DENXFPN: PULMONARY PATHOLOGIES DETECTION BASED ON DENSE FEATURE PYRAMID NETWORKS

Jun Xiao[†]*, Yuanxing Zhang*[†]*, Kaigui Bian*[†]*, Guopeng Zhou*[¶]*, Wei Yan*[†]

[†]Peking University, Beijing, China
[¶]Peking University First Hospital, Beijing, China

ABSTRACT

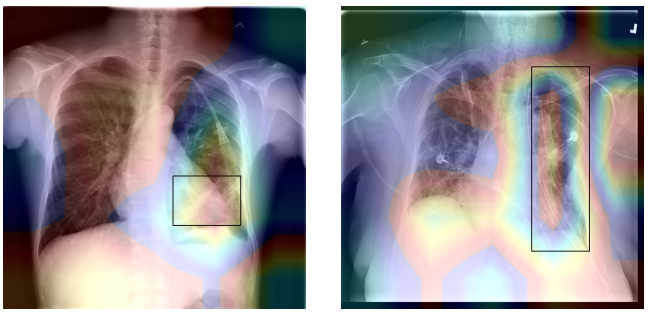
Computer-aided detection and diagnosis (CAD) have been applied to many departments of medical institutions, and early detection of diseases can prevent serious health loss. Pulmonary diseases generate negative effects on human health, even leading to death. The chest X-ray is a common examination for diagnosis of pulmonary diseases. The experienced radiologist can quickly infer patients' symptoms by screening the chest X-ray image. While in some developing countries or remote rural areas, due to the lack of experienced radiologists or doctors, patients may be misdiagnosed. Many efforts have been spent on developing an effective auxiliary detection system to provide medical workers with evidence on diseases. In particular, detecting pulmonary complication via chest X-ray images is one of the most challenging tasks. In this paper, we transform the pulmonary complications detection task into a multi-binary classification task for each pulmonary pathology, and propose a new classification model, DenXFPN (for X-ray). DenXFPN combines multiple feature maps at different scales extracted through a densely convolutional neural network. Our model achieves 0.827 on the area under the receiver operating characteristic curve (AUC) metric on average, which outperforms the state-of-the-art results on most of all pathologies in the Chest X-ray14 dataset.

Index Terms— chest X-rays classification, convolutional neural network, feature pyramid

1. INTRODUCTION

The past decade has witnessed the rapid development of the intelligent medical diagnosis system [1]. It has been widely used for assisting the doctors to provide the accurate and comprehensive diagnosis [2]. The chest X-ray is one typical application which greatly benefits from the intelligent medical diagnosis system. For example, the chest X-ray is the most commonly available radiological examinations for screening and diagnosis of many pulmonary diseases, e.g., Pneumonia, Nodule, Pneumothorax, Infiltration. According to a recent report [3], pneumonia kills nearly 1 million children younger than five years of age each year globally, which is greater than the number of deaths from any infectious diseases such as HIV infection, malaria, or tuberculosis. Hence, the early detection of these kinds of diseases could save lives, and the accurate results from the X-rays play a crucial role in helping patients obtain timely treatment.

Recently, many efforts have been spent on improving the accuracy in identifying diseases from the bio-signal. The state-of-the-art e-health techniques have been verified to show good performance in fine-grained disease analysis [4] and medical feature extraction [5]. Furthermore, the X-rays image analysis benefits from the prosperity of the deep learning methods [6, 7]. Regarding the analysis of the chest X-ray, it is applied to detect pulmonary diseases in chest X-ray images by using convolutional neural networks which are pre-trained on large-scale datasets for general image classification. Prior works of AlexNet [8] and Densely Connected Convolutional Networks (DenseNet) [9] can help identify the detailed evidence in the image for diagnosis of the chest-related diseases. For example, Yao et al. [10] and Pranav Rajpurkar et al. [11] both use DenseNet to conduct the feature extraction. In addition, it is an important method to enhance the accuracy of the detection by using feature pyramid, e.g., Spatial Pyramid Pooling (SPP net) [12], Single Shot MultiBox Detector (SSD) [13], Feature Pyramid Network (FPN) [14].



(a) Pneumonia (b) Pneumothorax

Fig. 1. Illustration of two examples of NIH Chest X-ray14 images. The diseases are usually diagnosed via small hints in the image.

However, due to the uniqueness of medical images compared to regular images and the scarcity of medical region annotation in the image, clinical diagnosis by chest X-ray images is more difficult than general classification tasks via reg-

ular images. It usually faces the following three challenges:

- **No region annotation in images**. Doctors provide comprehensive diagnoses of X-ray images through their extensive experience. In the X-ray dataset, there is usually only the label information of the disease, but there lack indications of the specific annotation of the disease in the image. Therefore, an effective machine-assisted decision-making model requires automated feature extraction of X-ray images and gives accurate diagnoses.

- **High accuracy requirements**. Medical diagnosis requires high accuracy, as misdiagnosis can lead to negative effect. However, not all doctors have sophisticated experience, especially in developing countries and remote rural areas. The CAD should provide the confident diagnosis so that the inexperienced medical workers can refer to the comment generated by the CAD.

- **Pulmonary complications**. Early and accurate detection of pulmonary complications guarantees that the patients would receive proper treatment. The CAD is expected to identify the evidence for the pulmonary complications without false alarms so that the medical workers can receive complete information of the diseases and suit the remedy to the case.

This paper is conducted on the largest open source chest X-ray dataset, namely Chest X-ray14 [15], which contains more than 100 thousand images. The dataset was proposed by Wang et al. [15] by using natural neural language techniques for extracting disease categories from medical reports. Fig.1 illustrates two example of chest X-ray images, where we can see that the diagnosis of diseases is usually determined via detailed information in the image. Based on the dataset, we transform the task into 14 binary classifications corresponding to the 14 pulmonary complications, as the patients may have more than one diseases at the same time. In order to address the above challenges, we proposed DenXFPN. DenXFPN leverages four serial Dense Blocks, Bottleneck Layers and Transition layers to learn knowledge of different scales. The features generated by each transition layer are then concatenated by an FPN and form a feature map. The feature map is treated as the evidence for identifying whether a disease occurs in the X-ray image. We compare the performance of our proposed network with several state-of-the-art models on the Chest X-ray14 dataset. Our network outperforms the state-of-the-art results on 12 of 14 pathologies in the Chest X-ray14, and achieves the best performance on the AUC metric on average over 14 pathologies. Further, the performance of DenXFPN on the remaining two pathologies are also close to the state-of-the-art results.

2. DESIGN OF DENXFPN

2.1. Overview of DenXFPN

We propose DenXFPN for the chest X-rays classification task. As shown in Fig.1, DenXFPN consists of three steps: feature extraction, feature pyramid and combined feature classification.The details are described in Section 2.2 - 2.4.

The input image is fed into a densely-connected convolutional neural network to learn the representation of the image. We can get four different representation of the image at different scales. By using feature pyramid, we upsample the upper layer and add it to the lower layer. The merged feature maps are then fed into a convolutional layer with kernel size 3×3 and global average pooling for the final representation. Finally, we concatenate all feature maps as a vector to classify the 14 pathologies.

2.2. Feature Extraction

The visual features of the image can be extracted by the convolutional neural network of which the greatest advance is local connectivity and weight sharing, compared with the fully-connected network. A convolutional layer is composed of multiple feature maps and each feature map include multiple neurons. In the high-level neural network, convolution operation can break through the limitation of the traditional filter and extract desired features according to an objective function. Generally, we use the output of the final convolutional layer as the representation of the input image.

In this work, we adopt Densely Connected Convolutional Networks (DenseNet) [9] to extract features of the input images. DenseNet is composed of Dense Blocks, Bottleneck Layers, and Transition layers. Dense block is the combination of convolutional layers, batch normalization (BN) and rectified linear unit activation. And in each dense block, the input from each layer comes from the output of all the previous layers. In addition, bottleneck layer refers to the 1×1 convolution operation in front of each dense block's 3×3 convolution. Bottleneck layer can effectively reduce the number of feature maps, diminish computation by reducing dimensions and integrate features of each channel. Transition layer is a 1×1 convolution layer and a 2×2 average pooling layer between two dense blocks to further compress parameters. Therefore, DenseNet alleviates gradient vanishing, strengthens the transmission of feature and retrieves features effectively. We use the output of each transition layer in DenseNet-121 [9] as the representation of input images. Finally, we integrate all the individual feature maps as the final feature map by a Feature Pyramid Networks (FPN).

2.3. Feature Pyramid

Feature Pyramid is a fundamental component of multi-scale target detection system. Detection on different feature scales is equivalent to ensemble learning which enables a model to detect objects across a broad range of scales by scanning the

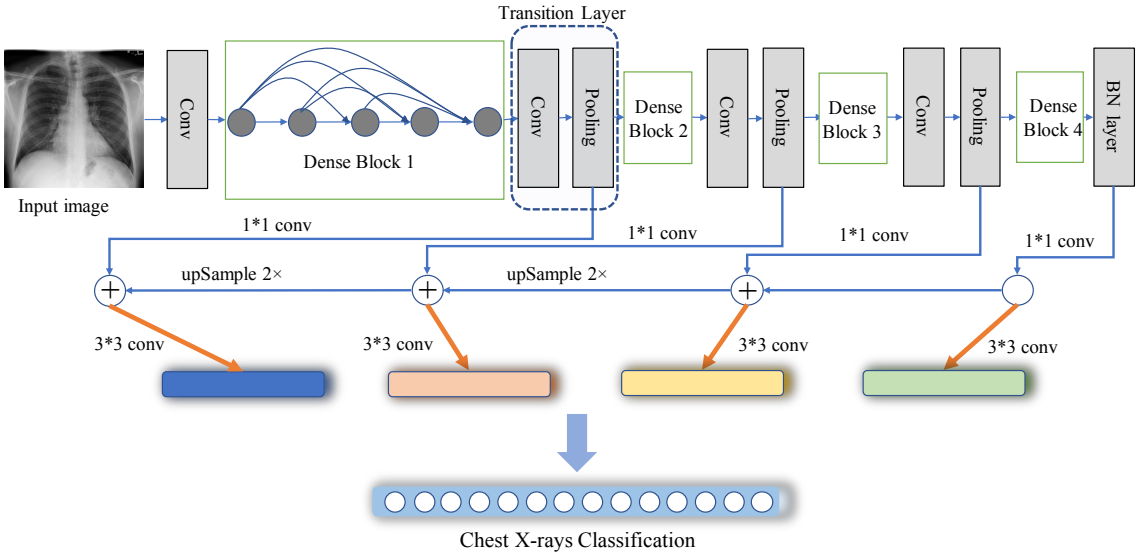


Fig. 2. Structure of DenXFPN. DenXFPN learns the image features at different scales via dense blocks and transition layers. The image features are fused to a representation by an FPN.

model over both positions and pyramid levels. However, typical feature pyramid networks, e.g., Featurized image pyramid, Single feature map, and Pyramidal feature hierarchy have obvious limitations which make computational time too long or don't take full advantage of the underlying features. Feature Pyramid Networks (FPN) link the high-level features of low-resolution and high-semantic information with the low-level features of high-resolution and low-semantic information from top to bottom, which make the features at all scales own rich semantic information.

We upsample the abstract and semantic high-level feature map, and then horizontally connect the feature to the previous layer that enhances the representation of the high-level feature. It is worth noting that the two-layer features of the lateral connection are the same in spatial dimensions which take advantage of the underlying location details. After getting feature representation on the transition layer, we add a 1×1 convolution kernel to produce a rough feature map. Then, we upsample the upper layer which can get the same dimension. We add all the features maps to fuse the knowledge learned from different scales. Last but not least, we use a 3×3 convolution kernel to process the merged feature map and a global average pooling to generate the final feature map, in order to eliminate the aliasing effect of upsampling.

As shown in Fig.1, we get four feature vectors by upsampling and adding operations on the different feature map scales. Because of the same dimension on the last two feature maps, we don't upsample the final feature map. Consequently, a 1024-dimensional vector is obtained by concatenating all feature vectors.

2.4. Combined Feature Classification

Chest X-rays classification is formulated as a multi-binary classification problem corresponding to each pulmonary pathology. The images labeled with the diseases will be treated as positive examples in the corresponding binary classification task, and all the other images are seen as negative examples. The outputs of our proposed framework are a vector of binary labels indicating the diagnosis of the following 14 pathology classes: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening, Hernia. We apply an element-wise sigmoid nonlinearity after 14-dimensional fully connected layer. During training, we take the sum of unweighted binary cross entropy losses as our loss function to optimize, i.e.,

$$L = \sum_{c=1}^{14} [-y_c log \hat{y_c} - (1 - y_c) log(1 - \hat{y_c})], \qquad (1)$$

where $\hat{y_c}$ stands for the predicted probability that the image is diagnosed with the c-th pathology and y_c indicates whether the disease does exist.

3. EVALUATION

3.1. Settings

We conduct experiments on the currently largest released chest X-ray dataset, NIH Chest X-ray14 dataset. We randomly split the dataset into training set (70%), validation set (10%) and test set (20%) and we ignore the personal attributes attached to the images. The weights of the network are initialized with that from a model pre-trained on ImageNet [16].

Pathology	Wang et al.(2017) [15]	Yao et al.(2017) [10]	CheXNet [11]	Ours
Atelectasis	0.716	0.772	0.795	**0.811**
Cardiomegaly	0.807	0.904	0.889	**0.911**
Effusion	0.784	0.859	0.871	**0.881**
Infiltration	0.609	0.695	0.693	**0.704**
Mass	0.706	0.792	0.815	**0.833**
Nodule	0.671	0.717	0.721	**0.740**
Pneumonia	0.633	0.713	0.746	**0.761**
Pneumothorax	0.809	0.841	0.837	**0.859**
Consolidation	0.708	0.788	0.782	**0.799**
Edema	0.835	0.882	**0.884**	0.883
Emphysema	0.815	0.829	0.876	**0.901**
Fibrosis	0.769	0.767	0.797	**0.815**
Pleural Thickening	0.708	0.765	0.759	**0.773**
Hernia	0.767	**0.914**	0.883	0.907
Average	0.738	0.798	0.810	**0.827**

Table 1. Fourteen pathologies and their AUCs, including the average AUC over all pathologies.

We use an initial learning rate of 0.001 that is decayed by 10 when the loss on the validation set no longer decreases after one epoch. The size of the mini-batches is set to 32 and we use Adam algorithm [17] with standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) to train the network end-to-end. In our experiment, we first downscale the images to 224×224 and enhance the images by normalization based on the mean and standard deviation of images in the ImageNet training set.

3.2. Details of Dataset

The chest X-ray14 dataset is the currently largest released chest X-ray dataset sponsored by the National Institutes of Health(NIH). It provides 112,120 frontal-view X-ray images with 14 different disease labels collected from 30,805 unique patients. To create these labels, Wang et al. [15] used Natural Language Processing (NLP) to text-mine disease classifications from the associated radiological reports. Importantly, this dataset demonstrates that these commonly occurred thoracic diseases can be detected and even spatially-located via a unified weakly-supervised multi-label image classification framework. The labels are expected to be >90% accurate and suitable for weakly-supervised learning. Chest X-ray14 can enable the data-hungry deep neural network paradigms for fostering clinically essential applications.

3.3. Performance Metric

We utilize the area under the curve (AUC) of receiver operating characteristic (ROC) to evaluate the performance of our proposed framework. Based on the definition of four basic measurements, namely true positive (TP), false positive (FP), true negative (TN), and false negative (FN), we can derive AUC metrics as listed:

$$\text{TPR} = \text{TP}/(\text{TP} + \text{FN}),$$

$$\text{FPR} = \text{FP}/(\text{FP} + \text{TN}),$$

$$\text{AUC} = \int_{-\infty}^{+\infty} \text{TPR}(T)\text{FPR}'(T)dT.$$

3.4. Experiment Results

We compare the proposed DenXFPN with three state-of-the-art pulmonary pathologies detection algorithms, namely CheXNet [11] and the models proposed in [15] and [10]. We reproduce the algorithms and conduct them on our splitting scheme of the dataset. We report the score of AUC metric of all algorithms on each of the 14 pathologies, and the results are listed in Tab. 1. We see that the proposed DenXFPN outperforms the state-of-the-art results on 12 of 14 pathologies in the Chest X-ray14, and achieves the best score on AUC metric on average over all 14 pathologies. Meanwhile, the results of the remaining two pathologies are also close to the state-of-the-art results.

4. CONCLUSIONS

In this paper, we proposed a classification network for the pathologies detection task, named as DenXFPN. DenXFPN learns multiple feature maps at different scales via densely-connected layers and fuses the feature maps via FPN. On the chest X-ray 14 dataset, compared with several state-of-art pulmonary pathologies detection algorithms, DenXFPN obtains the best performance on most pathologies and achieves the highest score on the AUC metric.

Acknowledgment

This work is partially supported by the National Key Research and Development Program No. 2017YFB0803302, the National Natural Science Foundation of China under Grant No. 61572051 and CERNET Innovation Project NGII20160709.

5. REFERENCES

[1] Manjeevan Seera and Chee Peng Lim, "A hybrid intelligent system for medical data classification," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2239–2249, 2014.

[2] Annan Li, Yunhong Wang, Jun Cheng, and Jiang Liu, "Combining multiple deep features for glaucoma classification," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 985–989.

[3] CDC report, "Pneumonia can be preventedvaccines can help," Accessed 6 November, 2017. https://www.cdc.gov/features/pneumonia/index.html, 2017.

[4] Hossein Nejati, Hamed Alizadeh Ghazijahani, Milad Abdollahzadeh, Tooba Malekzadeh, Ngai-Man Cheung, Kheng Hock Lee, and Lian Leng Low, "Fine-grained wound tissue analysis using deep neural network," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.

[5] Rohan Banerjee, Sakyajit Bhattacharya, and Shahnawaz Alam, "Time series and morphological feature extraction for classifying coronary artery disease from photoplethysmogram," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 950–954.

[6] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram Van Ginneken, and Clara I Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[7] Hoo-Chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, and Ronald M Summers, "Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2497–2506.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks.," in *CVPR*, 2017, vol. 1, p. 3.

[10] Li Yao, Eric Poblenz, Dmitry Dagunts, Ben Covington, Devon Bernard, and Kevin Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," *arXiv preprint arXiv:1710.10501*, 2017.

[11] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al., "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *European conference on computer vision*. Springer, 2014, pp. 346–361.

[13] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[14] Tsung-Yi Lin, Piotr Dollár, Ross B Girshick, Kaiming He, Bharath Hariharan, and Serge J Belongie, "Feature pyramid networks for object detection.," in *CVPR*, 2017, vol. 1, p. 4.

[15] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017, pp. 3462–3471.

[16] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. Ieee, 2009, pp. 248–255.

[17] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.