Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/4/2023416

Comparison of deep learning algorithms using non-local methods for lung nodule classification

Peiyu Wang

School of Computer Science, University of Nottingham, Ningbo, China, Zhejiang, 315000, China

scypw1@nottingham.edu.cn

Abstract. Lung cancer has been identified as a serious and fatal disease due to its high morbidity and mortality. It is of vital importance for lung cancer patients to obtain early detection of the disease so that the later treatments may bring good effects. Lung computed tomography (CT), as a normal method to diagnose the disease, can be used to recognize typical lung cancer, but it is possible to confuse cancer with some other diseases, such as innocent tumour and phthisis. Therefore, an accurate diagnostic tool is required to help clinical disease recognition. Machine learning (ML), especially deep learning (DL) is an ideal technique to classify CT images thanks to the great capability of image processing and feature recognition. However, the task for ML faces several challenges that have made a negative difference in the accuracy of algorithms. The main problem is that lung nodules can have heterogeneous sizes and shapes varying in a wide range, thus both local and global features of data should be considered to enhance the classification results. In this work, the author investigated advanced works in the territory of lung cancer identification using ML and the comparison results of newly proposed models and proper analysis are provided. In addition, possible future improvements are discussed.

Keywords: Computer Vision, Deep learning, Non-local network, Computer-Aided Diagnoses, Lung Nodule classification.

1. Introduction

The high mortality of lung cancer has led to its long-standing classification as a severe type of cancer. Typically, lung cancer does not cause symptoms until it has extended to other body parts or beyond the lungs. This suggests that the prognosis is less favourable than it is for many other cancer types. After being diagnosed, one in three people with the illness survive for at least a year, and one in twenty survive for at least ten years [1]. In these circumstances, early diagnosis of this illness is thought to be important since it can increase the efficacy of the available treatments and the likelihood of survival.

However, the diagnosis is normally accomplished by doctors through observing CT images, which is defective since it is time-consuming and prone to confusion with other diseases. Computer-Aided Diagnoses (CAD) is therefore regarded as a potential way to provide early detection [2]. One of the most successful methods for assisting doctors in this area is deep learning-based CAD.

In this work, lung nodule classification is mainly considered. Convolutional Neural Network (CNN) is now integrated into many computer vision works and the capability of it also increases as the structure of the network becomes deeper. From the LeNet [3] to VGG [4] and ResNet [5], the networks now can have hundreds of layers. But the classification accuracy did not increase as much as the number of layers. As of the writing date, the highest accuracy, which is 95.28%, is reached by ProCAN [6] model. This can be caused by several factors. First, the dataset used (LIDC-IDRI [7]) for the task is small-scale compared to other tasks. On the other hand, the size and shape of lung nodules vary substantially, and sometimes it is hard to identify whether a nodule is benign or malignant.

To solve the problem, methods that may take more features of data into consideration are introduced. In 2018, DeepLung [8] model extract features by using a 3D dual-path network and separate the task into nodule detection and classification. The experiment result of DeepLung shows its advanced capabilities in both nodule and patient, which reached an accuracy of 90.44%. In addition, Local_Global [9] used a network with the ability to capture both local and global features and brought an accuracy of 88.46%. One recently proposed model called ProCAN [6] improved the non-local mechanism in Local_Global by utilizing channel attention and integrated progressive learning in the training session has increased the accuracy to 95.28%. These existing results show that such a method augmenting the effects of original data can enhance the prediction ability of neural network models.

The work mainly considers and compares Local_Global and ProCAN models on both the feature extraction mechanism and other techniques used. The author also analyzed the possible feasibility and reason for the improvements brought by these mechanisms. The research is conducted as the following:
1. a summary of the current state of knowledge and issues in the area of lung nodule identification and categorization
2. analyze and introduce the specific structural design of Local_Global and ProCAN
3. study and compare the performance of each model, analyze the reasons why good models are able to produce high accuracy, and provide future improvement trends

2. Description of Local_Global

2.1. Residual blocks

Local-Global [9] is a neural network model used for pulmonary nodule classification using residual block and non-local network. Residual Block was first used in ResNet [5] and is now widely used in computer vision problems solving, especially integrated in Convolutional Neural Networks (CNNs). Specifically, several existing works for categorizing lung nodules adopted this network architecture [10]. CNN has been developing in the direction of "Deep", which means the number of layers of a network experienced considerable growth from LeNet [3] with 5 layers to VGG [4] with 19 layers. In ResNet where a residual block is proposed, the number of layers of CNN is enlarged to 152. However, as the networks become deeper, gradient vanishing and exploding problems appear and lead to the optimization of the network harder and harder. To make training deep networks possible, residual network rises. This architecture transfers the optimization objective from the desired mapping output to another function as formula 1:

$$F(x) = H(x) - x \tag{1}$$

From Figure 1 we can see that residual blocks are used to connect non-local blocks in the network. For simplicity in computing, the kernel size of the residual block is selected to be 3×3. Figure 2 shows that the input x is passed to the output of two convolutional layers, namely $F(x)$ and the task is to fit for a suitable $F(x)$ using the two layers.

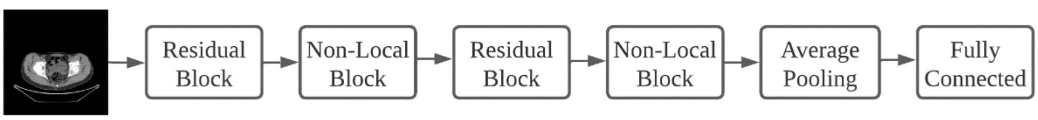

Figure 1. The overall network architecture of Local-Global.

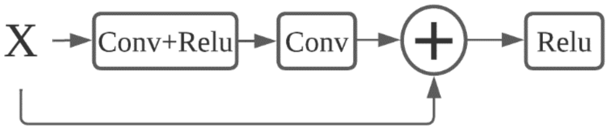

Figure 2. Residual block in Local-Global network.

2.2. Non-Local network

In the Local-Global model, non-local neural networks [11] are used to obtain global features. The mechanism, sometimes called Self-Attention Layers [12], aims to amplify the respective field in convolutional operation in CNN. Normally, the CNN can only extract features in the scope of kernel size, which is believed to be local and restrictive. Since nodule size can vary greatly, it is essential to consider the overall characteristics of the picture when classifying pulmonary nodules. The non-Local block initially leverages a linear convolution of the sample data, as seen in Figure 1. Each of these convolutions has a kernel size of 1. By this means, dimension reduction of original input is implemented so that later computation for the non-local network is simplified. These linear transformations provide the values $f(x)$, $g(x)$ and $h(x)$. Then, using Equation (2), the fundamental notion of non-local networks is applied, which is essentially matrix multiplications between the features:

$$x + h(x)Softmax\left(f(x)^T g(x)\right)^T \gamma \tag{2}$$

Figure 3 is an illustration of how the matrix computation is organized. The advantage of matrix multiplication is that it makes it possible for the network to gather general spatial characteristics through non-linear interaction. Without the need for parameters, every element in $f(x)$ is multiplied by every element in $g(x)$. Values are then mapped into the 0–1 range, and the result is processed using the $Softmax$ function to establish that the total of each feature map is equal to 1. The network may concentrate on certain $h(x)$ locations using these feature maps as attention masks. Multiplying $h(x)$ by the output of $Softmax$ causes the regions receiving little attention to be deleted and therefore free the next layer from the calculations of these features. γ in Equation (2) is a changeable argument that is used to control the influence of the non-local layer. In addition, dropout regularization is used to avoid overfitting.

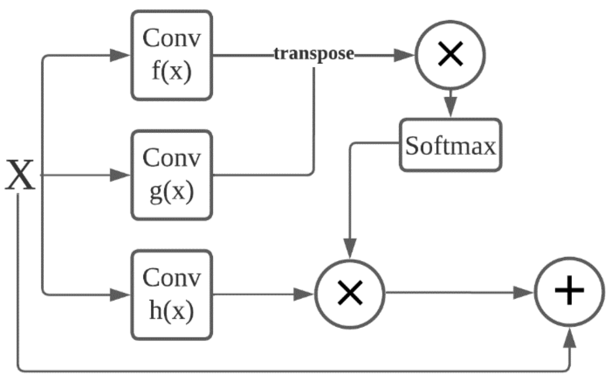

Figure 3. This figure shows how the feature matrix is calculated.

3. Description of ProCAN

ProCAN [6] is a new model proposed recently. The model adopts a strengthened non-local network which is called a channel attentive non-local network and a progressive growing method for the

DOI: 10.54254/2755-2721/4/2023416

training process. Given an accuracy of 95.28% and an AUC of 98.05%, the model achieved the state-of-the-art standard on the LIDC-IDRI dataset.

3.1. Non-local network with channel attentive mechanism

ProCAN uses similar non-local processing methods for sample images with Local-Global and adds a Channel attentive operation to consider channels of the image. Figure 4 demonstrates how the channel

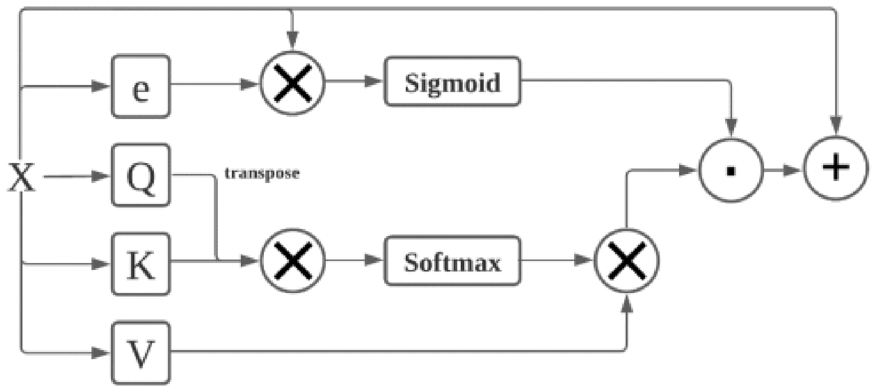

Figure 4. This figure shows how the input X is processed to extract features in multiple dimensions including the channel attentive mechanism.

attentive non-local (CAN) network is structured. By linear transformation toward input X, three feature spaces obtained:

$$Q_{i,j} = M^{q}_{i,c} X_{c,j} \tag{3}$$

$$K_{i,j} = M^{k}_{i,c} X_{c,j} \tag{4}$$

$$V_{i,j} = M^{v}_{i,c} X_{c,j} \tag{5}$$

where c, j, i are dimensions of input X and M^q, M^k, M^v are learnable arguments. Equations (3) to (5) deal with straightforward matrix multiplications between matrices M and X, and they may be efficiently carried out using convolution. The output of multiplication matrix Q^T and K is then subjected to $Softmax$ to produce a spatial attention matrix B. The attention matrix is finally computed using matrix multiplication once again. The entire procedure may be pictured as follows:

$$V_{i,j} Softmax(Q^{T}_{i,j} K_{i,j}) \tag{6}$$

It is worth mentioning that the attention obtained here does not change for different channels since each channel is multiplied by the same matrix after $Softmax$. Instead, the attention may have variant values by pixels.

In addition to channel-similar attention, ProCAN also uses an attention mechanism that has variations depending on channels. First, the task is to find attention for different channels. The step is as follows:

$$g_c = Sigmoid\left(X_{c,j}(m^{e}_{c} X_{c,j})\right) \tag{7}$$

where m^{e}_{c} is a learnable parameter and the $Sigmoid$ function is used to map values of g in the range between zero and one. After that, a residual connection of input is introduced:

$$\Psi_{c,n} = A_{c,n} g_c + X_{c,n} \tag{8}$$

Spatial attention A is multiplied by channel attention g. Finally, in order to control output channels and dimensions, a convolutional layer using 3x3 is used:

$$o = relu(\Psi * M^{o}) \tag{9}$$

It is helpful to employ non-local networks to capture the general features of lung nodules since they might vary in size and form, making it challenging to classify lung nodules. To generate features that may describe the size and shape of nodes, global feature extraction utilizing non-local networks is necessary. On the other hand, local features emphasize minute details like node density and texture.

3.2. A progressive training method employing a two-dimensional Bernoulli matrix.

The ProCAN model also uses progressive growing method to handle the limitation caused by network depth and structure. This method is firstly proposed by ProGAN [13] and is used in generative adversarial networks to generate images with high revolution easier. The fundamental concept behind it is to progressively raise the network's complexity to provide better results. In ProCAN paper, the specific action is adding new CAN blocks when training. In the previous introduction, there two main components for the network: a series of CAN layers f_0 and a classifier u. Denote a CAN layer as α, the attention actor can be shown by:

$$f_0 = \alpha_k(\alpha_{k-1}(\dots, \alpha_1(X)) \tag{10}$$

where k is the number of CAN blocks, and to augment features, another set of CAN blocks is connected to the end of f_0:

$$f_l = \alpha_l\big(f_0(X)\big) \tag{11}$$

here l is the number of added CAN blocks. Given that adding these blocks may affect the archaic training effectiveness and result in unexpected delay, the gradual growing is integrated to control the impact of newly added blocks. To implement this, a scalar p whose value varies from 0 to 1 is used. According to the training state, the actual value of p increases to take more features from α_l into account. Specifically, at the start state, the network does not use the new blocks at all and the value of p is zero. As the training goes on, it is required to enhance the impact of new features gradually. Thus, the value is set to be 0.25, 0.5, 0.75 respectively in the transitional state. Finally, the network may only accept features from α_l, which means p is set to be one at the final stage. The way to take both set of features into consideration can be expressed as follow:

$$f = p\alpha_l\big(f_0(X)\big) + (1-p)f_0(X) \tag{12}$$

However, multiplying the feature matrices by the scalar may affect the performance of non-local network because the concrete values may be distorted. Therefore, the ProCAN uses a $Bernoulli$ matrix with a probability of p to pass the original appearance of the features that exceed the threshold value. Also, feature values less than the probability are prevented from making a difference on the network. With the $Bernoulli$ matrix described as $\Omega \sim Bernoulli(p)$, the updated version of features passed is:

$$f = \Omega\alpha_l\big(f_0(X)\big) + (1-\Omega)f_0(X) \tag{13}$$

4. Comparison and revelation

4.1. Effectiveness of non-local network in lung nodule classification

Both models mentioned above adopt non-local methods. The term "Local" refers to the receptive field, which is the feature capture capability of network in observation. Taking the convolution operation as an example, the size of its receptive field is the size of the convolution kernel, and we typically use convolution kernels such as 3×3, 5×5, which only examine local regions, making them all local operations. Pooling operation is another example of local operations. It moves a forward step to extract features of images based on the convolution results. However, some tasks may need more information from the original picture. In contrast to a local field, non-local neural networks [11] indicates that the receptive field might be vast.

The use of non-local network in lung nodule classification is shown to be successful. As shown in Table1, Model Local-Global [9] reached a relatively high accuracy and AUC by simply embedding non-local blocks in the network, while other models shown in this table adopt multiple other methods such as knowledge-based collaborative submodel [14] and multi-view [14][15] method. In addition, ProCAN [6] introduce a new channel-attentive method for non-local and integrate progressive growth architecture to it, which reaches the highest accuracy so far.

Table 1. Comparisons in terms of accuracy and AUC of several models, Gated-Dilated.

Name	Accuracy	AUC
ProCAN	95.28	98.25
MV-KBC	91.6	95.7
MK-SSAC	92.53	98.51
DeepLung	90.77	
Local_Global	88.46	95.62
Gated-Dilated [16]	92.57	95.14

The new attention mechanism CAN in ProCAN also contributes to better performance of the network. The CAN block exhibits superiority over the dual attention networks utilized in DeepLung [8], which employ two independent blocks and the non-local paradigm for spatial and channel attention.

4.2. Effectiveness of training method

In addition to improvement in terms of non-local mechanism, ProCAN [6] also uses progressive growing method to train the model. ProGAN [13] first suggests a progressive growing network to gradually generate images with higher resolutions starting at lower resolutions. The integration of this method enhances the final performance according to Table 2 and Table 3. The compounded network has obvious advantages over other strategies without integration of multiple technologies.

Table 2. Performance of Local-Global and ProCAN, together with the ablation experiment without progressive growing network.

Model	**Accuracy**	**AUC**	**Precision**	**Sensitivity**
ProCAN	**94.11**	**97.13**	**94.54**	**93.12**
None	89.98	94.22	90.84	87.93
Local-Global	88.46	95.62	87.38	88.66

5. Conclusion

The author illustrates Local-Global model and ProCAN model in the previous sections and made a comparison trying to analysis the reason why the latter could produce higher accuracy and reaches the state-of-the-art standard. The former utilizes non-local network in lung nodule classification to enable the network to capture full appearance of training data and therefore reduce the negative effects brought by huge variance of lung nodule sizes and shapes. The latter expands the global perception to channel dimension inventively and adopts progressive learning method in training. Both non-local methods and newly proposed channel-attentive mechanism strengthen the use of training data. This is implemented by enhancing the effects of data on the original dimension or extending the data to a higher dimension. Under this circumstance, the model can identify nodules with multifarious size and form. In addition, the increase of accuracy of ProCAN can be attributed to progressive training method. A comparison is made between ProCAN and a model without using progressive training, in which the effectiveness of progressive learning is shown. Hence, the author could make a conclusion that a model with high classification accuracy is normally a technical cluster of several newly developed mechanisms or technologies from other applications of machine learning.

References

[1] Lung cancer. nhs.uk. 2022. Retrieved 19 August 2022, from

DOI: 10.54254/2755-2721/4/2023416

https://www.nhs.uk/conditions/lung- cancer/.

[2] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging, 35(5), 1299-1312.

[3] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324. (LeNet)

[4] Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. (VGG)

[5] He, K., Zhang, X., Ren, S., and Sun, J. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). (ResNet)

[6] Al-Shabi, M., Shak, K., and Tan, M. 2022. ProCAN: Progressive growing channel attentive non- local network for lung nodule classification. Pattern Recognition, 122, 108309.

[7] Armato III, S. G., McLennan, G., Bidaut, L., McNitt - Gray, M. F., Meyer, C. R., Reeves, A. P., ... and Clarke, L. P. 2011. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical physics, 38(2), 915-931.

[8] Zhu, W., Liu, C., Fan, W., and Xie, X. 2018. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 673-681). IEEE. (DeepLung)

[9] Al-Shabi, M., Lan, B. L., Chan, W. Y., Ng, K. H., and Tan, M. 2019. Lung nodule classification using deep local–global networks. International journal of computer assisted radiology and surgery, 14(10), 1815-1819. (Local-Global)

[10] Beini, Z., Xuee, C., Bo, L., and Weijia, W. 2021. A new few-shot learning method of digital PCR image detection. IEEE Access, 9, 74446-74453.

[11] Wang, X., Girshick, R., Gupta, A., and He, K. 2018. Non-local neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7794-7803). (non-local nn)

[12] Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. 2019. Self-attention generative adversarial networks. In International conference on machine learning (pp. 7354-7363). PMLR. (selfattention)

[13] Karras, T., Aila, T., Laine, S., and Lehtinen, J. 2017. Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196. (ProGAN)

[14] Xie, Y., Xia, Y., Zhang, J., Song, Y., Feng, D., Fulham, M., and Cai, W. 2018. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. IEEE transactions on medical imaging, 38(4), 991-1004. (MV-KBC)

[15] Xie, Y., Zhang, J., and Xia, Y. 2019. Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT. Medical image analysis, 57, 237-248. (MV- SSAC)

[16] Al-Shabi, M., Lee, H. K., and Tan, M. 2019. Gated-dilated networks for lung nodule classification in CT scans. IEEE Access, 7, 178827-178838.