Deep learning SPECT lung perfusion image classification method based on attention mechanism

Sitao Zeng¹,², Yongchun Cao²*, Qiang Lin², Zhengxing Man², Tao Deng², Rong Wang³

¹Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou 730030, Gansu, China
²Key Laboratory of Streaming Data Computing Technologies and Application, Northwest Minzu University, Lanzhou 730030, Gansu, China
³Department of Nuclear Medicine, Gansu Provincial Hospital, Lanzhou 730030, Gansu, China
Email:15713531294@163.com

Abstract: SPECT lung perfusion is an important functional imaging technology. It can capture the functional lesions of the lung in a non-invasive manner and has become an important clinical detection method for diseases such as pulmonary embolism. In order to realize the automatic detection of the degree of pulmonary embolism, this paper studies and constructs a deep classification model based on the attention mechanism. First, the normalization technique is used to convert the original lung perfusion file into a SPECT image; secondly, in view of the over-fitting phenomenon of the deep learning model caused by the small amount of medical image data and the unbalanced data, the image translation and rotation techniques are used to perform effective expansion; then, in order to improve the model's feature extraction ability, the attention mechanism is combined with the depth classification model to build a SPECT lung perfusion image classification model; finally, a set of real SPECT lung perfusion images were used to carry out comparative experiments on various depth classification models. The experimental results show that the model proposed in this paper can effectively detect the extent of lung disease lesions, and the classification accuracy rate exceeds 88%, which verifies the effectiveness and reliability of the classification model.

1. Introduction
Pulmonary Embolism (PE) is a kind of pulmonary vascular disease caused by various emboli blocking the main pulmonary artery and its branches. The case fatality rate can reach 3.5%~25%. When the patient experiences shock, the mortality rate is as high as 58%[1]. Because the clinical manifestations are usually relatively insidious, different clinical manifestations appear with the degree and time of blockage of the pulmonary artery and its branches, leading to a higher rate of misdiagnosis. Therefore, early diagnosis is extremely important for patient treatment.

In recent years, with the development of biology and imaging, the early diagnosis of PE has been greatly improved. Biological methods include cytopathology examination, bronchial endoscopic fluorescence cell examination and molecular pathology examination. Imaging methods are divided into X-ray examination, Computer Tomography (CT), and Magnetic Resonance Imaging (MRI) and SPECT radionuclide perfusion imaging. Compared with biological examination methods, imaging methods are...
a non-invasive diagnosis method that can minimize the harm to patients. Structural imaging methods such as X-ray, CT, and MRI may have the disadvantages of being blocked by other tissues and affected by radiation in lung examination, while SPECT radionuclide perfusion imaging can diagnose diseases based on the distribution characteristics of pulmonary blood flow damage and to judge the severity of the disease, it is easier to show the lesions caused by micro thrombosis. Therefore, it is widely used in the clinical examination of PE[2].

However, a large number of imaging images consumes the doctor's time and requires high professionalism. Secondly, due to the influence of subjective factors such as work status, personal experience and knowledge level, coupled with the complexity of the lung tissue structure, misdiagnosis and missed diagnosis are very easy to occur. With the development of deep learning technology, Computer Aid Diagnosis (CAD) has become a research hotspot in medical imaging. Since Convolutional Neural Network (CNN) has achieved excellent results in image classification, deep learning is widely used in medical lung image classification. Shen et al.[3] proposed a multi-scale convolutional neural network in the benign and malignant classification of lung nodules, and achieved good classification results. Liauchuk et al.[4] used CT images of 13 pulmonary tuberculosis patients to train the GoogLeNet model, and the AUC results obtained 0.969. Zhu et al.[5] designed a fully automatic lung image computer-aided diagnosis system, which proved the advantages of deep learning technology in lung image classification. Li et al.[6] proposed a complex contour convolutional neural network for the classification of lung nodules, and the classification accuracy reached 86.03%. At present, researchers use deep learning technology to do a lot of work on lung CT, X-ray and MRI and other structural images and have achieved good results. However, there are few researches on lung SPECT functional image recognition based on deep learning technology.

According to the defect area of SPECT lung perfusion image[7], the degree of lung lesions can be divided into four types: grade 0, grade 1, grade 2 and grade 3. In order to efficiently and accurately detect the degree of lung lesions, this paper studies the multi-class recognition of SPECT lung perfusion images, and proposes a multi-class recognition model of lung perfusion images that combines the Attention Mechanism (AM) with the VGGNet model. It aims to improve the classification effect of SPECT lung perfusion images to better meet the needs of clinicians in assisting diagnosis.

2. Lung perfusion medical image data

2.1. Data Normalization

SPECT lung perfusion imaging data is initially stored in DICOM file. Each component value in the file is a digital record of the amount of radionuclide or isotope (16-bit unsigned integer) pre-injected into the patient's body, not a pixel value. The DICOM file radiation value generated by SPECT lung perfusion examination not only varies widely, but also has obvious individual differences. Therefore, it is necessary to transform the emission value before model classification. Specifically, this paper uses Min-Max standardization to perform the following linear transformations on DICOM files:

$$\text{norm} = \frac{x - \text{min}(x)}{\text{max}(x) - \text{min}(x)}$$  \hspace{1cm} (1)

Where $x$ represents the image pixel value, and max(x) and min(x) represent the maximum and minimum values of image pixels, respectively.

2.2. Data Status

The all data collected from the diagnosis of various physiological diseases from SPECT lung perfusion imaging using Siemens SPECT-ECAM equipment in Gansu Provincial People's Hospital from January to December 2018. The patient stays in a supine position and injects slowly 99Tcm-MAA 117 ~ 185MBQ (3 ~ 5mCi) 5 minutes later, 360° plane static imaging was performed. The data storage format followed the Digital Image and Communications in Medical (DICOM) protocol. All images are normalized from the original DICOM format and converted to PNG images, contains SPECT lung
perfusion images of 197 patients (including 55 cases of level 0, 74 cases of level 1, 36 cases of level 2 and 32 cases of level 3). The entire data set contains four views of lung perfusion, namely Anterior, Left Lateral (LLAT), Posterior and Right Lateral (RLAT). The image size is 128×128, all of which are single-channel. The image is shown in Figure 1 below.

2.3. Data Preprocessing
The classification of lung perfusion images based on deep learning not only relies on a good classification model, but also on the number of images. Insufficient sample size will lead to problems such as over-fitting and poor generalization ability of the model. Because medical images involve patient privacy and require manual labeling by imaging experts, it is difficult to obtain large-scale labeled lung perfusion image data sets. Therefore, this paper expands the data based on the existing data, and sequentially performs a moderate translation and a random rotation of 0-15°. However, considering the possible sample imbalance in the original data set, unequal expansion between classes is introduced when the data is expanded.

In order to ensure the accuracy of data labeling, the data in this paper was completed under the guidance of three nuclear doctors based on the diagnosis report. The lungs are divided into left and right lungs. The right lung is composed of the upper lobe, the middle lobe and the lower lobe, and is divided into 10 lung segments, and the left lung is composed of the upper lobe and the lower lobe, and is divided into 8 lung segments. The images of different body positions show the lung segments incomplete, the following figure 2 (Anterior position) is a brief description: 1-3 constitutes the right upper lobe, which is the apical, posterior and anterior segment; 4-5 constitutes the right middle lobe, which is the lateral and medial ends; 10 constitutes the right lower lobe, which represents the anterior basal segment; 11-14 constitutes the left upper lobe, which is the posterior apical segment, the anterior segment, the upper lingual segment, and the lower lingual segment; 16 constitutes the left lower lobe, which represents the anterior medial bottom segment. According to SPECT images, lung perfusion defects are divided into four grades: grade 0, no perfusion defect or insignificant defect; grade 1, defect in surrounding local area; grade 2, defect area up to a lobe; grade 3, defect area more than a lobe. This paper takes the Anterior position as an example to introduce: level 0, no perfusion defect or insignificant defect in each lung segment on the left and right; level 1, you can see 1/3 of the defect around the fourth lateral segment of the right middle lobe; level 2, the fourth right lobe There is a 1/2 defect area in the lateral segment and the 12th anterior segment of the upper left lobe, and there is a 1/4 defect in the third anterior segment of the right upper lobe. The total defect area reaches a lobe; grade 3, the first apical segment of the right upper lobe and the third anterior segment, and the right middle lobe The 4th and 5th medial segments, the 10th anterior basal segment of the right lower lobe, and the 16th anterior medial basal segment of the left lower lobe are almost completely defective, the 2nd posterior segment of the right upper lobe, the 11th apical segment posterior and the 12th anterior segment There are partial defects in the 13th upper tongue segment and the 14th lower tongue segment, and the total defect area exceeds a lobe.
3. VGGNet model based on attention mechanism

3.1. Attention mechanism

In recent years, attention models have been widely used in various fields of deep learning, and good results have been achieved in visual question answering[8-9] and image detection[10-11]. By using attention mechanism, important information can be obtained more quickly and accurately, ignoring useless information. At present, many researchers have added the attention mechanism module to the neural network and have achieved good results.

The attention mechanism training process is divided into hard attention mechanism and soft attention mechanism. Since soft attention mechanism can realize end-to-end automatic training through network gradient and feedback, it is widely used in image classification tasks.

In the soft attention mechanism, according to different attention domains, it is divided into spatial domain and channel domain. The channel attention mechanism optimizes the feature classification effect by extracting the importance of different channel features to key information. By mining the dependency between the channel graphs, the importance of each characteristic channel is obtained, and selectively focus on information with large weights according to importance. The channel attention module first uses maximum pooling and average pooling to transform the feature map into two $1 \times 1 \times C$ features, then inputs the feature map to a perceptron containing a hidden layer for dimensionality reduction and dimensionality upgrade operations, then the two outputted feature graphs are summed up, and the channel attention weight is obtained using the sigmoid function. The channel attention module structure is shown in Figure 3.

![Channel attention module structure diagram](image)

The spatial attention module focuses attention on the spatial location information of the key features. By assigning different weights to the location information of the features, the network learns useful feature information for classification tasks. The spatial attention module first takes the output feature map of the channel attention module as input, performs maximum pooling and average pooling in the channel dimension, and then concatenates the two channel information into a feature map, feature extraction is further performed on a convolution layer with a convolution kernel size of 7, and finally passes through the sigmoid function to obtain the spatial attention feature map. The structure of the spatial attention module is shown in Figure 4.
3.2. **AM-VGGNet model**

The VGGNet model[13] was proposed by Visual Geometry Group of Oxford University, and won the second place in the ILSVRC2014 classification task. VGGNet is developed by AlexNet. It uses three 3 × 3 convolution kernels instead of one 7 × 7 convolution kernel. As the depth of the model increases, it has better resolving power. However, the number of VGGNet network layers continues to increase. On the one hand, it puts forward higher requirements for training samples, training time and memory. On the other hand, the maximum pooling layer in the VGGNet model will cause loss of spatial information. In response to the above problems, Sahil[13] proposed the model structure of sort_pool2d to improve the maximum pooling layer; Yu Lichun et al.[14] changed the maximum pooling layer in VGGNet to mean pooling. Although they have solved the defects of VGGNet to a certain extent, they have not achieved good results in the field of SPECT lung perfusion medical imaging. Therefore, in view of the characteristics of the data in this paper, proposes AM-VGGNet based on the attention mechanism, using the VGGNet model as the backbone network, and then adding the attention mechanism to the model. The attention mechanism module can effectively improve the model's sensitivity to features, so that the model can capture subtle differences in features of different degrees, thereby improving the classification performance of the model. The model structure is shown in Figure 5.

![Figure 4: Spatial attention module structure diagram](image)

![Figure 5: AM-VGGNet model structure diagram](image)

4. **Experimental design and result analysis**

4.1. **Experimental design**

After the data expansion operation of SPECT lung perfusion images, a total of 6304 images are obtained. The training set and the testing set are divided according to the ratio of 8:2. The specific division is shown in Table 1.

| Grade   | Training set | Testing set | Total  |
|---------|--------------|-------------|--------|
| Grade 0 | 1408         | 356         | 1764   |
| Grade 1 | 1352         | 336         | 1688   |
| Grade 2 | 1104         | 276         | 1380   |
| Grade 3 | 1180         | 292         | 1472   |
| Total   | 5044         | 1260        | 6304   |
In order to verify the effectiveness and superiority of the model proposed in this study, the classification effect of AM-VGGNet and DesNet121, GoogLeNet, InceptionV4, ResNet50, ResNext50, SE-ResNext50 and VGGNet are compared and analyzed. Train each model from scratch on the SPECT lung perfusion data set until the model converges, and ensure that each model is trained under the same experimental conditions. In the training process, the training data is tested with the test set every iteration cycle, and the classification accuracy is output and recorded.

The learning rate is one of the most important parameters in model training. Too high and too low learning rates may have an adverse effect on the results of the model. In order to ensure the effect of the experiment in this paper, we introduced the learning rate optimization function-Warmup. At the beginning of model training, a smaller learning rate is selected, after training for a period of time, use the preset learning rate to learn. In this paper, the initial learning rate is set to 0.01, when the training period is greater than 50 cycles, the learning rate is set to 0.0001.

Since the diagnosis report is based on a comprehensive analysis of the four views of the patient, this paper changes the model structure. In this paper, each patient has four views of images. We stitched the patient's images into four-channel data in the order of Anterior, LLAT, Posterior, and RLAT as input to the model. Therefore, set the number of input channels for the first convolution of the model to 4, set the Average pooling parameter to 1x1, and finally set the model's last fully linked category parameter to 4.

4.2. Analysis of Experimental Results

In order to better evaluate the classification performance of different models, this paper uses Precision, Recall and Accuracy as indicators for evaluation and comparison. When evaluating multi-category problems, the multi-category is usually decomposed into multiple two-category problems. Each time one of the categories is regarded as a positive category, and the other categories are unified as a negative category. The indexes of each category are calculated respectively, and finally the evaluation indexes of multiple categories are calculated on average.

Accuracy is a common index to evaluate the accuracy of classification model. Its value indicates the proportion of the number of samples with accurate classification to the total number of samples. The calculation formula is shown in (2). The closer the general accuracy is to 1, the better the classification effect is.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

(2)

Precision rate refers to the proportion of samples predicted to be positive in real classes, and the calculation formula is shown in (3).

\[
\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}
\]

\[
\text{Precision}_{\text{macro}} = \frac{\sum_{i=1}^{L} \text{Precision}_i}{|L|}
\]

(3)

Recall rate refers to the proportion of samples that are actually positive classes that are predicted to be positive classes. The calculation formula is shown in (4).

\[
\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}
\]

\[
\text{Recall}_{\text{macro}} = \frac{\sum_{i=1}^{L} \text{Recall}_i}{|L|}
\]

(4)

Micro-F1 is the harmonic mean of the precision rate and recall rate. The calculation formula is shown in (5).
By comparing eight classic CNN models in Table 2, it can be found that for the classification task of SPECT lung perfusion image lesions, the VGGNet model has relatively better classification effect and training time than the other seven models. Although models such as DesNet121, ResNext50 and SE-ResNext50 have deeper network layers and slightly higher classification accuracy than the VGGNet model, the deep model makes the training time longer, and the amount of data for each type of SPECT lung perfusion image is not large enough, the model easily falls into overfitting. The VGGNet model not only has relatively shallow layers, but also has relatively few model parameters and training time. Therefore, in order to verify the effectiveness of the method in this paper, an attention mechanism is introduced on the basis of the VGGNet model for comparative analysis.

As shown in Table 2. It can be seen that various neural networks can accurately distinguish the degree of lesions in SPECT lung perfusion images. Among them, the VGGNet based on the attention mechanism proposed in this paper has the best classification effect, and its classification accuracy rate reaches 88.1%. Compared with the DesNet121, ResNext50, ResNext20 and SE-ResNet models with a deeper network structure, the accuracy is more than 3%. In VGGNet, GoogLeNet and InceptionV4 with relatively shallow network layers, the classification accuracy of AM-VGGNet has increased by 5%-10%. Compared with the original VGGNet model, the classification accuracy is improved by 5.5%. The introduction of the attention mechanism enhances the expressiveness of the features, thereby improving the classification effect of the model.

Only one classification accuracy rate for evaluating the quality of a classification model cannot fully explain the effectiveness of the model classification. Therefore, this paper calculates the macro average precision rate, the macro average recall rate and the Micro-F1 value of the multi-classification evaluation index. As can be seen from Table 2, the Precisionmacro of the AM-VGGNet model has achieved an improvement of 2%-13% relative to the classic neural network model, and the Recallmacro has achieved an improvement of 1%-5% relative to the classic neural network model, Micro-F1 value has achieved a 3%-9% improvement over the classic neural network model. Compared with the original VGGNet model, Precisionmacro has increased by 4.1%, Recallmacro has increased by 4.4%, and Micro-F1 has increased by 4.3%. The improvement of these index values further illustrates the good classification performance of AM-VGGNet.

As shown in Figure 6. The abscissa is the iteration period, and the ordinate is the loss value and classification accuracy. In this experiment, set the same epoch parameters and observe the loss value and Acc change trend of different classification models. It can be seen from the figure that in the first 50 iteration cycles, the accuracy and loss curves of all models fluctuate relatively large, and the overall trend is upward; as the number of training increases and the Warmup learning rate optimization function is introduced, the accuracy of each model has begun to improve significantly, and gradually stabilized.
The Acc curve of the AM-VGGNet model has almost reached the state of convergence at 50 epoch, while the Acc curve of VGGNet is still in the rising state and reached a stable state at 120 epoch. When the other seven models, such as Desnet121 and GoogLeNet, were trained to 50 epoch, the Acc curves were all in the upward state, and none of them reached the state of convergence. AM-VGGNet has more than doubled the convergence speed of VGGNet. This is because the introduction of the attention mechanism strengthens the model's ability to extract features of SPECT lung perfusion images.

Figure 6 The change curve of loss and Accuracy of eight models
5. Conclusion
This paper aims at multi-class classification of SPECT lung perfusion images. On the basis of VGGNet, a lung perfusion medical image classification model AM-VGGNet with an attention mechanism is proposed. The specific work includes, first, normalize the data according to the characteristics of the SPECT radionuclide lung perfusion data, and then use the data expansion method to improve the problem of the small amount of medical data; finally use VGGNet and introduce the attention mechanism to construct the classification model, which improves the model's ability to extract features. The experimental part is compared with a variety of classical neural networks. The experimental results show that the AM-VGGNet lung perfusion image classification model proposed in this paper can improve the classification performance of SPECT lung perfusion images, it can effectively assist doctors in the diagnosis of lung diseases.

Acknowledgments
This study was funded by the National Natural Science Foundation of China (61562075), the Gansu Provincial First-class Discipline Program of Northwest Minzu University (11080305), and the Program for Innovative Research Team of SEAC ([2018] 98), Gankeji([2020] 11) Natural Science Foundation (20JR5RA511).

Recommend reason:
At present, researchers have made good achievements in CT, X-ray and MRI images using depth learning technology. However, under the category of SPECT nuclear medicine imaging modes, it is rarely found to construct lung disease identification based on deep learning. On the one hand, it can overcome the phenomenon of inconsistent diagnosis results caused by differences in clinical experience and subjective understanding of doctors and related practitioners; On the other hand, it can improve the efficiency and accuracy of doctors' examination, thus assisting doctors in the diagnosis of lung diseases.

References
[1] Wood KE. The presence of shock defines the threshold to initiate thrombolytic therapy in patient with pulmonary embolism[J]. Intensive Care Med. 2002, 28(11):1537 - 546.
[2] Wang H, Li G H, Fu X H, et al. Evaluation of thrombolytic therapy and anticoagulant therapy by pulmonary perfusion imaging in patients with acute pulmonary thromboembolism[J]. Digest of the world's latest medical information, 2015(28): 159-161 (in Chinese).
[3] Shen W, Zhou M, Yang F, et al. Multi-scale convolutional neural networks for lung nodule classification[C]. International Conference on Information Processing in Medical Imaging. Springer, Cham, 2015: 588-599.
[4] Liauchuk V, Kovalev V, Kalinovsky A, et al. Examining the ability of convolutional neural networks to detect lesions in lung CT images[C]. Journal of computer assisted radiology and surgery, CARS-2017 International Congress, Barcelona. 2016: 20-25.
[5] Zhu W, Liu C, Fan W, et al. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification[C]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018: 673-681.
[6] Lingling Li, Liyuan Ma, Licheng Jiao, et al. Complex Contourlet-CNN for polarimetric SAR image classification[J]. Pattern Recognition, 2019, 192: 1-39.
[7] Sun H F, Wang Z T, Zhou T, et al. Evaluation of lung function with SPECT/CT lung perfusion imaging and the potential factors of perfusion defects in patients with non-small-cell lung cancer[J]. Chinese Journal of Nuclear Medicine and Molecular Imaging, 2016, 36(4):322-325 (in Chinese).
[8] Shih K J, Singh S, Hoiem D. Where to look: focus regions for visual question answering[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 4613-4621.
[9] Yang Z, He X, Gao J, et al. Stacked attention networks for image question answering[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 21-29.
[10] Ji C, Liu H Y, Sun J F, et al. Image salient region detection based on spatial and frequency domains[J]. Journal of Jilin University: Engineering Edition. 2014, 44(1): 117-183(in Chinese).
[11] Dong C, Liu J H, Xu F, et al. Fast ship detection in optical remote sensing images[J]. Journal of Jilin University: Engineering Edition, 2019, 49(4): 1369-1376(in Chinese).
[12] Sahil S. A new kind of pooling layer for faster and sharper convergence[EB/OL]. [2020-07-25]. https://github.com/singlasahil14/sortpool2d/blob/master/sortpool2d_test.py.
[13] Yu L C, Liu J Q, Face recognition based on improved VGGNet algorithm[J]. Journal of Changchun University of Technology, 2018, 39(4): 378-385(in Chinese).