ISBN: 978-1-948012-16-4

Contemporary Issues in Computing (CIC)

DOI: <http://doi.org/10.26480/cic.01.2020.72.75>

A MULTI-STAGE GAN BASED FRAMEWORK FOR THE GENERATION OF HIGH-RESOLUTION MAMMOGRAPHY IMAGES

Rangan Das^a, Pradip Basak^a, Utsav Bandyopadhyay Maulik^b, Saumik Bhattacharya^c, Ujjwal Maulik^{a*}^aDepartment of Computer Science and Engineering, Jadavpur University, Kolkata^bDepartment of Computer Science and Engineering, Techno Main Salt Lake, Kolkata^cDepartment of Electronics & Electrical Communication Engineering, IIT Kharagpur*Corresponding Author e-mail: umaulik@cse.jdvu.ac.in

This is an open access article distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 26 October 2020

Accepted 27 November 2020

Available online 03 December 2020

ABSTRACT

Generative adversarial networks (GANs) have garnered a lot of attention because of their ability to generate data. This synthetic data generation is crucial for training machine learning models for domains where data is scarcely available, such as medical imaging, where the generation of data is not only expensive but also has privacy concerns. In these cases, such synthetic data can be used for data augmentation, for detecting out-of-distribution samples or for domain transfer tasks. Generating high-resolution medical images, such as mammograms are challenging because of their detailed features as well as their tissue-specific properties. In this work, we have implemented a multi-stage GAN based framework that can be used to synthesize high-resolution mammography images. In the first stage, a low-resolution image is generated from latent noise, while in the second stage, an Enhanced Super Resolution GAN (ESRGAN) is used to upscale the image.

KEYWORDS

Generative Adversarial Networks, Deep Learning, Mammography, Medical Imaging.

1. INTRODUCTION

Breast cancer is one of the major causes of death from cancer among women (Ferlay et al., 2010; Ribli et al., 2018). Mammography is, at present, one of the best available non-invasive techniques that allow early detection and diagnosis of breast cancer (Broeders et al., 2012). This uses X-ray imaging to look for abnormalities in the breast tissue. Abnormalities that are indicative of breast cancer are generally calcification and masses (Hupse et al., 2013). However, there are also other types of abnormalities that can indicate breast cancer.

Indicators such as architectural distortions or bilateral symmetry are subtle and can be easily overlooked by radiologists (Bozek et al., 2009). Over the last two decades, Computer-Aided Detection (CAD) has been implemented in multiple clinical scenarios to help radiologists screen mammograms (Christoyianni et al., 2002). At present, CAD-based approaches make use of hand-crafted features as well as machine learning to help radiologists screen the images (Birdwell et al., 2001). CAD-based approaches have often been compared to be as accurate as double reading. However, these traditional methods have also been shown to not have improvements and CAD systems need to be improved before the technology is fully realized in everyday practice (Lehman et al., 2015).

However, since 2012, deep learning-based methods have shown a lot of promise in the field of medical image analysis (Krizhevsky et al., 2012). Several studies have implemented deep learning to analyse mammographic images. However, one of the biggest drawbacks of deep learning methods is the requirement of copious amounts of data. Deep

convolutional neural networks for classification or segmentation tasks require a large amount of data. However, there is a lack of publicly available datasets when it comes to mammograms. The dataset INbreast [Breast Research Group] is presently no longer maintained, the DDSM dataset (Lee et al., 2017) contains uncompressed images that require specialized tools for preprocessing, while other datasets such as the BCSC digital mammography dataset [BCSC] or the UCI Machine Learning Data Repository Mammographic Mass Data Set (Elter et al., 2007) does not contain any mammography images. The MIAS Mini-Mammographic Database (Suckling, 1994) contains 322 images only. The number is often not enough to train complex deep learning models.

Therefore, for the synthesis of realistic mammographic images is quite a crucial task as the increased dataset size allows deep learning models to learn the data better. Traditional data augmentation techniques such as flipping, rotating, elastic deformations or translation do not account for the change in location of the masses or calcifications in breast tissues. They also do not capture the variation that may arise because of the difference in processes or protocols. Hence, GANs are the best choice for image synthesis (Yi et al., 2019). Not only can they generate images, but they can also upscale low-resolution images.

In this paper, we propose a multi-stage framework which can be used to synthesize mammography images for training machine learning algorithms. The framework consists of two GAN (Radford et al., 2015) models, the first stage generates low-resolution images from latent noise while the second GAN is used to upscale the output of the first GAN to a high-resolution image.

Quick Response Code	This paper was presented at	Access this article online	
	International Conference on Contemporary Issues in Computing (ICCIC-2020) - Virtual IETE Sector V, Salt Lake, Kolkata From 25th-26th July 2020	Website: www.intelcomp-design.com	DOI: 10.26480/cic.01.2020.72.75

2. METHODOLOGY

2.1 Data Description

In this experiment, we have used the MIAS Mini-Mammographic Database. The database contains high-resolution grayscale images that are digitized from films using a Joyce-Loebl scanning microdensitometer. The database consists of 322 images that are padded, clipped and cantered, making all the images 1024 pixels x 1024 pixels. The database also contains areas of anomaly annotated by experts. Three classes of background tissues as well as seven different classes of anomalies are present. This provides a diverse yet compact set of images that can be used for training deep learning models from scratch.

2.2 Network Architecture

In this section, the architecture of the multi-stage generative adversarial networks that are implemented in this paper is described. The model works in two stages. At the core of these models are convolutional neural networks that are used to implement the generator and discriminator in two stages: The synthesis or the generation stage, and the upscaling stage. In the first stage, the generator (Figure 1) is implemented as a convolutional neural network that performs a series of fractionally-stride convolutions or deconvolution operations with interleaved up-sampling operations. The final layer is an untrainable median filter that is used to reduce the artifacting effects of the fractionally-strided convolution operation. The model takes an input a noise vector of size 100 and outputs an 80x80 image.

The Discriminator network consists of 6 convolutional layers and 4 batch normalization layers. Leaky ReLU is used as the activation function for the first 5 layers and Sigmoid for the last layer. Batch normalization is done only for layer 2, layer 3 and layer 4. The discriminator network outputs the probability of the input images in the batch being real or fake.

Figure 1: First stage of the GAN: (a) Generator architecture; (b) Discriminator architecture

Figure 2: (a) Dense block; (b) Residual-in-Residual Dense Block; (c) Generator network for the second stage of the GAN

Figure 3. Discriminator network for the second stage of the GAN.

For the second stage, an ESRGAN [Wang et al., 2018] is used for upscaling the 80x80 images to 320x320. In the ESRGAN, the generator network is constructed using Residual-in-Residual Dense Blocks (RRDBs) as illustrated in Figure 2. RRDBs are created using Residual Dense Blocks, which contain densely connected convolutional layers. Each RRDB block uses residual learning to using a group of five dense blocks. Since the RRDBs do not change the dimension of the feature maps, multiple RRDBs can be stacked to make the network size variable. In our experiment, the generator of the ESRGAN uses 23 RRDBs. The discriminator network is illustrated in Figure 3. Here, a patch-based network is used to ensure that the subtle textural details are reconstructed effectively. It is based on a simple feed-forward convolutional neural architecture with 4 discriminator blocks. Each block consists of two convolution layers, each followed by batch normalization and Leaky ReLU activation. Strided convolutions are used to eliminate the need for pooling. The number of feature maps is gradually increased from 64 to 512 iteratively in each block and then is reduced to 1.

2.3 Training

The training is done in two phases for each GAN. In the first case, for discriminator training, an original image and random noise are passed through the discriminator and generator network respectively. The real loss is calculated from the discriminator output and the generator output is fed into the discriminator to obtain the fake loss. Both the losses are then backpropagated through the discriminator network. For generator training, random noise is fed into the generator network and the output of it is fed into the discriminator. The loss from the discriminator is then backpropagated through the generator network. This process is repeated for 100 epochs with binary cross-entropy loss as the loss function

Overall, the GAN loss function is,

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

where, p_{data} is the original image distribution and p_z is the normal distribution from where the random noise is sampled and fed into the generator. The GAN is trained using 3000 images that are created by augmenting the MIAS dataset. The GAN was trained for 100 epochs with a learning rate of 0.0002. For optimization, Adam (Kingma et al., 2014) was used with $\beta_1 = 0.5$ and $\beta_2 = 0.99$.

The ESRGAN makes use of a Relativistic average Discriminator (RaD) that attempts to predict the probability that a real image I_r is relatively more realistic than a generated image I_f . Then the RaD is given as

$$D_{Ra}(I_r, I_f) = \sigma(C(I_r) - E_{I_f}[C(I_f)]) \quad (2)$$

Where $C(I_r)$ is the discriminator output, $\sigma(\cdot)$ is the sigmoid function and $E_{I_f}[\cdot]$ is the average of all the fake data in the mini-batch. Then the discriminator loss can be given as:

$$L_D = -E_{I_r} [\log (D_{Ra}(I_r, I_f))] - E_{I_f} [\log (1 - D_{Ra}(I_f, I_r))] \quad (3)$$

Similarly, the generator loss is given as:

$$L_G = -E_{I_r} [\log (1 - D_{Ra}(I_r, I_f))] - E_{I_f} [\log (D_{Ra}(I_f, I_r))] \quad (4)$$

The generator loss is further modified to includes perceptual loss as defined by Blau et al, 2018, as

$$L_{G_{total}} = L_p + \lambda L_G + \eta L_1 \quad (5)$$

Where $L_{G_{total}}$ is the total generator loss, L_p is the perceptual loss, $L_1 = E_{I_f} \|G(I_f) - y\|_1$ is the 1-norm distance between the generated image $G(I_f)$ and the ground truth y . The coefficients λ and η are used to

balance the different loss terms.

The ESRGAN was trained for 250 epochs on the MIAS dataset. A warm-up of 5,000 batches was done where only the generator was trained. After that, the generator and the discriminator were trained simultaneously. A learning rate of 0.0002 was used. For optimization, Adam was used with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. For the generator loss function, the coefficients λ and η was set to 0.0005 and 0.001 respectively. The models were implemented using PyTorch framework and were trained using an Nvidia Tesla K80 GPU.

3. RESULTS AND DISCUSSION

The multi-stage GAN framework has been trained on the MIAS mammography dataset. Figure 4(a) and 4(b) shows the training losses during the first and second stage respectively. Table 1 summarizes the final loss values that are achieved. Figure 5 and 6 show the qualitative results of the proposed model.

Table 1: The losses of the generator and the discriminator networks of the different stages of the framework after completion of the training.

	Stage 1	Stage 2
Generator	0.996912	0.240168
Discriminator	1.101999	0.000026

Figure 4: Training loss curves of the GAN architectures for the (a) synthesis stage; (b) upscaling stage.

Figure 5: Sample images generated from the first GAN.

Figure 6: Results of upsampling of the (a) synthesized images and (b) downsampled ground truth images were done by the ESRGAN.

3.1 Ablation study

An ablation study was done to select the number of residual-in-residual dense blocks. Table 2 summarizes the results from the study and Figure 7 shows the qualitative effect of the varying number of RRDBs.

Table 2: The losses of the generator and the discriminator networks of the different stages of the framework after completion of the training.

	1 st	2 nd	3 rd	4 th
Number of RRDBs	15	20	23	25
Generator parameter count	25.2M	33.5M	38.5M	41.8M
Criterion pixel loss	0.0328	0.0267	0.0222	0.0225

Figure 7: A qualitative analysis of the effect of different number of RRDBs during the upscaling stage. (a), (g) Ground truths; (b), (h) down-sampled inputs to the network; (c), (i) 1st; (d), (j) 2nd; (e), (k) 3rd; (f), (l) 4th network settings.

3.2 Perceptual quality analysis

Perceptual quality analysis was done by crowdsourcing responses to ensure that the image synthesized in the first stage is as good as the images that were used to train the second stage. This was done because there is no standard metric for perceptual quality that can be used to validate the fact that the second GAN was indeed trained to upscale the images that are generated by the first.

A set of 20 images of size 80x80 was prepared that contains 10 images generated from the first GAN and 10 images that were down-sampled from the MIAS dataset and was used for training the second GAN. This was shown to 22 respondents. The respondents were shown five separate real mammography images and then were asked to find out which of the images in the image set were fake. Out of 22 respondents, 18 incorrectly classified more than 75% of the samples. This shows that the synthesized images and the real images were indeed perceptually similar.

4. CONCLUSION AND FUTURE SCOPE

In this paper, we propose a multi-stage GAN based framework to generate mammography images. The framework initially generates synthetic images and then upscales it. This makes it more stable than a single GAN that generates a high-resolution image. The proposed method can also be used for the generation of other types of images as well. During training, we have noticed the presence of box artifacting that is caused by the deconvolution operations. Future work is needed to address this instead of using naïve noise removal methods.

ACKNOWLEDGEMENTS

The research work is partially supported by the RUSA 2.0 funded project "Developing Computational Systems Biology Approaches Applying Multimodal Networks and Advanced Machine Learning Techniques" and the Department of Computer Science and Engineering, Jadavpur University, Kolkata.

REFERENCES

- BCSC. Retrieved February 13, 2020, from https://www.bcs-research.org/data/mammography_dataset
- Birdwell, R.L., Ikeda, D.M., O'Shaughnessy, K.F. and Sickles, E.A., 2001. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology*, 219(1), pp.192-202.
- Blau, Y., Mechrez, R., Timofte, R., Michaeli, T. and Zelnik-Manor, L., 2018. The 2018 PIRM challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 0-0).
- Bozek, J., Mustra, M., Delac, K. and Grgic, M., 2009. A survey of image processing algorithms in digital mammography. In *Recent advances in multimedia signal processing and communications* (pp. 631-657). Springer, Berlin, Heidelberg.
- Breast Research Group. Retrieved February 13, 2020, from http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_1Nbreast_Database

- Broeders, M., Moss, S., Nyström, L., Njor, S., Jonsson, H., Paap, E., Massat, N., Duffy, S., Lynge, E. and Paci, E., 2012. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *Journal of medical screening*, 19(1_suppl), pp.14-25.
- Christoyianni, I., Koutras, A., Dermatas, E. and Kokkinakis, G., 2002. Computer aided diagnosis of breast cancer in digitized mammograms. *Computerized medical imaging and graphics*, 26(5), pp.309-319.
- Elter, M., Schulz-Wendtland, R. and Wittenberg, T., 2007. The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. *Medical physics*, 34(11), pp.4164-4172.
- Ferlay, J., Héry, C., Autier, P. and Sankaranarayanan, R., 2010. Global burden of breast cancer. In *Breast cancer epidemiology* (pp. 1-19). Springer, New York, NY.
- Hupse, R., Samulski, M., Lobbes, M., Den Heeten, A., Imhof-Tas, M.W., Beijerinck, D., Pijnappel, R., Boetes, C. and Karssemeijer, N., 2013. Standalone computer-aided detection compared to radiologists' performance for the detection of mammographic masses. *European radiology*, 23(1), pp.93-100.
- Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M. and Rubin, D.L., 2017. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data*, 4, p.170177.
- Lehman, C.D., Wellman, R.D., Buist, D.S., Kerlikowske, K., Tosteson, A.N. and Miglioretti, D.L., 2015. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA internal medicine*, 175(11), pp.1828-1837.
- Radford, A., Metz, L. and Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
- Ribli, D., Horváth, A., Unger, Z., Pollner, P. and Csabai, I., 2018. Detecting and classifying lesions in mammograms with deep learning. *Scientific reports*, 8(1), pp.1-7.
- Suckling J, P., 1994. The mammographic image analysis society digital mammogram database. *Digital Mammo*, pp.375-386.
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y. and Change Loy, C., 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 0-0).
- Yi, X., Walia, E. and Babyn, P., 2019. Generative adversarial network in medical imaging: A review. *Medical image analysis*, p.101552.