ConvNet-CA: A Lightweight Attention-Based CNN for Brain Disease Detection

Hengde Zhu[1 [0000−0001−7027−3969]], Jian Wang[1 [0000−0001−8387−9505]], Shui-Hua Wang[1 [0000−0003−4713−2791]], Yu-Dong Zhang[1 [0000−0002−4870−1493]], and Juan M Górriz[2 [0000−0001−7069−1714]]

[1] School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE1 7RH, UK
yudongzhang@ieee.org
[2] Department of Signal Theory, Networking and Communications, University of Granada, Granada, 52005, Spain
gorriz@ugr.es

Abstract. Attention-based convolutional networks have attracted great interest in recent years and achieved great success in improving representation capability of networks. However, most attention mechanisms are complicated and implemented by introducing a large number of extra parameters. In this study, we proposed a lightweight attention-based convolutional network (ConvNet-CA) that has a low computation complexity yet a high performance for brain disease detection. ConvNet-CA weights the importance of different channels in features maps and pays more attention to important channels by introducing an efficient channel attention mechanism. We evaluated ConvNet-CA on a publicly accessible benchmark dataset: Whole Brain Atlas. The brain diseases involved in this study are stroke, neoplastic disease, degenerative disease, and infectious disease. The experimental results showed that ConvNet-CA achieved highly competitive performance over state-of-the-art methods on distinguishing different types of brain diseases, with an overall multi-class classification accuracy of 94.88±3.64%.

Keywords: Deep learning · Attention mechanism · Medical image.

1 Introduction

Brain disease is one of the most dangerous diseases which threaten human's health. It can cause headaches, coma, visual impairment, and movement disorder. There are varieties of brain diseases, and usually, doctors cannot distinguish them on the surface. Magnetic resonance imaging (MRI) is acknowledged as an ideal method to scan the brain's inner structure. Through MRI, clinicians can observe and assess the inner condition of the brain. In this way, clinicians decide whether the patient has a brain disease and which one the patient is suffering [1].

Although MRI scans enable clinicians to examine the brain's inner structure and condition, it is still not easy to detect pathological brains. The difference

between healthy brains and pathological brains can be very subtle [12]. Even the most experienced clinicians cannot avoid mistakes in brain examination. Therefore, using a computer-aided diagnosis system to help brain disease detection is becoming increasingly necessary.

There exist some methods that apply artificial intelligence to detect pathological brains automatically. Wang, et al. [15] proposed a pathological brain detection method based on stationary wavelet entropy. Nayak, et al. [9] designed a brain disease detection approach using improved particle swarm optimization and evolutionary extreme learning machine. These methods were effective in the model training and have achieved relatively high accuracy, but to extract more representative features from brain medical images, we opt to utilise deep learning to process medical images. Deep learning has been widely applied in image classification tasks in recent years. Due to its flexibility, it can be adapted to special problems [3]. As one of the most commonly used models, convolutional neural networks (CNNs) are often adopted as the backbone model [13]. Although they have shown superiority over traditional machine learning-based methods, the training is usually time-consuming, especially when a CNN has lots of layers. Our goal is to propose a lightweight CNN that is easy to train, while achieving better results than popular CNNs.

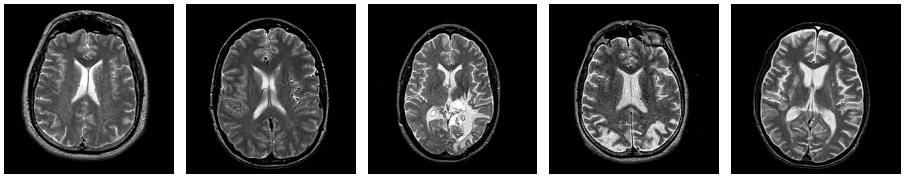
In this study, we introduced an efficient channel attention module that can be easily integrated into CNN architectures and proposed a novel lightweight attention-based CNN for brain disease detection. Our proposed model, named ConvNet-CA, has a concise structure and performs better than many large-scale CNNs [2, 5, 7, 11]. The main contribution of this study is we designed an efficient and lightweight model for pathological brain detection with high accuracy.

2 Data

In this study, we include a total number of 197 axial-oriented MRI T2-weighted images of healthy brain and pathological brains. Each image has a consistent dimension of 256×256 pixels. These images are acquired from a publicly accessible dataset: the Whole Brain Atlas [8]. There are five categories of brain images in our study, including the normal brain, stroke, neoplastic disease, degenerative disease, and infectious disease. Statistic information of the dataset is listed in Table 1. Some samples from the dataset are shown in Fig. 1. The detection of brain disease is considered as a 5-class classification task.

3 Methodology

This study applies advanced deep learning technology to diagnose brain disease based on MRI scans. Our model is based on a CNN architurecture and the Efficient Channel Attention (ECA) module [14]. The overall structure of our proposed model ConvNet-CA is shown in Fig. 3. It consists of three convolutional layers for feature extraction, with each layer followed by a ECA module to refine feature maps. The ECA modules only introduce a few parameters while allowing



(a) Normal Brain (b) Stroke (c) Neoplastic Disease (d) Degenerative Disease (e) Infectious Disease

Fig. 1. Brain MRI samples from the dataset

Table 1. Statistic information of the dataset.

Categories	Samples
Normal brain	20
Stroke	72
Neoplastic disease	31
Degenerative disease	41
Infectious disease	33

the model to focus on more important channels of feature maps. Pooling layers are added to summarise re-fined feature maps and reduce the size of feature maps. In this section, Conv-CA is explained in details.

3.1 Convolutional neural network

Researchers applied CNNs and developed hundreds of variants intending to solve almost every vision task. Although there existed some competitors in recent years such as graph neural networks and transformers, CNNs remain the mostly-adopted backbone of network architecture in the computer vision field. A CNN architecture typically has four elements, convolutional layer, pooling, activation function, and fully connected layer. Besides, dropout is one of the most applied techniques used in CNN training.

Convolutional layer. Convolutional layers are the most important parts of CNNs. They are used for feature extraction. In most cases, a CNN consists of many convolutional layers. These convolutional layers form a hierarchy that enables a CNN to extract deeper and deeper features of input data. In a convolutional layer, kernels move upon every region of the input data in a fixed order to produce feature maps. These feature maps are then fed to the next convolutional layer as new input.

Pooling. As a down-sampling method, pooling is used for removing redundant information in feature maps and preserving the most valuable information. Max pooling and average pooling are two of the most commonly pooling methods. Max pooling selects the maximal value as the representation of a local region, while average pooling utilises the mean value of a local region.

Activation function. A convolutional layer is often followed by an activation function that introduces nonlinearity. In this way, it enables CNNs to better portray the data distribution of the real world. The most commonly used activation function and also what we adopted in this study is ReLU.

Fully connected layer. A fully connected layer aims to map the distributed feature representation to sample annotation space. A fully connected layer is usually placed at the end of CNNs. It acts as a classifier of the model. In this research, we applied a fully connected layer at the end of the model structure.

Dropout. When training CNNs, we randomly select parts of neurons in hidden layers and set them to zero [10]. This operation is called dropout. Dropout is widely used in training deep neural networks for relieve overfitting. In this study, the dropout method is applied in the fully connected layer.

ConvNet-CA adopts a lightweight CNN as the backbone which consists of three convolutional layers with ReLU, three max pooling layers, and one fully connected layer with a dropout strategy.

3.2 Channel attention mechanism

Attention mechanism has been one of the most prominent progress researchers made in the deep learning over the past few years. It enable neural networks to focus on the most valuable information. In image analysis, spatial attention and channel attention are two of the most widely used attention mechanisms. Since Hu, et al. [6] proposed SENet where a channel attention module proved its potential, researchers have been looking for more advanced methods to make the best use of channel attention in feature extraction.

The channel attention module of SENet, named SE module, is divided into two steps, squeeze and excitation (see Fig. 2a). In the squeeze step, global average pooling summarises each feature map and compresses its dimension to 1×1. Supposing that the input block ω has a dimension of $H \times W \times C$, where H denotes height, W denotes width, and C denotes the number of channels, the squeeze step can be described as

$$G_{squeeze}(\omega_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \omega_c(i,j) \tag{1}$$

The squeeze step produces a global representation vector $z \in \mathcal{R}^{1\times1\times C}$. This vector does not directly weight the importance of channels. Instead, two fully connected layers are employed to learn importance of different channels. The first fully connected layer is applied to the global representation vector, which reduces its dimension to $1 \times 1 \times \frac{C}{r}$ with a reduction ratio r. Then a ReLU is adopted for nonlinear activation, and a second fully connected layer is applied to regain its dimension to $1 \times 1 \times C$. In the end, a *sigmoid* function is employed to activate the vector, which is later multiplied by the original feature maps to acquire weighted feature maps. Supposing that the weights of the first and second fully connected layer are denoted as $V_1 \in \mathcal{R}^{\frac{C}{r}\times C}$ and $V_2 \in \mathcal{R}^{C\times\frac{C}{r}}$, the

ReLU function is denoted as δ and the *sigmoid* function is denoted as σ, the excitation step can be described as

$$G_{excitation}(z, V) = \sigma(V_2\delta(V_1 z)). \tag{2}$$

As a SE module contains two fully connected layers that contain a massive number of parameters, it can result in slowing down the model training. In addition, the SE module endures a dimension reduction and a dimension increment, which might cause loss of information from original feature maps [14]. To limit model complexity and avoid dimension variation, we find this ECA module [14] that applies one convolution operation to generate the channel attention vector. The ECA module is more efficient as it has much fewer parameters. It can also preserve information better than the SE module as it avoids dimension reduction.

Similar to the SE module, the ECA module first applies a global average pooling layer to compress feature maps to a global representation vector $z \in \mathcal{R}^{1\times1\times C}$. A convolution layer with a kernel size of k and zero-padding of $\lfloor \frac{k}{2} \rfloor$ is then applied to the compressed vector to produce a channel attention vector with the size of the global representation vector. This operation captures local interaction across channels with a coverage of k. It is efficient that it only introduces k number of extra parameters. Given the channel dimension C, to adaptively determine the size of coverage k, the authors of the ECA module proposed a nonlinear mapping [14] as below

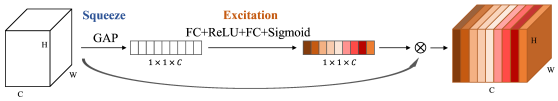
$$k = |\frac{\log_2(C)}{\gamma} + \frac{b}{\gamma}|_{odd}. \tag{3}$$

where γ and b are two constants.

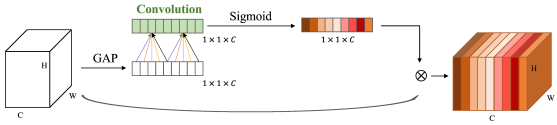
The process of the ECA module is illustrated in a more understandable way in Fig. 2b. The input block is down-sampled by a global average pooling layer. A convolution operation is performed on the compressed vector and generates the channel attention vector without dimension reduction. Then, a *sigmoid* function is applied to introduce nonlinearity. Finally, the channel attention vector is multiplied by original feature maps to obtain the weighted feature maps. It is observed that the ECA module has fewer steps and does not have the process of dimension reduction. Therefore, ConvNet-CA adopts ECA modules to perform channel attention.

3.3 ConvNet-CA

As it is shown in Fig. 3, ConvNet-CA has a very simple and elegant architecture. It consists of three blocks and one fully connected output layer. In the first block, we employ a 3×3 convolution followed by ReLU. Then we apply the first channel attention module. At the end of the first block, there exists a max-pooling layer. Convolution, channel attention module, and max-pooling layer constitute the first block. The second and the third blocks repeat the first block's structure. The main difference is the number of kernels in convolutoinal layers. After these three main blocks, we adopt a fully connected layer with dropout to output classification probabilities.



(a) SE module



(b) ECA module

Fig. 2. Comparison of ECA module with SE module

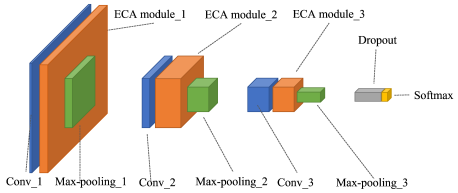


Fig. 3. Overall structure of ConvNet-CA

3.4 Evaluation metrics

Due to the small size of the dataset, we adopt the 5-fold cross-validation scheme to evaluate models' performance and generalization ability. The dataset is partitioned into five subsets. Each subset is called a fold. We perform five iterations of training and testing. In each iteration, one fold of data is used in testing, while the remaining folds are used in training. The averaged model performance on five iterations is calculated as the 5-fold cross-validated performance.

Four common evaluation metrics, including accuracy (ACC), precision (PRE), sensitivity (SEN), and f1-score (F1s), are used to compare classification performance among different methods. True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) of predicted results are introduced to define above evaluation metrics. The definitions are as follows

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$PRE = \frac{TP}{TP + FP} \tag{5}$$

$$SEN = \frac{TP}{TP + FN} \tag{6}$$

$$F1s = 2 \times \frac{PRE \times SEN}{PRE + SEN} \tag{7}$$

As this is a multi-class classification task, PRE, SEN, and F1s are first calculated for each class separately, and then the unweighted means of them, termed as PRE_{macro}, SEN_{macro}, and F1s_{macro}, are used to give an overall performance.

4 Experiments and results

4.1 Experiment set-up

The study conducted experiments on a NVIDIA TESLA P100 GPU with 16 GB RAM provided by Kaggle. The hyperparameter settings of ConvNet-CA are given in Table 2. It is worth noting that other methods of comparison used in this study are trained under the same settings of optimizer, learning rate, batch size and trainning epochs (see Table 3).

Table 2. Hyperparameter settings of ConvNet-CA.

Hyperparameter	Value	Hyperparameter	Value
Optimizer	Adam	Training epochs	100
Learning rate	0.0001	Activation function	ReLU
Dropout	0.2	Kernel size	3
Batch size	1	Number of Conv Layers	3

4.2 Performance on multi-class classification

We compared ConvNet-CA with four popular CNNs with strong representation capability. These networks are derived from Xception [2], Inception [11], ResNet50 [5], DenseNet121 [7]. As the dataset used in this study is a small dataset, directly training deep networks on it can lead to the overfitting problem. Thus, these networks were pre-trained on a large dataset, ImageNet, to learn how to capture representative features from images and then re-trained on our medical dataset. To adapt our dataset, the original fully connected layers at the end of these networks are replaced with a global average pooling layer and two new fully connected layers to perform a 5-class classification task. Except for the new fully connected layers, parameters of other layers of networks are frozen. These networks are re-trained to perform a domain-specific task.

The performance of different networks is summarised in Table 3. ConvNet-CA achieved the best performance over all metrics, with the classification accuracy of 94.88±3.64%, the macro precision of 96.50±2.74%, the macro sensitivity of 93.34±5.01%, and the macro f1-score of 94.21±4.45%. The deviation of accuracy and precision of ConvNet-CA are the smallest, indicating that its performance was more stable than other methods. In contrast, most deep networks achieved over 80.00% of all metrics, apart from ResNet50. The total number of parameters and FLOPs of different networks are shown in Table 4. It shows that ConvNet-CA has much fewer parameters and requires less computational resources. Its performance indicates that it can capture useful pattern more effectively.

Table 3. Classification performance. Metrics displayed in 'mean±standard deviation' format.

Method	ACC(%)	PRE_{macro}(%)	SEN_{macro}(%)	$F1s_{macro}$(%)
Xception	82.24±5.31	80.73±5.68	81.90±4.82	78.54±4.82
InceptionV3	83.73±2.20	86.98±3.18	82.92±3.80	82.67±3.52
ResNet50	60.96±8.09	56.24±4.03	57.66±4.64	49.09±4.03
DenseNet121	81.28±7.76	80.99±7.48	81.86±5.61	80.52±7.49
ConvNet-CA	**94.88±3.64**	**96.50±2.74**	**93.34±5.01**	**94.21±4.45**

Table 4. Total number of parameters and FLOPs.

Method	Params(M)	FLOPs(G)
Xception	22.96	11.95
InceptionV3	23.91	7.73
ResNet50	25.69	10.13
DenseNet121	8.09	7.45
ConvNet-CA	**0.80**	**1.75**

4.3 The effectiveness of channel attention mechanism

We designed experiments to study the effectiveness of channel attention used in ConvNet-CA. We denote ConvNet as the backbone of the ConvNet-CA, where channel attention is removed. The comparison of performance between ConvNet and ConvNet-CA is shown in Fig. 4. Experiment results show that the introduction of channel attention led to an overall performance increase of around 3.00%. The channel attention allows the network to focus on more useful information. It is worth noting that channel attention only introduces a few parameters. The total number of parameters for ConvNet and ConvNet-CA are 798,085 and 798,098, respectively. The slight difference in the number of parameters and the significant performance improvement demonstrated the effectiveness of channel attention.

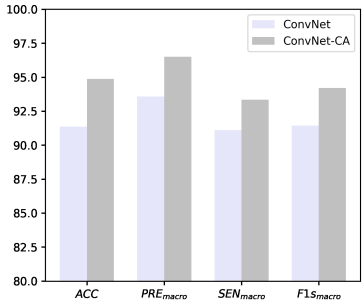


Fig. 4. Comparison between ConvNet-CA and ConvNet

4.4 Comparison with state-of-the-art methods

The overall performance of ConvNet-CA is compared with several state-of-the-art methods using the same dataset: VMD+SNPE+ANOVA [4] and FCEntF-II + K-ELM [9]. All mentioned methods above classify the brain MRI scans into five categories. To avoid the overfitting problem, both of our study and the study of Nayak, et al. [9] adopted a 5-fold cross validation scheme while a 10-fold cross validation scheme is adopted in the study of Gudigar, et al. [4]. From the Table 5, it can be observed that our method ConvNet-CA obtained the highest classification accuracy of 94.88% among the three methods.

Table 5. Comparison with state-of-the-art methods.

Study	Method	ACC(%)
Gudigar, et al. [4]	VMD+SNPE+ANOVA	90.68
Nayak, et al [9]	FCEntF-II + K-ELM	93.00
Our approach	ConvNet-CA	**94.88**

5 Conclusion

This study proposed a lightweight attention-based CNN for brain disease detection. Compared with other deep networks with a large amount of parameters, our lightweight model can capture features more efficiently and effectively from a specific dataset, with higher performance over four evaluation metrics in the study. This network integrates an efficient attention mechanism to assign different importance to different channels of feature maps. It enables the model to pay more attention to most important channels. Experimental resutls demonstrated the effectivess of the channel attention mechanism used in this study which led to significant performance enhancements. In the future, we shall apply our proposed model to more medical datasets and improve the model's generalization ability. The visualisation of the model is another research direction that would help us understand how the network works.

Acknowledgements

This paper is partially supported by Medical Research Council Confidence in Concept Award, UK (MC_PC_17171), Royal Society International Exchanges Cost Share Award, UK (RP202G0230), British Heart Foundation Accelerator Award, UK (AA/18/3/34220), Global Challenges Research Fund (GCRF), UK (P202PF11), Sino-UK Industrial Fund, UK (RP202G0289), and Hope Foundation for Cancer Research, UK (RM60G0680).

References

1. Atlas, S.W.: Magnetic resonance imaging of the brain and spine, vol. 1. Lippincott Williams & Wilkins (2009)
2. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1251–1258 (2017)
3. Górriz, J.M., Ramírez, J., Ortíz, A., Martinez-Murcia, F.J., Segovia, F., Suckling, J., Leming, M., Zhang, Y.D., Álvarez-Sánchez, J.R., Bologna, G., et al.: Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications. Neurocomputing **410**, 237–270 (2020)
4. Gudigar, A., Raghavendra, U., Ciaccio, E.J., Arunkumar, N., Abdulhay, E., Acharya, U.R.: Automated categorization of multi-class brain abnormalities using decomposition techniques with mri images: a comparative study. IEEE Access **7**, 28498–28509 (2019)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
6. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
7. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
8. Johnson, K.A., et al.: The whole brain atlas (2001)
9. Nayak, D.R., Dash, R., Majhi, B.: Discrete ripplet-ii transform and modified pso based improved evolutionary extreme learning machine for pathological brain detection. Neurocomputing **282**, 232–247 (2018)
10. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
11. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016)
12. Tofts, P.: Quantitative MRI of the brain: measuring changes caused by disease. John Wiley & Sons (2005)
13. Wang, J., Zhu, H., Wang, S.H., Zhang, Y.D.: A review of deep learning on medical image analysis. Mobile Networks and Applications **26**(1), 351–380 (2021)
14. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: efficient channel attention for deep convolutional neural networks, 2020 ieee. In: CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2020)
15. Wang, S., Du, S., Atangana, A., Liu, A., Lu, Z.: Application of stationary wavelet entropy in pathological brain detection. Multimedia Tools and Applications **77**(3), 3701–3714 (2018)