Anomaly detection for medical images based on a one-class classification

Qi Wei, Yinhao Ren, Rui Hou, Bibo Shi, Joseph Lo, et al.

Qi Wei, Yinhao Ren, Rui Hou, Bibo Shi, Joseph Y. Lo, Lawrence Carin, "Anomaly detection for medical images based on a one-class classification," Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis, 105751M (27 February 2018); doi: 10.1117/12.2293408

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

Anomaly detection for medical images based on a one-class classification

Qi Wei[a], Yinhao Ren[b], Rui Hou[b], Bibo Shi[b], Joseph Y. Lo[a,b], and Lawrence Carin[a]

[a]Department of Electrical and Computer Engineering, Duke University, Durham, USA
[b]Department of Radiology, Duke University, Durham, USA

ABSTRACT

Detecting an anomaly such as a malignant tumor or a nodule from medical images including mammogram, CT or PET images is still an ongoing research problem drawing a lot of attention with applications in medical diagnosis. A conventional way to address this is to learn a discriminative model using training datasets of negative and positive samples. The learned model can be used to classify a testing sample into a positive or negative class. However, in medical applications, the high unbalance between negative and positive samples poses a difficulty for learning algorithms, as they will be biased towards the majority group, i.e., the negative one. To address this imbalanced data issue as well as leverage the huge amount of negative samples, i.e., normal medical images, we propose to learn an unsupervised model to characterize the negative class. To make the learned model more flexible and extendable for medical images of different scales, we have designed an autoencoder based on a deep neural network to characterize the negative patches decomposed from large medical images. A testing image is decomposed into patches and then fed into the learned autoencoder to reconstruct these patches themselves. The reconstruction error of one patch is used to classify this patch into a binary class, i.e., a positive or a negative one, leading to a one-class classifier. The positive patches highlight the suspicious areas containing anomalies in a large medical image. The proposed method has been tested on InBreast dataset and achieves an AUC of 0.84.

The main contribution of our work can be summarized as follows. 1) The proposed one-class learning requires only data from one class, i.e., the negative data; 2) The patch-based learning makes the proposed method scalable to images of different sizes and helps avoid the large scale problem for medical images; 3) The training of the proposed deep convolutional neural network (DCNN) based auto-encoder is fast and stable.

Keywords: anomaly detection, medical images, one-class classification, unsupervised learning, autoencoder, convolutional neural network

1. DESCRIPTION OF PURPOSE

The target is to detect anomaly in medical images using patch based one-class classification. The classification is achieved by thresholding the reconstruction error derived from an autoencoder which characterizes negative patches. The autoencoder is constructed using a deep neural network which is trained using only negative patches decomposed from normal medical images.

2. INTRODUCTION AND MOTIVATION

In machine learning, conventional classification tries to classify instances into one of the multiple classes, i.e., more than two classes. However, in medical imaging, the imbalanced data, i.e., there are many more negative data than positive one, is always a big issue.[1] For mammography screening, the low prevalence of cancer means that 99.5% of cases are actually cancer free. The imbalanced learning problem has drawn a significant amount of attention from target detection, outlier detection, anomaly detection, novelty detection etc. One straightforward issue with the imbalanced learning problem is the ability of imbalanced data to significantly degrade the performance of most standard learning based classification algorithms. Most conventional algorithms developed on popular benchmark datasets, such as MNIST, CIFAR-10, ImageNet, assume or expect balanced class distributions or

(send correspondence to Qi Wei)
Qi Wei: E-mail: qi.wei@duke.edu, Telephone: 1 919 433 7379

Medical Imaging 2018: Computer-Aided Diagnosis, edited by Nicholas Petrick, Kensaku Mori, Proc. of SPIE Vol. 10575, 105751M · © 2018 SPIE · CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2293408

equal mis-classification costs. This nontrivial assumption makes these algorithms fail to properly learn the distributive characteristics of the complex imbalanced data sets and thus provide unfavorable results. Another problem is that in medical imaging, the positive data is visually very similar to the negative data, requiring experienced doctors which have been trained intensively over years. Besides, medical images usually have large size and high dimensionality, which makes the image wise classification quite challenging. For example, a typical mammogram image has 2000×4000 pixels, which makes the direct analysis or processing of it awkward.

To solve this imbalanced data problem and simultaneously leverage the huge amount data from one class, i.e., the negative one, we propose to use the one-class classification, also known as unary classification and first introduced by Moya & Hush (1996),[2] to try to identify anomaly amongst all images, by learning from a training set containing only negative data. We hypothesize that a model can learn the diversity of normal anatomy from negative images only, such that any anomaly not represented in the training (including cancer) would stand out as being abnormal and can thus be classified as suspicious. In this study, we propose to develop a one-class classifier for mammography, focusing initially on the task of detection of suspicious microcalcifications. To overcome the large scale problem of medical images, we propose to decompose one medical image into patches of much smaller size, i.e., 128×128 in this work.

How to identify a patch as a normal (negative) one or an abnormal (positive) one is task dependent. For example, the decomposed patches can be identified as an abnormal one if it contains a or part of a targeted object, such as a tumor or a nodule. More specifically, for the mammogram images, which will be addressed in this work, the widely used BI-RADS (Breast Imaging-Reporting and Data System) scheme, which is originally designed for use with mammography, can be exploited to name the negative and positive patches. For this study, we will regard BI-RADS 1 and BI-RADS 2 as being negative and BI-RADS 3 to BI-RADS 6 as being positive.

3. METHODS

To exploit the advantage of huge amount of negative images for one-class classification, we propose to learn an autoencoder to characterize patches extracted from these negative images.[3] An autoencoder is a data compression algorithm used for unsupervised feature learning of efficient coding, comprising an encoder and a decoder. In general, encoding a set of data is for the purpose of dimensionality reduction. More recently, with the development of the variational autoencoder (VAE), the autoencoder concept has become extended for learning generative models of data. The two parts, i.e., the encoder and the decoder, of an autoencoder can be defined as two mappings as

$$\begin{aligned} \phi &: \mathcal{X} \to \mathcal{Z} \\ \psi &: \mathcal{Z} \to \mathcal{X} \end{aligned} \tag{1}$$

where $\mathcal{X}$ represents the data space, $\mathcal{Z}$ represents the hidden space, ϕ is an encoder and ψ is an decoder. The loss function to be minimized is as follows.

$$\phi, \psi = \arg\min_{\phi,\psi} \sum_{i=1}^{N} \|X_i - (\psi \circ \phi) X_i\|^2 \tag{2}$$

A stochastic gradient descent (SGD) based back propagation can be used to update all the learnable parameters in ϕ and ψ. Based on the so-called universal approximation theorem and considering the recent success of deep neural networks in image analysis, ϕ and ψ are approximated by deep convolutional neural networks.

4. NETWORK STRUCTURE

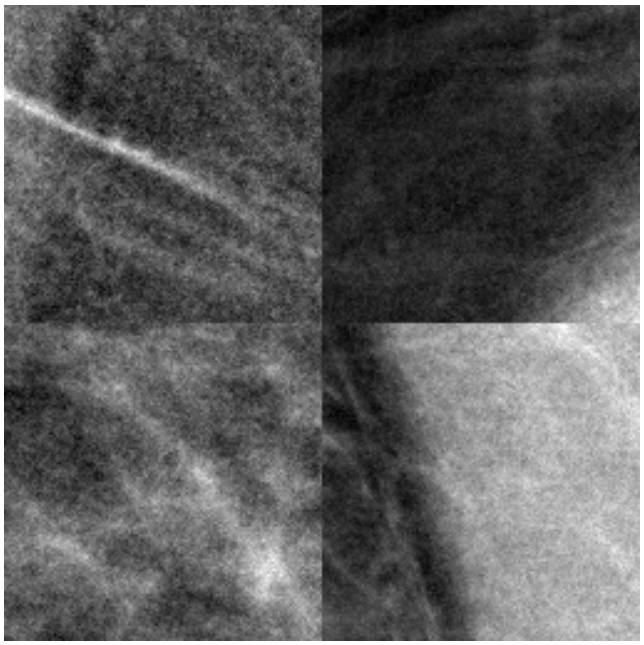
A convolutional neural network consisting of strided convolution neural network and RELU nonlinear transformation has been designed to realize the autoencoder. The autoencoder has eleven layers and the dimensionality in each layer is summarized in the following equation. Note that for each layer, the size of a hidden data point

is width × height × number of feature maps.

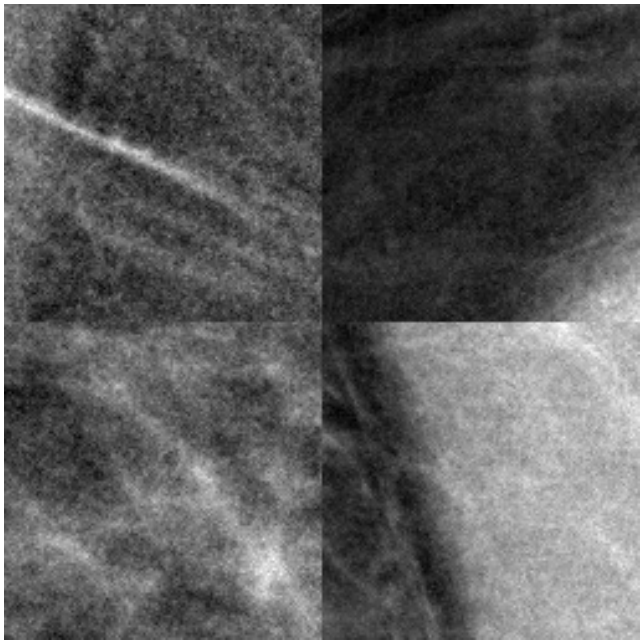
$$\begin{aligned}&\underbrace{128 \times 128 \text{ (input)} \rightarrow 64 \times 64 \times 2 \rightarrow 32 \times 32 \times 4 \rightarrow 16 \times 16 \times 8 \rightarrow 8 \times 8 \times 16 \rightarrow 4 \times 4 \times 32}_{\text{encoder}} \\ &\underbrace{\rightarrow 8 \times 8 \times 16 \rightarrow 16 \times 16 \times 8 \rightarrow 32 \times 32 \times 4 \rightarrow 64 \times 64 \times 2 \rightarrow 128 \times 128 \text{ (reconstruction)}}_{\text{decoder}}\end{aligned} \tag{3}$$

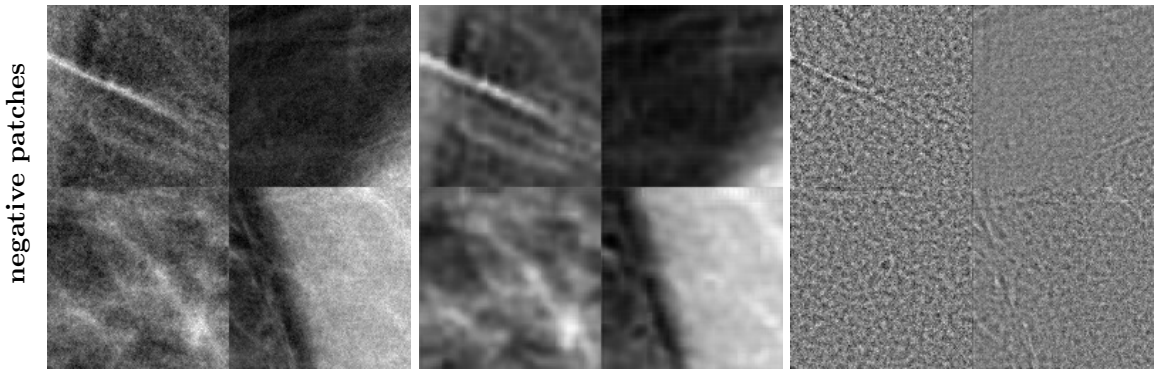
The dimensionality is decreasing as the layer goes deeper. This prevents the autoencoder from learning the identity function to improve its ability to capture important information and learn richer representations.

4.1 Model Architecture for MNIST

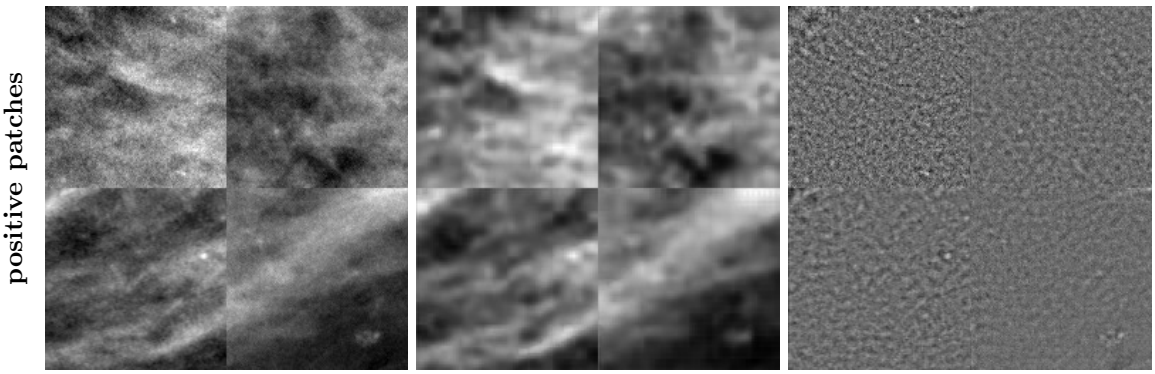


4.2 Model Architecture for CIFAR





(b) Input (zoomed) (c) Reconstruction (zoomed) (d) Error map (zoomed)

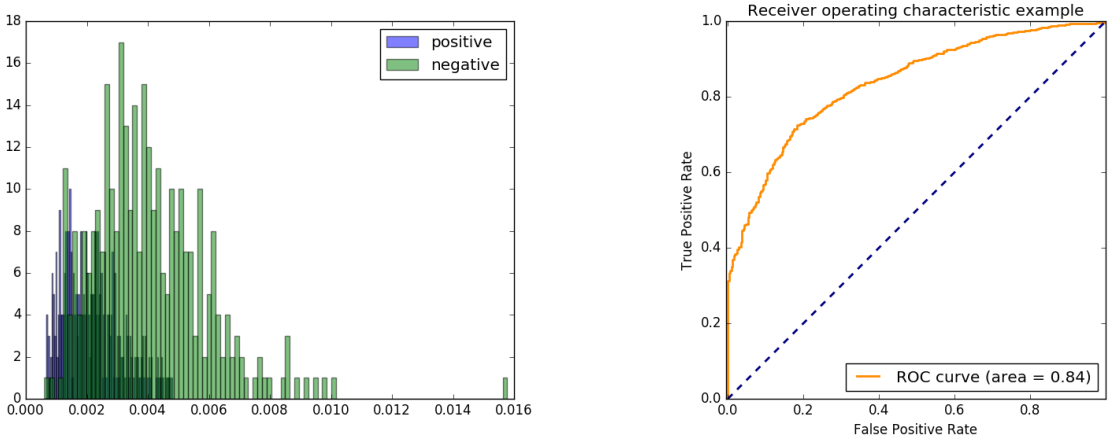


(f) Input (zoomed) (g) Reconstruction (zoomed) (h) Error map (zoomed)

Figure 1. Patches, their reconstructions and error maps (zoomed).

5. RESULTS

In this experiment, we test the proposed method in the widely used InBreast dataset.[4] The mixture of patches extracted from 67 BI-RARDS 1 images are used as negative data and the patches containing ROIs (region of interest) extracted from 49 BI-RADS 5 images are used as positive data. In total, we have 36013 negative patches from BI-RARDS 1 and 321 positive patches from BI-RADS 5. The proposed autoencoder has been trained on 90% of negative patches using stochastic optimization algorithm *Adam*[5] with 500 epochs, a batch size of 200 examples, a learning rate of 0.001, an exponential decay rate for the 1st moment estimates of 0.9 and an exponential decay rate for the 1st moment estimates of 0.999. After fixing the trained autoencoder, testing was performed by using it to reconstruct the negative (on a 10% held-out test set) and positive (not used in training) patches. Error maps portray the difference between the original patch image and its resulting, reconstructed version. During testing, a perfect autoencoder would reconstruct the image with no error, which in this study means that the image belongs to the same class as the purely negative training class. Conversely, presence of visible features in the error map suggests that the testing image may contain abnormalities. Fig. 3 shows some of the testing patches and the corresponding reconstructions, while Fig. 1 shows zoomed-in examples of those patches, reconstructions, and the error maps. Note that the error maps for negative images tend to show no important features, except for some high-frequency edges corresponding to Coopers ligaments that were not captured by the autoencoder. In contrast, however, the error maps for positive images depict suspicious microcalcifications, which as intended stand out because they were not represented in the purely negative training images. The reconstruction error can be quantified by compute the root mean square error (RMSE) over all pixels in one patch. The histogram of reconstruction error has been plotted in the left of Fig. 2 and the ROC curve by classifying reconstruction errors is plotted in the right of Fig. 2. Note that using the autoencoder model learned from only negative patches, we have got an AUC value of 0.84.



(a) Histogram of reconstruction errors (green is positive and blue is negative)

(b) ROC curves by thresholding autoencoder reconstruction errors for classifying negative and positive patches.

Figure 2. Analysis of reconstruction errors

6. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

To the best knowledge of the authors, this is the first work to try to detect anomaly in medical images using only negative data. This is also the first time that an autoencoder based on a deep convolutional neural network is used to achieve the one-class classification. In other words, the unsupervised learning framework is successfully applied to classify the decomposed patches from a testing image, leading to a patch-based anomaly detection.

7. CONCLUSIONS

This work proposed a method to detect an anomaly in medical images based on one-class classification. The proposed strategy requires only the negative patches to learn an autoencoder characterized by a deep convolutional neural network. The learned autoencoder is used to reconstruct given patches to compute their reconstruction errors. Thresholding these derived reconstruction errors can help separate the negative and positive patches. All the processing has been made in the patch level instead of the image level to avoid the large scale issue in medical images as well as to detect the position of an anomaly. This initial demonstration focused on microcalcification detection, but in ongoing work we will investigate more challenging tasks such as diagnosis of microcalcifications, and extend this work to analyze image-level performance.

REFERENCES

[1] Krawczyk, B., "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence* **5**(4), 221–232 (2016).

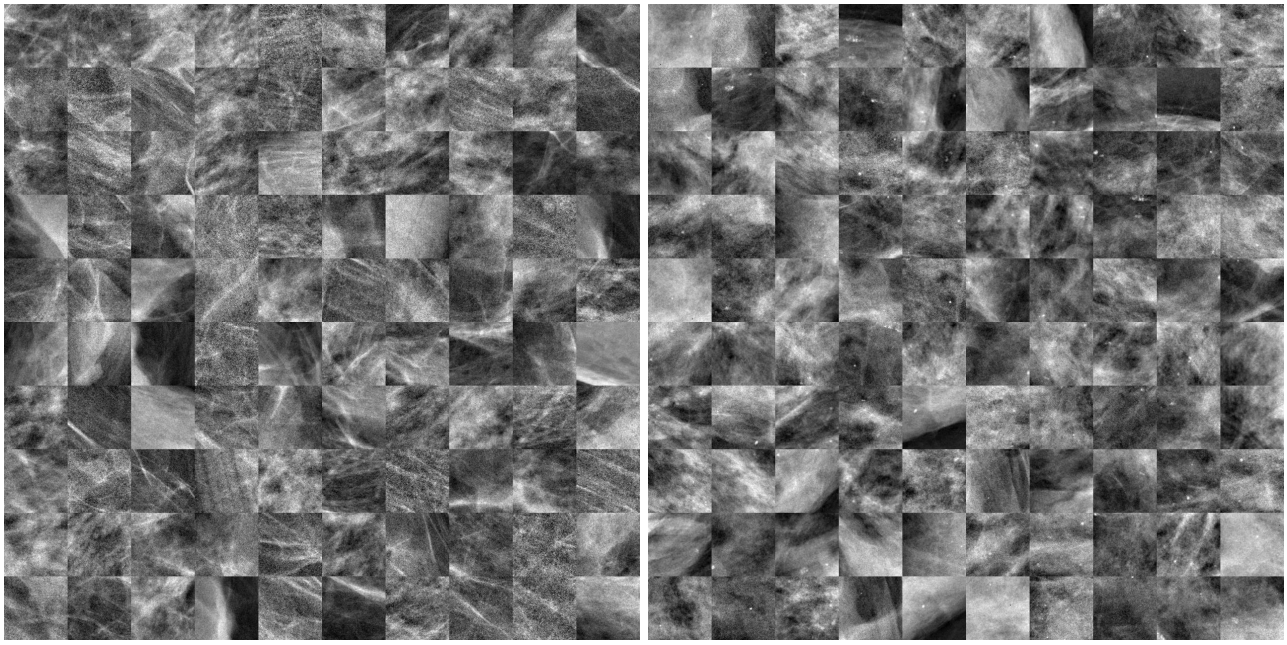
[2] Moya, M. M. and Hush, D. R., "Network constraints and multi-objective optimization for one-class classification," *Neural Networks* **9**(3), 463–474 (1996).

[3] Bengio, Y. et al., "Learning deep architectures for ai," *Foundations and trends® in Machine Learning* **2**(1), 1–127 (2009).

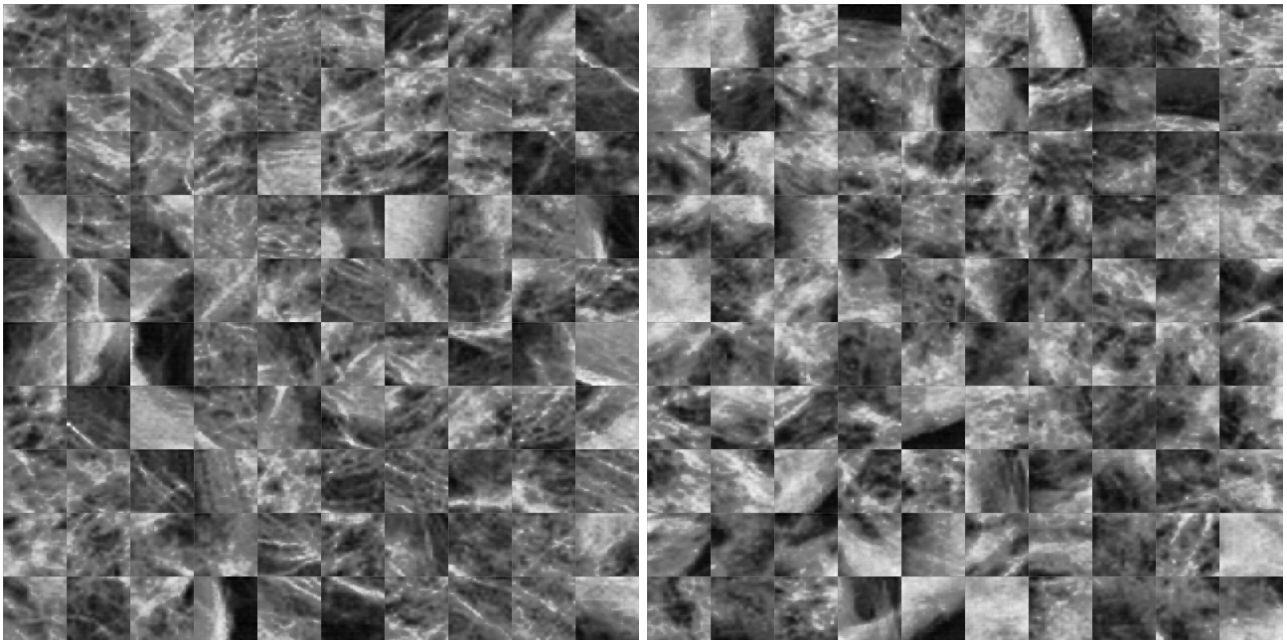
[4] Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., and Cardoso, J. S., "Inbreast: toward a full-field digital mammographic database," *Academic radiology* **19**(2), 236–248 (2012).

[5] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).

negative patches **positive patches**



(c) Input (d) Input



(e) Reconstruction (f) Reconstruction

Figure 3. Patches and their reconstructions