Deep Mask For X-ray Based Heart Disease Classification

Xupeng Chen
Department of Life Science
Tsinghua University
Beijing, 100084
xp-chen14@mails.tsinghua.edu.cn

Binbin Shi
Department of Life Science
Tsinghua University
Beijing, 100084
ltbyshi@gmail.com

Abstract

We build a deep learning model to detect and classify heart disease using $X-ray$. We collect data from several hospitals and public datasets. After preprocess we get 3026 images including disease type VSD, ASD, TOF and normal control. The main problem we have to solve is to enable the network to accurately learn the characteristics of the heart, to ensure the reliability of the network while increasing accuracy. By learning the doctor's diagnostic experience, labeling the image and using tools to extract masks of heart region, we train a U-net to generate a mask to give more attention. It forces the model to focus on the characteristics of the heart region and obtain more reliable results.

1 Introduction

Heart related disease kills millions of people every year, and better detection and diagnosis of heart disease can help prevent and treat diseases. Chest $X-rays$ is commonly used as an early diagnosis tool. It is inexpensive and convenient compared with other tools like CT and f-MRI. There are already many studies using automated analysis of chest $X-rays$ to help diagnosis and analysis in heart disease, especially early diagnosis.

In this study, we propose to use a novel automated method to locate the mask of heart and give more attention to the region of interest(ROI). We use a deep encoder decoder model(U-net) to extract masks of heart and classify hearts by paying more attention to the ROI. Our contributions are i) a good approach to detect ROI, ii) an automated method to classify heart disease in a more reliable way.

2 Methods

2.1 Problem depiction

There has been considerable research in the application of deep learning to medical imaging, especially in the field of cancer detection. However, there are not many studies on heart disease, especially the use of chest X-ray to diagnose heart disease. Compared with the amount of data generated by technologies such as CT and MRI, the amount of information that chest radiographs can provide is much less. The above is even more challenging.

We chose to use chest X-ray data, which is considered to be the simplest, most effective and low-cost diagnostic method, which makes sense for cheap and quick diagnosis. However, the difficulty of analysis also brings us great challenges. We can't put aside the physiological meaning, and try to solve this problem directly by applying the model. We found that some of the existing chest radiographs have been used to diagnose heart disease. Most of them use some of the more classic deep learning image processing networks directly. Good results, but did not delve into the problems behind this.

Most of the papers use VGG, Resnet, Densenet and other models. They have not considered the segmentation problem, nor have they considered the reliability of the network. Whether to learn the characteristics of the heart and use these features for diagnosis is not discussed.

And we believe that reliability and transparency are the most important issues that need to be solved for medical imaging problems. We believe that it is not enough to just collect data and send it directly into the network. If the entire process of the model is in a black box state that cannot be analyzed and controlled, then such a technology cannot be really promoted.

An interesting example can illustrate how important

accuracy and reliability are. Suppose the model has a detection rate of 0.8 for heart disease, ie, the probability of a positive test for a patient and a negative test result for a non-ill is 0.8. The probability of having a congenital heart disease in the population is 0.0074. A person is randomly selected from the population for testing. The probability of the person being sick when the test result is positive can be calculated by the Bayes formula as follows:

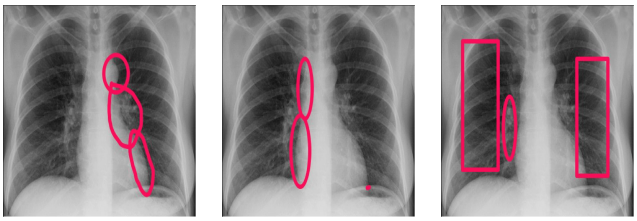
$$\begin{aligned}\mathbb{P}(A|B) &= \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)} \\ &= \frac{0.0074 \times 0.8}{0.0074 \times 0.8 + (1 - 0.0074) \times 0.2} \\ &= 0.02874903 \approx 2.87\%\end{aligned} \tag{1}$$

It can be seen that the detection rate of true positives in a model with an accuracy of 0.8 is quite low, because the overall incidence is low, unless the diagnostic accuracy is very high, the probability of true positive will be increased, and one is not reliable enough. Detection methods make it easier for patients involved in testing to suspect, oppose and reject.

We took this issue very seriously from the beginning. At the beginning we tried to distinguish one of the main categories (VSD) of normal samples and heart disease. After the fine-tune of the VGG16 model, we can get the test accuracy of 0.99. It is difficult for us. Knowing that such a good result is obtained by learning the contour features of the heart, or by learning other unrelated areas. Most of the existing studies are satisfied with the use of existing models, and have obtained a higher accuracy rate, which we believe has not solved the problem.

We used some visual methods to analyze the weights of each area of the network learning. The result is expected: the neural network is actually classified using some features of the lungs. In fact, it does not use the characteristics of the heart in the prediction. Heart disease, in fact, this is a model that looks good but not reliable enough.

In order to solve the problem of letting the model learn the true heart features, we consulted some of the doctor's diagnosis experience, manually labeled, combined with opencv for a series of pre-processing methods, labeled the heart contour, and used U-net to help segmentation The image is such that the model pays attention to the characteristics of the heart part. The image classification model is combined with the prediction.



(a) ROI of left heart region (b) ROI of right heart region (c) ROI of lung

Figure 1: ROI of diagnosis

2.2 Automatic Pipeline

In order to achieve the goal of using chest X-ray to diagnose heart disease, we have gone through many processes such as processing pictures, analyzing problems, and building models.

2.2.1 preprocessing

First of all, because the specifications of the chest radiograph are not uniform, we have processed the image into several different sizes and resolutions. At the time of specific training, because of the limited memory of the machine. We continue to resize the image based on the size of the network.

The problem we want to solve is to diagnose the patient's disease more accurately and reliably through the chest X-ray, so we think we must find a way to make the network focus on the characteristics of the heart. In the process of communicating with experts in heart disease, we already know that the characteristics of several parts of the heart occupy most of the weight in the diagnosis (Figure 1), so we hope to make the model pay attention to the characteristics of the heart.

Our images are lack of manual annotation, and if we want to use U-net to segment the contours of the heart and then focus on learning, we need to do some manual annotation. However, if we do several parts of the picture in Figure 1, we need a lot of time, so we thought of using opencv to help us solve some of the problems of the annotation. Starting from the idea of fitting the outline of the heart, we hope to be able to mark the outline of the heart. We looked for a series of methods to treat a chest map with only the outline of the heart. Because the chest image is binarized, there is a lot of interference with tissue and bone, the noise is very large, and the heart contour of a large number of pictures is not closed and does not meet the requirements. It is also unrealistic to manually mark the outline of the heart for a long time to complete.

We finally found a suitable treatment, because there is no obvious outline of the upper and lower parts of the heart, so we first manually mark the upper and lower boundaries of thousands of chest radiographs, and the rest are treated with opencv to obtain the outline of the heart, and Fit the outline with a polygon.

Steps are as follows

- binarization: Contour detection sometimes better reflects the content of the image. To perform contour detection on an image, the image must be binarized first. The binarization of the image is to set the gray value of the pixel on the image to 0 or 255, which will make the whole image appear obvious. Black and white effect.
- Expansion and Corrosion: Corrosion and expansion are the most basic operations in the morphological processing of an image. The open and closed operations that are encountered later are combined operations of corrosion and expansion operations. Corrosion and expansion are very versatile and work well: Corrosion can separate individual image elements, and expansion is used to connect adjacent elements, which is the most intuitive representation of the image after corrosion and expansion.
- Fill and Area Selection: The flood filling method is a method of filling the Unicom area with a specific color, and setting the upper and lower limits of the connectable pixels and the connected mode to achieve different filling effects. Diffuse fill is often used to mark or separate a portion of an image for further processing or analysis. It can also be used to obtain a masked area from the input image. The mask speeds up the process and the result is a continuous region. .
- polygon fitting: Use the opencv algorithm to fold a continuous smooth curve and finally achieve the purpose of approximating the polygon.

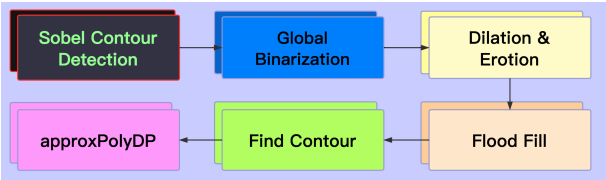


Figure 2: approximate heart contour

After such processing, each picture will generate a txt file of the corresponding vertices of the polygon for the next training use, and the coordinates of the contour need to be re-converted to regions (traces to masks) during specific training. By training U-net, we found that neural networks fulfill our requirements and can learn to accurately mark the outline of the heart. 8

The outline of the polygon is as follows:

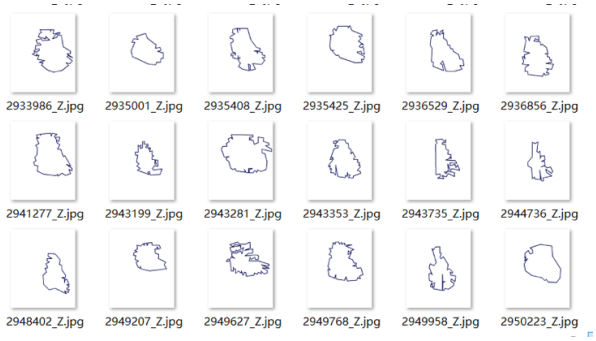


Figure 3: approximated heart contour

Since the U-net is used to shape the contour of the lateral heart with the contour of the front side, the effect is of course not good. We have manually labeled a number of lateral heart contours to improve the accuracy of the heart area of the chest segment on the side of the network.

2.2.2 Deep learning model

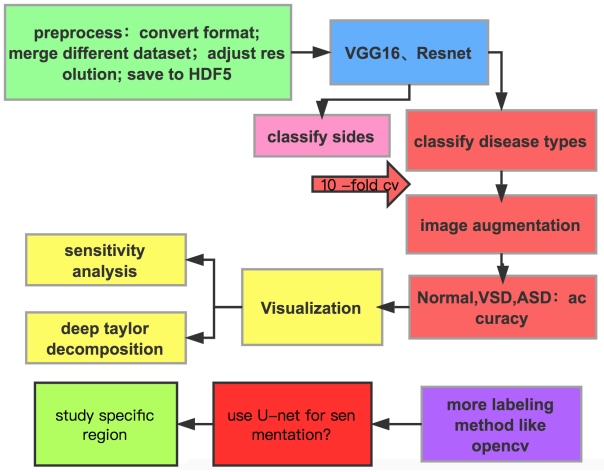


Figure 4: deep learning part pipeline

further process Because we collected data from three different sources, each of which sorted out some of the normal and sick types, and did various forms of processing (see the Preprocessing and Labeling section), including adjusting the resolution of the image, plus The

picture of the dividing line, the binarized picture, and the coordinates of the heart contour txt file (traces). In order to facilitate the training later, we have done some processing.

In the label of the picture, we directly correspond to the name of the picture plus the name of the disease. Because of the limitations of the machine memory, we have further processed the image into a 320*320 resolution in the trade-off between the number of images and the resolution of each batch of training. And use python's h5py library to convert the dataset into HDF5 format files for faster access and to group different data together for each training session.

VGG model for classification The VGG network is a network developed by Alex-net. Karen Simonyan and Andrew Zisserman of the Visual geometry group (VGG) of Oxford University wrote a paper in 14 years to discuss the importance of depth to the network; and built a depth of 19 layers. The network has achieved the first result in the ILSVRC and the second result in the classification.

VGG's goal is go deeper, and also inherits some frameworks of LeNet and AlexNet, especially like the AlexNet framework. VGG is also a convolution of 5 groups, 2 layers of fc image features, and a layer of fc classification features. According to the first five convolution groups, the different configurations in each group, the main difference between VGG and AlexNe is: use smaller filter size and interval in the first volume base layer, VGG (kernel size=3, stride=1), AlexNet(kernel size=11, stride=4), and VGG is training and testing images across the entire image and multi-scale

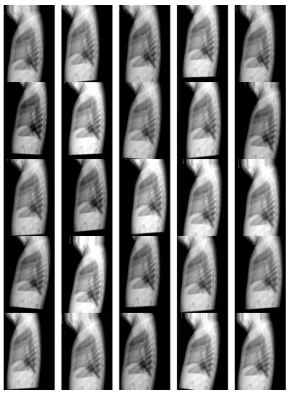
The basic structure of the VGG16 network we use is as follows:

We chose to use the VGG network first to try out the problems we faced. We used keras to define the network structure of VGG. We first made a simple distinction between the front and side of the chest radiograph to test the process. The accuracy rate is about 0.96. We think that this accuracy rate is not too high. After all, it only distinguishes the front side. There should be obvious distinguishing features.

In the following, we tried to use VGG to distinguish normal people, VSD, and ASD patients from other studies. We did not add more kinds of disease samples because of the limited number of chest radiographs we collected, so in the early attempts No samples were added to these patients. After solving the problem of learning reliability, you can consider solving the problem of small sample learning.

We also use the image transformation method, be-

(a) VGG model



(b) augmented images

cause the collected sample size is limited, after the layer screening and cleaning, the number of pictures that can be used is relatively small, we use a variety of commonly used affine transformations to increase the sample size. Taking a patient's lateral chest radio as an example, we think that some samples have been added by rotation, scaling, symmetry, translation, etc. This is after we consulted a doctor and learned that the relative positional information of the heart is not an important (factor) factor for diagnosis. We determined that this common image amplification method can be used.

We chose to use a VTG model that was pretrained on the imagenet dataset for classification and fine tune for the last three layers of the VGG network. We have also found other commonly used image recognition networks, but through our analysis, the key to this problem is not simply to use common networks to improve accuracy, so we have not stopped using various networks to find the 'highest Accuracy rate 'on.

As you can see in the results section, simply using the pre-trained VGG16 network, after fine tune, achieved "good results." But we also need to consider the question of what characteristics the model is learning.

Analysis of model performance In the difficulty and difficulty of the problem, we compared the work of some articles, almost no article analyzed what the neu-

ral network has learned, and we think that this analysis is meaningful and necessary.

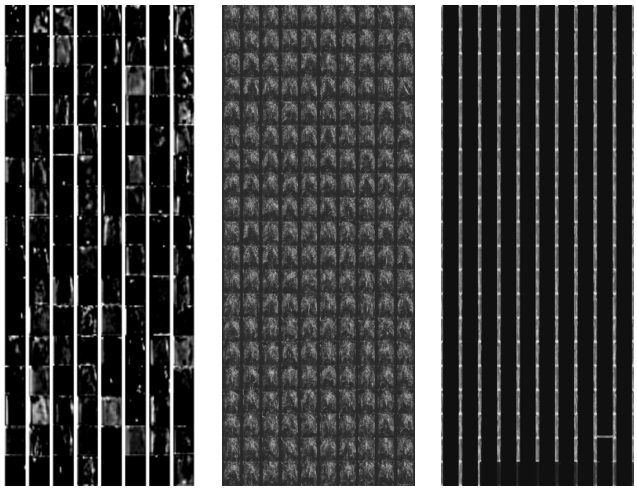
We started by trying to see what is output from each layer of the network, such as the output of the same batch of pictures in the block5_conv1 layer shown above. (Figure 6) But obviously this method does not tell us what the network is learning. Therefore, we have tried several methods to analyze what characteristics the network has learned. The first is sensitivity analysis. This is a gradient-based algorithm that calculates the sensitive relevance of an image and can see the importance of different pixels of the image. As shown in Figure 7. This approach allows us to visually analyze exactly which features become key features of e-learning and classification. The bright areas in the picture are the characteristics of the network's key learning. We found it very interesting that the network pays more attention to the area of the lungs and ignores the characteristics of the heart area. This is contrary to what we hope. Perhaps this network is more suitable for the detection of lung diseases than heart disease. Detection. Through such analysis, we can roughly conclude that it is unreliable to apply the image classification network to the diagnosis of chest X-ray.

We also tried the visual analysis method of deep taylor decomposition. Deep taylor decomposition is a kind of layer-wise relevance propagation, but we got a very strange result. As shown in Figure 8. Due to time, we have discovered through the sensitivity analysis that the neural network does learn from the facts, and there is no time to find the problem of deep taylor analysis here, but to find a way to solve the problem of incorrect learning characteristics.

U-net, Mask U-net and VGG model U-net is a deep learning network commonly used in the field of medical imaging for target detection. It can realize high-level complex features by combining low-level feature mapping to achieve precise positioning.

In 2014, Long et al. of the University of California at Berkeley proposed a Full Convolutional Network (FCN), which allows convolutional neural networks to perform dense pixel prediction without the need for a fully connected layer. This method can be used to generate image segmentation maps of any size. And this method is much faster than image block classification. Later, almost all advanced methods in the field of semantic segmentation used this model.

In addition to the fully connected layer, another major problem with semantic convolution using convolutional neural networks is the pooling layer. The pooling layer not only expands the receptive field but also aggregates the context, resulting in the loss of location information. However, semantic segmentation requires that the category map fits completely, so location information needs to be preserved. U-Net is an encoder-decoder structure. The encoder gradually reduces the spatial dimension of the pooling layer, and the decoder gradually repairs the details and spatial dimensions of the object. There is usually a quick connection between the encoder and the decoder, which helps the decoder to better fix the details of the target.



(a) intermediate output of model (b) sensitivity analysis (c) deep taylor decomposition

Figure 5

As mentioned above, U-Net is a FCN-based framework that is a full convolutional neural network with both input and output images and no fully connected layers. The shallower high-resolution layer is used to solve the problem of pixel positioning, and the deeper layer is used to solve the problem of pixel classification. However, U-Net does not add features like FCN, but concatenates to generate double-channel feature maps. , then convolution.

We used the U-net network shown in Figure 9.

We have learned some experience in chest radiography: left ventricle, right ventricle, left atrium, right atrium, large hilar, and widening of the lungs are areas of frequent concern for diagnosis. We want our model to better learn the really important features, and using U-net to segment first should be a good choice.

In the previous data preprocessing stage, we have introduced a series of processing of opencv by manual labeling, and obtained the coordinate values of the polygon vertices of thousands of heart contours, which can draw the contour of the heart. Since the polygon outline we are labeling is actually the value of the ex-

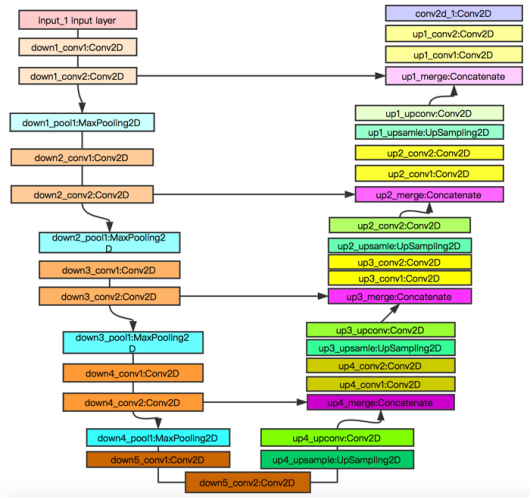
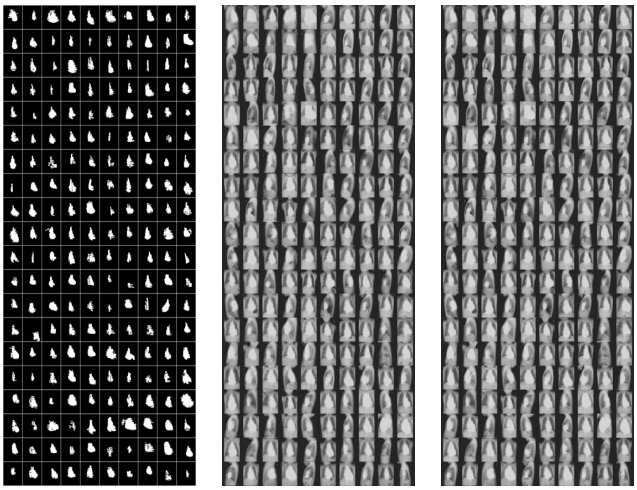


Figure 6: U-net architecture

act coordinate point, we need to convert the coordinate value into the traces to mask. Figure 8(a) is a sample of a number of samples converted from coordinate values to heart regions.

We use the original image and Mask to train the U-net network, and then use the trained U-net network to segment the unlabeled image and look for the heart area. The results are very ideal, as shown in Figure 11. First, let U-net learn the positive sample, then divide the area of the frontal heart, and also try to split the side area. The frontal segmentation effect is very good. The side of the front training network will be slightly worse, so we use manual A number of lateral heart contours are marked for training to improve the accuracy of side sample segmentation. In order to further improve the accuracy of the segmentation and reduce the influence of some strange rectangular patterns after segmentation (Fig. 8(b)), we changed the structure of U-net and spliced the pre-trained VGG network to U-net. To double the network and have more experience, this split is better than the last time. The new network structure is shown in Figure 9, and the new results are shown in Figure 8(c) above.

The heart area segmented by U-net is used as a feature to achieve better classification. We use a model called VGG+Mask, which uses U-net to select the outline of the heart, then uses the selected heart area as the Mask, and sets a fixed ratio to lower the gray value outside the heart (0.2). To achieve the purpose of letting the network focus on the heart area. Although our classification accuracy is reduced, we believe that the classification made using this method is obviously more reliable, and the improvement based on this principle makes sense.



(a) heart mask region (b) segment result using U-net (c) segment result using VGG-U-net

Figure 7

3 Performance Experiments

3.0.1 VGG classification result

Here are some of the results we have classified using VGG:

model: VGG dataset: Normal_az VSD 10-fold cross validation result

accuracy	variance	auroc	variance
0.940	0.00019629	0.988	1.94358E-05

model: VGG dataset: Normal_az VSD ASD 10-fold cross validation result

type	accuracy	variance
Overall	0.696	0.010905912
Normal	0.945	0.000448034
VSD	0.709	0.010017944
ASD	0.738	0.012352246

model: VGG dataset: Normal_indiana VSD 10-fold cross validation result

accuracy	variance	auroc	variance
0.975	0.000156748	0.999	1.15406E-06

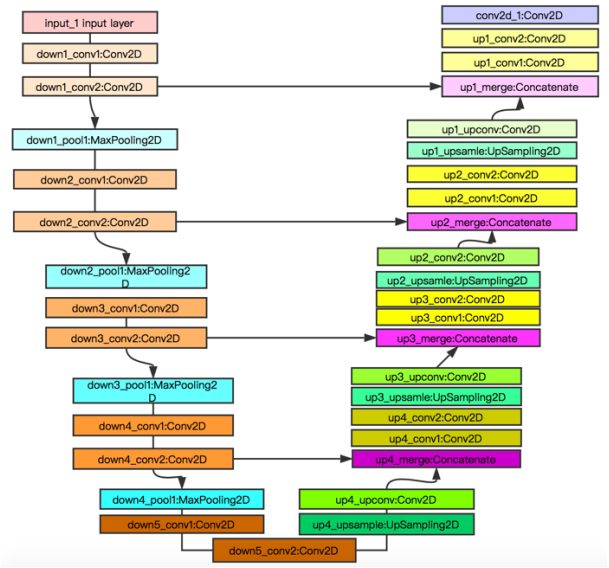


Figure 8: VGG-U-net architecture

model: VGG dataset: Normal_indiana VSD ASD 10-fold cross validation result

type	accuracy	variance
Overall	0.852	0.001002353
Normal	0.979	0.000124253
VSD	0.863	0.001000045
ASD	0.862	0.001117035

As I have analyzed before, although VGG and other studies can achieve high accuracy and auc_roc, we believe that the result is unreliable, so we used a series of preprocessing methods to obtain contours and Masks. It is used to train the U-net network and implement segmentation, and then use the image training that can segment the heart region. The results are as follows:

3.0.2 VGG16-U-net result

Here are some of the results we used to segment VGG16-U-net with mask and then use VGG16:

model: vgg16_mask_unet_vgg16 dataset: Normal_az VSD 10-fold cross validation result

accuracy	variance	auroc	variance
0.814	0.015732456	0.954	0.000215529

model: vgg16_mask_unet_vgg16 dataset: Normal_az VSD ASD 10-fold cross validation result

type	accuracy	variance
Overall	0.592	0.004522613
Normal	0.878	0.000124253
VSD	0.635	0.0037776
ASD	0.671	0.007881146

model: vgg16_mask_unet_vgg16 dataset: Normal_indiana VSD 10-fold cross validation result

accuracy	variance	auroc	variance
0.943	0.000708983	0.996	3.45306E-06

model: vgg16_mask_unet_vgg16 dataset: Normal_indiana VSD ASD 10-fold cross validation result

type	accuracy	variance
Overall	0.792	0.004999465
Normal	0.916	0.00347655
VSD	0.819	0.002836416
ASD	0.851	0.000999533

It can be seen that under such a network design, accuracy has declined, but we believe that such a decline is more real and more reliable, and optimization on this basis is the right direction. We will also discuss some new models and new ideas that may solve our problems.

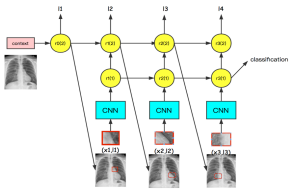
4 The Conclusions

We believe that the use of chest X-ray to diagnose heart disease is a very meaningful and challenging topic. The simple model used in previous studies cannot necessarily solve the problem. We have used a method that can analyze the characteristics of model learning and found it. Some ideas on how to solve the problem. The set of pre-processing plus U-net methods we use may solve the problem partially, and there are more models and methods worth trying.

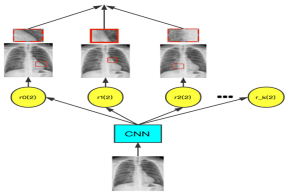
5 Future Work

We believe that in the use of chest data to diagnose heart disease, we have achieved better results than previous studies by using better data pre-processing and labeling, using better networks, but there are still many things. Waiting for us to continue exploring.

We think there are some models and methods that are worth trying in the future, such as models that try to improve image resolution, models for removing



(a) RNN-CNN architecture



(b) RCNN based model

bones and lung shadows. We also want to design an RNN-CNN network to implement the functions of the network to learn specific areas.

We believe that it is possible to carefully process parts of the heart contour every time, through the introduction of RNN. This is a bit like the attention mechanism in natural language processing. We think that by designing the network, the network can learn to start from a certain position and learn important features in a certain order - the outline of the heart, because only learning is small at a time. A region, and ultimately only a small amount of useful areas, the RNN + CNN method will not reduce the computational speed due to the introduction of RNN, but may lead to an increase in computing speed.

The following is a schematic diagram of the structure of the network we designed:

In the RNN-CNN network, CNN is mainly used for feature extraction of images, and transforms an overall image or a partial image into a vector with certain geometric transformation invariance. Each unit of the RNN uses the attention mechanism to find the most relevant local region for the classification from the overall image, and then extracts the features of the region image with CNN. Each RNN unit needs two aspects of information when selecting the area of focus: the state of the previous step RNN and the context that the CNN extracts from the overall image. The RNN terminates after finding all the areas of focus. Finally, the output generated by each step of the RNN is weighted by the fully connected network to give the final classification prediction.

In our question, we already know that doctors will focus on several areas of the heart contour, so we can focus on the key areas and achieve classification through the RNN-CNN network. Because the RNN-CNN network can achieve the selection of specific small areas, we can understand which areas of the model are learned, so the reliability of the model is guaranteed.

There are several advantages to using the RNN-CNN network to solve our medical image diagnosis problems. First, each step of the RNN requires only a small portion of the image information, and the amount of computation required is less than the computational amount of the overall image, so that higher resolution images can be input and the details in the image can be fully utilized. Second, the RNN can output the image regions and sequences that are focused on at each step, so it is easier to interpret the classification process of the model than to use a complete image for classification. For medical diagnosis, if the model's interpretability allows doctors to combine their own expertise and experience to make more reliable judgments. In addition, interpretability is also very helpful in model optimization. Third, in the model training process, the area of interest for each step can be prompted by the manually labeled image, and the weight of each partial region can also be set by the prior knowledge of the existing medical image. This way, the model can also give a more accurate prediction for small sample data.

We think that RNN-CNN can try several models, one is to give nothing, let the model completely rely on its own learning characteristics, as long as it can complete the classification task, it is a good model; the second is to do the annotation, how to find the model learning The marked part is predicted again; the third is not only the part of the label, but also the a priori weight, but does not strictly limit the range of weights. The second and third models may not be implemented without RNN, with a fixed model. The second model is somewhat similar to RCNN (Fig. 10(b)), which is to find K regions, select them with k rectangular boxes, weight them to learn, and finally summarize which diseases are predicted.

Another question to consider is whether the learning order of different small areas is important. It is different from human handwritten numbers or drawing pictures. It is not necessary in a certain order in heart disease diagnosis. We can consider these two situations and choose The result is a better model.

In short, the work on this topic is far from over, and there are still many things that can continue to be explored in the future.

References

[1] Sema Candemir, Stefan Jaeger, Wilson Lin, Zhiyun Xue, Sameer Antani, and George Thoma. Automatic heart localization and radiographic index computation in chest x-rays. In Medical Imaging 2016: Computer-Aided Diagnosis, volume 9785, page 978517. International Society for Optics and Photonics, 2016.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[3] Mohammad Tariqul Islam, Md Abdul Aowal, Ahmed Tahseen Minhaz, and Khalid Ashraf. Abnormality detection and localization in chest x-rays using deep convolutional neural networks. arXiv preprint arXiv:1705.09850, 2017.

[4] Atul Kumar, Yen-Yu Wang, Kai-Che Liu, I-Chen Tsai, Ching-Chun Huang, and Nguyen Hung. Distinguishing normal and pulmonary edema chest x-ray using gabor filter and svm. In Bioelectronics and Bioinformatics (ISBB), 2014 IEEE International Symposium on, pages 1–4. IEEE, 2014.

[5] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. Medical image analysis, 42:60–88, 2017.

[6] Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, and Klaus-Robert Müller. Explaining nonlinear classification decisions with deep taylor decomposition. Pattern Recognition, 65:211–222, 2017.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.

[8] MJ Yaffe and JA Rowlands. X-ray detectors for digital radiography. Physics in Medicine & Biology, 42(1):1, 1997.