Biomedical Data Science
2025, VOL. 5, NO. 3, 15-20
DOI: 10.47297/wspbdsWSP2752-630503.20250503

Interpretable GAN-based Breast Ultrasound Mass Augmentation for Improved Deep Learning BI-RADS Classification

Jie Zhu[1,2], Shaoyong Guo[2], Junfeng Ma[2], Ting Pang[2]*

[1]Henan Medical University, School of Health Management, Xinxiang, Henan, 453000, China

[2]Henan Medical University, Xinxiang High Performance Computing Medical Engineering Technology Research Center, Xinxiang, Henan, 453000, China

ABSTRACT

Generative adversarial network (GAN) synthesizing images has been widely applied to data augmentation. However, existing GAN-based imaging generation has drawbacks in explaining the synthetic principles. Hence, generative approaches resolving the limited medical images for deep learning-based computer aided diagnosis (CAD) which requires large amounts of data are scanty. In this paper, we propose an interpretable data synthesis network using modified triple generative adversarial network deployed into deep learning-based breast imaging reporting and data system (BI-RADS) classification of breast ultrasound mass. The mass synthetic network consists of an interpretable GAN that grasps the standardized characterizations described in BI-RADS for mass. Then diagnostic network with synthetic data determines the BI-RADS level of breast mass based on multi-classifier. Experiments were performed on a private breast ultrasound mass dataset to show that the synthetic network can provide interpretability of generative mass. In addition, we demonstrate that the synthetic masses improve the performance of breast cancer classification.

KEYWORDS

Interpretable GAN-based; Data augmentation; Breast ultrasound; BI-RADs classification; Deep learning

1 Introduction

Recently, deep learning technology, convolutional neural network (CNN), is now prevalently applied to medical imaging for computer aided diagnosis (CAD) [1-3]. Though deep learning requires large amounts of data for training, medical images are limited due to the complex and difficult annotations through collaborations between radiologists and researchers. Along with the popular applications of generative adversarial networks (GANs) in imaging synthesis, some scholars also have tried to deploy GAN for enlarging the medical imaging dataset [4-5]. However, existing GAN-based medical imaging generation has drawbacks in explaining the synthetic principles. Hence, generative approaches that augments the dataset in deep learning CAD are scanty in real applications.

Breast ultrasound mass is one of the most important signs for breast cancer which is the leading cancer among females [6-7]. Breast imaging reporting and data system (BI-RADS) is extensively used as the criteria to diagnose (BIRADS categories for breast ultrasound are listed in Table 1) and comment the breast abnormalities by radiologists [8]. To reduce the reviewing burdens of complicated breast cancer imaging, deep learning methods are widespread for CAD of breast abnormalities [9-10]. But the fact that publicly available breast cancer datasets are all mammography and extremely challenging data collection obstructs study on deep learning-based CAD of breast ultrasound without massive data. Therefore, synthetic augmentation utilizing interpretable GAN for breast ultrasound mass with low contrast is necessary.

Table 1 BI-RADS categories for breast ultrasound Imaging.

BI-RADS	Diagnosis
Category0	Need additional imaging evaluation.
Category 1	Normal ultrasound.
Category 2	Benign finding that does not require any further imaging.
Category 3	Probably benign finding that requires short term follow-up imaging.
Category 4	Suspicious abnormality, biopsy should be considered.
Category 5	Highly suggestive of malignancy that are almost always malignant.
Category 6	Known biopsy – proven malignancy- appropriate action should be taken.

In this paper, we propose an interpretable data synthesis network using modified triple generative adversarial network (TGAN) [11] deployed into deep learning-based BI-RADS classification of breast ultrasound mass. The mass synthetic network consists of an interpretable GAN that grasps the standardized characterizations described in BI-RADS for mass (shape, margin, echo pattern, posterior features). Then augmented dataset with synthetic masses are fed into deep learning-based diagnostic network to determine the BI-RADS level of breast mass based on multi-classifier. The benefits of

*** Corresponding Author:** Ting Pang, pt@xxmu.edu.cn

our method are: 1) the synthetic network can provide interpretability of generative mass by the explainable GAN; 2) we demonstrate that the interpretable synthetic masses improve the performance of breast cancer classification. This explainable network promotes the employment of GAN methods in deep learning CAD with enlarging dataset to improve the diagnostic performance of medical imaging.

2 Related works

2.1 Generative adversarial network

Generative adversarial network (GAN) inspired by the game theory learns the high-dimensional distributions of data [12]. It consists of two neural networks in competition: the generator G exploring to generate fake images and the discriminator D judging real or fake images repeatedly until obtaining synthetic images with high quality. At present, GAN emerges as a popular synthetic method to address the challenges of augmenting dataset in deep learning-based image analysis [13]. Besides, there are lots of variants of GAN architecture to improve the generative performance, like CGAN [14], WGAN [15], InfoGAN [16], LapGAN [17], SeqGAN [18], et al. In medical images, GAN have been developed in translation [19], segmentation [20] detection [21] and classification [22].

2.2 Breast ultrasound mass classification with deep learning

Ultrasound which uses non-ionizing radiation is more sensitive in dense breasts and effective in distinguishing the cysts from solid breast masses [6]. Deep learning has been used in breast ultrasound since 2016 [23] and most of the studies try to detect and classify the breast masses with binary category as benign or malignant [24-25]. But only one paper is about BI-RADS classification (listed in Table 1) of breast ultrasound mass as annotation is a tougher work. Huang et al. differentiated the identified regions of interest (ROIs) of breast tumors with ultrasound modality into five categories: category "3", category "4A", category "4B", category "4C", and category "5" based on CNN [26].

3 Method

The proposed architecture is shown in Fig.1 and the pipeline includes two parts: (1) modified TGAN as the interpretable GAN model to generate samples; (2) CNN-based diagnostic model based on a multi-BIRADS-classifier with synthetic data and real data.

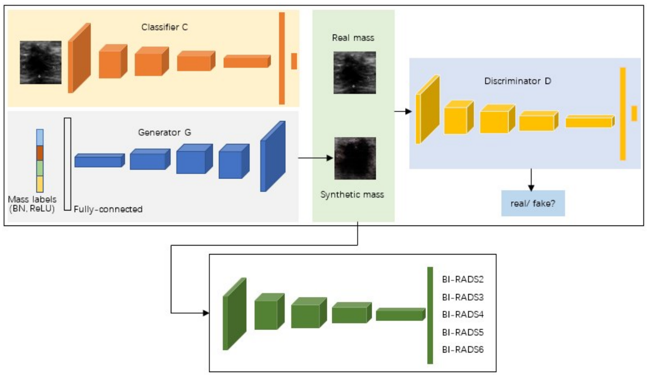

Figure 1 The proposed architecture. BN (Bach Normalization) and lReLU (Leaky Rectified Linear Unit).

3.1 Interpretable GAN model

According to the clinical reports written by experienced radiologists and the BI-RADS guidelines about the breast ultrasound masses, the shape, margin, echo pattern and posterior features are the main characterizations to measure the malignancy, seen in Table 2. Therefore, to enhance the explainability of the synthetic principles, we attempt to connect the generative model with the criterions that determine the BI-RADS level. Here, we utilize the TGAN as the generative model with advantages of stable training, easy convergence and abundant generated samples. Then four main features of masses are respectively treated as the label code $l^m \in R^M$ (m denotes the four characterizations and M denotes the number of masses). Each l^m is converted to label channel that is embedded to 128x128.

dimensional vector by passing through the BN (Bach Normalization), ReLU (Rectified Linear Unit) and one fully

Table 2 BI-RADS descriptions for the characterizations of breast ultrasound mass.

Characterizations	Descriptions
Shape	Round, oval (probably benign) or irregular (probably malignant).
Margin	Circumscribed (probably benign), microlobulated or obscured, indistinct, spiculated (probably malignant).
Echo pattern	Anechoic, hypoechoic, complex cystic and solid, isoechoic, hyperechoic or heterogeneous (together with other features).
Posterior features	Enhancement or shadowing (together with other features).

connected layer. Next, the label code is concatenated with noise z as the input of generator. Furthermore, the discriminator determines the mass whether is real or fake for the input image x.

In the generative model, generator G and discriminator D are trained simultaneously. The adversarial losses for the two parts are defined as formulas below:

$$L_G = \min \{-E_{z\sim p(z),\ I^m\sim p(I^m)}\left[\log(1 - D(G(z, I^m)))\right]\} \quad (1)$$

$$L_C = \min \{\alpha E_{(x_c,\ y_c)\sim p_c(x,\ y)}\left[\log(1 - D(x,\ y))\right]\} + R_L \quad (2)$$

$$L_D = \max \{E_{(x_l,\ y_l)\sim p(x,\ y)}\left[\log D(x,\ y)\right] + \alpha E_{(x_c,\ y_c)\sim p_c(x,\ y)}\left[\log(1 - D(x,\ y))\right] + (1-\alpha) E_{(x_g,\ y_g)\sim p_g(x,\ y)}\left[\log(1 - D(x, y))\right]\} \quad (3)$$

Where $\alpha \in (0, 1)$ is a constant which commands the relative importance between G and C. x is the input breast tumor image and y is the label (BI-RADS). For the generator G, the image x is generated by a noise $z\sim p(z)$ with the latent features code $I^m\sim p(I^m)$, which is $x = G(z, I^m)$. G minimizes the loss to learn the features for synthesis of similarly real masses with full to fool the discriminator. For the classifier C, a pseudo label y is assigned to image x by the model. For the discriminator D, it maximizes the loss to differentiate between the real and fake images x.

3.2 Diagnostic model

Here, we use inception-v3 having factorization designs to increase the diversity of features and reduce the number of parameters for avoiding overfitting [27]. According to the Table 1, BI-RADS 0 is abnormal breast ultrasound which needs another scanning and BI-RADS 1 does not reveal any indications of mass on breast ultrasound. So, considering the practical applications, we mainly focus on diagnosis from BI-RADS 2 to BI-RADS 6. The diagnostic result d is generated through the last softmax function as:

$$d = p(r) = \mathrm{softmax}(\omega r + b) \quad (4)$$

where r denotes the features obtained through the last fully connected layer. p(r) denotes the probability of each BI-RADS category. ω and b are the trained parameter matrix.

4 Experiments

4.1 Dataset and implementation details

The dataset of breast ultrasound mass images obtained in this research was from University of Malaya Medical Center (UMMC) under the institutional approval. We adopted 1447 images (original format is dicom) from 357 female patients (mean age, 51.6 y) between January 2018 and January 2019. All images were conducted and diagnosed by experienced radiologists according to BI-RADS. 1158 (80%) were used as training data and the rest were testing data (20%). The number of images with each BI-RADS for training and testing is shown in Fig. 2. To have a balance preserving resolution and computational complexity of the models, the images are cropped as ROIs and resized to the same size as 128×128 pixels. Furthermore, to initially resolve the prior problem of limited images, every ROI is flipped horizontally and vertically (seen in Figure 3). Hence, the number of imaging dataset becomes 4341.

We utilized the Tensorflow as framework and python language on a single GPU NVIDIA RTX 2070. For interpretable GAN, the batch size is 20, the size latent vector z is 100, epoch is 800, learning rate is 0.0002, optimizer is Adam (Adaptive Moment Estimation). For inception-v3, the batch size is 20, epoch is 100, learning rate is 0.0001, optimizer is Adam.

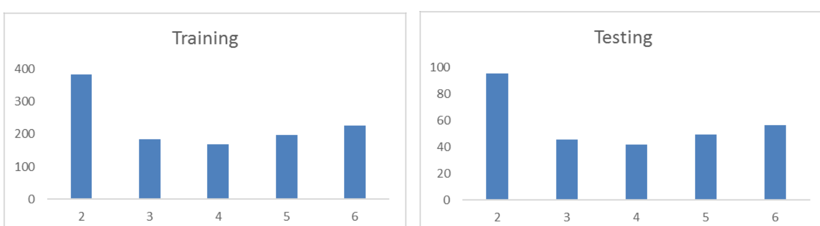

Figure 2 BI-RADS distribution of training data (a) and testing data (b).

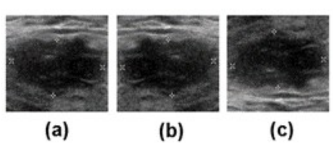

Figure 3 ROIs are flipped: original (a) horizontal (b) and vertical (c).

4.2 Results and analysis

4.2.1 Interpretable synthetic breast masses

The samples of generated breast ultrasound masses are exhibited in Fig.4. As shown in the figure, though the images have comparatively low resolution, they are visually similar with real breast lesions. The generated lesions captured most characteristics and elements, especially, the shape and margin are obvious while echo pattern and posterior features are obscure. Hence, the generative network mainly focuses on the features of shape and margin when generating masses. For the benign masses as BI-RADS2 and BI-RADS3, the shape is oval or round, the margin is circumscribed or lobulated. For the malignant masses as BI-RADS5 and BI-RADS6, the shape is irregular and the margin is lobulated, obscured, indistinct or spiculated. But for the masses as BI-RADS4, the shape and margin have all the features of benign or malignant masses which needs biopsy. Therefore, we can infer that it is meaningful that automatic classification training on the masses defined as BI-RADS4 and combining with histopathology report can avoid biopsy which are expensive and painful for patients. According to these visualizations, the proposed method effectively and interpretably synthesizes the breast masses with the characteristic labels.

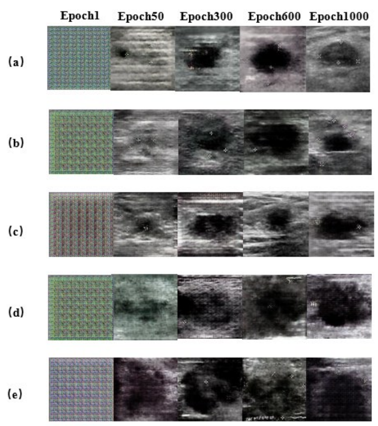

Figure 4 Generated samples of breast masses using proposed interpretable GAN. (a) Synthetic masses as BI-RADS2. (b) Synthetic masses as BI-RADS3. (c) Synthetic masses as BI-RADS4. (d) Synthetic masses as BI-RADS5. (e) Synthetic masses as BI-RADS6.

4.2.2 T-SNE visualization

To further analyze the synthetic masses, here deploys t-SNE (t-distributed stochastic neighbor embedding) algorithm for visualization. t-SNE is a machine learning algorithm for non-linear dimensionality reduction which is suitable for embedding high-dimensional data into 2 or 3 dimensions for visualization [28]. We visualized the synthetic images by comparing two situations: one with our interpretable GAN and one with original TGAN. Furthermore, we respectively and randomly selected 200 synthetic masses form the two scenarios for t-SNE visualization, as shown in Fig 5. We find that the benign masses as BI-RADS 2 and BI-RADS3 have many overlaps which are same with the malignant masses as BI-RADS5 and BI-RADS6. The benign masses are isolated with the malignant masses while the masses as BI-RADS4 have overlaps with the masses as BI-RADS3 and BI-RADS5. These results correlate with the obviously different features between benign and malignant masses while the masses of BI-RADS 4 have confusing features with either appearing benign or malignant masses. Besides, for the masses generated by interpretable GAN exhibit better separating localization than the masses generated by TGAN. This indicates that our method has increased synthetic performance by learning the representative features (shape and margin).

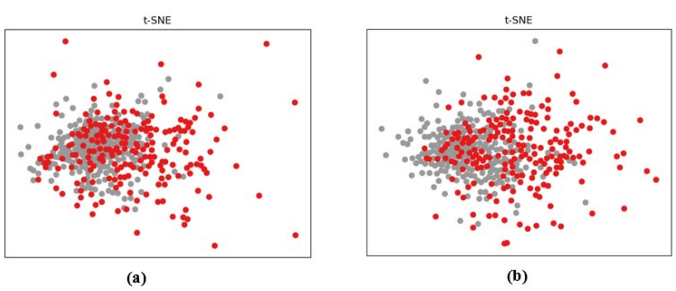

Figure 5 T-SNE visualization of synthetic masses. (a) Generated by interpretable GAN; (b) Generated by TGAN.

4.2.3 Improved AUC performance

In this section, we explored the effective mass classification based on BI-RADS using synthetic data augmentation by comparing with different inputs: 1) CNN with non-augmentation dataset; 2) AUG-CNN with only conventional data augmentation; 3) TL-CNN using transfer learning pretrained on ImageNet [29]; 4) GAN-CNN with synthetic data augmentation. Fig. 6 compares the classification performance in terms of AUC (area under curve) respectively with the four input datasets. AUC is the area under curve of receiver operating characteristic (ROC) with (1-specificity) for x-axis and sensitivity for y-axis. It can be seen from Fig. 6 that our method obtains the highest AUC. It is mainly due to the diagnosis network is trained with the generative lesions in adversarial strategy. Hence, from the compared and predicting classification results, we can easily discover that our method combination of synthetic classic data augmentation has overwhelming advantage in distinguishing breast masses.

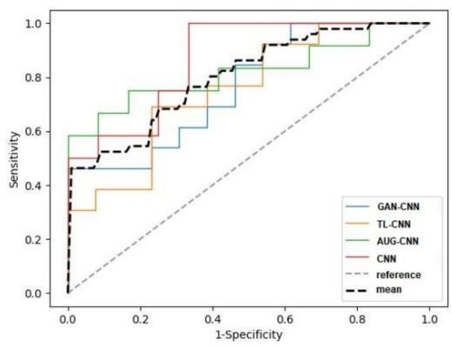

Figure 6 Comparison of AUC obtained from different methods.

5 Conclusion

In this paper, we proposed a novel interpretable GAN for breast ultrasound mass augmentation in improved BI-RADS classification to promote the employment of GAN methods in CAD. Our method utilizes modified TGAN to grasp the standardized characterizations (shape, margin, echo pattern, posterior features) described in BI-RADS for mass. After this, deep learning-based diagnostic network using augmented dataset with synthetic masses to judge the BI-RADS level based on multi-category of breast mass. Experiments on a small dataset are performed to validate the promising performance of the proposed method.

Funding

Selective Scientific Research Funding Poject for Overseas Returnees of Henan Province "Research on BI-RADS Multi-Class Classification Model for Mammographic Calcifications Based on Interpretable Deep Learning" (2025);International Science and Technology Cooperation Project of Henan Province "Research on Neuroimaging Diagnosis of Schizophrenia Based on Deep Reinforcement Learning and Attention Mechanism"(NO: 242102521012)

References

[1] Hu Z., Tang J., Wang Z., Zhang K., Zhang L., & Sun Q. (2018). Deep learning for image-based cancer detection and diagnosis– a survey. Pattern Recognition, 83, 134-149.

[2] Greenspan H., Van Ginneken B., & Summers R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging, 35(5), 1153-1159.

[3] Litjens G., Kooi T., Bejnordi B. E., Setio A. A. A., Ciompi F., Ghafoorian M., . . . Sanchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88

[4] Yi X., Walia E., & Babyn P. (2019). Generative adversarial network in medical imaging: A review. Medical image analysis, 101552.

[5] Wolterink J. M., Kamnitsas K., Ledig C., & Išgum I. (2018). Generative adversarial networks and adversarial methods in biomedical image analysis. arXiv preprint arXiv:1810.10352.

[6] Drukker K., Giger M. L., Vyborny C. J., & Mendelson E. B. (2004). Computerized detection and classification of cancer on breast ultrasound. Academic Radiology, 11(5), 526-535.

[7] Cheng H. D., Shan J., Ju W., Guo Y. H., & Zhang L. Automated breast cancer detection and classification using ultrasound images: A survey. Pattern Recognition, 2010, 43(1), 299-317.

[8] D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA, et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology; 2013.

[9] Yassin N. I. R., Omran S., El Houby E. M. F., & Allam H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using

different image modalities: A systematic review. Computer Methods and Programs in Biomedicine, 156, 25-45.

[10] Hamidinekoo A., Denton E., Rampun A., Honnor K., & Zwiggelaar R. Deep learning in mammography and breast histology, an overview and future trends. Medical Image Analysis, 2018, 47, 45-67.

[11] Chongxuan L. I., Xu T., Zhu J., & Zhang B. (2017). Triple generative adversarial nets. In Advances in neural information processing systems (pp. 4088-4098).

[12] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[13] Antonia Creswell, Tom White, Vincent Dumoulin, et al. Generative Adversarial Networks. Ieee Signal Processing Magazine, 2018,2:53-65.

[14] Mirza M., & Osindero S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

[15] Arjovsky M., Chintala S., & Bottou L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223).

[16] Chen X., Duan Y., Houthooft R., Schulman J., Sutskever I., & Abbeel P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In Advances in neural information processing systems (pp. 2172-2180).

[17] Denton E. L., Chintala S., & Fergus R. (2015). Deep generative image models using a laplacian pyramid of adversarial networks. In Advances in neural information processing systems (pp. 1486-1494).

[18] Yu L., Zhang W., Wang J., & Yu Y. (2017, February). Seqgan: Sequence generative adversarial nets with policy gradient. In Thirty-First AAAI Conference on Artificial Intelligence.

[19] Armanious K., Jiang C., Fischer M., Küstner T., Hepp T., Nikolaou K., ... & Yang, B. (2019). MedGAN: Medical image translation using GANs. Computerized Medical Imaging and Graphics, 101684.

[20] Xue Y., Xu T., Zhang H., Long L. R., & Huang X. (2018). Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. Neuroinformatics, 16(3-4), 383-392.

[21] Schlegl T., Seeböck P., Waldstein S. M., Schmidt-Erfurth U., & Langs G. (2017, June). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In International Conference on Information Processing in Medical Imaging (pp. 146-157). Springer, Cham.

[22] Salehinejad H., Valaee S., Dowdell T., Colak E., & Barfett J. (2018, April). Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 990-994). IEEE.

[23] Huynh B., Drukker K., & Giger M. (2016). Computer - aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks. Medical physics, 43(6Part30), 3705-3705.

[24] Shin S. Y., Lee S., Yun I. D., Kim S. M., & Lee K. M. (2019). Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. IEEE transactions on medical imaging, 38(3), 762-774.

[25] Zhou Y., Xu J., Liu Q., Li C., Liu Z., Wang M., ... & Wang, S. (2018). A radiomics approach with CNN for shear-wave elastography breast tumor classification. IEEE Transactions on Biomedical Engineering, 65(9), 1935-1942.

[26] Huang Y. Z., Han L. Y., Dou H. R., Luo H. H., Yuan Z., Liu Q., . . . Yin, G. F. (2019). Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images. Biomedical Engineering Online, 18(1), 8.

[27] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[28] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," Journal of Machine Learning Research, vol. 9, no. Nov, pp. 2579–2605, 2008.

[29] Krizhevsky A., Sutskever I., & Hinton G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).