

Deep Learning Based Mass Detection in Mammograms

Zhenjie Cao¹, Zhicheng Yang¹, Yanbo Zhang¹, Ruei-Sung Lin¹,

Shibin Wu², Lingyun Huang², Mei Han¹, Jie Ma^{3*}

¹PingAn Tech, US Research Lab, Palo Alto, CA, USA

²Ping An Technology, Shenzhen, Guangdong, China

³Shenzhen People's Hospital, Shenzhen, Guangdong, China

Abstract—Mammogram is the primary imaging technique for breast cancer screening, the leading type of cancer in women worldwide. While the clinical effectiveness of mammogram has been well demonstrated, the mammographic characteristics of breast masses are quite complex. As a result, radiologists certified for reading mammography are lacking, which limits the accessibility of mammography for more population. In this paper, we propose a Computer Aided Detection (CADe) method to automatically detect masses in mammography. Our method combines Faster R-CNN with Feature Pyramid Network, Focal Loss, Non-Local Neural Network to achieve the optimal mass detection performance. We comprehensively evaluated our method and baseline methods on three public datasets combined, namely the Digital Database for Screening Mammography (DDSM), INbreast, and Breast Cancer Digital repository (BCD). Our results demonstrate that our method outperforms the baselines by a large margin, reporting an Average Precision of 0.805, and a recall of 0.977, respectively.

I. INTRODUCTION

Medical imaging has been a revolutionary way for medical professionals to diagnose and treat medical diseases over the past decades. However, interpretation of medical images is a complex task, which can only be performed by medical professionals who have been extensively trained on reading medical images and have long-time clinical experience. With the recent advance of Artificial Intelligence (AI) with the emphasis of *deep learning* methods, numerous Convolutional Neural Networks (CNN) based approaches for computer-vision-related applications have been proposed [1], [2]. Many researches also apply the CNN-based methods to Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADx) of medical images for effectively assisting doctors (or even automatically) to locate lesions and determine if they are benign, or malignant (commonly known as *cancer*) in medical images [3]. This promising approach is expected to reduce radiologists' workload and to accelerate the diagnosis process while improving the diagnosis accuracy. These intelligent applications are also capable of integrating with the e-health system to generate a comprehensive clinical report, and/or to provide potential personal assistance, etc. [4].

Among these applications on medical images, detecting lesions in *mammography*, the primary imaging technique used for breast screening process, is gaining the increasing

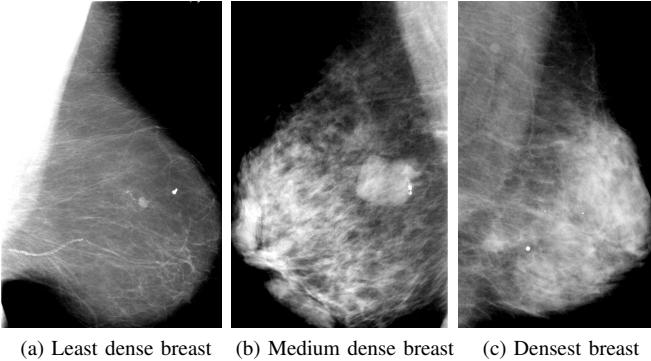
attention during recent years [5], because breast cancer is the leading type of cancer in woman, accounting for massive death worldwide every year [6]. There are mainly four types of abnormality patterns in mammograms: mass, calcification, asymmetry, and architectural distortion. In particular, *mass*, which refers to the hard but painless breast areas with irregular edges, is one of the most commonly concerned detection target in mammograms.

Authors in [7] leveraged conventional machine learning methods in K-means and support vector machine (SVM) classification. Abstract patterns were extracted from malignant and benign tumors, and were then used for the prediction model training. Recently, various deep-learning-based frameworks substantially outperform traditional machine learning methods for object detection tasks. An effective Region proposals with CNNs (R-CNN) model was first introduced in [8]. The enhanced version, Faster R-CNN, improved the overall detection performance and significantly reduced the processing overhead [9]. A recent work applied the Faster R-CNN framework on the large mixed public mammogram datasets to detect breast lesions [5].

In this paper, we present an improved Faster R-CNN model for automatic mass detection in mammography. We adopt three effective neural network techniques, namely Feature Pyramid Network [10], Focal Loss [11] and Non- Local Neural Network [12], and integrate them within the Faster R-CNN framework. We comprehensively evaluated the proposed method and compared with baseline methods on three public datasets combined, namely Digital Database for Screening Mammography (DDSM) [13], INbreast [14] and Breast Cancer Digital repository (BCD) [15]. We evaluate the individual or joint effectiveness of the three modules embedded in the Faster R-CNN model on our dataset. We demonstrate that the three modules employed all results in performance improvement, and the proposed method outperforms the baseline methods by a large margin, reporting an Average Precision of 0.805, and a recall of 0.977, respectively

The rest of this paper is organized as follows. Detailed descriptions of the proposed method are given in Sec. II. We describe the experiment setup and present result in Sec. III. Discussion of findings and future work are presented in Sec. IV. We conclude this paper in Sec. V.

* Corresponding authors: Jie Ma (email: majie688@hotmail.com).



(a) Least dense breast (b) Medium dense breast (c) Densest breast

Fig. 1: Mammograms with different breast densities.

II. METHODS

In this section, we describe the details of Faster R-CNN model and the modules of FPN, focal loss, and non-local neural networks.

A. Faster R-CNN

Faster R-CNN is a two-stage deep-learning-based object detection model. In the Faster R-CNN, a backbone network is adopted to generate feature maps. Based on feature maps, region candidates are first automatically generated by region proposal networks (RPN) instead of the previous slow selective search algorithm in the R-CNN version [8], then a CNN-based network is used to classify the object class and detect the bounding box. Moreover, the whole backbone convolution layers are shared layers which are used for both region proposal networks and classification head. This way, the overhead of the entire framework is significantly reduced, compared to other previous transformed versions of R-CNN.

B. FPN

FPN is a multi-scale algorithm for object detection proposed in [10]. Without introducing additional calculation overhead, FPN exploits the inherent hierarchical architecture in the process of generating the top feature map layer in the backbone network, and efficiently extracts multi-scale feature maps to constitute a pyramid structure for the following prediction tasks.

Since a typical mass does not have distinct outlines, a clear breast background is expected where the mass can be easily identified. However, due to the variation of radiation dose used for mammography, a clear background is not always achieved. As mammogram images have complex structures, identifying mass lesions from a dense background requires specialists to observe more areas and to distinguish the possible lesions from the normal dense breast texture. On the other hand, if a mass lies within a clear background, a smaller and more local receptive field is possibly adequate to locate it. FPN is a practical approach to facilitate this scalable detection, achieved by the fusion of multi-level feature maps with different sizes.

C. Focal Loss

As aforementioned in Sec. II-B, detecting lesions in dense breasts is a challenging task, because the sharpness of background will heavily impact the model's performance. Fig. 1 presents breast examples with three different densities. In this example, Fig. 1c can be regarded as a type of “hard examples” in the lesion detection task, as it has the densest breast among these three examples. This observation enables us to reconsider the importance of “easy examples” (masses with clear backgrounds) and “hard examples” (masses with bright backgrounds) during the training procedure. After the several beginning epochs, the easy examples tend to have less contribution to the loss. Instead, the contribution of “hard examples” (masses with bright backgrounds) to the loss should be increased. Since the hard examples are the ones considered with dense background, it's challenging to identify a mass lying within these fatty tissues with local features. Therefore, considering the global information will be necessary here. Moreover, the original Faster R-CNN usually obtains many false positives for mass detection. We thus replace the original loss function in Faster R-CNN with the state-of-the-art focal loss, which utilizes a weighting strategy and focuses on hard examples. The focal loss is defined as follows [11]:

$$FL(p) = -(1 - p)^\gamma \log(p), \quad (1)$$

where

$$p = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise,} \end{cases} \quad (2)$$

where $y \in \{-1, +1\}$ refers to the ground-truth binary class; $p \in [0, 1]$ denotes the predicted probability of the class with label $y = 1$; γ is a focusing variable, which is not less than 0.

D. Non-Local Neural Networks

The *Non-local Means* (NL-means) is a traditional computer vision algorithm originally used for image denoising [16]. Instead of calculating the mean value of a target pixel's “local” surrounding pixels, this algorithm computes the mean of all “non-local” pixels of the entire image, weighted by the similarity of each pixel to the target pixel. Therefore, the global details that might be ignored by the local mean approach can be maintained. Recently, authors in [12] proposed non-local neural networks that applied the NL-means idea to the modern deep learning architecture and has demonstrated its effective capacity of capturing long-range dependencies, i.e., global information addressed in Sec. II-C. Such dependency is also an important concern in mammograms. The original receptive field in the backbone network of Faster R-CNN model might not consider sufficient global information to calculate the response of a single position.

The key formula of the non-local neural networks [12] is:

$$\mathbf{v}_i = \frac{1}{C(\mathbf{u})} \sum_{\forall j} f(\mathbf{u}_i, \mathbf{u}_j)g(\mathbf{u}_j), \quad (3)$$

where \mathbf{u} refers to an image; i is the index of a target position; j denotes the index of all possible positions through the image \mathbf{u} ;

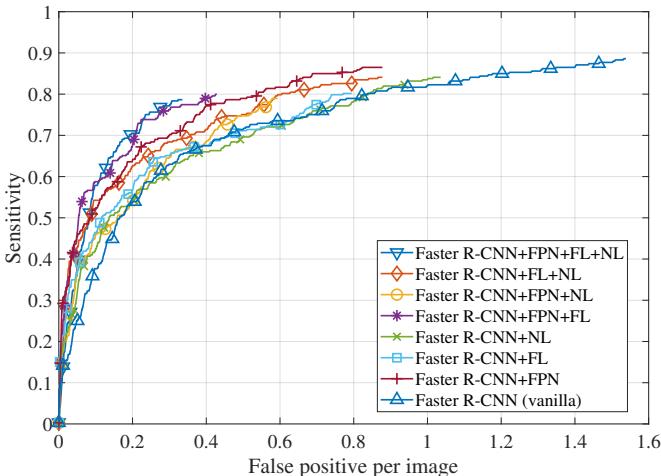


Fig. 2: FROC curves of different models. “FL” refers to focal loss, “NL” is non-local operation, and “vanilla” denotes the original Faster R-CNN model.

f calculates a scalar relationship between i and j ; g denotes an operation at the position j ; \mathbf{v} represents the calculated outcome of this equation, which is normalized by a factor $C(\mathbf{u})$. We here adopt the embedded dot product version as the function f .

$$f(\mathbf{u}_i, \mathbf{u}_j) = \theta(\mathbf{u}_i)^T \phi(\mathbf{u}_j), \quad (4)$$

where $\theta(\mathbf{u}_i) = W_\theta \mathbf{u}_i$; $\phi(\mathbf{u}_j) = W_\phi \mathbf{u}_j$; W_θ and W_ϕ denote the two target weight matrices, respectively. To avoid introducing much extra computation cost, we limit the number of non-local blocks in the late stage of our backbone network.

III. EVALUATION

A. Datasets

Our dataset comprises three public datasets, Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [13], INbreast [14], and BCD [15]. The original DDSM database [17] is a well-known public breast disease database. However, since the mammograms were collected using different equipment by several hospitals in the United States, the image quality was diverse. We adopt an updated version of DDSM, CBIS-DDSM [13], which eliminates the defective images and forms a standardized subset for more convenient use. The elimination process includes image decompression, image cropping and updating for the precision segmentation labels. CBIS-DDSM has $\sim 1,600$ images for mass lesions. INbreast has a total of 115 patient cases, of which 90 cases are from women with both breasts and 25 cases are from mastectomy cases. BCD dataset consists of 535 mass images.

In our experiment, the CBIS-DDSM data are randomly split as training, validation and test data with the percentage of 60%, 20%, and 20%, respectively. The other two public datasets (INbreast and BCD) along with the separated training subset of CBIS-DDSM constitute the training set, while the

TABLE I: Prediction results of different models. “FL” refers to focal loss, “NL” is non-local operation, and “vanilla” denotes the original Faster R-CNN model.

Method	AP	Recall
Faster R-CNN (vanilla)	0.762	0.950
Faster R-CNN+FPN	0.789	0.965
Faster R-CNN+FPN+FL+NL	0.805	0.977

remaining validation and test subsets of CBIS-DDSM are respectively sole validation and test sets.

B. Settings and Parameters

We next describe the settings and the key parameters of the Faster-RCNN model. ResNet-50 [18] is used as the backbone network of our Faster R-CNN, where the hyper-parameters are loaded from the pre-trained model on ImageNet [19]. Each original training image is down-sampled to a small size to ensure that the short edge has 1200 pixels. The Adam optimization method [20] is used in the model training. Initial learning rate is set to 0.01, with 4 epochs for warm up. There are 500 steps in each epoch, and the training process terminates after 200 epochs.

In the Faster R-CNN, the amount of proposals that are sent to the head function has an impact on both the computation resource and calculation workload for the head function. Since there could only be limited number of mass lesions in one mammogram, we set a small number of “candidate” proposals generated from RPN to the classification head. Since the amount of training images is not very large, there is a risk of over-fitting issue. We utilize several data augmentation methods, such as horizontally flip, translation, and scaling, in training to generate more data.

C. Effectiveness of Modules

We first evaluate the effectiveness of the three described modules embedded into the Faster R-CNN model. The commonly used paradigm, Free-response Receiver Operating Characteristic (FROC), is adopted as our metric. Fig. 2 shows that the FROC curves of various combinations of the modules with the vanilla Faster R-CNN model. We observe that (i) the Faster R-CNN model with all of three modules together outperform the others; (ii) the effectiveness of FPN is the most outstanding; (iii) the other two modules, focal loss and non-local operation, indeed improve the model. However, compared to their individual efficacy, the joint improvement is notable.

D. Prediction Results

We present the prediction results of three representative models on our dataset in Table I. Average precision (AP) is a widely used evaluation metric in object detection models. It is calculated as mean precision over several recall levels. When the short edge of an image is resized to 1,200 pixels, the Faster R-CNN model embedded with all three modules

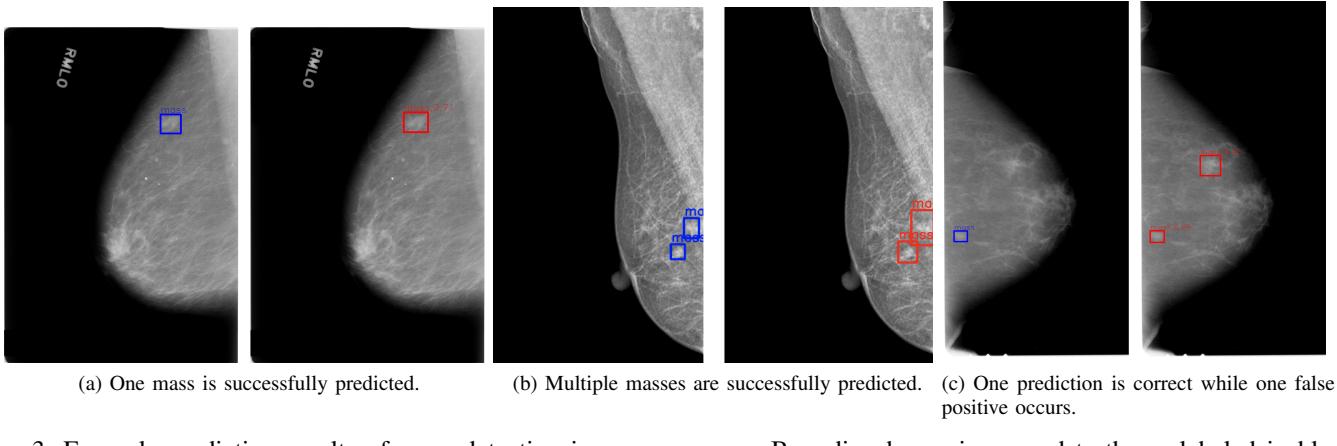


Fig. 3: Example prediction results of mass detection in mammograms. Bounding boxes in ground truth are labeled in blue, and the predicted masses are highlighted in red bounding boxes, respectively.

achieves the best detection results of 0.805 in AP, and 0.977 in recall, respectively. Fig. 3 provides three example prediction results of mass detection in mammograms. As we can see, one or multiple masses are successfully located (Fig. 3a,3b). One false positive shown in Fig. 3c occurs. This might be caused by the sharpness of the wrongly predicted area compared to the background.

IV. DISCUSSION AND FUTURE WORK

Studies have shown that breast density has been a risk of breast cancer in woman is in relation to the higher breast density. Chen et al. reported that when controlling for age and BMI, absolute mammographic density of Asian Americans is significantly lower than African Americans, but not compared with white women. Moreover, ethnic difference on breast density is especially significant on women older than 50 years old [21]. We are now constructing our private dataset by collaborating with Shenzhen People's Hospital, Guangdong, China. Investigating how much the race difference will impact the prediction performance of deep learning models is one of our next research directions.

Asymmetry and architectural distortion are two other major lesion types in breast cancer, and will be focused later when we obtain more labeled data from Shenzhen People's Hospital in the future.

V. CONCLUSION

In this paper, a Faster R-CNN model with a series of effective modules, which are FPN, focal loss, and non-local operation, is demonstrated for effectively detecting masses in mammograms. Our mammography dataset contains three public datasets, DDSM, INbreast, and BCD. The individual and joint effectiveness of modules are evaluated. The FROC curve shows that the Faster R-CNN model embedded with all three modules is the best model. On our dataset, it achieves the best detection results of 0.805 in Average Precision (AP), and 0.977 in recall, respectively.

ACKNOWLEDGMENTS

We would like to thank Dr. Le Lu and Dr. Shun Miao from PAII Inc. for their comments and suggestions.

REFERENCES

- [1] Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L.-J. Li, and L. Fei-Fei, "Thoracic disease identification and localization with limited supervision," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [2] S. Miao, Z. J. Wang, and R. Liao, "A cnn regression approach for real-time 2d/3d registration," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.
- [3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [4] P. Li, Z. Yang, W. Yan, M. He, Q. Yuan, K. Lan, J. Zheng, T. Liu, D. Cao, and Z. Zhang, "Mobicardio: A clinical-grade mobile health system for cardiovascular disease management," in *2019 IEEE International Conference on Healthcare Informatics (ICHI) Workshops (AI4CDM)*. IEEE, 2019.
- [5] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Scientific reports*, vol. 8, no. 1, p. 4165, 2018.
- [6] J. Ferlay, C. Héry, P. Autier, and R. Sankaranarayanan, "Global burden of breast cancer," in *Breast cancer epidemiology*. Springer, 2010, pp. 1–19.
- [7] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of k-means and support vector machine algorithms," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1476–1482, 2014.
- [8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.
- [9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.
- [10] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.
- [11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [12] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.

- [13] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, p. 170177, 2017.
- [14] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.
- [15] M. G. Lopez, N. G. Posada, D. C. Moura, R. R. Pollán, J. M. F. Valiente, C. S. Ortega, M. Solar, G. Diaz-Herrero, I. Ramos, J. Loureiro *et al.*, "Bcdr: a breast cancer digital repository," in *15th International conference on experimental mechanics*, 2012.
- [16] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 60–65.
- [17] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, p. 170177, 2017.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [21] Z. Chen, A. H. Wu, W. J. Gauderman, L. Bernstein, H. Ma, M. C. Pike, and G. Ursin, "Does mammographic density reflect ethnic differences in breast cancer incidence rates?" *American journal of epidemiology*, vol. 159, no. 2, pp. 140–147, 2004.