Project Report of
Machine Learning Course
July, 2018

Peng Cao, Rundi Wu
Peking University

Abstract

In this project, we implemented the task of abnormality detection, abnormality localization and image retrieval for musculoskeletal radiographs on MURA dataset. In task one, we trained some models based on ResNet and DenseNet to classify the input study as either normal or abnormal and use the ensembled prediction for output. We designed FuseNet, which fuses the outcome of the global image and that of the local area, to improve the performance of prediction. In task two, we tried to localize and visualize the abnormal areas on radiograph by applying gradient class activation map. In task three, we used the features extracted by our model as codes to output the images in the training data that is most similar to the input image.

1 Task One: Abnormality Detection

1.1 DenseNet / ResNet

1.1.1 Configuration

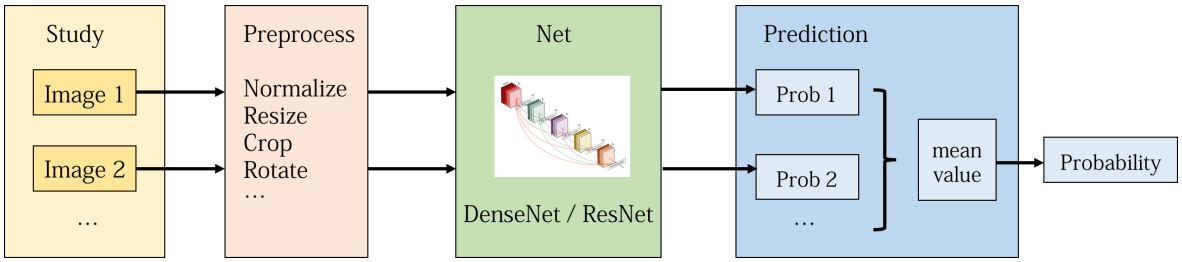


Figure 1: Prediction pipeline of the model

We firstly tried to reimplement the work by Rajpurkar et al.[1]. Figure 1 shows the prediction pipeline of the model. The model takes a study as input. It firstly feeds each image in the study into the network and gets the probability of abnormality of each image. Then it calculates the arithmetic mean value of the probabilities and outputs it as the final probability of the study. If the probability is higher than 0.5, we can regard it as an abnormal study.

The network uses DenseNet-169 [2]. The final fully connected layer is replaced with one that has a single output, and a sigmoid nonlinearity is added. To solve the problem of class imbalance, the loss function of each image X is defined by the weighted binary cross entropy

$$L(X, y) = -w_{T,1} y \log P(Y = 1|X) - w_{T,0}(1 - y) \log P(Y = 0|X)$$

where y is the label of the study, $P(Y = 1|X)$ is the probability that the network classifies the image as abnormal, $P(Y = 0|X)$ is the probability that the network classifies the image as normal, $w_{T,1}$ is the proportion of normal images and $w_{T,0}$ is the proportion of abnormal images in the dataset.

We also tried ResNet-50 [3] under the same configuration.

1.1.2 Training

Unlike the work by Rajpurkar et al.[1], we do more data-augmentation at training stage because using only horizontal flip and rotation will cause overfitting. So we further do random sized crop in pre-prossing stage. In other word, each input image is resized to 256×256 and then random-cropped to 224×224 before feeding into the network. At test time, we do center crop instead.

We use the weights pretrained on Imagenet and tune the hyper-parameters. After trying different batch sizes 8, 16, 32, 64, 128 and the corresponding learning rate, we chose to use batch size 16 and an initial learning rate 0.0001 that is decayed by a factor of 10 each time the validation loss plateaus after an epoch as the hyper parameters, which perform highest accuracy and AUC of ROC on validation.

1.2 FuseNet

Considering that the abnormalities of bones usually happens in small areas in the image and inspired by the Attention Guided Convolutional Neural Network by Qingji Guan et al.[6], we designed a network to combine the global features and the local features on radiographs to make abnormality detection.

1.2.1 Configuration

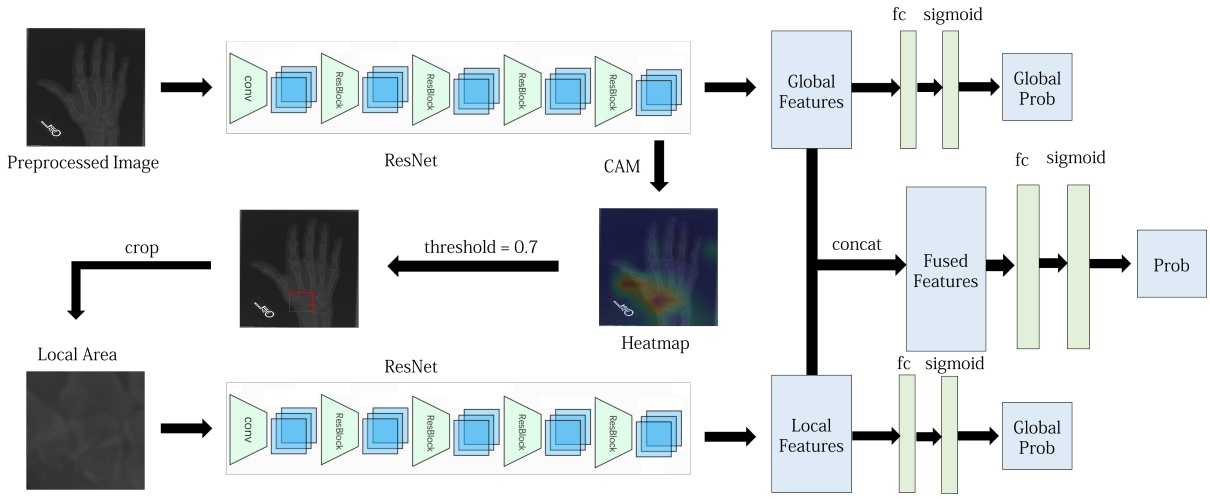


Figure 2: Architecture of FuseNet

The network consists of three parts: global branch, local branch and classifier. An input image first goes through the global branch and output the global feature (the 1-d feature after global average pooling). Then a heatmap is generated based on the feature maps(the 7×7 features before global average pooling). After that, the heatmap is turned into a binary map under a threshold. By locating the maximum connected component on the binary map, the local area is cropped from the input image(preprocessed image). Afterwards, the cropped local image, after the same procedure of preprocessing, goes through the local branch and output the local feature. Finally, the global feature and the local feature are concatenated and fed into the fully-connected layer for classification.

For the generation of heatmap, we tried to use the maximum values for the heat map, which is defined as $H_g(x, y) = max_k(|f_g^k(x, y)|)$, where $f_g^k(x, y)$ represent the activation of spatial

location (x, y) in the kth channel of the heat maps[6]. However, the performance is awful. Thus, we change the method to calculate the class activation map[4] of the image for heat map, which is defined as

$$M_c = \sum_k w_k^c A^k$$

where c is the given class , w_k^c is the weight corresponding to class c and $A^1, A^2, \ldots, A^N$ are the feature maps.

For the generation of local area, we also tried to resize the heat map and generate the local area on the original image, which is high resolution, but it took a lot of time for computation thus hard for training, so we abandoned it.

We use ResNet-50 as backbone for global branch and local branch. Figure 2 shows the Architecture of FuseNet.

1.2.2 Training

We use the ResNet-50 model trained in section 1 as the initialization for global branch and local branch. Notice that the global branch needs no training.

We first fix the global branch and train the local branch only. After convergence, we fix the local branch and train the final fully-connected layer. Since the problem of under-fitting raises when training the final fully-connected layer, we reduced some data augmentation (we deleted random crop and random rotation,and remained center crop and horizontal flip) in preprocessing.

1.3 Results

Table 1 shows the performance of FuseNet, including the accuracy of each type of the global branch, the local branch and the final prediction of the FuseNet as well as the AUC of the ROC curve of them. The local model of ResNet-50 performs slightly worse, perhaps it is because the cropping process leads to much information loss. The FuseNet turns out to be higher on accuracy and AUC than the global branch, indicating that the idea of combine global and local features works under the circumstance. We did not spend much time on training this model since the training process is too time-consuming, but we believe it will provide better performance after fine-tuning.

Type	ResNet50(b16, global)	ResNet50(b64, local)	FuseNet (b16)
ELBOW	0.8688	0.8438	0.8688
FINGER	0.8400	0.8229	0.8171
FOREARM	0.8394	0.8321	0.8467
HAND	0.7784	0.7605	0.8383
HUMERUS	0.8897	0.8603	0.9118
SHOULDER	0.8000	0.7949	0.7897
WRIST	0.8950	0.8529	0.8908
ALL	0.8452	0.8237	0.8510
AUC	0.9049	0.8848	0.9076

Table 1: The performance of FuseNet

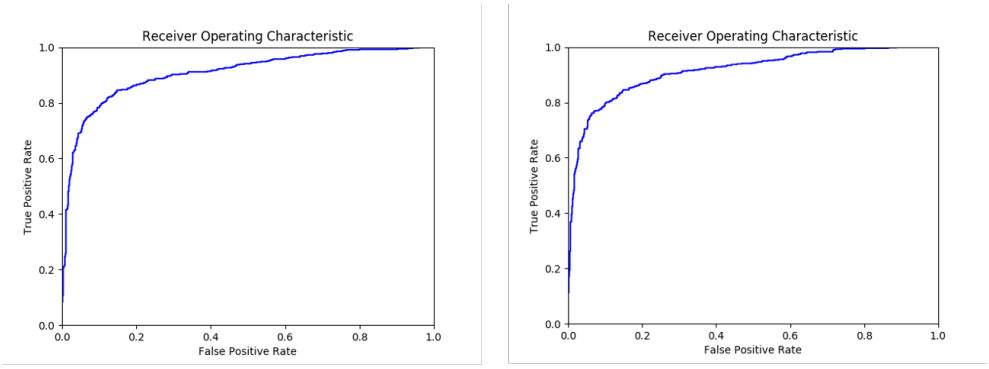
We ensemble five models(ResNet-50 (batch size 16), DenseNet-169 (batch size 16), DenseNet-169 (batch size 32), DenseNet-169 (batch size 64) and FuseNet) as our final model. Each models' result and the ensembled prediction's performances are shown in Table 2. The FuseNet turns

out to be the best single model. The numbers in red imply the highest accuracy among the five models. As can be seen in the table, different types' best classifiers are different models, and the ensembled model has the best performance on total accuracy and AUC.

Type	R50(b16)	D169(b16)	D169(b32)	D169(b64)	F(b16)	Ensembled
ELBOW	0.8688	0.8750	0.8875	0.8688	0.8688	0.8688
FINGER	0.8400	0.8514	0.8343	0.8400	0.8171	0.8514
FOREARM	0.8394	0.8540	0.8321	0.8321	0.8467	0.8467
HAND	0.7784	0.8084	0.8144	0.7964	0.8383	0.8024
HUMERUS	0.8897	0.8897	0.8971	0.9118	0.9118	0.9191
SHOULDER	0.8000	0.7846	0.7846	0.8256	0.7897	0.8000
WRIST	0.8950	0.8950	0.8824	0.8824	0.8908	0.8950
ALL	0.8452	0.8510	0.8469	0.8510	0.8510	0.8543
AUC	0.9049	0.9062	0.9058	0.9060	0.9076	0.9153

Table 2: The accuracy of each type and AUC of each model

Figure 3 shows the ROC of the single best model - FuseNet and the ensembled model.



FuseNet Ensembled

Figure 3: The Receiver Operating Characteristic of FuseNet and the ensembled model

2 Task Two: Abnormality Localization

2.1 Configuration

We applied Grad-CAM[5] to locate the abnormality on radiographs. We use model ResNet-50 with batch size of 16 in this task. We first compute the gradient of the output probability, S, with respect to feature maps A^k of a convolutional layer, i.e. $\frac{\partial S}{\partial A^k}$. These gradients flowing back are global-average-pooled to obtain the neuron importance weights w_k:

$$w_k = \frac{1}{Z}\sum_i\sum_j \frac{\partial S}{\partial A^k}$$

By performing a weighted combination of forward activation maps, following it a ReLU, we can to obtain the grad-CAM:

$$M = ReLU(\sum_k w_k A^k)$$

We used the grad-CAM as the heat map, it is a kind of localization. After printing it with the original image, the red region means high probabilities of abnormality and the blue region means low probabilites.

Then we tried to give more expilict results. We normalize the heatmap to $[0, 1]$ and then set a threshold 0.7 to turn it to a binary map, that is to say,

$$B(i, j) = \mathbf{I}(M(i, j) > 0.7)$$

Then we find the max connected component on the binary map B, acquire the boundary coordinates x_{min}, x_{max}, y_{min} and y_{max} and mark the rectangle according to the coordinates. The rectangle implies the region which has the highest probability of abnormality.

2.2 Results

Figure 4 shows some examples of the heat map and the localization window on the input. The input is random selected from each type in the valid dataset. The result is to our satisfactory.

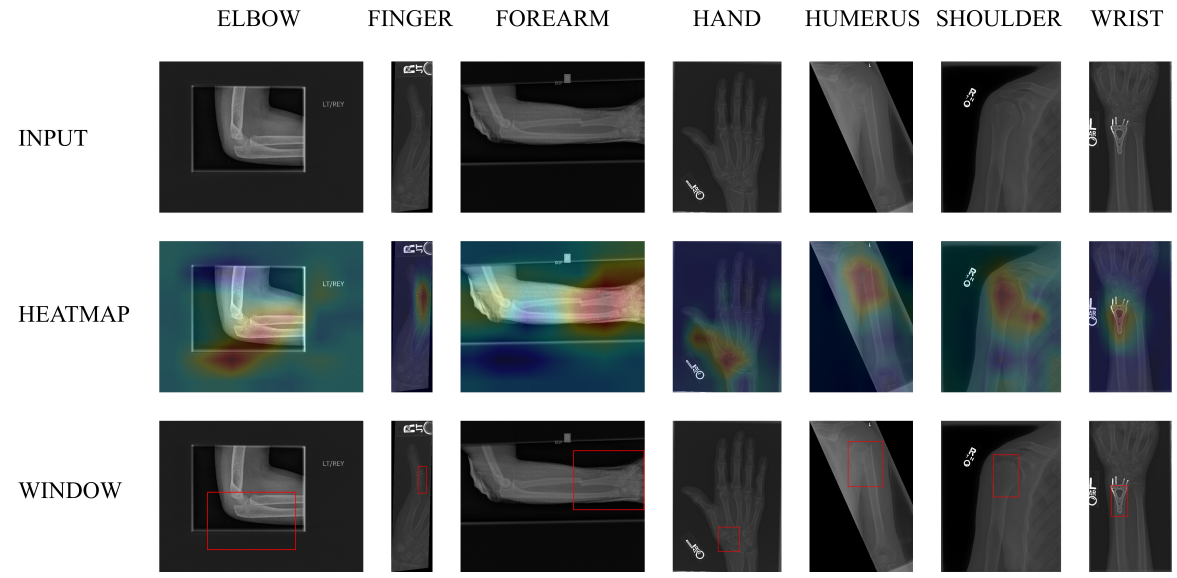


Figure 4: Examples of abnormality localization

3 Task Three: Image Retrieval

3.1 Configuration

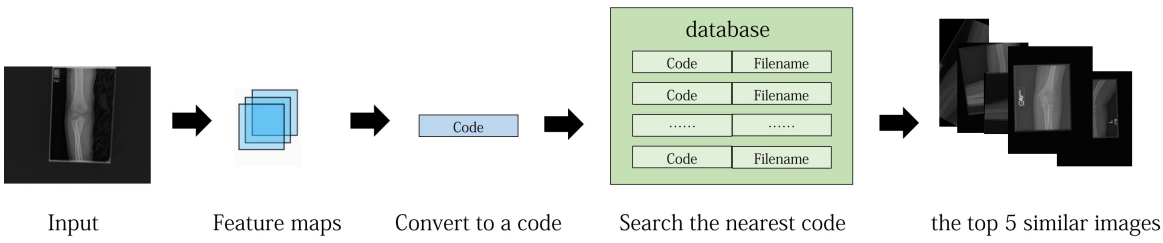


Figure 5: The pipeline of Image Retrieval

In this task, we represented each image in the training dataset by a code and established a database of the codes and their file names. The code, of the form $[c_1, c_2, \ldots, c_{2048}]$, is the feature extracted by the model. c_k is the mean value of the kth channel feature map. We use model ResNet-50 with batch size of 16 in this task.

Given an input image, it is firstly convert to a code. Then we use the k-nearest neighbor method to find codes with minimal distance in the database and output the top 5 similar images. The distance is defined by 2-norm.

Figure 5 illustrates the pipeline of this process. Notice that this pipeline may be time-consuming when the database is large, even though it performances well in our experiments.

Sometimes, the features from different types of data are similar, but finding them gives rare help to analysis the current patient. Therefore we implemented image retrieval both on the whole dataset and on the specific type dataset.

3.2 Results

Figure 6 shows examples of image retrieval on the whole dataset.

The input is random selected from each type in the valid dataset. In the most cases, most of the top-5 similar are the images of the same type as input, but sometimes not.

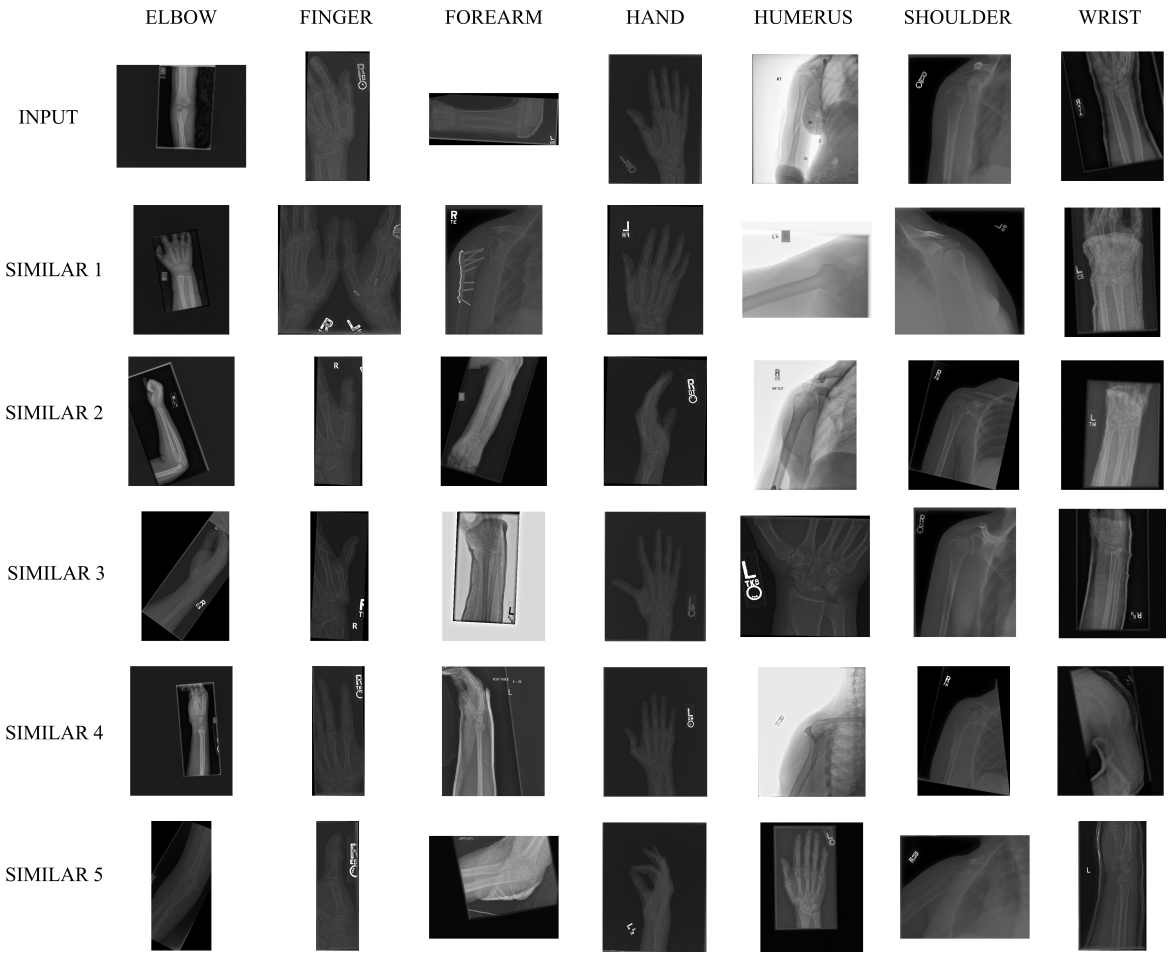


Figure 6: The examples of image retrieval from the whole data set

Figure 7 shows examples of image retrieval on the specific type dataset. The input is same with the above example. All of the outputs are from the same type of images as the input data,

which may provide more useful information.

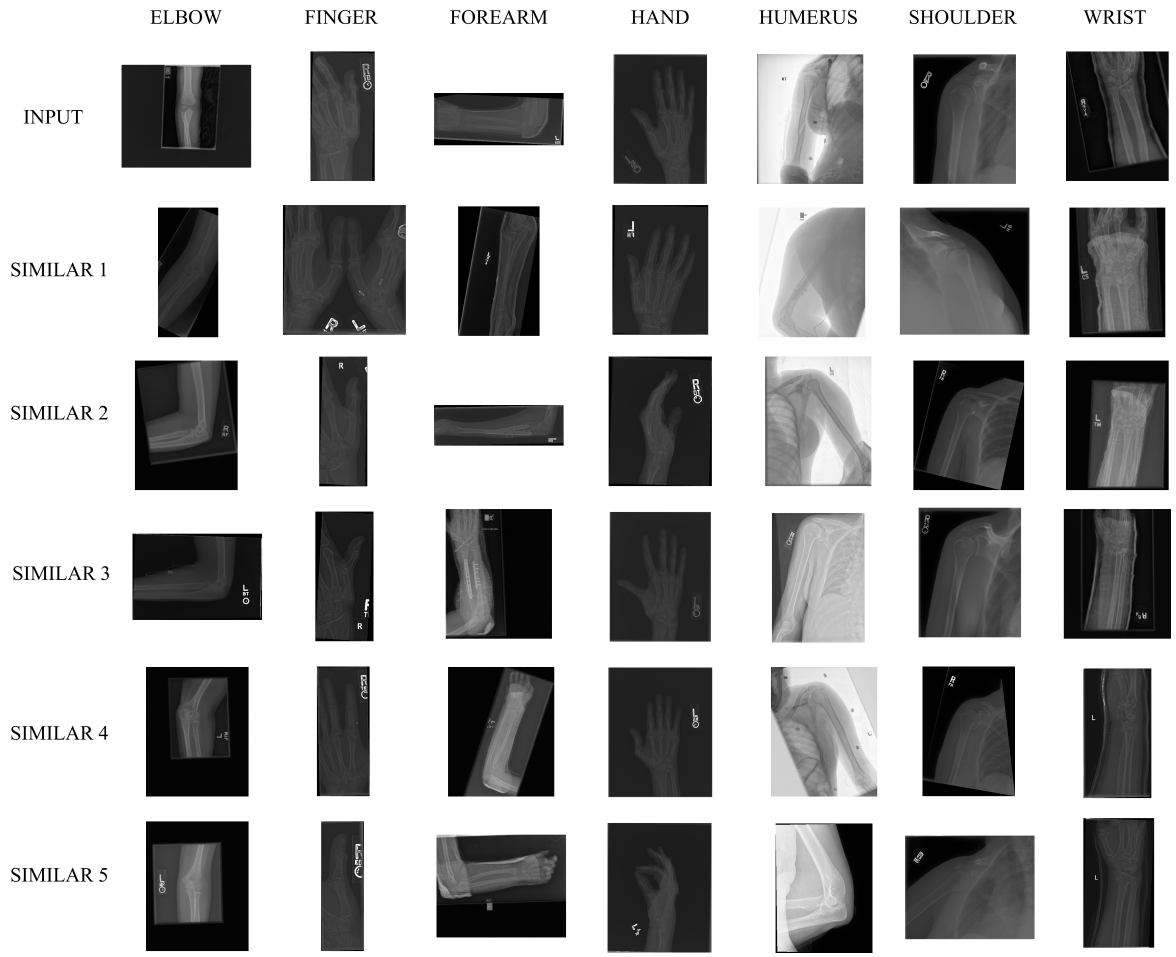


Figure 7: The examples of image retrieval from the specific type dataset

4 Conclusion and Discussion

In this project, we implemented some models based on ResNet and DenseNet for abnormality detection and gained the experience of fine-tuning their parameters. We designed FuseNet to combine the global features and the local features and it turns out to be the single best model. We ensembled the prediction of the models to make the final prediction of abnormality. The ensembled model makes the best performances of accuracy and AUC on validation dataset. We applied Grad-CAM to implement abnormality localization on radiographs. The output can be in the form of heatmap, which tells the probabilities of abnormality over the whole image, or window, which marks which region is most likely to be the abnormality. We implemented image retrieval by finding the nearest neighbor of the feature code in the databse. The 5-top similar images can be from whole or specific data set.

There are several questions left to be explore: There are some confusing data in the dataset, for example, nearly all the images with nails in it are labelled abnormal but there is no relation between nails and abnormality in the sense of pathology. There do exist normal radiographs with nails in it. How to deal with such problem? Our image retrieval model is time-consuming when the database is large and it is not accurate sometimes. How to improve it? ResNet and

DenseNet are networks firstly used for nature image classification, which is quite different from radiograph. Is there any architectures more suitable for radiographs? We are looking forward to explore these problems later.

5 Acknowledgement

We would like to thank Professor Liwei Wang, teaching assistant Xiaoyu Chen, teaching assistant Tiange Luo and the classmates in the machine learning course for their helpful guidance and suggestions during this project.

6 Attachments

README.md - notice and instructions of running codes
common.py - configurations of the project
dataset.py - the operations on data
model.py - the architecture of the models
train.py - the file to train the models
train_fuse.py - the file to train FuseNet
predict.py - the file to valid or test the models
locate.py - the file for abnormality localization queries
retrieval.py - the file for image retrieval queries
utils.py - other functions
models - a directory for the trained models
– densenet169_b16.pth.tar
– densenet169_b32.pth.tar
– densenet169_b64.pth.tar
– resnet50_b16.pth.tar
– fusenet_b16.pth.tar
database - a directory for the database for image retrieval
– database.hdf5
results - a directory for the results of the project

References

[1] Pranav Rajpurkar, Jeremy Irvin, Aarti Bagul, Daisy Ding, Tony Duan, Hershel Mehta, Brandon Yang, Kaylie Zhu, Dillon Laird, Robyn L. Ball, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Ng *MURA Dataset: Towards Radiologist-Level Abnormality Detection in Musculoskeletal Radiographs*

[2] Gao Huang, Zhuang Liu, Laurens van der Maaten *Densely Connected Convolutional Networks*

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun *Deep Residual Learning for Image Recognition*

[4] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba *Learning Deep Features for Discriminative Localization*

[5] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*

[6] Qingji Guan, Yaping Huang, Zhun Zhong, Zhedong Zheng, Liang Zheng and Yi Yang *Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification*

[7] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*

[8] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Robyn L. Ball, Curtis Langlotz, Katie Shpanskaya *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*