BME 548 Final Project: Towards Improving Classification of Suspected Abnormalities in Digital Breast Tomosynthesis Using Physical Layers

Hanxue Gu, Andy Yang, Tony Guo
Department of Biomedical Engineering
Duke University
Durham, NC 27708
jy168@duke.edu

Abstract

Breast cancer is one of the most common forms of cancer in the U.S. Developing an effective deep-learning-based diagnostic model to classify cases of breast cancer would greatly expedite the current screening process. Here, we develop a LeNet-based CNN assisted by an optimized illumination field and an optical filter to classify suspected cancerous tissue in low-dose digital breast tomosynthesis (DBT) images. Real-life low-dose X-ray images are simulated by adding noises into our dataset. The radius of the frequency domain aperture function and the values inside the aperture are used as trainable weights in the neural network; a similar process was performed for the spatial domain illumination pattern. The results indicate that the highest testing accuracy is achieved with noisy images in the absence of any of our proposed physical layers. Such observation demonstrates that adding noise in the training dataset makes the model more generalizable. While our proposed physical layers have limited benefit on the classification accuracy, they significantly improve the readability of images by human eyes and protect the model from overfitting. Our optimized frequency domain layer also reconfirm that low frequency information is much more critical than high frequency ones in classification, and it may serve as a template for optical filter design.

1 Introduction

Breast cancer is the second most common type of cancer in the United States. Screening for breast cancer cases has been a focus of public health effort in the past over 50 years, and has been shown to be able to decrease the mortality rate of breast cancer by over 20% [1]. Digital breast tomosynthesis (DBT) is a novel modality of breast imaging that helps overcome the high false positive/negative rate of conventional mammogram [4]. In addition, this 3D imaging technique also enables medical professionals to more accurately locate the site of tumor for possible surgical procedures [2].

Computer-assisted detection and diagnosis (CAD) software has been utilized in mammogram-based breast cancer diagnosis for several decades. With the recent developments in deep learning tools, deep-learning-based CAD systems for mammograms have been shown to match radiologists in the accuracy of diagnosing breast cancer [5]. However, due to DBT being a relatively young technology, CAD algorithms that cater to DBT are few. To add to the matter, DBT images obtained from different vendors vary significantly more compared to conventional mammograms obtained from different vendors, making an accurate, widely applicable CAD system difficult to develop [2].

In this project, we explore the possibility of using physical layers to improve the accuracy of a DBT-based breast cancer nodule classifying network. Our system is designed to work in conjunction with a nodule detection/segmentation system, and is capable of classifying the suspected abnormalities into

normal tissue and cancerous tissue. To simulate images obtained with low radiation dose, we added a combination of Gaussian noise and "salt-and-pepper" noise to mimic real-life X-ray images acquired under low radiation dose. The physical layers we explore are an optical filter and a illumination field. We envision these physical layers to be utilized in a visualizing machine analogous to a typical X-ray reader. Suspected nodules may be cropped out from the original DBT image and run through our system to verify or double check the malignancy of the nodule in question.

2 Related Work

Deep learning network used for mammogram screening has been constantly evolving over the past two decades. Recent publications have achieved promisingly high accuracy in classification, and some groups have proposed algorithms that can be adapted across different types of breast imaging data. In 2017, an algorithm designed by Shen et al. using an end-to-end training approach reached an area under the curve (AUC) of 0.91 when using digitized film mammograms [6]. The algorithm can also be transferred to full-field digital mammogram (FFDM) dataset, reaching high accuracy with a subset of the data. The authors, however, did not mention transferrability to DBT images. In 2018, Chougrad et al. utilized transfer learning to propose a deep learning framework that achieves 97% accuracy in classification [7]. In early 2021, Lotter et al. designed a classification algorithm that can be generalized to DBT images, and allegedly outperforms human radiologists consistently [8]. There is, however, few work related to designing physical layer as a means to improve mammogram/DBT classification.

3 Data and Method

3.1 Data Acquisition and Preprocessing

We acquired the original dataset from the Cancer Imaging Archive (`https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=64685580`). This dataset of DBT images is gathered from Duke University Hospital and Duke University. The dataset contains 22,032 reconstructed volumes that belonged to 5,610 studies from 5,060 patients. It is randomly split into training, validation, and test sets in a way that ensures no overlap of patients between the subsets. The test set included 460 studies from 418 patients. For the validation set we selected 312 studies from 280 patients and the remaining 4,838 studies from 4,362 patients were in the training set.

Digital tomosynthesis data are three dimensional, which means that each set of study consists of around 60 slices of 2D mammograms. The annotation of this dataset contains information about which slices indicate abnormalities. Hence, for those studies that contain abnormalities, we use the annotated slice as the representation. And for those studies that do not contain abnormalities, we select a random slice in the middle as the representation. In this way, we created a 2D dataset that includes similar features as the original 3D dataset. An example of this dataset is shown in Figure 1.

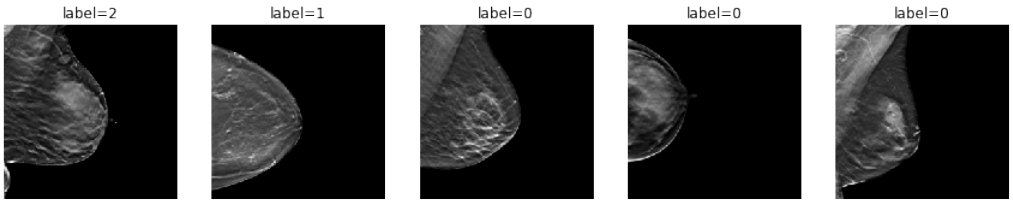

Figure 1: 2D dataset examples

In early exploration of this dataset, we noticed two major problems. The first one is that this dataset is highly imbalanced. Around 90 percent of the data does not contain any abnormality. This imbalance might cause the neural network to largely predict zero (normal) and still get an accuracy of over 90. To solve this problem, we decided to over sample the positive cases ten times to balance the

classes. We only did this to the training dataset and the validation dataset for better performance, and left the testing dataset unchanged for final performance evaluation. The result is shown in Figure 2.

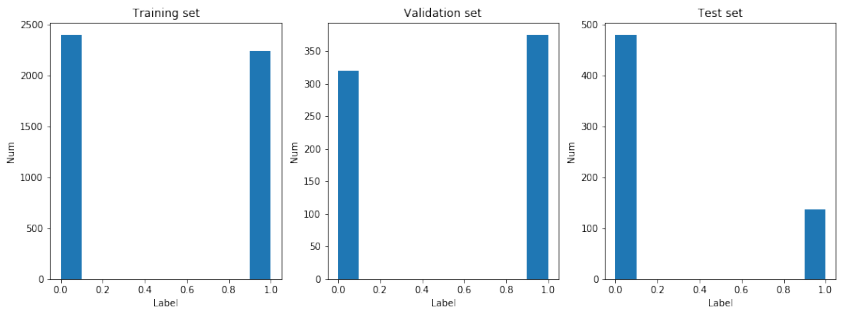

Figure 2: re-balanced dataset

The other problem is that the abnormalities are significantly smaller than the overall image. Our neural networks are therefore having a hard time to do any useful classification in our first round of tests. To solve this problem, we decided to crop out the abnormalities as a 224*224 square, and use the cropped image as the input to the network. For images that do not have any abnormality, we will randomly crop a 224*224 square on the breast. In this way, we successfully generated our preprocessed baseline dataset (Figure 3). It is also because of this step we decided to design this network and physical layers to work in conjunction with a separate detection/segmentation system, since our network by itself cannot classify full images.

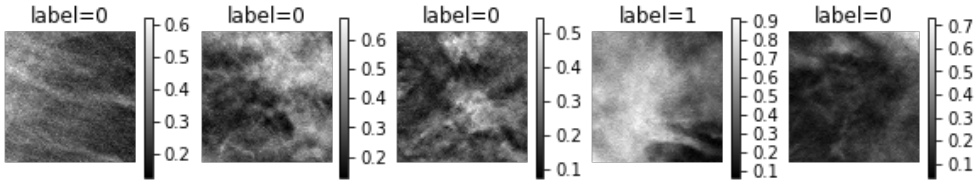

Figure 3: preprocessed baseline dataset

To create the modified dataset, we try to simulate the noise in low-exposure rate X-ray images. The noise pattern will be a 'salt and pepper' pattern. We simulate this effect by adding both 'salt and pepper' noise and Gaussian noise into the dataset, resulting in the modified dataset (Figure 4).

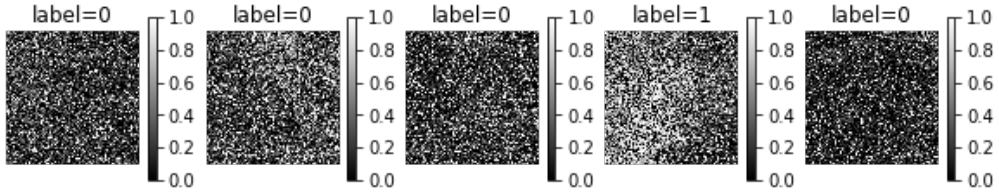

Figure 4: modified dataset

3.2 Physical Layer

The two physical layers that we set out to optimize are an optical filter and an illumination field. As mentioned in the introduction, these layers are designed to work post image acquisition in a X-ray-reader-like device. The parameters of interest are incorporated in the central neural network as trainable weights.

3.2.1 Frequency Domain

To find the optimal optical filter design, we optimize a circular aperture function in the frequency domain. A circular aperture is chosen because of its role as a low-pass filter, and would be the most effective in filtering out the high frequency salt-and-pepper noise while preserving the fundamental information in the original image.

Aperture radius. We first optimize the aperture radius by setting all values within the aperture to 1 and everything outside to 0. It means that the information higher than a certain frequency will be filtered out completely, and only the information within this radius can pass. The radius of the aperture is set as a single trainable variable at this point in the convolutional network.

Filter pattern. After setting the aperture radius equal to this optimal value, we optimize the pattern within the aperture as values between 0 and 1. We performed this step because we hypothesized that not all frequency information has the same weight in the final decision making. This aperture layer is added as an optimizable layer in the network.

3.2.2 Spatial Domain

Based on our hypothesis that nodules in breast-cancer-indicated DBT images frequently appear in some relatively fixed areas, we speculated that focusing illumination in certain areas might "highlight" the high-risk regions, thus improving the accuracy of our classification.

Illumination pattern. We make an element-wise multiplication to mimic the interaction between the illumination field and the image. Taking a reductionist approach, we set the illumination field as a trainable layer in the network, and only consider absorption at this point. The trained result would indicate what an ideal illumination pattern looks like.

Illumination radius. After optimizing the illumination pattern, we find that it shows a centralized feature. This is an expected outcome as we consider our hypothetical detection algorithm to be able to place the nodule in question near the center. Thus, we adopt a circular light source at this point, and optimize the illumination radius.

Illumination pattern under set radius. We implemented a uniform light source is when optimizing illumination radius. After we obtain a reliable radius, we adapt the light source's intensity as a trainable layer to further optimize the illumination system.

3.3 CNN Architecture

We use LeNet as the base architecture of our deep learning network. Both AlexNet and VGG-16 architecture were initially tested on, but the testing accuracy did not improve over the epochs, and the model severely overfits. Using LeNet, however, we observe clear progression of training and testing accuracy. The average pooling layer also contributes to eliminating the noise in the original picture.

3.4 Experiment/Training

We implement all our training evaluations under Ubuntu 20.04.1 environment, and used Adam optimizer throughout the experiments. We designed the following combinations of experiments (trained models):

- Clean image, simple LeNet architecture
- Modified dataset with noise, simple LeNet architecture
- Modified dataset with noise, LeNet + physical layer (optimizing aperture radius)
- Modified dataset with noise, LeNet + physical layer (fixed aperture size, optimizing aperture filter)
- Modified dataset with noise, LeNet + physical layer (optimizing illumination)
- Modified dataset with noise, LeNet + physical layer (with fixed aperture radius, optimizing illumination)
- Modified dataset with noise, LeNet + physical layer (with fixed illumination, optimizing illumination radius)

4 Results

4.1 Basic results

The table below shows the experiment result from each of the above models. We are selecting the best models based on validation accuracy.

Table 1: result table

Model	Training Accuracy	Validation Accuracy	Testing Accuracy
Clean image	84.28	82.01	77.4
Noisy image	86.19	80.86	84.1
Noisy image, physical layer optimizing aperture radius	76.91	81.87	77.9
Noisy image, physical layer optimizing aperture filter	77.07	81.43	78.1
Noisy image, physical layer optimizing illumination	74.65	78.56	78.6
Noisy image, fixed aperture radius, optimizing illumination	66.63	75.68	78.4
Noisy image, optimizing illumination radius	75.88	78.28	77.9

4.2 Performance of Physical layers

The first physical layer we optimized is the aperture radius in frequency domain. Below are the trained aperture function and the images obtained after implementing this layer.

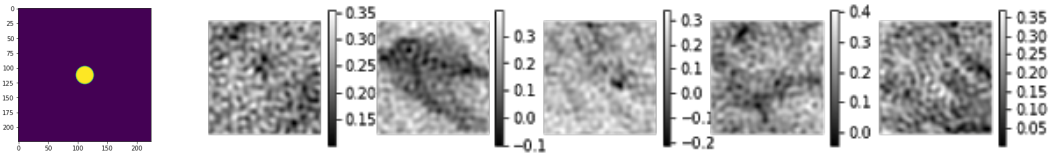

Figure 5: aperture radius (left); filtered images(right)

The second physical layer we optimized is the aperture pattern (passage rates). Below are the trained filter and the resulting images.

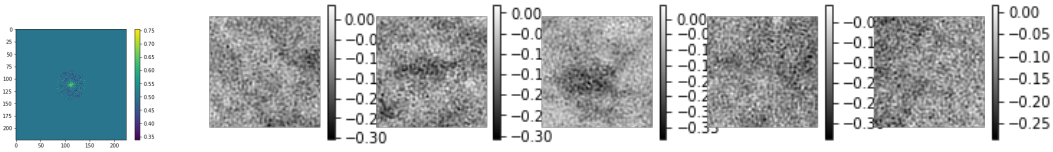

Figure 6: aperture pattern (left); filtered images(right)

The third physical layer is the illumination pattern. Below are the trained illumination field and the resulting images.

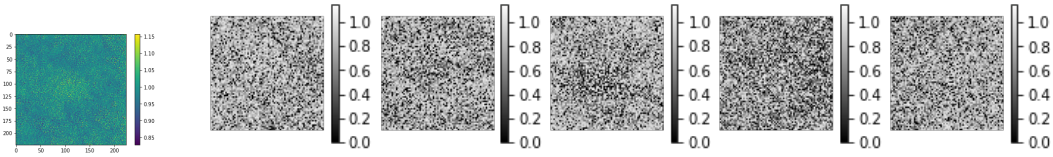

Figure 7: illumination pattern (left); filtered images(right)

The forth physical layer is the illumination pattern under a preset radius. Below are the trained filter and the resulting images.

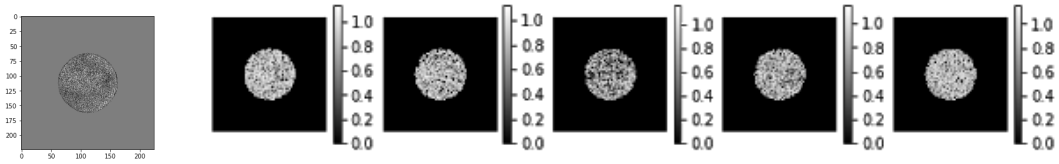

Figure 8: illumination pattern under a certain radius (left); filtered images(right)

The fifth physical layer we optimized is the illumination radius. The optimized radius of illumination is 0.76937926. Below are the images after the physical layer.

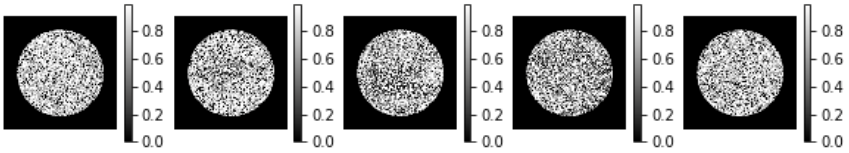

Figure 9: filtered images

5 Discussion

Comparing the classification accuracy for clean images and noisy images, we find that the noisy images model performs better on training accuracy and testing accuracy, though it has slightly lower validation accuracy. It indicates that even the noisy data are unrecognizable to human-beings and almost all the radiologists will meet obstacles on these low quality data, deep learning methods with CNN can get around the noise easily, and achieve a decent classification accuracy.

In contrary to our expectation, adding noise to the data increases the generalization ability without hurting the accuracy. After analyzing the Le-net structure, we realize that the average-pooling layer, which works similarly as a averaging filter, is capable of practically eliminating the negative impact of the noise. Such observation matches our previous understanding, and it may partially explain our network's robust ability to denoise.

While adding the physical layer does not hurt the performance dramatically, it also fails to improve the accuracy as we initially expected. However, through analyzing the loss and accuracy curves, we find that the no-physical-layer version of models is more susceptible to over-fitting despite its higher accuracy. This shows that to some extent, the physical layers we implement serves to protect the network from overfitting.

After implementing the frequency domain optimized physical layers, the images are significantly more recognizable. It meets our assumption that the noisy signal are mainly high frequency information, and that the high frequency domain contains little useful information. It is also shown from our optimized frequency pattern which has concentrated high passage rates at the center, while higher values on the peripheral are scattered.

Lastly, our optimized illumination field demonstrates that the centers of the cropped images are more useful, which is a natural result of having a hypothetical detection system that is capable of placing the suspected abnormality at the center of the image. We believe that optimizing the illumination layer will also allow our system to correct for the detection system's imperfect centering of the nodule in the cropped images.

6 Conclusion and Future Work

In conclusion, our physical layers are highly effective in improving image readability, although they have limited usefulness in improving classification accuracy of our neural network. Despite the physical layers' ability raise accuracy, they are helpful in avoiding overfitting in the training process. This project also confirms that low frequency information is much more important than their high frequency counterpart, and indicates a potential design for an optical filter.

In the future, we hope to further fine-tune or modify the VGG-16 architecture that we initially wanted to use. Despite LeNet's various advantages, it is after all a less powerful model than VGG and has suboptimal classification accuracy. We also hope to figure out a way to let our network take in inputs of variable shapes, in order to fully utilize the three-dimensional feature of the DBT dataset.

References

[1] S. Narod, J. Iqbal, A.B. Miller, Why have breast cancer mortality rates declined?, Journal of Cancer Policy **5**, 8-17 (2015).

[2] K.J. Geras, R.M. Mann, L. Moy. Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Concepts and Future Perspectives. Radiology **293**(2), 246–259 (2019).

[3] M.A. Helvie. Digital mammography imaging: breast tomosynthesis and advanced applications. Radiologic clinics of North America **48**(5), 917-29 (2010).

[4] J. Chubak, et al. Cost of breast-related care in the year following false positive screening mammograms. Medical care **48**(9) 815-20 (2010).

[5] B. Boyer, C. Balleyguier, O. Granat, C. Pharaboz, CAD in questions/answers: Review of the literature, European Journal of Radiology **69**(1), 24-33 (2009).

[6] L. Shen, L.R. Margolies, J.H. Rothstein, et al. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Sci Rep **9**, 12495 (2019).

[7] H. Chougrad, H. Zouaki, O. Alheyane, Deep Convolutional Neural Networks for breast cancer screening, Computer Methods and Programs in Biomedicine **157**, 19-30 (2018).

[8] W. Lotter, A.R. Diab, B. Haslam, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nat Med **27**, 244–249 (2021).