Breast Mass Detection in Mammograms via Blending Adversarial Learning

Chunze Lin[1], Ruixiang Tang[1] Darryl D. Lin[2], Langechuan Liu[2], Jiwen Lu[1]*,
Yunqiang Chen[2], Dashan Gao[2], and Jie Zhou[1]

[1] Tsinghua University, Beijing, China,
[2] 12 Sigma Technologies, San Diego, USA
{lcz16,trx14}@mails.tsinghua.edu.cn, {lujiwen,jzhou}@tsinghua.edu.cn,
{dlin,pliu,yunqiang,dgao}@12sigma.ai

Abstract. Deep learning approaches have recently been proposed for breast cancer screening in mammograms. However, the performance of such deep models is often severely constrained by the limited size of publicly available mammography datasets and the imbalance of healthy and abnormal images. In this paper, we propose a blending adversarial learning method to address this issue by regularizing the imbalanced data with synthetically generated abnormal samples. Unlike most existing data generation methods that require large-scale training data, our approach is carefully designed for augmenting small datasets. Specifically, we train a generative model to simulate the growth of mass on normal tissue by blending mass patches into healthy breast images. The resulting synthetic images are exploited as complementary abnormal data to make the training of deep learning based mass detector more stable and the resulting model more robust. Experimental results on the commonly used INbreast dataset demonstrate the effectiveness of the proposed method.

Keywords: Mammogram synthesis, mass detection, adversarial deep learning, digital mammography.

1 Introduction

Breast cancer is among the most common cancers affecting women around the world. Mammography has been demonstrated to be an effective imaging modality for early detection and diagnosis, and has contributed to substantial reduction of mortality due to breast cancer. Over the past few years, computer-aided detection of breast masses in mammography has attracted much attention from the medical imaging community [1, 3, 9, 8, 15].

Recently, with the prevalent success of deep learning in natural image applications, there has been keen interest in the medical imaging community to apply these methods to mammogram screening. However, deep convolutional neural networks (CNN) based approaches require a large amount of annotated

---

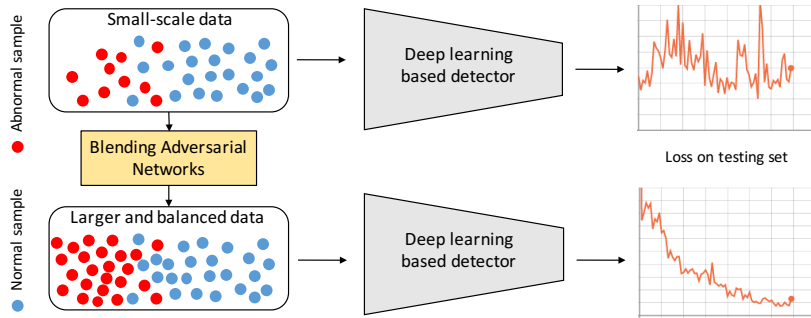* The corresponding author is Jiwen Lu.

Fig. 1. Training a detector with small-scale and imbalanced datasets leads to unsatisfactory results. Our blending adversarial networks help to increase the scale of the datasets and address the class imbalance issues for training more robust detector.

data. The lack of such data has become the main obstacle impeding deep learning methods from achieving impressive performance for breast cancer screening. In contrast to the natural image domain, collecting annotated breast mammograms is very expensive due to the need for expert annotation and oftentimes difficult or even impossible because of privacy restrictions. In addition to the lack of large-scale datasets, the natural class imbalance in mammography samples, where "normal" (or healthy) images significantly outnumber abnormal samples, further limits the performance of deep CNN based methods for breast cancer detection, as illustrated in Fig. 1.

A common way to alleviate these issues involves applying a series of transformations such as flipping, rotation or resizing to augment the training images. However, data augmentation using image transformation is limited in its ability to expand the manifold the positive samples occupy. More recently, generative adversarial networks (GANs) [4] have demonstrated the capability to synthesize realistic images that can be used for data augmentation. For example, Korkinof *et al.* [9] utilized the progressive generative adversarial network to generate high resolution mammograms. Wu *et al.* [15] proposed the conditional infilling GANs to generate lesions on non-malignant patches. One major drawback of these methods is that they rely on a large amount of data to train the generator, making them unsuitable for small-scale datasets.

In this paper, we propose blending adversarial networks to address the limitation of small-scale and imbalanced data for mass detection in mammogram. GANs based methods usually train a generator to synthesize mammograms from Gaussian distributed random values. This demands the generator to learn the texture, the shape and the size of the breast and lesion. Learning to synthesize these features requires inevitably a large-scale training dataset. As opposed to such a heavy task, we simplify the burden of the generator: we provide both the real lesion and "normal" image at the input, and train a model to imagine how this lesion will grow on the normal breast tissue. Since the information about the lesion and breast are given, the generator can focus on integrating the lesion
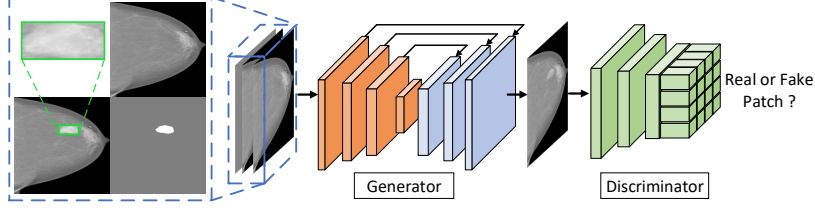
Fig. 2. Overview of the pipeline of the proposed blending adversarial networks. Given a real "normal" breast image and a lesion patch, the generator aims to blend these images at the indicated location. The discriminator verifies the quality of the generated data at patch level forcing the generator to produce highly realistic images.

into healthy breast tissue. By simplifying the task of the generator, we can train it even with a very small dataset. Therefore, we are able to utilize "normal" images to artificially generate abnormal mammograms to increase the data size and alleviate class imbalance. Extensive experiments on widely-used INbreast dataset [11] demonstrate the effectiveness of the proposed method, where the mass detector becomes significantly more robust when trained with the complementary synthetic samples.

2 Methodology

Mathematically, given a set of "normal" images $\mathbf{X} = \{X_1, ..., X_N\}$ and a set of lesion patches $\mathbf{E} = \{E_1, ..., E_M\}$, our goal is to learn a network to seamlessly blend the lesions into the "normal" images to form a new set of images containing lesions $\tilde{\mathbf{Y}} = \{\tilde{Y}_1, ..., \tilde{Y}_L\}$. The proposed blending adversarial networks can be learned from a small-scale dataset to generate new images. We then include them as the complementary training samples to train a deep learning based breast mass detector, making it more robust and effective.

2.1 Blending Adversarial Networks

In Fig. 2, we illustrate the pipeline of our blending adversarial networks, composing mainly of a generator and a patch discriminator. Unlike the vanilla GANs [4] which rely on large-scale datasets for training, we incorporate prior knowledge into the generator input and supervision signals to make the adversarial model robust to small-scale training data.

Generator: Rather than using a Gaussian distributed vector as input like vanilla GANs, we carefully design a three-channel input for our generator as shown in Fig. 2. The first channel corresponds to a "normal" image that provides contextual information about the overall breast. The second channel consists of a lesion patch directly pasted into the "normal" image to provide the texture of the mass. The third channel is a binary mask indicating the location where

we aim to blend the lesion, as a way to inform the model to pay more attention to this region. With these strong clues, we purposefully ease the task of our generative model, such that it can focus on the task of seamless blending alone. The generator is a fully convolutional network (FCN) that takes these three channels as input and produces an image with mass. The generator has an hourglass architecture with an encoder and a decoder. The encoder assimilates and fuses information associated with the normal breast and lesion patch. The decoder then expands the encoded information to generate a realistic image. In order to preserve the prior knowledge, we add the skip connections between the encoder and decoder layers to facilitate the propagation of prior clues given at the input.

Patch Discriminator: The generated images from the generator are then fed into a discriminator whose purpose is to verify the quality of the synthesized mammogram. In most existing GANs based method the discriminator performs an image level classification to distinguish real from fake images. However, in medical imaging, the texture details are crucial and a global classifier could overlook such information. We therefore propose to explore a patch discriminator [7] that can focus on the texture details in local image patches. This discriminator aims to classify whether each $N \times N$ patch in an image is real or fake. If a region looks fake or lacks texture details, it will result in a large value of loss forcing the generator to improve the quality. We run this discriminator convolutionally across the image and average all responses to provide the final output. In our experiments, we empirically set $N = 16$.

2.2   Adversarial Seamless Blending Supervision

Given a lesion patch E, a "normal" breast image X and a random position (x, y), the blending generator $\mathcal{G}$ aims to generate an artificial mammogram image $\tilde{Y}$ that contains the lesion at the position (x, y):

$$\tilde{Y} = \mathcal{G}(X, E, x, y; \theta_g)$$

where θ_g corresponds to the parameters of the network that should be optimized. Since our goal is to incorporate the lesion into a breast image, the network needs to pay attention to two parts: the region Ω where we want to integrate the lesion, and the remainder of the image, $\mathcal{B}$, where we want to keep unaltered (Fig. 3). The network needs to imagine how the lesion should grow according to the characteristics of the breast tissue inside the target region Ω, while keeping the background portion $\mathcal{B}$ identical. To this end, we propose the adversarial seamless blending supervision signals to guide the training process.

$$L = L_{adv} + \lambda_p L_{prior} \tag{1}$$

Adversarial Loss: We aim to generate synthetic images that are indistinguishable from real images, in order that the generated mammograms can be used as training samples. More specifically, we apply the adversarial loss [4] to

supervise the generator and the discriminator in an adversarial manner:

$$L_{adv} = \mathbb{E}_Y[\log D(Y)] + \mathbb{E}_{X,E}[\log(1 - D(G(X,E)))] \tag{2}$$

where the generator G is constrained to produce realistic images to confuse the discriminator D, while the discriminator D should correctly distinguish real images Y from generated ones $G(X,E)$. However, the min-max game between the generator and discriminator is not easy to converge during training, especially with a small-scale dataset. Therefore, we introduce the additional prior knowledge loss $L_{prior} = L_S + L_B$ to help guide the training process. It is composed of the following two loss functions and balanced with the parameter λ_p.

Seamless Blending Loss: As the lesions are often of small size in mammograms and the adversarial loss specifies only a high-level goal for the authenticity of the entire image, the generator may generate mammogram without any lesions. To circumvent this issue, we propose the following loss to make the generator pay attention to the region Ω where we aim to blend the lesion.

$$L_S = \|\nabla G(X,E)_\Omega - \mathbf{v}\|_2 \tag{3}$$

where

$$\text{for all } i,j \in \Omega, \ \mathbf{v}(i,j) = \begin{cases} \nabla X_\Omega(i,j) & \text{if } |\nabla X_\Omega(i,j)| > |\nabla E(i,j)| \\ \nabla E(i,j) & \text{otherwise} \end{cases} \tag{4}$$

This supervision signal makes the generator take into account the intensity variations of both the source lesion E and the normal tissue patch X_Ω for seamless blending. The notation $\nabla(\cdot)$ represents the gradient operation.

Background Loss: In order to preserve the background region $\mathcal{B}$ of the breast as given at input, we constrain the generator to output an image that maintains identical intensity and gradient. We employ both the L1 distance loss and the gradient difference loss to supervise the generator:

$$L_B = \left\|(G(X,E) - X)_\mathcal{B}\right\|_1 + \left\|(|\nabla G(X,E)| - |\nabla X|)_\mathcal{B}\right\|_2 \tag{5}$$

The subscript $\mathcal{B}$ indicates that the supervision operates at the background region of the breast excluding the location where we aim to incorporate the lesion. We use L1 distance rather than L2 as the former encourages less blurring [7]. The gradient loss plays a complementary role to L1 loss and forces the generator to better preserve the texture variations in background regions.

Note that both the adversarial loss L_{adv} and the prior loss L_{prior} should be used together to ensure the quality of synthetic images. Without the adversarial loss, the generated images may lack the texture details and appear fake. Without the prior loss, the generator will not be able to perform the seamless blending.

2.3 Mass Localization

As training a detection model requires a large-amount of data, most existing deep learning based methods for breast cancer diagnosis using mammograms
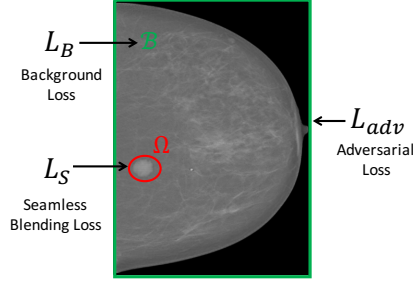
Fig. 3. Different regions on the image where the loss functions are applied. The background loss affects the background regions, the seamless blending loss focuses on the mass, and the adversarial loss controls the quality of the whole image.
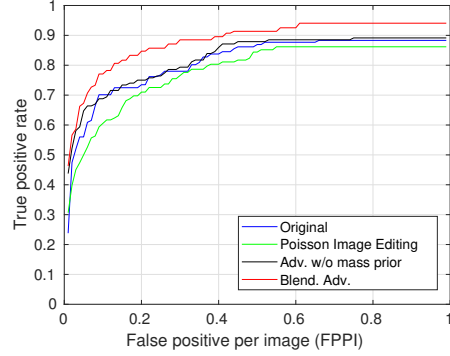

Fig. 4. FROC curves comparing the detection performance of detectors trained with original data and complementary data generated by PIE, adv. w/o prior knowledge of mass and our blend. adv. methods, respectively.

are limited to image-level classification [14, 16, 2]. In this paper, we adjust the state-of-the-art object detection framework, Mask R-CNN [5], for mass detection in mammography and improve the detection performance with our generated data. Note that recently more and more deep detector based methods have been proposed for mass detection, but they require large training datasets. Given an input mammogram, the proposed model aims to detect the lesion with both bounding boxes and segmentation masks. In addition to the classification loss and bounding box regression loss, we supervise the network with the segmentation loss to exploit pixel-wise information. The multi-task loss for each region of interest (RoI) is defined as:

$$L_{det} = L_{cls} + \lambda_{loc}L_{loc} + \lambda_{seg}L_{seg} \tag{6}$$

where L_{cls} is the classification loss, L_{loc} corresponds to the bounding box regression loss, and L_{seg} is the segmentation loss. λ_{loc} and λ_{seg} are weighting factors for different components of the loss function.

3   Experiments and Results

Database: We conducted our experiments on the widely used digital mammography dataset INbreaset [11]. This dataset comprises of a set of 115 cases containing 410 images, where 116 images contain benign or malignant masses. While INbreast is among the highest quality public mammography dataset with accurate annotations, there are only a limited number of images. We computed the results using a 5-fold cross validation experiment by carefully dividing the 115 cases into 80% for training and 20% for testing at the patient-level to avoid any positive bias.

Table 1. Comparison of detection performance with state-of-the-art methods in terms of true positive rate (TPR) versus false positive per image (FPPI) on INbreast dataset.

Method	TPR@FPPI	Run-time
Kozegar *et al.* [10]	0.65@1.1, 0.87@3.67	108s
Dhungel *et al.* [1]	0.87 ± 0.14@0.8, 0.96 ± 0.03@1.2	20s
Dhungel *et al.* [3]	0.90 ± 0.02@1.3, 0.95 ± 0.02@5.0	39s
Jung *et al.* [8]	0.88 ± 0.07@0.5, 0.91 ± 0.07@1.3	1.8s
Ours-Blend. Adv.	0.91 ± 0.07@0.5, 0.94 ± 0.07@0.8	0.5s
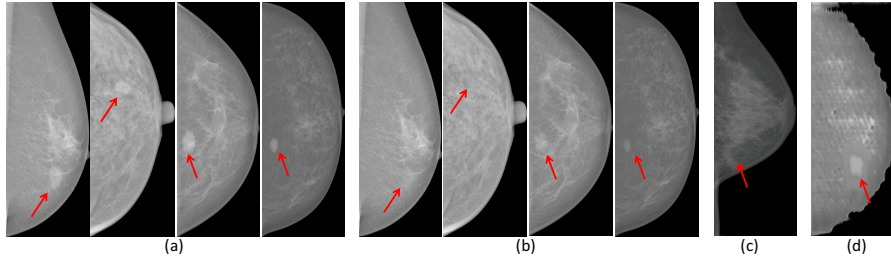
Fig. 5. Visualization of the images generated by (a) our blending adversarial networks, (b) Poisson image editing method, (c) adversarial model without mass prior knowledge and (d) vanilla GANs, respectively. The arrows indicate the position of the masses.

Implementation Details: We adopted U-Net [13] as the backbone for our generator and a series of four convolutional layers for our patch discriminator. For our mass detector, we employed ResNet50 [6] as backbone and initialized the parameters with COCO pretrained model. To facilitate better convergence, the training process consists of three steps: (1) only the top layers are learned for the first 30 epochs, (2) all layers from stage 4 of ResNet are fine-tuned for 30 epochs, and (3) we optimize all layers for 40 epochs.

In order to test the ability of the model to localize lesions, we evaluate the predictions using the Free-Response Operating Characteristic curve (FROC). The FROC curve depicts the true positive rate as a function of the number of false positives per image (FPPI). A mass is considered to be correctly localized if the intersection over union (IoU) ratio between the ground truth bounding box and the predicted bounding box is higher than 0.5.

Results and Analysis: To evaluate how well the generated images can enhance the mass detector performance, we trained a few variants of the detector, using either the original data, or the original data plus one of three types of generated complementary data: 1) using conventional Poisson Image Editing (PIE) blending method [12]; 2) using an adversarial network without prior information of mass; and 3) using our blending adversarial networks. For each case, we generate 200 complementary data, making the training data approximately three times larger than original data set. For generating a mammogram with lesion, we randomly select a region in a "normal" breast mammogram and a real lesion as input for our generator. Note that we randomly resize and rotate the input lesion to augment the data.

The FROC curves in Fig. 4 depict the performance of the model trained with different sets of data. We can see that the detector trained with the additional data generated by our blending adversarial networks performs significantly better than the detector trained only with the original data. We observe an improvement of $\sim 10\%$ on the true positive rate for the same number of false positives per image, clearly demonstrating the effectiveness of the proposed method. Using the data generated by our blending adversarial networks as complementary training data makes the detector more robust due to expanded sample space of the training data. Some examples of the images generated by our blending adversarial networks are shown in Fig. 5(a).

On the other hand, we observe a degradation of performance when the detector is trained with the additional data generated by the conventional Poisson Image Editing method [12]. These results suggest that naively increasing the number of training images may potentially lead to adverse effects. The adversarial learning process guides our generator to approximate the underlying distribution of the authentic data. In contrast, the conventional image processing methods are unable to control the quality of the generated sample. As illustrated in Fig. 5(b), the lesions are often invisible in images synthesized with the PIE approach. This "over-blending" effect tends to mislead the detector. Furthermore, prior knowledge is crucial for learning a data generative model with a small-scale dataset. Fig. 5(c) shows that without the prior knowledge of mass, *i.e.* replacing the real lesion patch with random noise at the input to the generator, the generated images lack fine texture details inside the mass region. Additionally, as shown in Fig. 5(d), the vanilla GANs (generating both breast and lesion from noise input) fail for dataset of this size and is only able to generate a mean shape of the breast with severe visual artifacts.

We compare the performance of our detector with several mass detection methods in the literature [8, 1, 10], and tabulate the results in Table 1. We follow the evaluation metrics as in previous works by giving the true positive rate at some acceptable false positive per image rates (TPR@FPPI). Our detector correctly localizes 91% of masses with a FPPI rate as of 0.50, while the existing mass detection approaches achieve similar true positive rates only with much larger number of false alarms. The run-time efficiency of the detector is also a key criterion for users. Without any cascaded structures or post refinement, our detector can execute at significantly higher speed of $0.5s$ per image.

4   Conclusion

In this paper, we proposed blending adversarial networks to help address the issue of class imbalance and data scarcity in mammography. We made full use of the existing "normal" images to generate breast mammograms with synthetic masses that could be used as positive samples for training deep learning based mass detector. As testament to the effectiveness of the proposed method, extensive experiments on the widely-used INbreast dataset demonstrated significant improvement of the detection performance.

References

1. Dhungel, N., Carneiro, G., Bradley, A.P.: Automated mass detection in mammograms using cascaded deep learning and random forests. In: DICTA (2015)
2. Dhungel, N., Carneiro, G., Bradley, A.P.: The automated learning of deep features for breast mass classification from mammograms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 106–114 (2016)
3. Dhungel, N., Carneiro, G., Bradley, A.P.: A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Medical image analysis (2017)
4. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NIPS (2014)
5. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: ICCV (2017)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5967–5976. IEEE (2017)
8. Jung, H., Kim, B., Lee, I., Yoo, M., Lee, J., Ham, S., Woo, O., Kang, J.: Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. PloS one (2018)
9. Korkinof, D., Rijken, T., O'Neill, M., Yearsley, J., Harvey, H., Glocker, B.: High-resolution mammogram synthesis using progressive generative adversarial networks. arXiv preprint arXiv:1807.03401 (2018)
10. Kozegar, E., Soryani, M., Minaei, B., Domingues, I., et al.: Assessment of a novel mass detection algorithm in mammograms. Journal of cancer research and therapeutics (2013)
11. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: toward a full-field digital mammographic database. Academic radiology (2012)
12. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. ACM TOG (2003)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
14. Shen, L.: End-to-end training for whole image breast cancer diagnosis using an all convolutional design. NIPS workshop (2017)
15. Wu, E., Wu, K., Cox, D., Lotter, W.: Conditional infilling GANs for data augmentation in mammogram classification. In: Image Analysis for Moving Organ, Breast, and Thoracic Images (2018)
16. Zhu, W., Lou, Q., Vang, Y.S., Xie, X.: Deep multi-instance networks with sparse label assignment for whole mammogram classification. In: MICCAI (2017)