Computers in Biology and Medicine 119 (2020) 103698

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

Improving breast mass classification by shared data with domain transformation using a generative adversarial network

Chisako Muramatsu [a,*], Mizuho Nishio [b], Takuma Goto [c], Mikinao Oiwa [d], Takako Morita [e], Masahiro Yakami [b], Takeshi Kubo [f], Kaori Togashi [f], Hiroshi Fujita [g]

[a] Faculty of Data Science, Shiga University, 1-1-1 Banba, Hikone, Shiga, 522-8522, Japan
[b] Preemptive Medicine and Lifestyle-Related Disease Research Center, Kyoto University Hospital, 53 Shogoin Kawaharacho, Sakyo-ku, Kyoto, 606-8507, Japan
[c] Department of Intelligence Engineering, Graduate School of Natural Science and Technology, Gifu University, 1-1 Yanagido, Gifu, 501-1194, Japan
[d] Department of Radiology, Nagoya Medical Center, 4-1-1 Sannomaru, Naka-ku, Nagoya, Aichi, 460-0001, Japan
[e] Department of Breast Surgery, Nagoya Medical Center, 4-1-1 Sannomaru, Naka-ku, Nagoya, Aichi, 460-0001, Japan
[f] Department of Diagnostic Imaging and Nuclear Medicine, Kyoto University Graduate School of Medicine, 54 Shogoin Kawaharacho, Sakyo-ku, Kyoto, 606-8507, Japan
[g] Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Gifu University, 1-1 Yanagido, Gifu, 501-1194, Japan

ARTICLE INFO

Keywords:
Mammography
Classification
Deep learning
Generative adversarial network
ROC curve

ABSTRACT

Training of a convolutional neural network (CNN) generally requires a large dataset. However, it is not easy to collect a large medical image dataset. The purpose of this study is to investigate the utility of synthetic images in training CNNs and to demonstrate the applicability of unrelated images by domain transformation. Mammograms showing 202 benign and 212 malignant masses were used for evaluation. To create synthetic data, a cycle generative adversarial network was trained with 599 lung nodules in computed tomography (CT) and 1430 breast masses on digitized mammograms (DDSM). A CNN was trained for classification between benign and malignant masses. The classification performance was compared between the networks trained with the original data, augmented data, synthetic data, DDSM images, and natural images (ImageNet dataset). The results were evaluated in terms of the classification accuracy and the area under the receiver operating characteristic curves (AUC). The classification accuracy improved from 65.7% to 67.1% with data augmentation. The use of an ImageNet pretrained model was useful (79.2%). Performance was slightly improved when synthetic images or the DDSM images only were used for pretraining (67.6 and 72.5%, respectively). When the ImageNet pretrained model was trained with the synthetic images, the classification performance slightly improved (81.4%), although the difference in AUCs was not statistically significant. The use of the synthetic images had an effect similar to the DDSM images. The results of the proposed study indicated that the synthetic data generated from unrelated lesions by domain transformation could be used to increase the training samples.

1. Introduction

Breast cancer is the most frequently diagnosed cancer and one of the leading causes of cancer deaths in women in many developed countries [1]. Efficient and accurate image diagnosis is important for screening program success, and early detection of cancers by periodic screening may improve mortality rate and patient outcome. Computer analysis of medical images can assist radiologists in achieving such goals.

Studies have reported the potential of deep learning techniques in medical image analysis for various tasks. A general data shortage is one of the obstacles in successful training of deep networks for medical images. In such cases, a model trained with natural images and data augmentation techniques are used for improving performance [2–4]. However, the general augmentation methods, including rotation, contrast transformation, and noise addition, can only increase samples with characteristics that are similar to the original ones. Although the use of pretrained models is reported to be beneficial, it may be more natural to use medical images rather than unrelated images [5,6].

Another way to compensate for small sample size is to use simulated samples. Recent studies reported the use of generative adversarial networks (GANs) to synthesize samples in order to increase training samples. Salehinejad et al. [7] proposed the use of a deep convolutional GAN to synthesize chest X-ray images with and without pathologic conditions to balance the number of training samples in different pathologic

* Corresponding author.
E-mail address: Chisako-muramatsu@biwako.shiga-u.ac.jp (C. Muramatsu).

https://doi.org/10.1016/j.compbiomed.2020.103698
Received 11 November 2019; Received in revised form 8 March 2020; Accepted 8 March 2020
Available online 10 March 2020

classes. Senaras et al. [8] proposed the image generation method for pathologic slides using a conditional GAN [9]. By providing ground truth label images of nuclei as an input, the trained GAN outputs synthetic images, which may be used for teaching purposes and for training and evaluation of algorithms. Frid-Adar et al. [10] used a GAN to synthesize liver lesions on CT, which were used for data augmentation in the classification of cysts, metastases, and hemangiomas. For breast lesions, Guan and Loew [11] investigated the usefulness of GAN-generated mammograms for data augmentation in a classification network of normal and abnormal mammograms. Al-Dhabyani et al. [12] synthesized breast ultrasound images using a GAN and compared the performance of classification between normal, benign, and malignant images using the traditional augmentation and GAN augmentation without and with transfer learning.

Costa et al. [13] synthesized retinal fundus images from synthetic retinal vessel tree images using a GAN. The synthetic vessel images were created by an adversarial autoencoder, and the synthetic pairs were employed as training data for a segmentation network. Gadermayr et al. [14] proposed the use of a cycle GAN [15] to translate images with healthy conditions to fake images with affected conditions for augmenting labeled segmentation data. Similarly, Cai, et al. [16] used the cycle GAN to synthesize CT from magnetic resonance imaging (MRI) and MRI from CT to improve the segmentation network. Becker et al. [17] used the cycle GAN to inject malignant features on mammograms and evaluated the detectability of cancer on real and synthetic images and the ability of radiologists to discriminate between real and fake images.

One of the drawbacks of GAN-generated images is image quality. When images are synthesized from random noise vectors as in the general GAN, checkerboard-type artifacts are often apparent from upsampling [7,10,11,17]. When using the cycle GAN or ones that take an image input, such as the autoencoder and pix2pix [15], such an artifact is less apparent. However, proper input images are required such as vessel tree images. In addition, when lesions are created from scratch, GAN tends to generate similar lesions (mode collapse). To solve these problems, in this study, we investigated the possibility of sharing lesion images for augmenting training samples. Tumor-like lesions in various organs have some common characteristics, such as a blob-like shape and spicularity. In this study, we synthesized mammograms with breast masses from CT images with lung nodules using the cycle GAN and examined whether synthetic images can help in improving classification accuracy when the sample size is small. To our knowledge, no one has used the cycle GAN to synthesize lesions from different organs for data augmentation. Our goals are to examine the possibilities of generating lesion images with better quality and larger variety without requiring paired lesion data and to investigate the utility of these images in training a CNN model.

2. Materials and methods

2.1. Digital mammography dataset

Institutional review board (IRB) approval was obtained for use in this retrospective study. Research information was released on the Nagoya Medical Center's website, and the requirement for patient informed consent was waived.

A digital mammography dataset was employed for training and evaluation of the classification network. Mammograms acquired between December 2006 and Jan 2011 were collected from a single institution. They were acquired using four imaging units, including a phase-contrast mammography system (Konica Minolta Holdings, Inc.), a direct conversion mammography system (Fujifilm Corporation), and computed radiography systems (Siemens with Konica Minolta or Fujifilm imaging plates). Regions of interest (ROIs), including masses, were obtained from previous studies [18,19], in which malignant lesions with the following diagnostic results were included: invasive ductal carcinomas, ductal carcinomas in situ, invasive lobular carcinomas, and mucinous carcinomas. For benign lesions, cysts, fibroadenomas and benign phyllodes tumors were included. Malignant cases were confirmed by biopsy or surgery, whereas benign cases were confirmed by biopsy or follow-up using mammography and ultrasound. As a result, the database consisted of 414 ROIs, including mass lesions obtained from 247 patients. Because of the small dataset, training samples were augmented by rotation with an interval of 15° and a horizontal flip. However, if the rotated image included areas outside of the mammographic field of view, these samples were excluded. The classification performance was evaluated by a 3-fold cross validation. Benign and malignant cases were randomly split to 3 groups, and one was used for training, another was used for validation and the remaining was employed for testing. The process was repeated for three groups. The number of samples is listed in Table 1.

2.2. Lung nodule dataset

The lung nodule dataset was used for training a cycle GAN and for generating synthetic data. The CT data were originally collected retrospectively for a different research project. An IRB approval was obtained for use in this study. The research information was released on Kyoto University's website and the patients' informed consent was waived. The images were acquired from 2009 to 2015 using Aquilion ONE or Aquilion 64 (Toshiba Medical Systems). A single most clinically important nodule per case was included in the database. In this study, an ROI at the central slice of each nodule was employed. In the previous study, the cases were split into training, validation, and test datasets for the purpose of developing a classification algorithm. In this study, a training dataset was used for training the GAN, and the validation dataset was employed for image generation. The number of samples is listed in Table 2.

2.3. Digitized mammography database

Public databases can also be used to supplement a small dataset. An open-access dataset, the Digital Database for Screening Mammography (DDSM) [20], was employed in this study for two purposes: to pretrain the classification network to compare the result with that of the synthesized data; and to train the GAN. For pretraining of the classification network, the cases were randomly split into training and validation datasets. For the cycle GAN, 125 test lesions were semirandomly selected on the basis of the Breast Imaging Reporting and Data System (BI-RADS) shape and margin categories. The remaining cases were employed to pretrain the network. The numbers of samples are listed in Table 3.

2.4. Image generation by cycle GAN

When using a general GAN, images are synthesized from random noise vectors. The characteristics of the generated images are difficult to controll unless some conditions are given. In this study, a cycle GAN was employed for the generation of breast mass and lung nodule images by

Table 1
Number of samples for training, validation, and testing in 3-fold cross-validation groups for breast mass classification.

		Training set			Validation set			Test set	
		Case	ROI	Aug	Case	ROI	Aug	Case	ROI
CV1	B	40	64	2808	34	64	2792	45	74
	M	42	71	2848	43	71	2856	43	70
CV2	B	45	74	2830	40	64	2808	34	64
	M	43	70	2848	42	71	2848	43	71
CV3	B	34	64	2792	45	74	2830	40	64
	M	43	71	2856	43	70	2848	42	71

*B: benign, M: malignant, Aug: augmented by rotation and flipping.

Table 2
Number of lung nodule samples for training the cycle GAN and for generating synthetic images.

	Training			Image generation		
	Primary	Metastasis	Benign	Primary	Metastasis	Benign
Number of ROIs	278	131	190	91	43	64

Table 3
Number of the DDSM samples for pretraining the classification network and for training the cycle GAN.

	Training		Validation	
	Benign	Malignant	Benign	Malignant
Number of ROIs for pretraining	750	653	84	68
Number of ROIs for cycle GAN	765	665	–	–

transforming the images from each other so that the generated lesion image inherited the characteristics of the input lesion to some extent. This may make it easier to synthesize images with the desired characteristics. The schematic diagram is shown in Fig. 1.

The details of the network architecture were specified in the original paper [15]. Briefly, the network has two generators each consisting of two 3x3 down-sampling convolution layers with 64 and 128 filters, 9 resnet blocks which consist of two 3x3 convolutional layers with 128 filters, and two 3x3 up-sampling convolution layers with 64 and 32 filters. Two discriminators consist of three 4x4 convolution layers with stride 2 and 64, 128, and 256 filters; a 4x4 convolution layer with 512 filters; and an output convolutional layer. The input and output image size for the generators was 256 x 256 pixels. The deep learning framework used was PyTorch. The network optimizer was Adam and the loss function consisted of the GAN loss and the cycle consistency loss as specified in the original paper [15]. Batch size was 1 and the number of epochs was 200.

A radiologist with 25 years of experience who specializes in breast imaging visually evaluated the generated mass images. The radiologist was shown 20 real mass images and 40 synthetic mass images in a randomized order and was asked to rate whether they looked real using a 10-point scale, where 1 and 10 corresponded with definitely fake and real, respectively. Receiver operating characteristic (ROC) analysis was used to evaluate discrimination ability.

2.5. Breast mass classification

A CNN was trained for classification between benign and malignant breast masses. In this study, a 50-layer residual network (ResNet) [21] was used. The input images were resized to 224 x 224 pixels. The deep learning framework was Keras with TensorFlow backend. For this network, the pretrained model using the ImageNet [22] dataset was provided. The classification performances in terms of accuracy and area under the ROC curve (AUC) were compared for the following training strategies as shown in Fig. 2: (a) CNN trained with the original dataset, (b) CNN trained with the augmented dataset, (c) ImageNet pretrained model fine-tuned with the augmented dataset, (d) CNN pretrained with the synthesized dataset and fine-tuned with the augmented dataset, (e) ImageNet pretrained model fine-tuned with the mixed data of augmented and synthesized images, (f) CNN pretrained with DDSM and fine-tuned with the augmented dataset, and (g) ImageNet pretrained model fine-tuned trained with the DDSM and fine-tuned with the augmented data. The statistical test was performed for the AUCs using ROC software (University of Chicago).

3. Results

3.1. Generation of mass images

The cycle GAN was trained using 1430 breast masses and 599 lung nodules. After the training, the fake mass images were generated using the 198 nodules (Table 2). Fig. 3 shows real mass images, the original nodule images and the synthetic mass images. Many of the synthetic images resembled the real mass images. When the chest wall was present, the network tried to mimic breast tissue overlap. However, the rib regions looked somewhat unnatural in the synthetic mammograms. The fake masses, in general, had the shape and margin characteristic of the original nodules.

The radiologist evaluated the generated images shown in Fig. 3. The average ratings for real and synthetic data were 5.9 and 4.9, respectively. The accuracy in discriminating between real and synthetic images in terms of AUC was 0.743, which is not very high but better than a random guess.

The synthesized images were used for pretraining the classification network. Because of the small sample size, generated images were augmented by rotation with an interval of 90° and horizontal flipping. The number of pretraining synthetic mass samples was 1584, which was comparable to the pretraining DDSM samples (1403), although the latter were all original samples.

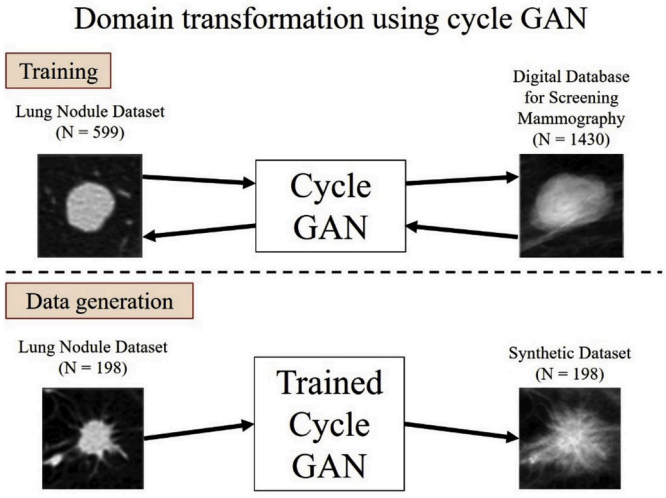

Fig. 1. Schematic diagram of domain transformation using a cycle GAN.

Table 4
Validation and test accuracies and the AUC for classification of benign and malignant breast masses.

	Validation accuracy	Test sensitivity	Test specificity	Test accuracy	Test AUC	p-value vs (a)	p-value vs (b)
(a)	67.9% (281/414)	72.8% (147/202)	59.0% (125/212)	65.7% (272/414)	0.724	–	–
(b)	70.8% (293/414)	86.6% (175/202)	48.6% (103/212)	67.1% (278/414)	0.741	0.4	–
(c)	80.5% (333/414)	81.2% (164/202)	77.4% (164/212)	79.2% (328/414)	0.869	<0.0001	<0.0001
(d)	69.5% (288/414)	89.1% (180/202)	47.2% (100/212)	67.6% (280/414)	0.739	0.2	0.6
(e)	81.3% (337/414)	86.1% (174/202)	76.9% (163/212)	81.4% (337/414)	0.884	<0.0001	<0.0001
(f)	69.6% (288/414)	75.2% (152/202)	69.8% (148/212)	72.5% (300/414)	0.796	0.0005	0.004
(g)	82.3% (341/414)	83.2% (168/202)	78.8% (167/212)	80.9% (335/414)	0.886	<0.0001	<0.0001

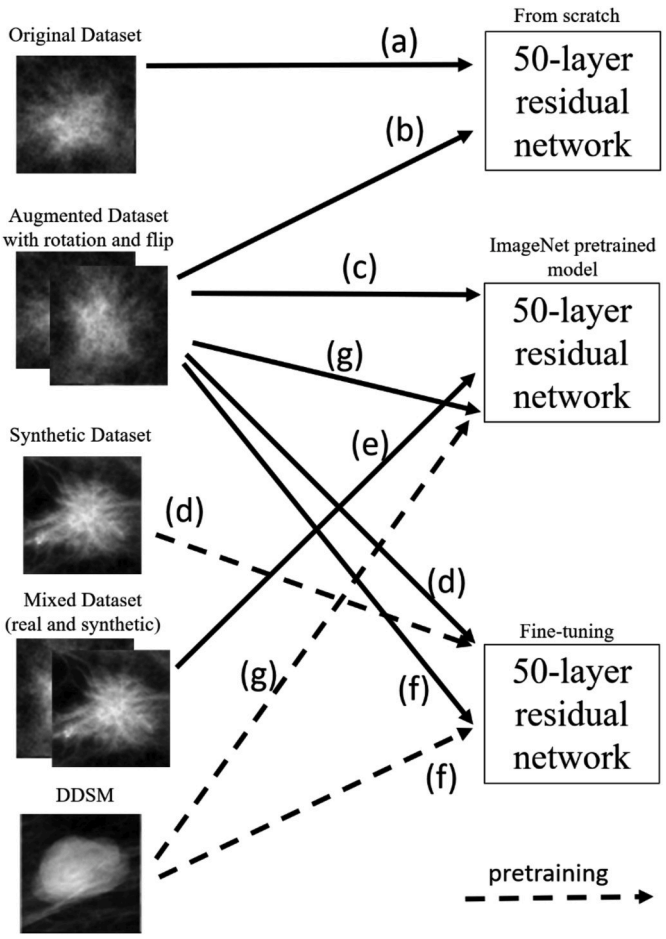

Fig. 2. Training strategies for classification network.

3.2. Classification performance

The classification accuracy and the AUC by the network trained with the original data from scratch were 65.7% (272/414) and 0.724, respectively. When the training data were augmented by rotation and flipping, the performance improved to 67.1% (278/414) and 0.741, respectively. The difference in AUC was not statistically significant ($p = 0.4$). The training curves are shown in Fig. 4. When using the original samples only for training, both the training and validation accuracies were low and unstable. When the data augmentation technique was used, the training loss became stable, but the validation loss was still unstable.

Using the synthetic masses for pretraining, the accuracy was 67.6% (280/424), and the AUC was 0.739. Compared with the result of no pretraining, the test accuracy and AUC were comparable. There was no statistically significant difference in AUCs ($p = 0.25$). When the ImageNet pretrained model was used, the accuracy and AUC were slightly improved from 79.2% (328/414) to 81.4% (337/414) and 0.869 to 0.884, respectively, by training with the synthetic images. The training curves are shown in Fig. 5. The validation loss became more stable using the pretrained model.

When the DDSM cases were used for pretraining, the accuracy and the AUC were 72.5% (300/414) and 0.796, respectively. They were improved to 80.9% (335/414) and 0.886, respectively, using the ImageNet pretrained model. Compared with the result of the regular data augmentation, the results obtained using the ImageNet pretrained model were all statistically significant. The validation and test accuracies as well as the test AUC and the corresponding p-values are listed in Table 4. The ROC curves are shown in Fig. 6. The effects of using the synthetic data and the public data were comparable with the pretrained model.

4. Discussion

We investigated the use of synthetic images in training the CNN for the classification of mammographic masses in the case of limited training samples. For creating realistic samples with controlled characteristics, a cycle GAN was employed for synthesizing the breast mass images from CT lung nodule images. Using this method, it may be possible to share samples of mass-like lesions in different organs. Furthermore, it is possible to use simulated images with various shapes and textures for generating realistic samples that are similar to the study cases by domain transformation. The advantage of using the cycle GAN is that it does not require paired ground truth images. Therefore, it was possible to synthesize images of lesions in the breast from lesions in the lung.

In the classification of benign and malignant masses, regular data augmentation was useful. Classification performance was improved by pretraining the network with natural images and training with the synthesized images. The results showed statistically significant improvement against regular augmentation, although we failed to find the statistically significant difference between without and with synthetic data. The result could be partly due to the small test size. The use of synthetic data had an effect similar to the use of the digitized mammography data for pretraining. When there is no public dataset available, the proposed method could be an alternative way to provide training samples.

In this study, the classification performance decreased when using the synthetic data only for pretraining. There could be two reasons for this result. Some synthetic images are not very similar to the masses with the corresponding labels. In a study by Costa et al. [13], synthetic images used for training the model were selected based on image quality. In this study, all generated images were used. It may be appropriate to select synthetic images with satisfactory quality.

In this study, the synthetic masses were labeled based on the pathologic labels of the original nodules. We believe that in general, output images from the cycle GAN inherit the characteristics of the input images, as shown in Fig. 3. To more explicitly generate benign and malignant masses, two separate cycle GANs were trained for benign and malignant lesions. The generated images were very similar to those by one cycle GAN, as shown in Fig. 7. The classification results for (d) and (e) were 72.5 (300/414) and 80.9 (335/414), respectively, and AUCs were 0.814 and 0.892, respectively. The result was comparable using the ImageNet pretrained model, but the result improved when the pretrained model was not used compared to results using one cycle GAN. The results were statistically significant (<0.0001) compared to that of the regular augmentation. In both cases, the nodule dataset includes metastatic lesions, some of which may share characteristics that are similar to those in the benign masses, such as the round shape and clear margin. “Truth” labeling of the synthetic images must be further investigated so that originally unlabeled data and simulated nonlesion images can also be used. In addition, in this study, we employed the digitized mammographic data as the teacher of the cycle GAN. These images are very old. However, the DDSM is the largest publicly available database and is still useful. The quality of the synthetic images may improve if the same digital mammographic data are used.

Another reason for the decrease in classification performance when using the synthetic data only for pretraining is the small number of synthetic samples. We only generated 198 synthetic masses in this study. Even with the data augmentation, the number of training samples was about 1500. When the DDSM samples were augmented to more than 60,000 samples by rotation and horizontal flipping, the classification result was comparable to that without use of the ImageNet pretrained model. It is possible that increasing synthetic samples has a similar effect. For creating more samples, providing more input data and/or

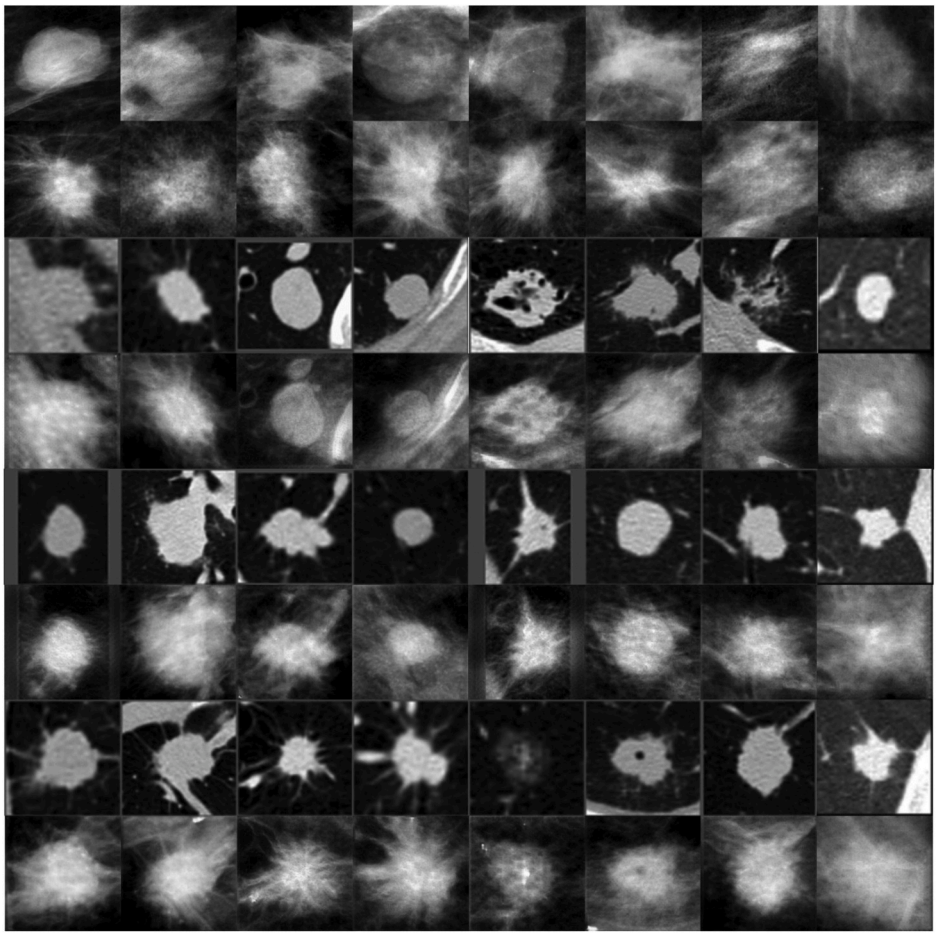

Fig. 3. Real and synthetic images. From first to last rows: real benign masses, real malignant masses, real benign lung nodules, synthetic masses generated from the prior row, real metastatic nodules, synthetic masses generated from the prior row, real primary cancers, and synthetic masses generated from the prior row.

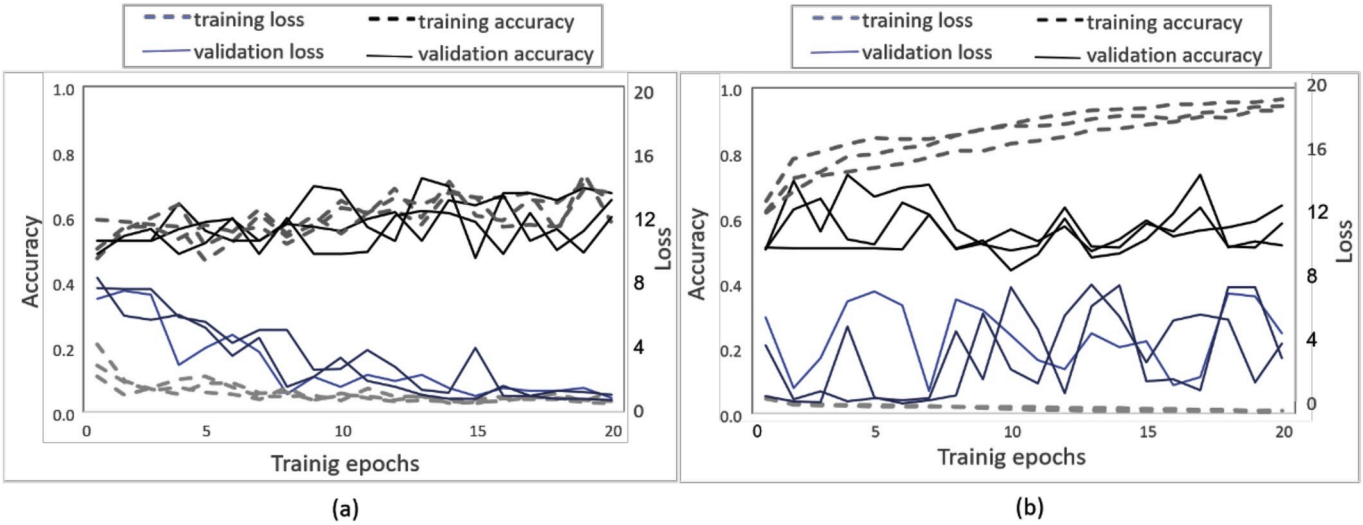

Fig. 4. Training curves for (a) original dataset and (b) augmented dataset.

adding variations through the network are necessary.

In this study, we hypothesized that tumor-like lesions in different organs have common characteristics, such as shape and spicularity. In fact, Fig. 3 shows that the shape and spicularity of lung nodules are retained in the generated breast mass when domain transformation is performed using GAN. In addition, Table 4 shows that the generated breast mass was useful for improving the accuracy and AUC of CNN for the classification of benign and malignant breast masses. These results support our hypothesis, and it is speculated that synthetic datasets obtained using GAN can be useful for a pair of tumors not only in the lung and breast but also in other organs.

Several studies have investigated the usefulness of GAN-generated

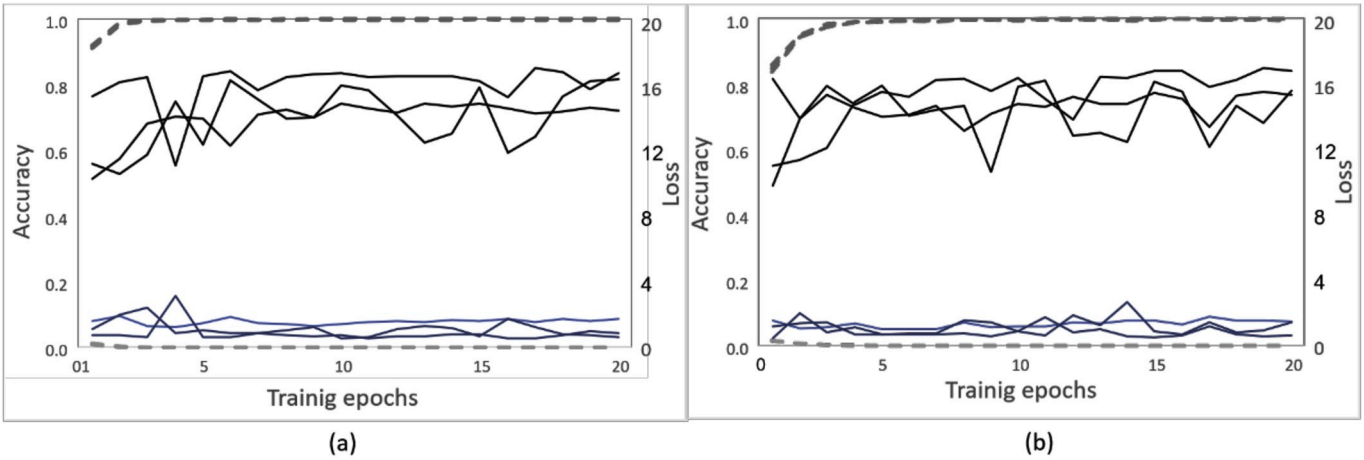

Fig. 5. Training curves for fine-tuning using (a) ImageNet pretrained model and (b) ImageNet pretrained model trained with the synthetic and real mix data.

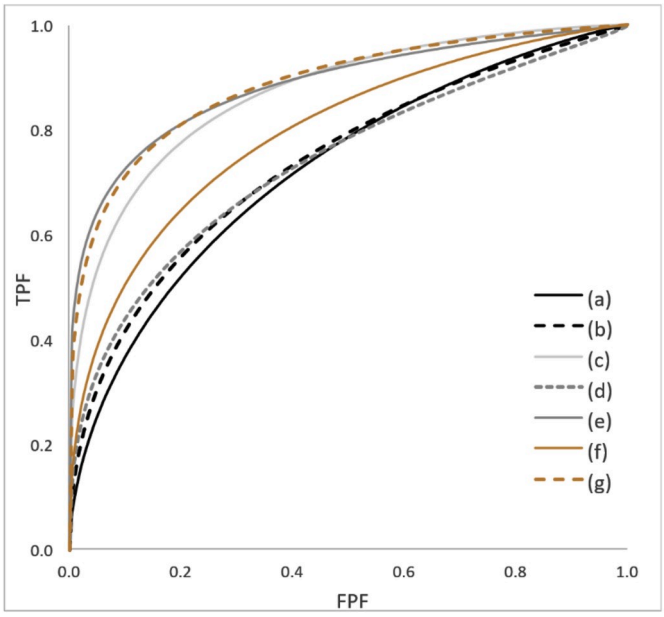

Fig. 6. ROC curves for the results in Table 4.

images for increasing training data of CNN models. Salehinejad et al. [7] generated chest radiographs with pathologies to mitigate the unbalanced dataset. The classification accuracy improved from 71% to 92%, although the images apparently had low resolution. Frid-Adar et al. [10] generated the images of liver lesions, including cysts, metastases, and hemangiomas. When using synthetic images for data augmentation, the classification accuracy improved from 63% with the regular augmentation to 86%. The synthetic images were evaluated by radiologists. They reported about 60% accuracy for 182 real and 120 fake images. However, no detail is given on how many synthetic images were assessed as real or fake. For mammography, Guan and Loew [11] synthesized normal and abnormal mammograms. By comparing the CNNs trained with the original data, augmented data with affine transformation, and GAN-generated data, there was a slight improvement in accuracy from 73.5% to 75.0% by adding GAN-generated data. However, affine transformation was not useful compared with the original data alone. Becker et al. [17] generated mammograms with and without lesions using a cycle GAN. They did not use the synthetic images for data augmentation but instead assessed the discrimination ability of the radiologists. In this study, a similar result was obtained: the synthetic images are useful for increasing the training data. The results cannot be directly compared because the image modalities and tasks investigated are different. Our study is unique in that we employed the cycle GAN to transfer lung nodules in CT to breast masses on mammograms to solve the image quality and mode collapse problems. Such domain transformation may be applied to other types of lesions and image modalities.

This study had several limitations. One was a small dataset size. The numbers of test samples as well as the synthetic samples were small, which may have been a reason for the insignificant result. Another limitation was that we only tested the classification performance of the breast masses. To validate our hypothesis, utility of domain transferred images in other various tasks must be investigated. The third limitation

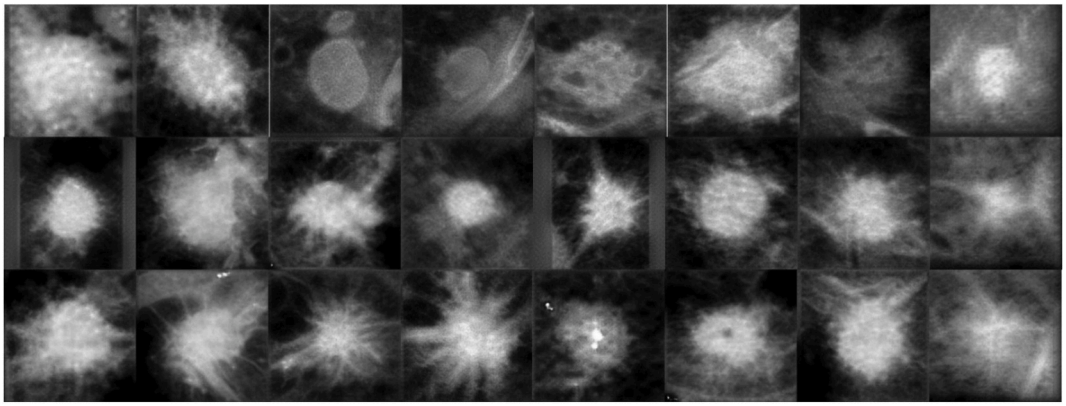

Fig. 7. Synthetic images generated by two cycle GANs separately trained with benign and malignant lesions.

was the lack of evaluation of the synthetic images. Most of the images generated in this study looked similar to the breast mass images. However, some images included artifacts from ribs and nonmass-like lesions generated from nodule with air cavites. Network optimization or exclusion of undesired samples may be necessary.

This study indicated that synthetic images generated by lesions in different organs by domain transformation could be used to share data between different tasks in order to increase CNN training samples when it is difficult to collect clinical data. Further investigation with a larger dataset and for different tasks is needed.

Funding

This study was supported in part by a Grant-in-Aid for Scientific Research (C) JSPS KAKENHI Grant Numbers JP17K09061 and JP16K19883, a Grant-in-Aid for Scientific Research (B) Grant Number JP19H035599, and a Grant-in-Aid for Scientific Research on Innovative Areas, MEXT KAKENHI Grant Number JP26108005.

Declaration of competing interest

None declared.

Acknowledgements

The authors would like to thank Enago (www.enago.jp) for the English language review.

References

[1] D. Forman, J. Ferlay, The global and regional burden of cancer, in: B.W. Stewart, C. P. Wild (Eds.), World Cancer Report 2014, the International Agency for Research on Cancer, Lyon, France, 2014, pp. 16–53.
[2] Z. Gao, L. Wang, L. Zhou, J. Zhang, HEp-2 cell image classification with deep convolutional neural networks, IEEE J. Biomed. Health. Inform. 21 (2016) 416–428.
[3] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, J. Colin, S.J. van Riel, M.M.W. Wille, M. Naqibullah, C.I. Sanchez, B. van Ginneken, Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks, IEEE Trans. Med. Imag. 35 (2016) 1160–1169.
[4] D.H. Kim, T. MacKinnon, Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks, Clin. Radiol. 73 (2018) 439–445.
[5] R.K. Samala, H.P. Chan, L. Hadjiiski, M.A. Helvie, J. Wei, C. Kenny, Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography, Med. Phys. 43 (2016) 6654–6666.
[6] S. Pang, Z. Yu, M.A. Orgun, A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images, Comput. Methods Progr. Biomed. 140 (2017) 283–293.
[7] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, J. Barfett. Generalization of Deep Neural Networks for Chest Pathology Classification in X-Rays Using Generative Adversarial Networks. arXiv:1712.01636.
[8] C. Senaras, M.K.K. Niazi, B. Sahiner, M.P. Pennell, G. Tozbikian, G. Lozanski, M. N. Gurcan, Optimized generation of high-resolution phantom images using cGAN: application to quantification of Ki67 breast cancer images, PloS One 13 (2018), e0196846.
[9] T.C. Wang, M.Y. Liu, J.Y. Zhu, A. Tao, J. Kautz, B. Catanzaro. High-resolution Image Synthesis and Semantic Manipulation with Conditional GANs. arXiv: 1711.11585.
[10] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, Neurocomputing 321 (2018) 321–331.
[11] S. Guan, M. Loew, Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks, J. Med. Img. 6 (2019), 031411.
[12] W. Al-Dhabyani, A. Fahmy, M. Gomaa, H. Khaled, Deep learning approaches for data augmentation and classification of breast masses using ultrasound images, Int. J. Adv. Comput. Sci. Appl. 10 (2019) 618–627.
[13] P. Costa, A. Galdran, M.I. Meyer, M. Neimeijer, M. Abramoff, A.M. Mendonca, A. Campilho, End-to-end adversarial retinal image synthesis, IEEE Trans. Med. Imag. 37 (2018) 781–791.
[14] M. Gadermayr, K. Li, M. Muller, D. Truhn, N. Kramer, D. Merhof, B. Gess, Domain-specific data augmentation for segmenting MR images of fatty infiltrated human thighs with neural networks, J. Magn. Reson. Imag. 49 (2019) 1676–1683.
[15] J.Y. Zhu, T. Park, P. Isola, A.A. Efros. Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. arXiv:1703.10593.
[16] J. Cai, Z. Zhang, L. Cui, Y. Zheng, L. Yang, Towards cross-modal organ translation and segmentation: a cycle- and shape-consistent generative adversarial network, Med. Image Anal. 52 (2019) 174–184.
[17] A.S. Becker, L. Jendele, O. Skopek, N. Berger, S. Ghafoor, M. Marcon, E. Konukoglu. Injecting and Removing Malignant Features in Mammography with cycleGAN: Investigation of an Automated Adversarial Attack Using Neural Networks. arXiv:1811.07767.
[18] C. Muramatsu, K. Nishimura, T. Endo, M. Oiwa, M. Shiraiwa, K. Doi, H. Fujita, Representation of lesion similarity by use of multidimensional scaling for breast masses on mammograms, J. Digit. Imag. 26 (2013) 740–747.
[19] C. Muramatsu, T. Hara, T. Endo, H. Fujita, Breast mass classification on mammograms using radial local ternary patterns, Comput. Biol. Med. 72 (2016) 43–53.
[20] M. Heath, K. Bowyer, D. Kopans, R. Moore, P. Kegelmey, Current States of the Digital Database for Screening Mammography. Digital Mammography, Kluwer Academic, Dordrecht, 1998.
[21] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. arXiv;1512.03385.
[22] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, IEEE Comput. Vis. Pat. Recog (CVPR) (2009) 248–255.