and large size of histopathological images, they generally cannot be processed directly, but a sufficient number of samples can be extracted for training, and the image is cut into several images. Each image patch is called an image patch [3] [4]. For each image patch, the corresponding features can be obtained, and then the features of these image patches can be aggregated to predict the entire image. The size of pathological images usually reaches tens of thousands of pixels, and the input size required by neural networks is usually around 500 × 500 pixels. Therefore, ultra-high-resolution histopathological images, they cannot be input into the network for calculation at one time, and block sampling is usually required.

3) Dataset production process

The experiment uses the pathological images provided by the Thoracic Surgery Department of Shanghai Pulmonary Hospital to create a pathological image dataset of lung cancer metastasis. The images are stored in SVS format and contain images of multiple resolution versions. Level 0 (level 0) is the highest resolution. High-resolution images can observe the internal structure of cells more clearly, which is conducive to the discovery of tiny lesions and the metastasis of a single cell. The resolution of images above Level 1 gradually decreases, and the resolution of images above Level 2 (level 2) is low and can be observed. Considering the overall contour features of the pathological image, it is impossible to distinguish whether the cells are normal or cancerous. Therefore, the experiment selected level 0 and level 1 images to create a lung cancer metastasis pathological image data set.

The specific production process is shown in **Figure 6**. Considering that the predicted image input into the network may contain blank areas, the area selected in the entire pathological image should contain both normal tissues, abnormal tissues and blank areas. After determining the horizontal and vertical coordinates of the area boundary, start from the upper left corner, and then cut the patch. The size of each patch is 256 × 256, and the overlap rate is 50%. 9890 patches were obtained at level 0 and level 1 respectively.

The doctor's annotation information is an XML file, and for training the segmentation model, a corresponding mask needs to be generated. ASAP also provides a Python package module that can be used for reading and writing WSI images, and a multi-resolution image-interface Python package that converts annotations into masks document.

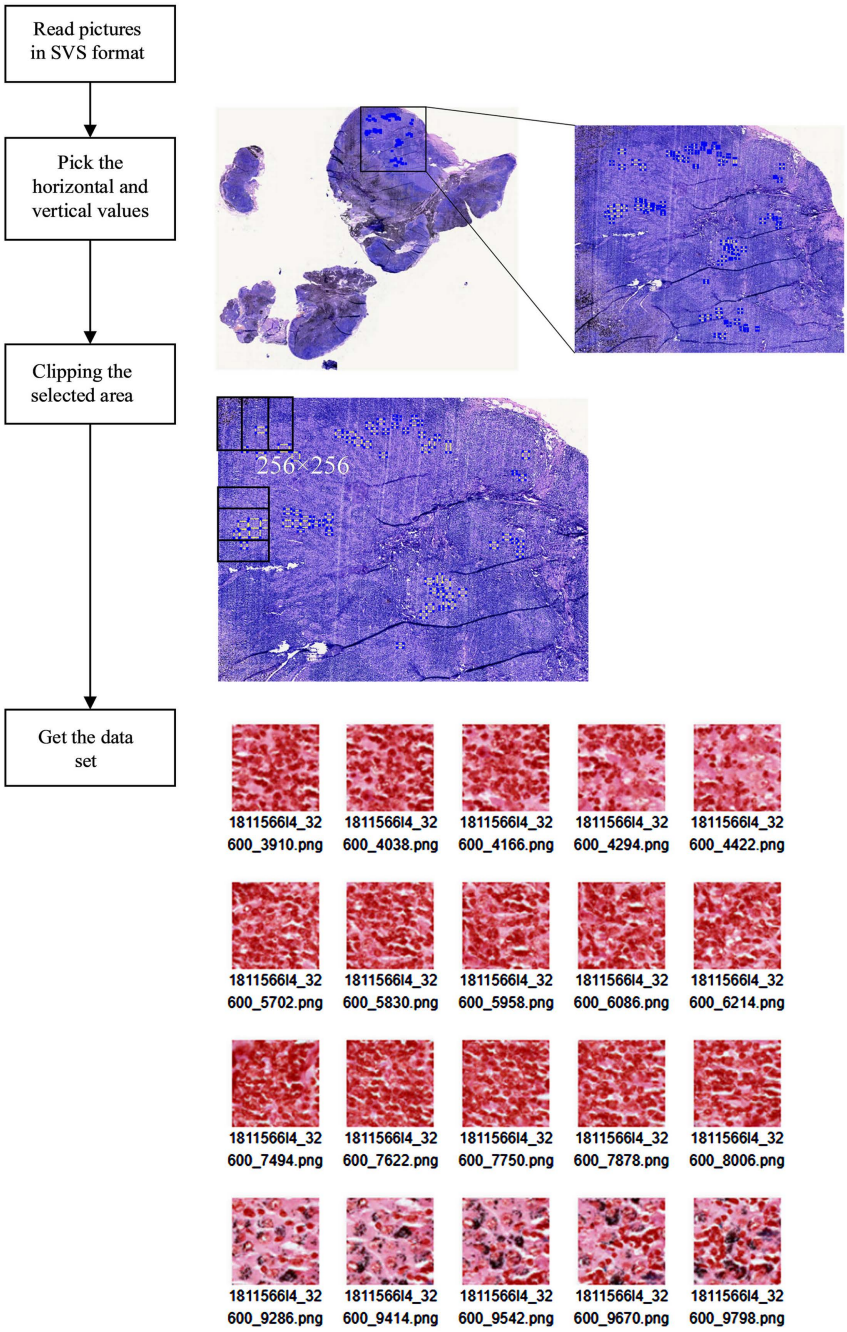

Figure 6. Lung cancer metastasis dataset production process.

First, use the Python package provided by ASAP to convert the xml file into a corresponding mask file, and select the horizontal and vertical coordinates of the area in the mask to be consistent with the original image. In the selected area, the cropping and block processing is performed, the size of each block is still 256 × 256, and the overlap rate (overlap) is 50%, and the corresponding mask data set is made. The mask (Mask) production process corresponding to the data set is shown in **Figure 7** (the images shown in the figure are all low-resolution images).

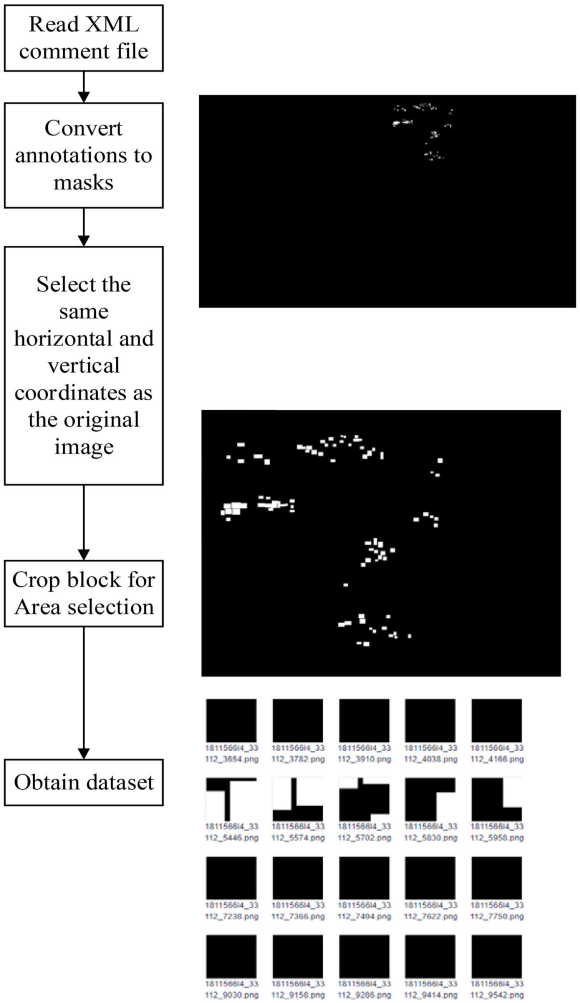

Figure 7. Mask production process.

3. Validation Dataset

1) Introduction of Camelyon17

The Camelyon17 dataset comes from the Camelyon17 Challenge, which is an automatic detection and assessment of breast cancer development based on complete slide images of histological lymph node sections. Lymph node metastasis has an important impact on the prognosis of breast cancer. When cancer cells spread to lymph nodes, it will have a great impact on the survival and prognosis of patients. To improve the understanding of the lymphatic system, a method that can automatically detect and identify lymph node metastases is needed. An automated solution would greatly reduce pathologist' workload and also reduce diagnostic subjectivity. Currently, accurate pathological grading of breast cancer remains a major challenge. In the case of breast cancer, the TNM stage is made up of the size of the tumor (T stage), the spread of the cancer to regional lymph nodes (N stage), and whether the cancer has metastasized to other parts of the body (M stage). Camelyon17 mainly studies the N-stage problem of breast cancer.

2) Dataset production process

Camelyon17 is a dataset containing 1399 annotated whole-slide images of lymph nodes, including lymph nodes with and without metastasis, with a total data volume of 3 TB. The data were collected from five different medical centers and covered different image appearance and staining changes, and each intact slide image had an index indicating whether it contained metastases (including large metastases, small metastases) or isolated tumor cells. Label, where detailed hand-drawn outlines of their metastatic lesions are provided for 209 full-slide images.

In order to make a block pathological image data set, this paper selects 5 pieces of breast cancer metastatic cell data from different institutions, covering full-slide images of small-area cancer cell metastasis and large-area cancer cell metastasis. The processing process of the slide image is the same as that of the lung cancer metastasis pathological image data set, and the flow chart is the same as Figure 2. 11,621 patches were obtained at level 0 and level 1 respectively. A partial screenshot of the Camelyon17 dataset is shown in Figure 4. The mask production process of Camelyon17 is the same as in Figure 5, and some screenshots of the mask are shown in Figure 8.

4. CNN-Based Classification Model for Blocky Pathological Images

Directly performing end-to-end pixel clustering makes the model tend to only focus on low-level image features, such as color, contrast, etc., and lose the semantic discriminativeness of features. Two-stage segmentation can avoid this

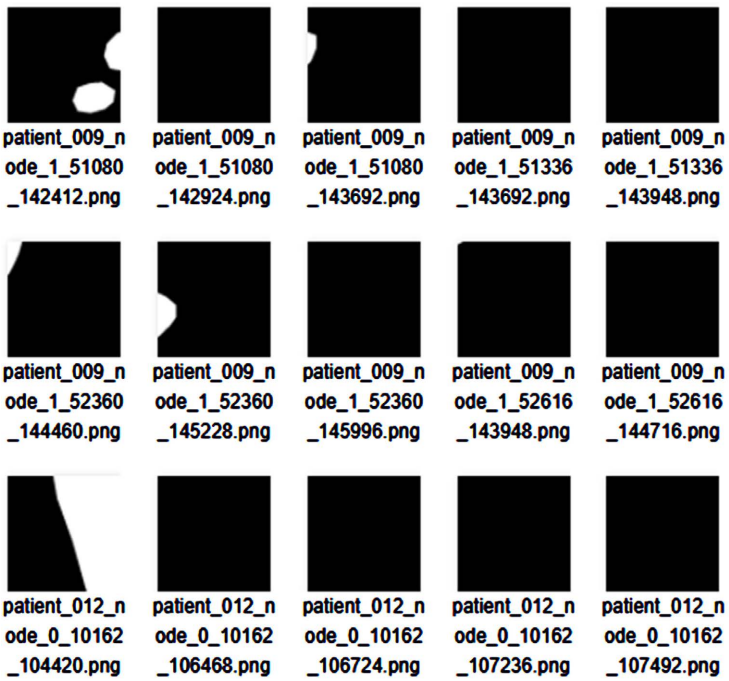

Figure 8. Camelyon17 breast cancer metastasis dataset's mask.

shortcoming. Moreover, most of the tissues in the pathological image are non-cancerous tissues and blank areas that do not contain cells. Classification is performed before the segmentation of lung cancer metastatic lesions, which has the advantage of reducing the false positive rate and improving the detection efficiency.

Convolutional neural networks are widely used in target detection and recognition, image classification, etc., and they show the best results in image classification models. Especially for pathological images, CNN can be used to achieve classification, improve the accuracy of the algorithm and reduce the complexity of the algorithm. The classification process of blocky pathological images is shown in Figure 9.

Convolutional neural networks generally include three types of layers: convolutional layers, pooling layers, and fully connected layers. The combination of the convolutional layer and the pooling layer forms a feature extractor, which is responsible for the extraction of high-level and low-level features of the image. The fully connected layer is responsible for sending the features mapped by the extractor to the final output layer.

1) Convolutional Layer

The process of convolution is actually the process of feature extraction. The most important concept is the convolution unit (also called convolution kernel). Each convolution layer is composed of convolution kernels of different sizes. The size, stride and padding of the convolution kernel together determine the features extracted by the convolution layer and the size of the output feature map. For example, a 5×5 image is convolved with 3×3 non-zero padding and the stride is set to 1 to obtain a 3×3 feature map. Usually the first layer of convolution is responsible for extracting low-level semantic features, such as edges, contours, corners, etc., and the deeper convolutional layers are continuously iterated from low-level features to extract more complex high-level semantic features.

2) Pooling Layer

The pooling process is actually a down-sampling process, which simulates the human visual system to achieve data dimensionality reduction. In convolutional neural networks, pooling layers are periodically inserted between convolutional layers. The purpose of pooling includes: reducing the amount of data to be processed in the next layer to avoid overfitting; enhancing the scale invariance and rotation invariance of the model. Commonly used truths include mean pooling, maximum pooling, random pooling, median pooling, combined pooling, etc. The most commonly used pooling step size is 2, that is, every 2 pixels are divided into 2×2 image blocks, and then the maximum value of the 4 numbers in each image block is taken, that is, the maximum pooling operation, so that the amount of data is reduced by 75%. Since pooling layers reduce the data size too quickly, most of the research tends to use fewer pooling layers or even no pooling layers.

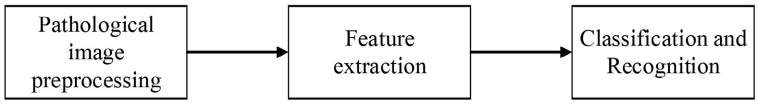

Figure 9. Pathological image classification process.

3) Fully connected Layer

In the convolutional neural network structure, after multiple convolutional layers and pooling layers, one or more fully connected layers are connected. When the features extracted by the feature extractor are sufficient, it is necessary to play the role of the fully connected layer for classification. Each neuron in the fully connected layer is fully connected to all neurons in the previous layer. The function is to reduce the data from high-dimensional to low-dimensional, and capture the class-discriminative parts of the convolutional layer or pooling layer. information is integrated. The fully connected layer of the last layer uses Softmax logistic regression for classification, and this layer can also be called a Softmax layer. If the target task is divided into 5 categories, then the last layer can be set with 5 fully connected units.

The convolutional neural network has the characteristics of high accuracy, fast classification speed, and stable model, so the convolutional neural network is used as a block pathological image classifier. The classifier consists of 2 5 × 5 convolutions, 1 pooling, and 3 full connections, and the stacking method is shown in **Figure 1**.

During training, a high-resolution dataset is used as input. See Equation (1) when the output is $P_{abnormal}$.

$$P_{abnormal} = Classifier\left(I_{level0}\right) \tag{1}$$

First, image features are extracted through two 5 × 5 convolutional layers and a maximum pooling layer, and the equation is shown in Equation (2).

$$F_{level0} = C_{5\times5}\left(f_{mp}\left(C_{5\times5}\left(I_{level0}\right)\right)\right) \tag{2}$$

Then use 3 fully connected layers to classify the extracted features, the formula is shown in Equation (3).

$$P_{abnormal} = fc_3\left(fc_2\left(fc_1\left(F_{level0}\right)\right)\right) \tag{3}$$

5. Segmentation Model of Lung Cancer Metastases

Image segmentation belongs to the visual task of pixel-level classification. According to different granularity levels, it can be divided into: semantic segmentation, instance segmentation, and panoptic segmentation. Semantic segmentation is to divide all pixels in the input image into different categories according to the objects of interest they belong to. Instance segmentation is based on semantic segmentation to segment different objects of the same category. Panoramic segmentation combines semantic segmentation and instance segmentation. Panoramic segmentation assigns semantic labels and instance labels to each pixel.

Semantic segmentation is often used in the segmentation of medical images, such as the segmentation of lesion areas in medical images, which belongs to the special case of semantic segmentation, binary segmentation, that is, semantic segmentation with only a single category in the foreground.

Aiming at the difficulty of segmenting lesion regions in pathological images, first, pathological images are characterized by the background of purple-red tissue fluid and ordinary cell nuclei. Cancerous cells generally appear to have blurred nuclei and slightly larger outlines than normal cells. Other areas such as large areas of white are background information. The shape and size of normal cells and cancer cells vary greatly with the degree of differentiation. Normal cells are mostly oval or round, while cancer cells are irregular in shape and vary in size. Second, cancerous cells are often difficult to distinguish from normal cells, and even the naked eye requires sufficient magnification to detect them.

Therefore, this paper uses images of multiple resolutions to train the network model, and designs a segmentation model that incorporates multi-resolution features. High-resolution images also bring a lot of detailed information. Adding the hole convolution residual module to U-Net increases the depth of the network model and improves the ability of the model to learn the detailed features of high-resolution images.

1) Dilated Convolutional Residual U-Net Network Model

High-resolution pathological images contain rich details of cancerous cells and normal cells, such as nuclei and cell outlines. Studies have shown that the deep network model has a strong ability to extract feature information, and the U-Net [5] model has fewer layers, which is not conducive to processing deep semantic information of high-resolution images [6], and there is a small local receptive field that cannot express The problem with long range dependencies.

Aiming at the problem of insufficient feature extraction ability of U-Net, the model replaces the original codec structure block by stacking the empty space pyramid pooling module and the convolutional layer, and designs the empty convolution residual U-Net as shown in **Figure 1** (Atrous Convolution Residual-Unet, ACR-Unet) network model to improve the model's ability to locate the lesion area.

When the low-resolution data set is used as input for training and the output is I_{level1}^{o}, the overall expression is shown in Equation (4).

$$I_{level1}^{o} = ACRUnet\left(I_{level1}\right) \tag{4}$$

For the encoder part, the encoder receives the input I_{en}^{i-1} from the previous layer and outputs feature information I_{en}^{i}, where the values of i are 1, 2, 3, 4, 5, the formula is shown in Equation (5).

$$I_{en}^{i} = \sum_{i=1}^{5} Encoding_{i}\left(I_{en}^{i-1}\right) \tag{5}$$

Each layer of the encoder contains two ACRB modules, a 3 × 3 convolutional layer and a 3 × 3 deconvolutional layer, the formula is shown in Equation (6).

$$Encoding_i\left(I_{en}^{i-1}\right)=C_{3\times3}^{T}\left\{C_{3\times3}\left[\sum_{i=0}^{n}H_{ACRB}\left(I_{en}^{i-1}\right)\right]\right\} \tag{6}$$

In Equation (6), n is the number of ACRB modules, which is 2 in this paper.

For the decoder part, each layer of the decoder is the same as the encoder, and at the same time, the input I_{de}^{i-1} of each layer is fused with the feature information of the corresponding encoder layer, and the output feature information I_{de}^{o}:

$$I_{de}^{o}=\sum_{i=1}^{5}Decoding_i\left\{©\left[I_{de}^{i-1},I_{en}^{5-(i-1)}\right]\right\} \tag{7}$$

2) Dilated Convolutional Residual Module

Atrous Convolution Residual Block (ACRB), the structure is shown in Figure 10. Among them, each layer of the codec contains two repeated ACRB modules, and each module contains a 1 × 1 convolution, an ASPP module, a 3 × 3 convolution and a skip connection structure. The ACRB module generates rich feature information by increasing the width and depth of the network, and multi-scale fusion can make full use of feature information. When the module receives the input feature F, the feature information F_c is obtained, and the Equation is as follows:

$$F_C=\sum_{k=0}^{n}H_{ACRB}\left(F\right) \tag{8}$$

First, the feature map undergoes a 1 × 1 convolution, a normalization processing layer β (Batch Normalization, BN), and a Relu activation function δ to obtain the prediction result, and the expression form is shown in Equation (9).

$$F_{pre}=\delta\left(\beta\left(C_{1\times1}\left(F\right)\right)\right) \tag{9}$$

The preprocessed feature map F_{pre} is input to the ASPP module, and the multi-scale fusion feature map F_A is output, the expression is as follows:

$$F_A=ASPP\left(F_{pre}\right) \tag{10}$$

Secondly, perform 3 × 3 convolution on the multi-scale fusion feature map, extract deep-level features, and perform normalization processing to improve the feature extraction ability of the model, and obtain the feature F_{AD}, the expression is shown in Equation (11).

$$F_{AD}=\beta\left(C_{3\times3}\left(F_A\right)\right) \tag{11}$$

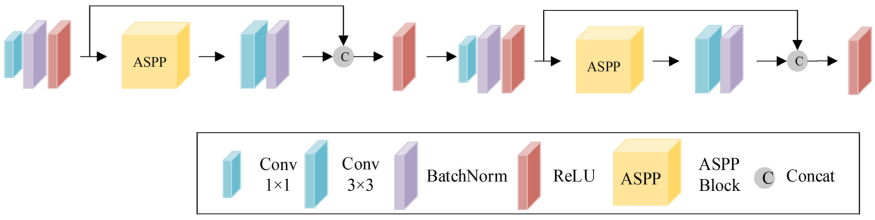

Figure 10. ACRB module structure.

The multi-scale feature and the preprocessing result are fused, and then the Relu activation function is used to obtain the fused feature F_C, the expression is as follows:

$$F_C = H_{ACRB}\left(I_{en}^{i}\right) = \delta\left\{\copyright\left[F_{AD} + F_{pre}\right]\right\} \tag{12}$$

3) Atrous Spatial Pyramid Pooling Block

Atrous Spatial Pyramid Pooling Block (ASPP) contains four parallel branch structures, as shown in Figure 11. The ASPP module can increase the receptive field without downsampling and enhance the network's ability to recognize multi-scale contexts. The input image is sampled in parallel by dilated convolutions with different sampling rates, and the associated features of the larger neighborhood range between pixels are extracted. Finally, the feature maps are added to compensate for the grid effect caused by dilated convolutions. For the task of this paper, the sampling rate is too large to produce meaningless weights. The selected sampling rates are 6, 12, 18 and a 1 × 1 convolution to obtain the multi-scale feature of the preprocessing feature, see Equation (13).

$$F_A = C_{1\times1}\left(F_{pre}\right) + C_6\left(F_{pre}\right) + C_{12}\left(F_{pre}\right) + C_{18}\left(F_{pre}\right) \tag{13}$$

4) Model Training Strategy

According to the characteristics of pyramid structure storage of digital pathological images, two data sets of high and low resolution were produced using one pathological slice. The idea of transfer learning is applied in the training process, where I_{level1} is a low-resolution pathological image and I_{level1} is a high-resolution pathological image. First use to train all layers of the network, then freeze the deep parameters of the encoder and decoder, and use to fine-tuning the shallow codec.

The training process is shown in Figure 12. Green represents the codecs involved in training during fine-tuning, and blue represents the frozen part during fine-tuning. Among them, the deep convolution is responsible for extracting abstract features, including the overall shape of the lung cancer metastasis area; the shallow convolution is responsible for extracting basic features, including the outline and edge of a single cancer cell.

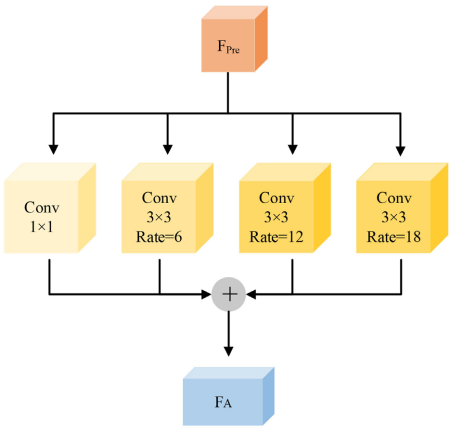

Figure 11. ASPP module structure.

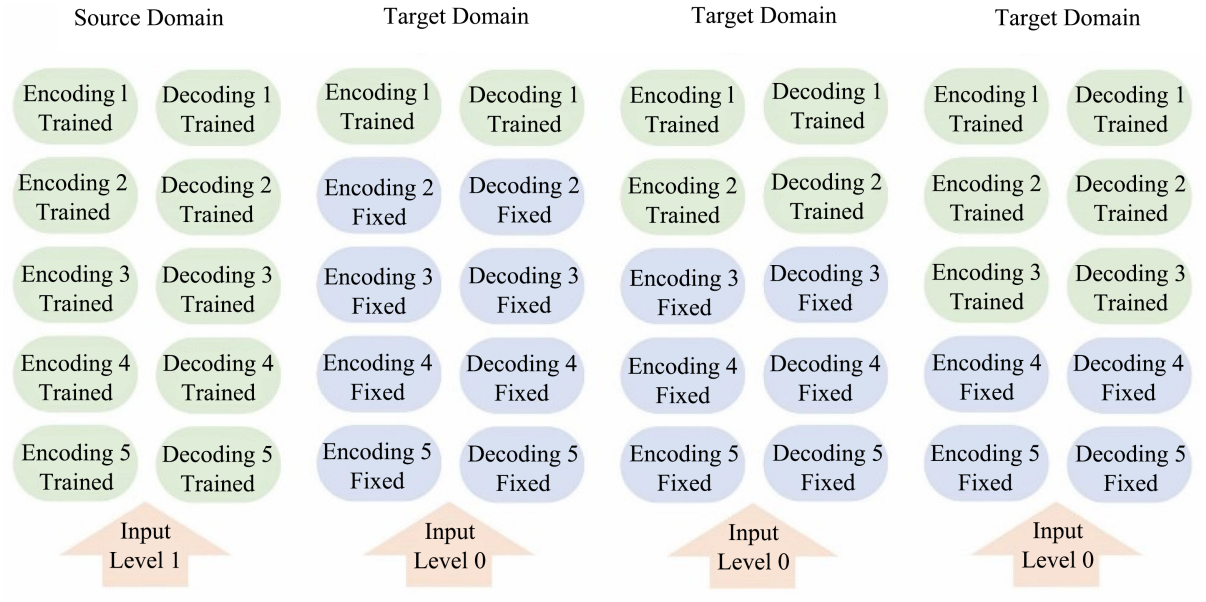

Figure 12. Model training process.

The trained network fuses high- and low-resolution features to ensure that the model can detect not only large areas of cancer metastases, but also single-cell metastases.

6. Experimental Results and Analysis

6.1. Experimental Environment and Evaluation Indicators

The experiment uses Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20 GHz processor, memory size is 16 G, independent graphics card uses two GeForce RTX 2080ti, system type is Ubuntu 18.04, uses Python3.6.9, Tensorflow framework to model Build and train.

In order to preserve the details of pathological images as much as possible and reduce the calculation amount of the model, the cropping size of the images in the dataset is 256 × 256. The network uses the Adam optimizer to optimize the parameters, the loss function uses the binary cross entropy loss function (Binary Crossentropy), the batch size (batch size) is 30, and the number of iterations is 500 and 1000 for different training strategies. The data set is divided into a training set, verification set, and test set according to 6:2:2.

In the two-stage detection network framework, the classifier part uses precision (Precision), recall (Recall), accuracy (Accuracy), F1-score to evaluate the performance of the model, and the formulas correspond to Equations (14)-(17).

$$\text{Precicion} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{16}$$

$$\text{F1-score} = \frac{2\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \tag{17}$$

where (True Positive) denotes true positive, (False Negative) indicates false negative, (False Positive) is false positive, (True Negative) denotes true negative. False Positive Rate refers to the proportion of all negative cases identified as positive after detection. The equation is as follows:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{18}$$

In the stage of tumor metastasis segmentation, the pixel classification accuracy (Pixel Accuracy, PA), image segmentation intersection over Union (mIoU), and Dice similarity coefficient (Dice Similarity Coefficient, DSC) were used to evaluate the performance of the model, and the calculation formulas corresponded to Equations (19)-(21).

$$\text{PA} = \frac{\sum_{i=0}^{k} P_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} P_{ij}} \tag{19}$$

$$\text{mIoU} = \frac{I}{k+1} \sum_{i=0}^{k} \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{20}$$

$$\text{Dice} = \frac{2\text{TP}}{\text{FP} + 2\text{TP} + \text{FN}} \tag{21}$$

6.2. Ablation Experiment

1) Effect of Classification Model on Experimental Results

This paper evaluates the performance of four well-known deep learning networks in this classification task, including: ResNet34, VGG16, CNN and AlexNet, and the classification results are shown in Table 1. Due to the single feature of pathological images, it has better performance in classification models with fewer layers. CNN has the highest classification accuracy, fast speed and stable performance, so CNN is used as the classifier.

In order to verify the improvement of the performance of the model by the classifier, two sets of experiments were set up. The first group directly input the unclassified test set into the segmentation model, and the second group input the images classified as abnormal tissues into the segmentation model, the results are shown in Table 2. It can be seen that the two-stage segmentation reduces the false positive rate by 13.3%, and the prediction completion time of the

Table 1. Comparison of classifier performance.

Model	Precision	Recall	Accuracy	F1-score
ResNet34	0.925	**0.898**	0.907	**0.911**
VGG16	0.897	0.899	0.896	0.898
CNN	**0.965**	0.847	**0.957**	0.902
AlexNet	0.896	0.855	0.889	0.875

Table 2. The effect of classifier on model performance.

Experiment	FPR	predicted time/s
Group 1	0.320	1249
Group 2	**0.187**	**364**

entire test set is increased by 70.8%, and the performance of the two-stage model is better.

2) The effect of the dilated convolution residual module on the experimental results

In order to verify the effect of the dilated convolution residual module on the experimental results, three sets of comparative experiments were set up. The first group only used the trained U-Net, the second group used the trained Res UNet, and the third group used the proposed ACRU-Net. In order to maintain the principle of controlling variables, the training strategy is the same as the third group in (3) of this section. Figure 13 shows the prediction results of the three experiments for the same pathological image, where the black area represents the normal tissue area, and the white area represents the cancerous tissue area. The results show that U-Net is greatly affected by the background information and has a large boundary error when segmenting the lesion area. When using Res UNet with a deeper network layer, the interference of background information is weakened to a certain extent. The former two have the problem of mis-segmenting the background area into lesions. Using the ACRU-Net proposed in this paper, both large-area lesions and small-area lesions can be accurately located, which greatly reduces the influence of background information on the prediction results.

3) Effects of Different Training Strategies on Experimental Results

In order to explore the influence of the number of fine-tuning model layers on the experimental results, four groups of experiments shown in Table 3 were set up under the premise of controlling variables, where √ indicates that the module is fine-tuned, × indicates that the module only uses training, and other conditions remain unchanged. Figure 14 shows the change of the loss function during the training process of the 4 groups of experiments. The change of the loss function of the first and second groups is consistent. The decline speed is faster in the first 100 rounds of training, and it tends to be stable after 100 rounds, and the gradient update is smaller. The difference is that the fluctuation range of Group 2 is larger than that of Group 1. The loss function of Group 4 has only small fluctuations in the first 100 rounds of training, and then it tends to remain almost unchanged. In contrast, the loss function of the third group decreased gradually, fluctuated greatly within the first 500 iterations, and the loss function reached stability after 700 iterations, and its mIoU value combined with Table 3 was better than the other three groups of experiments. This shows that the third group of experimental training has the best effect and can be used as the final training strategy.

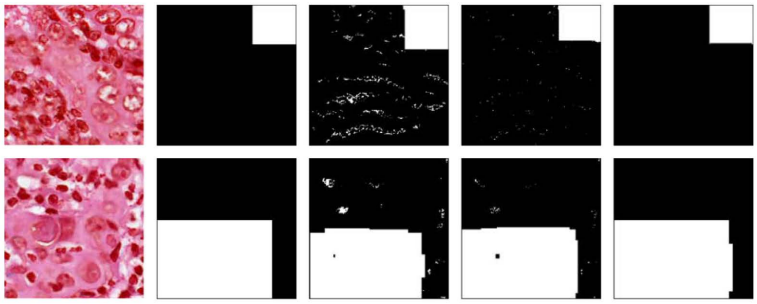

Figure 13. The effect of ACRB module on model performance. (a) Original, (b) Ground truth, (c) U-Net, (d) Res Unet, (e) ACRU-Net.

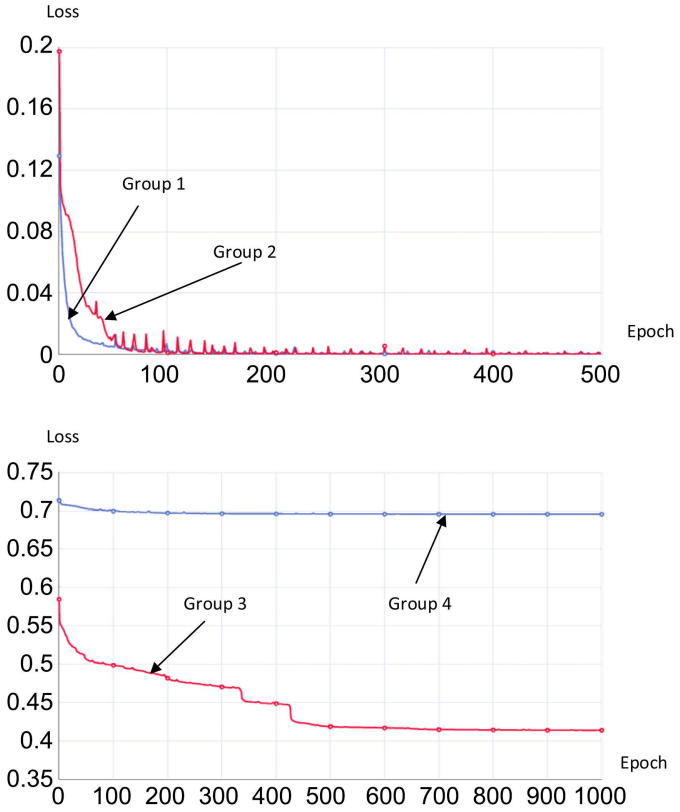

Figure 14. Convergence process of loss function for models with different fine-tuning layers.

Table 3. Different fine-tuning layers training process and results.

Network	Group 1	Group 2	Group 3	Group 4
Encoding 1 Decoding 1	×	√	√	√
Encoding 2 Decoding 2	×	×	√	√
Encoding 3 Decoding 3	×	×	×	√
Encoding 4 Decoding 4	×	×	×	×
Encoding 5 Decoding 5	×	×	×	×
mIoU	0.801	0.875	**0.913**	0.904

4) Model generalization verification

To verify whether the model is applicable to other datasets, the method is applied to the Camelyon17 public dataset. For the first stage, high-resolution data set training is used, and the training strategy for the segmentation model is the same as the third group in (3) in this section. The parameter settings are the same as in Section 5.1, where the number of iterations is 1000. The test set was input into the model, the accuracy rate in the classification stage was 97.7%, and the mIoU in the segmentation stage was 93.5%. As shown in Figure 15, the model performed well in locating the lesion area, which further proved the effectiveness of the model.

6.3. Comparative Experiment

Compare the two-stage segmentation model with 9 end-to-end models on the lung cancer metastasis pathological image dataset (D1) and the Camelyon17 dataset (D2). The training strategy remains the original method, and only one dataset (made under Level0 Data set) for training, did not adopt the training strategy proposed in this paper. The experimental results of each model are shown in Table 4. It can be seen that the evaluation indicators PA, mIoU, and Dice of the method in this paper are significantly better than the mainstream model, and the false positive rate FPN lowest.

Figure 16 shows the segmentation results of the above model on the pathological map of lung cancer metastasis. It can be seen that the classic segmentation network models include FCN [7], DeepLabv3 [8], SegNet [9], Mask RCNN [10], nn U-Net [11], BCDU-Net [12], Medical Transformer [13], TransUNe [14], CaraNet [15], which are not accurate enough for the location of the lesion area, and the false positive rate is high. The latter five methods are medical image segmentation models proposed in recent years, and they also have the problem

Table 4. Comparison of different model performance.

Method	PA		mIoU		Dice		FPN	
	D1	D2	D1	D2	D1	D2	D1	D2
FCN [7]	0.899	0.875	0.905	0.869	0.812	0.844	0.311	0.315
DeepLabv3+ [8]	0.931	0.890	0.898	0.901	0.902	0.917	0.210	0.291
SegNet [9]	0.929	0.911	0.906	0.899	0.932	0.890	0.294	0.332
Mask RCNN [10]	0.891	0.879	0.884	0.903	0.878	0.889	0.295	0.344
nn U-Net [11]	0.877	0.855	0.781	0.873	0.924	**0.927**	0.209	0.254
BCDU-Net [12]	0.879	0.907	0.859	0.891	0.891	0.866	0.195	0.265
Medical Transformer [13]	0.904	0.890	0.901	0.900	0.918	0.879	0.214	0.277
TransUNe [14]	0.925	0.917	0.863	0.895	0.894	0.905	0.203	0.253
CaraNet [15]	0.911	0.869	0.892	0.897	0.909	0.890	0.218	0.217
Ours	**0.948**	**0.937**	**0.913**	**0.935**	**0.934**	0.915	**0.187**	**0.197**

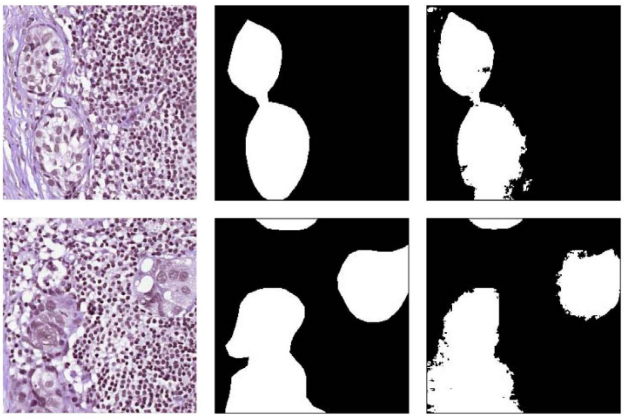

Figure 15. Validation results of Camelyon17 dataset. (a) Original, (b) Ground truth, (c) Experimental results.

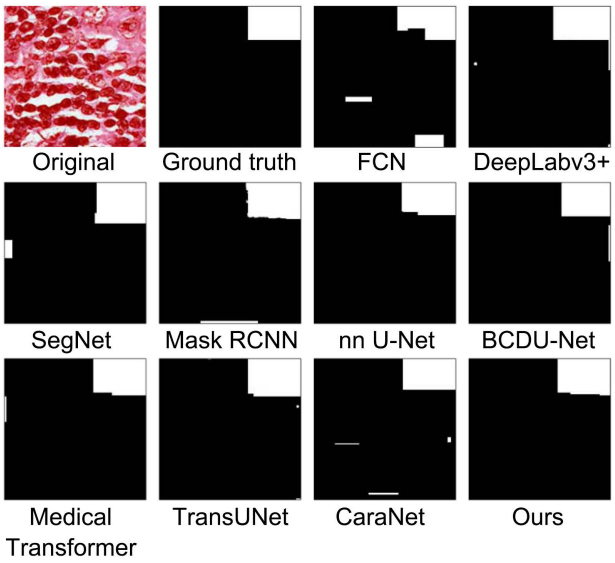

Figure 16. Lung cancer metastasis pathological image compare experimental results.

of misclassifying normal cells as cancerous cells, and their marginal performance is not good. Compared with other networks, the method proposed in this paper is more accurate in locating the lesion area on the pathological image data set, and the edge outline is clear. That is suitable for lesion location.

Figure 17 shows the segmentation results of the above model on the pathological map of lung cancer metastasis. It can be seen that the classic segmentation network models include FCN, DeepLabv3, SegNet, and Mask RCNN, which are not accurate enough for the location of the lesion area, and the false positive rate is high. The last five methods are medical image segmentation models proposed in recent years, and they also have the problem of misclassifying normal cells as cancerous cells, and their marginal performance is not good. Compared with other networks, the method proposed in this paper is more accurate in locating the lesion area on the pathological image data set, and the edge outline is clear.

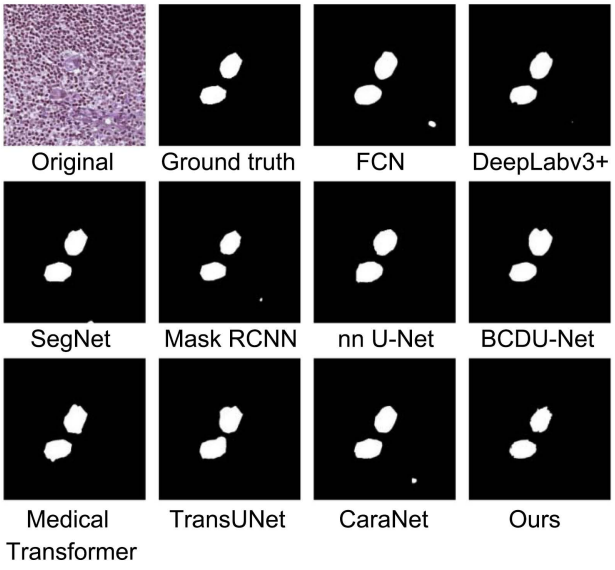

Figure 17. Camelyon16 compare experimental results.

7. Conclusion and Discussion

This article introduces the process and principle of obtaining tissue from lung biopsy to making digital pathological images. Then, the pathological image dataset of lung cancer metastasis was made using the pathological images acquired by lymph node aspiration, and the block dataset was made using the Camelyon17 public dataset for the validation of model validity. According to the comparative test above, the method proposed in this paper is more accurate in locating the lesion area on the pathological image data set, and the edge outline is clear, which is suitable for this task.

Fund

This work was supported by 1) Science and Technology Commission of Shanghai Municipality: No. 19ZR1421500 and 2) National Natural Science Foundation of China: Grant No. 61175036.

Conflicts of Interest

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

References

[1] Zhang, C.X., Wu, W.J., Yang, J. and Sun, J.Y. (2022) Application of Artificial Intelligence in Respiratory Medicine. *Journal of Digital Health*, **1**, 30-39. https://doi.org/10.55976/jdh.1202215330-39

[2] Fan, L., Xia, Z.Q., Zhang, X.B. and Feng, X.Y. (2017) Lung Nodule Detection Based on 3D Convolutional Neural Networks. 2017 *International Conference on the Frontiers and Advances in Data Science*, Xi'an, 23-25 October 2017, 7-10.

[3] Van Gansbeke, W., Vandenhende, S., Georgoulis, S. and Van Gool, L. (2021) Un-

supervised Semantic Segmentation by Contrasting Object Mask Proposals. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 10032-10042. https://doi.org/10.1109/ICCV48922.2021.00990

[4] Wang, D.Y., Khosla, A., Gargeya, R., Irshad, H. and Beck, A.H. (2016) Deep Learning for Identifying Metastatic Breast Cancer. *arXiv*.

[5] Weng, W.H. and Zhu, X. (2021) INet: Convolutional Networks for Biomedical Image Segmentation. *IEEE Access*, **9**, 16591-16603. https://doi.org/10.1109/ACCESS.2021.3053408

[6] Lassen, B.C., Jacobs, C., Kuhnigk, J.M., van Ginneken, B. and van Rikxoort, E.M. (2015) Robust Semi-Automatic Segmentation of Pulmonary Subsolid Nodules in Chest Computed Tomography Scans. *Physics in Medicine & Biology*, **60**, Article 1307. https://doi.org/10.1088/0031-9155/60/3/1307

[7] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

[8] Chen, L.C., Zhu, Y.K., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, Cham, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49

[9] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615

[10] He, K.M., Gkioxari, G., Dollár, P. and Girshick, R. (2017) Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/ICCV.2017.322

[11] Isensee, F., Petersen, J., Kohl, S.A.A., Jäger, P.F. and Maier-Hein, K.H. (1904) nnU-Net: Breaking the Spell on Successful Medical Image Segmentation. *arXiv*.

[12] Azad, R., Asadi-Aghbolaghi, M., Fathy, M. and Escalera, S. (2019) Bi-Directional ConvLSTM U-Net with Densley Connected Convolutions. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Seoul, 27-28 October 2019, 406-415. https://doi.org/10.1109/ICCVW.2019.00052

[13] Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I. and Patel, V.M. (2021) Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Strasbourg, 27 September 2021, 36-46. https://doi.org/10.1007/978-3-030-87193-2_4

[14] Chen, J., Lu, Y., Yu, Q., *et al.* (2021) Transunet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv*.

[15] Lou, A., Guan, S.Y., Ko, H. and Loew, M.H. (2021) Caranet: Context Axial Reverse Attention Network for Segmentation of Small Medical Objects. *arXiv*.