Towards to Reasonable Decision Basis in Automatic Bone X-Ray Image Classification: A Weakly-Supervised Approach

Jianjie Lu,[1] Kai-yu Tong[2]

The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, China, 999077
[1] jacklu@link.cuhk.edu.hk, [2] kytong@cuhk.edu.hk

Abstract

A weakly-supervised framework is proposed that cannot only make class inference but also provides reasonable decision basis in bone X-ray images. We implement it in three stages progressively: (1) design a classification network and use positive class activation map (PCAM) for attention location; (2) generate masks from attention maps and lead the model to make classification prediction from the activation areas; (3) label lesions in very few images and guide the model to learn simultaneously. We test the proposed method on a bone X-ray dataset. Results show that it achieves significant improvements in lesion location.

Introduction

Deep learning, particularly convolutional network (CNN), has achieved high accuracies in classification of medical images (Litjens et al. 2017). However, it is still risky to leave diagnosis to machines entirely. Patients and clinicians would prefer to have reliable decision basis from classifiers rather than simple classification results, i.e., normal or abnormal. This is challenging in some cases since it is expensive to obtain lesion labels from experts for classification datasets with tens of thousands of images.

The unsupervised or weakly-supervised approach can be a better choice in this situation. Class activation map (CAM) is one of the widely-used interpretation methods for CNN-based medical image classification models (Feng et al. 2017; Hicks et al. 2018). It can produce heatmaps for localizing the most indicative areas of the input, which is first proposed by (Zhou et al. 2016).

In this work, we propose a 3-stage weakly-supervised framework for providing more reasonable decision basis instead of using CAM directly. Experimental results show that our method achieves highly competitive improvements in lesion location.

Dataset

We evaluate our method on MURA, a large dataset of musculoskeletal radiographs containing 40561 images, where each study is manually labeled as normal or abnormal and no lesion label is provided (Rajpurkar et al. 2018).

Methodology

The framework is designed in three stages progressively. In each stage, we first train a 23-layer ResNet with around 98% accuracy in the validation set. Note that our goal is not to achieve state-of-the-art classification accuracy but rather to output more reasonable decision basis.

Stage 1: Positive Class Activation Map The classification network essentially maps an image to class scores. For MURA datasets, our model outputs two values. The weights of fully-connected (FC) layers will learn to assign high scores to the correct class. Before flowing into the classification layers, features are 2-dimensional and hence we can visualize them to see which part of inputs contribute to the classification results. We use h_i to denote the i-th feature map and $w_{c,i}$ to denote the weight in the FC layer for h_i leading to class c. There are totally 128 filters in the last convolution layer. The CAM for class c is given by:

$$H_c = \sum_{i=1}^{128} w_{c,i} h_i \tag{1}$$

where h_i is a 20×20 matrix and $w_{c,i}$ is a scalar. We observe that when weight $w_{c,i}$ is negative, the corresponding h_i has a negative effect on the class scores. Therefore, we only visualize the features maps that have positive weights. We name this method as positive class activation map (PCAM), which has the formulation as (2). It provides clearer lesion location in our work.

$$H'_c = \sum_{i=1}^{128} \max(0, w_{c,i}) h_i \tag{2}$$

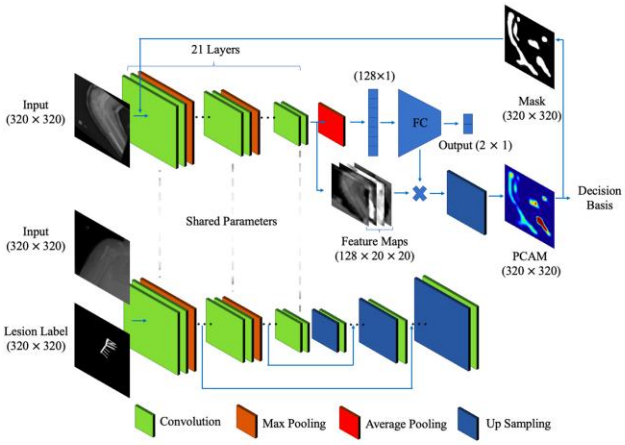

Figure 1. General overview of the proposed framework.

Stage 2: Learnable Attention Map Inspired by the self-guided attention in (Li et al. 2018), we generate masks and guide the network to classify from areas obtained by PCAM as much as possible. Fine-tuning the pre-trained model in the first stage directly can only generate shrinking attention maps. If we retrain the network, the classification result is not accurate in the first epochs, which causes inaccurate and unstable location early on. To address this dilemma, we introduce the confidence parameter $\lambda = e^{P} / e$, which is controlled by classification accuracy P. We use L_1 to denote the cross-entropy loss in the first stage and L_2 to denote that in the second stage. The loss function has the following formulation:

$$L = (1-\lambda)L_1 + \lambda L_2 \tag{3}$$

Stage 3: Weakly-supervised Learning So far, the decision basis is obtained in an unsupervised way. Although lesion probably lies in these attention maps, we cannot guarantee that the learning process of classification network is consistent with clinicians' decisions. For combining clinicians' priors, we marked the lesion location of 100 images, a small proportion (100/40561) of the dataset. After that, we train an encoder-decoder fully convolutional network, where the encoder has a same typology as the classification network and shares parameters. Figure 1 shows an overview of the proposed framework. We use L_3 to denote the MSE loss between the output of FCN and clinicians' labels. In this stage, we optimize the following multi-part objective function:

$$L = (1-\lambda)L_1 + \lambda L_2 + \lambda' L_3 \tag{4}$$

where λ' determines the importance of clinicians' priors. The larger value contributes to a closer result to lesion labels. We set it at 1 in the experiment.

Experimental Results

We present qualitative example results in Figure 2. As can be seen, attention maps generated by CAM are slightly homogeneous in spite of containing some lesions. In comparison, PCAM can provide more clear attention maps than CAM. By learning attention maps, the lesion location is highlighted but not ideal as expected. In stage 3, we achieve significant improvement in lesion location comparing with other approaches.

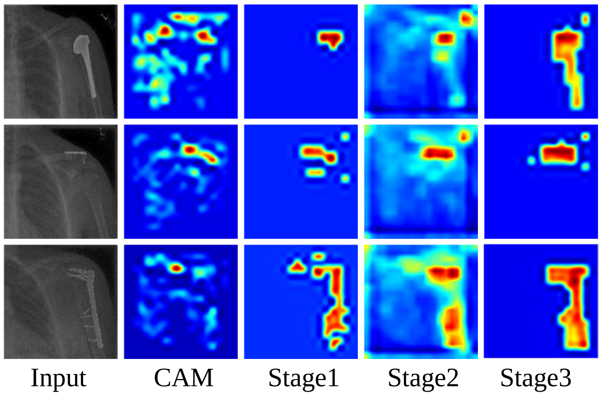

Figure 2. Qualitative example results of attention maps generated by different methods.

Conclusion

In conclusion, we propose a weakly-supervised approach to obtain reasonable decision basis in classification models. Experiments demonstrate that our framework achieves better lesion location. In future, we would like to extend this work in more kinds of medical images and apply quantitative comparison metrics for evaluation.

References

Feng, X.; Yang, J.; Laine, A. F.; and Angelini, E. D. 2017. Discriminative Localization in CNNs for Weakly-supervised Segmentation of Pulmonary Nodules. In *MICCAI*, 568-576.

Hicks, S. A.; Eskeland, S.; Lux, M.; and et al. 2018. Mimir: An Automatic Reporting and Reasoning System for Deep Learning Based Analysis in the Medical Domain. In *Proceedings of the 9th ACM Multimedia Systems Conference*, 369-374.

Li, K.; Wu, Z.; Peng, K. C.; Ernst, J.; and Fu, Y. 2018. Tell Me Where to Look: Guided Attention Inference Network. In *CVPR*, 9215-9223.

Litjens, G.; Kooi, T.; Bejnordi, B. E.; and et al. 2017. A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis* 42: 60-88.

Rajpurkar, P.; Irvin, J.; Bagul, A.; and et al. 2018. MURA Dataset: Towards Radiologist-level Abnormality Detection in Musculoskeletal Radiographs. In *International Conference on Medical Imaging with Deep Learning*.

Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; and Torralba, A. 2016. Learning Deep Features for Discriminative Localization. In *CVPR*, 2921-2929.