Applied Intelligence (2021) 52:11058–11067
https://doi.org/10.1007/s10489-021-02935-w

Identifying cardiomegaly in chest x-rays using dual attention network

Lifang Chen[1,2] · **Tengfei Mao**[1] · **Qian Zhang**[1]

Accepted: 14 October 2021 / Published online: 20 January 2022

Abstract
The chest X-ray (CXR) is one of the most commonly available radiological examinations for identifying chest diseases. The application of deep learning methods in computer vision is becoming more and more mature, it provides new methods for automatic analysis of medical images and assisting doctors in high-precision intelligent diagnosis. In this paper, we propose a dual attention network to identify cardiomegaly (CXRDANet) on CXR images. CXRDANet is equipped with channel attention module (CAM) and spatial attention module (SAM), which selectively enhance features highly related to lesion area. We select CXR images of cardiomegaly and normal from ChestX-ray14 and NLM-CXR, without overlapping images, as the training set and the test set. Experimental results show that our method attains the accuracy of 0.9050, the sensitivity of 0.9445, the specificity of 0.8610, the F1 score of 0.9059, the AUC of 0.9588, which is a new state-of-the-art performance. In addition, we apply our method to the multi-label CXR image classification, and its performance has reached an excellent level.

Keywords Cardiomegaly · Strip pooling · Dual Attention Network · Chest X-rays

1 Introduction

Cardiomegaly is an important sign of potentially severe heart diseases and will increase the risk of heart diseases that seriously threatens patient's life. It is estimated that there are more than 200,000 new cases of cardiomegaly every year in the US, which is an important health issue for the middle-aged and senior population [1].

Chest X-rays (CXR) are widely used in clinics, with approximately 2 billion people being examined every year around the world. It is essential for the screening, diagnosis and management of various diseases including cardiomegaly [2].

However, as shown in Fig. 1, only subtle differences exist between the CXR of diseased and normal condition. Even for radiologists, distinguishing different types of diseases from CXRs is a tedious and challenging task [3].

Recently, many studies [4, 5] show that deep learning methods have been widely used in the field of medical image processing, such as breast cancer detection [6], pulmonary nodule detection [7], diabetic retinopathy detection [8]. With the development of deep learning [9, 10] and the release of large-scale CXRs dataset [11, 12], many related works have been carried out in developing deep learning based method for diseases detecting and classification.

Wang et al. [11] evaluate four traditional CNN networks [9, 13–15] to determine the existence of multiple pathologies using CXR images. Rajpurkar et al. [2] fine-tune a modified DenseNet-121 on CXR images. Due to the limitation of data volume, Zhou et al. [1] use transfer learning to solve the problem of cardiomegaly identifying, however, they ignore the factors that the lesion area of cardiomegaly is relatively small and may be affected by unrelated areas or noise during training. Candemir et al. [16] introduce a

✉ Lifang Chen
may7366@163.com

Tengfei Mao
6191611035@stu.jiangnan.edu.cn

Qian Zhang
6191611050@stu.jiangnan.edu.cn

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

[2] Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China

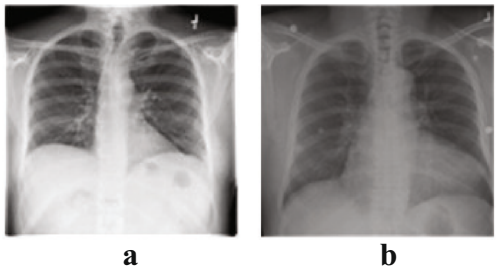

Fig. 1 **a**:X-ray Images for Patients without Cardiomegaly. **b**:X-ray Images for Patients with Cardiomegaly

model pre-trained in chestX-ray14 [11], then fine-tune the network on CXR images labeled as cardiomegaly. However, it only uses multiple deep networks to process images independently during training, and does not make full use of the features extracted by the deep network.

Compared with natural image classification, it is more difficult to classify diseases from CXR images due to the need to recognize the texture and features from the lesion areas. In clinical practice, the diagnosis of cardiomegaly is usually interfered by irrelevant areas or image noise. As shown in Fig. 2, we design the channel attention module (CAM) and the spatial attention module (SAM) to address the above problems.

Specifically, in the CAM, we adopt grouping strategy to improve the performance of the network while reducing the number of parameters, and use global average pooling and global max pooling to capture spatial statistics and distinctive object features, respectively. In the SAM, the input feature maps is split into two branches. Then, we encode the spatial information by performing horizontal and vertical strip pooling in two branches respectively. To obtain spatial attention in the horizontal and vertical direction, we process the encoded spatial information by FC layer and sigmoid activation. The final output is obtained by multiplying the input feature maps and the spatial attention.

We evaluated our method on ChestX-ray14 and National Library of Medicine CXR Indiana Collection (NLM-CXR) and reached an excellent level. In addition, we used the chestX-ray14 dataset to evaluate our method on the multi-label classification task of chest diseases, and its AUC reached 0.8186, becoming one of the most advanced methods. To facilitate further research, the source codes of CXRDANet will be released at https://github.com/TenfoldM/CXRDANet.

The main contributions of our work are:

(1) We proposed CXRDANet, which can achieve accurate cardiomegaly diagnosis in CXR image.
(2) We designed two attention modules, CAM and SAM, which suppress noises and enhance the correct semantic feature areas simultaneously.
(3) Experiments on two large public datasets show that CXRDANet reached the state-of-the-art.

2 Related work

In this section, we will discuss the application of attention mechanism in medical image processing and grouping strategy in deep learning.

2.1 Grouping strategy in deep learning

Grouping strategy has become one of the common methods for constructing deep networks. Krizhevsky et al. [14] use Group Convolution to group the input feature maps and then perform convolution operations separately, which reduces the number of model parameters and can also be deployed on multiple GPUs. Xie et al. [17] proposed ResNeXt, which used cardinality to control the number of groups, several convolution operations are adopted to obtain

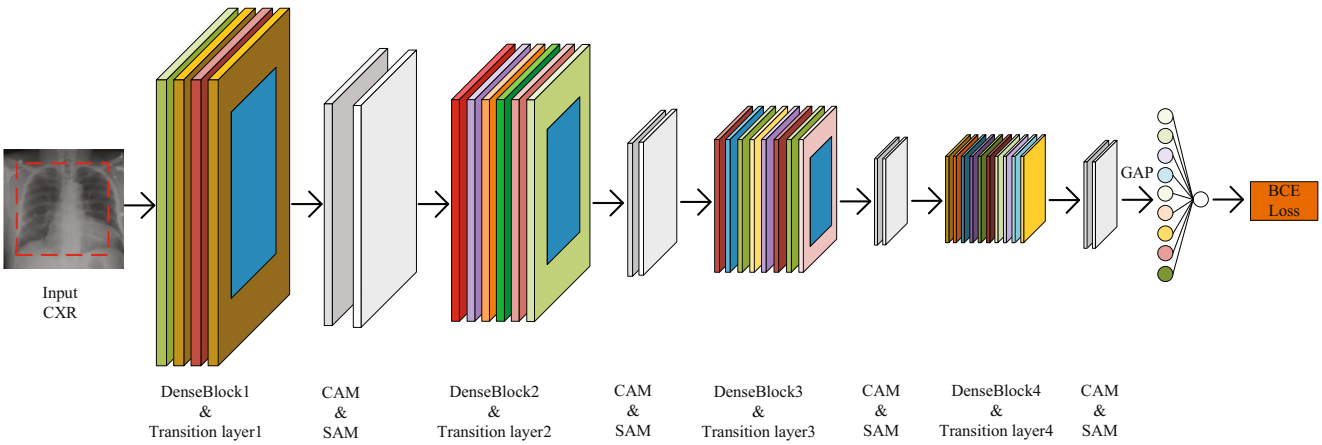

Fig. 2 Overall framework of CXRDANet, the input features pass through the Channel Attention Module (CAM) and Spatial Module (SAM) successively

higher-level representations of the input feature maps. Howard et al. [18] used depthwise separable convolution in MobileNetv3 to treat each channel as a group and model the spatial relationships in these groups. Li et al. [19] divided feature maps into multiple groups according to channel dimensions and learned well-distributed semantic feature representations in the space through grouping strategy.

2.2 Attention mechanism

The attention mechanism has been extensively studied in previous work. It assigns higher weights to the correct semantic feature areas while suppressing irrelevant information, thereby improving the effectiveness of the model. SE block [20] uses global average pooling for feature maps through spatial dimensions to obtain the spatial information of each feature map and then captures the importance of each channel through two Full Connection (FC) layers. ECA-Net [21] uses 1D convolution with adaptive kernel size based on SE block to obtain the relationship between channels and reduce the number of parameters. Wang et al. [22] proposed the Non-local module, which constructs long-range dependencies by calculating the correlation matrix between each spatial point in the feature map. SK-Net [23] adaptively selects the size of the core according to the input feature map and uses softmax attention to fuse multiple branches of different core sizes.

DANet [24] uses the self-attention mechanism to design the Position Attention Module (PAM) and Channel Attention Module (CAM) to capture the feature dependencies in the spatial dimension and the channel dimension. It is worth noting that due to the self-attention mechanism in PAM and CAM, DANet needs to generate the matrices of Query, Key, and Value for each of the two modules, and the size of each matrix depends on the input feature maps, which will significantly increase the number of network parameters. Different from DANet, the SAM and CAM we designed in CXRDANet are lightweight and effective, which will be introduced in deatil in Section 3.

2.3 Medical disease diagnosis using attention mechanism

Disease diagnosis needs to distinguish subtle differences between different conditions. Generally, diseases are characterized by lesion areas, containing key feature information for disease diagnosis. Tang et al. [25] propose AGCL, the framework using iterative attention-guided curriculum learning which leverages the severity-level attributes mined from radiology reports, for thoracic disease classification. Guan et al. [26] especially establish a three-branch network to avoid noise and gain the discriminative feature for disease classification. Yao et al. [27] predict the pathologies through the combination of multi-resolution saliency map and Long Short-Term Memory Network (LSTM). Recently, Ma et al. [28] use the hard example attention module to combine the misclassified positive cases with the original dataset to train the network to alleviate the class-imbalance problem.

3 Our method

The identification of cardiomegaly requires network to master the long-range dependencies of spatial dimension between the heart and lungs, meanwhile, pay attention to the dependencies between the channels features. In CXRDANet, we use DenseNet as the backbone network, as shown in Fig. 2, in each transition layer of the DenseNet, CAM and SAM are embedded to make the network learn more related features.

Specifically, CAM divides the feature maps into sub-tensors. For each sub-tensor, we spilt it into two branches in the channel dimension, and perform global average pooling and global max pooling respectively. After the FC layer and concatenate operation, CAM will capture the relationship between channels in the input feature maps. SAM divides the feature maps into two branches in the channel dimension and encodes the spatial information by performing horizontal and vertical strip pooling [29] operations on each branch. After passing through the FC layer, the two branches will capture the attention of the feature maps in the horizontal and vertical directions, and perform element-wise multiplication with the input feature maps to obtain the final spatial attention feature maps. Finally, to ensure the exchange of information between the two branches, we also used the channel shuffle operation in [30].

In addition, we found that through experiments, our method also has a good performance in the multi-label classification of chest diseases.

3.1 Channel attention module

As illustrated in Fig. 3, CAM aims to capture the relationship between channels in the input feature maps. We use the grouping strategy in CAM, which reduces model parameters, prevents overfitting due to insufficient data, and improves network performance [14]. Specifically, after grouping operation the parameters of FC layer is $(C/2G \times C/2G) \times 2G$, which $2G$ times less than before. The input feature maps $F \in \mathbb{R}^{C\times H\times W}$, where C, H, W, refer to the number of channel, height and width, respectively. Thus, CAM divides F into G groups, i.e., $F = [F_1, F_2, ..., F_G]$, $F_k \in R^{C/G\times H\times W}$. Table 5, shows the

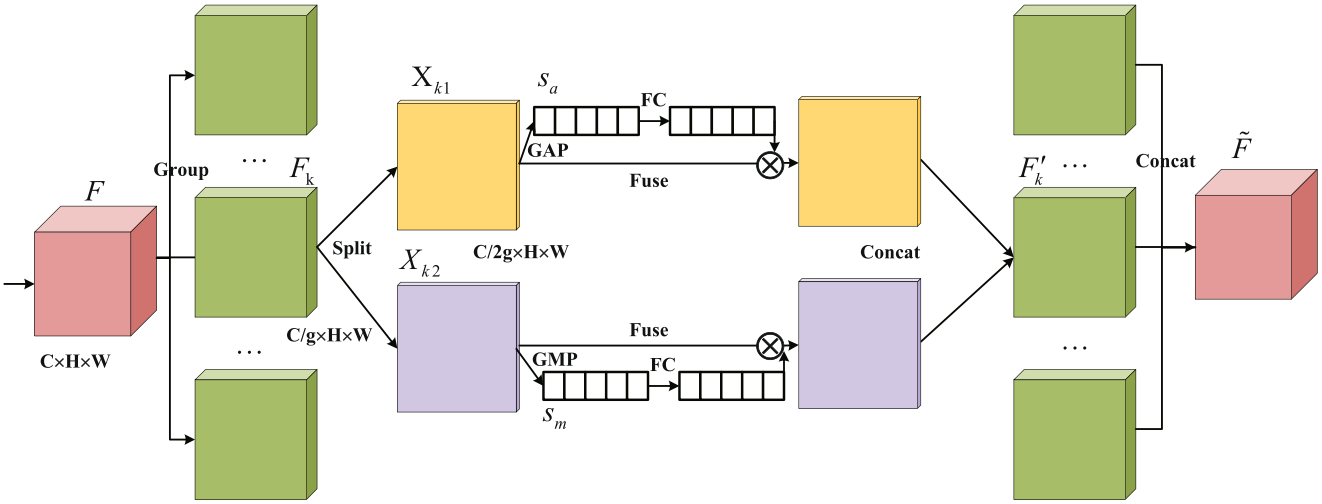

Fig. 3 Structure of Channel Attention Module

number of different groups and their corresponding CAM parameters and performance. In this paer, G is set to 4. The sub-tensor F_k spilt into two branches alone the channel dimension, i.e., $F_k = [X_{k1}, X_{k2}]$, $X_{k1}, X_{k2} \in \mathbb{R}^{C/2G \times H \times W}$. We use global average pooling and global max pooling on the two branches to aggregate spatial average features and distinctive object features:

$$s_a = F_{GAP}(X_{k1}) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} X_{k1}(i, j) \tag{1}$$

$$s_{\mathrm{m}} = F_{GMP}(X_{k2}) \tag{2}$$

Where $s_a, s_m \in \mathbb{R}^{C/2G \times 1 \times 1}$, F_{GAP} denotes global average pooling, F_{GMP} denotes global max pooling.

Then, the channel-wise statistics s_a and s_m will generate channel attention maps after FC layer, sigmoid activation and concatenate operation:

$$F'_k = Cat(\sigma(F_C(s_a)) \cdot X_{k1}, \sigma(F_c(s_m)) \cdot X_{k2}) \tag{3}$$

Where $Cat(.,.)$ refers to concatenate through channel dimensions, σ denotes sigmoid activation and F_C denotes FC layer.

Finally, CAM will obtain the weighted feature maps $\tilde{F}$ to express the importance of feature information, enhance useful information in the network and suppress useless information:

$$\tilde{F} = Cat(F'_1, F'_2, ..., F'_G) \tag{4}$$

3.2 Spatial attention module

In CXRDANet, SAM focuses on the correct spatial semantics areas. As shown in Fig. 4, we spilt input feature maps $\tilde{F}$ into two branches, i.e., $\tilde{F} = [\tilde{F}_1, \tilde{F}_2]$, $\tilde{F}_1, \tilde{F}_2 \in \mathbb{R}^{C/2 \times H \times W}$. In order to obtain long-range dependencies, we perform horizontal and vertical strip pooling to encode spatial information on the two branches respectively. Specifically, we deployed pooling kernels of

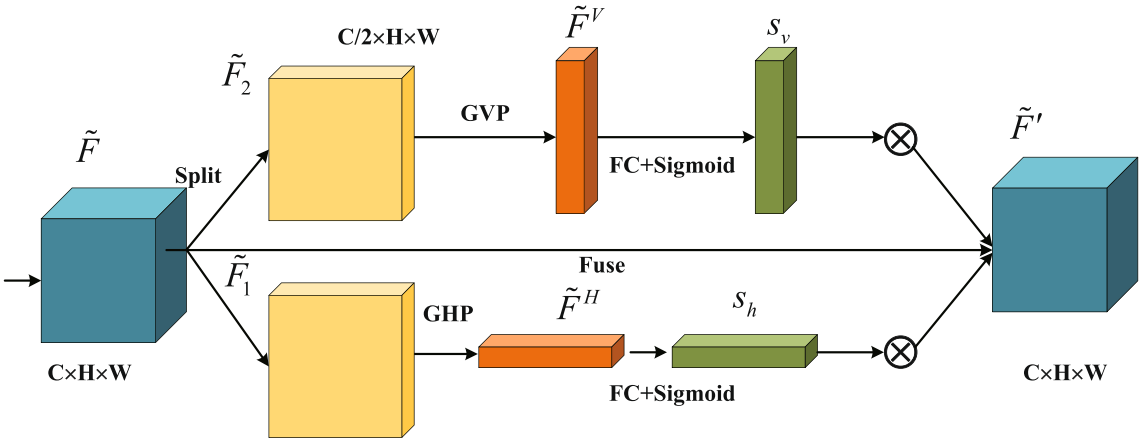

Fig. 4 Structure of Spatial Attention Module

$(1, H)$ or $(W, 1)$ to achieve horizontal and vertical strip pooling:

$$\tilde{F}^H = \frac{1}{H}\sum_{i=0}^{W}\tilde{F}_1(i, H) \tag{5}$$

$$\tilde{F}^V = \frac{1}{W}\sum_{j=0}^{H}\tilde{F}_2(W, j) \tag{6}$$

Equations (5) and and (6) represent horizontal strip pooling and vertical strip pooling, respectively $\tilde{F}^H \in \mathbb{R}^{C/2\times H\times 1}$, $\tilde{F}^V \in \mathbb{R}^{C/2\times 1\times W}$.

After $\tilde{F}^H$, $\tilde{F}^V$ are processed by the full connect operation and sigmoid activation respectively, they will obtain spatial attention in the horizontal and vertical directions:

$$s_h = \sigma(F_C(\tilde{F}^H)) \tag{7}$$

$$s_v = \sigma(F_C(\tilde{F}^V)) \tag{8}$$

In (7) and (8), $s_h \in \mathbb{R}^{C/2\times H\times 1}$, $s_v \in \mathbb{R}^{C/2\times 1\times W}$ SAM will get the final output after multiplying s_h , s_v and $\tilde{F}$:

$$\tilde{F}' = \tilde{F} \cdot \mathrm{s}_h \cdot \mathrm{s}_v \tag{9}$$

3.3 Loss Function

In the task of identifying cardiomegaly, we use Binary Cross Entropy as the loss function:

$$L = -y\log\widehat{y} - (1-y)\log(1-\widehat{y}) \tag{10}$$

Where y is the ground truth and $\widehat{y}$ is the predicted label. In particular, τ is the final output value of the network. We set the threshold of τ to 0.5. When $\tau \geq 0.5$, $y = 1$, otherwise $y = 0$.

In the task of multi-label CXR image classification, due to the imbalance of data and the differences in the feature and texture information of different diseases, the difficulty of learning is different. We use Focal loss [31] instead of BCE loss to solve mentioned problems:

$$L = -(1-\widehat{y})^{\beta} y\log\widehat{y} - \widehat{y}^{\beta}(1-y)\log(1-\widehat{y}) \tag{11}$$

Where β is a hyperparameter, we set $\beta = 2$, and the average loss of 14 diseases is taken as the overall loss of the network.

4 Experiment

4.1 Dataset

We use ChestX-ray14 [11] dataset and NLM-CXR dataset as the data source of the experiment.

The ChestX-ray14 dataset is a large CXR dataset released by the National Institutes of Health (NIH) in 2017 which contains 112,120 CXR images taken from 30,805 unique patients. Each CXR is marked with binary labels for 14 different diseases, of which 2772 are marked as cardiomegaly.

However, the disease labels are obtained using natural language processing technology from the radiology reports, claiming 90% accuracy. Thus, ChestX-ray14 dataset may not suitable for testing.

NLM-CXR is a set of images that are collected from various hospitals affiliated with the Indiana University School of Medicine and contains about 4000 CXRs. Each X-ray has a corresponding radiologist report. In the experiment for identifying cardiomegaly, we manually labeled 268 CXRs with cardiomegaly and 268 normal CXRs based on radiologists' reports.

Specifically, in the task of identifying cardiomegaly, similar to Candemir et al. [16], we use the CXRs labeled as cardiomegaly in the entire ChestX-ray14 dataset and use the same number of CXRs labeled as normal for training.

We randomly select 200 normal and cardiomegaly CXRs from the NLM-CXR dataset as the test set, and the rest were used for training. In the task of multi-label classification of chest diseases, we follow the official split standards of ChestX-ray14 dataset.

4.2 Implementation details

During training, we downscale the original images of size 256 × 256, and apply randomly cropping, the size of CXR images that will feed into the network is 224 × 224. We use the mean and standard deviation of the images in ImageNet for normalization. Random horizontal flipping is used in training set for data augment, and using Adam optimizer to optimize the network. In the phase of testing, the size of image is also downscaled to 256 × 256, then we perform center cropping to obtain images of size 224 × 224.

The backbone of CXRDANet is a 121-layer DenseNet pre-trained on ImageNet. In the first ten epochs of training, we use the warm-up [32] strategy to gradually increase the learning rate from 0.0001 to 0.001, and then use cosine annealing [33] to decay the learning rate to 0 after 50 epochs.

Our mothed is implemented by Pytorch 1.2 and trained on NVIDIA Gefoce Rtx 3080 GPU.

4.3 Comparisons with the state-of-the-art methods in identifying cardiomegaly

The comparison results between CXRDANet and the state-of-the-art methods [1, 16] in identifying cardiomegaly are

Table 1 Comparison of our method and the state-of-the-art methods

Method	Accuracy	Sensitivity	Specificity	F1	AUC
Zhou [1]	0.8636	0.9390	0.8103	0.8508	0.9327
Candemir [16]	0.8824	0.9258	0.8392	0.8873	0.9487
Ours	**0.9050**	**0.9445**	**0.8610**	**0.9059**	**0.9588**

Table 2 CXRDANet is compared with other methods on ChestX-ray14 dataset

Disease	Wang [11]	Yao [27]	Zhou [1]	Candemir [16]	Ma [28]	Guendel [35]	Guan [36]	Ours
Atel	0.7003	0.733	0.7562	0.7474	0.7627	0.767	0.781	0.7590
Card	0.8100	0.865	0.8496	0.8560	0.8835	0.883	0.880	**0.8983**
Effu	0.7585	0.806	0.8067	0.8152	0.8159	0.828	**0.829**	0.8266
Infi	0.6614	0.673	0.6828	0.7066	0.6786	0.709	0.702	**0.7099**
Mass	0.6933	0.718	0.7952	0.8042	0.8012	0.821	**0.834**	0.8320
Nodule	0.6687	0.777	0.7477	0.7397	0.7293	0.758	**0.773**	0.7533
Pneu1	0.6580	0.684	0.7073	0.7233	0.7097	0.731	0.729	**0.7356**
Pneu2	0.7993	0.805	0.8269	0.8637	0.8377	0.846	0.857	**0.8760**
Cons	0.7032	0.711	0.7232	0.7203	0.7443	0.745	**0.754**	0.7512
Edema	0.8052	0.806	0.8228	0.8260	0.8414	0.835	0.850	**0.8636**
Emph	0.8330	0.842	0.8782	0.8938	0.8836	0.895	**0.908**	0.8898
Fibr	0.7859	0.743	0.7979	0.7808	0.8007	0.818	0.830	**0.8359**
P_T	0.6835	0.724	0.7541	0.7463	0.7536	0.761	0.778	**0.7894**
ernia	0.8717	0.775	0.8872	0.8267	0.8763	0.896	0.917	**0.9404**
Mean	0.7451	0.761	0.7881	0.7727	0.7941	0.807	0.816	**0.8186**

We calculated the AUC score of each disease and the average AUC score of 14 diseases

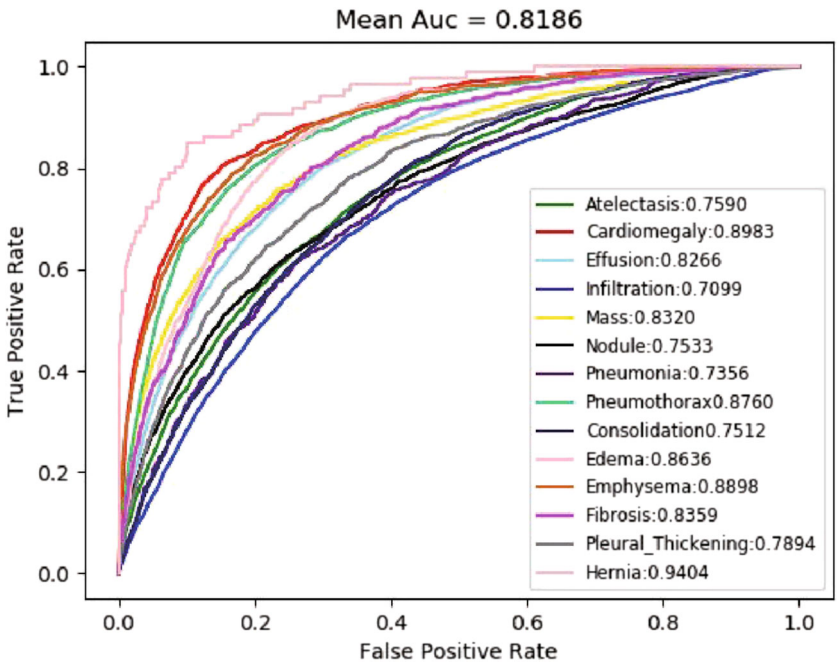

Fig. 5 ROC curve and AUC scores of CXRDANet on Chest X-ray14 dataset

Table 3 Comparison of ablation studies with different network structures

	Accuracy	Sensitivity	Specificity	F1	AUC
Complete model	0.9050	0.9445	0.8610	0.9059	0.9588
-SAM&CAM	0.8561	0.8615	0.8507	0.8550	0.9325
-SAM	0.8725	0.8901	0.8500	0.8696	0.9493
-CAM	**0.8825**	**0.8923**	**0.8706**	**0.8810**	**0.9571**

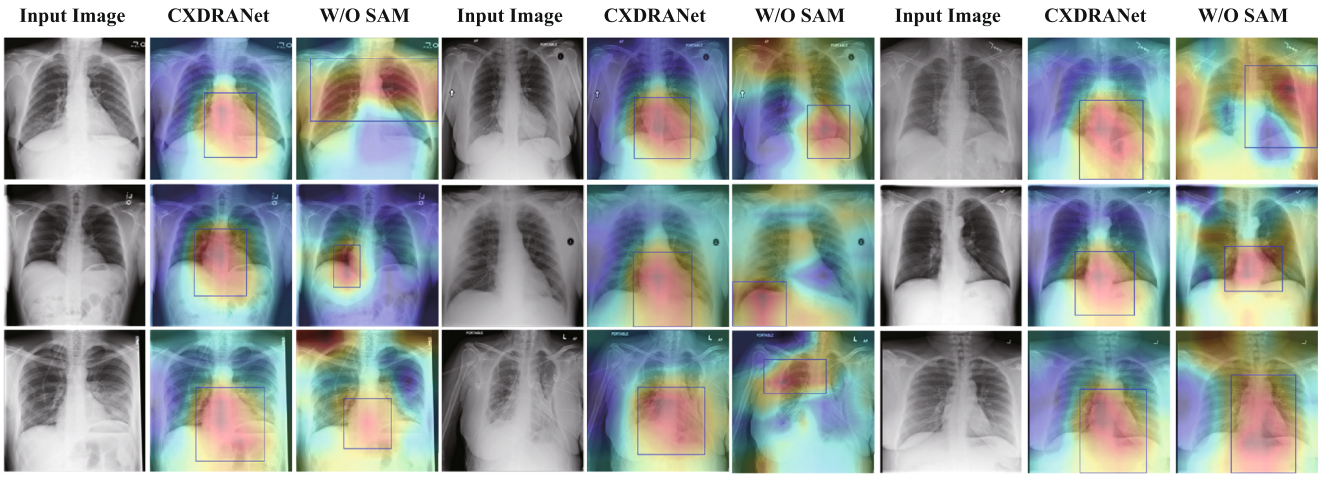

Fig. 6 Grad-CAM visualization results. All target layer selected is the last convolutional layer

Table 4 Combining methods of CAM and SAM

Method	Accuracy	Sensitivity	Specificity	F1	AUC
CAM&SAM in parallel	0.8896	0.8968	0.8135	0.8653	0.9031
SAM-CAM	0.8912	0.9342	0.8463	0.8842	0.9337
CAM-SAM	**0.9050**	**0.9445**	**0.8610**	**0.9059**	**0.9588**

Table 5 CAM parameters of different grouping numbers(G) and corresponding performance

G	Param. of CAM	Accuracy	Sensitivity	Specificity	F1	AUC
1	176,284	0.9037	0.9410	0.8523	0.8964	0.9453
4	44,392	**0.9050**	**0.9445**	**0.8610**	**0.9059**	**0.9588**
8	22,314	0.9025	0.9398	0.8372	0.8853	0.9212
16	11,284	0.8934	0.9244	0.8389	0.8772	0.9039

shown in Table 1. We use the same training set and test set to train CXRDANet and other models (Table 2).

We use accuracy, sensitivity, specificity, F1 and AUC to evaluate the performance of clinical diagnosis system. Accuracy is the most commonly used evaluation metric to measure the performance of a classifier. Compared with previous work, the accuracy of our proposed method for identifying cardiomegaly is increased by more than 2%.

Generally, the significance level of diagnosis system is measured with specificity and sensitivity and the consistency of the system is observed with Area Under Curve (AUC) [34]. The sensitivity and specificity of our model are 0.9445 and 0.8610, respectively. F1 score is the harmonic mean of the precision and recall which reflect the comprehensive performance of the model. F1 score of CXRDANet is 0.9059, which better than other methods. The larger area under ROC curve (AUC), the better the effect. In our model, the AUC reached 0.9588, which is the best among all methods (Fig. 5).

4.4 Ablation study

As shown in Table 3. We conducted additional ablation experiments to evaluate the effectiveness of different modules. In this work, we regard our method as a complete model, remove the SAM or CAM individually to prove the effectiveness of CXRDANet.

We removed SAM from the complete model, AUC decreased from 0.9588 to 0.9493, accuracy and F1 score decreased by 3.25% and 3.63% respectively. Sensitivity decreased by 5.44%, specificity only decreased by 1.10%. Then, the CAM is removed from the model, the AUC of the model decreased from 0.9588 to 0.9571, accuracy and F1 score decreased by 2.25% and 2.49% respectively. Sensitivity decreased by 5.22%, but specificity increased by 0.96%. The model without CAM is better than that without SAM in all the 5 metrics, which indicates that SAM has played a greater role.

In clinical practice, the diagnosis of cardiomegaly usually requires the radiologist to calculate the patient's cardiothoracic ratio(CTR) [37]. Therefore, the network needs to focus on the heart and lungs in the CXR image. For the qualitative analysis, as shown in Fig. 6, we adopt Grad-CAM [38] to CXRDANet and CXRDANet without SAM. It's clearly seeing that the Grad-CAM masks of the CXRDAnet cover the heart regions better than the CXRDANet removed SAM. That is, the SAM can help CXRDANet learn long-range dependencies in CXR image, thereby gathering informative features from heart regions.

CAM and SAM can be placed in parallel or sequentially and there are three different ways of arranging the CAM and SAM: sequential CAM-SAM, sequential SAM-CAM, and parallel use of both CAM and SAM. Theoretically, the channels of the feature map represent different features, and the CAM adaptively recalibrates the of the channe-wise feature responses through learning. SAM attempts to highlight the semantic features related to the disease in each channel. If the SAM-first order is adopted, due to SAM weighted some semantic features, CAM's learning of the importance of channels will be affected; if CAM and SAM are used in parallel, CAM's recalibration of the channels will not benefit SAM. In Table 4, our experimental result shows that the CAM-SAM order is better than the SAM-CAM and parallel use of both CAM and SAM (Table 5).

4.5 Comparisons with state-of-the-art methods in multi-label classification of chest diseases

As shown in Table 2, We compare CXRDANet with some state-of-the-art methods [1, 11, 16, 27, 28, 35, 36] to prove the effectiveness of the proposed method based on the ChestX-ray14 dataset. In Fig. 5, we can see that CXRDANet reached the mean AUC of 0.8186 for all 14 diseases in the ChestXray-14 dataset.

In terms of the AUC of each disease, except for the AUC of atelectasis, effusion, mass, nodule, consolidation and emphysema, which are lower than the previous work, the AUC of the other diseases are higher than the other three methods. In summary, our method can well complete the task of multi-label classification of chest diseases.

5 Conclusion

We propose a novel dual attention network (CXRDANet) that can collect rich global context information and reduce the interference of irrelevant areas or noise. The experimental results prove that CXRDANet achieves state-of-the art performance on the diagnosis of heart disease and the multi-label classification of chest diseases. However, our method currently only processes X-rays. In the future, we will combine other more advanced diagnostic methods to further improve diagnostic efficiency.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

In Case Animals Were Involved Ethical Approval Animals were not involved.

And/or in Case Humans Were Involved Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

Competing of Interests The authors have no conflict of interest.

References

1. Zhou S, Zhang X, Zhang R (2019) Identifying cardiomegaly in chestx-ray8 using transfer learning. Stud Health Technol Inf 264:482–486
2. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, et al. (2017) Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv:1711.05225
3. Abiyev RH, Ma'aitah MKS (2018) Deep convolutional neural networks for chest diseases detection. J Healthcare Eng 2018:1–11
4. Shen D, Wu G, Suk H-I (2017) Deep learning in medical image analysis. Ann Rev Biomed Eng 19:221–248
5. Ker J, Wang L, Rao J, Lim T (2017) Deep learning applications in medical image analysis. IEEE Access 6:9375–9389
6. Kadam VJ, Jadhav SM, Vijayakumar K (2019) Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression. J Med Syst 43(8):263
7. Rocha J, Cunha A, Mendonça AM (2020) Conventional filtering versus u-net based models for pulmonary nodule segmentation in ct images. J Med Syst 44(4):1–8
8. Doshi D, Shenoy A, Sidhpura D, Gharpure P (2016) Diabetic retinopathy detection using deep convolutional neural networks. In: 2016 International Conference on Computing, Analytics and Security Trends (CAST). IEEE, pp 261–266
9. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
10. Huang G, Liu Zx, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
11. Wang X, Peng Y, Lu L, Lu Zx, Bagheri M, Summers RM (2017) Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2097–2106
12. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghgoo B, Ball R, Shpanskaya K et al (2019) Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 33, pp 590–597
13. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
14. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105
15. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
16. Candemir S, Rajaraman S, Thoma G, Antani S (2018) Deep learning for grading cardiomegaly severity in chest x-rays: an investigation. In: 2018 IEEE Life Sciences Conference (LSC). IEEE, pp 109–113
17. Xie S, Girshick R, Dollár P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1492–1500
18. Howard A, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V et al (2019) Searching for mobilenetv3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 1314–1324
19. Li X, Hu X, Yang J (2019) Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. arXiv:1905.09646
20. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
21. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q (2020) Eca-net: Efficient channel attention for deep convolutional neural networks, 2020 IEEE. In: CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE
22. Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7794–7803
23. Li X, Wang W, Hu X, Yang J (2019) Selective kernel networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 510–519
24. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H (2019) Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 3146–3154
25. Tang Y, Wang X, Harrison AP, Lu L, Xiao J, Summers RM (2018) Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp 249–258
26. Guan Q, Huang Y, Zhong Z, Zheng Z, Zheng L, Yang Y (2020) Thorax disease classification with attention guided convolutional neural network. Pattern Recogn Lett 131:38–45
27. Yao L, Poblenz E, Dagunts D, Covington B, Bernard D, Lyman K (2017) Learning to diagnose from scratch by exploiting dependencies among labels. arXiv:1710.10501
28. Ma Y, Zhou Q, Chen X, Lu H, Zhao Y (2019) Multi-attention network for thoracic disease classification and localization. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp 1378–1382
29. Hou Q, Zhang L, Cheng M-M, Feng J (2020) Strip pooling: Rethinking spatial pooling for scene parsing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 4003–4012
30. Zhang X, Zhou X, Lin M, Sun J (2018) Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6848–6856
31. Lin T-Y, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp 2980–2988
32. Goyal P, Dollár P, Girshick R, Noordhuis P, Wesolowski L, Kyrola A, Tulloch A, Jia Y, He K (2017) Accurate, large minibatch sgd: Training imagenet in 1 hour. arXiv:1706.02677
33. He T, Zhang Z, Zhang H, Zhang Z, Xie J, Li M (2019) Bag of tricks for image classification with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 558–567
34. Agnes SA, Anitha J, Pandian SIA, Peter JD (2020) Classification of mammogram images using multiscale all convolutional neural network (ma-cnn). J Med Syst 44(1):30
35. Guendel S, Grbic S, Georgescu B, Liu S, Maier A, Comaniciu D (2018) Learning to recognize abnormalities in chest x-rays with location-aware dense networks. In: Iberoamerican Congress on Pattern Recognition. Springer, pp 757–765

36. Guan Q, Huang Y (2020) Multi-label chest x-ray image classification via category-wise residual attention learning. Pattern Recogn Lett 130:259–266
37. Frishman WH, Nadelmann J, Ooi WL, Greenberg S, Heiman M, Kahn S, Guzik H, Lazar EJ, Aronson Miriam (1992) Cardiomegaly on chest x-ray: prognostic implications from a ten-year cohort study of elderly subjects: a report from the bronx longitudinal aging study. Amer Heart J 124(4):1026–1030
38. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp 618–626