Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks

Conference Paper · July 2018

DOI: 10.1117/12.2318100

2 authors, including:

Shuyue Guan
George Washington University
32 PUBLICATIONS **91** CITATIONS

SEE PROFILE

Breast Cancer Detection Using Synthetic Mammograms from Generative Adversarial Networks in Convolutional Neural Networks

Shuyue Guan, Murray Loew
Medical Imaging and Image Analysis Laboratory, Department of Biomedical Engineering, George Washington University, 800 22nd Street NW, Washington DC, 20052, USA

ABSTRACT

The Convolutional Neural Network (CNN) is a promising technique to detect breast cancer based on mammograms. Training the CNN from scratch, however, requires a large amount of labeled data. Such a requirement usually is infeasible for some kinds of medical image data such as mammographic tumor images. Because improvement of the performance of a CNN classifier requires more training data, the creation of new training images -- image augmentation – could be one solution to this problem. In this study, we applied the Generative Adversarial Network (GAN) to generate synthetic mammographic images from the Digital Database for Screening Mammography (DDSM). From the DDSM, we cropped two sets of regions of interest (ROIs) from the images: normal and abnormal (cancer/tumor) Those ROIs were used to train the GAN, and the GAN then generated synthetic images. To compare the GAN with the affine transformation augmentation methods, such as rotation, shifting, scaling, etc., we used six groups of ROIs (three simple groups: affine augmented, GAN synthetic, real (original), and three mixture groups of each pair of the three simple groups) for each to train a CNN classifier from scratch. And, we used real ROIs that were not used in training to validate classification outcomes. Our results show that, to classify the normal ROIs and abnormal (tumor) ROIs from DDSM, adding GAN-generated ROIs to the training data can reduce overfitting of the classifier. But the affine transformations performed slightly better than GAN. Therefore, GAN could be an optional augmentation approach. The images augmented by GAN or affine transformation cannot substitute entirely for real images to train CNN classifiers because the absence of real images in the training set will cause serious over-fitting with more training.

Keywords: breast mass classification, deep learning, convolutional neural networks, generative adversarial networks, image augmentation, image synthesis, mammogram, computer-aided diagnosis

1. INTRODUCTION

In the U.S., breast cancer is the second leading cause of death among women, and it will be diagnosed in about 12% of U.S. women during their lifetime [1,2]. The commonly used mammographic detection based on computer-aided detection (CAD) methods can improve treatment outcomes for breast cancer and increase patient survival times [3]. These traditional CAD tools, however, have a variety of drawbacks because they rely on manually designed features. For example, hand-crafted features tend to be domain-specific, and the process of feature design can be tedious, difficult, and non-generalizable [4]. In recent years, developments in machine learning have provided alternative methods for feature extraction; one is to learn features from whole images directly through a Convolutional Neural Network (CNN) [5,6]. Usually, training the CNN from scratch requires a large number of labeled images [7]; for example, the AlexNet (a classical CNN model) was trained by using about 1.2 million labeled images [8]. For some kinds of medical image data such as mammographic tumor images, to obtain a sufficient number of images to train a CNN classifier is difficult because the true positives are scarce in the datasets and expert labeling is expensive [9]. The shortcomings of an insufficient number of images to train a classifier are well-known [8,10], so it is worthwhile to examine image augmentation as a way to create new training images and thus to improve the performance of a CNN classifier.

As with CNN, the Generative Adversarial Network (GAN) is a neural network-based learning method introduced by Goodfellow *et al.* in 2014 [11], and it is a state-of-the-art technique in the field of deep learning [12]. GAN has many novel applications in the field of image processing, including image translation [13,14], object detection [15], super-resolution,[16] and

image blending [17]. Recently, various GANs have been developed for medical imaging, such as GANCS [18] for MRI reconstruction, SegAN [19], DI2IN [20] and SCAN [21] for medical image segmentation. As far as we aware there is no study about using GAN as a data augmentation method on mammograms to train CNN classifier for breast cancer detection. Our study aims to fill that gap.

Previous approaches to image augmentation used original images modified by rotation, shifting, scaling, shearing, and/or flipping. We name the original images **ORG images**, and the images augmented by affine transformation **AFF images** in the rest of this paper. The potential problem with such processing is that slightly changed images are similar to original ones; they may not be used as new training images to improve the performance of a CNN classifier. Large changes, on the other hand, may change the structure or pattern of objects in training images and degrade the performance of the classifier. An alternative image augmentation method is to generate synthetic images using the features extracted from original images. Such generated images are not exactly like the original ones but could keep the essential features, structures, or patterns of the objects in the original images. Therefore, GAN is a good candidate as a method for augmenting the training dataset. We name those images the **GAN images**.

In this study, we firstly cropped the regions of interest (ROIs) from images in the Digital Database for Screening Mammography (DDSM) [22] database as original ROIs. Second, by using these ORG ROIs, we applied GAN to generate the same number of GAN ROIs. To compare with the affine transformation augmentation, we also used ORG ROIs to generate the same number of AFF ROIs. Then, we used six groups of ROIs: GAN ROIs, AFF ROIs, ORG ROIs and three mixture groups of each pair of the three simple ROIs to train a CNN classifier from scratch for each group. We used the ORG ROIs that had never been used in augmentation and training to validate classification outcomes. Our results illustrate that to classify the normal ROIs and abnormal (tumor) ROIs from DDSM, adding GAN ROIs to the training data can prevent over-fitting of the classifier. The maximum validation accuracy for training by only GAN ROIs is about 83.1%; it shows that the synthetic ROIs generated from a GAN can retain some important features, structure, or patterns from ORG ROIs. But the GAN ROIs may have more different features than ORG ROIs, thus, the serious over-fitting occurred. Adding ORG ROIs in the training set can help correct this problem. Affine transformations perform slightly better than GAN; GAN could be an alternative augmentation option. The images augmented by GAN or affine transformation cannot substitute for real images to train CNN classifiers, however, because the absence of real images from the training set will cause serious over-fitting with further training.

2. METHODS

2.1 The Mammogram Databases and Image Pre-processing

Mammography is the process of using low-energy X-rays to examine the human breast for diagnosis and screening. The goal of mammography is the early detection of breast cancer [23], typically through detection of masses or abnormal regions from the formed X-ray images. Usually, such abnormal regions are detected by doctors or expert radiologists. In this study, we used mammograms from the Digital Database for Screening Mammography (DDSM) [22]. The DDSM is a widely used mammographic images resource by the U.S. Mammographic Image Analysis Research Community. It is a collaborative effort between Massachusetts General Hospital, Sandia National Laboratories, and the University of South Florida Computer Science and Engineering Department.

DDSM describes the location and boundary of actual abnormalities by chain-codes, which are recorded in OVERLAY files for each breast image containing abnormalities. The DDSM Utility [24] provides the tool to read boundary data and display them for each image having abnormalities. We used the ROIs instead of entire images to train CNN classifiers. To define abnormal ROIs from images containing abnormalities, we used the minimum bounding box areas surrounding the whole given ground truth boundaries. Normal ROIs, are defined similarly; their sizes are approximately the average size of abnormal ROIs. Locations of normal ROIs are selected randomly on normal breast areas. These images cropped from mammogram are ORG ROIs.

2.2 Image Augmentation by Affine Transformation

The image augmentation by affine transformations that we applied on ORG ROIs are: rotation, width shifting, height shifting, shearing, scaling, horizontal flipping and vertical flipping. All transformations are applied randomly, and some have limited ranges. The range of rotation is 0-30 degrees and of width shifting, height shifting, shearing, scaling is 0%-20% according to the total image size. Since the input image size and position will be changed after affine transformations, padding (filling) points outside the boundaries is needed to maintain the size of the output image. Three

padding methods are considered: set a **constant** value for all pixels outside the boundaries, copy the values at the **nearest** pixel on the boundaries and **reflect** the image by the boundaries. Fig. 1 displays the results of the three padding methods. We will choose to use one padding method that can obtain the best classification accuracy.

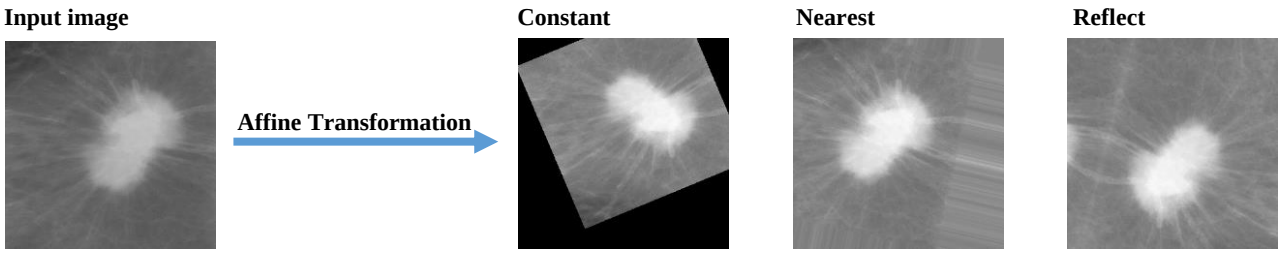

Fig. 1. The three affine transformations.

2.3 Image Augmentation by GAN

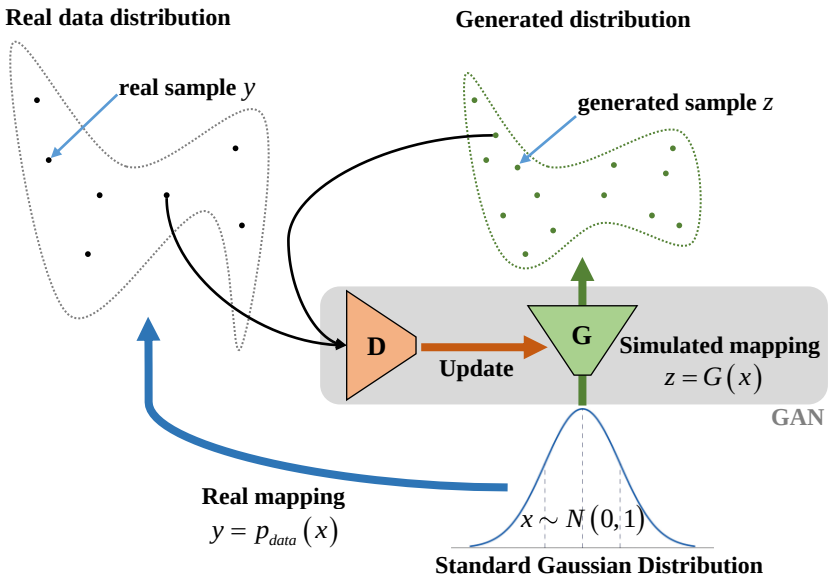

Fig. 2. The principle of GAN.

The GAN is a neural-network-based generative model that learns the probability distribution of real data and creates simulated data samples with a similar distribution (Fig. 2). Formally, in d -dimensional space, for $x \in R^d$, $y = p_{data}(x)$ is a mapping from x to real data y . We create a neural network called the **generator** G to simulate this mapping. If sample y comes from p_{data} , it is a real one; and if sample z comes from G , it is a synthetic one. Another neural network **discriminator** D is used to detect whether a sample is real or synthetic. Ideally, $D(y)=1; D(z)=0$. The two neural networks G and D compose the GAN. We can find G and D by solving the two-player minimax game [11], with value function $V(G,D)$:

$$\min_G \max_D V(G,D) = \mathrm{E}\left[\log D\left(p_{data}(x)\right)\right] + \mathrm{E}\left[\log\left(1 - D\left(G(x)\right)\right)\right] \tag{1}$$

This min-max problem has a global optimum (Nash equilibrium) solution for $G(x) = p_{data}(x)$. That is the goal to find the distribution of real data. At equilibrium, discriminator D can no longer distinguish the real from the synthetic

sample, where $D(y) = D(z) = 0.5$. Synthetic samples can be generated from G by changing the input x. In this study, the input x for G we used was a noise vector having 100 elements from a Gaussian distribution $\sim N(0,1)$. The key point of a well-trained GAN is that it could generate seemingly real-like data samples using noise vectors. To train a GAN, we used a limited number of real samples. Ideally, a GAN could generate unlimited different synthetic samples.

To implement GAN, we built the generator and discriminator neural networks. The generator consists of four up-sampling layers to double the size of image, and five convolutional layers. The activation function for each layer is the ReLU function [25] except the last one for output, which is a tanh function. The function of the generator is to transform a 100-length vector to a 320x320x3 image. The input of the discriminator is a 320x320x3 image and its output is a value between 0 and 1, which '0' stands for the synthetic image and '1' for the real one. Like a typical CNN, the discriminator has four convolutional layers with max-pooling layers and one fully-connected (FC) layer. The activation function for each convolutional layer is also the ReLU function and the output layer is a sigmoid function, which maps the output value to the range of [0, 1].

The training methods of GAN are:

- Step 1: Randomly initialize all weights for both networks.
- Step 2: Input a batch of 100-length noise vectors to the generator to obtain synthetic images.
- Step 3: To train the discriminator by a batch of synthetic images labeled '0' and real images labeled '1'.
- Step 4: To train the generator: input a batch of 100-length noise vectors to the generator to obtain synthetic images and label them as '1'. Then, input these synthetic images to discriminator to obtain the predicted labels. The differences between predicted labels and '1' will be the loss for updating the generator. It is noteworthy that in this step, only the weights in the generator are changed; weights in the discriminator are fixed.
- Step 5: Repeat Step 2 to Step 4 until all real images have been used once, that defines one **epoch**. When the number of epochs reaches a certain value, training stops.

Actually, for Step 5, the ideal situation to stop training is when the classification accuracy of the discriminator converges to 50%. It means the discriminator no longer can distinguish the real images and synthetic images generated from a well-trained generator. The discriminator plays a role as an assistant in GAN. After training, we will use the generator neural networks to generate synthetic images for usage in the next step.

2.4 CNN for classification

A CNN was designed as the discriminator in GAN. Its function is to distinguish real and synthetic mammographic ROIs. We also built a CNN to classify abnormal ROIs and normal ROIs, it is called CNN tumor classifier. This CNN classifier consists of three convolutional layers with max-pooling layers and two FC layers. The activation function for each layer is the ReLU function except the last one for output. The output layer uses a sigmoid function, which maps the output value to the range of [0, 1]. Its input is the image in size 320×320-pixel. Since the sigmoid function was used in the output layer, the predicted outcome from the CNN classifier is a value between 0 and 1. By default, the classification threshold is 0.5, meaning that if the value is less than 0.5 it will be considered as "0" (normal), otherwise it will be considered as "1" (abnormal). The optimizer for training is Nadam using default parameters [26] (except the learning rate changed to 1e-4), the loss function is Binary Cross Entropy, the updating metric is Accuracy, the batch size is 26 and the number of total epochs is set to be 750. To train this CNN classifier from scratch, we used the labeled ROIs of abnormal and normal mammographic images. All training data include ORG ROIs, AFF ROIs and GAN ROIs, but validation data are only the ORG ROIs.

3. EXPERIMENT AND RESULTS

3.1 Experiment Plan

We collected 1300 real abnormal ROIs (O_{abnorm}, 'O' for original) and 1300 real normal ROIs (O_{norm}) in total. After taking off 10% for validation, there are 1170 O_{abnorm} and 1170 O_{norm}. We firstly did the data augmentation to 1170 O_{abnorm} and 1170 O_{norm} by affine transformations to obtain 1170 A_{abnorm} ('A' for affine) and 1170 A_{norm}; the details are shown in Sec. 2.2. To test the three padding methods, we trained three CNN classifiers from scratch by three types of

augmented datasets respectively and found that the CNN classifier trained by nearest padding AFF ROIs has the best overall performance. Therefore, we used the **nearest padding** AFF ROIs for the remaining experiments.

We then used the ORG ROIs to train two generators: GAN_{abnorm} and GAN_{norm} for generating GAN ROIs. As shown in Fig. 2, during the training process, the generator G provided synthetic ROIs to discriminator D. D was trained to distinguish the real from the synthetic ROIs by real and synthetic ROIs. And, once synthetic ROIs were distinguished, D gave feedback loss to G for G's updating. Then G will generate synthetic ROIs more like the real ones. By inputting noise vectors to GAN_{abnorm} and GAN_{norm}, we obtained 1170 G_{abnorm} and 1170 G_{norm}.

Table 1. Training plans.

Set#	Dataset for training	Set#	Dataset for training	Set#	Dataset for training
1	1170 O_{abnorm} labeled '1' 1170 O_{norm} labeled '0'	**2**	1170 A_{abnorm} labeled '1' 1170 A_{norm} labeled '0'	**3**	1170 G_{abnorm} labeled '1' 1170 G_{norm} labeled '0'
4	1170 O_{abnorm} + 1170 A_{abnorm} labeled '1' 1170 O_{norm} + 1170 A_{norm} labeled '0'	**5**	1170 O_{abnorm} + 1170 G_{abnorm} labeled '1' 1170 O_{norm} + 1170 G_{norm} labeled '0'	**6**	1170 A_{abnorm} + 1170 G_{abnorm} labeled '1' 1170 A_{norm} + 1170 G_{norm} labeled '0'

We repeated training the CNN classifier from scratch using different datasets of labeled ROIs shown in Table 1. In each set, the number of abnormal ROIs and normal ROIs is equal. We used 130 O_{abnorm} and 130 O_{norm} that were never used in the training process as validation data to evaluate those CNN classifiers.

3.2 Classification Results

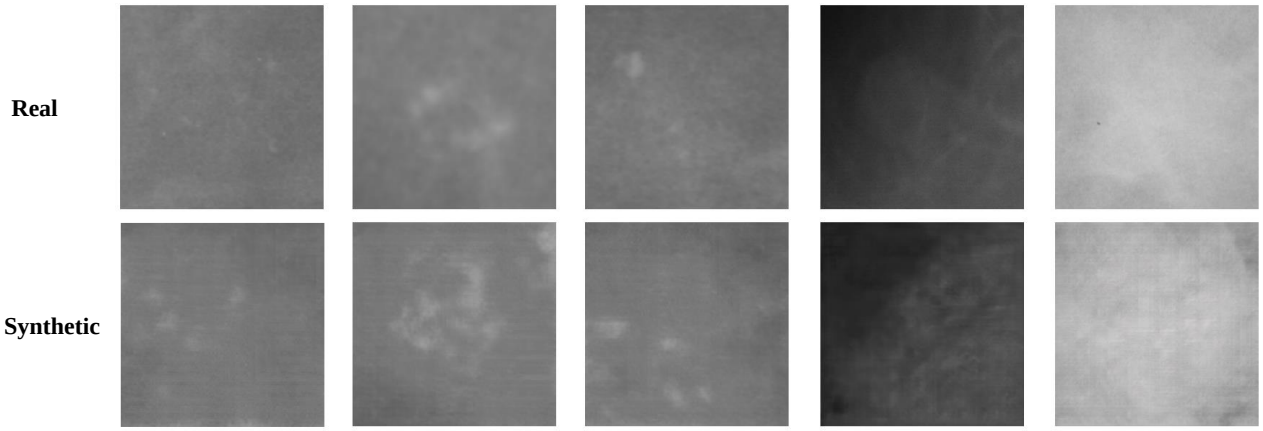

Fig. 3. (Top row) Real **abnormal** ROIs; (Bottom row) synthetic **abnormal** ROIs generated from GAN.

Specifically, for training GAN, we used 1170 real abnormal ROIs to obtain the generator GAN_{abnorm}, and used 1170 real normal ROIs to obtain the generator GAN_{norm}. Fig. 3 shows some synthetic abnormal ROIs (G_{abnorm}) generated from GAN_{abnorm}. Then, we generated 1170 G_{abnorm} and 1170 G_{norm} by generators.

Table 2. Analysis of validation accuracy for CNN classifiers.

Set#	Best performance (%)	**Stable performance (%)**	Over-fitting?
1 (ORG)	88.85	**85.02**	Yes
2 (AFF)	81.15	**67.99**	Yes
3 (GAN)	83.08	**60.72**	Yes
4 (ORG + AFF)	89.62	**87.07**	No
5 (ORG + GAN)	87.31	**85.40**	No
6 (AFF + GAN)	85.38	**68.32**	Yes

The results of training accuracy and validation accuracy after each training epoch (it is defined in Sec. 2.3, training methods, Step 5; the total epochs are 750) are shown in Fig. 4. It is evident that Sets 1, 4, and 5 perform well, and 3 is the worst. To analyze those results quantitatively, we show the maximum validation accuracy, average validation accuracy after 600 epochs and whether over-fitting occurred. The maximum validation accuracy can indicate the **best performance** of the classifier, but it may be reached fortuitously. The average validation accuracy after 600 epochs can show the **stable performance** of the classifier. For a good classifier, this value will be monotone increasing and convergent. The criterion for **occurrence of over-fitting** is defined by the value: average validation accuracy after 400 epochs minus (-) average validation accuracy before 400 epochs; if it is negative, then we consider that the over-fitting occurred because of the decrease of validation accuracy during training. Table 2 shows these quantitative results.

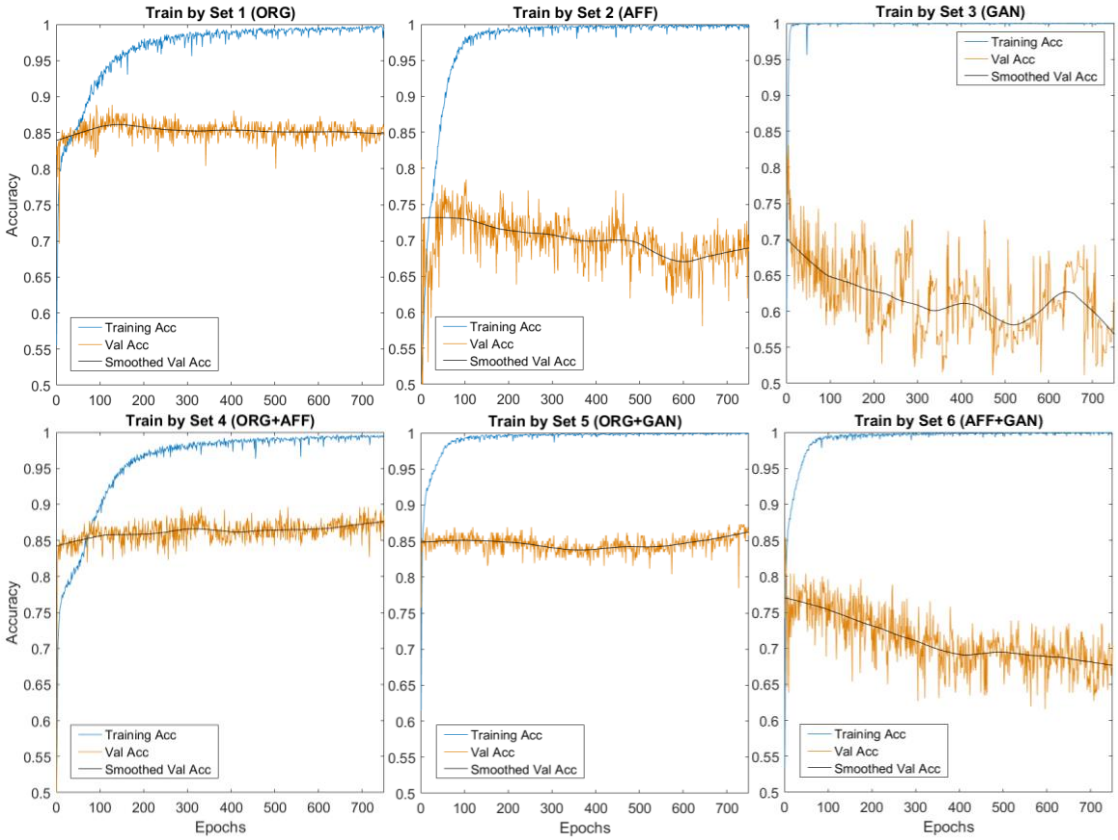

Fig. 4. Training accuracy and validation accuracy for six training datasets.

Since the maximum validation accuracy may be fortuitous, the stable performance during training is a more-reliable measure of a classifier. We can see from Table 2, that:

- ORG ROIs must be included in the training set because the stable performances of sets without ORG ROIs are lower than 70%.
- By comparing Set 1 with Set 4 and Set 5, to avoid over-fitting, image augmentation is required.
- By comparing Set 2 with Set 3, affine-transformed images could have features closer to real images than GAN-generated images. And, by comparing Set 4 with Set 5, we see that AFF ROIs are better than GAN ROIs for image augmentation. Examination of the synthetic ROIs in Fig. 3 indicates that they have some artificial characteristics.
- Since the performance of the affine transformation is similar to GAN for image augmentation (their validation accuracies in Fig. 4 are close and no over-fitting occurred), GAN could be an alternative augmentation method for training CNN classifiers.

For training by only real ROIs, the validation accuracy is as good as training by adding GAN ROIs. But adding GAN ROIs can help avoid over-fitting. Adding AFF ROIs can not only avoid over-fitting, but also improves the validation accuracy. Therefore, image augmentation is necessary to train CNN classifiers and although the affine transformation performs slightly better than GAN, GAN still could be an alternative option.

4. DISCUSSION

The hypothesis of GAN is that, in d -dimensional space, there exists a mapping function $p_{data}(x)$ from vector x to real data y; a GAN can learn and simulate the mapping function $G(x)$ using samples from the distribution of real data. $G(x)$ is also called a generator. The ideal outcome is $G(x) = p_{data}(x)$. The maximum validation accuracy for training by GAN ROIs is about 83.1%, which shows that the generator acquired some important features from ORG ROIs. But the GAN ROIs may have more different features compared to ORG ROIs, thus, serious over-fitting occurred with more training. Adding ORG ROIs in the training set can help correct this problem.

Theoretically, a well-trained GAN could generate images having the same distributions as real images. If so, the performance of a CNN classifier trained by GAN ROIs will be as good as when trained by ORG ROIs. Our results, however, show that, based on training performance, GAN did not meet theoretical expectations. For image augmentation, GAN is not even as good as affine transformations. The problem could be found by looking at the synthetic images (Fig. 3): they have clear artificial characteristics. One possible reason is that GAN adds some features or information not belonging to real images. Those new features lead classifiers to detect abnormal features in real images and make the validation accuracy lower than when using affine transformations. The possible solution is to change the architecture of the generator or/and discriminator in GAN. In this paper, the architecture we used is DCGAN [27]. Actually, there are about 300 architectures of GAN [28] proposed recently. We believe that some of them may achieve a better performance than affine transformations for image augmentation.

In future work, we could apply transfer learning to train the classifier because transfer learning is another important approach to deal with insufficient amounts of training data, besides data augmentation. Since the DDSM provides truth labels for benign and malignant tumors, we could also do classification for benign and malignant ROIs instead of abnormal and normal ROIs. Also, we may examine performances of other architectures of GAN in terms of image augmentation.

5. CONCLUSION

In this paper, we applied GAN to generate synthetic mammograms. GAN can be used as an image augmentation method for training and to improve the performance of CNN classifiers. Our results show that, to classify the normal ROIs and abnormal (tumor) ROIs from DDSM, adding GAN generated ROIs in training data can help the classifier prevent from over-fitting (Table 2). Another traditional image augmentation method -- affine transformation, however, performs slightly better than GAN; GAN could be an alternative augmentation option. By comparing GAN ROIs with affine-transformed ROIs in their distributions of mean, standard deviation, skewness and entropy, we found that except the mean's, affine transformed ROIs are more similar to real ROIs than GAN ROIs in terms of the other three distributions. Our results also show that images augmented by GAN or affine transformation cannot substitute for real images to train CNN classifiers because the absence of real images in training set will cause serious over-fitting with more training.

REFERENCES

[1] Siegel, R. L., Miller, K. D. and Jemal, A., "Cancer statistics, 2016," CA. Cancer J. Clin. **66**(1), 7–30 (2016).

[2] DeSantis, C. E., Fedewa, S. A., Goding Sauer, A., Kramer, J. L., Smith, R. A. and Jemal, A., "Breast cancer statistics, 2015: Convergence of incidence rates between black and white women," CA. Cancer J. Clin. **66**(1), 31–42 (2016).

[3] Rao, V. M., Levin, D. C., Parker, L., Cavanaugh, B., Frangos, A. J. and Sunshine, J. H., "How Widely Is Computer-Aided Detection Used in Screening and Diagnostic Mammography?," J. Am. Coll. Radiol. **7**(10), 802–805 (2010).

[4] Yi, D., Sawyer, R. L., Cohn III, D., Dunnmon, J., Lam, C., Xiao, X. and Rubin, D., "Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors," ArXiv170506362 Cs (2017).

[5] Lo, S.-C. B., Chan, H.-P., Lin, J.-S., Li, H., Freedman, M. T. and Mun, S. K., "Artificial convolution neural network for medical image pattern recognition," Neural Netw. **8**(7–8), 1201–1214 (1995).

[6] Jamieson, A. R., Drukker, K. and Giger, M. L., "Breast image feature learning with adaptive deconvolutional networks," presented at Medical Imaging 2012: Computer-Aided Diagnosis, 23 February 2012, 831506, International Society for Optics and Photonics.

[7] Erhan, D., Manzagol, P.-A., Bengio, Y., Bengio, S. and Vincent, P., "The Difficulty of Training Deep Architectures and the Effect of Unsupervised Pre-Training" (2009).

[8] Krizhevsky, A., Sutskever, I. and Hinton, G. E., "ImageNet Classification with Deep Convolutional Neural Networks," [Advances in Neural Information Processing Systems 25], F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 1097–1105 (2012).

[9] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. and Summers, R. M., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Trans. Med. Imaging **35**(5), 1285–1298 (2016).

[10] Pinto, N., Cox, D. D. and DiCarlo, J. J., "Why is Real-World Visual Object Recognition Hard?," PLOS Comput. Biol. **4**(1), e27 (2008).

[11] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., "Generative Adversarial Nets," [Advances in Neural Information Processing Systems 27], Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2672–2680 (2014).

[12] Hong, Y., Hwang, U., Yoo, J. and Yoon, S., "How Generative Adversarial Networks and its variants Work: An Overview of GAN," ArXiv171105914 Cs (2017).

[13] Wang, C., Xu, C., Wang, C. and Tao, D., "Perceptual Adversarial Networks for Image-to-Image Transformation," ArXiv170609138 Cs (2017).

[14] Yi, Z., Zhang, H., Tan, P. and Gong, M., "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," ArXiv170402510 Cs (2017).

[15] Li, J., Liang, X., Wei, Y., Xu, T., Feng, J. and Yan, S., "Perceptual Generative Adversarial Networks for Small Object Detection," ArXiv170605274 Cs (2017).

[16] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," ArXiv160904802 Cs Stat (2016).

[17] Wu, H., Zheng, S., Zhang, J. and Huang, K., "GP-GAN: Towards Realistic High-Resolution Image Blending," ArXiv170307195 Cs (2017).

[18] Mardani, M., Gong, E., Cheng, J. Y., Vasanawala, S., Zaharchuk, G., Alley, M., Thakur, N., Han, S., Dally, W., Pauly, J. M. and Xing, L., "Deep Generative Adversarial Networks for Compressed Sensing Automates MRI," ArXiv170600051 Cs Stat (2017).

[19] Xue, Y., Xu, T., Zhang, H., Long, R. and Huang, X., "SegAN: Adversarial Network with Multi-scale L_1 Loss for Medical Image Segmentation," ArXiv170601805 Cs (2017).

[20] Yang, D., Xiong, T., Xu, D., Huang, Q., Liu, D., Zhou, S. K., Xu, Z., Park, J., Chen, M., Tran, T. D., Chin, S. P., Metaxas, D. and Comaniciu, D., "Automatic Vertebra Labeling in Large-Scale 3D CT using Deep Image-to-Image Network with Message Passing and Sparsity Regularization," ArXiv170505998 Cs (2017).

[21] Dai, W., Doyle, J., Liang, X., Zhang, H., Dong, N., Li, Y. and Xing, E. P., "SCAN: Structure Correcting Adversarial Network for Organ Segmentation in Chest X-rays," ArXiv170308770 Cs (2017).

[22] Heath, M., Bowyer, K., Kopans, D., Moore, R. and Kegelmeyer, W. P., "The digital database for screening mammography," Proc. 5th Int. Workshop Digit. Mammogr., 212–218, Medical Physics Publishing (2000).

[23] Friedewald, S. M., Rafferty, E. A., Rose, S. L., Durand, M. A., Plecha, D. M., Greenberg, J. S., Hayes, M. K., Copit, D. S., Carlson, K. L., Cink, T. M., Barke, L. D., Greer, L. N., Miller, D. P. and Conant, E. F., "Breast Cancer Screening Using Tomosynthesis in Combination With Digital Mammography," JAMA **311**(24), 2499–2507 (2014).
[24] Sharma, A., [DDSM Utility], GitHub (2015).
[25] Nair, V. and Hinton, G. E., "Rectified linear units improve restricted boltzmann machines," Proc. 27th Int. Conf. Mach. Learn. ICML-10, 807–814 (2010).
[26] Kingma, D. P. and Ba, J., "Adam: A Method for Stochastic Optimization," ArXiv14126980 Cs (2014).
[27] Radford, A., Metz, L. and Chintala, S., "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," ArXiv151106434 Cs (2015).
[28] Hindupur, A., [the-gan-zoo: A list of all named GANs!] (2018).