Anomaly Detection Model Combined With Attention Mechanisms for Chest X-ray Images

Tan Yanli[1,2], Azliza Mohd Ali[2,*], Sharifalillah Nordin[2], Wang Jin[1,2]

[1] Department of Electronic Engineering, Taiyuan Institute of Technology, 030008 Taiyuan, Shanxi, China;

[2] College of Computing, Informatics and Mathematics , Universiti Teknologi MARA, Selangor, Shah Alam, 40450, Malaysia;

azliza@tmsk.uitm.edu.my

Abstract. This study introduces CB-ResNet and CB-DenseNet anomaly detection models that utilise attention mechanisms. The objective is to overcome the limitations of classic anomaly detection algorithms, including low detection accuracy and unstable model performance. The backbone network utilises the convolutional block attention mechanism (CBAM) to improve the extraction of target feature information in both spatial and channel dimensions during shallow feature extraction. By assigning importance to the extracted features of the CNN network, the attention mechanism can eliminate irrelevant information and enhance the model's ability to learn anomalous data features. The experimental findings demonstrate that both CB-ResNet and CB-DenseNet models, which utilise attention mechanisms, may reach detection accuracy over 99%. Furthermore, these models exhibit strong stability and possess a high level of generalisation capability. We gain superior efficiency compared to conventional models such as ResNet and DenseNet.

Keywords: Anomaly detection; Attention mechanism; ResNet; DenseNet.

1 Introduction

Anomaly identification in medical images is a prominent area of research in image processing and a crucial component of medical diagnosis [1]. Conventional anomaly detection algorithms rely on collections of images and struggle to differentiate the tiny distinctions between normal and abnormal characteristics in medical imaging. Deep learning has demonstrated significant potential in acquiring representations of complex data, It has expanded the limits of numerous learning tasks. The objective of this study is to develop a model for detecting anomalies in chest X-ray images using deep learning algorithms. The model utilises a convolutional neural network to identify anomalies by performing image preprocessing, image augmentation, and image feature extraction.

Pan Liyan et.al [2] proposed an improved method utilising AlexNet to categorise various types of chest x-rays demonstrating pneumonia. Horry M J suggested employ-

N. A. S. Abdullah et al. (eds.), *Proceedings of the International Conference on Innovation & Entrepreneurship in Computing, Engineering & Science Education (InvENT 2024)*, Advances in Computer Science Research 117,
https://doi.org/10.2991/978-94-6463-589-8_8

ing transfer learning as a means to automatically identify individuals with COVID-19 [3]. The study analysed the effectiveness of state-of-the-art convolutional neural network (CNN) models that have been employed for medical image categorization in recent years. These models include VCG19, MobileNetV2 [4], Xception [5], Inception [6], and Inception-ResNetV2 [7]. After performing a thorough analysis, we determined that VGG19 and MobileNet had better results when compared to the other results.

The paper is organized as follows: In Sect.2, we present the convolutional neural networks-ResNet and DenseNet. Section 3 describes two anomaly detection models: CB-ResNET and CB-DenseNet. Section 4 presents the experimental setup and results. Finally, conclusions and future work are drawn in Sec.5.

The main contributions of this article are: proposing anomaly detection models CB-ResNet and CB-DenseNet based on attention mechanisms, and applying this model to the field of chest X-ray anomaly detection for the first time. Compared with traditional anomaly detection models, the network model proposed in this article has higher detection accuracy and shows strong generalization ability.

2 Convolutional Neural Network – CNN

In this section, we introduce the two classical convolutional neural network models mainly used in this paper: ResNet and DenseNet.

2.1 Deep residual network — ResNet

In recent years, the deep residual network (ResNet) has been widely used to solve the degradation problem in deep neural networks by introducing cross-layer connections [8], which deepens the network and improves its performance. As shown in Figure 1, ResNet uses a cut-off design to input data from previous layers directly into later data layers. If the input is x, the desired underlying mapping is $H(x)$, and a residual mapping is defined as $F(x) = H(x) - x$, the original mapping function $H(x)$ can be expressed as $F(x) + x$. The introduction of disconnected designs does not add other parameters and has no impact on the original network . ResNet can also be solved by feedback training through deep neural networks [9]. In the training process, errors at the bottom layer can be propagated to the upper layer through the breaking mode, effectively avoiding the problem of gradient disappearance caused by too many layers and ultimately improving the accuracy of training.

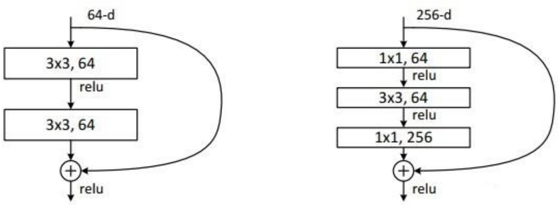


Fig. 1. Residual block structures of ResNet at different depths

2.2 Dense Convolutional Networks — DenseNet

DenseNet is a deep neural network connected between layers [10]. In this dense connection model, the output feature maps by all layers before the current layer are connected along the channel dimension as the input of the current layer, and then the output of this layer is used as the input for the subsequent layers. The dense connection between feature layers can maximize the information exchange among them. The dense convolution block and transition layer are the two most important modules in the network structure of DenseNet201, in which the growth rate is used to control the output of each feature layer. The network structure of DenseNet201 is shown in Figure 2.

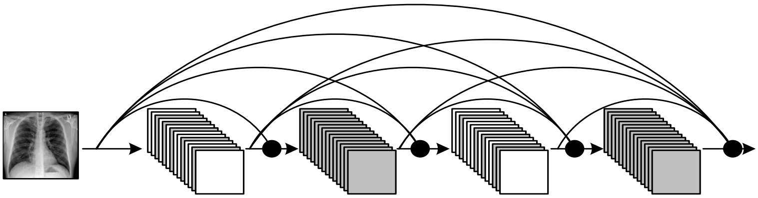


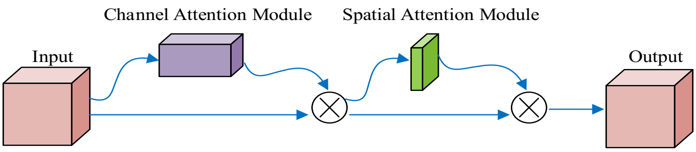
Fig. 2. Network structure of DenseNet201

3 Proposed Model of Attention-based Anomaly Detection

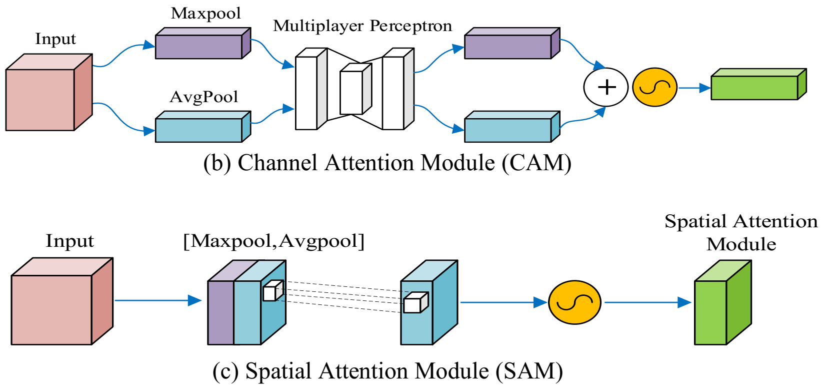
A suggested anomaly detection model, CB-ResNET and CB-DenseNet, integrates the pre training network CBAM (Convolutional Block Attention Module) and attention mechanism. It is based on ResNet and DenseNet.

3.1 CBAM（Convolutional Block Attention Module）

CBAM is a general-purpose and lightweight module that can be seamlessly integrated into any CNN architecture without significant overheads and trained end-to-end together with base CNNs [11]. There are two sequential sub-modules in CBAM, including Channel Attention Module (CAM) and Spatial Attention Module (SAM). CBAM is shown in Figure 6(a). CAM is shown in Figure 6 (b). The output feature map of the CAM is taken as the input feature map of the SAM, as shown in Figure 3 (c).



(a) Convolutional Block Attention Module (CBAM)



(b) Channel Attention Module (CAM)

(c) Spatial Attention Module (SAM)

Fig. 3. Schematic diagrams of (a)CBAM, (b) CAM, and (c)SAM.

3.2 CB-DenseNet model

In order to improve the performance of CNN-based anomaly detection models for chest X-ray images, the CB-DenseNet network was constructed according to the contents above, as shown in Figure 4. The CBAM is embedded into a dense block in the DenseNet201 channel, with the CBAM being in front of each 1×1 convolutional layer and behind the first three dense blocks. By using the convolutional attention mechanism to extract features at shallow layers, the network efficiency is further improved.

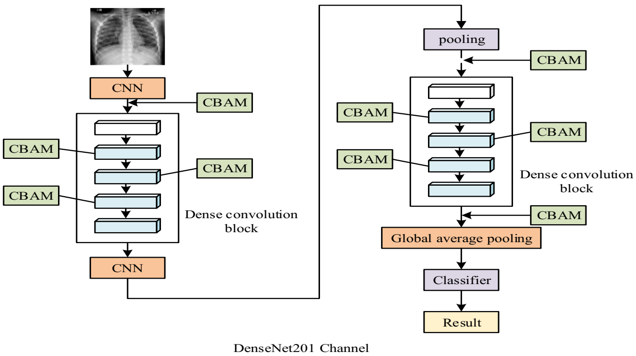


Fig. 4. Network structure of CBAM+DenseNet201

The output from the previous convolution block is sequentially fed to the convolution layer (Conv), the batch normalization (BN) layer, the rectified linear unit (ReLU) layer, and the pooling layer. The output is called an activation map (AM). The channel optimization activation maps can be obtained:

$$E = CAM(D) \otimes D \quad (1)$$

The final output:

$$F = SAM(E) \otimes E \quad (2)$$

where E represents the channel optimization activation map, F represents the final output, and $\otimes$ represents the multiplication operation between elements, The improvement of the elements from E to F is achieved through the channel attention mechanism and the space attention mechanism, and the output enters the next level of the convolution block for operation.

3.3 CB-ResNet model

Based on the above working principle, the CB-ResNet model is also proposed. In this model, ResNet50 with attention mechanism is used as the backbone network. Channel attention mechanism and spatial attention mechanism of CBAM are applied in a block structure, i.e., channel attention and spatial attention are successively passed before the original block and residual structure are connected.

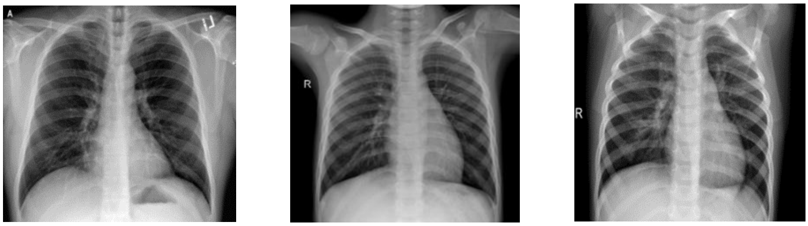
4 Experiment

4.1 Experimental Setting

The experiment was performed using Python 3.6 on the Windows 10 operating system. The model was built based on PyTorch. The computer is equipped with a CPU of R7-5800H and a RAM of 16GB. The algorithm was trained and tested with GPU acceleration. In the training process, Adam and SGD optimizers were used. All models iterated the entire training set (Epoch) for 20 and 50 epochs. The Batch Size of forward propagation and backpropagation was 32 each time, and the learning rate was set to 0.0001.

4.2 Data

This study uses the publicly available Covid-19 dataset. The Covid19-dataset contains 456 CXR images from confirmed COVID-19 patients, 374 CXR images from patients with viral pneumonia, and 285 normal CXR images in PNG format with a resolution of 1024×1024 pixels. The dataset is divided into a testing set and a training set. The testing set contains 230 CXR images of COVID-19 patients, 226 CXR images of patients with viral pneumonia, and 145 normal CXR images. The training set contains 226 CXR images of COVID-19 patients, 148 CXR images of patients with viral pneumonia, and 140 normal CXR images. The typical CXR images from patients with COVID-19 and patients with non-COVID-19 are shown in Figure 5.



(a) (b) (c)

Fig. 5. Chest X-ray images of (a) normal people, (b) patients with COVID-19, and (c) patients with viral pneumonia (non-COVID-19).

4.3 Evaluation indicators

Precision (P), Recall (R) and F1 scores were selected as evaluation indexes. Precision (*P*) represents the probability that all samples detected positive are actual positive samples [12]. This index reflects the detection capability and is calculated by Equ. 3. Recall (R) represents the probability of being detected as a positive sample in an actual positive sample, and it is calculated by Equ. 4. In anomaly detection and evaluation, Precision (P) and Recall (R) are commonly used performance evaluation indexes, but they are often contradictory . To better evaluate the performance of anomaly detection algorithms, the F1 score, also known as harmonic mean, is proposed by combining Precision (P) and Recall (R) into a single index [13], and it can be regarded as a weighted average of *P* and *R*. The F1 score is calculated by Equ. 5 with a range of 0 to 1 The closer the F1 score is to 1, the higher the model's performance is.

$$\mathrm{Precision} = \frac{TP}{TP+FP} \tag{3}$$

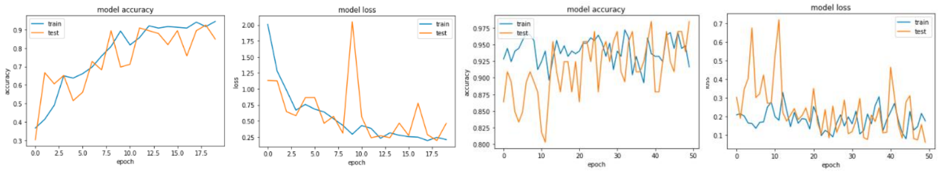
$$\mathrm{Recall} = \frac{TP}{TP+FN} \tag{4}$$

$$F1 = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{5}$$

where TP (Ture Positive) means that the Positive sample is correctly detected as a positive sample; FP (False Positive) means that the Negative sample is incorrectly detected as a positive sample; TN (True Negative) means that the negative sample is correctly detected as a negative sample; FN (False Negative) indicates that the positive sample is incorrectly detected as negative [14].

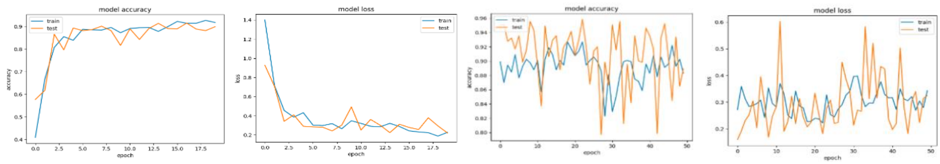
4.4 Model verification and result analysis

In this work, four deep learning network models are developed, trained, and tested on the same dataset, and the performance of these models is compared. Figure 6 shows the accuracy and loss of the training set and testing set obtained using the ResNet50 and DenseNet121 models for 20 and 50 epochs. After classifying the images in the testing set for 20 epochs, the classification accuracy of the ResNet50 and DenseNet121 models is 92.74% and 95.58%, respectively. However, after classifying the images for 50 epochs, the classification accuracy of these two models varies largely, indicating that the increasing number of epochs does not improve the performance of these two models.



epoch=20 epoch=50.

(a) Accuracy and loss of the ResNet50 model after training for 20 and 50 epochs

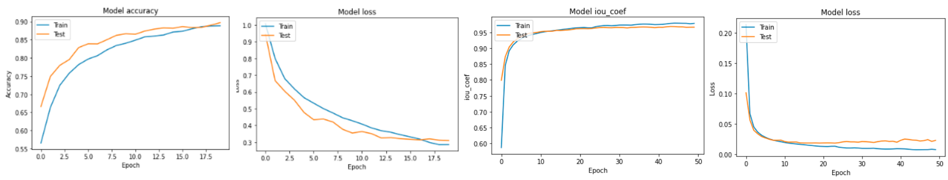


epoch=20 epoch=50

(b) Accuracy and loss of the DenseNet121 model after training for 20 and 50 epochs

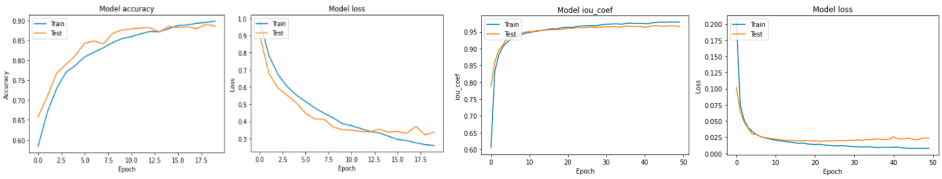
Fig. 6. Accuracy and loss of the ResNet50 and DenseNet121 models after training for 20 and 50 epochs

Figure 7 shows the accuracy and loss rate of the training set and testing set obtained using the CB-ResNet and CB-DenseNet models in 20 and 50 epochs. After the model with 20 epochs was used to classify the images in the testing set, the classification accuracy of CB-ResNet and CB-DenseNet was 97.89% and 98.82%, respectively, both of which were higher than those of ResNet and DenseNet. With the increasing number of epochs, the stability of the model remained high, the accuracy showed a steady upward trend and finally tended to converge, suggesting that the model had a strong generalization ability.



epoch=20 epoch=50

(a) Accuracy and loss of the CB-ResNet model after training for 20 and 50 epochs



epoch=20 epoch=50

(b) Accuracy and loss of the CB-DenseNet model after training for 20 and 50 epochs

Fig. 7. Accuracy and loss of the CB-ResNet and CB-DenseNet models after training for 20 and 50 epochs

Table 1 shows the training results of four models and their performance indexes on the testing set. The accuracy and loss of the four models on the same training set and testing set are different. The introduction of attention mechanism in the backbone network can effectively improve the anomaly detection accuracy of the model on the testing set, with CB-ResNet and CB-DenseNet reaching 99.55% and 99.75% accuracy, respectively. With the increasing epochs, the accuracy of the CB-ResNet and CB-DenseNet models is steadily improved, indicating that the model has a good generalization ability.

Table 1 Comparison of performance indexes among different models

Model	Iterations	train_loss	train_accuracy	test_loss	test_accuracy	F1
ResNet	20	0.045367	91.52%	0.002865	92.74%	0.9236
	50	0.052742	91.85%	0.004731	92.81%	0.9287
DenseNet	20	0.031763	93.50%	0.002127	95.58%	0.9428
	50	0.056732	94.36%	0.003466	96.44%	0.9472
CB-ResNet	20	0.024739	95.46%	0.002109	97.89%	0.9521
	50	0.022531	98.21%	0.001783	99.55%	0.9653
CB-DenseNet	20	0.022761	97.46%	0.00412	98.82%	0.9836
	50	0.021842	98.54%	0.001356	99.76%	0.9875

Finally, the abnormal results of CXR images were displayed, and the abnormal results were divided to obtain the Mask of chest X-ray images of COVID-19 patients, as shown in Figure 8.

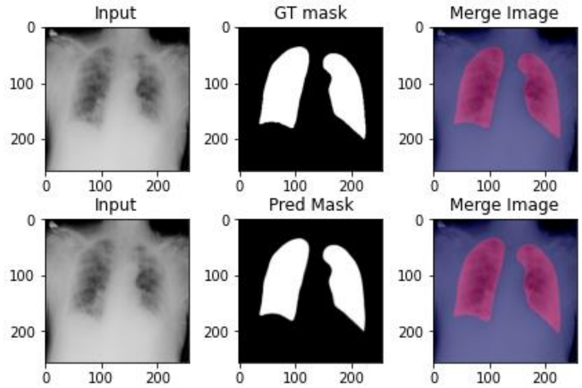


Fig. 8. Mask of chest X-ray images of COVID-19 patients

5 Conclusion

Based on ResNet and DenseNet models, two anomaly detection models (CB-ResNet and CB-DenseNet) were proposed by introducing attention mechanism, and their performance was verified on the training and testing sets. Compared with ResNet and DenseNet, CB-ResNet and CB-DenseNet exhibit an increase in anomaly detection accuracy, and their performance is stable. The results indicate that CB-ResNet and CB-DenseNet models have better generalization ability.The limitation of this algorithm is that it runs slowly due to the large number of network model parameters. In the future, it will be necessary to continue to optimize the proposed network model structure. How to reduce network model parameters and shorten network running time is the next research direction.

Acknowledgments. The authors would like to express gratitude to the School of Computing Sciences, College of Computing, Informatics and Mathematics, Department of Electronic Engineering, Taiyuan Institute of Technology for all the given support.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Pathak, Yadunath, et al. "Deep transfer learning based classification model for COVID-19 disease." Irbm 43.2 (2022): 87-92.
2. Pan, Liyan, and Huiying Liang. "Classification model for pathogen types in pediatric pneumonia based on deep learning." China Digit. Med 14 (2019): 59-61.
3. Horry, Michael J., et al. "COVID-19 detection through transfer learning using multimodal imaging data." Ieee Access 8 (2020): 149808-149824.
4. Dong, Ke, et al. "MobileNetV2 model for image classification." 2020 2nd International Conference on Information Technology and Computer Application (ITCA). IEEE, 2020.

5. Lo, Wai Weng, Xu Yang, and Yapeng Wang. "An xception convolutional neural network for malware classification with transfer learning." 2019 10th IFIP international conference on new technologies, mobility and security (NTMS). IEEE, 2019.
6. Si, Chenyang, et al. "Inception transformer." Advances in Neural Information Processing Systems 35 (2022): 23495-23509.
7. Alruwaili, Madallah, Abdulaziz Shehab, and Sameh Abd El-Ghany. "COVID-19 Diagnosis Using an Enhanced Inception-ResNetV2 Deep Learning Model in CXR Images." Journal of Healthcare Engineering 2021.1 (2021): 6658058.
8. Rajpal, Sheetal, et al. "Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images." Chaos, Solitons & Fractals 145 (2021): 110749.
9. Farooq, Muhammad, and Abdul Hafeez. "Covid-resnet: A deep learning framework for screening of covid19 from radiographs." arXiv preprint arXiv:2003.14395 (2020).
10. Hasan, Najmul, et al. "DenseNet convolutional neural networks application for predicting COVID-19 using CT image." SN computer science 2.5 (2021): 389.
11. Wang, Wei, et al. "A CBAM based multiscale transformer fusion approach for remote sensing image change detection." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15 (2022): 6817-6825.
12. Yacouby R, Axman D. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models[C]//Proceedings of the first workshop on evaluation and comparison of NLP systems. 2020: 79-91.
13. Hashi, Emrana Kabir, and Md Shahid Uz Zaman. "Developing a hyperparameter tuning based machine learning approach of heart disease prediction." Journal of Applied Science & Process Engineering 7.2 (2020): 631-647.
14. Jaiswal, Aayush, et al. "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning." Journal of Biomolecular Structure and Dynamics 39.15 (2021): 5682-5689.